



**Universitat**  
de les Illes Balears

**DOCTORAL THESIS**  
**2024**

**Theoretical and data-driven models in  
Ecology**

**Àlex Giménez Romero**





**Universitat**  
de les Illes Balears

# DOCTORAL THESIS 2024

Doctoral programme in Physics

## Theoretical and data-driven models in Ecology

Àlex Giménez Romero

**Thesis Supervisor:** Manuel A. Matías  
**Thesis Tutor:** Cristóbal López Sánchez

Doctor by the Universitat de les Illes Balears

**Supervisors:**  
Manuel Matías

Àlex Giménez Romero.  
*Theoretical and data-driven models in Ecology.* ©  
Palma de Mallorca, July 2024

A en Manuel Miranda  
pel seu suport i ajuda  
durant tots aquests anys.  
Sempre estaràs amb mi.  
i recordare sempre  
el que em vas ensenyar.



Dr Manuel A. Matías of the Consejo Superior de Investigaciones Científicas (CSIC)

I DECLARE:

That the thesis title *Theoretical and data-driven models in Ecology*, presented by Àlex Giménez Romero to obtain a doctoral degree, has been completed under my supervision and meets the requirements to opt for an International Doctorate.

For all intents and purposes, I hereby sign this document.

Signature

Dr. Manuel A. Matías  
Thesis Supervisor

Palma de Mallorca, July 2024



## **Acknowledgements**

Acknowledgements go here.



## **Resum**

El resum va aquí.

## **Resumen**

El resumen va aquí.

## **Abstract**

Abstract goes here.



# Contents

I

## Introduction

<b>1</b>	<b>The global biodiversity crisis</b>	0
1.1	Introduction	0
1.2	The Mass Mortality Event of <i>Pinna nobilis</i>	0
1.3	<i>Xylella fastidiosa</i> : an emerging global threat	0
1.4	The global decline of seagrasses: the case of <i>Posidonia oceanica</i>	0
1.5	Coral reefs: a global crisis	0

<b>2</b>	<b>Models in Ecology</b>	1
----------	--------------------------	---

2.1	Introduction	1
2.2	Why do we need models?	1
2.3	Theoretical modelling	1
2.3.1	Dynamical systems	1
2.3.2	Compartmental models	1
2.3.3	Individual-based models	1
2.3.4	Reaction-difussion models	1
2.4	Data-driven modelling	1
2.4.1	Machine learning	1
2.4.2	Climate data	1
2.4.3	Species Distribution Models	1

II

## Modelling parasite-produced marine diseases

<b>3</b>	<b>The case of the mass mortality event of <i>Pinna nobilis</i></b>	7
----------	---	---

3.1	Introduction	8
3.2	The SIRP model	10
3.2.1	Model structure and initial considerations	10
3.2.2	General SIRP model	11
3.2.3	Model reduction	13

<b>3.3</b>	<b>Numerical analysis of the model</b>	<b>15</b>
3.3.1	The basic reproduction number $R_0$	15
3.3.2	Final state of the epidemic	16
3.3.3	Maximum of infected individuals	17
3.3.4	Numerical verification of the fast-slow approximation	18
3.3.5	Numerical verification of the model approximation from the exact reduction	19
<b>3.4</b>	<b>Model validation with data of the <i>Pinna nobilis</i> Mass Mortality Event</b>	<b>19</b>
<b>3.5</b>	<b>Conclusions</b>	<b>23</b>
<b>3.6</b>	<b>Appendix</b>	<b>24</b>
3.6.1	Finding a conserved quantity for the SIRP model	24
3.6.2	Stability analysis of the fixed points of the SIRP model	25
3.6.3	Calculation of $R_0$ using the Next Generation Matrix method	25
3.6.4	Sensitivity Analysis	26
3.6.5	General rate change with temperature	27
<b>4</b>	<b>Spatial effects in parasite-induced marine diseases</b>	<b>29</b>
<b>4.1</b>	<b>Introduction</b>	<b>30</b>
<b>4.2</b>	<b>The SIRP spatial model</b>	<b>31</b>
<b>4.3</b>	<b>Results</b>	<b>33</b>
4.3.1	Non-spatial limit	33
4.3.2	Approximate relation between parasites and infected hosts	34
4.3.3	Spatial threshold	36
4.3.4	Spreading speed of the infected population and time to extinction	37
<b>4.4</b>	<b>Conclusions</b>	<b>39</b>
<b>4.5</b>	<b>Appendix</b>	<b>40</b>
4.5.1	Derivation of the non-spatial equation for $R_\infty$	40

### III

## Modelling vector-borne plant diseases

<b>5</b>	<b>Non-stationary vector populations</b>	<b>45</b>
<b>5.1</b>	<b>Introduction</b>	<b>46</b>
<b>5.2</b>	<b>The model</b>	<b>47</b>
5.2.1	Preliminary analysis of the model	48
<b>5.3</b>	<b>Results</b>	<b>49</b>
5.3.1	The effect of non-stationary vector populations into the epidemic threshold and disease dynamics	49
5.3.2	The basic reproduction number for non-stationary vector populations	52
5.3.3	Fast-slow approximation	53
5.3.4	Reduction to a SIR model	55
<b>5.4</b>	<b>Conclusions</b>	<b>55</b>
<b>5.5</b>	<b>Appendix</b>	<b>56</b>
5.5.1	Calculation of $R_0$ from standard methods	56
5.5.2	Calculation of $R_0$ for non-stationary vector populations	58

<b>6</b>	<b>A compartmental model for <i>Xylella fastidiosa</i> diseases</b>	<b>59</b>
----------	---	-----------

<b>6.1</b>	<b>Introduction</b>	<b>60</b>
<b>6.2</b>	<b>Materials and Methods</b>	<b>61</b>
6.2.1	Epidemic model: the SEIR-V model	61
6.2.2	Basic reproductive number	62
6.2.3	Epidemiological data	63

6.2.4	Model fitting through Bayesian Inference . . . . .	63
6.2.5	Sensitivity Analysis . . . . .	63
<b>6.3</b>	<b>Results</b> . . . . .	<b>64</b>
6.3.1	Model fit and parameter estimates . . . . .	64
6.3.2	Global Sensitivity Analysis . . . . .	68
6.3.3	Epidemic control through vector management . . . . .	69
<b>6.4</b>	<b>Discussion</b> . . . . .	<b>70</b>
<b>6.5</b>	<b>Appendix</b> . . . . .	<b>72</b>
6.5.1	Vector population dynamics . . . . .	72
6.5.2	Determination of $R_0$ . . . . .	72

## IV

## Modelling the risk of vector-borne plant diseases

<b>7</b>	<b>Global predictions for the risk of establishment of Pierce's disease of grapevines</b>	<b>77</b>
7.1	Introduction . . . . .	78
7.2	Results . . . . .	79
7.2.1	Thermal requirements to develop PD . . . . .	79
7.2.2	MGDD/CDD distribution maps . . . . .	81
7.2.3	PD global risk . . . . .	82
7.2.4	PD risk projections for 2050 . . . . .	85
7.2.5	Risk based on vector information . . . . .	86
7.2.6	Combining vineyard land cover across Europe with the model output . . . . .	86
<b>7.3</b>	<b>Discussion</b> . . . . .	<b>87</b>
<b>7.4</b>	<b>Methods</b> . . . . .	<b>89</b>
7.4.1	Inoculation tests . . . . .	89
7.4.2	Modified Growing Degree Days . . . . .	89
7.4.3	Disease progress with temperature . . . . .	90
7.4.4	Disease recovery through winter curing . . . . .	90
7.4.5	Global climate data, MGDD/CDD computation . . . . .	91
7.4.6	Disease Model . . . . .	91
7.4.7	Distribution of wine-grape production areas . . . . .	93
7.4.8	<i>Philaenus spumarius</i> SDM . . . . .	93
7.4.9	Risk assessment by 2050 . . . . .	93
<b>8</b>	<b>Pierce's disease epidemic risk under global warming</b>	<b>95</b>
<b>8.1</b>	<b>Introduction</b> . . . . .	<b>96</b>
<b>8.2</b>	<b>Methods</b> . . . . .	<b>97</b>
8.2.1	Climate datasets . . . . .	97
8.2.2	Climate-driven epidemiological model . . . . .	97
8.2.3	Model adaptation to daily temperature data . . . . .	99
8.2.4	Vector climatic suitability . . . . .	99
8.2.5	Risk velocity . . . . .	100
8.2.6	Maps . . . . .	100
<b>8.3</b>	<b>Results</b> . . . . .	<b>100</b>
8.3.1	Present and future climate suitability for <i>Xylella fastidiosa</i> (PD) and <i>Philaenus spumarius</i> . . . . .	100
8.3.2	Pierce's disease risk projections under climate change . . . . .	101
<b>8.4</b>	<b>Discussion</b> . . . . .	<b>105</b>

---

<b>9 Pierce's Disease risk with high-resolution climate data .....</b>	<b>109</b>
<b>9.1 Introduction .....</b>	<b>110</b>
<b>9.2 Results .....</b>	<b>111</b>
9.2.1 Global differences in PD risk between coarse and fine-grain climate data .....	111
9.2.2 Pierce's Disease risk surges in previously unresolved microclimates .....	114
<b>9.3 Discussion .....</b>	<b>116</b>
<b>9.4 Methods .....</b>	<b>117</b>
9.4.1 Climate data .....	117
9.4.2 Vector climatic suitability .....	117
9.4.3 Vineyard data .....	117
9.4.4 Climate-driven epidemiological model .....	118
9.4.5 Model adaptation to daily temperature data .....	119

**V****Data-driven methods for global ecological problems**

<b>10 Universal spatial properties of coral reefs .....</b>	<b>123</b>
<b>10.1 Introduction .....</b>	<b>124</b>
<b>10.2 Results .....</b>	<b>124</b>
10.2.1 Coral reefs macroecological patterns .....	124
10.2.2 The fractal nature of coral reefs .....	126
10.2.3 Fractality extends up to coral provinces .....	126
<b>10.3 Discussion .....</b>	<b>129</b>
<b>10.4 Methods .....</b>	<b>130</b>
10.4.1 Global coral reef data .....	130
10.4.2 Coral reefs as clusters of connected coral/algae class polygons .....	130
10.4.3 Coral reefs area, perimeter and inter-reef distance .....	130
10.4.4 Coral reef size distribution .....	131
10.4.5 Fractal dimensions from area-perimeter relation .....	131
10.4.6 Box-Counting fractal dimension .....	131
10.4.7 Compactness and elongation index .....	131
<b>11 Reconstructing coastal pH time-series with machine learning .....</b>	<b>133</b>
<b>11.1 Introduction .....</b>	<b>134</b>
<b>11.2 Results .....</b>	<b>136</b>
11.2.1 Time series data .....	136
11.2.2 Reconstruction pH time series with Deep Learning .....	138
<b>11.3 Discussion .....</b>	<b>139</b>
<b>11.4 Methods .....</b>	<b>141</b>
11.4.1 Data collection .....	141
11.4.2 Data processing .....	142
11.4.3 Computing the trend of seasonal data .....	142
11.4.4 Selecting the best neural network architecture .....	142
<b>12 Mapping the distribution of seagrass meadows from space with deep convolutional neural networks .....</b>	<b>145</b>

## List of publications

1. Àlex Giménez-Romero et al. "Modelling parasite-produced marine diseases: The case of the mass mortality event of *Pinna nobilis*". In: *Ecological Modelling* 459 (2021), page 109705. ISSN: 0304-3800. DOI: <https://doi.org/10.1016/j.ecolmodel.2021.109705>. URL: <https://www.sciencedirect.com/science/article/pii/S030438002100260X>
2. Susana Flecha et al. "pH trends and seasonal cycle in the coastal Balearic Sea reconstructed through machine learning". In: *Scientific Reports* 12.1 (July 2022), page 12956. ISSN: 2045-2322. DOI: [10.1038/s41598-022-17253-5](https://doi.org/10.1038/s41598-022-17253-5). URL: <https://doi.org/10.1038/s41598-022-17253-5>
3. Àlex Giménez-Romero et al. "Spatial effects in parasite-induced marine diseases of immobile hosts". In: *Royal Society Open Science* 9.8 (2022), page 212023. DOI: [10.1098/rsos.212023](https://doi.org/10.1098/rsos.212023). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.212023>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsos.212023>
4. Àlex Giménez-Romero, Rosa Flaquer-Galmés, and Manuel A. Matías. "Vector-borne diseases with nonstationary vector populations: The case of growing and decaying populations". In: *Phys. Rev. E* 106 (5 Nov. 2022), page 054402. DOI: [10.1103/PhysRevE.106.054402](https://doi.org/10.1103/PhysRevE.106.054402). URL: <https://link.aps.org/doi/10.1103/PhysRevE.106.054402>
5. Alex Giménez-Romero et al. "Global predictions for the risk of establishment of Pierce's disease of grapevines". In: *Communications Biology* 5.1 (Dec. 2022), page 1389. ISSN: 2399-3642. DOI: [10.1038/s42003-022-04358-w](https://doi.org/10.1038/s42003-022-04358-w). URL: <https://doi.org/10.1038/s42003-022-04358-w>
6. Àlex Giménez-Romero, Eduardo Moralejo, and Manuel A. Matías. "A Compartmental Model for *Xylella fastidiosa* Diseases with Explicit Vector Seasonal Dynamics". In: *Phytopathology®* 113.9 (2023). PMID: 36774557, pages 1686–1696. DOI: [10.1094/PHYTO-11-22-0428-V](https://doi.org/10.1094/PHYTO-11-22-0428-V). eprint: <https://doi.org/10.1094/PHYTO-11-22-0428-V>. URL: <https://doi.org/10.1094/PHYTO-11-22-0428-V>
7. Àlex Giménez-Romero et al. "Contrasting Patterns of Pierce's Disease Risk in European Vineyards Under Global Warming". In: *bioRxiv* (2023). DOI: [10.1101/2023.07.17.549293](https://doi.org/10.1101/2023.07.17.549293). eprint: <https://www.biorxiv.org/content/early/2023/07/18/2023.07.17.549293.full.pdf>. URL: <https://www.biorxiv.org/content/early/2023/07/18/2023.07.17.549293>

8. Àlex Giménez-Romero, Eduardo Moralejo, and Manuel A Matias. "High-resolution climate data reveals increased risk of Pierce's Disease for grapevines worldwide". In: *bioRxiv* (2024). DOI: [10.1101/2024.03.06.583743](https://doi.org/10.1101/2024.03.06.583743). eprint: <https://www.biorxiv.org/content/early/2024/03/11/2024.03.06.583743.full.pdf>. URL: <https://www.biorxiv.org/content/early/2024/03/11/2024.03.06.583743>
9. Universal spatial properties of coral reefs
10. Mapping the distribution of seagrass meadows from space with deep convolutional neural networks

## Other publications

1. Clara Lago et al. "Degree-day-based model to predict egg hatching of *Philaenus spumarius* (Hemiptera: Aphrophoridae), the main vector of *Xylella fastidiosa* in Europe". In: *Environmental Entomology* 52.3 (Apr. 2023), pages 350–359. ISSN: 0046-225X. DOI: [10.1093/ee/nvad013](https://doi.org/10.1093/ee/nvad013). eprint: <https://academic.oup.com/ee/article-pdf/52/3/350/50615564/nvad013.pdf>. URL: <https://doi.org/10.1093/ee/nvad013>
2. Leaf susceptibility of Macaronesian laurel forest species to *Phytophthora ramorum*
3. A comprehensive dataset on global coral reefs size and geometry



# Introduction

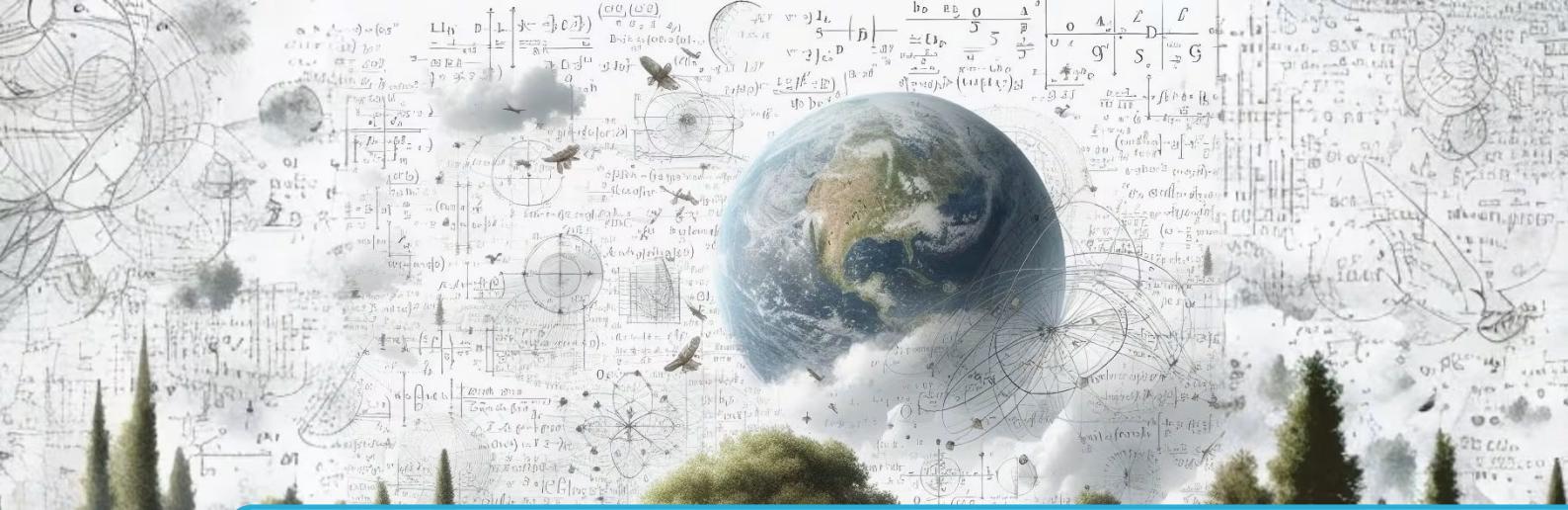
<b>1</b>	<b>The global biodiversity crisis .....</b>	<b>0</b>
1.1	Introduction .....	0
1.2	The Mass Mortality Event of <i>Pinna nobilis</i> .....	0
1.3	<i>Xylella fastidiosa</i> : an emerging global threat .....	0
1.4	The global decline of seagrasses: the case of <i>Posidonia oceanica</i> 0	0
1.5	Coral reefs: a global crisis .....	0
<b>2</b>	<b>Models in Ecology .....</b>	<b>1</b>
2.1	Introduction .....	1
2.2	Why do we need models? .....	1
2.3	Theoretical modelling .....	1
2.4	Data-driven modelling .....	1





## 1. The global biodiversity crisis

- 1.1 Introduction
- 1.2 The Mass Mortality Event of *Pinna nobilis*
- 1.3 *Xylella fastidiosa*: an emerging global threat
- 1.4 The global decline of seagrasses: the case of *Posidonia oceanica*
- 1.5 Coral reefs: a global crisis



## 2. Models in Ecology



### 2.1 Introduction

### 2.2 Why do we need models?

### 2.3 Theoretical modelling

#### 2.3.1 Dynamical systems

#### 2.3.2 Compartmental models

#### 2.3.3 Individual-based models

#### 2.3.4 Reaction-diffusion models

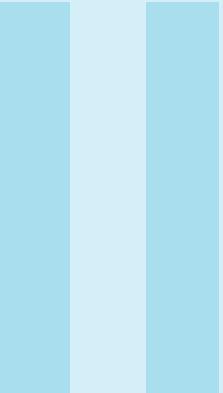
### 2.4 Data-driven modelling

#### 2.4.1 Machine learning

#### 2.4.2 Climate data

#### 2.4.3 Species Distribution Models





# Modelling parasite-produced marine diseases

<b>3</b>	<b>The case of the mass mortality event of <i>Pinna nobilis</i> . . . . .</b>	<b>7</b>
3.1	Introduction . . . . .	8
3.2	The SIRP model . . . . .	10
3.3	Numerical analysis of the model . . . . .	15
3.4	Model validation with data of the <i>Pinna nobilis</i> Mass Mortality Event . . . . .	19
3.5	Conclusions . . . . .	23
3.6	Appendix . . . . .	24
<b>4</b>	<b>Spatial effects in parasite-induced marine diseases . . . . .</b>	<b>29</b>
4.1	Introduction . . . . .	30
4.2	The SIRP spatial model . . . . .	31
4.3	Results . . . . .	33
4.4	Conclusions . . . . .	39
4.5	Appendix . . . . .	40



## Preface

### Objectives

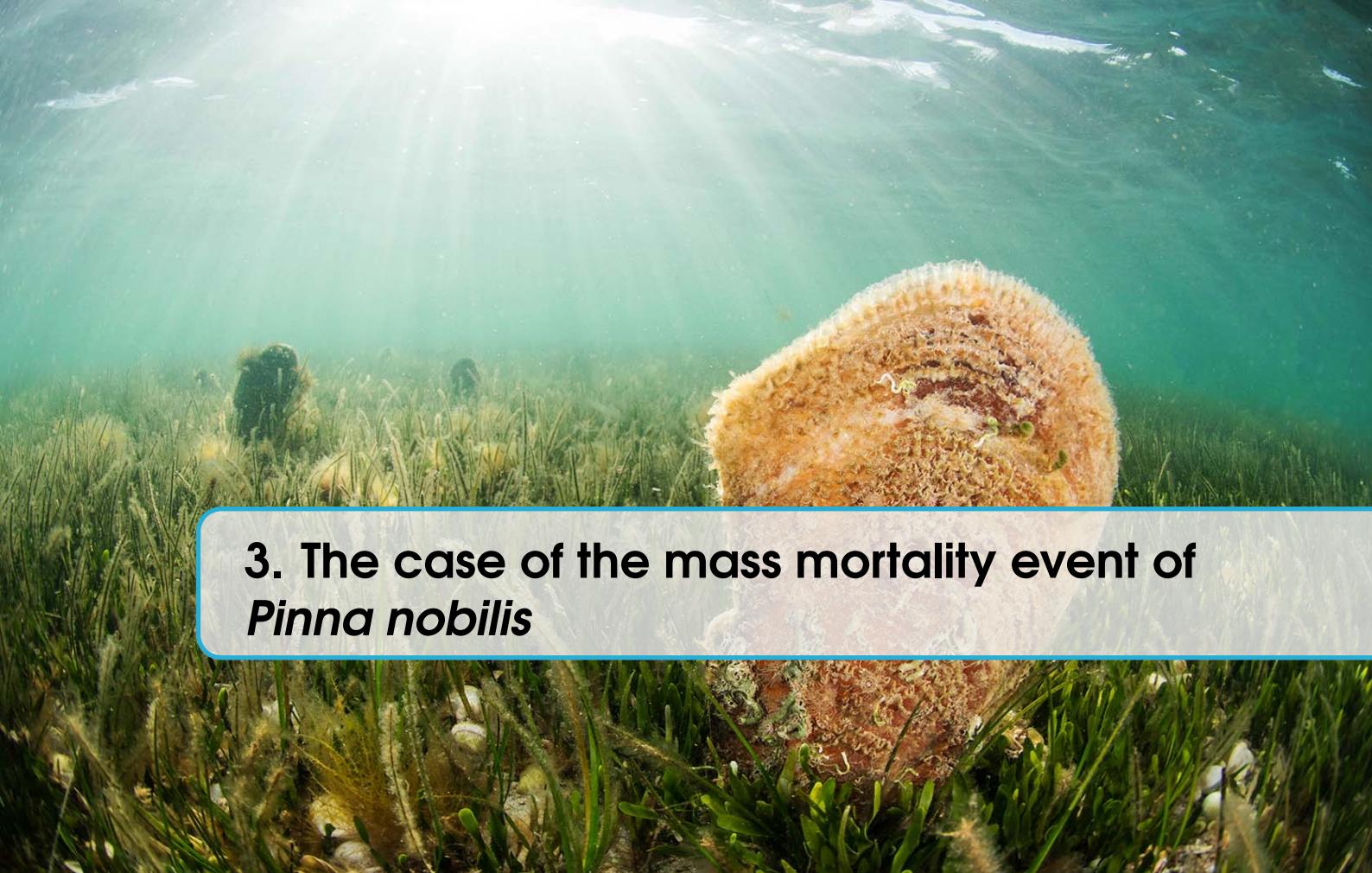
- First objective
- Second objective

### Contributions

- First objective
- Second objective

**Note:** This is a note.





### 3. The case of the mass mortality event of *Pinna nobilis*

Àlex Giménez-Romero<sup>1</sup>, Amalia Grau<sup>2,3</sup>, Iris E. Hendriks<sup>4</sup>, Manuel A. Matías<sup>1</sup>

1. Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Palma de Mallorca 07122, Spain
2. Laboratori d'Investigacions Marines i Aqüicultura (LIMIA), Govern de les Illes Balears, Av. Gabriel Roca 69 E-07158 Port d'Andratx, Mallorca, Spain
3. INAGEA (INIA-CAIB-UIB), Campus UIB, E-07122 Palma de Mallorca, Spain
4. Instituto Mediterráneo de Estudios Avanzados, IMEDEA (CSIC-UIB), E-07190 Esporles, Mallorca, Spain

#### Published as

Àlex Giménez-Romero et al. "Modelling parasite-produced marine diseases: The case of the mass mortality event of *Pinna nobilis*". In: *Ecological Modelling* 459 (2021), page 109705. ISSN: 0304-3800. DOI: <https://doi.org/10.1016/j.ecolmodel.2021.109705>. URL: <https://www.sciencedirect.com/science/article/pii/S030438002100260X>

### 3.1 Introduction

Marine organisms, like their terrestrial counterparts, can serve as hosts for a diversity of parasites and pathogens present in the ecosystem, which are directly responsible for disease outbreaks. Disease induced mortality affects not only the host population, but can cascade through the whole ecosystem, altering its structure and functioning (441). Furthermore, climate change can increase the spread range and impact of parasites and pathogens (72). In fact, marine infectious diseases are recently increasing due to climate change and other anthropogenic pressures, like pollution and overfishing (268). This, in turn, threatens many valuable ecologically habitats and can also result in substantial economic losses in e.g. aquaculture (269). Analysing the impact of these events at appropriate scales (spatial and temporal) and biological organisation levels (species, populations and communities) is crucial to accurately anticipate future changes in marine ecosystems and propose adapted management and conservation plans (342). Thus, there is a strong need to address the mechanism of disease propagation in marine organisms.

However, the state of the art of epidemiological studies in marine ecosystems lags behind that of terrestrial ecosystems (221). Contact and vector-borne based infectious diseases of terrestrial vertebrates and their epidemiology are typically studied using variations of the classical formulation of Kermack and McKendrick (260, 261, 262), the SIR model. Among other things, this formalism allows to understand why epizootics spread and stop, as the propagation of a disease is a threshold phenomenon (18), regulated by the now commonplace  $R_0$  dimensionless number. Within this framework the initiation of epidemic transmission occurs when an infected individual is in close contact with a susceptible host or through a transmission vector, as typically pathogens can only survive for a very limited time outside the host in an aerial environment. On the other hand, as air is typically a much harsher medium for pathogens than water, the sea is expected to host a large number of pathogens (viruses, bacteria and parasites) for a relatively long time. The longer life span of pathogens in a water medium, together with the increased buoyancy arising from the different physical properties of seawater and air, coupled to the existence of marine currents that can transmit pathogens for long distances away, allows diseases to spread faster and reach further distances in marine environments compared to epidemics in terrestrial systems (80). As a result, the possible long-term transmission of parasites by currents in marine environments make them more prone to suffer from persistent zoonotics compared to terrestrial ecosystems, where for an epidemic outbreak to occur the presence of an initial infected host (or vector) is necessary within a susceptible population. Until quite recently, marine zoonotics were mostly studied using different models compared to terrestrial diseases and it was not even clear whether the same tools could be applied (300). The abundance of pathogens in marine ecosystems is one of the reasons why proliferation models, that do not focus on transmission and assume a widespread occurrence of the pathogen and a rapid transmission problem, have been most popular in the field (354). In fact, compartmental models are starting to be used only recently in the study of marine epizootics (44).

An important subset of marine organisms are sessile, e.g. bivalves, which means that they can not move. In the case of sessile terrestrial organisms disease transmission occurs mostly through vectors, insects that transmit the pathogens causing the disease. Instead, in marine ecosystems disease transmission is most often waterborne, in particular in passive water filtering feeders, as is the case of bivalves. Recently, some compartmental models considering the pathogen population have been proposed to study particular bivalve epidemics (42, 43, 44). In the present work we analyse a model that is aimed to describe disease transmission from an infected immobile host to a susceptible one of the same species through waterborne parasites, that are explicitly described. The model is closely related to the SIP model introduced in (44). In this first study we analyse in detail the properties of the mean-field version of the model, that aims to describe spatially homogeneous (i.e. well mixed) populations. The well mixed approximation will be valid whenever the mean distance among hosts is smaller than the mixing length of the parasites before they get inactivated or absorbed. The model is written such that waterborne transmission is the only mechanism by which one infected immobile host can infect a healthy one, and, thus, does not describe infection through direct contact. It is also assumed that the infected hosts, as invertebrates, do not have immune memory, and that the probability that an infected individual recovers is small and can be neglected. Thus, the model is not adequate to study infection of highly aggregated molluscs (like some mussels) or other passive filters like corals,

as for these hosts one should also include the possibility of infection through direct contact. A first very relevant question is whether the model, describing infection of immobile (sessile) hosts through waterborne parasites can be reduced to a simpler version in which the parasite compartment is not needed. One exact and two approximate reductions are presented. We believe that the model can be most useful in the rapid characterisation of emergent marine epidemics if the right data from a well mixed system are available.

A very timely case study of such emerging epidemics is the noble fan mussel (or pen shell) *Pinna nobilis*. This fan mussel is the largest endemic bivalve in the Mediterranean Sea, and is under a serious extinction risk due to a Mass Mortality Event (MME) that has occurred throughout the whole Mediterranean basin very recently (175, 434, 459). Right before this MME, it was distributed across a wide type of habitats including coastal and paralic ecosystems at depths between 0.5 to 60 m (75, 355). In open coastal waters, the distribution of the species is mainly associated with seagrass meadows, typically of *Posidonia oceanica*, which has been indicated as its optimal habitat (227). Its lifespan is up to 50 years in favourable conditions and its size can get up to 1.2 m, placing it among the largest bivalves of the world (76). These fan mussels play a crucial ecological role in their habitat, as *P. nobilis* individuals filter water, thus retaining a large amount of organic matter from suspended detritus, contributing to water clarity (421). Furthermore, it is a habitat-forming species, because its shell provides a hard-surface within a soft bottom ecosystem, which can be colonised by different benthic species, augmenting biodiversity (76). In addition, at very dense populations, the species can function as an ecosystem engineer, creating biogenic reefs (255).

Despite *P. nobilis* populations have greatly declined due to anthropogenic activities in the 20th century (434), the ongoing MME is the most worrying and widespread threat to *P. nobilis* throughout the Mediterranean Sea. As a consequence, the species has been declared as critically endangered (239). Although different aetiological agents have been proposed, including Mycobacteria and other bacteria (81, 381, 385), there is evidence that the main cause of this mortality is the protozoan *Haplosporidium pinnae* (56, 92, 125), a new species that belongs to the genus *Haplosporidium*, one of the four genera of the protist order *Haplosporida*, where it has been found that other Haplosporidian parasites are behind the extensive mortality of several oyster species (22, 73). Life stages include uninucleate and binucleate cells, plasmodia, and spores. A group of experts following up the event predicted a high risk that the disease would be spread by marine currents through the Mediterranean basin, which could cause the extinction of the species (76) as it is endemic. This has helped to better understand the spread of the disease, and identified surface currents as the main factor influencing local dispersion, whereas environmental factors influence the disease expression, which seems to be favoured by temperatures above 13.5 °C and a salinity range between 36.5 and 39.7 psu.

In summary, we introduce and study in detail the properties of the mean-field version of a general compartmental model to study marine epidemics for bivalve populations, namely passive filtering sessile invertebrate hosts infected through waterborne parasites. There are two main hypotheses, the first one that a population level description (i.e. without the consideration of spatial effects) is able to describe well the dynamics of the epidemic in a relatively dense population in small bounded regions. A second assumption is that the host becomes infected with some probability, but that there is not a critical parasite load in the infection process. After presenting the full SIRP model, then three different reductions are discussed, one exact, an approximate reduction of the former and a third reduction based on a timescale approximation. The study is closed with a validation with the available experimental data for the infection process of *Pinna nobilis* kept in tanks. We wish to point out, that being a highly endangered and protected species, the reported data correspond to an *unintended experiment* that cannot be repeated, and maybe these data represent the only opportunity to estimate the fundamental parameters of the model. In addition, the setup in which the *Pinna nobilis* were kept in tanks, represent themselves the ideal implementation of the conditions under which the mean-field model SIRP is valid.

## 3.2 The SIRP model

### 3.2.1 Model structure and initial considerations

In this work we analyse the SIRP model, a deterministic multi-compartmental mean-field model, continuous in time and unstructured in spatial or age terms, to study infection in bivalve populations. In particular, we stress that the model as it is written describes spatially-homogeneous populations. Compartmental models are the most frequently used class of models in terrestrial epidemiology (139), and originated in the classic study of SIR models by Kermack and McKendrick (261). The use of compartmental models in the study of infectious processes in marine systems is quite rare until very recently (221). As already advanced in Section 3.1, there are relevant features in the description of epidemic processes in marine ecosystems that are different with respect to the case of terrestrial ecosystems (300), and their study in marine environments is dominated so far by so called proliferation models (354), which do not address the transmission of the pathogen. See (44) for a discussion of several compartment models for the study of marine epizootics.

Compartmental models of diseases in terrestrial ecosystems caused by microparasites (i.e. viruses, bacteria and protozoans) do not consider a compartment to describe the dynamics of the parasite (299), describing just the different stages of the host. Infection typically occurs in 2 ways: i) as a contact process, in which the microparasite is transmitted directly from a infected host, I, by contact or through air in close proximity, to a susceptible host, S; ii) through a vector, that has acquired the microparasite by biting an infected host, I, and passes the microparasite to a susceptible host, S. In the first case one can describe the infection process through some probability that the individuals come close, while in the second it is very relevant to describe the vector mobility, and at least 2 compartments, susceptible and infected vector, are typically needed. Once the microparasite enters the host, it proliferates inside it, so the infection process can be described by using compartments for susceptible individuals, S, infected individuals, I, and possible exposed individuals, E. In particular, transmission in terrestrial sessile organisms (e.g. plants) is generically vector-borne. In the case of marine ecosystems, infection typically occurs through water-borne parasites, in particular in filter-feeders sessile organisms, while vector-transmitted diseases are much less frequent. Parasites may be transported by diffusion, sea currents, or even active motion (i.e. if they have flagella). In any case, infection between sessile hosts is not through direct contact, but instead through the production and excretion of parasites by infected individuals and the assimilation by filtering of parasites by a healthy (susceptible) host. So, parasites are produced and excreted to the marine medium, in which they stay infective until they become deactivated (i.e die) or are absorbed by hosts. In a way, in parasite transmitted marine diseases parasites have a dual role: they are not only agents that induce infection but also act as vectors that transmit disease from an immobile infected host to a susceptible one.

The SIRP model is a general mean-field compartmental model to describe epidemic transmission through water-borne parasites, that we think is specially adequate to describe epidemic transmission in sparsely located passive filter feeders, like many bivalves. We exclude the case of colonies in which individuals are in close proximity, e.g. mussels, corals, etc, in which direct contact could be relevant and should be included in the model. In the SIRP model hosts are described through 3 different compartments, as in the SIR model, that represent different evolution stages of the disease: a susceptible class of healthy individuals that may contract the disease, S, an infected class of individuals that may pass the disease through excretion of the parasite, I, and a class of removed (namely dead) individuals, R, that cannot be infected any more and that cannot transmit the disease plus an extra compartment, P, for the parasite population in the medium. It is important to note that invertebrates do not develop long-term immunity in the mammalian sense (354), and so, no compartment of individuals “recovered with immunity” is considered. However, bivalves have a first line of defense with hemocytes being able to fight parasites and reduce their internal population. Nevertheless, available evidence indicates that the number of individuals that can achieve a full recovery is usually small and can be neglected, and so it is not necessary to consider a process in which individuals in the I compartment return to the S compartment at some rate, like in the SIS model. Instead, the population’s long-term response to disease, when it occurs, is through natural selection for genotypes characterised by increased resistance to or tolerance for the pathogen. As already advanced, the SIRP model

includes a fourth compartment that represents the parasite population in the water medium, whose population needs to be described explicitly. An explicit compartment allows to model the situation in which the population of parasites may evolve dynamically in time, although in Sec. 3.2.3.3 we will consider the case in which the parasite populations accommodates almost instantaneously to the infected host population, and the description of the parasite can be simplified.

Infection occurs when the host enters in contact with the parasite in the marine medium. It involves a process of filtration of water by the bivalve, and although a detailed representation has been discussed in the literature (42), in the current SIRP model it is represented in an effective way. Infection is modelled through a nonlinear term, typical in compartmental models, but that now depends on the parasite population and not on the population of the infected compartment,  $I$ . In terrestrial epidemiology there are two alternative ways to represent infection (297): i) mass action incidence, in which infection grows as the population gets larger,  $\beta SI$ ; ii) standard incidence, in which the infection is bounded as the population grows,  $\beta SI/N$ , where  $N$  is the total (host) population. One must look at these two choices as limit cases, with the possibility that in reality the system is best described by an intermediate form, closer to one of the limit cases, for example the modified SIR model in (61) in which the infection term has  $S + I$  in the denominator instead of the total population  $N$ , because the  $R$  compartment is removed. Modelling infection with an explicit representation of the parasite population encounters the same basic *dilemma* about whether the incidence grows as the (host) population increases or is bounded. We will model infection as  $\bar{\beta}SP$ , where the two possibilities are equivalent to the two different incidences just mentioned: i)  $\bar{\beta} = \beta$ ; ii)  $\bar{\beta} = \beta/N$ , where  $\beta$ , the disease transmission rate, is a *constant*, that can depend at most on external parameters, like temperature and salinity, but not on the variables defining the model (populations of host compartments or parasites). The model is valid considering both types of incidence, and in the case study we will see which incidence seems more adequate in this case.

### 3.2.2 General SIRP model

In this section we will write a mean-field compartmental model to describe epidemics of immobile (sessile) hosts in a marine medium through infection by a water-borne parasite. Being a mean-field model implies that the model is compartmental and does not include an explicit space dependence, and so, describes a well mixed system. The mean-field model describes a spatially homogeneous system, but we hope that it will be the basis for spatially inhomogeneous situations, by adding suitable terms accounting for the mobility of the parasite.

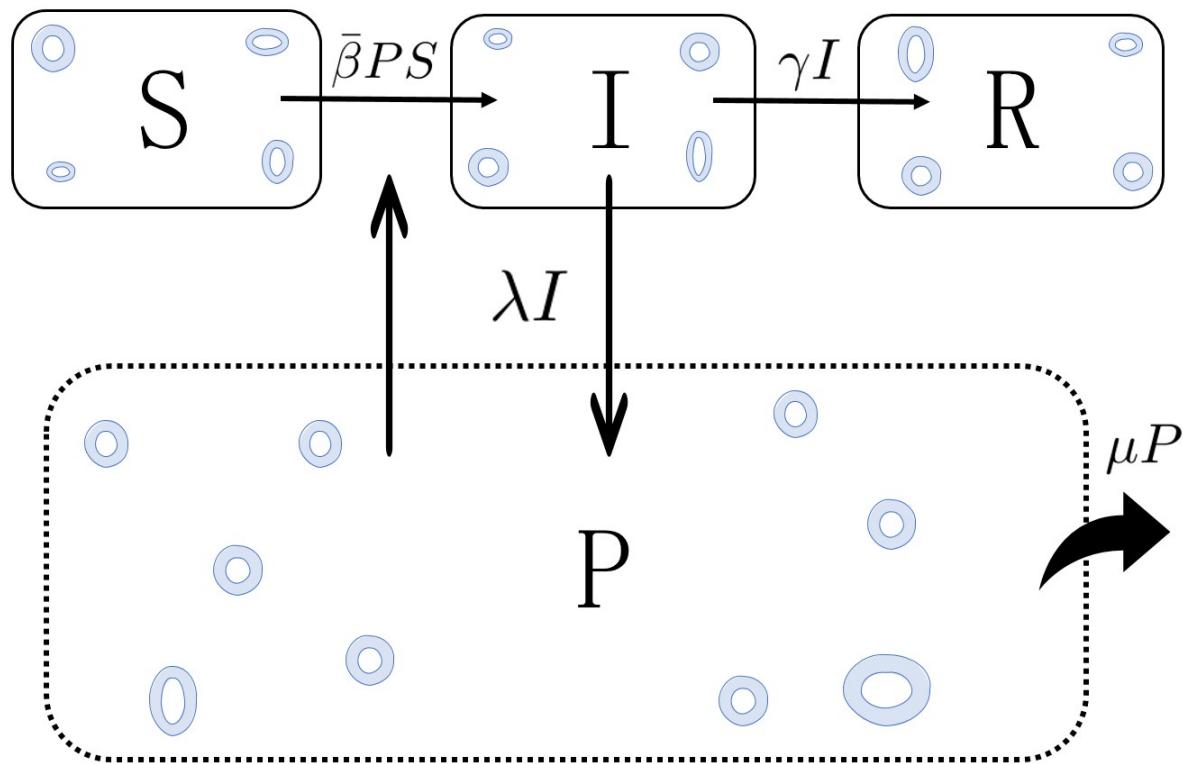
It is also assumed that the hosts become infected with some probability when exposed to a parasite, i.e., that there is not a critical parasite load needed for infection. The model is defined according to the following reaction processes,



which is graphically summarised in Fig. 3.1.

According to the scheme in Fig. 3.1, we consider the host in 3 possible states: susceptible,  $S$ , infected by the parasite,  $I$  and removed (dead),  $R$ . Then we introduce the parasite population in the medium (water),  $P$ . In the model, the  $\bar{\beta}$ , the disease transmission rate parameter regulates the infection rate of susceptible hosts and accounts, among other mechanisms, for the parasite intake rate,  $\gamma$  the mortality of infected hosts, being the inverse of the typical mean time for an infected host to die;  $\lambda$  the production rate of parasites from infected hosts, and  $\mu$  the inverse of the typical life time of the parasite.  $\mu$  can be related to several processes, like biological deactivation (or survival time) or other general losses, like dilution due to renewal of water in a closed experiment, natural losses in open ecosystems or absorption by other filter feeders. We are not considering the possibility of spontaneous parasite gain, i.e. immigration, in this version of the model. A summary of the model parameters can be found in Table 3.1. We do not consider vital dynamics for the hosts, and this implies that the sum of the 3 host subpopulations is constant,  $N = S + I + R$ , as the time scale of the disease evolution is much faster than the typical life cycle of fan mussels. The model is similar to the SIP presented in (44), except for an extra term in  $\dot{P}$ ,  $-\bar{\beta}PS$ , accounting for the fact that when a parasite infects a host it is absorbed by it. The conditions under which the SIRP model can be simplified to the SIP model are discussed in Section 3.3.5.

In order to build the deterministic model we consider that the population is large enough to neglect fluctuations and that it is well mixed, so that spatial effects can be neglected. In this situation, we consider the infection process to be proportional to the number of parasites in the medium, so that the average number of contacts between susceptible fan mussels and the average parasite population is given by  $\bar{P}S$ , and, thus, the change in the number of susceptible fan mussels takes the form  $\dot{S} = -\bar{\beta}PS$ , where the dot over a variable indicates a differentiation with respect to time:  $\dot{S} = dS/dt$ .



**Figure 3.1:** SIRP model flow diagram. The model variables are represented by capital letters: susceptible hosts ( $S$ ), infected hosts ( $I$ ), dead hosts ( $R$ ) and the population of parasite ( $P$ ). Arrows represent the processes in the model with their rates indicated next to them and blue rings represent parasites. The flow follows the scheme in Eq. (3.1), that leads to the system of differential equations in Eq. (3.2).

**Table 3.1:** Model parameters description

Variable	Definition	Parameter	Definition
$S$	Susceptible host	$\beta$	Disease transmission rate
$I$	Infected host	$\gamma$	Host mortality rate
$R$	Removed host	$\lambda$	Production rate of parasites by infected hosts
$P$	Parasite in the medium	$\mu$	Parasite deactivation/dilution rate

Following this argumentation, the scheme in Eq. (3.1) and Fig. 3.1, one can write the evolution

equations of the SIRP model,

$$\begin{aligned}\dot{S} &= -\bar{\beta}PS \\ \dot{I} &= \bar{\beta}PS - \gamma I \\ \dot{R} &= \gamma I \\ \dot{P} &= \lambda I - \bar{\beta}PS - \mu P.\end{aligned}\tag{3.2}$$

Model (Eq. (3.2)) lives in the 4-dimensional (S,I,R,P) phase space, representing the variables the populations of individuals in the susceptible, infected and removed host compartments and of parasites, respectively. These variables could be redefined so that S, I and R represent proportions of hosts in each compartment and P the population of parasites per host.

The fixed points of Eq. (3.2) are determined by the conditions<sup>1</sup>  $I = P = 0$ , to be fulfilled simultaneously. We will study the stability of the fixed point defined by  $S(0), I(0) = P(0) = 0$  and  $R(0) = N - S(0)$ . A linear stability analysis of this fixed point reveals that it has two null eigenvalues, that stem from the condition  $N = S + I + R$  and the conserved quantity of 3.6.1. The first condition,  $S + I + R = N$ , implies that it is enough to consider two of the host populations, e.g.  $S$  and  $I$ , as the third one can be obtained from the other two. The implications of the conserved quantity reported in 3.6.1 are more subtle, as it implies that fixed points are not isolated, as it happens in ordinary dissipative dynamical systems, and there is an infinite number (a line of) fixed points for the final state of the epidemic, depending on the initial conditions. This also implies that the phase space is foliated by the conserved quantity,  $C$  of Eq. (4.13), and every initial condition,  $S_0$ , with a different value of  $C$  leads to a different asymptotic condition,  $S_\infty$ , just as shown in (327) for the SIR model (cf. Fig. 10.1 in *op. cit.*). The third eigenvalue, that is the largest of the two non-zero eigenvalues, can be positive if  $\beta S_0 \lambda > \gamma(\beta S_0 + \mu)$  and negative if the inequality is reversed, defining the conditional stability of the fixed point. The fourth eigenvalue is always negative and all the eigenvalues are always real (cf. 3.6.2). The instability of the fixed point along the third eigenvalue drives the beginning of the epidemic.

An extremely important result in epidemiology is the so-called *basic reproduction number*,  $R_0$ , a dimensionless number which represents the number of secondary infections produced by a primary infection in a fully susceptible population.  $R_0 = 1$  defines the threshold for epidemic propagation: an epidemic will occur when  $R_0 > 1$ , and the number of infected individuals will grow, at an exponential rate in the early phases of the epidemic (91), while if  $R_0 < 1$  the infection will wane naturally. This quantity can be formally obtained making use of the Next Generation Matrix (NGM) method (138, 143). Applying this formal method to our system of ordinary differential equations (ODE's) one obtains the following relation for the basic reproduction number (cf. 3.6.3),

$$R_0 = \frac{\lambda}{\gamma \left( 1 + \frac{\mu}{\bar{\beta}S(0)} \right)}.\tag{3.3}$$

The threshold condition provided by  $R_0$  (Eq. (3.3)) is equivalent to the linear stability condition for the third eigenvalue of the initial, pre-epidemic, fixed point, as  $\bar{\beta}S(0)\lambda > \gamma(\beta S(0) + \mu)$  implies that this eigenvalue is positive and the disease-free equilibrium state unstable being this equivalent to  $R_0 > 1$  (cf. 3.6.2). Thus, if  $R_0 > 1$  the fixed point is unstable, and an epidemic will ensure if infected hosts,  $I$ , or parasites,  $P$  appear in the system. An epidemic will propagate until the system reaches an stable fixed point, that signals the end of the epidemic (cf. 3.6.2).

### 3.2.3 Model reduction

The SIRP model lives in a 4-dimensional phase space and depends on 4 parameters, what makes difficult to confront it with experimental data. Thus, we will discuss here three alternative ways of reducing the model. The first involves an exact reduction of the model, based on the conserved quantity derived in 3.6.1. The second reduction consists of an approximation to the previous exact reduction, that turns out to be equivalent to an exact reduction of a slightly simplified model (without the  $-\bar{\beta}SP$  term in the equation of  $\dot{P}$ ). The third one is based on an approximation valid if the system parameters fulfil certain conditions.

<sup>1</sup>We do not consider the trivial fixed point  $S = I = R = P = 0$  that would imply  $N = 0$  and  $P = 0$  at all time.

### 3.2.3.1 Exact reduction of the SIRP model

From the conserved quantity derived in 3.6.1, it is possible to write the parasite population in the SIRP model as a function of the host states as follows,

$$P(S, I) = -\frac{\lambda}{\gamma}(S + I) + \frac{\mu}{\bar{\beta}} \ln(S) + S + C(0), \quad (3.4)$$

where  $C(0) = P(0) + \frac{\lambda}{\gamma}(S(0) + I(0)) - \frac{\mu}{\bar{\beta}} \ln(S(0)) - S(0)$ .

Substituting Eq. (3.4) into the general SIRP model of Eq. (3.2) we obtain the following nonstandard SIR model,

$$\begin{aligned} \dot{S} &= \frac{\lambda \bar{\beta}}{\gamma} S(S + I) - \mu S \ln(S) - \bar{\beta} S^2 - S \bar{\beta} C(0) \\ \dot{I} &= -\frac{\lambda \bar{\beta}}{\gamma} S(S + I) + \mu S \ln(S) + \bar{\beta} S^2 + S \bar{\beta} C(0) - \gamma I \\ \dot{R} &= \gamma I. \end{aligned} \quad (3.5)$$

Although using the conserved quantity yields an exact reduction from a 4D dynamical system to a 3D one, the number of independent parameters and initial conditions remain unchanged, i.e. they still depend on 4 parameters and 4 initial conditions. Thus, although useful, (3.5) is not ideal when trying to fit experimental data, and this is the reason for trying a further approximation to Eq. (3.5) to be discussed next.

### 3.2.3.2 Further approximation to the exact reduction

A further approximation to Section 3.2.3.1, that is less restrictive and expected to be valid in a broader parameter range than Section 3.2.3.3 is possible. This approximation reduces the number of free parameters by one, what is useful in fitting available data. The approximation consists of neglecting the  $S$  term in Eq. (3.4), what is possible if  $\lambda/\gamma \gg 1$  and also  $\mu \ln N / (\bar{\beta} N) \gg 1$ , as  $S(t)$  decreases monotonically with time and is, at most,  $N$  at the initial time. Interestingly, this approximation is equivalent to the simplification of the equation for  $\dot{P}$  in (3.2) so that the  $-\bar{\beta}SP$  is skipped, what yields exactly the SIP model of Ref. (44). This reduced model has an exact conserved quantity,  $\mathcal{C}$ , that differs from that of the SIRP model in the linear  $S$  term (cf. 3.6.1). Using this approximation one can write,

$$\begin{aligned} \dot{S} &= \frac{\lambda'}{\gamma} S(S + I) - \mu S \ln(S) - S \bar{\mathcal{C}}(0) \\ \dot{I} &= -\frac{\lambda'}{\gamma} S(S + I) + \mu S \ln(S) + S \bar{\mathcal{C}}(0) - \gamma I \\ \dot{R} &= \gamma I, \end{aligned} \quad (3.6)$$

where  $\lambda' = \lambda \bar{\beta}$  and  $\bar{\mathcal{C}}(0) = \bar{\beta}(P(0) + \lambda/\gamma(S(0) + I(0)) - \mu/\bar{\beta} \ln S(0)) = \bar{\beta} \mathcal{C}(0)$  is a redefinition of the conserved quantity of the SIP model Eq. (3.18),  $\mathcal{C}(0)$ , a constant, such that it absorbs  $\bar{\beta}$  and all initial conditions of the model. The result is that Eq. (3.6) depends on 3 parameters and 1 constant, compared to Eq. (3.5) that depends on 4 parameters, facilitating, thus, the use of the model to fit experimental data.

### 3.2.3.3 Model reduction through fast-slow separation

The third reduction of the 4-D dynamical model Eq. (3.2) makes the assumption that the time scale in which the parasite population changes are faster than the one corresponding to the host. This means that pathogen deactivation in the medium must be faster than host mortality. In terms of the rates associated to each of these processes, this means  $\mu > \gamma$ . Taking  $\mu$  as common factor in  $\dot{P}$  one can write,

$$\varepsilon \dot{P} = \lambda I / \mu - \bar{\beta} SP / \mu - P \quad (3.7)$$

where  $\varepsilon = 1/\mu$  is small, as  $\mu$  is large. If furthermore  $\mu \gg \bar{\beta}N$  and  $\lambda \gg \beta P$  one arrives to,

$$P \approx \frac{\lambda}{\mu} I. \quad (3.8)$$

Under this approximation the slow subsystem can be written,

$$\begin{aligned}\dot{S} &= -\beta' IS \\ \dot{I} &= (\beta'S - \gamma)I \\ \dot{R} &= \gamma I,\end{aligned}\tag{3.9}$$

that is equivalent to the classical SIR model with  $\beta' = \bar{\beta}\lambda/\mu$  instead of the infection rate  $\bar{\beta}$ . The reduced 3-D model Eq. (3.9) from the original 4-D SIRP model Eq. (3.2) depends on 2 parameters instead of 4 as the original model had and 1 initial condition, e.g.  $I(0) = N - S(0)$  if  $R(0) = 0$ , and is much more amenable to be applied to the analysis of experimental data, as shown in Section 3.4. Furthermore,  $\gamma$  could be eliminated through a time rescaling,  $t \rightarrow t' = \gamma t$  with a redefinition of  $\beta' \rightarrow \beta'' = \beta'/\gamma = \beta\lambda/(\mu\gamma)$ , leaving the model as a function of a single effective parameter. However, we will keep both  $\beta'$  and  $\gamma$  for convenience when fitting the model to experimental data in Section 3.4, as we would need to know anyhow  $\gamma$  in order to analyse the experimental data. The validity of this approximation is checked numerically in Section 3.3.

### 3.3 Numerical analysis of the model

Due to the impossibility of solving the SIRP model analytically, in the present section we perform a numerical characterisation of the model<sup>2</sup>. Moreover, we show the validity range of the performed approximations to reduce the SIRP model to an effective SIR model. We start our numerical analysis by investigating the relative influence of the model parameters on some epidemiological quantities of interest: the basic reproduction number ( $R_0$ ), related to the existence of an epidemic outbreak, continuing with the final state of the epidemic, given by the final number of dead individuals ( $R(\infty)$ ) and the maximum of infected individuals ( $I_{\max}$ ) together with the time at which it occurs ( $t_{\max}$ ).

In order to identify the most influential parameters of our model Sensitivity Analysis (SA) will be performed. SA can be divided into two classes: Local Sensitivity Analysis (LSA) and Global Sensitivity Analysis (GSA). LSA represents the assessment of the local impact of input factors variation on model response by concentrating on the sensitivity in the vicinity of a set of factor values. Such sensitivity is often evaluated through gradients or partial derivatives of the output functions at these factor values, such that other input factors are kept constant. Since epidemic models exhibit a threshold behaviour, controlled by the dimensionless quantity  $R_0$ , it is relevant to study its robustness with respect to small perturbations by means of the LSA explained above, as its analytical expression is known.

On the other hand, GSA will be applied to study the influence of the parameters in the final state of the epidemic and the epidemic peak by exploring a large domain of the parameter space. In turn, GSA is the process of apportioning the uncertainty in outputs to the uncertainty in each input factor over their entire range of interest. A sensitivity analysis is considered to be global when all the input factors are varied simultaneously and the sensitivity is evaluated over the entire range of each input factor, in clear contrast to LSA. Within GSA, first order indices are a measure of the contribution to the output variance given by the variation of the parameter alone averaged over variations in other input parameters while second order indices take into account first order interactions between parameters. While LSA is carried out analytically (if exact expressions are available), GSA is a purely numerical approach. Further mathematical details on Sensitivity Analysis can be found in 3.6.4.

For all the sensitivity analysis performed in the following sections, and in order to avoid ambiguities associated to the definition of  $\bar{\beta}$  as a function of  $N$ , we assume  $N = 1$ , so that both possible incidences yield  $\bar{\beta} = \beta$  and the numerical results are equivalent.

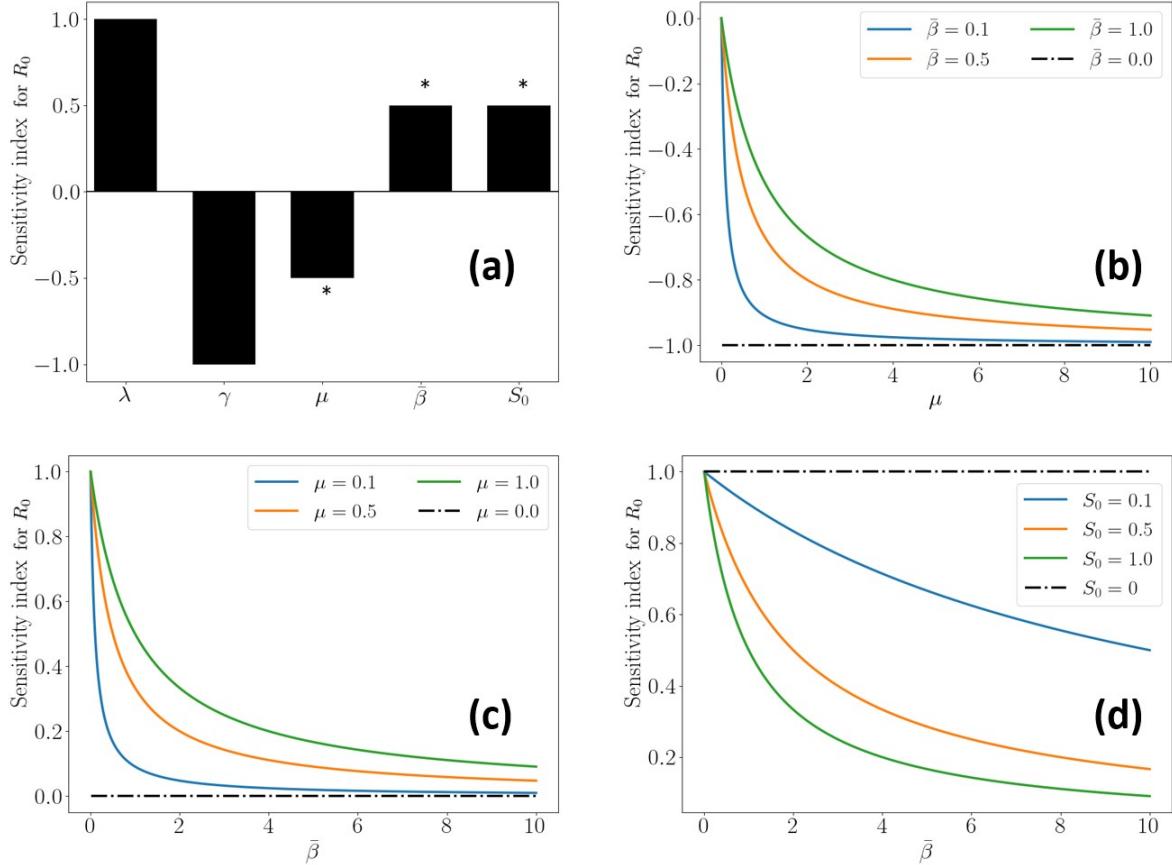
#### 3.3.1 The basic reproduction number $R_0$

To study the relevance of parameters involved in an epidemic outbreak a LSA was performed. We analyse the local sensitivity of  $R_0$  through the normalised sensitivity index, so that the function  $F(\vec{p})$  of Eq. (3.26) is substituted by the analytical expression of  $R_0$ , Eq. (3.3).

---

<sup>2</sup>All numerical simulations of the dynamical system Eq. (3.2) have been carried out using a Runge-Kutta 4th order method, with a temporal step  $\Delta t = 0.001$ . Numerically stable results are obtain with  $\Delta t \leq 0.01$ .

**Fig. 3.2(a)** shows the sensitivity index for  $R_0$  for specific baseline parameters, where  $\lambda$ ,  $\bar{\beta}$  and  $S_0$  contribute to increase the basic reproduction number while  $\alpha$ ,  $\gamma$  and  $\mu$  contribute to decrease it, as expected. Moreover, we can see that  $\lambda$  and  $\gamma$  are the most influential parameters while  $\mu$ ,  $\bar{\beta}$  and  $S_0$  depend on each other. These dependencies cause varying influences on  $R_0$ , which are fully depicted in panels **Fig. 3.2(b-d)**. It can be seen that the influence of  $\bar{\beta}$  increases with the increase of  $\mu$  and the decrease of  $S_0$ . Similarly, the importance of  $S_0$  increases with  $\mu$  and decreases with  $\bar{\beta}$ . On the other hand, the impact of  $\mu$  increases with the decrease of both  $S_0$  or  $\bar{\beta}$ .

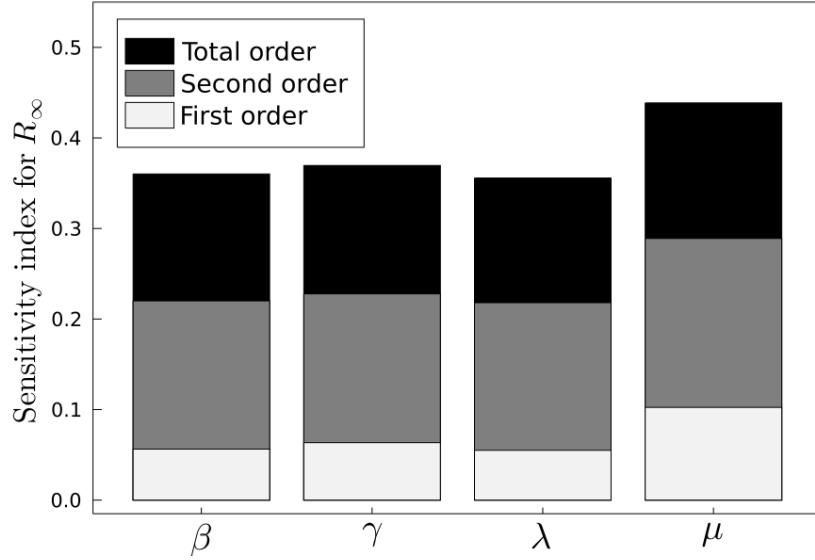


**Figure 3.2:** Panel (a): Local sensitivity analysis of  $R_0$  for the baseline parameters  $\lambda = 1$ ,  $\gamma = 1$  and  $\mu = \bar{\beta} = S_0 = 1$ . The asterisks mark parameters for which the sensitivity index is not constant, depending on, at least, another parameter. Panels (b-d): Local sensitivity analysis of  $R_0$  with respect to parameters with an asterisk, showing the different dependence with a second parameter and the effect on the varying sensitivity index.

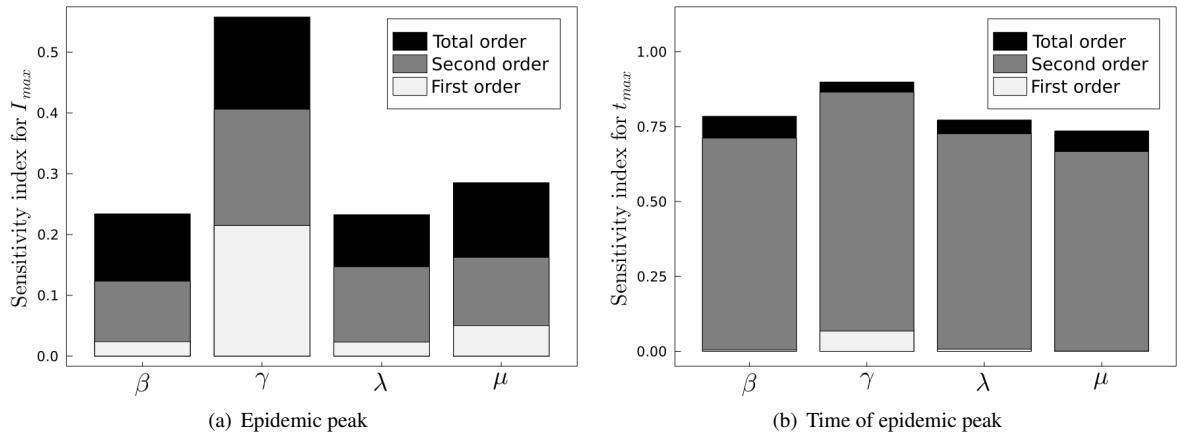
### 3.3.2 Final state of the epidemic

Another important quantity in epidemiology is the final state of the epidemic, which can be characterised by the final number of dead individuals,  $R_\infty$ . Within our general SIRP model it is not possible to find an analytical expression of  $R(t)$  so that we need to tackle the problem numerically. To this end, we perform GSA for the final number of dead individuals in order to determine the most influential parameters for this quantity. In particular, we apply the Sobol method, discussed in [3.6.4](#). The Confidence Interval, CI, obtained in our study is less than 1% of the index value, indicating a very high accuracy, therefore it is not shown in the figures. The results of the explained procedure are shown in [Fig. 3.3](#), where the total order (black), first order (white) and second order (gray) sensitivity indices for each of the model parameters are detailed. It can be observed that  $\mu$  has a slightly greater influence than the other parameters with respect to the final number of dead individuals. Note that the second order indices are larger than the first order ones for all the parameters, which indicates a high influence of the

nonlinearities in our model, at least for the particular quantity under study.



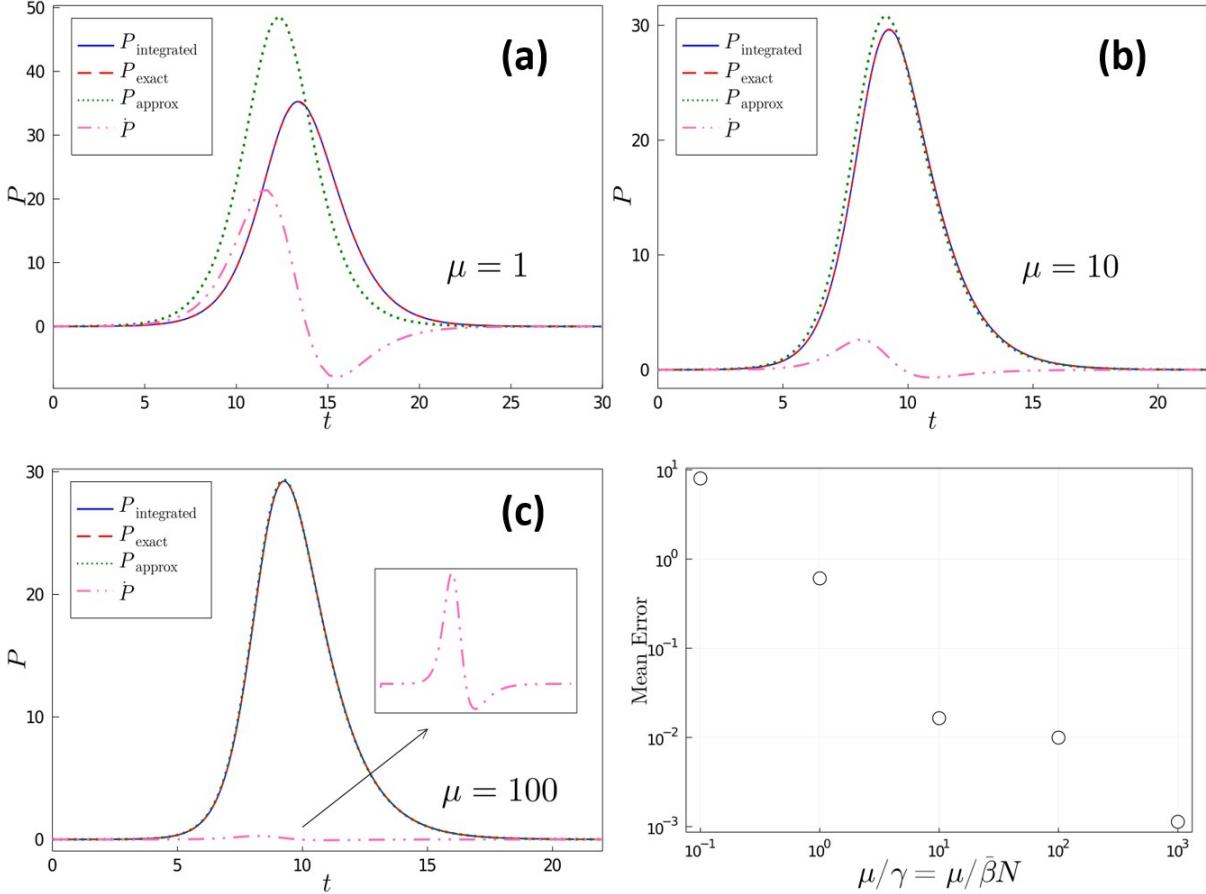
**Figure 3.3:** Sensitivity indices (LSA) for the final number of dead individuals ( $R_\infty$ ) for each one of the indicated parameters. The black bars represent the total order indices of sensitivity while white (grey) colour represents the contribution of the first (second) order indices.



**Figure 3.4:** Global sensitivity analysis for the maximum of infected individuals  $I_{max}$  (a) and its time occurrence  $t_{max}$  (b). The black bars represent sensitivity at all orders, while white (grey) colour represents the contribution of the first (second) order indices.

### 3.3.3 Maximum of infected individuals

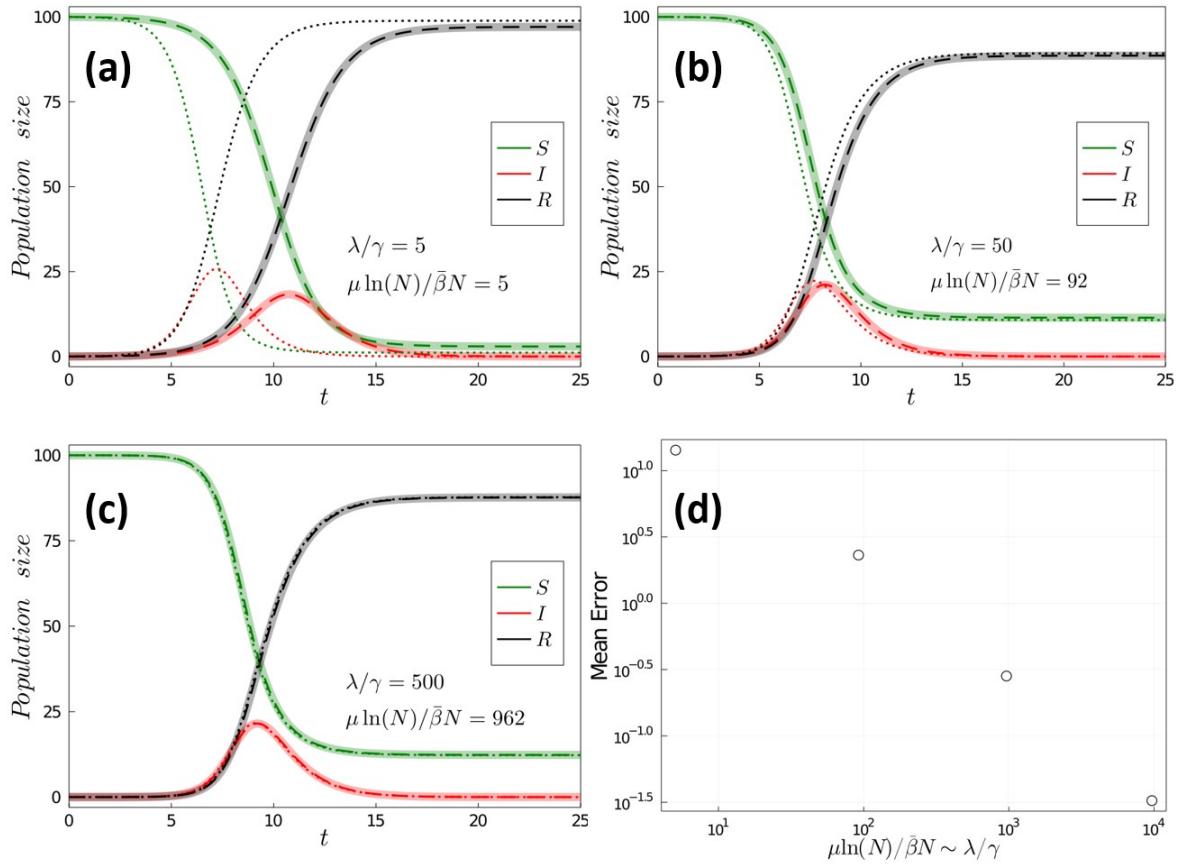
A GSA of the maximum number of infected individuals,  $I_{max}$  and the time it occurs,  $t_{max}$  is performed to study the influence of the model parameters regarding these quantities. In this case, Fig. 3.4,  $\gamma$  has greater influence in the epidemic peak than any of the other parameters, while for the time at which the peak takes place, all the parameters have basically the same influence. Again, the second order indices (the first order interactions between parameters) account for most of parameter sensitivity, in particular in the time of the epidemic peak, indicating the high degree of nonlinearity of this effect.



**Figure 3.5:** Numerical check of the approximate expression for the pathogen concentration, (Eq. (3.8)),  $\bar{\beta} = 1/50$  and  $\gamma = 1$ : (a)  $\mu = 1$ ; (b)  $\mu = 10$ ; (c)  $\mu = 100$ , while  $\lambda$  is varied to keep  $R_0 = 2.5$ , defined in (Eq. (3.3)) (with  $S_0 = N = 50$ ), i.e.,  $\lambda = 5, 27.5, 252.5$  respectively for (a)-(b)-(c), respectively. The blue solid line represents the numerically integrated quantity, the red dashed line (superimposed to the blue solid one as they are identical) is the exact solution for this quantity, (Eq. (3.4)) and the green dotted line accounts for the approximate expression from the timescale separation (Eq. (3.8)). The dash-dotted pink line represents the derivative of  $P$ ,  $\dot{P}$ , in the scaled time frame. Panel (d): Mean error between the approximate and exact solutions for increasing  $\mu/\gamma = \mu/\bar{\beta}N$ .

### 3.3.4 Numerical verification of the fast-slow approximation

The parasite concentration approximation, based on a timescale separation discussed in Section 3.2.3.3, is now verified by computational means. The verification was performed using both mass action and standard incidence, but for the sake of simplicity we show only the results for the standard incidence case. Worth is to say that, mathematically, changing from standard incidence to mass action involves only a rescaling of the  $\beta$  parameter, so that the numerical results are exactly the same. Fig. 3.5 contains a comparison for 3 different values of the parasite deactivation rate,  $\mu$ . It can be seen that the approximation is poor when  $\mu \sim \gamma, \bar{\beta}N$  Fig. 3.5(a), as it could be expected. On the other hand, the approximation is quite good when  $\mu$  is one order of magnitude larger than  $\gamma$  and  $\bar{\beta}N$  Fig. 3.5(b), while it is extremely accurate when  $\mu$  is two orders of magnitude larger than  $\gamma, \bar{\beta}N$ , Fig. 3.5(c). The figure also shows the numerical value of  $\dot{P}$  (pink dashdot), and it can be checked how it becomes smaller as  $\mu$  increases compared to  $\gamma, \bar{\beta}N$ , justifying the timescale separation of Section 3.2.3.3. Finally, Fig. 3.5(a-c) also shows (dashed red line) the analytical value for  $P(S,I)$  derived in Eq. (3.16), that matches perfectly the result of the numerical integration of Eq. (3.2), as should be the case.



**Figure 3.6:** Numerical check of the exact model reduction along with the subsequent approximation shown in Section 3.2.3.1 with  $N = 100$ ,  $\bar{\beta} = 1/100$ ,  $\gamma = 1$ , (a)  $\lambda = 5$ ,  $\mu = 1.1$ ; (b)  $\lambda = 50$ ,  $\mu = 20$ ; (c)  $\lambda = 500$ ,  $\mu = 209$ .  $R_0 = 2.38$  for all the panels. The solid semitransparent lines represent the original 4D model, the dashed lines the exact reduction and the dotted lines the approximate model from the exact reduction. Panel (d): Mean error between the approximate and exact solutions for increasing  $\mu \ln(N)/\bar{\beta}N$  and  $\lambda/\gamma$  while  $R_0 = 2.38$  is kept constant.

### 3.3.5 Numerical verification of the model approximation from the exact reduction

The numerical verification was performed for both mass action and standard incidence, but for the sake of simplicity in Fig. 3.6 we show only the results for the standard incidence case. First, and as it should be because it is an exact result, the exact reduction of the SIRP model discussed in Section 3.2.3.1 matches perfectly the numerical results obtained from the full model for all possible parameter values, Fig. 3.6(a-c). Regarding the approximation to the exact reduction, one can see how the approximation converges to the exact solution as the parameters fulfil the conditions indicated in Section 3.2.3.2, namely that both  $\gamma/\lambda \gg 1$  and  $\mu \log(N)/\bar{\beta}N \gg 1$ , becoming very accurate if these ratios are larger than 1 by two orders of magnitude or more (cf. Fig. 3.6(c)). We recall that in this case the SIRP model converges to the SIP model of (44). Conversely, the approximation is poor when any of these two ratios is of order 1 (cf. Fig. 3.6(a))), while Fig. 3.6(b) presents the result in an intermediate case, in which the approximation is fair.

## 3.4 Model validation with data of the *Pinna nobilis* Mass Mortality Event

In this section, the general SIRP model is validated against collected data from the *Pinna nobilis* Mass Mortality Event. As explained in Section 3.1, the disease is caused by the parasite *Haplosporidium pinnae* and the hosts, *P. nobilis*, are sessile bivalves endemic of the Mediterranean Sea. Thus, this epidemic is a perfect candidate to be described by the SIRP model. In the model, parasite production occurs only inside infected hosts, and parasites are released to the medium, either through their respiratory or digestive system. The simultaneous occurrence of the different

possible stages of the parasite (uni- and bi-nucleate cells, multinucleate plasmodia, sporocysts and uninucleate spores) in the same host individual is not common among haplosporidans and makes *H. pinnae* different from previously known haplosporidian species (92). The occurrence of uni- and binucleate stages suggest possible direct transmission from infected to healthy fan mussels, as observed in *B. ostreae* and *B. exitiosa* (24, 122, 229). Additionally, the presence of spores (a dormant, resistant stage) could allow long persistence in the environment and the hypothetical involvement of an intermediate host as suggested for *H. nelsoni* and *H. costale* (19, 222, 353). While uninucleate cells are always detected in infected fan mussels, sporulation has been only detected sporadically (92). Thus, we assume that infection occurs mostly through uninucleate (or binucleate) cells by direct transmission (as the experimental observations in captivity point out, see (175)). We do not consider disease transmission through other stages. We do not consider spores, given the infrequent observation of spores and the current lack of experimental information about spore transmission (that could involve another intermediate host species). Regarding plasmodia and sporocyst stages, these stages are too large to be released through the epithelium. The distinction between uninucleate and binucleate cells seems unnecessary at this level of representation, as these phases only participate in parasite proliferation inside infected hosts, a process that we consider in an effective way. Finally, the evidence of the time course of the disease compared to the long life cycle of *P. nobilis* suggests host vital dynamics (i.e. recruitment (reproduction) and natural death) can be neglected.

After an epidemic outbreak that took place in Portlligat, in the north east of Catalonia, 215 *Pinna nobilis* individuals were extracted from their natural medium in order to be preserved as a genetic reserve in several controlled water tanks of different institutions in Spain (175). The institutions that participated in this preservation effort were IFAPA, IEO, IRTA, IMEDMAR-UCV and Oceanogràfic of Valencia. The original idea was to rescue the individuals before infection, however, the subsequent evolution of the rescued *Pinna nobilis* populations indicates that some individuals were already infected at the time of extraction (and/or in contact with some amount of the parasite transferred from sea water). This allowed the opportunity to use the data of the time evolution of the epidemic in the controlled water tanks, reported in (175), to evaluate the described SIRP model<sup>3</sup>. The empirical data consists of the proportion of survivors as a function of time in the controlled water tanks with a temporal resolution of one month. Despite the fact that the temperature of the water in the tanks was controlled, it was sharply lowered in most of the tanks when mortality started to appear within the population, as a last effort to keep the rest of the population safe and alive, since keeping the temperature below approximately 13.5°C is a known strategy to preserve *Pinna nobilis* individuals as disease expression is minimal (76). Fortunately, two of the tanks kept its temperature approximately constant during the full recorded time. This is the case of the tanks in IFAPA in Huelva and the Oceanogràfic of Valencia (OCE), both Spanish institutes. These water tanks have been selected to validate our model, maintaining constant temperatures of 14°C and 17°C, respectively.

First we will fit the exact reduction of the SIRP model, assuming  $\mu \log(N) \gg \bar{\beta}N$  and  $\lambda/\gamma \gg 1$  as discussed in Section 3.2.3.2, namely Eq. (3.6). This reduced model depends on three parameters ( $\lambda'$ ,  $\mu$ ,  $\gamma$ ) and one constant,  $\mathcal{C}(0)$ , cf. Section 3.2.3.2, that is related to the initial conditions of the model. The order of magnitude of the mortality rate can be deduced from data, with an estimate value of  $\gamma \approx 1 \text{ month}^{-1}$ . We fix this parameter in order to give some biological information to our model prior to the computational fit. We focus on the *R* compartment, as it can be retrieved directly from data in (175)<sup>4</sup>. We use a box-constrained variant<sup>5</sup> of the well known BFGS optimisation algorithm (164) with a common L2 loss function, also known as Residual Sum of Squares (RSS)<sup>6</sup>. By running this algorithm one observes that the optimal parameters tend to be the ones in the boundary of the box-constrained parameter space. Furthermore, if the box size is increased (or decreased) the optimal parameters continue to be in the boundary of the box-constrained parameter space. This indicates that there exist several parameter combinations that optimally fit the data, and the combination parameters found by the optimisation algorithm are only marginally optimal with respect to other parameter values. The locus (actually a valley)

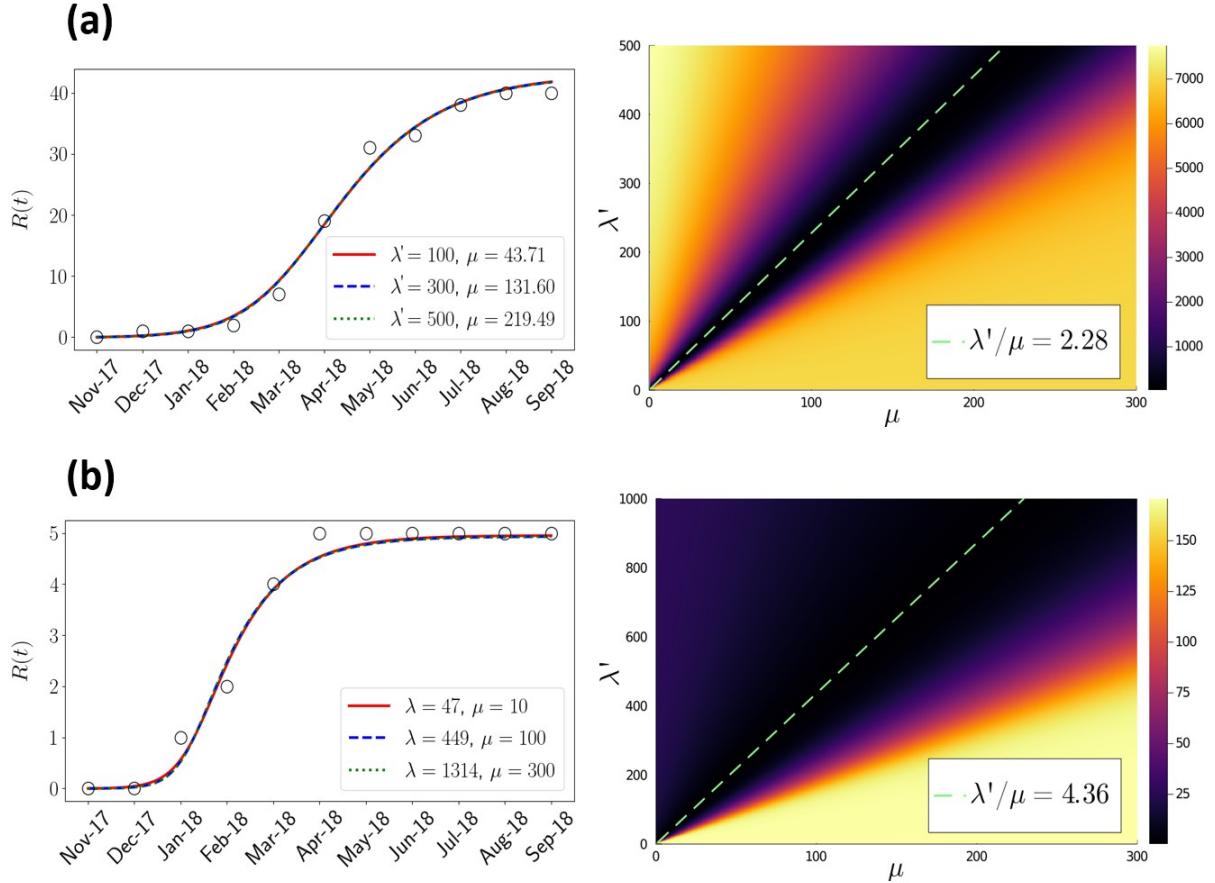
<sup>3</sup>Data use in this work with the purpose of validating and fitting parameters for the SIRP model have been taken from the Supplementary Information of [175]

<sup>4</sup>The number of dead individuals can be obtained as  $R = N - S$ , where  $S$  is the population of survivors and  $N$  is the total number of individuals in the tanks, 50 (IFAPA) and 5 (Oceanogràfic), respectively

<sup>5</sup>We constrain the optimisation because the unconstrained optimisation to the full range of the parameters, i.e., from 0 to  $\infty$  is not practical.

<sup>6</sup>The algorithm is implemented within the Julia high-level programming language [40] using the DifferentialEquations.jl package [362].

of marginal optimal parameters can be seen in the right hand side panels of Fig. 3.7, where the cost function value of the optimisation algorithm is plotted as heat map.



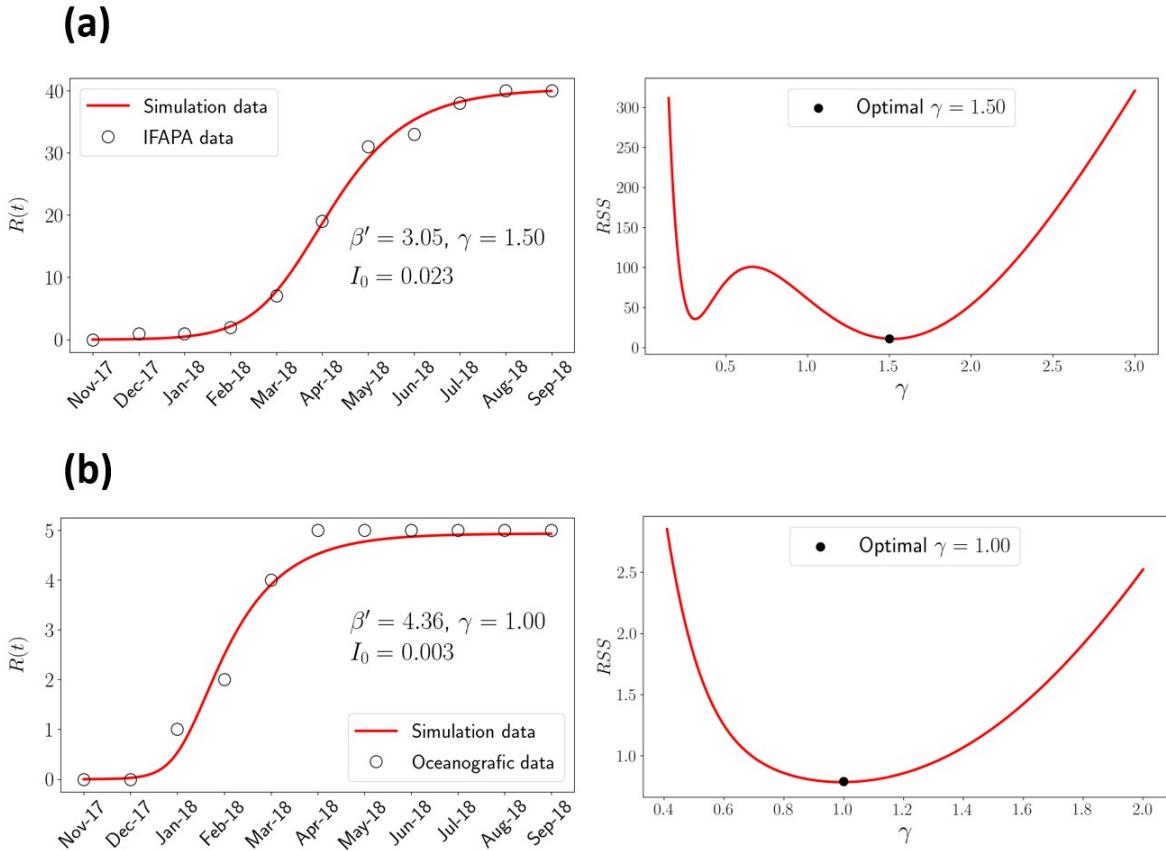
**Figure 3.7:** Parameter estimation for the approximation from the exact reduction of the SIRP model (Eq. (3.6)) using data from IFAPA (panel (a)) and OCE (panel (b)) water tanks, at 14 °C and 17 °C respectively. Left figures represent several fits of the model to empirical data of the number of dead hosts ( $R(t)$ ) using different optimal combinations of the parameters. Right figures are the RSS errors as a function of the input parameters, where the green dashed line represents the set of optimal combinations of the parameters with  $RSS = 60, 0.8$  for IFAPA and OCE, respectively.

Now we reach the point regarding the dilemma between mass action and standard incidence discussed in Section 3.2.1. If one does not correct the  $\bar{\beta}$  parameter with the size of the host population,  $N$ , that is equivalent to assuming the mass action incidence  $\bar{\beta} = \beta$ , the values that one would obtain for  $\beta' = \beta\lambda/\mu = \lambda'/\mu$  for both populations take disparate values in both tanks:  $\beta' = 0.046$  for the IFAPA data set and  $\beta' = 0.87$  for the Oceanogràfic (OCE) data set, a factor of 19 between them while their temperatures differ only by 3 °C. These numbers indicate that the standard incidence is more reasonable, what amounts to choosing  $\bar{\beta} = \beta/N$ , where the final values of the reported parameter  $\beta'$  should be multiplied by  $N = 50$  for the IFAPA tank and  $N = 5$  for the OCE tank. The final result is then  $\beta' = 2.28$  and  $\beta' = 4.36$  for IFAPA and OCE tanks, that are the values reported in Fig. 3.7, implying that an almost twofold increase of the  $\beta'$  parameter corresponds to an increase of 3 °C. This relation is in good agreement with the typical changes in rates of a wide range of organisms with a 3 °C change in temperature, while a 19-fold change in the rate would imply at least a 30 °C change in temperature (cf. 3.6.5).

The fact that there is an infinite number of combinations of the parameters that optimally fit the real data suggests that, as two parameters are slaved one to each other, that the model admits a further reduction. This reduction corresponds exactly to the approximate SIR model derived in Eq. (3.9), with the relationship  $\beta' = \lambda'/\mu$ , as anticipated. So, this gives further corroboration to the use of the SIR model Eq. (3.9) to fit  $\beta'$  as the free parameter (fixing the value

of  $\gamma$  and with  $I_0$  as the initial condition determined by the fit). For consistency with the previous fitting we expect to obtain  $\beta' = 2.28$  and  $4.36$  as the optimal parameters for the IFAPA and OCE water tanks, respectively, and this is the case.

Interestingly, as reduced model Eq. (3.9) has fewer parameters to fit we can relax our initial assumption of  $\gamma = 1 \text{ month}^{-1}$  and check how the fit improves or worsens when varying  $\gamma$ . In Fig. 3.8 a fit of the reduced SIR model Eq. (3.9) is shown for the IFAPA (top) and Oceanogràfic (bottom) controlled water tanks<sup>7</sup>. Fig. 3.8(c-d) shows the RSS error as  $\gamma$  is varied. It can be seen that for the IFAPA water tanks  $\gamma = 1.5 \text{ month}^{-1}$  yields more accurate results, while for the Oceanogràfic water tanks  $\gamma = 1 \text{ month}^{-1}$  remains optimum. This shows a decrease in the mean removal time  $1/\gamma$  for lower water temperatures, with the finite size errors inherent to the OCE tank (as  $N = 5$ ). In the left panels the simulated curve of dead individuals,  $R$  compartment, as a function of time for the optimal fitted parameters is confronted to the experimental data, showing a remarkable agreement. With the optimum values of  $\gamma$ , in the IFAPA tank (now with  $\gamma = 1.5 \text{ month}^{-1}$ ) a new value of  $\beta' = 3.05$  is obtained, implying a probably more reasonable ratio of 1.43 for  $\beta'$  in both tanks (it was 1.91 in the original fit). From the optimal parameters we obtain the basic reproduction number, since  $R_0 = \beta'/\gamma$  we have that  $R_0^{\text{IFAPA}} \simeq 2$  and  $R_0^{\text{OCE}} \simeq 4$ , clearly above the epidemic threshold.



**Figure 3.8:** Parameter fitting for the R compartment to model (Eq. (3.9)) using data from IFAPA (panel (a)) and Oceanogràfic (panel (b)). The left part of both panels of the figure shows the optimal fit of the model to empirical data with  $\text{RSS} = 10.9, 0.8$  for IFAPA and OCE, respectively. The right panels show the variation of the RSS error for some values of  $\gamma$ . The  $\beta'$  values have been obtained assuming a standard incidence, as explained in the main text.

Summarising, the SIRP model is able to fit two sets of experimental data, agreeing with a standard incidence, according to which the infection rate depends on the amount of parasites per pen shell individual. *Pinna nobilis* individuals in the IFAPA experiment were actually

<sup>7</sup>The  $N$  correction corresponding to standard incidence has already been applied to these values.

distributed in 4 tanks, and the standard incidence is compatible with this experimental aspect. The temperature dependence of the fitted parameters in this range ( $14 - 17^\circ\text{C}$ , appears to be compatible (although experiments at different temperatures would be needed) with an Arrhenius dependence of the infection parameters, also known as Boltzmann-Arrhenius (70, 311), that can be extended to account for the expect unimodal dependence on temperature, with a maximum infectivity at a characteristic temperature for the parasite (311). Therefore, we can assume that global change (or temperature shifts) is expected to have complex effects on infectious diseases, causing some to increase, others to decrease, and many to shift their distributions (371). In the particular case of pen shell mortality, our model results suggest the proposed mechanism of lower disease expression at lower temperatures. This might have direct consequences for the development of the mortality event and offers a bleak perspective for the future and specifically in the eastern Mediterranean basin, where the mortality was observed later due to current patterns but average temperatures tend to be higher than in the western part of the Mediterranean.

### 3.5 Conclusions

In this work we have analysed a compartmental model to study marine epizootics for sessile hosts assuming infection by direct transmission through waterborne parasites. Moreover, we have used data from the recent mass mortality event of *Pinna Nobilis* in the Mediterranean Sea as a case study to validate our model. Compartmental models are routinely used in the study of disease infection and propagation in terrestrial ecosystems, including the study of the current Covid-19 pandemic (see, e.g., (91)). However, these models are starting to be used only recently in the study of marine epizootics (44), while proliferation models have been the most popular in the field (354). A reason for the low popularity of compartment models in the study of marine epizootics is that there are some aspects in its modelling that differ from the now standard application to terrestrial ecosystems (300). An important difference is that, in principle, (micro)parasites need to be modelled explicitly in marine ecosystems, while often they are not included in the description in terrestrial ecosystems (299).

The SIRP model has 4 compartments and depends on 4 parameters, so that it is not quite amenable to theoretical analysis. At the same time, due to the large number of parameters of the model, using it to analyse experimental observations can be cumbersome in practice if the parameter values are unknown. Nevertheless, we have shown three reductions of the model, one exact and two approximate ones, that can be useful to overcome these limitations that are typically present at the first stages of emergent epidemics. Indeed, the timescale approximation is able to fit the collected data of our case study for some optimal parameters, as shown in Section 3.4. This approximation is particularly useful as it only depends on 2 parameters, the death rate of infected hosts,  $\gamma$  and an effective infection rate,  $\beta'$ . Although this approximation simplifies the fitting procedure, there is a price to be paid in this analysis. The infection parameter,  $\beta$ , and the parameters regulating proliferation,  $\lambda$ , and deactivation/dilution of the parasite,  $\mu$ , become entrained into a single effective parameter,  $\beta'$ . Thus, the full understanding of the different effects at play in the system requires further work. Furthermore, we have shown that an epidemic model for immobile hosts can be reduced to the standard SIR model, which assumes direct contact among the hosts, i.e. that the hosts are mobile. This reduction is only valid when the time scale of the parasites is much faster than that of the hosts, i.e.  $\mu \gg \beta N, \gamma$ . Thus, our work provides a ground to apply the SIR model in marine epidemics of sessile hosts that fulfil the required conditions.

In a world with many possible new epizootics, we believe that our reduced model can be specifically useful to understand key features of those emerging diseases characterised by the spreading of waterborne parasites in a relatively fast way, provided that the temporal evolution of the disease can be determined for, at least, some set of individuals. Thus, some of the key parameters can be fitted to the available experimental data as shown in Section 3.4. Still, the fitted relevant parameters may need to be supplemented with further information or targeted experiments. We hope that this approach can be useful in understanding emerging diseases in shellfish species of economic not only ecological value, and also, with suitable modifications, in aquaculture. It is noteworthy that our case study is a haplosporidian waterborne parasite. In fact, waterborne haplosporidans have been responsible for some of the most significant and

consequential marine disease epizootics on record and are considered the major pathogens of concern for aquatic animal health shellfish industries around the world (22). The SIRP model is the simplest model that one could think of having in mind its practical application, but could be extended to incorporate further effects that are so far described in an effective way.

## 3.6 Appendix

### 3.6.1 Finding a conserved quantity for the SIRP model

Starting with the SIRP model,

$$\begin{aligned}\dot{S} &= -\bar{\beta}PS \\ \dot{I} &= \bar{\beta}PS - \gamma I \\ \dot{R} &= \gamma I \\ \dot{P} &= \lambda I - \bar{\beta}PS - \mu P,\end{aligned}\tag{3.10}$$

from the  $\dot{S}$  equation,  $P$  can be written as follows,

$$P = -\frac{1}{\bar{\beta}} \frac{\dot{S}}{S},\tag{3.11}$$

and summing up the equations for  $\dot{S}$  and  $\dot{I}$  the following relation for  $I$  is obtained

$$I = -(\dot{S} + \dot{I})/\gamma.\tag{3.12}$$

Replacing Eq. (3.11), Eq. (3.12) and the differential equation for  $\dot{S}$  in the 4th differential equation in Eq. (3.10) one obtains,

$$\dot{P} = -\frac{\lambda}{\gamma}(\dot{S} + \dot{I}) + \dot{S} + \frac{\mu}{\bar{\beta}} \cdot \frac{\dot{S}}{S}\tag{3.13}$$

As  $\dot{S}/S = d(\ln S)/dt$ , all terms in the previous equation are exact differentials with respect to time, and the equation can be integrated yielding,

$$P + \frac{\lambda}{\gamma}(S + I) - S - \frac{\mu}{\bar{\beta}} \ln S = C\tag{3.14}$$

with the integration constant  $C$ , that is a conserved quantity, i.e., it takes the same value at one time of the dynamical evolution of the system.  $C$  is related to the initial conditions by,

$$C = P(0) + \frac{\lambda}{\gamma}(S(0) + I(0)) - \frac{\mu}{\bar{\beta}} \ln S(0) - S(0) = P(0) + \frac{\lambda}{\gamma}(N - R(0)) - \frac{\mu}{\bar{\beta}} \ln S(0) - S(0)\tag{3.15}$$

It is possible to use Eq. (4.13)-Eq. (3.15) to express one of variables as a function of the others, for example the parasite concentration  $P$  as,

$$P(S, I) = P(0) - \frac{\lambda}{\gamma}(S + I - N + R(0)) + \frac{\mu}{\bar{\beta}} \ln \frac{S}{S(0)} + S - S(0),\tag{3.16}$$

or equivalently as,

$$P(S, R) = P(0) + \frac{\lambda}{\gamma} \left[ R - R(0) + \frac{\mu\gamma}{\bar{\beta}\lambda} \ln \frac{S}{S(0)} \right] + S - S(0)\tag{3.17}$$

From Eq. (4.13), it is easy to show that the SIP model of Ref. (44), that differs from the SIRP model in that the fourth equation is simplified to  $\dot{P} = \lambda I - \mu P$ , has as exact conserved quantity,

$$P + \frac{\lambda}{\gamma}(S + I) - \frac{\mu}{\bar{\beta}} \ln S = \mathcal{C}\tag{3.18}$$

as the extra term in the SIRP model  $-\bar{\beta}SP$  is equal to  $\dot{S}$  from the first equation Eq. (3.10).

The SIR model has a conserved quantity (327), that in the case of Eq. (3.9) takes the form,

$$I + S - \frac{\gamma}{\beta'} \ln S = C. \quad (3.19)$$

Rewriting Eq. (4.13) in the alternative form,

$$\frac{\gamma}{\lambda} P + \left(1 - \frac{\gamma}{\lambda}\right) S + I - \frac{\mu\gamma}{\lambda\bar{\beta}} \ln S = C' \quad (3.20)$$

it can be seen that if  $\lambda \gg \gamma$  Eq. (3.20) reduces to Eq. (3.19), remembering that in Eq. (3.9)  $\beta' = \lambda\bar{\beta}/\mu$ . The assumptions used to arrive to Eq. (3.19) in Section 3.2.3.3 where  $\mu \gg (\gamma, \bar{\beta})$ , and taking into account the expression for  $R_0$  Eq. (3.3), that  $\lambda \gtrsim \mu$  is most plausible to keep  $R_0$  above the epidemic threshold ( $R_0 > 1$ ).

### 3.6.2 Stability analysis of the fixed points of the SIRP model

Here we will assume the initial fixed point of our SIRP model, with  $I(0) = P(0) = 0$  right before the introduction of the infection, either through  $I$  or  $P$ . We will assume that  $R(0) = 0$ , so that  $S(0) = N$ . To study the linear stability of the model we need to write the Jacobian, that takes the form,

$$J = \begin{pmatrix} -\bar{\beta}P & 0 & 0 & \bar{\beta}S \\ \bar{\beta}P & -\gamma & 0 & \bar{\beta}S \\ 0 & \gamma & 0 & 0 \\ -\bar{\beta}P & \lambda & 0 & (\bar{\beta}S - \mu) \end{pmatrix} \quad (3.21)$$

and obtain the eigenvalues for both fixed points, where we have already used the standard incidence,  $\bar{\beta} = \beta/N$ , from the evidence of the validation with experiments. For the pre-epidemic fixed point, the Jacobian becomes,

$$\begin{pmatrix} 0 & 0 & 0 & \bar{\beta}S(0) \\ 0 & -\gamma & 0 & \bar{\beta}S(0) \\ 0 & \gamma & 0 & 0 \\ 0 & \lambda & 0 & (\bar{\beta}S(0) - \mu) \end{pmatrix} \quad (3.22)$$

Matrix Eq. (3.22) has two null (0) eigenvalues and a pair of eigenvalues given by,

$$\Lambda_{1,2} = -\frac{1}{2} \left( \gamma + \mu + \bar{\beta}S(0) \pm \sqrt{\gamma^2 + \mu^2 + (\bar{\beta}S(0))^2 + 2\mu\bar{\beta}S(0) - 2\gamma\mu - 2\gamma\bar{\beta}S(0) + 4\lambda\bar{\beta}S(0)} \right) \quad (3.23)$$

from which one can determine that the fixed point is unstable whenever

$$\lambda\bar{\beta}S(0) > \gamma(\mu + \bar{\beta}S(0)) \quad (3.24)$$

and stable if the inequality is reversed. It can be easily shown that Eq. (3.24) is equivalent to  $R_0 > 1$ , with  $R_0$  given by Eq. (3.3).

The final point of the epidemic,  $S(\infty)$ , can be found by solving the transcendental equation,

$$\left( \frac{\lambda}{\gamma} - 1 \right) S(\infty) - \frac{\mu}{\bar{\beta}} \ln(S(\infty)) = C \quad (3.25)$$

where  $C$  is determined from the initial conditions (Eq. (3.15)) and  $I(\infty) = P(\infty) = 0$ . (Eq. (3.25)) has two roots, where  $S(\infty)$  represents the smallest one.

### 3.6.3 Calculation of $R_0$ using the Next Generation Matrix method

The so called Next Generation Method (NGM) is a method to obtain  $R_0$ , the basic epidemiological quantity that measures the number of secondary cases produced by a typical infected individual during its entire period of infectiousness in a completely susceptible population. It was discussed in 3.6.2 that  $R_0$  is related to the largest non-zero eigenvalue, say  $\Lambda$ , of the fixed point corresponding to the infection-free equilibrium. An outbreak occurs when  $\Lambda > 0$  (or equivalently

when  $R_0 > 1$ ) and the NGM is an ingenious method to obtain directly  $R_0$  in a reduced linear system. In more concrete terms, within the NGM method  $R_0$  is the dominant eigenvalue of a suitably defined linear operator (a linear matrix in a suitable basis). This operator is obtained from a decomposition of the Jacobian,  $J$  of the infected/infecting compartments (i.e. excluding susceptible and removed compartments) in the form  $J = T + \Sigma$ , where  $T$  is the *transmission part*, that describes the production of new infections, and  $\Sigma$  the *transition part*, that describes changes of state (including death). Then, it can be proved (138) that the *basic reproduction number*  $R_0$  is given by the spectral radius (i.e. the largest eigenvalue) of the (next generation) matrix  $K = -T\Sigma^{-1}$ .

In the case of the SIRP model the decomposition is applied to the  $2 \times 2$  Jacobian corresponding to the dynamical evolution of the  $(I, P)$  infectious compartments, being the decomposition,

$$J = \begin{pmatrix} -\gamma & \bar{\beta}S_0 \\ \lambda & -(\bar{\beta}S_0 + \mu) \end{pmatrix} \quad T = \begin{pmatrix} 0 & \bar{\beta}S_0 \\ 0 & 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} -\gamma & 0 \\ \lambda & -(\bar{\beta}S_0 + \mu) \end{pmatrix}$$

where the  $\bar{\beta}PS$  term in  $I$  is the only one that contributes to the transmission matrix, as it is the only process involving infection, while all the other terms in the dynamical equations of  $I$  and  $P$  imply transitions (to another compartment, like  $I \rightarrow R$  or birth and death of  $P$ ).

Then, the next generation matrix is given by,

$$K = -T\Sigma^{-1} = \begin{pmatrix} \frac{\lambda\beta S_0}{\gamma(\beta S_0 + \mu)} & \frac{\beta S_0}{\beta S_0 + \mu} \\ 0 & 0 \end{pmatrix} \implies R_0 = \frac{\lambda\beta S_0}{\gamma(\beta S_0 + \mu)},$$

This result coincides with the expectation that  $R_0$  should correspond to the number of hosts infected in a single generation by the appearance of an infected host in a completely susceptible population. This can be obtained from the number of parasites produced by an infected host,  $\lambda$ , times the time in which the infected host is alive producing parasites,  $1/\gamma$ , multiplied by the number of infected hosts produced per parasite,  $\beta S_0$ , times the time the parasite is alive available to infect,  $1/(\mu + \beta S_0)$ , taking into account that parasites are inactivated at a rate  $\mu$  and also die when infecting at a rate  $\beta S_0$ , where this result assumes that the susceptible population does not change from its initial value  $S_0$ .

### 3.6.4 Sensitivity Analysis

One particular way to analyse the local sensitivity (LSA) of a given model function,  $F(\vec{p})$ , for each of the parameters that conform it,  $p_i$ , is through the normalised sensitivity indexes (82),

$$\Omega_{p_i}^F = \frac{\partial F}{\partial p_i} \frac{p_i}{F} \Big|_{p_i=p^0}. \quad (3.26)$$

where the partial derivatives in Eq. (3.26) are determined analytically in our case.

GSA works by studying the influence of a large domain of parameter space in the final state of the epidemic and in the epidemic peak. In our case this will be achieved by means of a variance based analysis, known as Sobol method (403). This particular method provides information no only on how a particular parameter alone influences the model outputs (as happens with LSA), but also on the influence of its interactions with other parameters. This information is organised in what are known as Sobol indices, that have been implemented within the Julia high-level programming language (40) using the DifferentialEquations.jl package (362), and in particular through its subpackage DiffEqSensitivity.jl. This implementation allows the user to sample the parameter space using QuasiMonteCarlo methods and thus obtain confidence intervals (CI) for the sensitivity indices, which are directly related to the committed statistical error.

The total order indices are a measure of the total variance of the output quantity caused by variations of the input parameter and its interactions. First order (or “main effect”) indices are a measure of the contribution to the output variance given by the variation of the parameter alone, but averaged over variations in other input parameters. Second order indices take into account first order interactions between parameters. Further indices can be obtained, describing the influence of higher-order interactions between parameters, but these are not going to be considered. More detailed information about sensitivity analysis can be found in (377).

### 3.6.5 General rate change with temperature

In [187] the metabolic rate of a wide variety of organisms was studied, showing that the change in the metabolic rate with temperature was similar among them. In particular, the natural logarithm of the metabolic rate linearly depends on the inverse of absolute temperature,

$$\log(R(T)) = a \cdot \left( \frac{100}{T} \right) + b \quad (3.27)$$

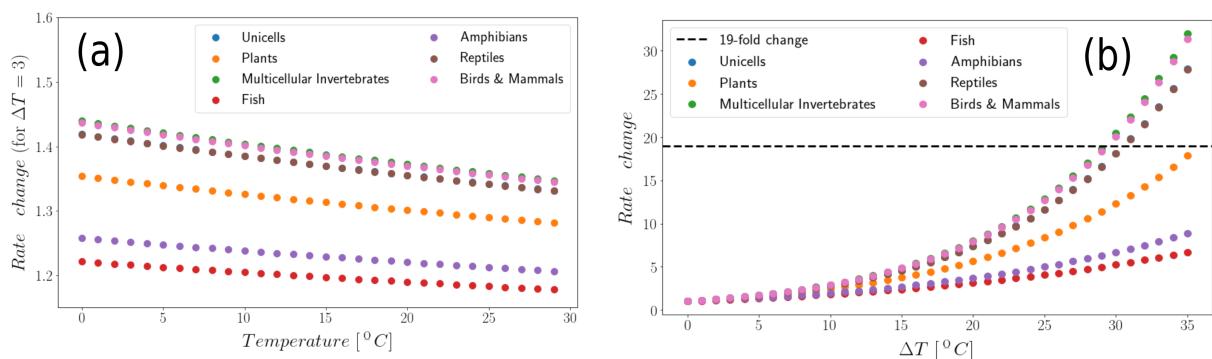
and for all the analysed organisms they found that  $a$  lies between  $-5$  and  $-10$  and  $b$  between  $14$  and  $30$ . From their analysis, we can compute the change in the rate for a given increase of temperature,

$$\frac{R(T + \Delta T)}{R(T)} = \exp\left(a \cdot \left(\frac{100}{T + \Delta T}\right) + b\right) / \exp\left(a \cdot \left(\frac{100}{T}\right) + b\right) = \exp\left(a \cdot \frac{-1000}{T + \Delta T} \cdot \frac{\Delta T}{T}\right). \quad (3.28)$$

Substituting  $T = 287K$  and  $\Delta T = 3K$ , that correspond to our available data (cf. [Section 3.4](#)) in [Eq. \(3.28\)](#), using both the upper and lower limit of  $a$ , we obtain that the expected increase in the effective transmission rate is between  $1.2$  to  $1.45$ . This is far from the  $19$ -fold increase that we obtained with the mass action hypothesis in [Section 3.4](#) while it is in good agreement with either the  $1.92$  ratio we obtained for  $\bar{\beta}$  with the reduction of [Section 3.2.3.2](#) or the  $1.43$  ratio obtained with the fast-slow approximation of [Section 3.2.3.3](#), both obtained using the standard incidence choice.

[Fig. 3.9\(a\)](#) shows the change in the rate with an increase of  $3^\circ C$  for different base temperatures and for all the organisms analysed in [187], and using their fit. Note that for all temperatures between  $0^\circ C$  and  $30^\circ C$  the rate change lies between  $1.2$  and  $1.45$ . [Fig. 3.9\(b\)](#) shows the change in the rate for different temperature increases, with a base temperature of  $T = 287K$ . Note that in order to obtain a  $19$ -fold increase the temperature change should be at least of  $30^\circ C$ <sup>8</sup>. The temperature dependence of metabolic rates has been reported in the context of epidemic parameters [107, 398]

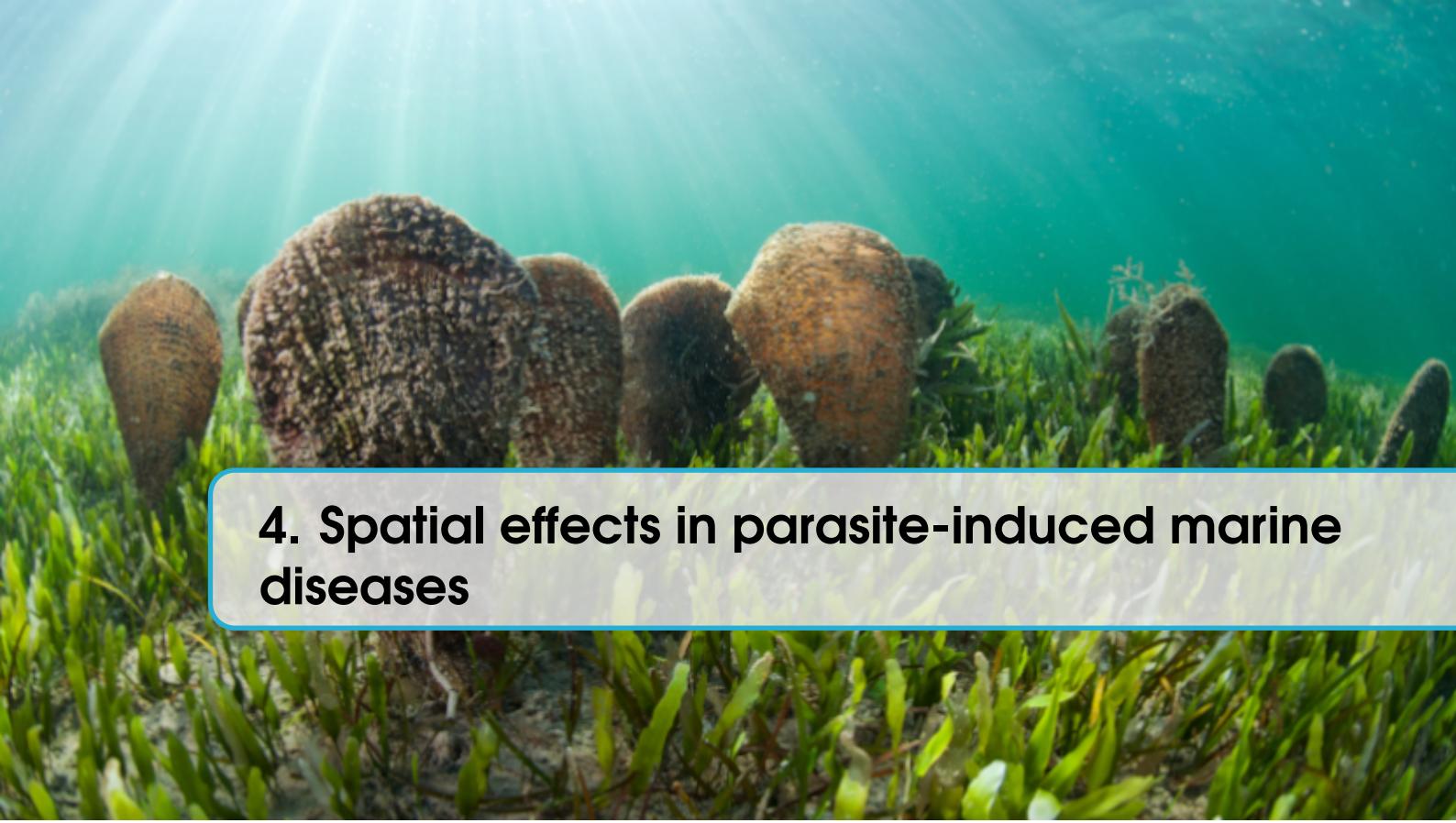
The behavior of the metabolic rates re-analysed here has been also found experimentally in epidemic contexts such as [107, 398], i.e. the increase of the rates with temperature fulfill the ranges shown here.



**Figure 3.9:** Graphical representation of change in the rate (in ordinates) for different reference temperatures (in abscissae) for: (a) a temperature increase of  $3^\circ C$ ; (b) a temperature increase of  $14^\circ C$ . The black dotted line in (b) corresponds to a  $19$ -fold increase in the rate.

<sup>8</sup>A temperature change of  $30^\circ C$  could fall outside the range in which the study of [187] is valid. We just stress that a  $19$ -fold rate change is unlikely for the case of a  $3^\circ C$  that correspond to the 2 data sets that we compare in this section.





## 4. Spatial effects in parasite-induced marine diseases

Àlex Giménez-Romero<sup>1</sup>, Federico Vázquez<sup>2,1</sup>, Cristóbal López<sup>1</sup>, Manuel A. Matías<sup>1</sup>

1. Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Palma de Mallorca 07122, Spain
2. Instituto de Cálculo, FCEyN, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina

### Published as

Àlex Giménez-Romero et al. "Spatial effects in parasite-induced marine diseases of immobile hosts". In: *Royal Society Open Science* 9.8 (2022), page 212023. DOI: [10.1098/rsos.212023](https://doi.org/10.1098/rsos.212023). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.212023>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsos.212023>

## 4.1 Introduction

Wildlife emergent infectious diseases represent a substantial threat to ecosystems and the conservation of their biodiversity (127). Their effects can be devastating at the ecological level, causing local extinctions (127) and in some cases pushing endemic species to the verge of extinction, as is the case of *Pinna nobilis* (76); at the economic level, producing losses in agriculture, livestock and aquaculture (347, 420, 439), and impact human health, as is the case of the COVID-19 pandemic (376). For the past decades, parasites have been continuously emerging (124, 317), while globalisation and climate change have contributed to their evolution. This has allowed these parasites to enter in new ecological niches and spread further the diseases they produce (4). In particular, marine infectious diseases are recently increasing due to these and other anthropogenic pressures, like pollution and overfishing (268), inducing widespread mass mortalities in several species (150, 247, 434).

An important subset of marine organisms affected by infectious diseases are sessile (i.e. they cannot move), like bivalves, sponges or corals. An increasing number of outbreaks affecting marine mollusks have been reported, some of them causing mass mortalities in commercially important bivalves (216). Mainly due to the economic importance of some species (e.g. oysters), infectious diseases in bivalve populations have been deeply studied (301, 348, 349, 353). Recently, deterministic compartmental models have been used to describe parasite transmitted diseases in marine sessile bivalves (42, 43, 198), showing to be able to accurately predict disease transmission in some circumstances. The main limitation of these compartmental models is the assumption of a non-spatial description of the system under study. This underlying hypothesis assumes that any pair parasite-host of the system can interact at any time, which is unrealistic in general. A non-spatial description assumes well mixed populations, which implies that the mean distance among hosts is smaller than the typical distance explored by parasites in their lifetime. This assumption can be quite realistic in some situations, as it is in (198) where the hosts were kept in tanks with water renovation. However, a non-spatial model is not expected to yield a good description of spatially extended hosts in a natural setting.

The key quantity in mathematical epidemiology is the basic reproductive number,  $R_0$ , that represents the number of infected individuals generated in one generation by the appearance of a single infected individual in a fully susceptible population. Thus,  $R_0 > 1$  ensures the onset of an epidemic, as the number of infected individuals will grow exponentially producing a widespread disease (18). If we first disregard spatial effects and assume a non-spatial description,  $R_0$  can be obtained from standard methods, like the Next Generation Matrix method (138), and will only depend on *intrinsic* characteristics of the pathosystem (host-pathogen system) under study. However, this basic reproduction number is unable to characterise the threshold behaviour in many situations, including spatially extended systems (120, 281, 367). In these systems, the propagation of an epidemic to the entire system needs that a certain spatial threshold is exceeded (186). Otherwise the disease will only take place in suitable localised parts of the system, not being able to propagate to the total system. Thus, disease spread will be strongly affected by the host spatial distribution and pathogen mobility, which are not accounted for in non-spatial models.

In this work we will try to unravel the transmission mechanisms of a parasite-induced disease affecting immobile hosts in a spatially extended system. We will approach the problem both theoretically and through numerical simulation. The numerical study is based on Individual-Based Modelling (IBM), a method widely used to study ecological systems (214), so that individuals are treated as discrete entities, space is introduced explicitly and the dynamics are stochastic. Representative average behaviours can be obtained by averaging over a sufficient number of realisations, and the accuracy of the approach can be calibrated by deriving the corresponding non-spatial limit, that can be confronted with the suitable compartmental model on which a particular IBM is based. The IBM approach to our problem will allow to study in depth the relation between pathogen mobility and immobile host infection. As parasites move randomly over the space, tracking the position of each parasite at different times turns to be of fundamental importance to properly capture the stochastic dynamics of infections from parasites to hosts. Modelling parasites and hosts as individual entities allows to take into account the spatial and temporal heterogeneity of interactions between them. This heterogeneity and the level of control in microscopic interactions cannot be captured by other mathematical approaches

such as partial differential equations. On the other hand, IBMs are mathematically involved, and analytical treatments are normally cumbersome, while their numerical implementation is computationally expensive (64).

Here we introduce a spatially-explicit individual-based model to study parasite-induced marine diseases of immobile hosts. The model is applied to the case of diffusing parasites and uniformly distributed hosts. The system under study is an extension of the compartmental model presented in (198). As a main result, we find that the occurrence of an outbreak will depend on the balance between the intrinsic characteristics of the pathosystem, well represented by the above described non-spatial basic reproductive number,  $R_0$ , and features that characterise parasite mobility. We generalise the basic reproductive number, that we will refer to as  $\tilde{R}_0$ , such that it accounts for the number of hosts that get infected by the appearance of a single infected individual in a fully susceptible population in a spatially extended system.  $\tilde{R}_0$  characterises the global epidemic and can be written as a product between  $R_0$  and a factor describing parasite mobility. The latter factor is smaller and at most equal to 1, which implies that, as it could be expected, it is more difficult to induce a global outbreak in a spatially extended system (a two dimensional lattice in our case) than in a well mixed (non-spatial) population.

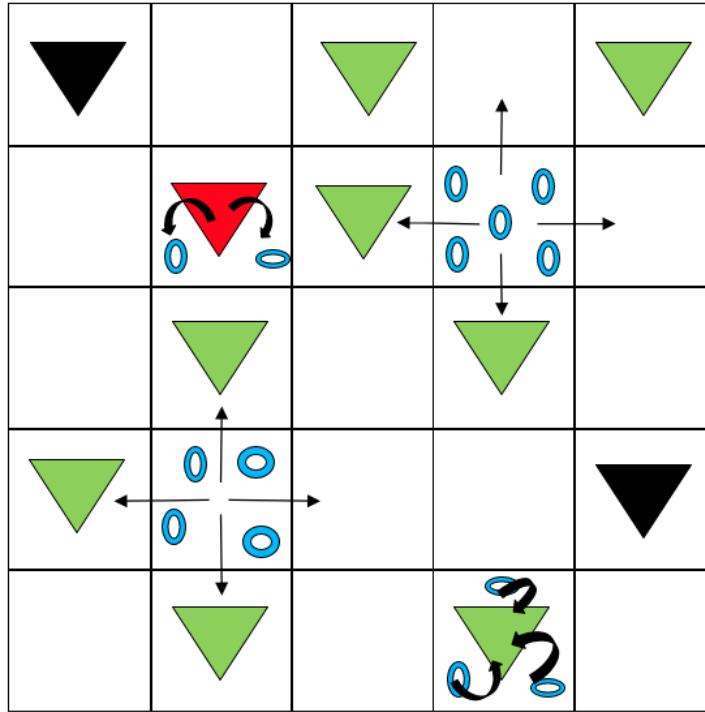
The paper is organised as follows: in Section 5.2, we introduce some biological considerations for bivalve epidemics, discussed in more detailed in Ref. (198), and build the spatially-explicit model. In Section 5.3, we present analytical results that are discussed and supported by numerical simulations. Specifically, the high mobility limit is discussed and connected to the the compartmental model. An approximation for the parasite population is discussed. Then, the effect of parasite mobility to the epidemic threshold is characterised, deriving an analytical expression for the basic reproduction number. Furthermore, the spreading speed of the disease and the time-scale to extinction is investigated. Finally, Section 4.4 contains some concluding remarks.

## 4.2 The SIRP spatial model

The most important biological features of the system under study are as follows. First, hosts are immobile, while the disease is transmitted by parasites produced by infected hosts. There are two mechanisms by which parasites are cleared from the medium: i) they have a finite life time after which they die; ii) they get absorbed after they infect a host and thus are no longer in the medium and cannot infect other hosts. Recruiting (birth) of hosts occur at a very slow rate compared to other timescales in the system, and accordingly it will be considered negligible in the model. Moreover, hosts do not show long-term immunity, as is typical of invertebrates, like mollusks (354). We also assume that recovery (healing) of infected hosts, if it occurs, can be neglected. Furthermore, we consider that dead hosts are not a source of parasites in the medium. See (198) for a detailed presentation of the non-spatial SIRP model, including these biological modelling considerations.

Under these considerations, we introduce an individual-based model with explicit space characterisation to study the effect of parasite mobility in disease transmission. We consider a square grid of length  $L$  with periodic boundary conditions and place a single host per site, so that there are  $N = L^2$  hosts. The hosts can be in three discrete states: susceptible,  $S$ ; infected,  $I$  and dead (or removed),  $R$ . Then, we introduce the parasite population as a new individual with a single state,  $P$ . Hosts are sessile (i.e. immobile), while parasites are allowed to move between the lattice sites. As initial condition we assume that the entire host population is susceptible,  $S(0) = N = L^2$ , and that a small initial number of parasites,  $P(0)$  is introduced in the system.

Infection occurs when susceptible hosts filter parasites in their close proximity. Accordingly, the infection process is implemented between parasites and susceptible hosts sharing the same lattice site. In particular, susceptible hosts in contact with a parasite become infected at rate  $\beta$ . As the infection event implies the filtering of a parasite by a susceptible host, when a new infection occurs a parasite of that particular site is removed. Infected individuals die at rate  $\gamma$  and produce parasites at rate  $\lambda$ , while parasites die at rate  $\mu$ . Parasites move randomly between the four neighbouring lattice sites at rate  $\kappa$ , which corresponds to a diffusive motion. Table 4.1 contains the definitions of the variables and parameters of the model and Fig. 4.1 shows a schematic representation of the dynamics.



**Figure 4.1:** Scheme of the individual based model. Green, red and black triangles represent susceptible, infected and dead hosts, respectively. Blue rings represent parasites, which move randomly between cells. Susceptible hosts get infected by filtering parasites while infected hosts produce them. Dead hosts do not participate in the dynamics of the system.

**Table 4.1:** Variables and parameters of the model

Variable/Parameter	Definition
S	Susceptible host
I	Infected host
R	Dead host
P	Parasite
$\beta$	Parasite-host transmission rate
$\gamma$	Host mortality rate
$\lambda$	Production rate of parasites by infected hosts
$\mu$	Parasite natural death rate
$\kappa$	Parasite dispersal rate (mobility)
$R_0$	Non-spatial basic reproductive number
$\tilde{R}_0$	Spatial basic reproductive number

Formally, the model is mathematically described by a system of  $N$  master equations for the probabilities of the states in each lattice site  $i$ . Eq. (4.1) summarises the reactive events. This is very difficult to manage analytically, so the time evolution of the model is numerically solved using Gillespie's algorithm (184) (the code can be found in (192)).



## 4.3 Results

In this section several features of the model are studied, both numerically (from IBM simulations) and analytically. All numerical results were obtained for a square lattice of length  $L = 100$ , with  $N = S(0) = 10^4$  hosts and using a small initial condition of  $P(0) = 50$  parasites in the centre site.

### 4.3.1 Non-spatial limit

An important test of the IBM implementation is to show that, under suitable circumstances, it converges to the non-spatial model on which the IBM is based. This occurs in the limit when the parasites move many times before dying or infecting a susceptible host. In this situation, each parasite visits typically all the hosts of the system and may infect any of them. This is equivalent to infecting a random host of the system, which happens with probability  $\beta S/N$ , being  $S$  the total number of susceptible hosts in the system. An equivalent picture is that parasites will end up uniformly distributed in the lattice, so that there will be  $P/N$  parasites in each lattice site at any time. One expects to reach these conditions when  $\kappa \gg \mu, \beta$ , and, thus the system as a whole can be described by the following system of ordinary differential equations (ODE's),

$$\begin{aligned}\dot{S} &= -\beta PS/N, \\ \dot{I} &= \beta PS/N - \gamma I, \\ \dot{R} &= \gamma I, \\ \dot{P} &= \lambda I - \beta PS/N - \mu P,\end{aligned}\tag{4.2}$$

that is precisely the SIRP non-spatial model (198), where  $S, I, R$  are the total number of susceptible, infected and recovered hosts in the system,  $P$  the total number of parasites and  $N$  is the number of hosts.

The basic reproduction number,  $R_0$ , of this non-spatial model is the dimensionless quantity that yields the number of secondary infections generated by the appearance of a single infected individual in a completely susceptible population, also indicating whether the system will exhibit an epidemic outbreak,  $R_0 > 1$ , or not,  $R_0 < 1$ . In our case it can be directly computed as the mean number of parasites produced by an infected host during its mean lifetime,  $\lambda/\gamma$ , times the mean number of susceptible hosts that get infected by parasites during their mean lifetime,  $\beta/(\mu + \beta)$ ,

$$R_0 = \frac{\lambda}{\gamma} \frac{\beta}{\mu + \beta}.\tag{4.3}$$

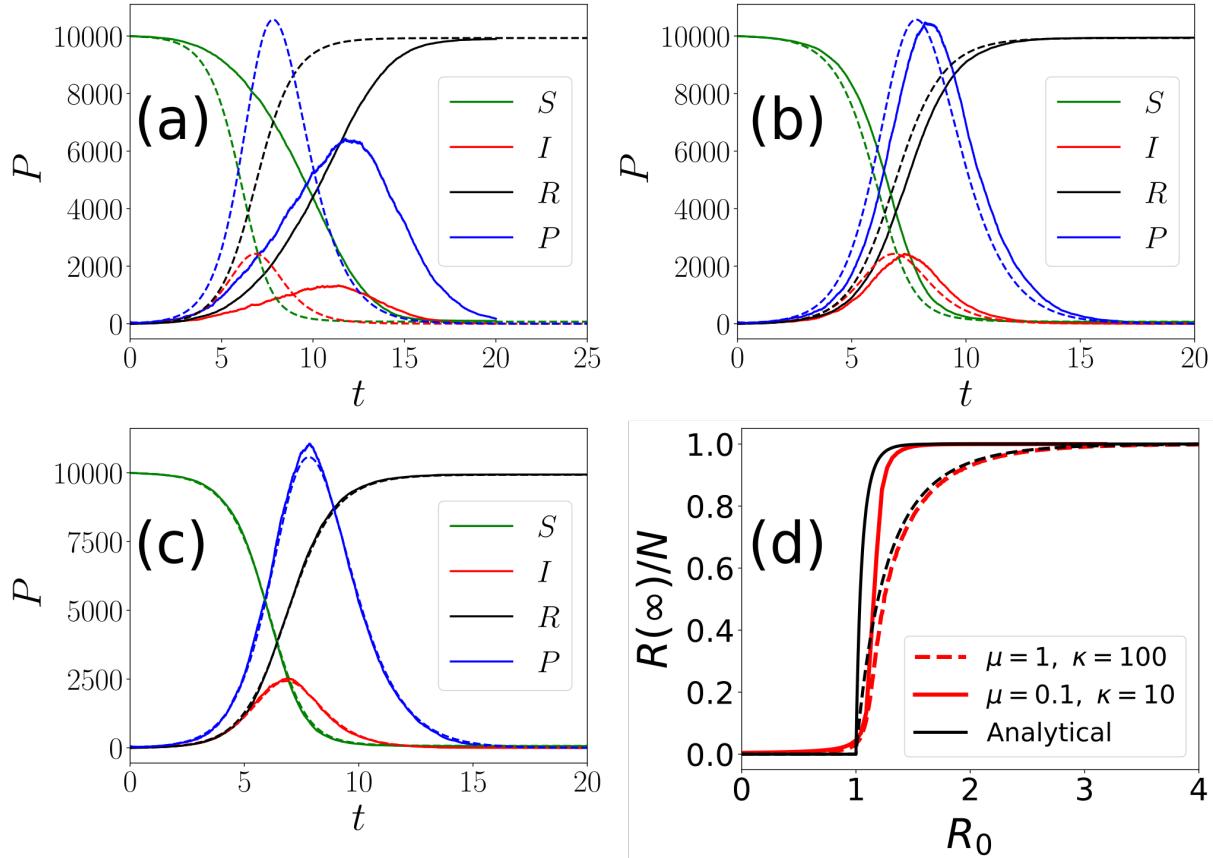
This result can be corroborated with standard methods such as the Next Generation Matrix method (138) (see (198)), where  $S(0) = N$  has been considered.

Moreover, the model has a conserved quantity  $\mathcal{C}$  (198) that allows to find an analytical expression for the final number of dead individuals (cf. Section 4.5.1),

$$R(\infty) = N + \frac{S(0)}{\xi} W_0 \left( -\xi \exp \left( -\frac{\beta}{\mu} C \right) \right),\tag{4.4}$$

$$\text{with } \xi = S(0) \frac{\beta(\lambda - \gamma)}{\mu\gamma} \text{ and } C = P_0 + \frac{\lambda}{\gamma} (S(0) + I(0)) - S(0).$$

The non-spatial limit of the model has been evaluated by comparing realisations of the stochastic model (in the limit  $\kappa \gg \mu, \beta$ ) with numerical solutions of the non-spatial ODE system of Eq. (4.2). Furthermore, the analytical expression for  $R(\infty)$  using the non-spatial model, Eq. (4.4), is also compared to the numerical results of the individual based model. As shown in Fig. 4.2 (a)-(c), as  $\kappa$  is increased compared to  $\mu$  the individual based model approaches the non-spatial one. Fig. 4.2(d) shows how the numerical results for  $R(\infty)$  for different  $R_0$  values approach the analytical solution in the non-spatial limit.



**Figure 4.2:** Numerical solution of the non-spatial model (Eq. (4.2), dashed lines) compared with numerical solutions of the individual based model (solid lines) approaching the non-spatial limit with fixed  $\gamma = \mu = \beta = 1$  and  $\lambda = 6$ . (a)  $\kappa = 10^2$ , (b)  $\kappa = 10^3$ , (c)  $\kappa = 10^4$ . Panel (d) shows the final fraction of dead hosts,  $R(\infty)/N$ , as function of  $R_0$  for  $\kappa/(\mu + \kappa) = 0.999$  with  $\mu = 1, 0.1$  compared to the analytical result.

### 4.3.2 Approximate relation between parasites and infected hosts

In the limit  $\kappa \gg \beta, \mu$  a time-scale approximation can be performed so that the parasite population dynamics directly relates to that of the infected hosts. In the non-spatial limit it was already shown in (198) that, if  $\mu \gg \beta, \gamma$  and  $\lambda \gg \beta P/N$ , the total parasite population of the system can be well described using the approximation (see (198) for a detailed discussion),

$$P(t) \approx \frac{\lambda}{\mu} I(t) , \quad (4.5)$$

Here we extend the validity of this approximation to spatial systems far from the non-spatial limit. Consider the local dynamics of the parasite population on a lattice site  $i$ . Note that when the host in the site is susceptible, parasites in this site can either infect the host, die, or move to another site. All these processes imply that a parasite will disappear from the current site. Once the host at site  $i$  gets infected, infection can no longer occur whereas parasite production is now possible. If  $\kappa$  is small enough compared to  $\lambda$  and  $\mu$ , the only competing processes in sites with infected individuals will be the production of parasites and their natural death, which can be fairly described by the following rate equation,

$$\frac{dP}{dt} = \lambda - \mu P , \quad (4.6)$$

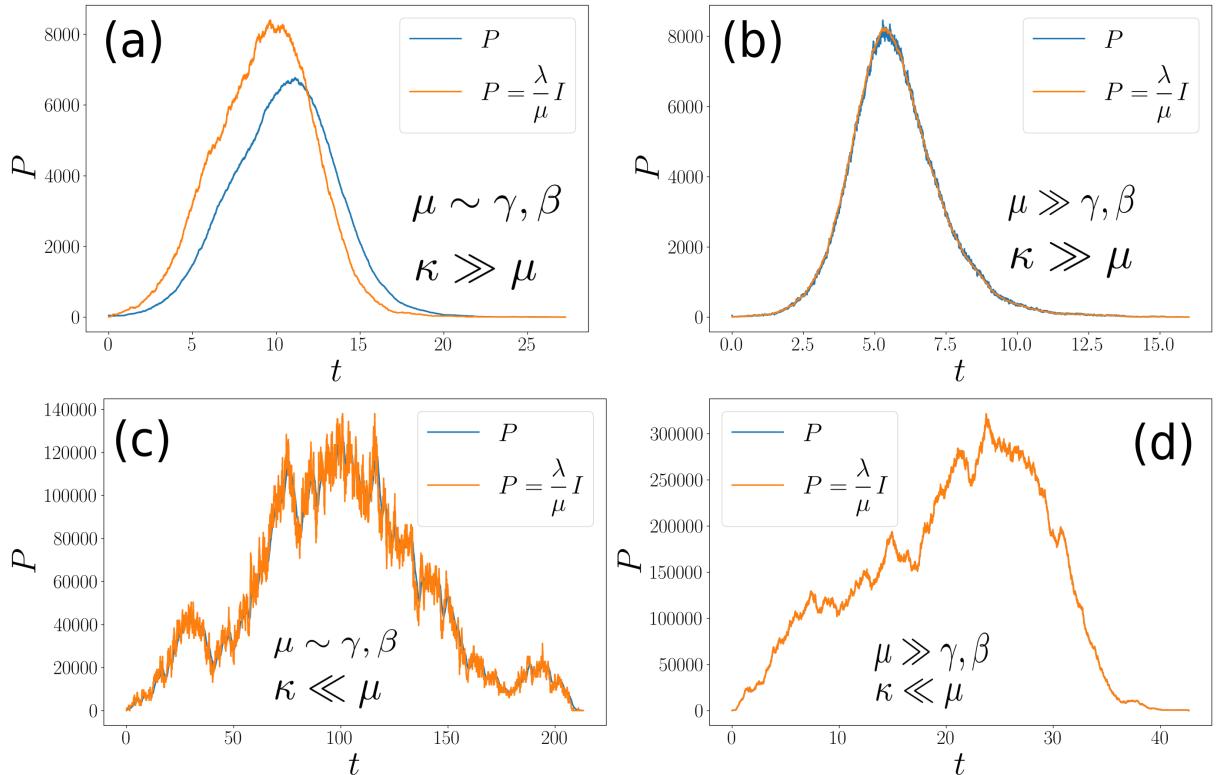
whose solution is

$$P(t) = \frac{\lambda}{\mu} + \left[ P(0) - \frac{\lambda}{\mu} \right] e^{-\mu t} . \quad (4.7)$$

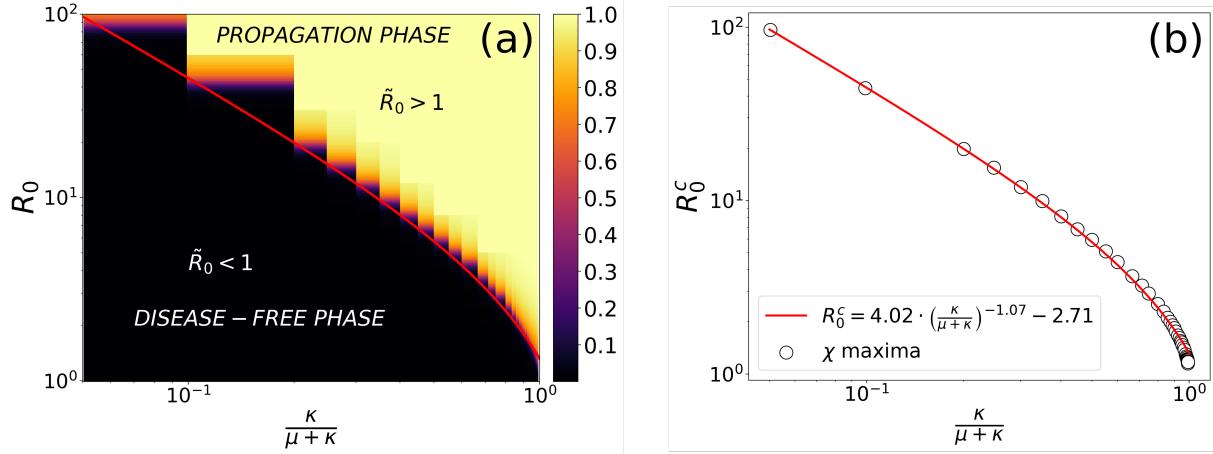
From Eq. (4.7) one may notice that the stationary value of  $P$ ,  $\lambda/\mu$ , is reached in a time proportional to  $t_{\text{eq}} \propto 1/\mu$ . This derivation allows to find a condition for which ?? is valid beyond the non-spatial limit. Basically, if the mean dispersal time,  $1/\kappa$ , is greater than the equilibrium time,  $t_{\text{eq}} \propto 1/\mu$ , parasites in sites with infected hosts will reach its stationary level before parasites enter or leave the sites. Thus, sites with infected hosts can be considered as a closed system and the approximation holds. In other words, if the dispersal rate of parasites is small compared to the parasite deactivation rate,  $\kappa \ll \mu$ , the local parasite population of the site will reach its stationary level  $P_i = \lambda/\mu$ . It is possible to extend the result to the entire system: if there are  $I(t)$  infected sites in the system at time  $t$  and  $\kappa \ll \mu$  is fulfilled, there will be a total parasite population of  $P(t) = (\lambda/\mu)I(t)$ , which is equivalent to ??.

Thus, for the non-spatial limit ( $\kappa \gg \mu$ ) we have that if  $\mu \gg \beta, \gamma$  ?? is valid, while for  $\kappa \ll \mu$  the approximation is also valid regardless of the value of  $\beta, \gamma$ , as the nature of the approximation is different. Thus, in general, as  $\kappa$  decreases over  $\mu$  (the lower the parasite mobility becomes) we expect the approximation to work better.

The parasite approximation to infected hosts dynamics, ??, is numerically verified for different mobility conditions. -Fig. 4.3(a)-(b) shows how the approximation improves as  $\mu$  grows over  $\beta, \gamma$  (mean errors are 0.18 and 0.0081, respectively) in the non-spatial limit, i.e.  $\kappa \gg \mu$ , as expected. This result is in perfect agreement with that found in (198). Then, Fig. 4.3(c)-(d) show that the approximation is valid in general when  $\kappa \ll \mu$  but improves anyway when  $\mu \gg \beta, \gamma$  (mean errors are 0.04 and 0.0026, respectively). Summarising, we see that the lower the value of  $\kappa$  is with respect to  $\mu$  the more valid ?? is, regardless of the value of  $\beta, \gamma$ , while in the non-spatial limit,  $\kappa \gg \mu$ , the condition  $\mu \gg \beta, \gamma$  is needed.



**Figure 4.3:** Numerical verification of the approximate expression for the parasite population dynamics, ??, for different mobility conditions. The simulations were performed fixing  $\beta = \gamma = 1$  for all panels. (a)  $\mu = 1$ ,  $\kappa = 10^2$ ,  $\lambda = 6.06$ ,  $\kappa/(\mu + \kappa) = 0.99$ ; (b)  $\mu = 100$ ,  $\kappa = 10^4$ ,  $\lambda = 306$ ,  $\kappa/(\mu + \kappa) = 0.99$ ; (c)  $\mu = 1$ ,  $\kappa = 0.01$ ,  $\lambda = 1200$ ,  $\kappa/(\mu + \kappa) = 0.01$ ; (d)  $\mu = 100$ ,  $\kappa = 1$ ,  $\lambda = 60600$ ,  $\kappa/(\mu + \kappa) = 0.01$



**Figure 4.4:** (a) Phase diagram showing the transition between the the disease-free phase and the propagation phase for several values of the parasite mobility and  $R_0$ . The colour code represents the fraction of dead individuals (i.e.  $R/N$ ) in the final state of the epidemic computed by the average over 1000 realisations. (b) Fit for the transition line following Eq. (4.10), where dots are the maximums of the “order parameter” fluctuations,  $\chi = \langle R(\infty)^2 \rangle - \langle R(\infty) \rangle^2$

#### 4.3.3 Spatial threshold

One of the main questions in epidemiology is to define the conditions under which an epidemic outbreak occurs, which usually is translated into the existence of a threshold. In a well mixed (non-spatial) system the basic reproduction number ( $R_0$ ), that characterises this threshold  $R_0 = 1$ , can be defined exclusively from *intrinsic* parameters of the pathosystem, as the host-pathogen interaction does not depend on the host spatial structure or pathogen mobility (see Eq. (4.3)). In stochastic spatial models this formulation of  $R_0$  breaks down. First of all, in stochastic models, even above the threshold there is a non-zero probability that the disease is unable to propagate initially, given by  $P_{\text{outbreak}} = 1 - (1/R_0)^{I(0)}$  (62). Furthermore, the discrete nature of the populations also modifies the estimates of  $R_0$  (257). On the other hand, the introduction of space changes completely the nature of epidemic outbreaks, modifying the host-pathogen interactions by means of specific host spatial distributions and pathogen mobility patterns. Even if the basic reproduction number of the non-spatial model is above the threshold ( $R_0 > 1$ ), if parasite mobility is not large enough, the epidemic will stay locally confined. Thus, one expects that the threshold at which an epidemic outbreak can propagate to the rest of the system will depend on the balance between the intrinsic pathosystem parameters in  $R_0$  and parasite mobility, defining a spatial basic reproduction number,  $\tilde{R}_0$ .

Having in mind the study in Section 4.3.1, we expect that in the high mobility limit the basic reproduction number is defined by the non-spatial formula, Eq. (4.3). On the other hand, the lower the parasite mobility is, the more difficult will be for a local outbreak to propagate through the system. Thus, it is natural to think of an spatial basic reproduction number of the form  $\tilde{R}_0 = R_0 f(\kappa)$ , where  $f(\kappa)$  is an increasing function of the parasite dispersal rate accounting for parasite mobility fulfilling the limit  $\lim_{\kappa \rightarrow \infty} f(\kappa) = 1$ .

Indeed, some authors recently showed that the spatial basic reproduction number can be defined as the product between the non-spatial value,  $R_0$ , and a factor accounting for spatially-dependent interactions,  $f(r)$ , in the form  $\tilde{R}_0 = R_0 f(r)$  (159, 413). However, these expressions are not analytical (159, 160) or are not directly related to pathogen mobility (413). Here we propose a simple expression for the spatial basic reproduction number regulating the spatial propagation of the epidemic,

$$\tilde{R}_0 = \frac{\lambda}{\gamma} \frac{\beta}{\mu + \beta} \frac{\kappa}{\mu + \kappa} = R_0 \frac{\kappa}{\mu + \kappa}. \quad (4.8)$$

The derivation of Eq. (4.8) accounts for the number of secondary parasites that are able to produce new infections, or equivalently, the number of secondary infections produced by an initial infected host. If we consider an initial infected individual, on average it will produce  $\lambda/\gamma$  parasites. Then, these parasites can only move to neighbouring sites or die, so that the dispersal

probability is given by  $\kappa/(\mu + \kappa)$ . Finally, considering that parasites do not affect each other trying to infect the same host, the infection probability is given by  $\beta/(\mu + \beta)$ . Joining all terms, we finally obtain Eq. (4.8). This expression is valid when parasites move only to sites with susceptible individuals and do not try to infect the same host. Thus, the derived  $\tilde{R}_0$  is only an approximation to the spatial basic reproduction number for the case of an initial introduction of a small quantity of parasites in a fully susceptible population.

Note that, as expected, the spatial basic reproduction number is nothing other than the basic reproduction number of the non-spatial model multiplied by an increasing function of the parasite mobility,  $\kappa/(\mu + \kappa)$ . Taking the limit  $\kappa \gg \mu$  in Eq. (4.8) (non-spatial limit) the basic reproduction number of the non-spatial model is recovered. Conversely, in the limit of very low mobility, the  $\kappa/(\mu + \kappa)$  factor is small, and this has to be compensated with a large value of the non-spatial basic reproduction number,  $R_0$ , in order that there is an outbreak, i.e.,  $\tilde{R}_0 > 1$ .

The spatial threshold,  $\tilde{R}_0 = 1$ , given by Eq. (4.8), has been numerically checked by computing the phase diagram between the absorbing phase  $R(\infty) \approx 0$  (no infection, i.e. disease-free state) and the active phase  $R(\infty) > 0$  (in which some level of infection has occurred, i.e. propagation phase) for several values of the parasite mobility and the basic reproduction number of the non-spatial model,  $R_0$  Eq. (4.3). The transition is expected to occur at  $\tilde{R}_0 = 1$ , implying from Eq. (4.8) that the dependence of the critical value of  $R_0$ , say  $R_0^c$ , is expected to take the form,

$$R_0^c \sim \left( \frac{\kappa}{\mu + \kappa} \right)^{-1}. \quad (4.9)$$

As discussed above, we expect that  $\tilde{R}_0 = 1$  with  $\tilde{R}_0$  given by Eq. (4.8) does not represent exactly the spatial threshold, and for this reason we suggest the more general functional form,

$$R_0^c \sim A \left( \frac{\kappa}{\mu + \kappa} \right)^{-B} - C, \quad (4.10)$$

to be fitted to numerical data, where  $A = 1$ ,  $B = 1$  and  $C = 0$  would imply a perfect agreement of numerical simulations of the IBM model with Eq. (4.8).

In order to obtain the phase diagram, we compute the absorbing state of the model as an average over 1000 realisations for each value of the mobility and  $R_0$  considered. Then, the critical value  $R_0^c$  is computed for each mobility value as the  $R_0$  value for which the fluctuations of the “order parameter”  $\chi = \langle R(\infty)^2 \rangle - \langle R(\infty) \rangle^2$  are maximal, as this would be an indication of a transition between the disease-free and the propagation phases.

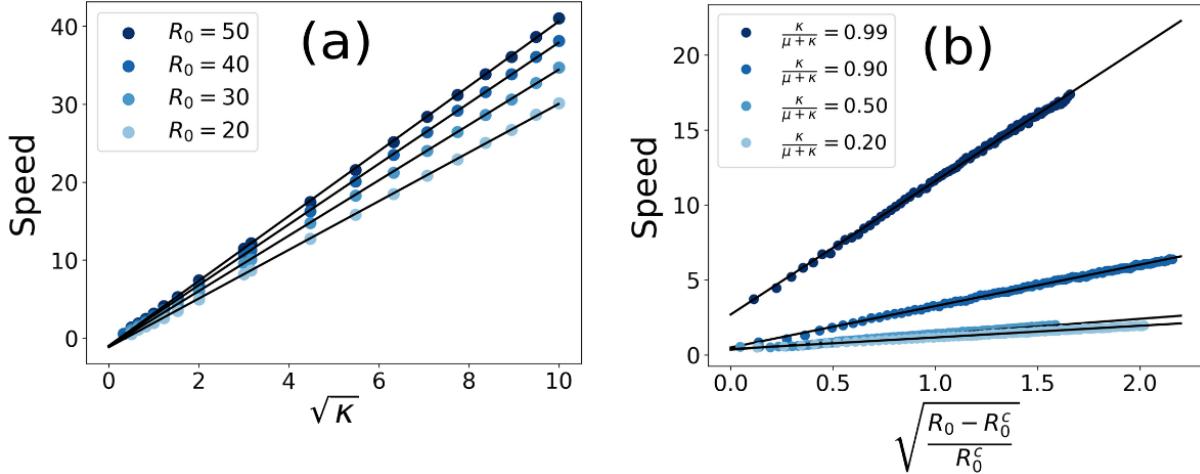
Fig. 4.4(a) shows the numerical results of the computed transition between the disease-free and propagation phases. The heatmap coding represents the average value of absorbing state  $\langle R(\infty) \rangle$  for several values of the mobility factor and  $R_0$ . As expected, the lower the mobility factor is, the higher the value of  $R_0$  is needed for the disease to invade the population. Fig. 4.4(b) shows the fit of Eq. (4.10) with less than a 1% of relative error. Interestingly, we obtain  $B = 1.07 \approx 1$  which validates our expression for the spatial threshold as a first approximation. However, the values for  $A = 4.02$  and  $C = 2.7$  show a significant deviation from Eq. (4.9) and indicate that the expression Eq. (4.8) is an approximation to the spatial basic reproduction number, which however seems to contain the right dependence on  $\kappa/(\mu + \kappa)$ , and where  $A$  could be a geometric factor for a lattice.

#### 4.3.4 Spreading speed of the infected population and time to extinction

Another relevant epidemiological question is how does an infected population spread after the onset of an epidemic. In order to obtain this spreading speed we computed the mean time needed for an infected individual to reach the boundary of the system. More specifically, for each particular choice of the model parameters, 1000 simulations were run for several system sizes ranging from  $L = 10$  to  $L = 60$ . The computed mean time was found to depend linearly with the system size, thus allowing to compute the speed from the slope of this relation. With this procedure, the spreading speed was computed for several values of the parasite mobility and  $R_0$ , large enough to ensure an epidemic outbreak that reached the boundary of the system. In this situation, the spreading speed is expected to depend linearly with the square root of the parasite mobility,

$$v \sim \sqrt{\kappa}. \quad (4.11)$$

**Fig. 4.5(a)** shows this square root dependence for different values of the fixed  $R_0$ . Similarly, the speed was also computed for several values of the basic reproduction number and a fixed mobility. In this case, it varies with the square root of the distance to the critical value of  $R_0$ ,  $R_0^c$ , as shown by **Fig. 4.5(b)**. This is in good agreement with other mathematically similar models (38).



**Figure 4.5:** (a) Disease spreading speed as function of the square root of the parasite mobility for several values of  $R_0$ . The plot shows a remarkable agreement with Eq. (4.11). (b) Disease spreading speed as function of the square root of the distance to the critical value of  $R_0$  for several values of the parasite mobility.

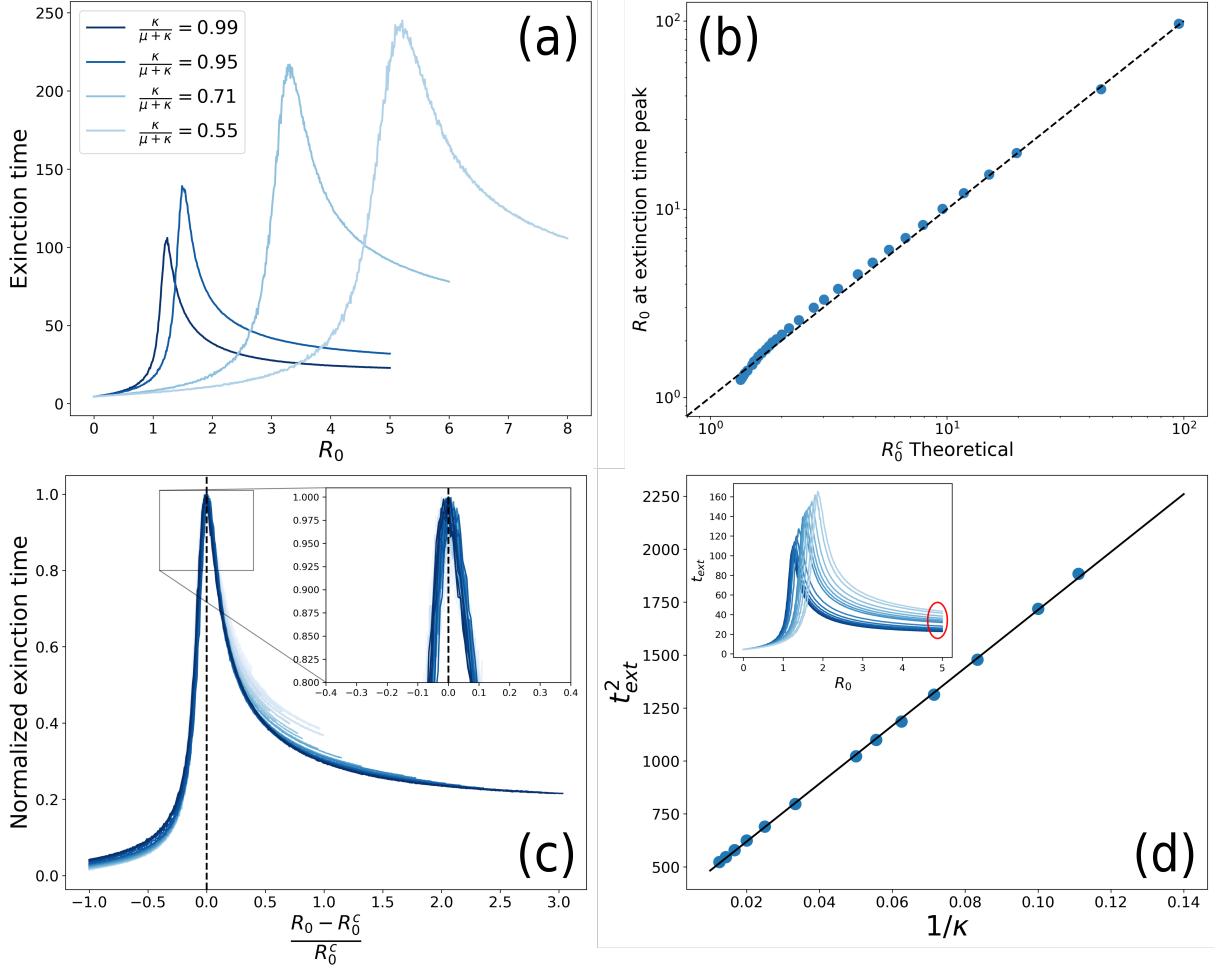
The extinction time is defined as the time elapsed from the beginning of the epidemic until the system reaches its absorbing state, that is, when no parasites or infected individuals are left. From Eq. (4.11), we expect the time to extinction to increase when the parasite mobility is decreased. Moreover, we expect the extinction time to decrease with the distance to the epidemic threshold, as we expect to reach faster the absorbing state for larger values of the spatial basic reproduction number.

In the limiting case where all (or almost all) hosts die, it is clear that the disease must have spread to the entire system. Thus, in this limit, the extinction time should be proportional to the inverse of the disease spreading speed,  $t_{\text{ext}} \sim 1/v$ . Then, in this limit, we can relate the extinction time with the parasite mobility as follows,

$$t_{\text{ext}} \sim \frac{1}{\sqrt{\kappa}} \quad \text{for } \tilde{R}_0 \gg 1 . \quad (4.12)$$

However, the absorbing state is not always reached after all hosts becoming infected and Eq. (4.12) is only expected to work far from the epidemic threshold, when the disease is expected to spread to the entire system.

In Fig. 4.6(a) the extinction time is plotted against the basic reproduction number for some values of the parasite mobility. As expected, the extinction time increases for lower values of the parasite mobility. The increasing behaviour before the peak can be understood as the increasing time needed for the initial perturbations to decay to the disease-free phase. After the peak, the greater the basic reproduction number the faster the epidemic will reach its absorbing state with a non-negligible number of dead individuals. So, with this interpretation, the peaks of the extinction time should coincide with the epidemic threshold for each value of the parasite mobility.



**Figure 4.6:** (a) Extinction time for some values of the parasite mobility. (b) Comparison of the critical  $R_0$  value computed with Eq. (4.8) compared to the values obtained numerically by computing the maximum of the extinction time. (c) Scaling of the extinction time with several values of the parasite mobility. (d) Representation of the square of the extinction time as function of the inverse of the parasite mobility. The inset shows the zone where this relation has been computed, showing a good agreement with Eq. (4.12).

In Fig. 4.6(b) we compare the numerical value of  $R_0$  at which the extinction time peaks with the theoretical value of  $R_0^c$ , computed with Eq. (4.10), showing good agreement. Thus, the dependence of the extinction time with  $R_0$  should vanish if plotted against the distance to  $R_0^c$ . Furthermore, if the extinction time is normalised (dividing each line by its maximum), all the lines should collapse near the transition point. In Fig. 4.6(c) the normalised extinction time is plotted against the distance to the critical value of  $R_0$ . The scaling is shown to be valid only near the transition point, as expected.

In the limiting case where the epidemic dies by infecting a large part of the host population, i.e. for a large enough  $R_0$  value, the extinction time should follow Eq. (4.12), as previously discussed. In Fig. 4.6(d) we show how the extinction time relates to the parasite mobility in this limit, following the predicted behaviour.

## 4.4 Conclusions

In this work we have developed a spatially-explicit individual-based model for parasite-produced marine epidemics of immobile hosts. This study has allowed us to tackle important questions in marine epidemiology, as how spatial constraints affect epidemic spreading in filter-feeder populations or how will the infected population of hosts change in space and time. While addressing the aforementioned questions, we have shown that there exists a regime of high

parasite mobility where the time progression of both host and parasite populations can be well described by the non-spatial version of the model (i.e. the system of ODE's presented in (198)). We have also shown that a fast-slow approximation for the time progression of the parasite population, already presented in (198), can be extended for spatial systems. Interestingly, the conditions under which this approximation is valid are less restrictive than in the non-spatial case, and regimes in which this approximation is valid for low mobility and comparable time scales are reported in this contribution.

We have derived an approximate analytical expression of the *spatial* basic reproduction number, that allows to predict the onset of a global epidemic in a spatial model. The obtained expression explicitly shows a trade-off between the intrinsic pathosystem dynamics (i.e.  $R_0$ ) and a factor accounting for parasite mobility. Moreover, the spatial threshold defined by  $\tilde{R}_0 = 1$  separates the final state of the system in two different phases, namely a disease-free phase and a propagation phase. In the propagation phase, any initial condition of infected individuals or parasites will propagate throughout the system, causing a proper outbreak. On the other hand, in the disease-free phase the conditions are not sufficient for a local introduction of parasites or infected individuals to spread through the system. The effect of the parasite mobility in the spatial basic reproduction number is clear, the more parasites move the more infections they cause.

The spatiotemporal behaviour of the system has been investigated in the propagation phase. First, we showed that the infected population spreads through the space with a speed directly proportional to the square root of the diffusion coefficient of parasites, showing good agreement between the derived analytical expression and numerical simulations. The time to extinction has been also studied by means of numerical simulations, showing that, if the system is far above enough of the spatial threshold, the time to extinction can be analytically computed, in good agreement with simulations. We obtained that larger values of the parasite mobility yield more severe epidemics in which there are more infections and the extinction is faster.

To summarise, in the present work we have introduced and analysed and Individual-Based approach to epidemic transmission in spatially extended systems of immobile hosts. The infection mechanism is due to mobile parasites, that are in turn produced by infected hosts. The study allows to answer some biologically relevant questions, like predicting the occurrence of a global epidemic outbreak or its velocity of expansion through the system. Thus, the analytical and computational results of the model shed light on the underlying mechanisms underpinning the emergence of a global epidemic outbreak and its spatial progression. This work provides a first step into the spatial-explicit individual-based modelling of marine epidemics of immobile hosts.

Although this work has considered the case of a spatially homogeneous distribution of hosts, we plan to extend the study to more general cases, discussing the effect of inhomogeneous spatial host distributions. Furthermore, other biological relevant effects could be added to the model to enhance the description of different epidemics, e.g. infected individuals could still filter parasites or parasite-load dependent infection process. The model could also describe epidemics on other immobile species such as filter feeders like sponges or other bivalves, corals, intertidal communities or starfishes provided that the necessary modifications in the model are properly included. Stochastic spatially-explicit descriptions like the one presented here could be also extended to the study of epidemics of other immobile hosts, like vector-borne diseases of plants. However, this would imply a quite different model to describe the different epidemic compartments of the vectors and also their ecological features. We hope these studies can be useful in conservation plans or ecosystem management and could serve as a basis for more sophisticated models.

## 4.5 Appendix

### 4.5.1 Derivation of the non-spatial equation for $R_\infty$

The model described by the ODE system in Eq. (4.2) has a conserved quantity  $\mathcal{C}$  given by (198).

$$\mathcal{C} = P + \frac{\lambda}{\gamma} (S + I) - S - \frac{\mu}{\beta} \ln S \quad (4.13)$$

At  $t = \infty$  the system reaches an absorbing state completely determined by  $S(\infty)$ , as  $P(\infty) =$

$I(\infty) = 0$  and  $N = S(\infty) + R(\infty)$ . Thus, from Eq. (4.13) we have

$$S(\infty) \left( \frac{\lambda}{\gamma} - 1 \right) - \frac{\mu}{\beta} \ln(S(\infty)) = \mathcal{C}_0 \quad (4.14)$$

The transcendental equation Eq. (4.14) can be solved by means of the Lambert's W function,

$$S(\infty) = -\frac{\mu\gamma}{\beta(\lambda-\gamma)} W_0 \left( -\frac{\beta(\lambda-\gamma)}{\mu\gamma} \exp(-\beta\mathcal{C}_0/\mu) \right) \quad (4.15)$$

which can be simplified to

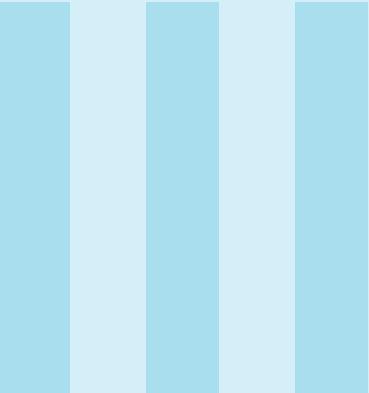
$$S(\infty) = -\frac{S(0)}{\xi} W_0 \left( -\xi \exp \left( -\frac{\beta}{\mu} C \right) \right), \quad (4.16)$$

with  $\xi = S(0) \frac{\beta(\lambda-\gamma)}{\mu\gamma}$  and  $C = P(0) + \frac{\lambda}{\gamma} (S(0) + I(0)) - S(0)$ .

Finally, the absorbing state fulfils the condition  $N = S(\infty) + R(\infty)$  so that the final number of dead individuals can be expressed as

$$R(\infty) = N + \frac{S(0)}{\xi} W_0 \left( -\xi \exp \left( -\frac{\beta}{\mu} C \right) \right). \quad (4.17)$$

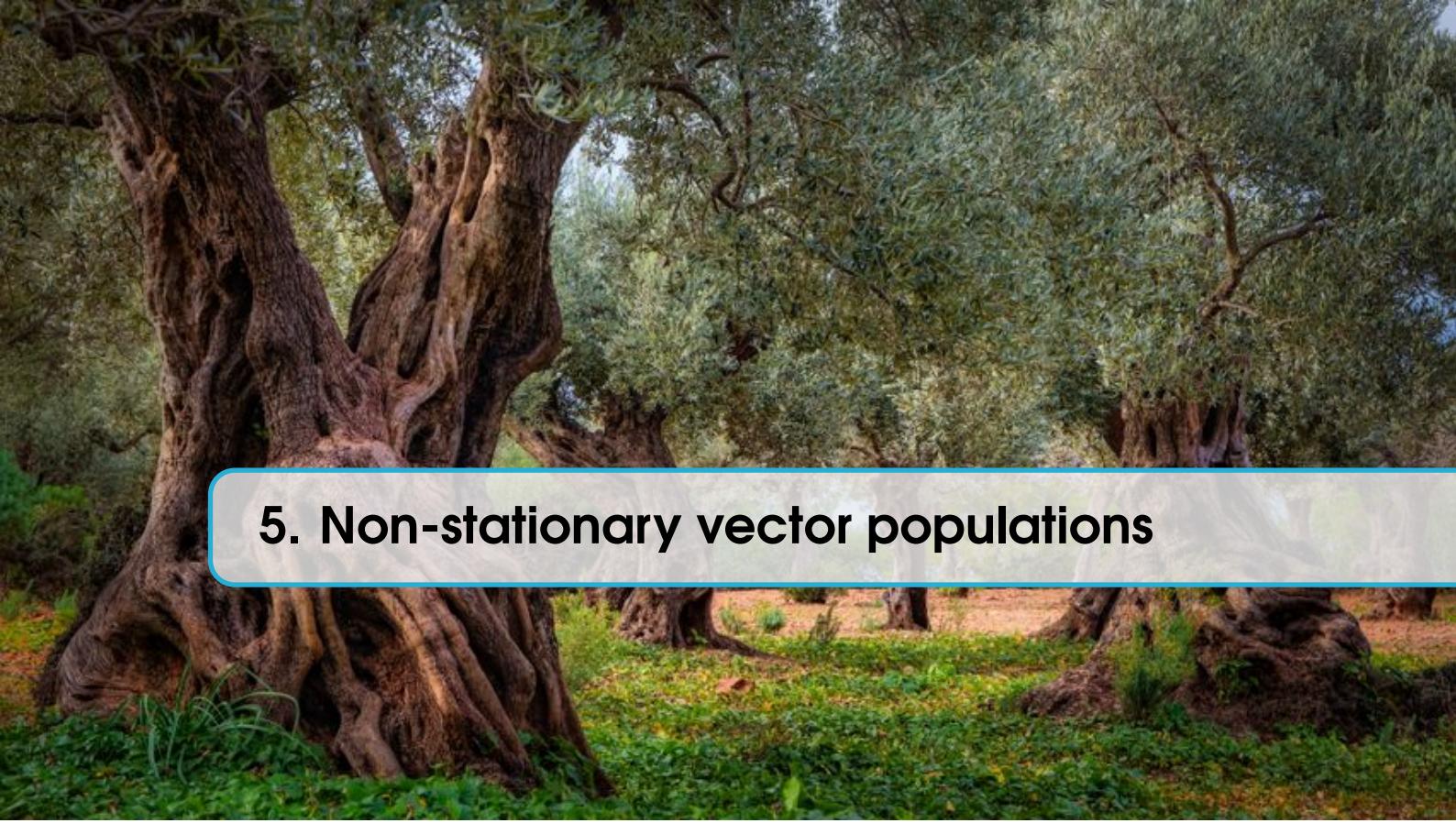




# Modelling vector-borne plant diseases

<b>5</b>	<b>Non-stationary vector populations .....</b>	<b>45</b>
5.1	Introduction .....	46
5.2	The model .....	47
5.3	Results .....	49
5.4	Conclusions .....	55
5.5	Appendix .....	56
<b>6</b>	<b>A compartmental model for <i>Xylella fastidiosa</i> diseases .....</b>	<b>59</b>
6.1	Introduction .....	60
6.2	Materials and Methods .....	61
6.3	Results .....	64
6.4	Discussion .....	70
6.5	Appendix .....	72





## 5. Non-stationary vector populations

Àlex Giménez-Romero<sup>1</sup>, Rosa Flaquer-Galmés<sup>1,2</sup>, Manuel A. Matías<sup>1</sup>

1. Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Palma de Mallorca 07122, Spain
2. Grup de Física Estadística, Departament de Física. Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

### Published as

Àlex Giménez-Romero, Rosa Flaquer-Galmés, and Manuel A. Matías. "Vector-borne diseases with nonstationary vector populations: The case of growing and decaying populations". In: *Phys. Rev. E* 106 (5 Nov. 2022), page 054402. DOI: [10.1103/PhysRevE.106.054402](https://doi.org/10.1103/PhysRevE.106.054402). URL: <https://link.aps.org/doi/10.1103/PhysRevE.106.054402>

## 5.1 Introduction

Vector-borne diseases are caused by infectious agents transmitted by living organisms, called vectors, frequently insects. These diseases represent a significant threat to global human health (23), causing diseases such as malaria, dengue, yellow fever, Zika, trypanosomiasis and leishmaniasis (392). Vector-borne human diseases are responsible of more than 17% of all human infectious diseases, causing millions of cases and more than 700 000 deaths annually (450). Moreover, crop production and farm profitability are also affected by bacterial (235) and virus (60) vector-borne diseases. Some examples are the Pierce's Disease of grapevines, that has resulted in an annual cost of approximately \$100 million in California alone (424), the olive quick decline syndrome, which could cause about 5 and 17 billion US\$ of losses in Italy and Spain over the next 50 years in the absence of disease control measures (391) and the multiple diseases caused by viruses (374), with diseases like the tobacco mosaic, tomato spotted wilt, etc., transmitted by aphids and other vectors.

Compartmental deterministic models, e.g. the well known SIR model (261), have been widely used in the modeling of vector-borne diseases after the seminal work of Ross and Macdonald (290), that opened the way to controlling malaria outbreaks by acting on the vectors of the disease (the *Anopheles* mosquito). These models consider that both host and vector populations can be divided into different compartments describing different states of the individuals, such as susceptible, infected or removed (recovered or dead) (62), and the time-evolution of these compartments is expressed as a system of ordinary differential equations, defining a dynamical system. Compartmental models provide a mean-field description, that imply well-mixed (in practice spatially homogeneous) populations. The well mixed approximation will be valid whenever the mean distance among hosts is smaller than the mixing length of vectors before they die. In the case of vector-borne diseases it is also equivalent to every vector effectively interacting with all the hosts and every host with all the vectors. A mean-field description is not always valid in spatially extended systems, but still it is often the first step before writing a spatially explicit description.

The most relevant piece of information about a disease is whether an epidemic outbreak will take place or not. The *basic reproduction number*,  $R_0$ , measures the number of secondary infections caused by an initial infected individual in a fully susceptible population, defining the epidemic threshold (18, 428), that determines the emergence (or not) of an outbreak. If  $R_0 > 1$  an epidemic outbreak will occur, while there will be no outbreak otherwise. The standard way of computing  $R_0$  in deterministic compartmental models is based on the existence of an initial disease-free (pre-pandemic) equilibrium, represented by the absence of infected hosts and vectors (249, 274). Some standard methods based on the linear stability condition of this equilibrium have been developed to allow the direct computation of  $R_0$ , such as the Next-Generation Matrix (NGM) method (138).

In the case of vector-borne diseases, some models assume that populations (both hosts and vectors) do not change with time (see e.g. (63, 290, 427)), thus assuming equal birth and death rates. This guarantees the existence of a disease-free equilibrium and the proper use of standard methods to determine  $R_0$ . However, this assumption could be far from reality in several pathosystems. For example, the interaction between temperature, precipitation variations and other factors may lead to strong variations in the vector population (176, 370), implying that the pre-pandemic state may not be an equilibrium state and that standard methods cannot be applied.

Compartmental models of vector-borne diseases have another feature that may hinder their practical applicability. Namely the fact that these models have many compartments, that describe the different states of both hosts and vectors, and as a consequence a relatively large number of parameters. This may lead to an issue known as *parameter identifiability and uncertainty* (101), depending on the available data, that is more likely to be found in models with many compartments and parameters (372). Usually, parameter estimation procedures are needed to connect the models with disease data, mainly using incidence or prevalence over time in the host population. Unfortunately, under many circumstances the underlying model parameters are unidentifiable, so that many different sets of parameter values produce the same model fit (250). Moreover, these parameters can be really difficult to determine from the available experimental data. Nevertheless, in some cases, mathematical manipulations can

be performed to reduce the model complexity using exact or approximate relations (198). In such cases, the number of parameters of the models can be usually reduced in terms of new parameters defined as combinations of some of the original parameters.

The plan of this paper is as follows. In Section 5.2 we develop a compartmental model of vector-borne transmitted diseases with constant, but different, birth and death rates for the vectors, that we will use to describe the case of growing and decaying vector populations. For simplicity, the model assumes that there is no host to host direct transmission and that the development of the disease is faster than host recruitment, which is also a realistic assumption in many cases, like plant diseases. Section 5.3 contains the main results of the study. In particular, we show that the asymptotic approach fails to estimate the  $R_0$  of the model, overlooking outbreaks if some conditions are fulfilled. Here, we provide an alternative method to compute  $R_0$  based on the average number of secondary host infections produced by a primary infected host in one generation. It turns out that the validity of the asymptotic approach depends, among other things, on some time-scales of the model. Furthermore, we discuss and apply some approximations that allow to reduce the model in favor of simpler ones, with both fewer compartments and fewer parameters. In particular, we show that if some of the parameters fulfill certain conditions, it is possible to reduce the original model with five compartments and four parameters to a SIR model, with three compartments and two parameters. It is expected that model reductions like this one significantly help in solving possible problems of parameter unidentifiability that plague these models. It is interesting to note that a model in which hosts do not interact directly, but only through vectors, in a certain limit becomes described as if hosts would infect directly one to each other, what is assumed in some studies without suitable confirmation. Finally, the main concluding remarks of the study are presented in Section 5.4.

## 5.2 The model

The compartment model for vector-borne diseases that we will use to illustrate the points to be discussed in this study consists of 5 compartments, 3 of which describe the host population (susceptible,  $S_H$ , infected,  $I_H$ , and removed,  $R_H$ ), while the other 2 describe the vector population (susceptible  $S_V$  and infected vectors,  $I_V$ ). Thus, we consider that the pathogen affects only the hosts and do not consider exposed compartments. In addition, no direct host to host or vertical (or mother to offspring for vectors) transmission is assumed. The model could be also generalized to include an exposed host compartment and the above mentioned transmission modes, which would hinder the theoretical analysis without altering the qualitative conclusions of the study. Finally, for the host population we consider neither recruitment nor natural death and then the total host population,  $N_H$ , is constant,  $N_H = S_H + I_H + R_H$ . Finally, we assume that infected hosts do not have a mechanisms to combat the disease and become susceptible again. These assumptions are reasonable in the case of many phytopathologies.

The model is defined according to the following processes,



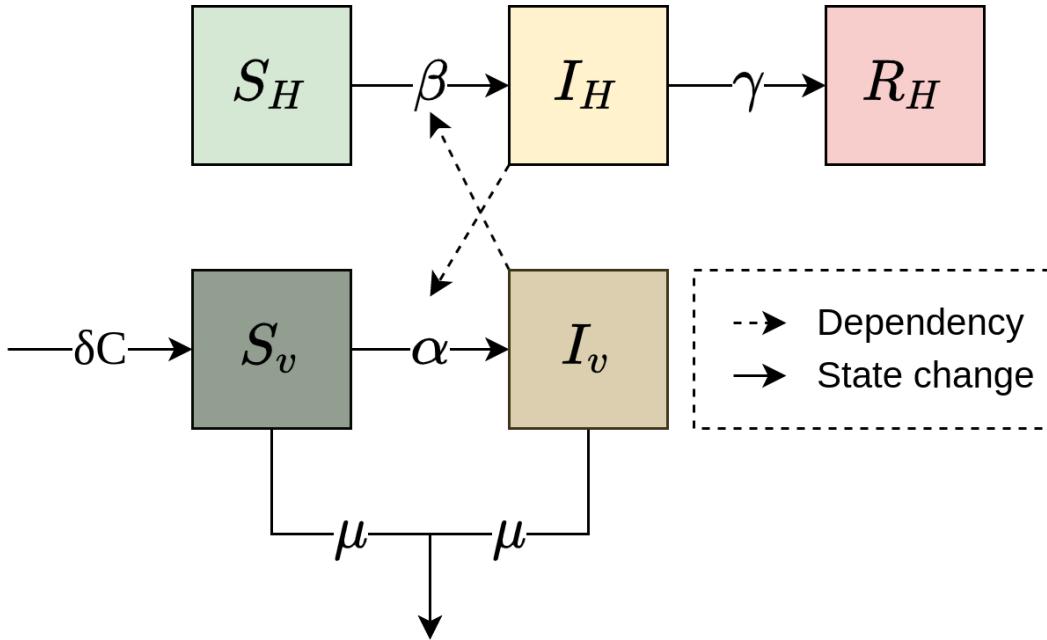
which are graphically described in Fig. 5.1, being the birth of new susceptible vectors described as a source term. Thus, the host-vector compartmental model is written as,

$$\begin{aligned} \dot{S}_H &= -\beta S_H I_V / N_H \\ \dot{I}_H &= \beta S_H I_V / N_H - \gamma I_H \\ \dot{R}_H &= \gamma I_H \\ \dot{S}_V &= \delta C - \alpha S_V I_H / N_H - \mu S_V \\ \dot{I}_V &= \alpha S_V I_H / N_H - \mu I_V, \end{aligned} \tag{5.2}$$

where the crossed nonlinear terms,  $S_x I_y$  are written divided by the total host population,  $N_H$ , what corresponds to the so called standard incidence, which differs of the purely bilinear form known as mass action incidence (297).

The model describes infection of susceptible hosts,  $S_H$ , at a rate  $\beta$  through their interaction with infected vectors,  $I_V$ , while susceptible vectors,  $S_V$ , are infected at a rate  $\alpha$  through their

interaction with infected hosts  $I_H$ . Infected hosts exit the infected compartment at rate  $\gamma$ , while infected vectors stay infected the rest of their lifetime, as we consider that the pathogen does not affect them, as it is customary. Vectors die naturally (or disappear from the population by some mechanism) at rate  $\mu$  and are born (appear) at a constant rate  $\delta$  being susceptible. The constant term  $C$  sets the scale of the stationary value of the vector population. Fig. 5.1 shows an schematic representation of the model and we refer to (63) for a similar model of vector-borne diseases. However, the model in (63) includes exposed compartments and direct host to host transmission, but assumes that the birth and death rate of vectors are identical, and thus, the population does not change with time and stays as fixed by the initial condition.



**Figure 5.1:** Schematic representation of the model n Eq. (5.2). Boxes are the compartments in which the population is divided, solid arrows represent changes in state (so transitions between compartments), and dashed arrows depict the crossed interaction between hosts and vectors.

### 5.2.1 Preliminary analysis of the model

From Eq. (5.2) it is straightforward to verify that the population of hosts remains constant over time,  $N_H = S_H + I_H + R_H$ , while the vector population fulfills,

$$\dot{N}_v = \dot{S}_v + \dot{I}_v = -\mu(S_v + I_v) + \delta C = -\mu N_v + \delta C, \quad (5.3)$$

with solution,

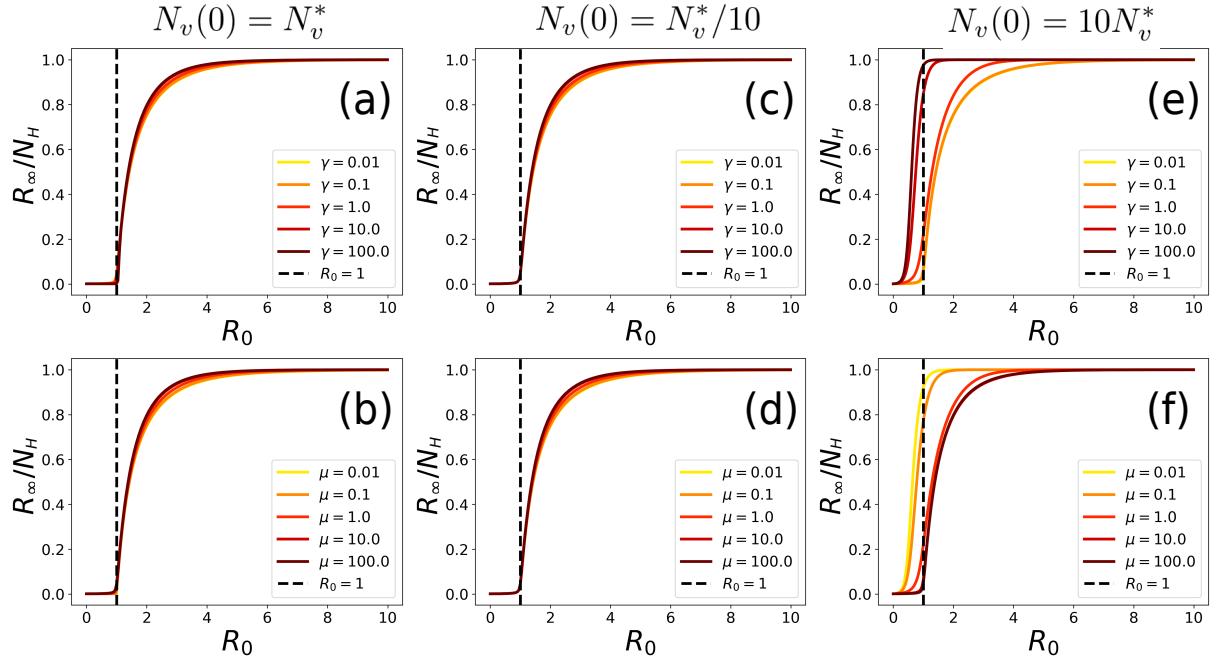
$$N_v(t) = \frac{\delta}{\mu}C + \left(N_v(0) - \frac{\delta}{\mu}C\right)e^{-\mu t}. \quad (5.4)$$

From Eq. (5.4) one can write the stationary value for the vector population,  $N_v^*$ ,

$$N_v^* = \lim_{t \rightarrow \infty} N_v(t) = \frac{\delta}{\mu}C. \quad (5.5)$$

Thus, if the initial population of vectors is below (above) the stationary value, the vector population will grow (decrease) until it reaches the stationary value. On the other hand, if  $N_v(0) = N_v^* = \delta C / \mu$  the initial population of vectors is already at the stationary state. The initial condition for the vector population can be written in terms of its stationary value Eq. (5.5),  $N_v(0) = fN_v^*$ , where both  $f < 1$  and  $f > 1$  are possible, so that one gets,

$$N_v(t) = N_v^* [1 + (f - 1)e^{-\mu t}]. \quad (5.6)$$



**Figure 5.2:** Numerical verification of the predictive power of the basic reproduction number relation Eq. (5.7), by plotting the final size of the epidemic,  $R_\infty/N_H$  as function of  $R_0$ . In panels (a),(b) the initial vector population is in the stationary value, in panels (c),(d) is below,  $N_v^*/10$ , and in panels (e),(f) above,  $10N_v^*$ . Panels (a),(c),(e) show realisations for different  $\gamma$  values with a fixed  $\mu = 1$  baseline value. Panels (b),(d),(f) show realisations for different  $\mu$  values with a fixed  $\gamma = 1$  baseline value.

We note that vector-borne disease models that assume constant vector populations (e.g.(63)) can be recovered by setting  $\delta = \mu$  and  $C = N_v(0)$ , so that any initial condition for the vector population is stationary, i.e.  $\dot{N}_v = 0$  in Eq. (5.3) and  $N_v(t) = N_v(0)$ . We note that our model describes populations with an asymptotic stationary vector population and cannot describe periodic vector populations.

## 5.3 Results

### 5.3.1 The effect of non-stationary vector populations into the epidemic threshold and disease dynamics

Let us start with the cases in which any initial condition for the vector population is stationary and the total vector population remains unchanged. This will happen when the birth  $\delta$  and death  $\mu$  vector rates are identical, independently on the initial condition of the vector population, or the case in which the initial condition of the vector population is already at its stationary value,  $N_v(0) = N_v^*$ , independently of the values of  $\delta$  and  $\mu$ . In such a case, the initial disease free state of the model, given by  $I_H(0) = I_v(0) = 0$ , is a fixed point (equilibrium state) of the dynamical system Eq. (5.2) independently of the other initial conditions for the host and vector populations. This allows the definition of the basic reproduction number,  $R_0$ , using standard methods such as linear stability analysis or the Next-Generation Matrix (NGM) method (138) (see 5.5.1).

In other cases, the total vector population will vary with time provided that the initial condition,  $N_v(0)$ , is not identical to the asymptotic value at large times,  $N_v^*$ . In these cases, an initial disease-free state is not an equilibrium (fixed point) of the model. However, in the literature it is customary to apply the standard techniques, i.e. NGM, to compute  $R_0$  using the vector population in the asymptotic state, that is the post-pandemic disease-free equilibrium (152, 273, 397, 442, 458). The use of these methods is supported by the fact that the asymptotic dynamics of the model converges to the dynamics of the subsystem where the vector population is stationary (90, 418).

In both cases the basic reproduction number is given by,

$$R_0 = \frac{\beta\alpha}{\mu\gamma} \frac{S_H(0)}{N_H^2} N_v^* . \quad (5.7)$$

As usual,  $R_0$  accounts for the number of secondary infections produced by an infected individual in one generation and controls the threshold behavior of the model: for  $R_0 < 1$  the epidemic dies out and for  $R_0 > 1$  an outbreak occurs. By one generation we refer to the typical time in which new infections can be produced, being the generation time in our model,

$$t_g = 1/\gamma + 1/\mu . \quad (5.8)$$

Now we will show that Eq. (5.7) is not always predictive about the onset of the epidemic. In Fig. 5.2 the final size of the epidemic,  $R_\infty/N_H$ , is plotted as a function of  $R_0$ , where  $R_\infty$  is the number of dead individuals at the end of the epidemic. Fig. 5.2(a)-(d) show that Eq. (5.7) does indeed regulate the onset of an epidemic when the initial vector population is in its stationary value or below it. This result is general and does not depend on the time-scales of the system,  $1/\gamma$  and  $1/\mu$ , and so all curves in these panels behave similarly. In contrast, Fig. 5.2(e)-(f) shows that Eq. (5.7) does not predict the onset of epidemic outbreak when the initial vector population is larger than the stationary value. Thus, for  $R_0 < 1$  (computed using Eq. (5.7)) severe outbreaks appear, yielding mortalities even above 80% of the total population. However, one can observe that as  $\mu$  is increased, or  $\gamma$  decreased, the predictive power of Eq. (5.7) is progressively recovered.

Thus, only if the vector population reaches its stationary value before infected hosts have produced new infections can the onset of an epidemic be characterized by Eq. (5.7). Let us discuss separately the cases  $f > 1$  and  $f < 1$ , with  $N_v(0) = fN_v^*$ , namely when the initial vector population is above and below its stationary value, this is, decaying and growing vector populations towards the asymptotic state.

If  $f > 1$  Eq. (5.6), the time to approach the stationary value,  $t^*$ , is,

$$(1+\varepsilon)N_v^* = N_v^* \left[ 1 + (f-1)e^{-\mu t^*} \right] , \quad (5.9)$$

where  $\varepsilon \rightarrow 0$  is a small parameter controlling the amount by which the vector population differs from its asymptotic value at time  $t^*$ . Thus, the time to approach the stationary value, with precision  $\varepsilon$ , is given by

$$t^* = -\frac{1}{\mu} \ln \left( \frac{\varepsilon}{f-1} \right) = \frac{1}{\mu} \left| \ln \frac{\varepsilon}{f-1} \right| , \quad (5.10)$$

where the last equality assumes that the small parameter  $\varepsilon$  satisfies  $\varepsilon < (f-1) > 0$ .

If the vector population reaches its stationary value before infected hosts have had time to generate new infections then  $R_0$  as determined from Eq. (5.7) is a good prediction of the onset for an epidemic, what is equivalent to the condition that  $t^*$  is much smaller than the hosts infectious period,  $t^* \ll 1/\gamma$ ,

$$\frac{1}{\gamma} \gg \frac{1}{\mu} \left| \ln \frac{\varepsilon}{f-1} \right| \quad \text{or} \quad \frac{\mu}{\gamma} \gg \left| \ln \frac{\varepsilon}{f-1} \right| . \quad (5.11)$$

Otherwise, Eq. (5.7) will not be predictive of the epidemic onset, and as shown in Fig. 5.2(e-f) one may have outbreaks with a substantial final size with  $R_0 < 1$ .

In the case of growing vector populations,  $f < 1$ , if  $R_0 < 1$  an outbreak cannot occur at all, because  $R_0$  is calculated with the asymptotic population,  $N_v^*$ , that is larger than the vector population at any finite time,  $N_v(t) < N_v^* \forall t$ , and so the threshold condition is never attained. In the  $R_0 > 1$  case the behaviour will be richer, and it will depend on the initial condition,  $N_v(0)$ . One can define an instantaneous basic reproductive number,

$$R_0^{(i)}(t) = \frac{\beta\alpha}{\mu\gamma} \frac{S_H(0)}{N_H^2} N_v(t) = R_0 \frac{N_v(t)}{N_v^*} , \quad (5.12)$$

using  $N_v(t)$  instead of  $N_v^*$ , with  $R_0^{(i)}(t) < R_0 \forall t$  because the vector population grows. In particular, if  $R_0^{(i)}(0) > 1$  there will be an outbreak occurring for short times, and the population of infected

hosts will start growing. If instead,  $R_0^i(0) < 1$ , and as  $R_0 > 1$  with  $R_0$  being calculated with the asymptotic state, there must be an intermediate time, say  $t_D$ , for which  $R_0^{(i)}(t_D) = 1$ . Thus, from  $t > t_D$  an outbreak will occur, not initially but after a finite time, that induces a delay in the outbreak, and the infected host population will start growing.

The difference between the original and the delayed dynamics stems from the waiting time to reach  $R_0^{(i)} = 1$ ,  $t_D$ , plus the non-linear effect associated to a new initial condition for the epidemic outbreak at  $t_D$ . Thus, in the case that  $R_0 > 1$  and  $R_0^i(0) < 1$ , from Eq. (6.5) and Eq. (5.6) we can analytically approximate the delay as the time needed to reach  $R_0^{(i)}(t_D) = 1$ ,

$$1 + (f - 1)e^{-\mu t_D} = \frac{1}{R_0}, \quad (5.13)$$

which yields the relation,

$$t_D = -\frac{1}{\mu} \ln \left[ \frac{1 - R_0}{(f - 1)R_0} \right], \quad (5.14)$$

where the argument of the logarithm is always positive because  $R_0 > 1$  and  $f < 1$ . Eq. (5.14) is only valid if  $f < 1/R_0$ , for  $R_0^{(i)}(0) = fR_0 < 1$ , as if otherwise  $R_0^{(i)} > 1$  the outbreak would already occur initially.

From Eq. (5.14) one can see that when the initial vector population is far enough from its stationary value,  $f \rightarrow 0$ , the delay saturates to a constant value, instead of increasing. This is,

$$\lim_{f \rightarrow 0} t_D = \frac{1}{\mu} \ln \left( \frac{R_0}{R_0 - 1} \right). \quad (5.15)$$

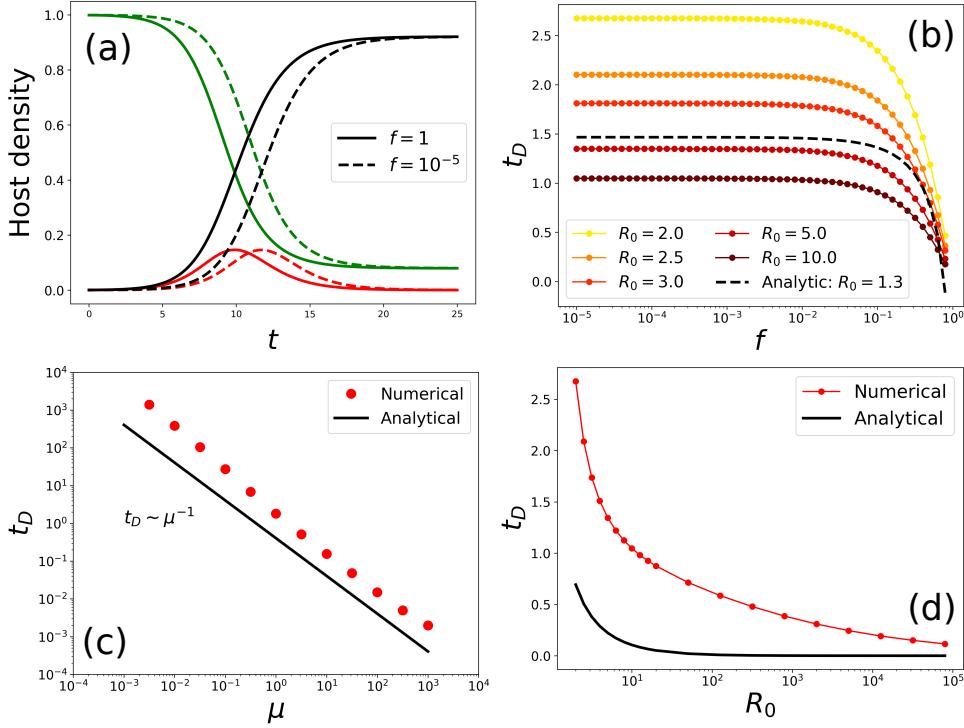
In addition, for increasing values of the basic reproduction number,  $R_0$ , the delay tends to vanish, and from Eq. (5.15). This is,

$$\lim_{R_0 \rightarrow \infty} t_D = \frac{1}{\mu} \ln(1) = 0, \quad (5.16)$$

where the limit  $f \rightarrow 0$  is taken simultaneously to guarantee that  $R_0^{(i)}(0) = fR_0 < 1$ . On the other hand the delay,  $t_D$ , scales with the vectors lifetime,

$$t_D \sim \frac{1}{\mu} = \tau_v. \quad (5.17)$$

Fig. 5.3(a) shows an example of the time delay caused in the hosts dynamics when the vector population grows from an initial condition far from the stationary value. In Fig. 5.3(b) we can qualitatively observe that all the predicted properties of the delay are fulfilled, namely, the time delay saturates for low  $f$  values and decreases with increasing  $R_0$ . Although the analytical expression (black dashed line) is clearly not exact due to nonlinear effects, Eq. (5.14) captures the basic trends of the time delay,  $t_D$ . This is clear from Fig. 5.3(c), that shows that the delay scales with  $1/\mu$  and in Fig. 5.3(d) that shows that the delay tends to 0 in the limit  $R_0 \rightarrow \infty$ , in agreement with the prediction of Eq. (5.16).



**Figure 5.3:** Numerical study of the delay induced by growing vector populations. (a) Comparison of hosts dynamics for a stationary vector population ( $f = 1$ ) and a growing vector population ( $f = 10^{-5}$ ). (b) Time delay as function of  $f$  for different values of the basic reproduction number  $R_0$ . (c) Time delay as function of the vector natural death rate. (d) Time delay as function of the basic reproduction number,  $R_0$ , with  $f = 10^{-5}$ .

### 5.3.2 The basic reproduction number for non-stationary vector populations

As shown in the previous section, traditional methods to compute the basic reproduction number fail in the case of epidemic models with decaying vector populations,  $f > 1$ , unless the time scale of vector population fulfills the strong inequality condition Eq. (5.11), as illustrated in Section 5.3.1. Here we introduce an effective, average definition of  $R_0$ , useful to predict the epidemic onset for vector-borne diseases with decaying vector populations, i.e. the case where traditional methods fail. It is defined as the average number of infections produced by an infected individual in one generation Eq. (5.8),

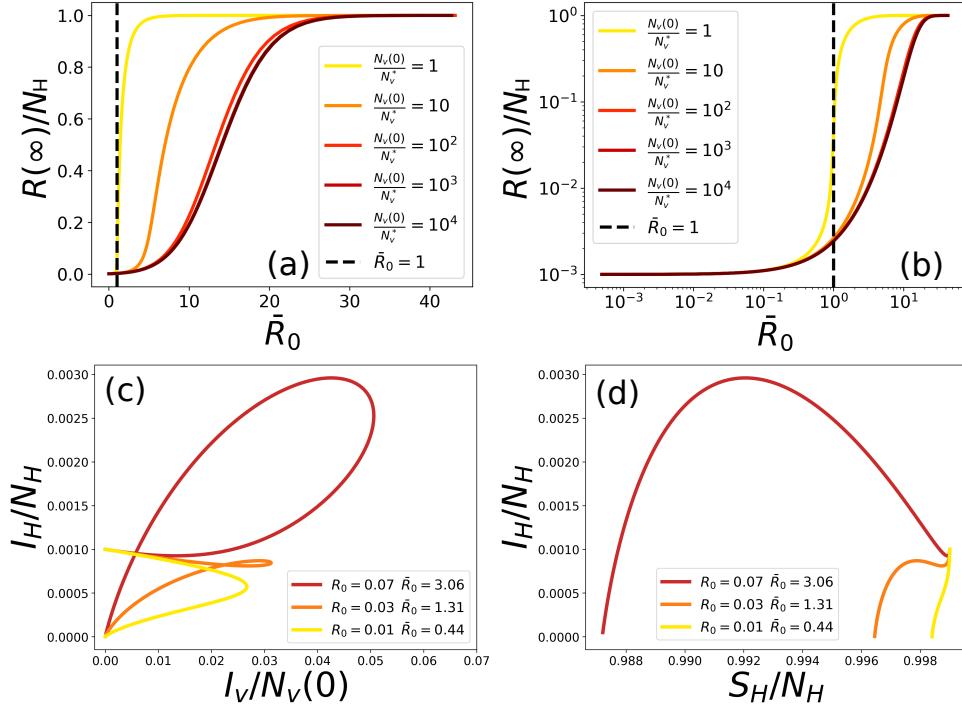
$$\overline{R}_0 = \langle R_0^i(t) \rangle \Big|_{0}^{t_g} = R_0 \left[ 1 - \frac{1}{\tau} (f-1) (e^{-\tau} - 1) \right] = R_0 \cdot \mathcal{F} \quad (5.18)$$

where  $\tau = 1 + \mu/\gamma$  and  $\mathcal{F}$  accounts for the effect of the decaying vector population on the stationary  $R_0$  (see Section 5.5.2 for the full derivation of Eq. (5.18)).

A first observation is that  $\overline{R}_0 > R_0$  always (for  $f > 1$ ). This stems from the fact that  $\tau = 1 + \mu/\gamma > 1$ , so that  $e^{-\tau} - 1 < 0$ , and  $f-1 > 0$ , which yields  $\mathcal{F} > 1$ . This discussion unravels why standard methods fail to predict the onset of an epidemic under decaying vector populations. Another important point is that if  $\mu/\gamma \gg 1$ , which implies  $\tau \gg 1$ ,

$$\lim_{\tau \gg 1} \mathcal{F} = \lim_{\tau \gg 1} \left[ 1 - \frac{1}{\tau} (f-1) (e^{-\tau} - 1) \right] = 1 + \frac{f-1}{\tau}, \quad (5.19)$$

and if furthermore  $\tau \sim \frac{\mu}{\gamma} \gg (f-1)$  then  $\mathcal{F} \rightarrow 1$  and  $\overline{R}_0 \rightarrow R_0$ . This is in agreement with the discussion in Section 5.3.1 showing that the  $R_0$  computed from standard methods works if  $\mu \gg \gamma$ .



**Figure 5.4:** Numerical verification of the expression for the basic reproduction number for vector-borne diseases with decaying vector populations Eq. (5.18). Final size of the epidemic as a function of the basic reproduction number in panels: (a) linear scale; (b) logarithmic scale. Phase space trajectories in panels: (c)  $I_H/N_H$  vs  $I_v/N_v(0)$  and (d)  $I_H/N_H$  vs  $S_H/N_H$ , where an initial condition  $I_H(0)/N_H = 0.01, S_H(0)/N_H = 0.99$  and  $I_v(0)/N_v(0) = 0$  has been used for the 3 cases.  $\mu = \gamma$  has been used in all the simulations.

Fig. 5.4(a-b) contrasts numerically the validity of Eq. (5.18) to predict the final size of the epidemic as a function of the general basic reproduction number,  $\bar{R}_0$ , in linear and logarithmic scale, respectively. We observe that, independently of the initial condition of vectors, the outbreak occurs for  $\bar{R}_0 > 1$ . However, we may notice that for large values of the initial condition of vectors the final size of the epidemic grows more slowly, so that larger values of  $\bar{R}_0$  are needed to produce a proper outbreak. This can be explained by the fact that for  $\bar{R}_0$  slightly above the threshold,  $\bar{R}_0 = 1$ , and large values of  $f = N_v(0)/N_v^*$ , infections are produced only in the transient period of the dynamics, as  $R_0 < 1$ . This is, while the vector population is decaying to its stationary value, the vectors are able to produce new infections, but once the vector population reaches the stationary value, the epidemics stops. This transmission mechanism is radically different to that of vector-borne diseases with stationary vector populations in which the pre-pandemic disease-free state is an equilibrium of the system. The phase-space plots in Fig. 5.4(c-d) show that the time-averaged basic reproduction number  $\bar{R}_0$  is able to accurately predict the conditions under which the infected host population will grow, in contrast with  $R_0$  computed in the post-pandemic fixed point. In essence, for  $\bar{R}_0 > 1$  the infected host population,  $I_H$ , grows before reaching the absorbing state,  $I_H = I_v = 0$ , while for  $\bar{R}_0 < 1$  the infected host population is monotonically decreasing. We note that Eq. (5.18) is similar to the time-averaged basic reproduction number presented in (445) for the periodic case, which is a first-order approximation to the true basic reproductive number (26).

### 5.3.3 Fast-slow approximation

The original 5-D Eq. (5.2) model is certainly not amenable to mathematical analyses due to its high phase-space dimensionality and the fact that it depends on 4 parameters. Moreover, in a real-case application, if the parameters conforming the model are not known, the model could suffer from parameter unidentifiability. However, some approximations can be performed to reduce the mathematical complexity of the model, as for instance a fast-slow (or adiabatic) approximation.

If the time-scale of the vector population evolution is much faster than that of the infected hosts, what is expected to be a good approximation in many practical cases, the vector population will almost instantaneously adapt to its stationary value. Thus, if  $1/\mu \ll 1/\gamma$ , or equivalently if  $\gamma/\mu \ll 1$ , we can rewrite the time derivative of the vector infected population as

$$\varepsilon \dot{I}_v = \frac{\alpha}{\mu} S_v \frac{I_H}{N_H} - I_v , \quad (5.20)$$

where time has been re-scaled to  $t' \rightarrow \gamma t$  and  $\varepsilon = \gamma/\mu$  is a small parameter. Then,  $\dot{I}_v$  can be neglected and the infected vector population can be obtained from the relationship,

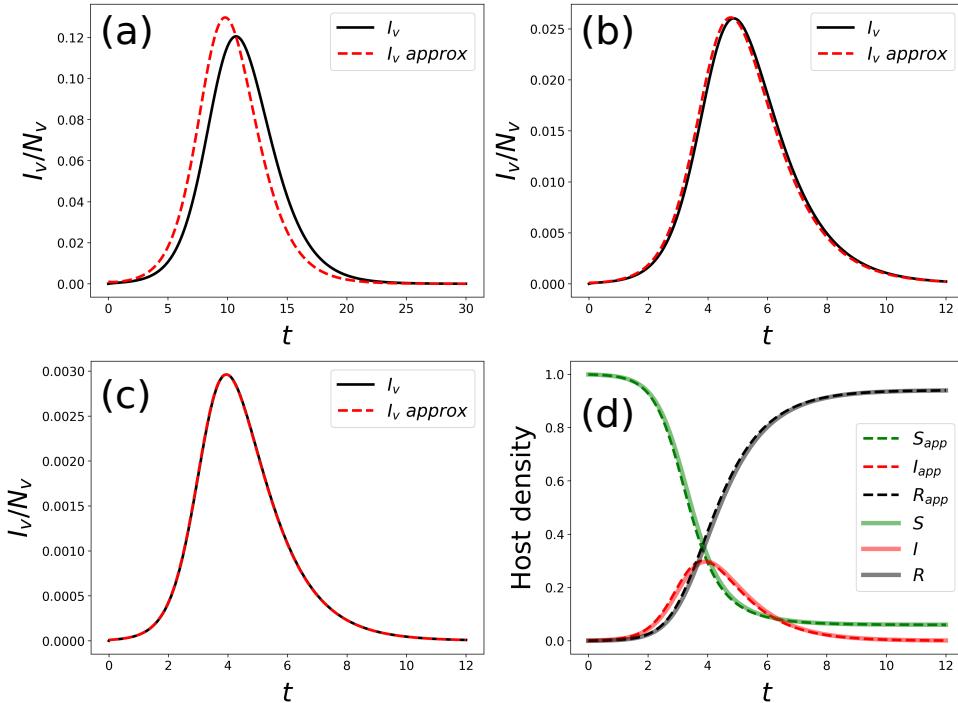
$$I_v \approx \frac{\alpha}{\mu} \frac{S_v I_H}{N_H} . \quad (5.21)$$

Substituting Eq. (5.21) into the original system Eq. (5.2) and the identity  $N_v(t) = S_v(t) + I_v(t)$ , while considering that the conditions for which the time-scale approximation is valid,  $\mu \gg \gamma$ , imply that the vector population will reach its stationary value almost instantaneously, so that  $N_v(t) \approx N_v^*$ , we obtain the following reduced system,

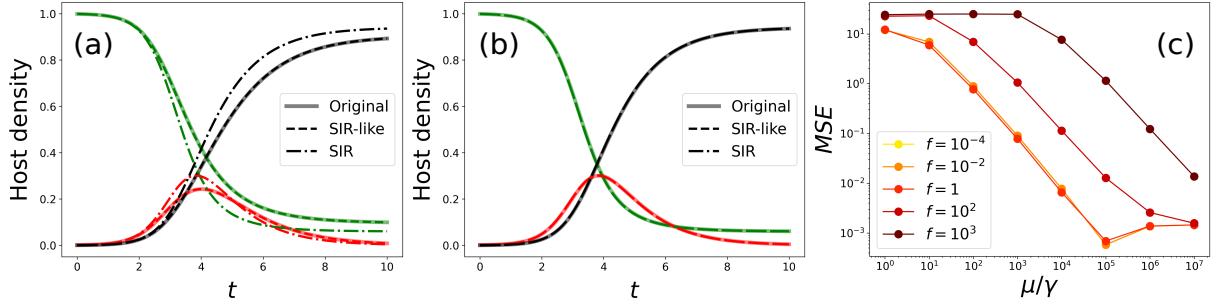
$$\begin{aligned} \dot{S}_H &= -\beta' \frac{S_H I_H}{\lambda N_H + I_H} \\ \dot{I}_H &= \beta' \frac{S_H I_H}{\lambda N_H + I_H} - \gamma I_H \\ \dot{R}_H &= \gamma I_H , \end{aligned} \quad (5.22)$$

where  $\beta' = \beta N_v^*/N_H$  and  $\lambda = \mu/\alpha$ .

Moreover, if  $f \neq 1$  the above mentioned timescales relationship must fulfil  $\frac{\mu}{\gamma} \gg \left| \ln \frac{\varepsilon}{f-1} \right|$  (cf. Eq. (5.11)) and not only  $\frac{\mu}{\gamma} \gg 1$ . It is important to notice that the presence of direct host to host transmission would simply re-scale the coefficient  $\beta'$ , and the SIR reduction Eq. (5.22) would keep its validity.



**Figure 5.5:** Numerical verification of the time-scale approximation (Eq. (5.21)) with  $N_H = 100$ ,  $\alpha = \gamma = 1$ .  $\beta$  is chosen such that  $R_0 = 3$ . (a)  $\mu = 1$ , (b)  $\mu = 10$ , (c)  $\mu = 100$ . Panel (d) shows a comparison between the approximate and original models for the parameters used in (c), where the approximated models is expected to represent well the original one.



**Figure 5.6:** Comparison between the original model and the reductions, Eq. (5.22) (SIR-like) and Eq. (5.23) (SIR) with  $N = 100$ ,  $\mu/\gamma = 10^3$  and  $f = 1$ .  $\beta$  was chosen such that  $R_0 = 3$ . (a)  $\lambda = 1$ , (b)  $\lambda = 10^3$ , (c) Mean Squared Error between the original model and the SIR approximations as function of the ratio  $\mu/\gamma$  and  $f$ .

In Fig. 5.5 we numerically verify the validity of the presented fast-slow approximation. As expected, we observe that the approximation breaks down for  $\mu \sim \gamma$  (Fig. 5.5(a)), while as  $\mu$  becomes larger than  $\gamma$  the approximation improves Fig. 5.5(b) and it becomes quantitative when  $\mu \gg \gamma$ , Fig. 5.5(c). Finally, we show in Fig. 5.5(d) a comparison between the dynamics of the hosts using both the original and the approximated model using the same parameters than in Fig. 5.5(c), where the results of both models are expected to converge.

### 5.3.4 Reduction to a SIR model

In addition to the previous condition,  $\gamma/\mu \ll 1$ , if one has that  $\lambda N_H \gg I_H$  also holds (which is indeed plausible in this limit) Eq. (5.22), then the model can be written as a standard SIR model,

$$\begin{aligned} \dot{S}_H &= -\beta_{eff} \frac{S_H I_H}{N_H} \\ \dot{I}_H &= \beta_{eff} \frac{S_H I_H}{N_H} - \gamma I_H \\ \dot{R}_H &= \gamma I_H, \end{aligned} \quad (5.23)$$

$$\text{where } \beta_{eff} = \frac{\beta'}{\lambda} = \frac{\beta \alpha N_v^*}{\mu N_H}.$$

In Fig. 5.6 we show the validity of the reduced models Eq. (5.22) and Eq. (5.23). Fig. 5.6(a) shows that the SIR-like model (Eq. (5.22)) works when the time-scale approximation can be performed (as  $\mu/\gamma \gg 1$ ) but the SIR model fails when the condition  $\lambda N_H \gg I_H$  is not fulfilled. Conversely, in Fig. 5.6(b) we show that as  $\lambda N_H \gg I_H$  is fulfilled, then the SIR model perfectly matches the original model. Finally, Fig. 5.6(c) shows the decrease in the mean squared error of the approximation as the condition Eq. (5.11) is fulfilled for different values of  $f$ .

## 5.4 Conclusions

In the present work we have analyzed several features of a compartmental deterministic model for vector-borne diseases with 3 compartments for hosts and 2 for vectors, that does not consider neither direct host to host nor vertical transmission. The goal is to study the behavior of the model in the case that the vector population is not stationary. In this case, the pre-pandemic disease-free state is not a fixed point (equilibrium state) of the dynamical system, and, in principle, the methods that are customarily used to determine the basic reproduction number,  $R_0$  do not work. This is so because these methods determine the onset of an outbreak by performing a linear stability analysis of the disease-free state, assuming that it is a fixed point of the model. A common assumption made in the literature is to determine  $R_0$  from the asymptotic state for the vectors (if it is not an extinction state), a fixed point of the model.

We have analyzed several initial conditions of the vector population, characterizing different regimes. In the case that the initial condition for the number of vectors is below the asymptotic state, implying that the vector population overall grows, then  $R_0$  as determined from the asymptotic state correctly predicts the existence (or not) of an epidemic outbreak, but

with a temporal delay in its appearance. This result contrasts with the situation in which the initial state is above the asymptotic state, with an overall decrease in the vector population. In this case  $R_0$  determined from the asymptotic state may fail badly, predicting no outbreak while a large fraction of the population might get infected. We present a simple, albeit useful, generalization of  $R_0$  that is able to give a reasonable prediction of the epidemic threshold for decaying populations, including the case in which vectors become extinct, a case in which the asymptotic estimation to determine  $R_0$  cannot be applied.

Compartmental models of vector-borne diseases usually have many compartments and parameters, which can lead to a problem of parameter unidentifiability. The model analyzed here is not an exception, and when applied to real-world cases many different combinations of the parameters could be able to reproduce the available data. Thus, in order to facilitate the application of the model to experimental data, we have studied a useful fast-slow (or adiabatic) approximation that allows to reduce the model if the parameters fulfill certain conditions. In particular, our study shows that under quite realistic assumptions (the typical timescale of hosts infection and death is much slower than vector timescales) it is possible to obtain a reduced SIR model. We recall that this reduction implies that, under these assumptions, the process by which hosts (that could be immobile) get infected through the action of vectors is equivalent to a direct interaction among hosts.

The deterministic compartmental model analyzed here, with some modifications, is a clear candidate to study many vector-borne diseases, in particular phytopathologies. Furthermore, in case of parameter unidentifiability the model reductions performed in this work could be useful to solve this issue. In any case, this description is still idealized, as compartmental models imply a well-mixed assumption in which space is not explicitly described. This kind of representations are not always applicable to real-world scenarios although are useful as a first approximation. Thus, future research should focus on the integration of space and vector mobility in the model to account for more realistic situations.

## 5.5 Appendix

### 5.5.1 Calculation of $R_0$ from standard methods

The standard methods of calculation of  $R_0$  are based in the linear stability analysis of the disease-free equilibrium, either directly, through the linear analysis of the fixed point, that yields the stability condition from which  $R_0$  can be obtained, or using the Next Generation Method (NGM) (138) that provides directly  $R_0$  by solving a suitable linear problem. Customarily these methods are applied to a pre-pandemic disease-free equilibrium, but as there is no such state in the case of non-stationary populations, here a similar approach is applied to a post-pandemic or asymptotic disease-free equilibrium.

#### Linear stability analysis

In order to perform the linear stability analysis of the fixed point ( $I_H = I_v = 0$ ) we first need to compute the Jacobian matrix,  $J$ ,

$$J = \begin{pmatrix} -\beta \frac{I_v}{N_H} & 0 & 0 & -\beta \frac{S_H}{N_H} \\ \beta \frac{I_v}{N_H} & -\gamma & 0 & \beta \frac{S_H}{N_H} \\ 0 & -\alpha \frac{S_v}{N_H} & -\alpha \frac{I_H}{N_H} - \mu & 0 \\ 0 & \alpha \frac{S_v}{N_H} & \alpha \frac{I_H}{N_H} & -\mu \end{pmatrix} \quad (5.24)$$

Then, we evaluate the Jacobian at the fixed point (or disease free equilibrium, DFE), yielding

$$J|_{DFE} = \begin{pmatrix} 0 & 0 & 0 & -\beta \\ 0 & -\gamma & 0 & \beta \\ 0 & -\alpha \frac{C}{N_H} \frac{\delta}{\mu} & -\mu & 0 \\ 0 & \alpha \frac{C}{N_H} \frac{\delta}{\mu} & 0 & -\mu \end{pmatrix} \quad (5.25)$$

where  $S_H = N_H$  has been considered.

The eigenvalues of Eq. (5.25) are,

$$\begin{aligned}\lambda_0 &= 0 \\ \lambda_\mu &= -\mu \\ \lambda_\pm &= -\frac{(\gamma+\mu)}{2} \pm \frac{1}{2} \sqrt{(\gamma-\mu)^2 + 4\beta\alpha \frac{C}{N_H} \frac{\delta}{\mu}}\end{aligned}\tag{5.26}$$

It is straightforward to see that all eigenvalues are real and the stability of the disease-free equilibrium is determined by the sign of the eigenvalues.  $\lambda_\mu = -\mu < 0$  as  $\mu$  is defined positive, so in order to discuss the stability of this fixed point, we need to study the  $\lambda_\pm$  eigenvalues.  $\lambda_-$  is always negative, but  $\lambda_+$  changes sign depending on the values of the parameters. The threshold condition  $\lambda_+ = 0$  leads to:

$$\lambda_+ = 0 \Rightarrow \frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} = 1\tag{5.27}$$

So, for  $\frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} < 1 \Rightarrow \lambda_+ < 0$  the fixed point is stable and for  $\frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} > 1 \Rightarrow \lambda_+ > 0$  a perturbation will grow in the direction of the eigenvector associated to  $\lambda_+$ . Thus, this threshold defines the basic reproduction number,

$$R_0 = \frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu}\tag{5.28}$$

If instead of  $S_H = N_H$  one considers any initial condition of hosts,  $S_H(0)$ , the basic reproduction number is given by,

$$R_0 = \frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} \frac{S_H(0)}{N_H}\tag{5.29}$$

### Next Generation Matrix method

The previous result can also be obtained by means of the NGM method, which is explained in detail in (138). Basically the method is based in decomposing the Jacobian in the form  $J = T + \Sigma$ , where  $T$  is the *transmission part*, that describes the production of new infections, and  $\Sigma$  the *transition part*, that describes changes of state (including death). Then, it can be proved (138) that the *basic reproduction number*  $R_0$  is given by the spectral radius (i.e. the largest eigenvalue) of the (next generation) matrix  $K = -T\Sigma^{-1}$ .

$$K = -T\Sigma^{-1} = \begin{pmatrix} \frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} & \frac{\beta}{\mu} \\ 0 & 0 \end{pmatrix}\tag{5.30}$$

with,

$$T = \begin{pmatrix} 0 & \beta \frac{N_H}{N_H} \\ 0 & 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} -\gamma & 0 \\ \alpha \frac{C}{N_H} \frac{\delta}{\mu} & -\mu \end{pmatrix}$$

$$\text{and } -\Sigma^{-1} = \begin{pmatrix} \frac{1}{\gamma} & 0 \\ \frac{\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} & \frac{1}{\mu} \end{pmatrix}$$

The basic reproduction number is the spectral radius of this matrix so:

$$\begin{aligned}\det(K - \sigma\mathbb{I}) = 0 &\implies \begin{vmatrix} \frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} - \sigma & \frac{\beta}{\mu} \\ 0 & -\sigma \end{vmatrix} = \\ &= (-\sigma) \left( \frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} - \sigma \right) = 0.\end{aligned}$$

Solving for  $\sigma$  one obtains the solutions,

$$\sigma_1 = \frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H}; \quad \sigma_2 = 0\tag{5.31}$$

Therefore, the basic reproduction number is

$$R_0 = \frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} \quad (5.32)$$

If instead of  $S_H = N_H$  one considers any initial condition of hosts,  $S_H(0)$ , the basic reproduction number is given by,

$$R_0 = \frac{\beta\alpha}{\gamma\mu} \frac{C}{N_H} \frac{\delta}{\mu} \frac{S_H(0)}{N_H} \quad (5.33)$$

### 5.5.2 Calculation of $R_0$ for non-stationary vector populations

We extend the computation of  $R_0$  in the case of non-stationary and non-periodic vector populations by following the natural definition of *basic reproductive number*. Thus,  $R_0$  is computed by averaging the number of secondary infections produced by an infected individual along one generation, that is equivalent to averaging the instantaneous definition of  $R_0$ , namely  $R_0^i$ , over one generation,

$$\overline{R_0} = \langle R_0^i(t) \rangle \Big|_0^{t_g} = \frac{R_0}{N_v^*} \langle N_v(t) \rangle \Big|_0^{t_g} = \frac{R_0}{N_v^*} \frac{1}{t_g} \int_0^{t_g} N_v(t) dt, \quad (5.34)$$

where the integral in Eq. (6.6) is solved as

$$\begin{aligned} \int_0^{t_g} N_v(t) dt &= \left[ N_v^* t - \frac{1}{\mu} (N_v(0) - N_v^*) e^{-\mu t} \right]_0^{t_g} = \\ &= N_v^* t_g - \frac{1}{\mu} (N_v(0) - N_v^*) [e^{-\mu t_g} - 1]. \end{aligned} \quad (5.35)$$

Thus, the basic reproduction number for non-stationary vector populations is given by

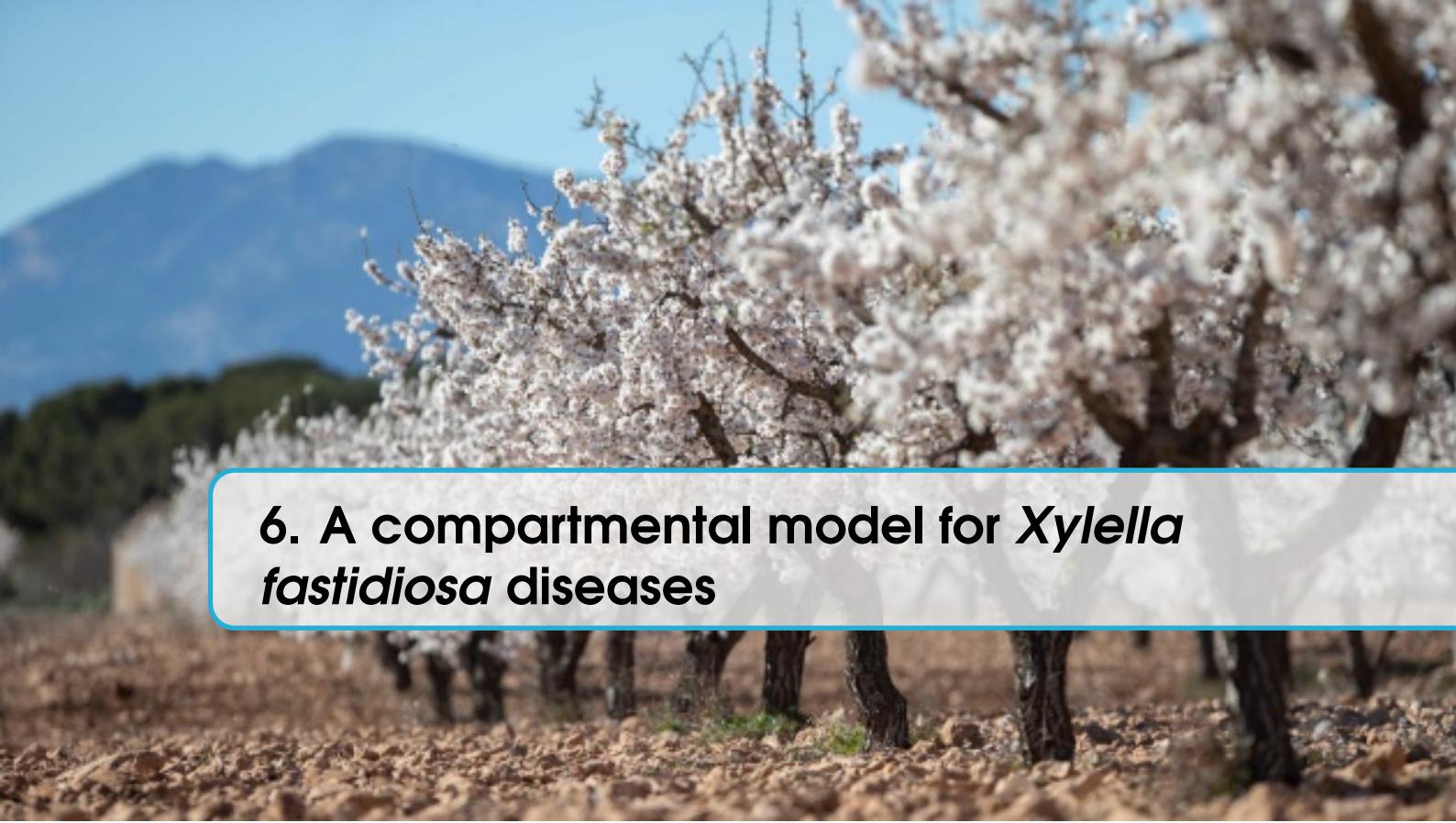
$$\overline{R_0} = \frac{R_0}{N_v^*} \left\{ N_v^* - \frac{1}{\mu t_g} [N_v(0) - N_v^*] [e^{-\mu t_g} - 1] \right\}, \quad (5.36)$$

where the generation time,  $t_g$ , is Eq. (5.8). Eq. (6.8) can be rewritten as,

$$\overline{R_0} = \langle R_0^i(t) \rangle \Big|_0^{t_g} = R_0 \left[ 1 - \frac{1}{\tau} (f - 1) (e^{-\tau} - 1) \right] = R_0 \cdot \mathcal{F}, \quad (5.37)$$

where  $\tau = 1 + \mu/\gamma$  and  $\mathcal{F}$  is the expression in brackets, which accounts for the effect of the decaying vector population on the stationary  $R_0$ .

In our approach, a generation is defined as the time elapsed in the following sequence of processes: 1) A host individual becomes infected; 2) The infected host passes the disease to a susceptible vector; 3) The infected vector dies. Basically, the time elapsed from the first to the last process is the time in which new infections can be produced, i.e.  $t_g$  Eq. (5.8).



## 6. A compartmental model for *Xylella fastidiosa* diseases

Àlex Giménez-Romero<sup>1</sup>, Eduardo Moralejo<sup>2</sup>, Manuel A. Matías<sup>1</sup>

1. Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Palma de Mallorca 07122, Spain
2. Tragsa, Passatge Cala Figuera 6, 07009 Palma de Mallorca, Spain

### Published as

Àlex Giménez-Romero, Eduardo Moralejo, and Manuel A. Matías. "A Compartmental Model for *Xylella fastidiosa* Diseases with Explicit Vector Seasonal Dynamics". In: *Phytopathology®* 113.9 (2023). PMID: 36774557, pages 1686–1696. DOI: [10.1094/PHYTO-11-22-0428-V](https://doi.org/10.1094/PHYTO-11-22-0428-V). eprint: <https://doi.org/10.1094/PHYTO-11-22-0428-V>. URL: <https://doi.org/10.1094/PHYTO-11-22-0428-V>

## 6.1 Introduction

Mathematical and computational modeling in Ecology and, in particular, Epidemiology have been recently recognized as powerful approaches to guide empirical work and provide a framework for the synthesis, analysis and development of conservation plans and policy-making (98, 279, 326, 382). Plant epidemics, mainly plant-virus diseases, have been often described by compartmental models, which deal with the overriding importance of transmission mechanisms in determining epidemic dynamics (243, 244, 291). These models have contributed to providing answers to some questions related to the ecology of plant diseases and have led to direct applications in disease control while guiding research directions (241).

The emergence of vector-borne plant pathogens in new areas causing huge economic impacts, such as *Xylella fastidiosa* and the *Candidatus Liberibacter* spp. (Huanglongbing or citrus greening), has sparked interest in modeling vector-transmitted plant-disease epidemics (99, 241). The vector-borne bacterium *X. fastidiosa* (Xf) is a multi-host pathogen endemic to the Americas that causes economically important diseases, mostly in woody crops (234). Xf is a genetically diverse species with three evolutionary well-defined clades forming the *pauca*, *fastidiosa*, and *multiplex* subspecies, native from South, Central, and North America, respectively (431). Within each subspecies, diverse genetic lineages with different host ranges are found. Xf is transmitted non-specifically by xylem-sap-feeding insects belonging to the sharpshooter leafhoppers (Hemiptera: Cicadellidae, Cicadellinae) and spittlebugs (Hemiptera: Cercopoidae) (363).

Recently, Xf has gained renewed interest due to the massive mortality of olive trees in Apulia, Italy (380). The first focus of the olive quick decline syndrome (OQDS) was detected in 2013 around Gallipoli (Apulia, Italy) (378) and since then has spread throughout the region by the meadow spittlebug, *Philaenus spumarius*. Although this was the first official detection of Xf in Europe, it has recently been demonstrated that the pathogen arrived much earlier in Corsica (407) and in the Balearic islands (314). Around 1993, two strains of the subspecies *fastidiosa* (ST1) and *multiplex* (ST81) were introduced from California to Mallorca (Spain) with infected almond plants (314). To date, over 80% of the almond trees in Mallorca show leaf scorch symptoms and the outbreak has changed the iconic rural landscape of this Mediterranean island (337).

The meadow spittlebug, *P. spumarius* (Hemiptera: Aphrophoridae), has recently been shown to be the main vector of Xf in Europe both in transmission experiments and in field studies (111, 113, 286, 313, 380). *P. spumarius* is a polyphagous species from the Palearctic region, presenting one generation per year (univoltine) and overwintering as eggs. Foam-forming nymphs emerge at the end of winter, feeding on herbaceous plants. The time required for their development to the adult stage depends mainly on temperature and humidity (48, 100, 109). In Mediterranean climates, *P. spumarius* adults generally move from the herbaceous cover to the crop canopy as evapotranspiration increases in late spring (May–June). In mid-summer, the populations of *P. spumarius* tend to decrease in the crop canopy, while the insects are captured more frequently in trees and shrubs interspersed in crops. Summer dispersal of spittlebugs to wild hosts as refugee seems a common general pattern in Mediterranean crops in Italy (48, 115) and Spain (318). Because the bacterium has not been detected in spring on insects feeding on the herbaceous cover or in weeds in Europe (48, 109, 337), it is assumed that all spittlebug adults acquire the bacteria from the main crop (olive, almond, vine, etc.). Once infected, Xf colonizes the insect foregut in a persistent and non-circulatory manner without transovarial (parent to offspring) or transstadial (inter-stage) transmission (9, 167, 358) and without a period latency after vector acquisition (8, 167).

Several epidemic models have been already developed for Xf-diseases, but they lack a realistic description of some relevant processes (241). Some of these models assume a simple general form for infected host dynamics (2, 128, 448) or use a simplified S-I compartmental scheme for hosts, disregarding important features such as the latent period or the host mortality rate (407). Models that do take these features into account, however, do not explicitly model the population of vectors responsible for disease transmission (449). Other more recent models have taken a step further in explicitly modeling the vector population (71, 194), but the characterization of its dynamics is still relatively simple, as it overlooks the known seasonal patterns of vector abundance. Several recent studies have provided new insights into the ecology and temporal dynamics of the transmission of Xf by *P. spumarius* in olive plants (48, 49). However, these

experimental data of the pathosystem have not been yet integrated at the population level. Thus, there is a need to continue advancing in the modeling of Xf diseases by developing more realistic models that can elucidate the fundamental processes involved in vector-host-pathogen interactions and help to design effective control strategies.

In this work, we develop a deterministic continuous-time compartmental model to describe the general epidemiological dynamics of diseases produced by Xf in Europe. We explicitly account for key biological aspects of the disease, including the seasonal dynamics of its main vector, *P. spumarius*. Our model is able to describe field data from the two major European outbreaks: the olive quick disease syndrome (OQDS) in Apulia, Italy, caused by the subspecies *pauca*, and the almond leaf scorch disease (ALSD) in Mallorca, Spain, caused by subspecies *multiplex* and *fastidiosa*. We aimed to find the most influential parameters in the model with respect to incidence and mortality in both diseases by performing a global sensibility analysis. With this information, the next goal was to explore control strategies acting especially on the vector population.

## 6.2 Materials and Methods

### 6.2.1 Epidemic model: the SEIR-V model

We developed a deterministic continuous-time compartmental model that incorporates the specific biological features of Xf diseases in Europe, including the dynamics of the main relevant vector *P. spumarius* (93). To build the model we took the following considerations: (i) we assume there is no winter recovery of infected hosts and thus they die sometime after infection; (ii) hosts show an asymptomatic period in which they are non-infectious in practice (exposed compartment) because the bacteria are not yet systemically extended (410, 416), while vectors are infectious immediately after acquiring the bacterium (158); (iii) vectors have an annual life cycle without mother-to-offspring disease transmission (167, 358), so we consider the annual emergence of susceptible newborn vectors and a constant death rate for both susceptible and infected vectors; (iv) infected vectors carry the bacterium during their entire lifespan without affecting their fitness; and finally, (v) we do not consider host recruitment or natural death given that the typical development time of Xf-epidemics is faster than the typical host's life cycle.

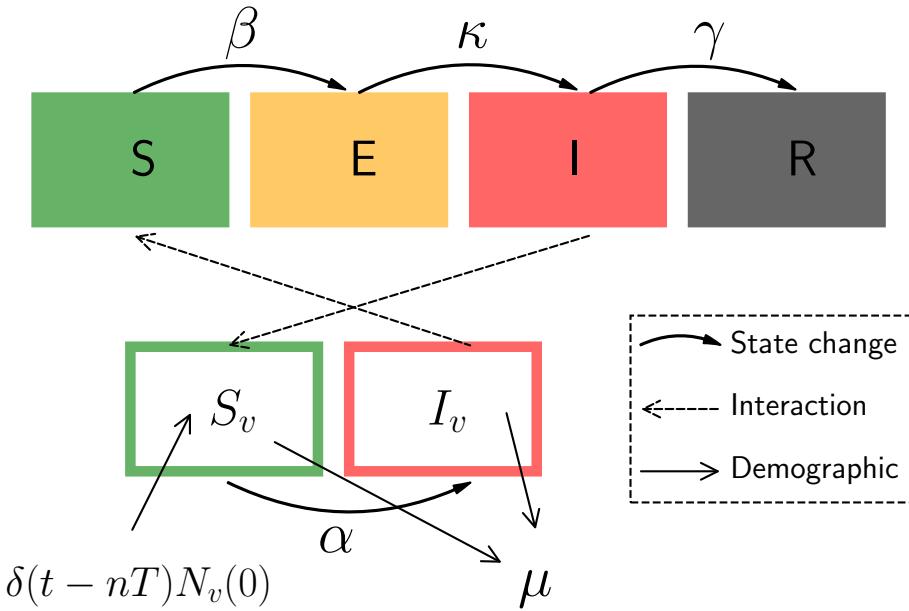
Altogether, our deterministic continuous-time compartmental model consists of six compartments, four describing the host population (susceptible,  $S_H$ , exposed,  $E_H$ , infectious,  $I_H$ , and removed,  $R_H$ ), and two describing the vector population (susceptible,  $S_V$ , and infected,  $I_V$ ). The model is defined according to the following processes,

$$\begin{aligned} S_H + I_V &\xrightarrow{\beta} E_H + I_V, \quad E_H \xrightarrow{\kappa} I_H, \quad I_H \xrightarrow{\gamma} R_H \\ S_V + I_H &\xrightarrow{\alpha} I_V + I_H, \quad S_V \xrightarrow{\mu} \emptyset, \quad I_V \xrightarrow{\mu} \emptyset, \end{aligned} \tag{6.1}$$

which are illustrated in Section 6.2.1, being the birth of new susceptible vectors described as a source term. Thus, the host-vector compartmental model is written as,

$$\begin{aligned} \dot{S}_H &= -\beta S_H I_V / N_H \\ \dot{E}_H &= \beta S_H I_V / N_H - \kappa E_H \\ \dot{I}_H &= \kappa E_H - \gamma I_H \\ \dot{R}_H &= \gamma I_H \\ \dot{S}_V &= N_v(0) \sum_{n=1}^{\infty} \delta(t - nT) - \alpha S_v I_H / N_H - \mu S_v \\ \dot{I}_V &= \alpha S_v I_H / N_H - \mu I_V. \end{aligned} \tag{6.2}$$

The model describes the exposure of susceptible hosts,  $S_H$ , at a rate  $\beta$  through their interaction with infected vectors,  $I_V$ , while susceptible vectors,  $S_V$ , get infected immediately at a rate  $\alpha$  through their interaction with infectious hosts  $I_H$ . Exposed hosts get infectious at rate  $\kappa$ , being the mean latent period  $\tau_E = 1/\kappa$ , while infectious hosts die at rate  $\gamma$ , having a mean infectious period of  $\tau_I = 1/\gamma$ . Infected vectors stay infected and infectious for the rest of their lifetime. Regarding the seasonal dynamics of vectors, we assume that new adults emerge synchronously each year



**Figure 6.1:** Schematic representation of the model Eq. (6.2). Boxes are the compartments in which the population is divided, solid curved arrows represent changes in state, i.e. transitions between compartments, dashed arrows depict the crossed interaction between hosts and vectors and solid straight arrows represent demographic changes in vector population.

in fields being all susceptible. This is represented by the term  $N_v(0) \sum_{n=1}^{\infty} \delta(t - nT)$  in Eq. (6.2), where  $T = 1 \text{ yr}$  is the period and  $\delta(t - nT)$  is Dirac delta function, and basically implements a yearly pulse of new vectors at a certain moment in the year. Vectors are removed (die, move to herbaceous vegetation and other non-host trees, exit the field, etc) at a given rate  $\mu$ , which we consider identical for susceptible and infected vectors. For simplicity, we consider that the quantity of annual newborn adults,  $N_v(0)$ , is constant. This outburst of new adults followed by an exponential decay resembles the temporal patterns on the abundance of *P. spumarius* observed in crop fields (20, 30, 111, 287) (see Fig. 6.8).

In our model (Eq. (6.2)) the crossed nonlinear terms in  $\dot{S}_H$  and  $\dot{S}_v$ ,  $S_H I_v$  and  $S_v I_H$ , are divided by the total host population,  $N_H$ . Thus, the vector-to-plant infection process is modeled using mass action incidence, which is density dependent, while the plant-to-vector infection process is modeled using standard incidence, which is frequency dependent (297). This implies that doubling the number of vectors in the crop field would double the number of resulting exposed (or infected) hosts, as this process is population-dependent (mass action incidence), while doubling the number of hosts would not result in more vectors per unit area being infected, as this process only depends on the contact probability, being frequency dependent (standard incidence). We think this is the most reasonable assumption because, for a given plantation framework, increasing the number of hosts is expected to also increase the area of the field, while the number of vectors is an independent quantity.

## 6.2.2 Basic reproductive number

The basic reproductive number,  $R_0$ , of the model cannot be trivially computed using standard methods such as the Next Generation Matrix (NGM) (138), as there is no pre-pandemic fixed point in the system of differential equations Eq. (6.2). For periodically varying vector populations, rigorous methods have been developed (25), but not for the case of growing or decaying vector populations. Here we use the simple method developed in the work of (195) (see Section 6.5.2), which effectively computes the average number of secondary infections produced by an initially infectious individual in one generation. Thus, the effective basic reproductive number is given by

$$R_0 = \frac{\beta\alpha}{\mu\gamma} \frac{S_H(0)}{N_H^2} \frac{N_v(0)}{\mu\tau} (1 - e^{-\mu\tau}) , \quad (6.3)$$

where  $\tau$  corresponds to the time length of one generation, in our case one year. This  $R_0$  is calculated using the initial susceptible host population,  $S_H(0)$ . Below we will also use a time-dependent  $R_0(t)$  using  $S_H(t)$ .

### 6.2.3 Epidemiological data

Epidemiological data from an ALSD outbreak in the island of Mallorca, Balearic Islands, Spain were taken from (314). Dated phylogenetic analysis and estimates of disease incidence showed that the introduction of both subspecies occurred around 1993 and ~79% almond trees were infected in 2017 (314). The annual proportion of infected individuals in the almond tree population between 1993 and 2017 was estimated by analyzing through qPCR the presence of Xf-DNA in the growth rings of 34 sampled trees (cf. Fig. 3 in (314)). The disease progression curve was estimated without distinguishing whether infections were caused by *multiplex* or *fastidiosa* subspecies. In addition, a two-sided bootstrap confidence interval for each data point was set using the SciPy bootstrap function in Python (438). On the other hand, epidemic data for OQDS were retrieved from (449). The data consisted of 2 to 3 yearly censuses of symptom prevalence in 17 olive groves infected with Xf subsp. *pauca* in Apulia, Italy, which were aggregated to fit our model as shown in Fig. 4 in (449). Because the compartments of our model are not in one-to-one correspondence with those shown in the work of White et al. (449), we used the sum of the symptomatic and desiccated infected trees in the dataset ( $I_S + I_D$ ) to fit the sum of the infected and dead trees ( $I + R$ ) and the sum of susceptible and asymptomatic hosts ( $S + I_A$ ) to fit the sum of susceptible and exposed hosts ( $S + E$ ). The processed data used to fit the model can be found in (192), while the raw data can be found in the supplementary data accessible online of the cited articles (314, 449).

### 6.2.4 Model fitting through Bayesian Inference

We employed an informative normal  $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$  prior distribution, with  $\hat{\mu}$  and  $\hat{\sigma}$ , the mean and standard deviation, respectively, for previously measured parameters in the literature, such as the infectious and latent periods for ALSD,  $\tau_I \sim \mathcal{N}(14, 4)$ ,  $\tau_E \sim \mathcal{N}(4, 1)$  (314, 416) and OQDS,  $\tau_I \sim \mathcal{N}(3.5, 1)$ ,  $\tau_E \sim \mathcal{N}(1.75, 0.5)$  (158). The corresponding rates are given by  $\gamma = 1/\tau_I$  and  $\kappa = 1/\tau_E$ , respectively. Similarly, a prior normal distribution was used for the removal rate of vectors,  $\mu \sim \mathcal{N}(0.02, 0.0075)$ , as the mean value  $\mu = 0.02$  already captures the vector dynamics observed in field-data (Fig. 6.8). Regarding the prior distribution for the transmission rates a very wide and uninformative uniform prior distribution,  $\beta \sim \mathcal{U}(0.001, 1)$  and  $\alpha \sim \mathcal{U}(0.001, 1)$ , was used for each parameter. The number of hosts,  $N_H$ , was already provided in the datasets, while, given the lack of information about the vector population, we assumed  $N_v(0) = N_H/2$  for the initial vector population of each year. However, we tested the robustness of our results by changing  $N_v(0)$ .

The posterior distributions of the parameters were approximated using the Markov Chain Monte Carlo algorithm No U-Turn Sampler (NUTS) with the recommended target acceptance rate of 65% (233). To ensure a proper convergence, we constructed three independent Markov Chains with  $10^5$  iterations each after a burn-in of  $10^4$  iterations and checked that the results were statistically equivalent. For each chain, we started at the maximum-likelihood parameters yielded by the Nelder-Mead algorithm with 1000 iterations.

The parameters of our compartmental model were determined by fitting the model to data by means of a Bayesian Inference framework using the Turing.jl package (181) in Julia (40). The scripts used to fit the model can be found in (192).

### 6.2.5 Sensitivity Analysis

We performed a Global Sensitivity Analysis (GSA) (377) of the model to assess the relative contribution of its parameters and their interactions with different features of the epidemic. In contrast to the Local Sensitivity Analysis (LSA), the GSA assesses the influence of a large domain of the parameter space in the desired outputs of the model. We performed GSA by means of a

variance-based analysis, the Sobol method (403). This particular method provides information not only on how a particular parameter alone influences the model outputs (as happens with LSA), but also due to the nonlinear interactions among two or more parameters. Briefly, the method considers the model output,  $Y$ , as a general function of the inputs,  $f(x_1, \dots, x_n)$ , so that the variance of the output,  $Var(Y)$  is decomposed as the sum of the variances given by the variations of the parameters alone and its interactions,  $Var(Y) = \sum_{i=1}^n Var(f(x_i)) + \sum_{i < j} Var(f(x_i, x_j)) + \dots$ . This information is organized in what are known as Sobol indices. The total order indices are a measure of the total variance of the output quantity caused by variations of the input parameter and its interactions,  $S_T = Var(f(x_1, \dots, x_n))/Var(Y)$ . First order (or “main effect”) indices are a measure of the contribution to the output variance given by the variation of the parameter alone, but averaged over the variations in other input parameters,  $S_i = Var(f(x_i))/Var(Y)$ . Second-order indices take into account first-order interactions between parameters,  $S_{ij} = Var(f(x_i, x_j))/Var(Y)$ . Further indices can be obtained, describing the influence of higher-order interactions between parameters, but these are not going to be considered.

Following the Sobol method, we analyzed the variation of the time at which the infectious population peaks,  $t_{peak}$ , the magnitude of this peak,  $I_{peak}$  and the final number of dead hosts,  $R_\infty$ , relative to variations of the model parameters. The method was implemented within the Julia high-level programming language (40) using the sub-package DiffEqSensitivity.jl in DifferentialEquations.jl package (362).

## 6.3 Results

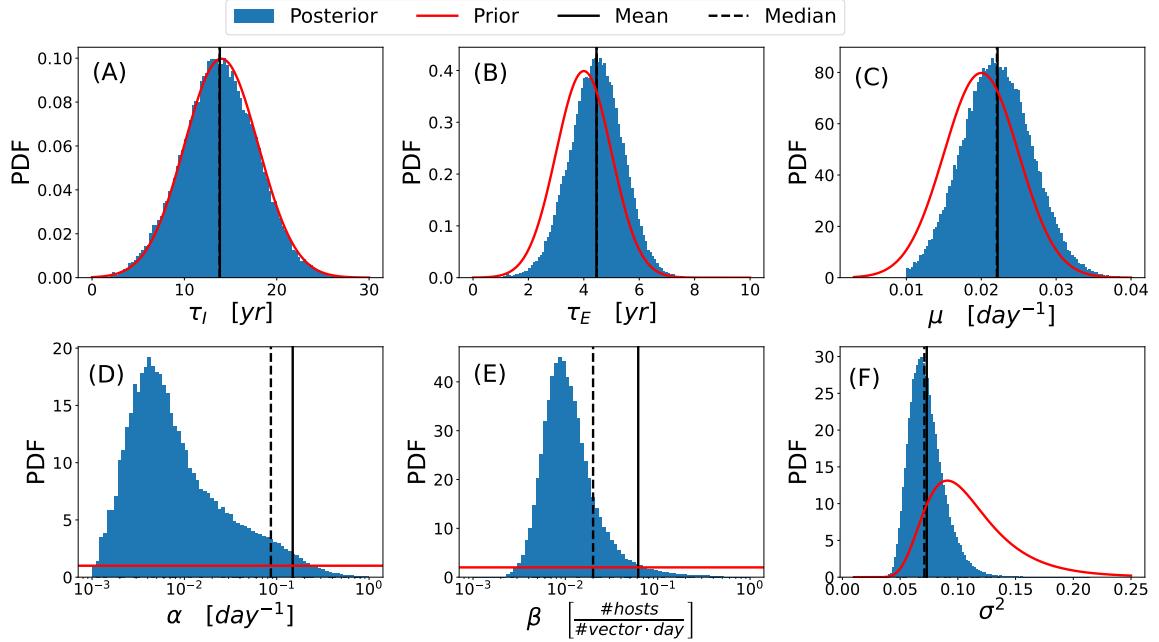
### 6.3.1 Model fit and parameter estimates

The posterior distributions of the fitted parameters including their estimated mean and median for ALSD and OQDS are shown in Figs. 6.2 and 6.3, respectively, together with the assumed prior distributions. We observe that the literature-driven priors for the latent and infectious period,  $\tau_E$  and  $\tau_I$ , were already very good guesses and changed slightly converging to the appropriate distribution that better fitted the epidemic data for both ALSD and OQDS (Fig. 6.2(A-B) and Fig. 6.3(A-B)). Similarly, the prior for the vector removal rate,  $\mu$ , obtained from field data, was good enough so that little changes were needed for convergence (Fig. 6.2(C) and Fig. 6.3(C)). On the other hand, we also observe that the completely uninformative priors for the transmission rates successfully converged to the posterior distributions (Fig. 6.2(D-E) and Fig. 6.3(D-E)).

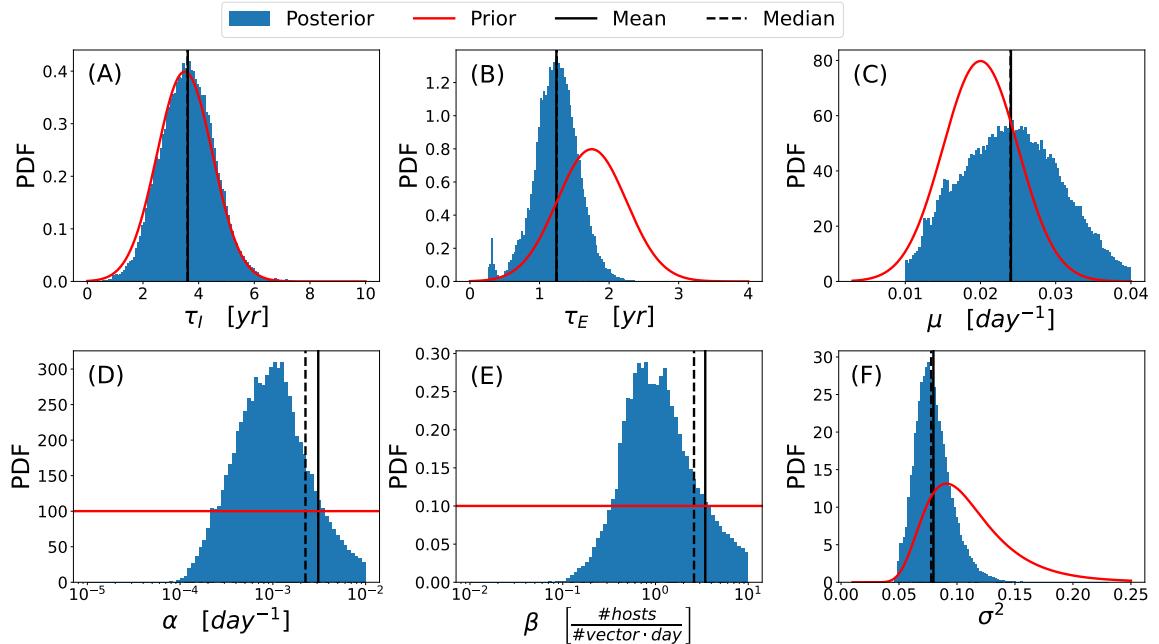
**Table 6.1:** Estimated epidemiological parameters from Bayesian model fitting to the disease progression curve of ALSD in Mallorca.

Parameter	Definition	Units	Posterior Mean	Posterior Median	95% C.I.
$\tau_I$	Host infectious period	yr	13.84	13.82	[7.12, 20.47]
$\tau_E$	Host latent period	yr	4.46	4.47	[2.88, 5.99]
$\beta$	Host infection rate	#host #vector-day	0.062	0.02	[0.0061, 0.3013]
$\alpha$	Vector infection rate	day <sup>-1</sup>	0.15	0.086	[0.0047, 0.54]
$\mu$	Vector removal rate	day <sup>-1</sup>	0.0222	0.0221	[0.015, 0.030]
$R_0$	Basic reproductive number	-	133	25	-

The latter distributions are far from a Gaussian-like shape (note that the x-axis is log-scaled), being heavy-tailed. This kind of distribution highly distorts the statistical measures of mean, median and standard error, indicating that the estimates for transmission rates are not as robust as the estimates for the other parameters. These rather uninformative distributions are most probably arising because of the lack of data about the vector, i.e.  $S_v(t)$  and  $I_v(t)$ , to constrain the fits. In essence, many combinations of  $\alpha$  and  $\beta$  can similarly fit the host data while yielding quite different time series for  $S_v(t)$  and  $I_v(t)$ , which cannot be contrasted due to the lack of field data. Nevertheless, the obtained best-fit mean and median parameters, although quite different, are able to perfectly fit the data (Fig. 6.4). Finally, we also observe that the variance for the field data also converged to a bell-shaped distribution.



**Figure 6.2:** Posterior (blue histograms) and prior (red line) distributions of the model parameters for ALSD. Solid and dashed black lines correspond to the mean and median of the posterior distributions. (A) Host infectious period  $\tau_I = 1/\gamma$ . (B) Host latent period  $\tau_E = 1/\kappa$ . (C) Vector removal rate  $\mu$ . (D) Vector infection rate  $\alpha$ . (E) Host infection rate  $\beta$ . (F) The variance of the field data  $\sigma^2$ .



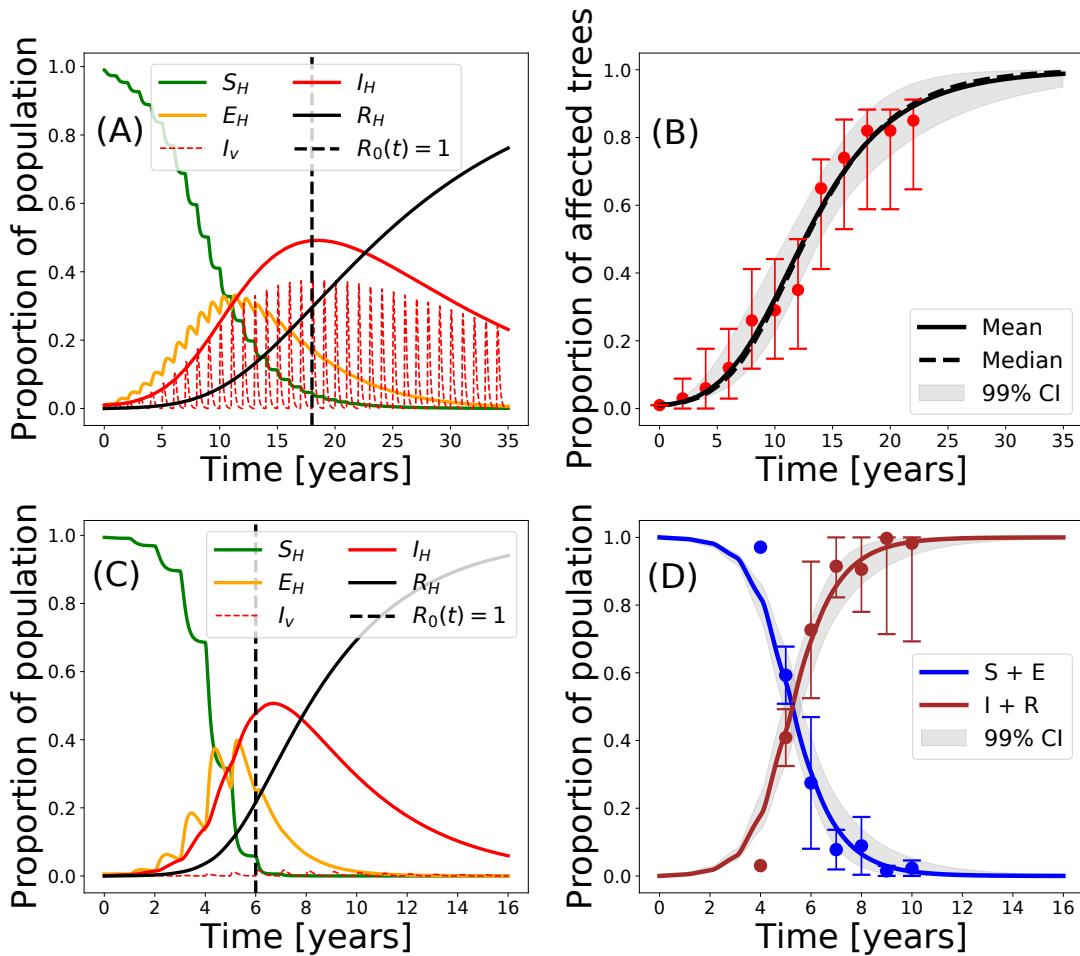
**Figure 6.3:** Posterior (blue histograms) and prior (red line) distributions of the model parameters for OQDS. Solid and dashed black lines correspond to the mean and median of the posterior distributions. (A) Host infectious period  $\tau_I = 1/\gamma$ . (B) Host latent period  $\tau_E = 1/\kappa$ . (C) Vector removal rate  $\mu$ . (D) Vector infection rate  $\alpha$ . (E) Host infection rate  $\beta$ . (F) Variance of the field data  $\sigma^2$ .

Mean and median parameter estimates, i.e. the best-fit parameter values for ALSD and OQDS, are summarized in Tables 6.1 and 6.2, respectively. As already seen from the posterior distributions, the best-fit values for  $\tau_E$ ,  $\tau_I$  and  $\mu$  are close to the ones given by literature and field data for both diseases. Conversely,  $\alpha$  and  $\beta$  are rather uninformative, as their 95% confidence

intervals cover almost two orders of magnitude. This again indicates that without some data about the evolution of the vector states in time,  $S_v(t)$  and  $I_v(t)$ , it is nearly impossible to derive the proper values for these parameters.

**Table 6.2:** Estimated epidemiological parameters from Bayesian model fitting to the disease progression curve of OQDS in Apulia.

Parameter	Definition	Units	Posterior Mean	Posterior Median	95% C.I.
$\tau_I$	Host infectious period	yr	3.61	3.60	[2.06, 5.20]
$\tau_E$	Host latent period	yr	1.24	1.25	[0.70, 1.75]
$\beta$	Host infection rate	#host / #vector-day	3.44	2.60	[0.55, 8.79]
$\alpha$	Vector infection rate	day <sup>-1</sup>	0.0031	0.0022	[0.0005, 0.0084]
$\mu$	Vector removal rate	day <sup>-1</sup>	0.0240	0.0240	[0.014, 0.035]
$R_0$	Basic reproductive number	-	33	21	-

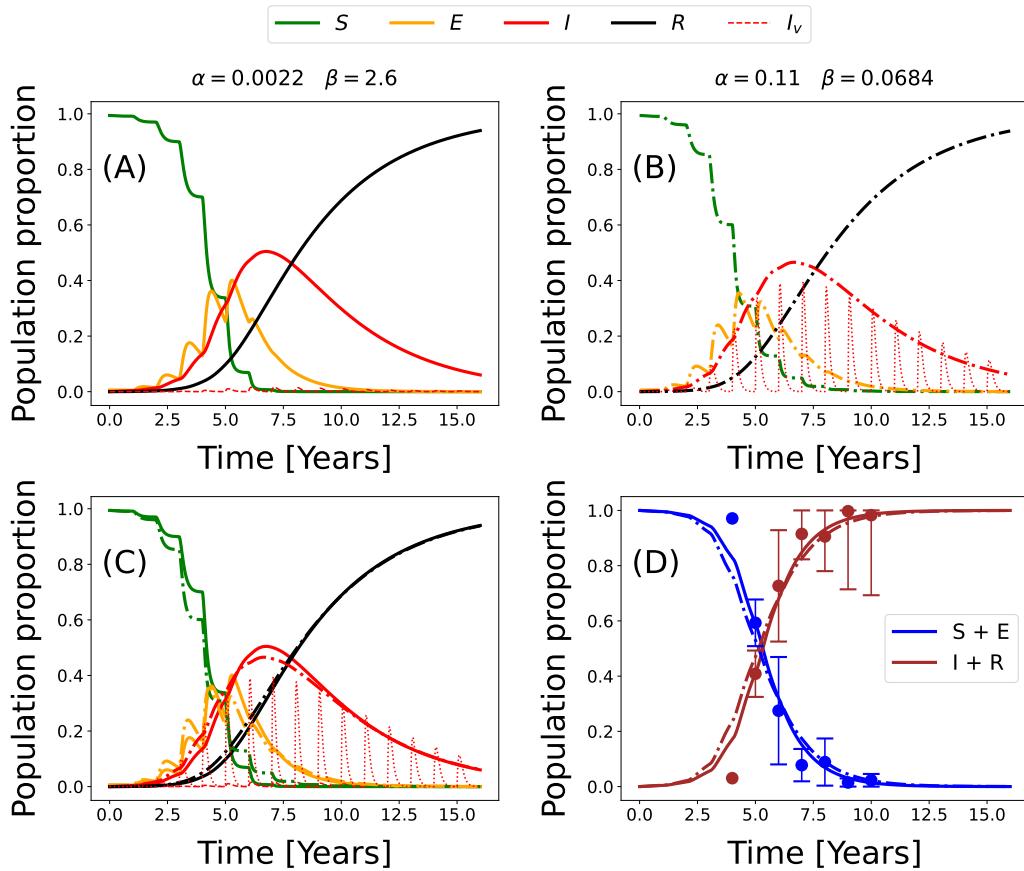


**Figure 6.4:** (A) Simulation of the model with best-fit parameters for ALSD. (B) Model fit to field data by means of the mean and median values of the posterior distributions of the parameters for ALSD. (C) Simulation of the model with best-fit parameters for OQDS. (D) Model fit to field data by means of the mean and median values of the posterior distributions of the parameters for OQDS. The gray-shaded area corresponds to the 99% confidence interval. The error bars for the field data correspond to their 95% confidence interval obtained with a bootstrapping technique.

Overall, the data falls within the 99% confidence limits of the fitted model for both the ALSD and OQDS outbreaks (Fig. 6.4(B,D)). We also computed the instantaneous reproductive number,  $R_0(t)$ , by using Eq. (6.3) with  $S_H(t)$  instead of only  $S_H(0)$  along the simulation. Noteworthy,  $R_0(t) = 1$

coincides with the stopping of new infections being produced, i.e. the number of exposed hosts does not increase (Fig. 6.4(A,C)). This supports our approximate method for computing the reproductive number for Xf diseases (Section 6.5.2, Eq. (6.3)). Due to the different time scales of both epidemics ( $\tau_I^{ALSD} + \tau_E^{ALSD} > \tau_I^{OQDS} + \tau_E^{OQDS}$ ), the OQDS outbreak dies out earlier than the one for ALSD.

We notice that for ALSD a large proportion of the vector population gets infected every year (Fig. 6.4(A)), while a very small proportion is needed in OQDS to produce a lethal outbreak (Fig. 6.4(C)). However, this last statement is rather unrealistic, as around 50% of the vectors that are captured in Apulia are indeed infected by Xf (93, 112). Thus, the evolution of the infected vector population should be qualitatively similar to that obtained for ALSD (Fig. 6.4(C)). As previously explained, different suitable combinations of  $\alpha$  and  $\beta$  parameters should give rise to similar progression curves for the hosts while different ones for the vectors, but the realistic values for these parameters cannot be obtained from the Bayesian fit due to the lack of data of the vector states,  $S_v(t)$ ,  $I_v(t)$ .



**Figure 6.5:** (A) Simulation of the model with the original best-fit parameters for OQDS. (B) Simulation of the model with the original best-fit parameters for OQDS but with different  $\alpha$ ,  $\beta$  values. (C) Comparison of the progression curves. Note that the curves for the hosts are very similar while the curve for the infected vector population is very different. (D) Comparison of the model fit to the data with both simulations. Solid lines correspond to results with the original best-fit parameters while dash-dot lines correspond to the results of the more realistic scenario with different  $\alpha$  and  $\beta$ .

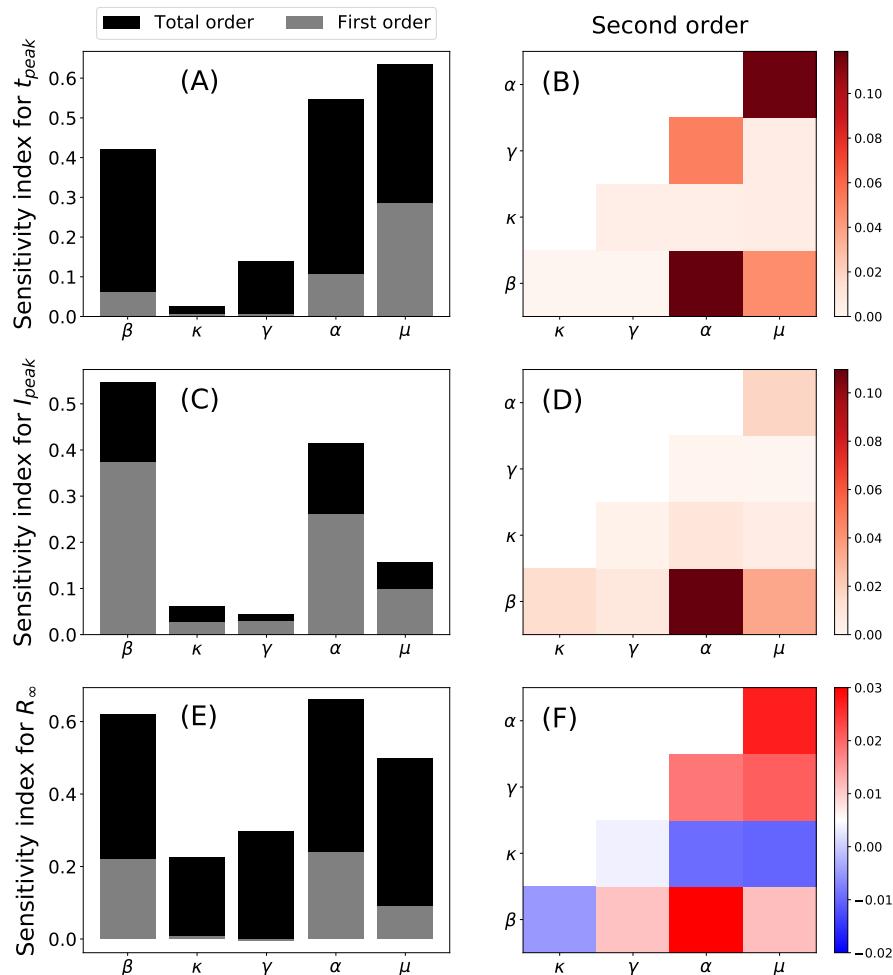
Nevertheless, by manually exploring other values for  $\alpha$  and  $\beta$  parameters, we can obtain a more biologically plausible scenario for the OQDS that is still able to fit the available data for the hosts. Fig. 6.5(A) shows a simulation of the model with previously inferred best-fit median parameters for OQDS. By changing the values of  $\alpha$  and  $\beta$ , we obtain a more realistic scenario, i.e. around a 50% of the vector population getting infected during the outbreak (Fig. 6.5(B)) (93, 112). Noteworthy, the  $\beta$  value obtained in this way is almost identical to the transmission

rate recently reported by (49) for OQDS. This change in the transmission parameters only affects the progression curve of the infected vector population, being the progression of the host compartments practically unchanged (Fig. 6.5(C)). Anyway, both sets of parameter values for  $\alpha$  and  $\beta$  can properly fit the field data, corresponding exclusively to the host population (Fig. 6.5(D)).

The model adjusted to the progression curves of both diseases indicates that the transmission rate  $\alpha$  must be greater than  $\beta$  when the proportion of infected vectors is relatively high ( $> 30\%$ ). We checked if the relation between  $\alpha$  and  $\beta$  held when changing the assumed  $N_v(0) = N_H/2$ , obtaining that it kept approximately the same for very different values of the initial vector population.

### 6.3.2 Global Sensitivity Analysis

We computed the sensitivity indices for the model parameters with respect to the more relevant quantities of interest, namely, the time at which the number of infectious hosts is maximal,  $t_{peak}$ , the maximum number of infectious hosts,  $I_{peak}$  and the final number of dead hosts,  $R_\infty$ . The results were obtained exploring the parameter space constrained to the intervals  $\{\beta \in (0.001, 0.1), \tau_E \in (3, 7), \tau_I \in (5, 25), \alpha \in (0.001, 1), \mu \in (0.01, 0.04)\}$  using  $10^4$  Quasi-Monte Carlo samples and are summarized in Fig. 6.6.



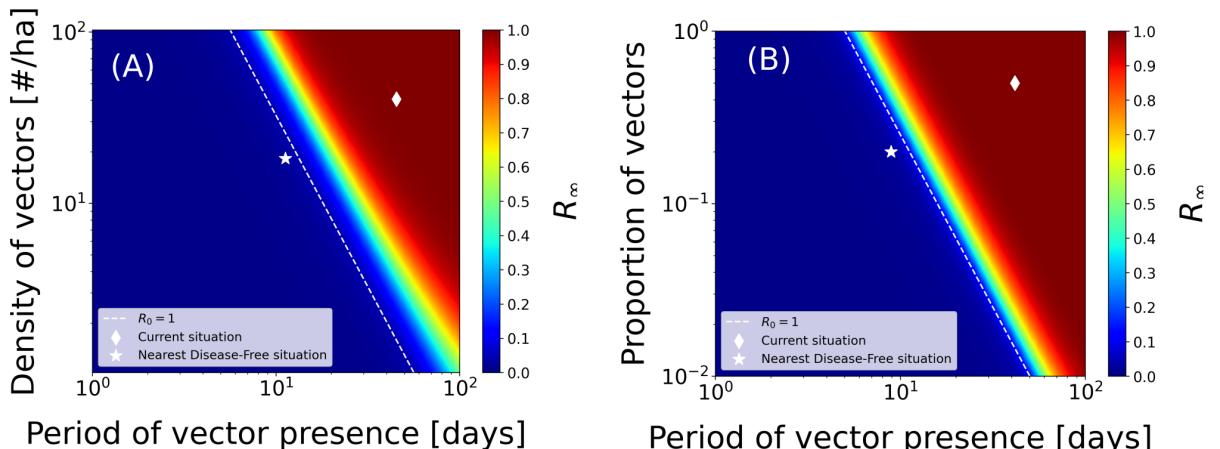
**Figure 6.6:** Global Sensitivity Analysis of the model parameters performed with the Sobol method with respect to the time at which the infectious population peaks,  $t_{peak}$  (A-B), the magnitude of this peak,  $I_{peak}$  (C-D) and the final number of dead hosts,  $R_\infty$  (E-F). The left column (A,C,E) shows the total and first-order indices and the right column (B,D,F) shows the second-order indices.

Parameters  $\alpha$ ,  $\beta$  and  $\mu$  are the most influential with regard to the time at which the infectious host population peaks,  $t_{peak}$ , the magnitude of the peak,  $I_{peak}$ , and the final number of dead hosts,  $R_\infty$ . The total output variance (total order indices) cannot be explained by the variances of the parameters alone (first order indices) (Fig. 6.6). Therefore, higher-order interactions among the parameters importantly affect the sensitivity of the quantities under study. Indeed, the contribution to the total output variance of  $\gamma$  and  $\kappa$  for  $t_{peak}$  and  $R_\infty$  come notably from higher-order interactions. This can be checked in panels (B,C,F) of Fig. 6.6, in which the contribution to the output variance from interactions between pairs of parameters (second order indices) is represented. Interactions among the parameters contribute to increasing the output variance with respect to  $t_{peak}$  and  $I_{peak}$ , while the effect is more heterogeneous in the case of  $R_\infty$ . In particular, the interactions between  $\alpha - \beta$  and  $\alpha - \mu$  produce the main contributions to the increase of output variance in all cases, while  $\kappa - \beta$ ,  $\kappa - \alpha$  and  $\kappa - \mu$  interactions decrease the output variance.

### 6.3.3 Epidemic control through vector management

The sensitivity analysis clearly indicates that acting on  $\alpha, \beta$  and  $\mu$  is the best strategy to lower disease incidence and mortality. However, controlling transmission rates is cumbersome so a different control strategy based only on vector control is considered in this section. In our model, there are two ways of implementing vector-population control: (i) decreasing the typical time,  $1/\mu$ , that vectors spend between crops each year by some mechanism (thus increasing  $\mu$ ) and (ii) reducing the initial number of vectors that invade crops each year (e.g. lowering  $N_v(0)$  via egg or nymph control (270)).

We analyzed the effect of vector management by simulating epidemic outbreaks using different values of  $\mu$  and  $N_v(0)$ , and keeping the rest of parameters as fitted for both ALSD and OQDS outbreaks (Fig. 6.7). In both epidemics, decreasing the presence time as well as the number of vectors contribute to controlling the epidemic by lowering  $R_0$  and, consequently, the final size of the epidemic,  $R_\infty$ . Furthermore, we observe that decreasing vector presence is more efficient than decreasing its annual initial population, i.e. we further reduce  $R_\infty$ , the final size of the epidemic, by applying a similar reduction in the residence time  $1/\mu$ . This could also be anticipated as  $R_0$  depends quadratically on  $1/\mu$  while only linearly on  $N_v(0)$  (Eq. (6.3)). However, the minimal intervention strategy, starting from the current situation in the  $(1/\tau, N_v(0))$  parameter space that yields an absolute control of the epidemic,  $R_0 < 1$ , involves a mixed strategy of lowering both  $1/\mu$  and  $N_v(0)$ .



**Figure 6.7:** Epidemic control through vector management for ALSD in Mallorca (A) and OQDS in Apulia (B). The white shaded line denotes  $R_0 = 1$ , the white diamond corresponds to the parameter values of the fitted model. The white star is the closest disease-free state to the current situation in this representation.

## 6.4 Discussion

In this work, we have developed a deterministic continuous-time compartmental model for *Xylella fastidiosa* vector-borne diseases in Europe. The model attempts to characterize the main biotic processes that lead to the development of epidemics, including the seasonal dynamics of the main vector, *P. spumarius*. We show how the model is sufficiently general to represent with some accuracy the parameters that determine the ALSD in Mallorca (Spain) and the OQDS in Apulia (Italy), both transmitted by *P. spumarius*. To our best knowledge, this is the first mathematical model describing Xf epidemics that considers the temporal pattern of vector abundance observed in field data, faithfully representing the known biological information about the pathosystem. It includes a dynamic approximation of the non-stationary populations of *P. spumarius*, mathematically represented by a sporadic source term through which vectors are born every year, and an exponential decay term. Due to the non-stationarity of the vector dynamics,  $R_0$  in the model cannot be computed with standard methods such as the Next Generation Matrix (138). To circumvent this problem, we applied an approximate method to compute it as previously proposed by (195). We show that this approximate  $R_0$  correctly characterizes the epidemic, further validating the method proposed by (195).

Nonlinear mathematical models of disease transmission enhance our understanding of the different mechanisms operating in an epidemic, especially compared with correlative or machine learning methods, often very useful in practice but offering very little understanding. A key aspect to render these models useful is the determination of the parameters from available data. If this step can be properly performed, these models become very predictive and especially helpful to design disease control strategies. However, an appropriate calibration of the model relies on access to good-quality field data, which is often the bottleneck for the application of this kind of models. In the present study, the parameters have been obtained using a Bayesian inference framework, which relies on probability distributions rather than point-like measures. This way, mean or median values can be considered together with their confidence intervals able to characterize the robustness of the obtained parameters. In general, we obtained different values of the parameters for the ALSD and OQDS outbreaks in Mallorca and Apulia, respectively. The fitted values, however, are in good agreement with previous field-based measures for each disease while the differences observed between both outbreaks may reflect differences between the Xf subspecies and crops involved (deciduous vs. evergreen).

One of the conclusions of the study is that the available data for both diseases is not enough to obtain robust estimates for all of the model parameters. The lack of data about the vector population compartments yields many possible values for the parameters that regulate transmission,  $\alpha$  and  $\beta$ , provided that the progression of the host compartments correctly fits the field data. In other words, very infectious vectors (high  $\beta$ ) that hardly ever get infected (low  $\alpha$ ) can produce a similar outbreak within the host population to that produced by very low infectious vectors (low  $\beta$ ) that get infected very often (high  $\alpha$ ). The great difference in these situations would be that, in the former, the infected vector population would be very low, while in the latter, it would be quite high. This is a manifestation of parameter unidentifiability from the fit (101, 372), which stresses the importance of transmission and calls for detailed measurements of the vector population, and not just of the hosts. Furthermore, to compare transmission rates between different diseases caused by Xf (e.g.  $\beta$ ,  $\alpha$ ), it is necessary to know the vector-host population ratio of the pathosystem ( $N_v/N_H$ ), since  $\beta$  is expressed as a number of hosts per vectors per day. Although, in general, populations of *P. spumarius* in the canopy of olive trees are much larger than those found in the almond trees of the Balearic Islands during the months of July and August (287), our work is based on data from studies in which information of the vector populations is not provided. Without this information, therefore, conclusive results regarding transmission cannot be obtained.

In any case, our model shows that the vector-to-plant transmission process, mediated by  $\beta$ , is somehow different from that from the plant-to-vector one, mediated by  $\alpha$ . In essence,  $\beta$  must be smaller than  $\alpha$  in order to reproduce the observed outbreaks and have a sufficiently large vector population getting infectious, being this fact independent of the particular choice of  $N_v(0)/N_H$ . This heterogeneity can be caused by several factors: differences in the efficiency of plant-to-vector transmission with respect to vector-to-plant transmission, differences in contact rates, i.e. susceptible vectors contact trees at a different rate than infected vectors; vector feeding

preferences, i.e. differences in the probability of contacting a susceptible host compared to an infectious host, etc. Indeed, our mathematical model assumes constant contact rates with no preferences over any host state, so that under these assumptions, it indicates that the probability of effectively transmitting the pathogen from plant to vector is greater than from vector to plant. However, this interpretation is subject to this particular assumption, so that to fully disentangle this question experimental work in form of transmission assays should be performed. Furthermore, we found that the timing and magnitude of the infectious host peak and the final number of dead hosts are mostly controlled by the vector-to-plant transmission rate,  $\beta$ , the plant-to-vector transmission rate,  $\alpha$  and the vector removal rate  $\mu$ . Because these parameters are strongly related to the vector, the analysis makes clear that enhancing the knowledge about the vector, as well as obtaining precise data, is crucial to improve the modeling of Xf diseases and pose important questions to be solved in specifically designed experiments.

The fact that the most influential parameters of the model are those related to the vector can be used to design appropriate disease control strategies. Because acting on transmission rates is rather cumbersome, we argue that control strategies should focus on reducing the vector population in crop fields. In our model, this depends on two parameters,  $\mu$ , the rate at which vectors die (or move to herbaceous vegetation and other non-host trees or exit the field) and  $N_v(0)$ , the number of newborns susceptible vectors every year (assumed constant in this study). Our results show that a mixed strategy acting on both parameters is optimal to lower disease prevalence and, eventually, eradicate the disease. Interestingly, we also show that acting on the vector removal  $\mu$  is more effective than controlling the newborn vector population  $N_v(0)$ . In fact, most control strategies carried out in practice for Xf diseases focus on the latter factor, reducing  $N_v(0)$  via egg or nymph control (113, 270, 286). However, our results indicate that alternative strategies based on increasing the removal (or dispersal) rate of vectors should be explored. Furthermore, the evolution of the population compartments of the hosts and vectors provides relevant information on the epidemiology of both diseases. In both cases, the newly defined basic reproductive number that accounts for a decaying vector population is very predictive of the moment in which new infections are not produced anymore, coinciding approximately with the peak of infectious hosts. Therefore, any intervention with control measures after this peak would have marginal effects on future disease progression.

Our mathematical model is still rather simple, implementing only a few relevant epidemic processes in contrast to the high complexity of the pathogen–vector–host interactions occurring in plant epidemics. Indeed, the model itself raises some questions about these interactions, for example, whether or not contact rates are homogeneous. Another simplification of the model is the fact that the spatial constraints and the intrinsic stochasticity of the transmission processes are neglected. A straightforward extension of the model would be to include a specific spatial setting and implement the explicit motion of the vector within a stochastic framework, such as Individual Based Models (214). With this, the effectiveness of current and further control strategies could be tested and improved controlling for the motion of the vector. For instance, the control strategy based on the removal of symptomatic trees together with their surrounding trees at a given distance could be implemented in the model, evaluate the current effectiveness according to the present protocols and even provide improved parameters to be implemented in the field. Of course, implementing a model in which the spatial degrees of freedom are explicitly represented would require access to further information about vector mobility and spatially resolved data to confront the model, which is not currently available.

Mathematical models tested against experimental data increase our understanding of the system under study. They also help to identify critical parameters that require better prior information to adjust functions relating to different variables and make the model predictions more accurate to suggest and test control strategies (123, 242). Our mathematical model suggests a certain lack of knowledge of the transmission processes and reveals that the currently available data is not enough to fit complex models dealing with the explicit dynamics of the vector population.

## 6.5 Appendix

### 6.5.1 Vector population dynamics

Fig. 6.8 shows a time series for the population of *Philaenus spumarius* in Mallorca, taken from (287) (in blue). Superimposed (in orange) is the assumption used in our model Eq. (6.2), the  $\delta(t - nT)$ , i.e. every year susceptible vectors appear in the system.

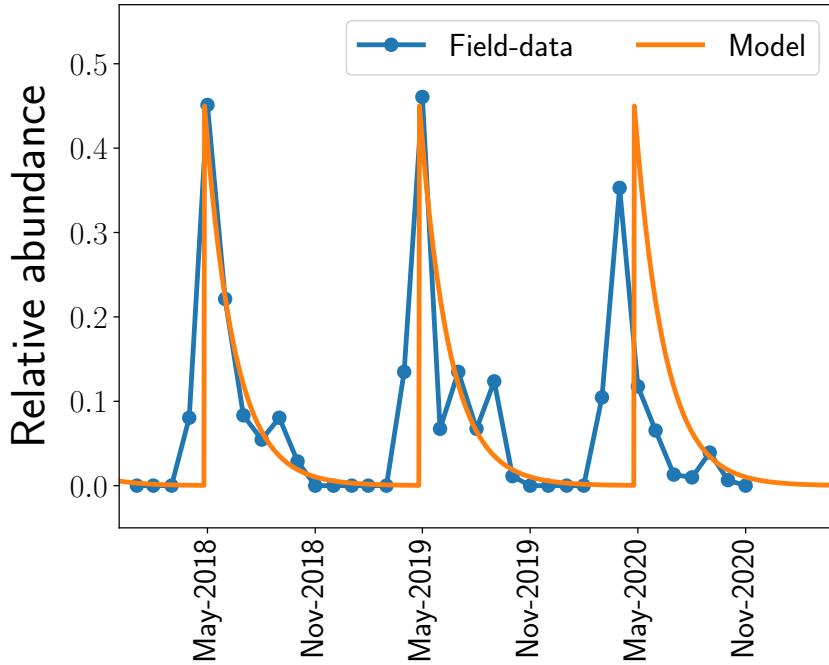


Figure 6.8: Vector dynamics produced by the model compared to field-data from [287].

### 6.5.2 Determination of $R_0$

The handicap of determining the basic reproductive number of the model Eq. (6.2) is that the pre-pandemic fixed point given by  $I_H = I_v = 0$  and  $S_H = S_H(0)$  is not a fixed point of the system of differential equations, because vector population decays, so that the standard methods to compute  $R_0$  such as the Next Generation Matrix (138, 195) do not apply. In (195) a method was suggested to determine the basic reproductive number in the case of compartmental models of vector-borne transmitted diseases in which the vector population grows or decays. It consists in averaging the instantaneous basic reproductive number over the time of a generation.

To proceed we consider that  $I_H = I_v = 0$ ,  $S_H = S_H(0)$  is indeed a fixed point of the system. Then, the basic reproductive number could be determined, e.g. as shown in (63). First, an infectious host infects vectors at a rate  $\beta S_H(0)/N_H$  for a time  $1/\gamma$ . This produces  $\beta S_H(0)/\gamma N_H$  infected vectors. The second stage is that these infectious vectors infect hosts at a rate  $\alpha N_v(0)/N_H$  for a time  $1/\mu$ , producing  $\alpha N_v/\mu N_H$  infectious hosts per vector. The net result of these two stages is

$$\tilde{R}_0 = \frac{\alpha\beta}{\mu\gamma} \frac{S_H(0)}{N_H^2} N_v(0) = R_0^* \cdot N_v(0) . \quad (6.4)$$

This result coincides with the value of  $R_0$  obtained using the standard NGM method, that can be applied in this case because we are assuming that we use a nongeneric initial condition that sits at the fixed point of the model.

In practice, our initial condition will never be a fixed point of the model, and, as mentioned above, we will obtain an approximate basic reproductive number, to which we will refer as  $R_0$  using the method suggested in (195), that consists in calculating the average number of secondary infections produced by an infectious host in one generation. One first defines an

instantaneous basic reproductive number,

$$R_0^{(i)}(t) = \frac{\beta\alpha}{\mu\gamma} \frac{S_H(0)}{N_H^2} N_v(t) = R_0^* N_v(t) , \quad (6.5)$$

from which the average is simply computed as

$$R_0 = \left\langle R_0^{(i)}(t) \right\rangle \Big|_0^\tau = R_0^* \langle N_v(t) \rangle \Big|_0^\tau = R_0^* \frac{1}{\tau} \int_0^\tau N_v(t) dt . \quad (6.6)$$

In our model, the time-dependent vector population can be obtained from Eq. (6.2),

$$\dot{N}_v = \dot{S}_v + \dot{I}_v = -\mu N_v \implies N_v(t) = N_v(0) e^{-\mu t} , \quad (6.7)$$

and introducing this expression for  $N_v(t)$  in Eq. (6.6) the integral can be solved

$$R_0 = \frac{\beta\alpha S_H(0)}{\mu\gamma N_H^2} \frac{N_v(0)}{\mu\tau} (1 - e^{-\mu\tau}) = R_0^* \frac{N_v(0)}{\mu\tau} (1 - e^{-\mu\tau}) , \quad (6.8)$$

that is an approximated expression to the basic reproductive number for our model, in which the vector population is nonstationary, where, in Eq. (6.5) and Eq. (6.8) it has been defined,  $R_0^* = (\beta\alpha S_H(0)) / (\mu\gamma N_H^2)$ .

Note that in our model one generation correspond to one year and that  $N_v(0)$  is reset every year.



# Modelling the risk of vector-borne plant diseases

<b>7</b>	<b>Global predictions for the risk of establishment of Pierce's disease of grapevines .....</b>	<b>77</b>
7.1	Introduction .....	78
7.2	Results .....	79
7.3	Discussion .....	87
7.4	Methods .....	89
<b>8</b>	<b>Pierce's disease epidemic risk under global warming .....</b>	<b>95</b>
8.1	Introduction .....	96
8.2	Methods .....	97
8.3	Results .....	100
8.4	Discussion .....	105
<b>9</b>	<b>Pierce's Disease risk with high-resolution climate data .....</b>	<b>109</b>
9.1	Introduction .....	110
9.2	Results .....	111
9.3	Discussion .....	116
9.4	Methods .....	117





## 7. Global predictions for the risk of establishment of Pierce's disease of grapevines

Àlex Giménez-Romero<sup>1</sup>, Javier Galván<sup>1</sup>, Marina Montesinos<sup>2</sup>, Joan Bauzà<sup>3</sup>, Martin Godefroid<sup>4</sup>, Alberto Fereres<sup>4</sup>, José J. Ramasco<sup>1</sup>, Manuel A. Matías<sup>1</sup>, Eduardo Moralejo<sup>2</sup>

1. Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Palma de Mallorca 07122, Spain
2. Tragsa, Passatge Cala Figuera 6, 07009 Palma de Mallorca, Spain
3. Departamento de Geografía, Universidad de las Islas Baleares, Campus UIB, 07122, Palma de Mallorca, Spain
4. Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas, ICA-CSIC, 28006, Madrid, Spain

### Published as

Alex Giménez-Romero et al. "Global predictions for the risk of establishment of Pierce's disease of grapevines". In: *Communications Biology* 5.1 (Dec. 2022), page 1389. ISSN: 2399-3642. DOI: [10.1038/s42003-022-04358-w](https://doi.org/10.1038/s42003-022-04358-w). URL: <https://doi.org/10.1038/s42003-022-04358-w>

## 7.1 Introduction

Emerging plant pathogens and pests are costly both economically and environmentally for society (87, 312, 351, 409). Among valuable crops recurrently affected by emerging diseases, the grapevine occupies a remarkable place in the history of plant pathology (51, 65, 373, 415). Nowadays, Pierce's disease (PD) is considered a potential major threat to winegrowers worldwide (234). The annual economic burden in California alone has been estimated at over \$100 million (423), and the disease is a well-recognised limiting factor in the cultivation of *Vitis vinifera* in the southeastern US (234). In Europe, despite strict quarantine measures to protect the wine industry (Directive 2000/29/EC), PD has recently been established for the first time in vineyards on the island of Majorca, Spain (209, 313). This finding, alongside the detection of PD in Taiwan (411), has raised concerns about its possible spread to continental Europe and other wine-producing regions worldwide.

The causal agent of PD (129), the bacterium *Xylella fastidiosa* (Xf) (444), is native to the Americas where it also causes vector-borne diseases on many economically important crops, such as citrus, almond, coffee and olive trees (8, 10). Xf is phylogenetically subdivided into three major monophyletic clades that correspond to the three formally recognised subspecies: *fastidiosa*, *multiplex* and *pauca*, native from Central, North and South America, respectively (400, 430). Although as a taxonomic unit Xf infects more than 560 plant species (130), it also shows genetic variation among subspecies and sequence types (STs) in both host specificity and host range (334). Since 2013, diverse STs of the three subspecies have been detected in Europe mainly associated with crop and ornamental plants (133, 336, 379); among these, the clonal lineage of the subsp. *fastidiosa* responsible for PD (hereafter termed Xf<sub>PD</sub>). The same genetic lineage also causes almond leaf scorch disease in California (9) and Majorca (Spain) (314), where it is widespread in almond plantations and vineyards, affecting more than 23 grape varieties (313).

A key trait in the understanding of Xf's invasive potential is its capacity of being transmitted non-specifically by xylem sap-feeding insects belonging to sharpshooter leafhoppers (Hemiptera: Cicadellinae) and spittlebugs (Hemiptera: superfamily Cercopidae) (7, 113) – e.g., at least eight species transmit PD in the southeastern US (341). Such non-specificity would have facilitated Xf<sub>PD</sub> invasion after being unwittingly brought to Majorca around 1993 with infected almond cuttings from California and its spread thereafter to grapevines through local populations of the meadow spittlebug, *Philaenus spumarius* (314). Recently, the role of *P. spumarius* in the transmission of PD in Majorca has been demonstrated (313) and its involvement in epidemic outbreaks in California, previously thought marginal (363, 396), is being revisited (30, 110). To date, the meadow spittlebug has been confirmed as the major vector in the olive quick decline syndrome, PD and the almond leaf scorch disease outbreaks in Europe (113, 114, 313, 314); therefore, its geographic distribution should be taken into account when assessing the risk of Xf-related diseases (204).

The tropical origin of Xf subsp. *fastidiosa* already suggests that PD is a thermal-sensitive disease, with the temperature being a range-limiting factor (89, 357). Thus, the accumulated heat units (i.e., growing-degree days) required to complete the process from Xf<sub>PD</sub> infection to symptom development is critical to predicting the probability of developing PD acute infections (155). Conversely, the effect of cold-temperature exposures in the recovery of Xf-infected grapevines is a well-established phenomenon (155, 283, 356), limiting the geographic range and damage of chronic PD in vineyards in the US (234). Such "winter curing" has been linked to the average  $T_{min}$  of the coldest month, to exposures to extreme cold temperatures for several days, or to the accumulation of chilling hours (17). The dynamics of chronic infections –i.e., those that persist from one year to the next year– are determined by the net balance between the number of new infections during the growing season and those infected plants recovered in winter. Because new infections late in the growing season are more likely to recover during winter than early-season infections, the vector's phenology greatly influences the dynamics of chronic infections and PD transmission (128, 154, 215, 363).

Several works have attempted to predict the potential geographic range of the subsp. *fastidiosa* (59, 203, 231) and other Xf subspecies in Europe (54, 391) and worldwide (231) using bioclimatic correlative species distribution models (SDMs). However, none of these works has explicitly included information on vectors' distribution or disease dynamics. They hence provide

little epidemiological insight into the underlying environmental causes underpinning or limiting a potential invasion. An alternative to overcome these limitations is to develop mechanistic models based on the physiology of the pathogen (256), coupled with epidemiological models that consider the disease dynamics while avoiding the difficulties of including transmission parameters for each of the PD potential vectors.

Risk maps often represent an average snapshot that overlooks interannual climate variability and the effects of climate change as limiting disease factors *per se*. This leads frequently to risk overestimation (31, 106, 387, 422). Increased availability of computational resources to deal with demanding climate databases now makes it possible to fit dynamic epidemiological models that include climate variability at broad spatiotemporal scales. For example, high-resolution satellite-based climate data have been employed for testing mechanistic models that relate critical physiological processes of coffee rust with climate variables in past outbreak events (32). Despite these important advances, no attempt of exploring mechanistic SDM has been performed yet for PD.

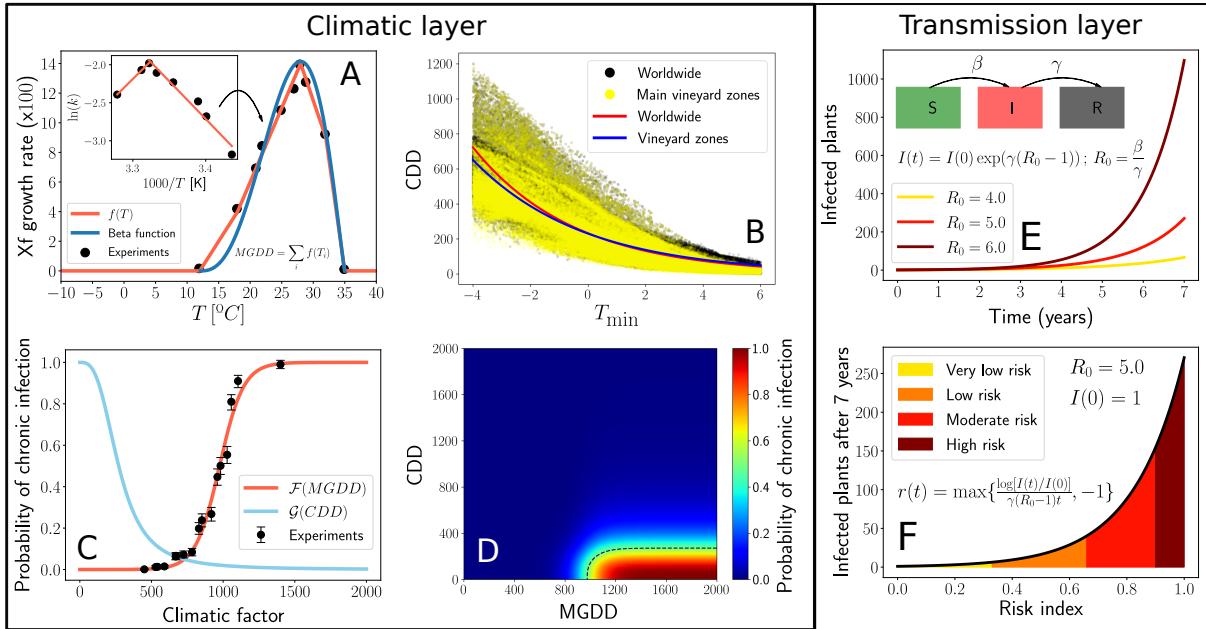
In this work, we present a temperature-driven dynamic epidemiological model to infer where PD would have become endemic in different wine-growing regions worldwide from 1981 onward if we forced the introduction of Xf-infected plants. We follow an invasive criterion as defined by Jeger & Bragard (241) to include, as far as we can, key plant, pathogen, and vector parameters and their interactions for estimating the risk of establishment, persistence, and subsequent epidemic development. The model assumes a local  $Xf_{PD}$  spatial propagation among plants mediated by the presence of potential vectors. Due to the limited knowledge about the vectors of PD in most wine-growing regions of the world (363), we employ a fixed maximal estimate for basic reproductive numbers ( $R_0$ ) in the epidemiological models, except for Europe, where there are precise estimations of climate suitability for the main vector *P. spumarius* (204). This heuristic approach to obtaining PD risk maps yields results that are consistent with all the relevant data available (59). It also allows us to quantitatively approximate the current potential growth rate of PD incidence in wine-growing regions under different transmission scenarios, as well as extrapolate the impact of PD by 2050 (188). By estimating a lower global risk of PD, our study casts doubts on the potential impact predicted for other Xf-related diseases transmitted by *P. spumarius* (391), specially in Europe when vector distribution is taken into account.

## 7.2 Results

### 7.2.1 Thermal requirements to develop PD.

We examined the response of a wide spectrum of European grapevine varieties to  $Xf_{PD}$  infection in three independent experiments conducted in 2018, 2019, and 2020. Overall, 86.1% ( $n = 764$ ) of 886 inoculated plants, comprising 36 varieties and 57 unique scion/rootstock combinations, developed PD symptoms 16 weeks after inoculation. European *V. vinifera* varieties exhibited significant differences in their susceptibility to  $Xf_{PD}$  (Online Supplementary Information). All varieties, however, showed PD symptoms to some extent, confirming previous field observations of general susceptibility to  $Xf_{PD}$  (234, 313, 357). We also found significant differences in virulence ( $\chi^2 = 68.73$ ,  $df = 1$ ,  $P = 2.2 \times 10^{-16}$ ) between two  $Xf_{PD}$  strains isolated from grapevines in Majorca across grapevine varieties (??). Full details on the results of the inoculation tests are available in Methods and Supplementary Information.

Growing degree days (GDD) have traditionally been used to describe and predict phenological events of plants and insect pests, but rarely in plant diseases (302). We took advantage of data collated in the inoculation trials together with temperature to relate symptom development to the accumulated heat units at weeks eight, 10, 12, 14 and 16 after inoculation (Supplementary Data 1). Rather than counting GDDs linearly above a threshold temperature, we consider Xf's specific growth rate *in vitro* within its cardinal temperatures. The empirical growth rates come from the seminal work by Feil & Purcell (155) shown in the inset of Fig. 7.1A. This Arrhenius plot is transformed, as explained in Online Supplementary Information, to obtain a linear approximation in a limited range of temperatures as shown in the main plot of Fig. 7.1A. Inspired by the fit in Fig. 3 of (155), we approximate the growth rates by a piece-wise function  $f(T)$  (see Methods below) formed by four segments. This Modified Growing Degree Day (MGDD) profile enables to measure the thermal integral from hourly average temperatures, improving the prediction scale



**Figure 7.1: Climatic and transmission layers composing the epidemiological model.** (A) MGDD profile fitted to the *in vitro* data of *Xf* growth rate in Feil & Purcell 2001 [155]. The original Arrhenius plot in Kelvin degrees (inset) was converted to Celsius, as explained in Online Supplementary Information, to obtain the main plot. (B) Correlation between CDD and the average  $T_{\min}$  of the coldest month between 1981 and 2019. Plotted black dots (worldwide) and yellow dots (main wine-producing zones) depict climatic data from 6,487,200 cells at  $0.1^{\circ} \times 0.1^{\circ}$  resolution, spread globally and retrieved from ERA5-Land dataset. The red solid line depicts the fitted exponential function for worldwide data and the blue solid line for main vineyard zones. (C) Nonlinear relationship between MGDD (red line) and CDD (blue line) and the likelihood of developing chronic infections. Black dots depict the cumulative proportion of grapevine plants in the population of 36 inoculated varieties showing five or more symptomatic leaves at each of the 15 MGDD levels (see Online Supplementary Information). Vertical bars are the 95% CI. (D) Combined ranges of MGDD and CDD on the likelihood of developing chronic infection. (E) Transmission layer in the dynamic equation (1) of the SIR compartmental model. (F) Relationship between the exponential growth of the number of infected plants with the risk index and their ranks.

of the biological process [74]. MGDD also provides an excellent metric to link *Xf<sub>PD</sub>* growth in culture with PD development as, once the pathogen is injected into the healthy vine, symptoms progression mainly depends upon the bacterial load (i.e., multiplication) and the movement through the xylem vessel network, which are fundamentally temperature-dependent processes [155, 170]. Moreover, MGDD can be mathematically related to the exponential or logistic growth of the pathogen within the plant (Online Supplementary Information).

Interannual infection survival in grapevines plays a relevant role when modelling PD epidemiology. In our model, we assumed a threshold of five or more symptomatic leaves for these chronic infections based on the relationship between the timing and severity of the infection during the growing season and the likelihood of winter recovery [154, 155, 283]. This five-leaf cut-off was grounded on: (i) the bimodal distribution in the frequency of the number of symptomatic leaves among the population of inoculated grapevines [??], whereby vines that generally show less than five symptomatic leaves at 12 weeks after inoculation remain so in the following weeks, while those that pass that threshold continue to produce symptomatic leaves, and (ii) the observed correlation between the acropetal and basipetal movement of *Xf* along the cane [??]. The likelihood of developing chronic infections as a function of accumulated MGDD among the population of grapevine varieties was modelled using survival analysis with data fitted to a logistic distribution  $\mathcal{F}(MGDD)$ . A minimum window of  $MGDD = 528$  was needed to develop chronic infections (var. Tempranillo), about 975 for a median estimate, while a cumulative  $MGDD > 1159$  indicate over 90% probability within a growing season (red curve in Fig. 7.1C and Methods).

Next, we intended to model the probability of disease recovery by exposure to cold tem-

peratures. Previous works had specifically modelled cold curing on Pinot Noir and Cabernet Sauvignon varieties in California as the effect of temperature and duration (283) by assuming a progressive elimination of the bacterial load with cold temperatures (154). In the absence of appropriate empirical data to formulate a general average pattern of winter curing among grapevine varieties, we combined the approach of Lieth et al. (283) and the empirical observations of Purcell on the distribution of PD in the US related to the average minimum temperature of the coldest month,  $T_{min}$ , isolines (17). To consider the accumulation of cold units in an analogy of the MGDD, we searched for a general correlation between  $T_{min}$  and the cold degree days (CDDs) with base temperature = 6 °C (see Methods). We found an exponential relation,  $CDD \sim 230 \exp(-0.26 \cdot T_{min})$ , where specifically,  $CDD \gtrsim 306$  correspond to  $T_{min} < -1.1^\circ\text{C}$  Fig. 7.1B. To transform this exponential relationship to a probabilistic function analogous to  $\mathcal{F}(MGDD)$ , hereafter denoted  $\mathcal{G}(CDD)$ , ranging between 0 and 1, we considered the sigmoidal family of functions  $f(x) = \frac{A}{B+x^C}$  with  $A = 9 \cdot 10^6$ ,  $B = A$  and  $C = 3$  (Fig. 7.1C), fulfilling the limit  $\mathcal{G}(CDD = 0) = 1$ , i.e. no winter curing when no cold accumulated, and a conservative 75% of the infected plants recovered at  $T_{min} = -1.1^\circ\text{C}$  instead of 100% to reflect uncertainties on the effect of winter curing.

## 7.2.2 MGDD/CDD distribution maps.

MGDD were used to compute annual risk maps of developing PD during summer for the period 1981-2019 (see Methods). The resulting averaged map identifies all known areas with a recent history of severe PD in the US corresponding to  $\mathcal{F}(MGDD) > 90\%$  (i.e., high-risk), such as the Gulf Coast states (Texas, Alabama, Mississippi, Louisiana, Florida), Georgia and Southern California sites (e.g., Temecula Valley) (Fig. 7.2A), while captures areas with a steep gradation of disease endemicity in the north coast of California ( $\mathcal{F}(MGDD) > 50\%$ ). Overall, more than 95% of confirmed PD sites ( $n = 155$ ) in the US (Supplementary Data 2) fall in grid cells with  $\mathcal{F}(MGDD) > 50\%$ .

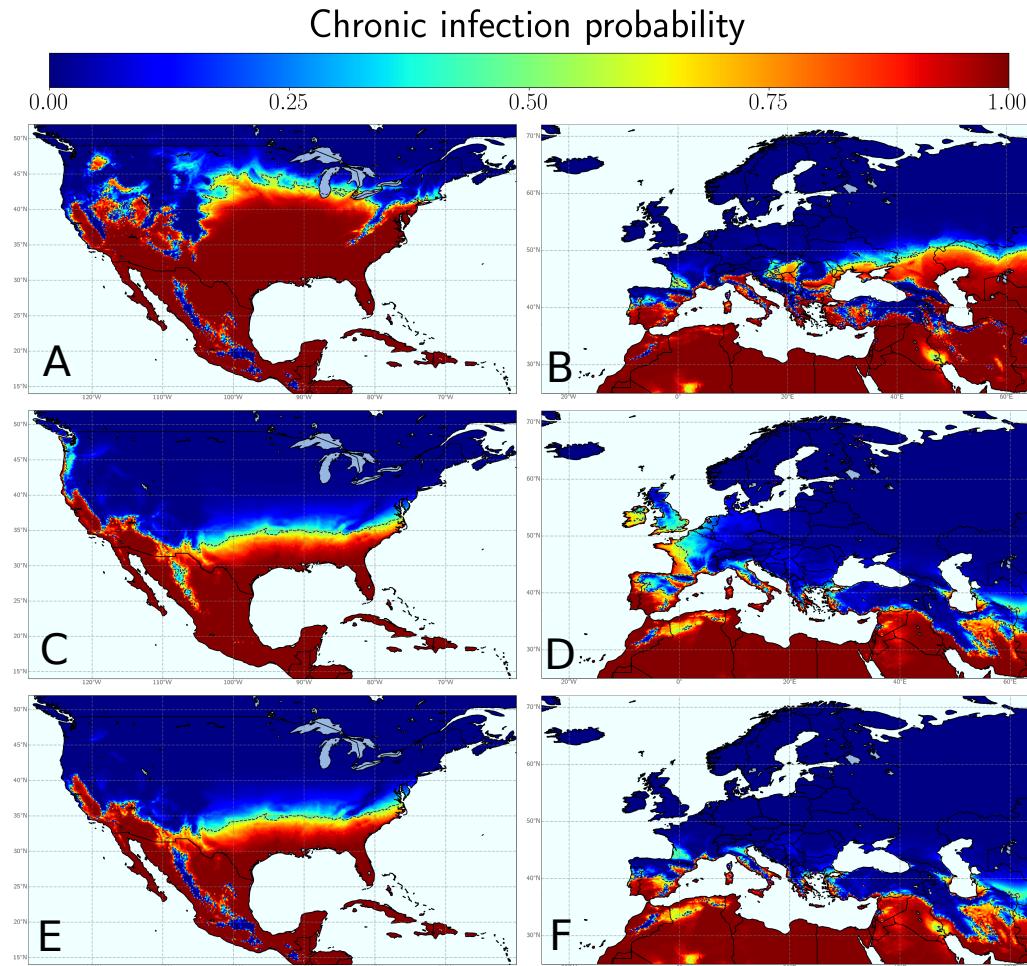
The average MGDD-projected map for Europe during 1981-2019 spots a high risk for the coast, islands and major river valleys of the Mediterranean Basin, southern Spain, the Atlantic coast from Gibraltar to Oporto, and continental areas of central and southeast Europe (Fig. 7.2B). Of these, however, only some Mediterranean islands, such as Cyprus and Crete, show  $\mathcal{F}(MGDD) > 99\%$  comparable to areas with high disease incidence in the Gulf Coast states of the US and California. Almost all the Atlantic coast from Oporto (Portugal) to Denmark are below suitable MGDD, with an important exception in the Garonne river basin in France (Bordeaux Area) with low to moderate MGDD (Fig. 7.2B).

Fig. 7.2A shows how the area with high-risk MGDD values extends further north of the current known PD distribution in the southeastern US, suggesting that winter temperatures limit the expansion of PD northwards (234). A comparison between MGDD and CDD maps (Fig. 7.2A vs. Fig. 7.2C, Fig. 7.2E) further supports the idea that winter curing is restricting PD northward migration from the southeastern US. However, consistent with growing concern among Midwest states winegrowers on PD northward migration led by climate change (171), we found a mean increase of 0.12%  $\text{y}^{-1}$  in the areal extent with  $CDD < 306$  ( $\sim T_{min} < -1.1^\circ\text{C}$ ) since 1981, comprising land areas between 103°W and 70°W of the US (??). Such an upward trend corresponds to 5090  $\text{km}^2 \text{y}^{-1}$  in the potential northward expansion of PD due to climate change and an accumulation of  $\sim 193420 \text{ km}^2$  of new areas at risk since 1981.

High-CDD values would also be expected to restrict the potential PD colonisation in continental Europe (Fig. 7.2D). Unlike North America, the East-West distribution of major European mountain ranges together with the warming effect of the Gulf Stream decreases the likelihood of cold winter spells reaching the western Mediterranean coast.  $\mathcal{G}(CDD)$  between 100% and 95% (i.e., recovery probability < 5% – low winter curing) are mostly prevalent below 40°N latitude in the southwest Iberian Peninsula and Mediterranean islands and coastlands (< 50 km away). Above 40°N latitudes,  $CDD < 100$  are encountered mainly in the Atlantic coast and Mediterranean coast and islands (Fig. 7.2D). In contrast, central and southeast Europe show high CDD values likely preventing  $Xf_{PD}$  winter survival on infected grapevines.

In Fig. 7.2E-F, we show the average climatic suitability for PD establishment only from the mechanistic relation between  $Xf_{PD}$  and temperature. Although all areas with current  $Xf_{PD}$ -related outbreaks are identified, risk predictions based only on the combination of MGDD and CDD could lead to overestimations, as this approach overlooks disease transmission dynamics and

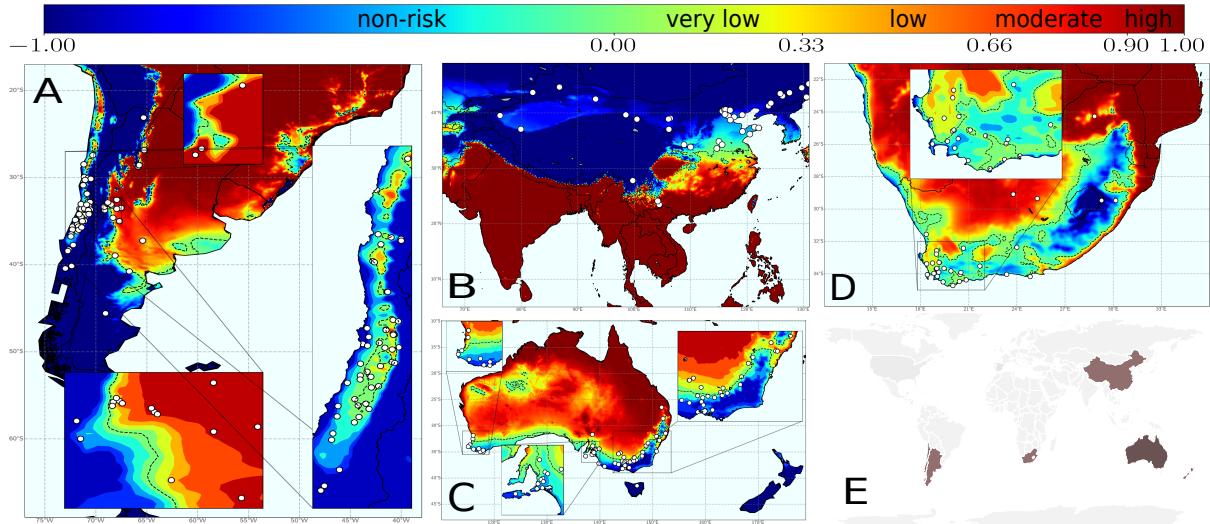
climate interannual variability.



**Figure 7.2: Average thermal-dependent maps for Pierce's disease (PD) development and recovery in North America and Europe.** PD development during the growing season based on average  $\mathcal{F}(\text{MGDD})$  estimations between 1981 and 2019 in North America (**A**) and Europe (**B**) derived from the results of the inoculation experiments on 36 grapevine varieties. Large differences in the areal extension with favourable MGDDs can be observed between the US and Europe. The winter curing effect is reflected in the distribution of the average  $\mathcal{G}(\text{CDD})$  for the 1981–2019 period in the United States (**C**) and Europe (**D**). A snapshot of the temperature-driven probability of chronic infection averaged for the 1981–2019 period is obtained from the joint effect of MGDD and CDD in North America (**E**) and Europe (**F**). Warmer colours indicate more favourable conditions for chronic PD and the dashed line highlights the threshold of chronic infection probability being 0.5.

### 7.2.3 PD global risk.

We ran several simulations of the model Eq. (7.7) with  $R_0$  values between 1 and 14 to validate PD spatiotemporal distribution in the US. We found  $R_0 = 8$  as the optimal parameter for maximising the area under a ROC curve (??), returning an accuracy of more than 80%, except for 2006, due to data obtained from an area at the transient-risk zone (?? and Table 7.1). For Europe and the rest of the world, we derived a  $R_0 = 5$ , as a maximal baseline estimate for modelling PD transmission (see Methods and Online Supplementary Information). These  $R_0$  values should be taken as operating estimates for the model. From the model simulations Eq. (7.7), we obtained a risk index  $r$  that measures the relative exponential growth rate in the population of infected plants at the epidemic onset with respect to the maximum growth,  $r = 1$ . This index served to rank the epidemic-risk zones in high ( $> 0.9$ ), moderate ( $0.66 – 0.9$ ), low ( $0.33 – 0.66$ ) and very low ( $\sim 0.075 – 0.33$ ) risks (see Fig. 7.1F, Methods, and Online Supplementary Information).



**Figure 7.3: Climate-driven risk maps for PD establishment in main viticulture regions worldwide under a baseline  $R_0 = 5$  scenario.** White dots indicate the main vineyard areas in the wine-growing regions of China and the Southern Hemisphere. (A) Chile and Argentina; (B) Asia with special attention to China; (C) Australia and New Zealand (wine areas are not marked as the whole country is without risk); and (D) South Africa. (E) Global distribution of main wine-producing areas analysed. The risk index  $r_j(t)$ , express the relative exponential growth rate of the disease incidence, and was scaled from 0.1 to 1 and ranked as very low (0.10-0.33), low (0.33-0.66), moderate (0.66-0.90) and high ( $> 0.90$ ).

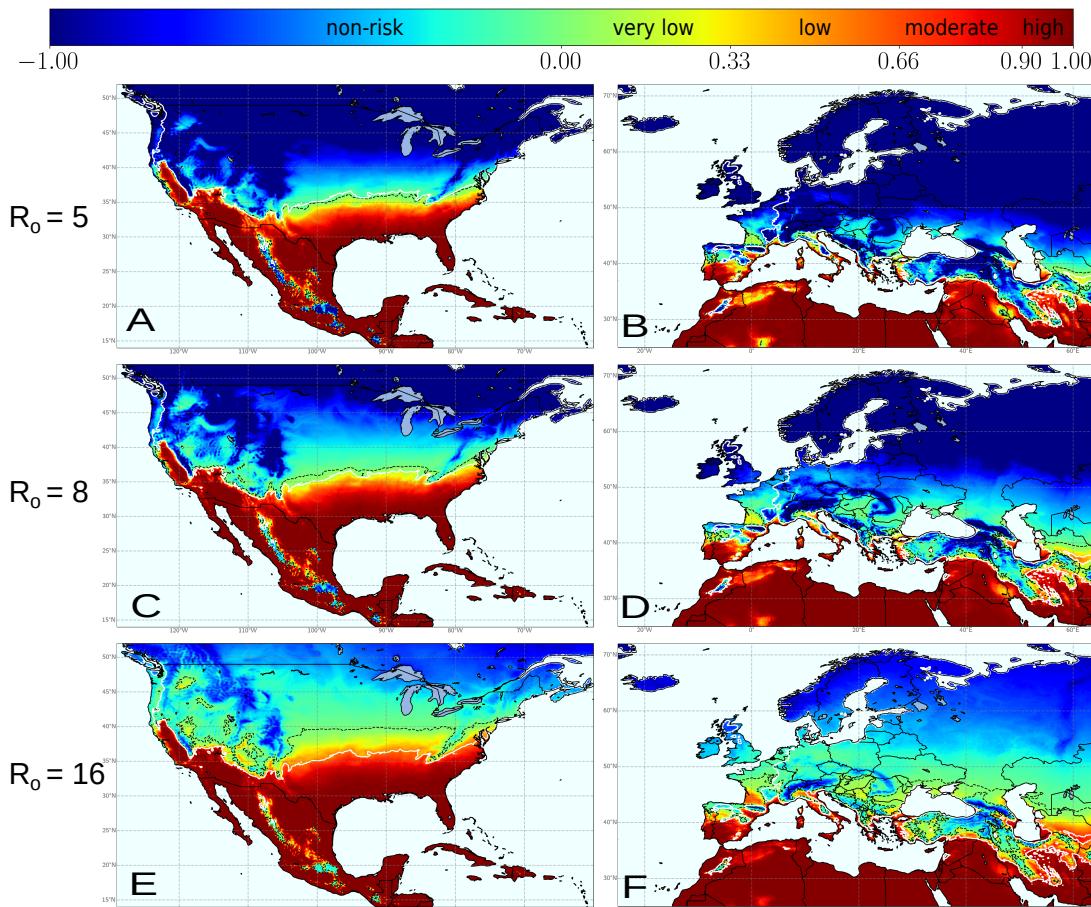
**Table 7.1: Validation of model predictions.** The items are locations where PD was present or absent. TP corresponds to true positives and TN to true negatives according to our model with  $R_0 = 8$ .

Year	Presence	Absence	TP	TN	Accuracy
2001	16	5	15	3	86%
2002	12	2	11	1	86%
2005	4	2	4	1	83%
2006	8	0	4	0	50%
2015	53	0	51	0	96%
<b>TOTAL</b>	<b>93</b>	<b>9</b>	<b>85</b>	<b>5</b>	<b>88%</b>

To date, PD is mainly restricted to the American continent with some unrelated introductions of  $Xf_{PD}$  to Taiwan and Majorca (Spain) from the United States (313, 411). To assess the risk of PD establishment elsewhere, we projected our epidemiological model into the main winegrowing regions of the Northern Hemisphere (US, Europe and China) and Southern Hemisphere (Chile, Argentina, South Africa, Australia and New Zealand)(Fig. 7.3A-E). We found that emerging wine-producing areas in China are predominantly located in non-risk zones, whereas only some vineyards in the Henan and Yunnan provinces fall in transition and moderate-high risk zones (Fig. 7.3B and Supplementary Data 3). In Europe, 92.1% of the territory is in non-risk zones and 6.1% is included in the epidemic-risk zone, with only 1.9% showing a high-risk index and 1.5% a moderate risk (Online Supplementary Information). The model also reveals a progressive transition from areas without risk ( $r(t) < 0$ ) before 1990 to epidemic-risk zones with low-risk indexes by 2019 ((188), see Movies), mainly affecting the basins of the rivers Po in Italy, Garonne and Rhone in France and Douro/Duero in Portugal and Spain. This represents a mean increase of  $0.21\% \text{ y}^{-1}$  in the epidemic-risk zone, a rate 3.5-times greater than that of the eastern US, which could increase the likelihood of PD establishment in Europe in the coming decades. In the US, most states

around the Gulf Coast show high-risk indexes, whereas, around 37.5% of California's surface is suitable for epidemics with high growth rate incidence (Online Supplementary Information).

In the Southern Hemisphere, vineyards at non-risk or transient epidemic-risk zones predominate – e.g., non-risk in New Zealand and Tasmania (Fig. 7.3C). Risk indexes in areas where PD can become established ( $r(t) > 0$ ) range from very low to low for most coastal vineyards in Australia (west, south and east) with somehow more suitable conditions in the interior of New South Wales, Greater Perth and Queensland (Fig. 7.3C); a general very-low or low-risk indexes are predicted in the Western Cape in South Africa (Fig. 7.3D); overall very-low but localised low to moderate risk indexes in some areas in Chile; and low to moderate growth of the number of infected vines in most of Argentina, being this the wine-growing country with the highest risk (Fig. 7.3A). Detailed information on areas with non-risk, transient risk and risk indexes (i.e., disease-incidence growth rates) in areas with the potential risk of establishment by country and regions is provided in Supplementary Information.



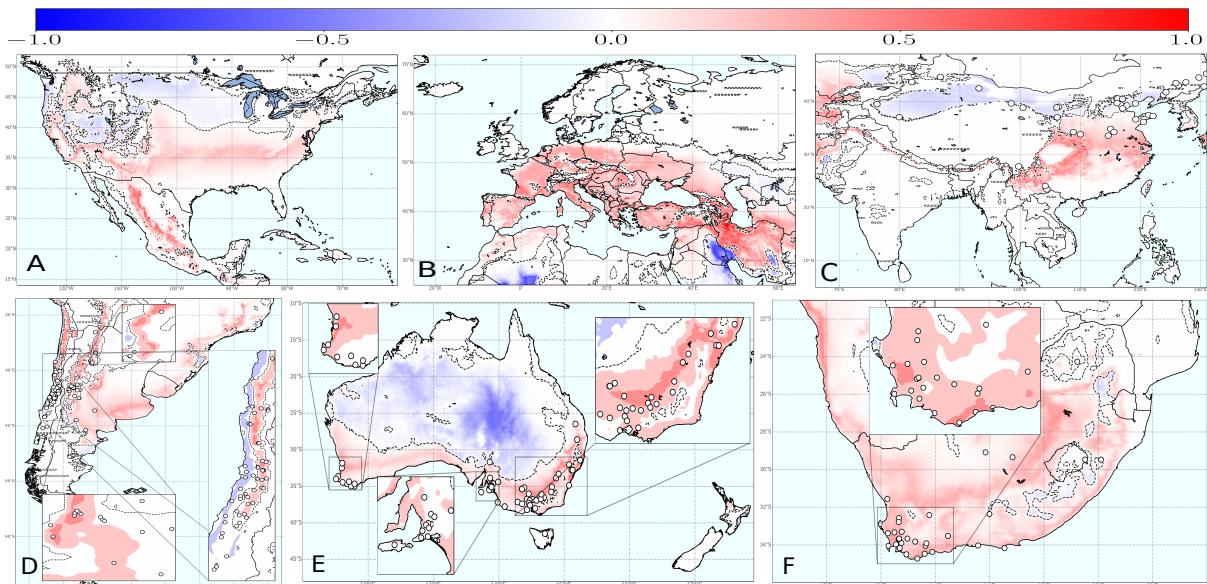
**Figure 7.4: Temperature-driven dynamic-model simulations for PD establishment from 1981 to 2019 under different  $R_0$  scenarios with a spatially homogeneous vector distribution.** For comparison, the baseline scenario with a  $R_0 = 5$  for Europe is projected to North America (A) capturing to some extent the distribution and severity of PD in that continent. In Europe (B) high-risk areas (i.e.,  $r_j(t) > 0.90$ ) are restricted to coastal Mediterranean and the south of the Iberian Peninsula; black dash line separate areas with  $r(t) > 0$  where theoretically PD can thrive. Under higher  $R_0$  scenarios,  $R_0 = 8$  for North America (C) and Europe (D), the dash lines tend to separate from isoline  $T_{\min} = -1.1^\circ\text{C}$  (white line); and even more in extreme transmission pressure  $R_0 = 16$  for North America (E) and Europe (F).

Risk indexes may vary within epidemic-risk zones if any of the epidemiological parameters governing transmission change. As expected,  $I(t) < I(0)$  boundaries increasingly displace to northern latitudes in the US and Europe under higher transmission scenarios, increasing the risk-epidemic zones significantly (Fig. 7.4A-F). The line representing the outbreak extinction i.e., the non-risk zone  $r(t) < -0.09$ , in the validated  $R_0 = 8$  scenario for the US, falls at some distance above

the isoline  $T_{\min} = -1.1^{\circ}\text{C}$  in comparison to the  $R_0 = 5$  scenario (Fig. 7.4, (188)). This distribution pattern holds and moves slightly northward over time in parallel to global warming, although the trend of PD latitudinal change is moderated by high-CDD values (i.e. cold accumulation). In addition, the disease extension also declines due to CDD interannual fluctuations in the simulations. Cold waves periodically occur that reach latitudes close to the Gulf, such as those that occurred in 1983, 1993, 1995, 2000, 2009 and 2013 (Movies at (188)), thus preventing PD expansion northward. The magnitude of this decrease is revealed after comparing the average annual increase of the areas between  $r(t) > 0$  and  $CDD < 306$  lines. From 1981 to 2019, the area with risk  $r(t) > 0$  increased at a rate of  $0.05\% \text{ y}^{-1}$ , while that of  $CDD < 306$  by  $0.12\% \text{ y}^{-1}$ , an important difference not explained alone by CDDs without considering climate fluctuations (??).

### 7.2.4 PD risk projections for 2050

. Global shifts in the risk index  $r_j(t)$  between 2019 and those projected for 2050 were calculated under the same baseline scenario (Fig. 7.5A-F, Methods). Our simulation shows a generalised increasing trend mainly due to shifts from transition zones to epidemic-risk zones with very low or low-risk indexes in the main wine-growing regions, except for the US. Here the epidemic-risk zone would increase by 12.8% with the greater increments in the high-risk index category (22.7%) and a decrease in the transition zones (Online Supplementary Information). Much less surface would be included in the epidemic-risk zone in Europe (8.6%) compared to the US (36.5%). However, the epidemic-risk zone would expand by 40.0% with respect to 2020, a rate more than three times higher than that of the US (Online Supplementary Information). Such increases are due to the emergence of previously unaffected areas in 2020 evolving into epidemic-risk zones by 2050, and epidemic growth-rate increases in already epidemic-risk zones in 12 of 42 countries (Online Supplementary Information). Among these 12 countries, however, there is substantial variation in the risk index increments within epidemic-risk zones with respect to 2019 (Online Supplementary Information). While non-risk zones still cover 87.6% of Europe's land area, epidemic-risk zones with high-risk indexes are expected to be almost two-fold higher than that of 2019, comprising 3.2% of Europe (Table 7.2).



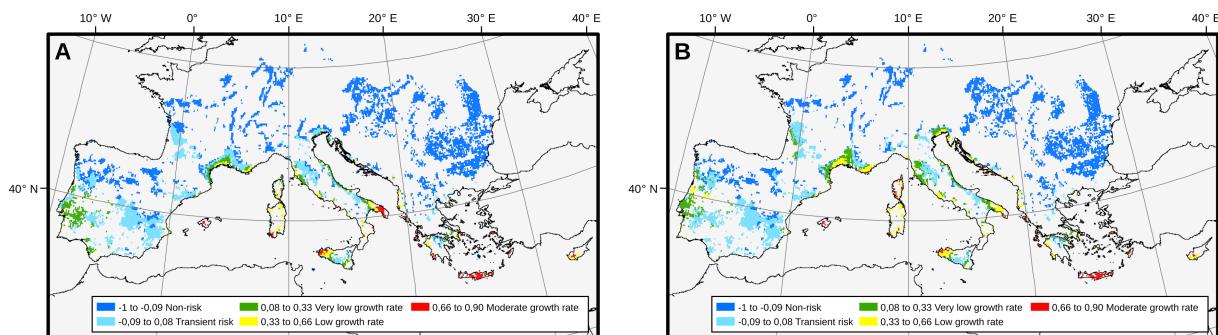
**Figure 7.5: Global shifts in PD risk index ( $r_j(t)$ ) from 2020 to 2050.** To build the maps, we have assumed a spatial homogeneous vector distribution and a  $R_0 = 5$  scenario, except for the US where a  $R_0 = 8$  has been used in the model simulations. (A) North America; (B) Europe; (C) Asia; (D) South America; (E) Australia and New Zealand; and (F) South Africa. Risk-index increases are in red and decreases in blue. The dashed line represents the spatial threshold where  $r_j(t)$  difference changes from negative to positive.

**Table 7.2: Shifts in risk areas for Pierce's disease in Europe projected for 2050 under a  $R_0 = 5$  scenario.** The model was run assuming the same homogeneous spatial distribution of the vector for the whole period.

Risk	2050 km <sup>2</sup>	2020 km <sup>2</sup>	Difference km <sup>2</sup>	Difference (%)	2050 (%) area	2020 (%) area
No risk	8885300.5	9334178.7	-448878.2	-4.8	87.6	92.1
Transition	381081.3	182872.6	198208.7	108.3	3.8	1.8
Very low	189025.3	179225.7	9799.6	5.5	1.9	1.8
Low	207599.4	104143.1	103456.3	99.3	2.1	1.0
Moderate	154780.5	148621.4	6159.0	4.1	1.5	1.5
High	322225.9	190971.4	131254.5	68.7	3.2	1.9

## 7.2.5 Risk based on vector information

. So far, we have ignored the distribution of known and potential vector species due to their large number in the Americas and the limited quantitative information generally available. In the case of Europe, given *P. spumarius* prevalence as a potential vector and its wide distribution, we added a vector layer in a spatially dependent  $R_0(j) = R_0^{\max} v(j)$ , where  $v(j)$  is the climatic suitability for the vector (Methods),  $v = 1$  implies optimal climatic conditions with no constraints for the vector population size, while  $v = 0$  implies unsuitable climatic conditions and its absence (Online Supplementary Information). According to the model, no European zone shows a high-risk index and barely 0.34% of the territory falls in areas with potential moderate exponential growth rates in disease incidence (Online Supplementary Information). Irrespective of vineyard distribution, we estimated that PD could potentially become established (i.e.,  $r(t) > 0$ ) at a maximum of 3.1% of the territory, while the area at moderate-risk index would be 5-times lesser than the model without the vector's climate suitability layer, this latter more in consonance with other proposed risk maps (59, 203). Such differences in the projected risks are mainly concentrated in the warmest and driest Mediterranean regions and are due to uncertainties concerning temperature-humidity interactions in the ecology of the vector (204).



**Figure 7.6: Intersection between Corine-land-cover vineyard distribution map and PD-risk maps for 2020 and 2050.** Data were obtained from Corine-land-cover (2018) and the layer of climatic suitability for *P. spumarius* in Europe from [204]. The surface of the vineyard contour has been enlarged to improve the visualisation of the risk zones and disease-incidence growth-rate ranks. (A) PD risk map for 2019 and its projection for 2050 (B). Blue colours represent non-risk zones and transient risk zones for chronic PD ( $R_0 < 1$ ). The 2050 map shows some contraction of epidemic-risk zones with moderate risk indexes in Mediterranean islands and Apulia (Italy) as the climate becomes hotter and drier.

## 7.2.6 Combining vineyard land cover across Europe with the model output.

When we integrate into the model a layer of vineyard surface from Corine-Land-Cover, we find that PD could potentially become established (i.e.,  $r(t) > 0.075$ ) in 22.3% of the vineyards in Europe. However, no vineyard is in epidemic-risk zones with a high-risk index and only 2.9% of the vineyard surface is at moderate risk (Online Supplementary Information). The areas with the highest risk index ( $r(t)$  between 0.70 and 0.88) are mainly located in the Mediterranean islands of Crete, Cyprus and the Balearic Islands or at pronounced peninsulas like Apulia (Italy) and Peloponnese (Greece) in the continent (Fig. 7.6A and Online Supplementary Information).

Most vineyards are in non-risk zones (42.1%), whereas 35.6% are located in transition zones with presently non-risk but where  $Xf_{PD}$  could become established in the next decades causing some sporadic outbreaks. In Online Supplementary Information, we provide full details of the total vineyard areas currently at risk for each country and region.

Our model with climate and vector distribution projections for 2050 indicates a 55.8% increase in the epidemic-risk zone in Europe (Fig. 7.6B). This increment would be mainly due to the extension of epidemic-risk zones with very low and low-risk indexes. However, within the epidemic-risk zones, areas with moderate risk indexes would decrease from 114 925 ha in 2020 to 43 114 ha in 2050, and no vineyards would be at high risk (Fig. 7.6B; see Online Supplementary Information). Counterintuitively, our model indicates a substantial increase in the area where PD could establish and become endemic for 2050, but a moderate decline in those areas where crop damage could be expected to be significant (e.g., Balearic Islands, Crete, Cyprus, Apulia).

## 7.3 Discussion

We introduce an epidemiological approach to assess the risk of PD establishment and epidemics in vineyards worldwide. The model includes the dynamics of the infected-host population, which enables estimating the initial exponential growth/decrease rate of the disease incidence. Unlike SDM correlative studies, Bayesian or, in general, machine learning black-box approaches, our model goes beyond by providing a mechanistic framework and thus explanatory power. In addition, it is flexible enough to simulate different climate and transmission scenarios, allowing, for instance, the incorporation of information on the spatial distribution of the vector. Comprehensive global PD risk maps result from the model simulations with historical climatic data. A web page is included, showing simulations with different parameters to estimate the risk of PD anywhere (188).

Temperature regulates key physiological processes of the ectothermic organisms involved in PD and thus limits the thermal range in which they can thrive (106).  $Xf_{PD}$  multiplication and survival within vine xylem vessels not only characterise PD, but also determine the bacterial population dynamics (155, 170). PD symptom development can be therefore characterised as a thermal-dependent continuous process within the range of  $Xf_{PD}$ 's cardinal temperatures (387). The combination of MGDD metrics with robust experimental data provides a reliable predictor of climatic suitability and the probability of developing PD during the summer, whereas CDD accounts for the effect of cold-temperature exposure on infected-plant recovery. This opposite contribution of MGDD and CDD in the demography of infected plants shapes the impact of climate variability on the epidemic dynamics in the early stages of the invasion (Fig. 7.1D). Given that the physiological basis of the plant-Xf interaction leading to symptoms development is poorly understood, we caution that other environmental factors, such as drought, nutrient status or crop management may modulate symptom expression and hence add an error in the MGDD parameter not measured in this work. Also, another limitation is that the relation between MGDD and  $Xf_{PD}$  growth rates come from in vitro experiments and we assume that it is valid in planta as well. Nonetheless, we deem the error range would be smaller than the differences in the accumulated MGDDs needed to reach the same disease level among varieties (i.e., regional differences) and smaller than the interannual MGDD oscillations found in most locations.

Knowledge of insect distribution is crucial for predicting epidemic outbreaks of endemic diseases, as well as the risk of invasion by emerging vector-borne pathogens ((78, 241), (cf. (391))). Given the great diversity of known and potential vectors that can transmit PD (363), it has not been possible to include each region's particular vectors in the model. Therefore, when evaluating the risk of PD on a global scale, we have considered a homogeneous spatial distribution of the vector (fixed  $R_0$ ), except in Europe where there is information on the main vector (??). As expected, the European case shows how models that assume a homogeneous spatial distribution of the vector generally produce epidemic risk zones with higher risk indexes than models that include a heterogeneous spatial distribution. This lack of information about vectors is one of the main reasons why the risk of vector-borne plant diseases is often overestimated.

Risk overestimations may involuntarily stem from other additional sources too. Using mean data as inputs in epidemiological models can lead to biased results when response functions are nonlinear and climate variability is not accounted for (387). This study presents experimental evidence of a non-linear relationship between MGDDs and PD chronic infections and indirect

empirical evidence of a non-linear relationship between CDDs and PD recovery (??). Such a non-linear response consequently greatly impacts reducing the risk of PD establishment and steeping the spatial gradients in risk maps (Figs. 7.4 and 7.6). Moreover, MGDDs and CDDs might help to explain why disease pressure is much higher in the southeastern US than in California and Europe (Figs. 7.2 and 7.4) or, for example, earlier reports of PD outbreaks in Kosovo (37). Cooler summer nights in California and a shorter growing season compared to those found in the Gulf states in the southeastern US explain the difference in the accumulated *MGDD* for both areas. In the case of Kosovo, *CDD* values above certain thresholds could have led to the extinction of incipient outbreaks driven by several years with *MGDD* in the conducive range of PD (Fig. 7.2).

Our PD risk map for Europe confirms previous predictions for the subsp. *fastidiosa* from SDMs (59). Both approaches make congruent predictions on PD potential distribution, providing convergent lines of independent evidence of climate suitability. However, our risk maps go further by incorporating in the epidemic-risk zones information on the relative exponential growth rates in the potential disease incidence. In general terms, the epidemic-risk map including vector information indicates low risk for chronic PD. Only ~ 0.34% of European vineyard surface, mainly located in Cyprus, Crete, Sardinia, part of Sicily and the Balearic Islands, meet climatic conditions for PD to become endemic and cause significant damage (Online Supplementary Information). Other regions such as Bordeaux, Portugal, Rhône Valley, and the Veneto region, would be included in epidemic-risk zones but with very low to low exponential growth rates in disease incidence. By contrast, notorious wine-growing regions in Spain (e.g., Rioja, Ribera del Duero), France (e.g., Burgundy) and Italy (e.g., Piedmont) currently fall within areas considered as non-risk zones, transient-epidemic zones or epidemic-risk zones with very-low risk indexes (Fig. 7.6).

The dynamic nature of the simulation outputs already points to a progressive global increase in the areal extension of PD epidemic-risk zones ( $r(t) > 0$ ) in the last decade, irrespective of vineyard distribution (see movies on (188)). This is even more accentuated in the model projections for 2050, which point out a global expansion of PD epidemic-risk zones at different velocities among continents due to climate change (Fig. 7.5). For example, many important viticulture areas in western Europe included in non-risk or transition zones before 1990 are progressively shifting to hotter summers and milder winters and hence would be increasingly suitable for the disease within the extrapolated current scenario. This is further illustrated by a 40% increase of the potential epidemic-risk zone by 2050 concerning 2020 for Europe and more moderate increases in the United States and the Southern Hemisphere (Fig. 7.5). Nonetheless, our model projection for 2050 that includes spatial heterogeneity in the vector distribution, as in Europe, would indicate lower transmissibility because global change is predicted to have negative effects on *P. spumarius* abundance in Europe (204, 252). At the global scale, there is certainly scientific consensus that climate change will follow a general pattern summarised in the paradigm "dry gets drier, wet gets wetter" (157). In agreement, our model projection for PD on vineyards of Majorca (Spain) suggests shifts to slightly less favourable conditions for  $Xf_{PD}$  transmission and an expected progressive decrease in the impact of the disease by 2050. This example and others in Mediterranean islands (see Supplementary Data 4) advocate for certain caution when interpreting climate change projections, especially in other Mediterranean climates of the world, where the complex interactions between humidity and temperature can limit the presence and abundance of vectors (??).

The scope of our study excludes location-specific complexities surrounding PD ecology due to scale limitations. The spatial distribution of the vector is considered only for the *V. vinifera*- $Xf_{PD}$ -*P. spumarius* pathosystem in Europe, so  $R_0$  estimations could locally differ in other wine-producing regions elsewhere (Fig. 7.3). Disease incidence thus could locally vary where the climate is conducive to PD. Such variation is because transmission rates tend to increase exponentially rather than linearly under environmental conditions favouring vector abundance (215), as has been observed at a local scale on vineyards of Majorca (313). Our study also does not contemplate likely changes within the PD pathosystem. To date, PD is caused by  $Xf_{PD}$  (i.e., ST1/ST2), but other genotypes of the subsp. *fastidiosa* or other subspecies and their recombinations could arise in the future with different ecological and virulence traits (430). On the other hand, new vector species could be accidentally brought in (363), as exemplified with the introduction of the glassy-winged sharpshooter (*Homalodisca vitripennis*) in California, modifying transmission rates and disease incidence in new areas (128). To capture these

uncertainties in relation to the vector, we have performed simulations with  $R_0 = 8$  and  $R_0 = 16$  (Fig. 7.4). Remarkably, a comparison of PD risk maps for Europe with different  $R_0$  suggests for non-Mediterranean areas the need to stress more surveillance on the introduction of alien vectors rather than in the pathogen itself. This is because, under the current scenario ( $R_0 = 5$ ) with *P. spumarius* as the main vector, most of the non-Mediterranean vineyards would not support the establishment of PD, but the introduction of new insect vectors with greater transmission efficiency ( $R_0 = 8$ ) could compensate the climatic layer and increase the risk index above 0. In addition, differences in grapevine varietal response alongside virulence variation among Xf strains may slightly modify PD thermal tolerance limits and therefore locally modulate epidemic intensity (see details in Online Supplementary Information). Such effect could be seen with cv. Tempranillo, a widely planted variety in northern Spain (Online Supplementary Information); the rate of symptom progress and systemic movement is higher than the average varietal response to  $Xf_{PD}$  (i.e., lower MGDD), which in addition might imply higher survival rates. This point calls for further testing in the field.

Our model partially explains why PD has not become established in continental Europe and other main wine-growing regions worldwide during the last 150 years, in contrast to other exotic diseases and pests brought in with native vines from the US (51, 65, 373, 415). We suggest that the underlying causes of this low-invasiveness risk in Europe are fundamentally two: (i) low climatic suitability for chronic PD and (ii) a climatic mismatch between environment conditions suitable for both the vector and the pathogen and their interplay in disease dynamics, similar to the situation recently described for the *V. vinifera*- $Xf_{PD}$ -*P. spumarius* pathosystem in northern California (30). Currently, suitable conditions for the pathogen's invasion mostly concur in Mediterranean islands and coastlands (Supplementary Data 4). Likewise, similar results would be expected in other Mediterranean climates of main winegrowing regions of the Southern Hemisphere if a vector spatial distribution layer is incorporated in the model simulations (see (188)). Finally, although increasing global warming will extent epidemic-risk zones in all continents, some caution is recommended to not incur risk overestimation, as we show in the PD risk projections for 2050 in Europe when taking into account the vector spatial distribution; complex interactions between temperature and humidity in the ecology of the vectors may have a great effect in their distribution, abundance and thus transmission capacity (204). There is an urgent need to fill the knowledge gap on the ecophysiology for each potential vector to downscale PD model predictions to local and regional situations.

## 7.4 Methods

### 7.4.1 Inoculation tests.

$Xf_{PD}$ -inoculation tests were conducted in 2018, 2019 and 2020. A sample of 36 local, regional and international wine-grape varieties was selected, which included nine of the 10 most cultivated wine-grape varieties representing more than 80% of the worldwide vineyard surface (<https://www.oiv.int>). Plants were randomly distributed in 12-plant rows along an insect-proof net tunnel and exposed to environmental temperature. In total, 57 rootstock-scion combinations were pin-prick mechanically inoculated (9) with two strains of *Xf*, subsp. *fastidiosa* (ST1) isolated from grapevines in Majorca. Disease severity was rated by counting the number of symptomatic leaves eight weeks after inoculation in mid-May and then every two weeks until the 16th week (313). Full details on the inoculation conditions, isolates, disease score and statistical analysis are provided in Supplementary Information.

### 7.4.2 Modified Growing Degree Days.

We generalized McMaster & Wilhelm's (302) formulation of growing-degree days to account for the growth rate of  $Xf_{PD}$  as a function of temperature under optimal culture conditions based on the well-known Arrhenius law valid in the relevant temperature range for Xf (Online Supplementary Information). Specific growth rate ( $k$ ) values at different temperatures were extracted from the publication of Feil & Purcell (155) to build the mathematical function  $f(T)$

describing the *Xf*'s instantaneous growth rate dependence on temperature according to

$$f(T) = \begin{cases} 0 & \text{if } T < T_{\text{base}} \\ m_1 \cdot T - b_1 & \text{if } T_{\text{base}} \leq T < T_1 \\ m_2 \cdot T + b_2 & \text{if } T_1 \leq T < T_{\text{opt}} \\ m_3 + b_3 & \text{if } T_{\text{opt}} \leq T < T_2 \\ m_4 + b_4 & \text{if } T_2 \leq T \leq T_{\text{max}} \\ 0 & \text{if } T \geq T_{\text{max}} \end{cases}$$

where  $T_{\text{base}} = 12^\circ\text{C}$ ,  $T_1 = 18$ ,  $T_{\text{opt}} = 28^\circ\text{C}$ ,  $T_2 = 32$  and  $T_{\text{max}} = 35^\circ\text{C}$ ; the slopes are  $m_1 = 0.66$ ,  $m_2 = 1$ ,  $m_3 = -1.25$  and  $m_4 = -3$  and the intercepts are  $b_1 = -8$ ,  $b_2 = -14$ ,  $b_3 = 4$  and  $b_4 = 105$ .

MGDD is then defined as:

$$MGDD(t) = \frac{1}{24} \sum_{\tau \in t} f(T(\tau)), \quad (7.1)$$

where  $\tau$  is expressed in hours,  $t$  in years and we divide by 24 to obtain  $MGDD(t)$  in degree days.

#### 7.4.3 Disease progress with temperature.

Hourly mean temperature data were recorded between April 1 and October 31 in 2018, 2019 and 2020 with an automated weather station (Quimisur, IQ2000). The temperature sensor was at a two-meter height from the bare ground and around five meters from the entrance of the insect-proof net tunnel. To characterise the progress of PD symptoms, we converted into *MGDD* units the cumulative hourly mean temperatures measured from the time of inoculation to the day of disease evaluation using Eq. (8.1). In total, 15 *MGDD* levels were estimated corresponding to weeks 8, 10, 12, 14, and 16 after inoculation in the years 2018, 2019, and 2020, respectively. Data on the number of symptomatic leaves (severity) for each plant and *MGDD* levels were pooled in a single database to seek a generalized average thermal response pattern among the population of *V. vinifera* varieties (see Supplementary Data 1). To model the probability of chronic infections (i.e., persistent year-to-year infections), we used a survival analysis, where the event of interest depends on the cumulative *MGDD* rather than time. First, we defined a chronic infection cut-off point to transform the number of symptomatic leaves into binary data. Previous research had evidenced that early grapevine infections, in addition to producing more extensive and severe PD symptoms, are more likely to survive the following year than late infections ((154, 155, 283)). Furthermore, susceptible cultivars generally show lower recovery percentages compared to the less susceptible ones in the field (359, 360). Similarly, we observed in our inoculation assays that the majority of infections that reach around five or more symptomatic leaves 12 weeks after inoculation continue to develop more symptomatic leaves the following weeks, while for plants that do not exceed that threshold, symptoms tend to remain stagnant. These results indicate a low probability of survival for infections showing few symptomatic leaves during the growing season and thus support our heuristic approach of assigning five or more symptomatic leaves as a threshold for chronic infection (see Supplementary Information and ?? for assumptions of chronic infection). Using the "survival" package in R (417), we analysed the cumulative probability of developing chronic infections as a function of *MGDD*.  $F(MGDD)$  was adjusted to the experimental data by the nonlinear least squares method. The 10th, 33rd, 50th, 66th, and 90th percentiles were used to scale the risk of the total *MGDD* in the logistic function,  $\mathcal{F}(MGDD)$  (Fig. 7.1C).

#### 7.4.4 Disease recovery through winter curing.

We modelled winter curing considering the effect of temperature duration below a threshold temperature, where we assume that the bacterial killing process increases in efficiency with decreasing temperatures (283). To adjust a probabilistic model to the accumulation of cold units, we took as reference the distribution and severity of PD in the US proposed by Purcell based on the isolines of the mean  $T_{\min}$  of the coldest month (available in (17)) where PD is rare ( $T_{\min}$  between  $-1.1^\circ\text{C}$  and  $1.7^\circ\text{C}$ ), occasional ( $1.7 - 4.5^\circ\text{C}$ ) and severe ( $> 4.5^\circ\text{C}$ ). Noteworthy, the projection of these isolines in Europe has predicted with some precision the distribution of the establishment of *Xf* in the continent (59). To capture the accumulation nature of the chilling process at different climatic zones, we determined the global average correlation between

$T_{\min}$  and the average accumulated CDD between November 1 and March 31 in the northern hemisphere and between April 1 and October 31 in the southern hemisphere using 6,487,200 points distributed throughout the planet. The *CDD* was estimated as

$$CDD(t) = \frac{1}{24} \sum_{\tau \in t} (6 - T(\tau)) \quad \text{for } T_i \leq 6^{\circ}\text{C}, \quad (7.2)$$

where the threshold  $6^{\circ}\text{C}$  comes from Ref. (283).

#### 7.4.5 Global climate data, MGDD/CDD computation.

Global mean hourly temperature data were downloaded from the ERA5-Land dataset (323) at  $0.1^{\circ}$  spatial resolution using GRIB format. The annual average  $T_{\min}$  of the coldest month was calculated from the hourly average temperature from the ERA5-Land dataset. To calculate the annual *MGDD* and *CDD* a simple Julia (40) library was built on top of GRIB.jl package (21). For the Northern Hemisphere, the accumulated *MGDDs* were computed from April 1 to October 31, whereas (*CDDs*) were estimated from November 1 to March 31, and the reverse for the Southern Hemisphere.

#### 7.4.6 Disease Model.

We used a standard susceptible-infectious/infected-recovered (SIR) compartmental model to assess the risk of PD establishment and epidemics worldwide, represented by the following three equations in the large population limit:

$$\begin{aligned} \dot{S} &= -\beta SI/N, \\ \dot{I} &= \beta SI/N - \gamma I, \\ \dot{R} &= \gamma I, \end{aligned} \quad (7.3)$$

where  $S$  is the susceptible host population,  $I$  is the infected population,  $R$  is the dead population and  $S + I + R = N$  is the total number of vines in the population. The reduction of a vector-borne disease model to a SIR model gives rise to a linear dependence of the basic reproductive number  $R_0$  on the vector population (see Online Supplementary Information). Vector-plant transmission of the pathogen is approximated with an effective plant-to-plant transmission rate  $\beta$  (Online Supplementary Information), as has been done previously for other Xf-related diseases (449), and the transition from the infected compartment to the recovered (dead) compartment is given by the recovery (mortality) rate  $\gamma$ . In a mean-field approximation of the onset of an outbreak, the basic reproductive number ( $R_0 = \beta/\gamma$ ) defines the exponential growth/decrease stage in the SIR model (Fig. 7.1E, Online Supplementary Information). Although the time from infection to vine death depends on the environmental conditions and the grape wine variety, we assigned a mortality rate of  $\gamma = 0.2 \text{ y}^{-1}$  based on the estimated median survival time of infected vines in California (9). The maximum growth rate of the epidemic, relevant for an estimation of the risk of establishment, occurs when  $S(t = 0) \sim N$ , and was approximated by the (linearized) differential equation,

$$dI/dt \approx \beta I - \gamma I = \gamma I (\beta/\gamma - 1) = \gamma I (R_0 - 1), \quad (7.4)$$

where we have assumed the initial conditions:  $S(t = 0) \approx N$ ,  $I(t = 0) = I(0) \approx 0$  and  $R(t = 0) = 0$ . This linear differential equation can be integrated exactly:

$$I(t) = I(0) \exp(\gamma(R_0 - 1)t). \quad (7.5)$$

To account for the effect of temperature in the epidemic process, we modify the previous expression as follows

$$\begin{aligned} I(t) &= I(0) \exp(\gamma(R_0 - 1)t) \mathcal{F}(MGDD(t)) \mathcal{G}(CDD(t)) \\ &= I(0) \exp(\gamma(R_0 - 1)t) \Pi(t), \end{aligned} \quad (7.6)$$

where  $\Pi(t) = \mathcal{F}(MGDD(t)) \mathcal{G}(CDD(t))$  is the cumulative probability of chronic infection dependence on temperature and  $R_0$  bears the information on the vector density.

The spatial unit of the model is given by the resolution of the ERA5-Land data, for which we assume uniform conditions within each of the grid cells (approximately  $9 \times 9 \text{ km}^2$ ) in terms of vector population, susceptible vines and parameters that define the model. Risk outcome is calculated for each cell of the spatial raster individually; i.e., there is no simulated spread from one cell to another. Altogether, the equation representing the number of individuals with chronic infections in each cell  $j$  at time  $t$  is written as

$$I_j(t) = \underbrace{I_j(t-1) e^{\gamma(R_0(j)-1)}}_{\text{transmission layer}} \underbrace{\Pi_j(t)}_{\text{climatic layer}}, \quad (7.7)$$

where  $I(t-1)$  is the number of chronic infections in the previous year ( $t-1$ ) and  $\Pi_j(t) = \mathcal{F}(MGDD)\mathcal{G}(CDD)$  is the climatic layer that modulates the growth term and describes the cumulative probability of new infections becoming chronic in the time period between  $t-1$  and  $t$ . The model assumes a homogeneous distribution of the vector population among the grid cells (same  $\beta$  and then same  $R_0(j) = R_0$ ) except for Europe, where information on the spatial distribution of *P. spumarius* is available (see Methods). In this latter case, a spatial dependent  $R_0(j)$  is incorporated into the model by considering the product of the homogeneous  $R_0$  and the spatially-dependent climate suitability for vectors (Online Supplementary Information).

To compute the epidemic-risk maps, we carried out a simple simulation summarized in three steps: (i) at the initial condition for the first year considered,  $t_0$ , each grid cell is seeded with a single infected plant,  $I(t_0) = 1$ ; (ii) the simulation runs for a year and the incidence is calculated following Eq. (7.7); (iii) we seed again the cells for which the number of infected plants has vanished. In the last seven years of the simulation, there is no reseeding to allow the system to relax. This process is repeated until the final year  $\mathcal{T}$ . Finally, the risk index  $r_j(\mathcal{T})$  is calculated from the final number of infected plants at grid cell  $j$  as

$$r_j = \max \left\{ \frac{\log(I_j(\mathcal{T})/I_j(t_0))}{\gamma(R_0(j)-1)\mathcal{T}}, -1 \right\}. \quad (7.8)$$

In this equation,  $r_j$  implicitly delimits three differential risk zones in the maps: 1) a non-risk zones where  $r_j \leq -0.09$ , and the number of infected plant decreases exponentially; 2) a transition areas where  $-0.09 < r_j \leq 0.075$ , and 3) an epidemic risk-zone where  $r_j > 0.075$  and PD can theoretically become established and produce an outbreak –the number of infected plants increases exponentially (see Online Supplementary Information for further details.)

Model performance was calibrated with observed records of PD presence in California and the southeast of the US, where the disease is well established. PD distribution data were collected from publications from 2001 to 2020. Publications were filtered by selecting only records where the pathogen detection on symptomatic grapevines was confirmed by PCR or Elisa. The exact coordinates of the records were taken when available in the publication or approximated to locality or county level to build the Supplementary Data 1 (5, 17, 217, 283, 328, 340, 430, 440). For modelling purposes and to attempt a general rough estimate of the  $R_0$  parameter valid for the entire US, we assumed a single vector with a uniform spatial distribution. We ran several model simulations with  $R_0$  ranging from 1 to 14. Model prediction performance was estimated using a ROC curve by plotting the true-positive rate (TPR), calculated as the ratio (TP/TP+FN), against the false-positive rate (FPR), calculated as the ratio (FP/TN+FP), where PD absence/presence fulfil the following conditions: true positive (TP), PD is positive and  $r > 0$ ; true negative (TN), PD is negative and  $r < 0$ ; false positive (FP), PD absent but  $r > 0$ ; and false negative (FN), PD positive and  $r < 0$  (246). A different approach was followed to estimate  $R_0$  for Europe given that PD is only present in Majorca and hence spatiotemporal data on the PD distribution is limited to the island. First, we estimated the transmission rate of the main European vector *P. spumarius* from the well-studied disease progress curve of the almond leaf scorch epidemic in Majorca. Then, using the known mortality rate of PD-infected vines  $\gamma \sim 0.2 \text{ y}^{-1}$  and the inferred transmission rate,  $\beta = 0.8 \text{ y}^{-1}$ , the basic reproduction number for PD in Majorca yields  $R_0 = \beta/\gamma \approx 4$ . Finally, using data on the climate suitability of the vector in Majorca,  $v = 0.8$ , and inverting the relation  $R_0(j) = R_0 v(j)$ , we estimated  $R_0 \approx 4/0.8 = 5$  as a maximal estimate baseline scenario for PD transmission in Europe (Online Supplementary Information). This figure is not intended to be an exact estimate of  $R_0$  but rather an average reference in our model in agreement with the lesser abundance of vectors

relative to the US. Furthermore, since there is no information on the distribution of the potential vectors and no PD distribution data to calibrate, we also used a conservative  $R_0 \approx 5$  scenario for the rest of the world.

#### 7.4.7 Distribution of wine-grape production areas.

Risk maps were focused solely on wine-grape regions excluding table and dried grapes producing areas. Data on the vineyard surface in Europe were obtained from the CORINE land-cover map (153) (Fig. 7.6). Nomenclature of Territorial Units for Statistics (NUTS) was used as a geocoding for the subdivisions of European countries for statistical purposes. To visualize the locations of the main growing regions in the risk maps, we included dots representing the distribution of the main wine-growing regions collected from official statistics and maps from the countries (Fig. 7.5).

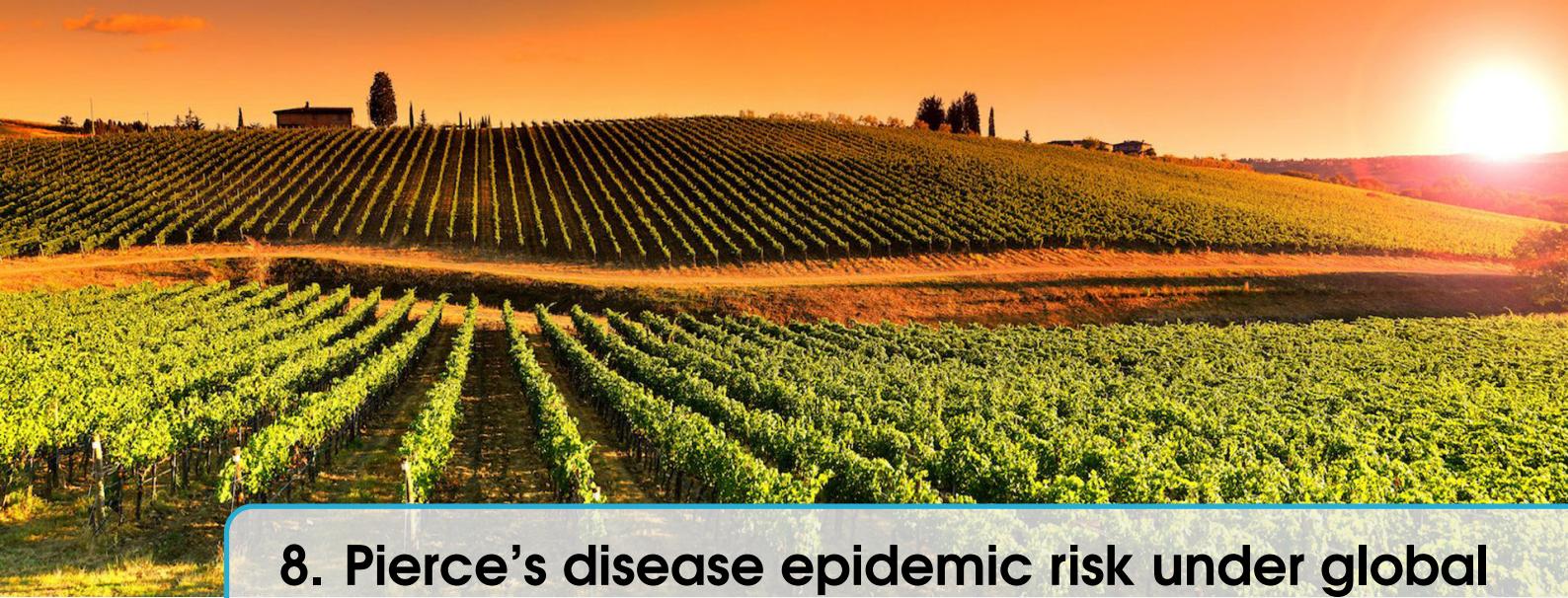
#### 7.4.8 *Philaenus spumarius* SDM.

The potential distribution of *P. spumarius* in Europe under current and future (i.e., 2050) climatic conditions was provided by Godefroid et al. (204). Predictions were obtained using a generalised additive model and two bioclimatic descriptors i.e., a climatic moisture index for the coldest 8-month period of the year and the average maximum temperature in spring (March, April and May). Both descriptors reflect physiological constraints acting on life stages of the meadow spittlebug, particularly sensitive to spring temperature and humidity (eggs and nymphs), and were identified as good predictors of *P. spumarius* distribution ((204)). We used the positive relationship between the climate suitability and spittlebug adult abundance ((204)) to assume no climatic constraints on vector population sizes at optimal climatic conditions ( $v=1$ ). Climatic suitability indexes,  $v(x)$ , were used to compute a spatially-dependent basic reproduction number,  $R_0(x) = R_0 v(x)$ . The linear dependence between the basic reproduction number and climatic suitability is justified by a vector-borne epidemic compartmental model (Online Supplementary Information).

#### 7.4.9 Risk assessment by 2050.

Climatic variables were obtained with annual resolution by extrapolating the computed  $MGDD(t)$  and  $CDD(t)$  time series up to 2050. The observed trends of the time series were captured using a machine learning-based linear regression model while the interannual fluctuations were modelled by Gaussian noise (Online Supplementary Information). Future risk extrapolations were obtained as the average of  $10^4$  simulations of this process. A correlative SDM was used to estimate vector spatial distribution in Europe using the global circulation model MIROC5 and greenhouse gas emission scenario RCP4.5, assuming moderate climate change (204). Afterwards, the risk was computed following the same simulation procedure previously explained.





## 8. Pierce's disease epidemic risk under global warming

Àlex Giménez-Romero<sup>1</sup>, Maialen Iturbide<sup>2</sup>, Eduardo Moralejo<sup>3</sup>, Jose M. Gutiérrez<sup>2</sup>, Manuel A. Matías<sup>1</sup>

1. Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Palma de Mallorca 07122, Spain
2. Instituto de Física de Cantábrria (CSIC-University of Cantabria), Avenida de los Castros, 39005 Santander, Spain
3. Tragsa, Passatge Cala Figuera 6, 07009 Palma de Mallorca, Spain

### Published as

Àlex Giménez-Romero et al. "Contrasting Patterns of Pierce's Disease Risk in European Vineyards Under Global Warming". In: *bioRxiv* (2023). DOI: [10.1101/2023.07.17.549293](https://doi.org/10.1101/2023.07.17.549293). eprint: [https://www.biorxiv.org/content/early/2023/07/18/2023.07.17.549293](https://www.biorxiv.org/content/early/2023/07/18/2023.07.17.549293.full.pdf). URL: <https://www.biorxiv.org/content/early/2023/07/18/2023.07.17.549293>

## 8.1 Introduction

Climate change is widely recognized as an important driver of shifts in the distribution and prevalence of plant diseases worldwide (94, 132, 147, 220, 370, 401). Although the impact of climate change on the distribution of plant diseases has been approached from various perspectives (36, 344), few studies have considered epidemiological dynamics in climate projections (33, 248). Modeling disease epidemics is a complex task, as they are emergent phenomena resulting from non-linear interactions between disease components. In addition, many of the processes involved in disease development also exhibit non-linear responses to changes in environmental variables (177, 386). This complexity is further exacerbated in the case of vector-borne plant diseases (245). While climate primarily determines the potential geographic range of each organism in the pathosystem, the development of epidemic outbreaks depends on favorable host-pathogen-vector-climate interactions that drive transmission chains. Consequently, modeling the risk of vector-borne plant diseases implies delimiting their epidemiological niche rather than the ecological niche of their parts, as is commonly done.

The emergence of *X. fastidiosa* in Europe has renewed interest in modeling vector-borne disease, particularly the risk they pose to the European wine industry. Pierce's disease (PD) is an endemic fatal disease of grapevines in the Americas transmitted unspecifically by sap-feeding insect vectors belonging to sharpshooter leafhoppers (Hemiptera: Cicadellinae) and spittlebugs (Hemiptera: superfamily Cercopidae) (364). In the US, PD causes huge economic losses to the wine sector estimated at 100 M\$ per year in California alone (424). The causative agent of PD is *Xylella fastidiosa* (Xf), a bacterium capable of colonizing the xylem vessels of more than 600 hosts, including important crops (131). As a taxonomic unit, Xf comprises three recognized subspecies, *fastidiosa*, *multiplex*, and *pauca*, and more than 90 sequence types (i.e. genetic lineages) with distinct host ranges. Specifically, the Xf clonal lineage that causes Pierce's disease (hereafter Xf<sub>PD</sub>) also causes almond leaf scorch in California (11).

Until the beginning of the 21st century, Xf was a pathogen officially restricted to the American continent (12). In 2013, the involvement of Xf subsp. *pauca* in the massive death of ancient olive trees in Apulia, Italy, and its rapid spread raised alarm in European agriculture (378). Today, all three Xf subspecies have been detected in the Balearic Islands (Spain), including Xf<sub>PD</sub>, and several clonal lineages have been found in Corsica and the PACA region of France, Alicante (Spain), Tuscany (Italy) and Portugal (134, 294, 338). Outside North America, Xf<sub>PD</sub> is only established on the islands of Mallorca and Taiwan, and has recently been detected in Israel, Lebanon and Portugal (88, 457). In all European outbreaks, the insect vector *Philaenus spumarius* is the main and almost unique carrier of Xf (109).

Despite recent studies agree that the current risk of PD establishment in Europe is primarily confined to the Mediterranean basin (194, 205, 206), its potential future progression is not yet clear. Some efforts have been made to characterize the geographical distribution of Xf-induced diseases in Europe under climate change, but these are limited to the use of species distribution models (SDMs) for the pathogen (55, 207, 391) and the vector (205), which have led to conflicting results. On the one hand, higher temperatures are expected to promote bacterial growth in susceptible crops in continental southern Europe, while on the other hand, these areas are progressively experiencing drier environmental conditions detrimental to vector populations (205). Furthermore, the use of SDMs to predict the potential distribution of vector-borne plant diseases, while capable of providing good approximations, is generally inadequate. The observed distribution of the pathogen cannot be separated from that of the vector, especially in the case of obligate pathogens such as Xf. Furthermore, the potential distribution of the pathogen or the vector alone have no epidemiological meaning, what implies that they do not provide quantitative predictions of the severity of the disease. In addition, a larger set of available climate models is desirable to properly deal with the inherent uncertainty in the predictions,

To overcome these limitations, here we use a novel climate-driven epidemiological model of PD epidemic risk (194). The model determines the spatio-temporal epidemic risk based on the spatial distribution of vectors, temperature-dependent bacterial growth and survival within hosts, and subsequent epidemiological dynamics. The model forces the introduction of the pathogen and examines whether the disease can establish and spread from previous states under climatic conditions of the location. In (194) the model was used to determine the

risk of PD under current climatic conditions in wine-growing regions worldwide, while a linear regression was used to obtain a crude first estimate of the risk in the future. This simple estimation is not expected to be reliable, because it overlooks the role of nonlinearities in the model and also does not take into account climate change scenarios (as described in the manuscript). To assess the potential distribution and relative impact of PD on European vineyards under different levels of global warming, here we used state-of-the-art regional climate projections from the EURO-CORDEX initiative (240). Our study takes into account uncertainties in climate projections and provides an updated and comprehensive assessment of PD risk in European wine-growing regions, addressing previous limitations. We posit that pest risk maps constructed from projections of epidemiological models driven by climate data provide a more realistic, quantitative and explanatory predictions than correlative and probabilistic models. Additionally, they offer valuable insights for anticipating and managing the potential impacts of PD and thus ensuring the resilience of viticulture despite future climate challenges.

## 8.2 Methods

### 8.2.1 Climate datasets

We used E-OBS version v21e (116) as the reference observational climatic dataset, providing daily gridded data for Europe at a resolution of 0.1 degrees ( $\sim 10$  km). Maximum and minimum temperature data was used to compute the MGDD and CDD indices involved in the growth and survival processes of the  $X_{PD}$  pathogen (see "Climate-driven epidemiological model" section below). To calibrate the distribution models of *P. spumarius* capturing the widest possible range (North America and Europe), we used the ERA5-Land reanalysis (324) due to its global (land) coverage and high resolution (0.1 degrees, as E-OBS). Daily precipitation and daily minimum and maximum temperature data were retrieved to calculate the moisture index and maximum temperature during spring index required for the vector suitability model (see "Vector suitability" section below). Historical and future projections of both indexes were calculated using regional climate simulations from the state-of-the-art large high-resolution (0.11 degrees) ensemble provided by EURO-CORDEX (202). This dataset includes daily simulations of precipitation and temperatures from a large ensemble of Regional Climate Models (RCMs) driven by Global Climate Models (GCMs) from the CMIP5 project (414). For this, we considered the RCP8.5 simulations for 40 combinations of GCMs-RCMs (Table 8.1). In order to calculate 20-year mean climatic indexes across the different global warming levels ( $+1.5^\circ\text{C}$ ,  $+2^\circ\text{C}$ ,  $+3^\circ\text{C}$  and  $+4^\circ\text{C}$ ), we relied on the time periods during which each CMIP5 driving model reaches the designated level within the RCP8.5 scenario (see (140)). This information is available at the IPCC WGI Atlas GitHub repository (238).

**Table 8.1: EURO-CORDEX GCM-RCM combinations used in this study. Numbers indicate the number of runs in each combination.**

	CNRM-CM5	EC-EARTH	HadGEM2-ES	IPSL-CM5A-MR	MPI-ESM-LR	NorESM1-M
CLMcom-CCLM4-8-17_v1	1	1	1		1	
DMI-HIRHAM5_v2	1	1	1			
GERICS-REMO2015_v2	1					
IPSL-WRF381P_v2	1					
KNMI-RACMO22E_v2	1		1			
SMHI-RCA4_v1	1	2	1	1		1
CLMcom-ETH-COSMO-crCLIM-v1-1_v1	2		1		1	1
DMI-HIRHAM5_v1		1		1	1	
IPSL-WRF381P_v1		1	1	1		1
KNMI-RACMO22E_v1		2		1	1	1
MOHC-HadREM3-GA7-05_v1		1	1		1	1
GERICS-REMO2015_v1				1		1
MPI-CSC-REMO2009_v1					1	
SMHI-RCA4_v1a					1	
DMI-HIRHAM5_v3						1

### 8.2.2 Climate-driven epidemiological model

We used the model developed in (194), which describes the initial exponential rise (or decrease) of infected plants at the onset of an epidemic based on two main features: the

spatial distribution of the vector and the bacterial growth and survival processes mediated by temperature. Temperature-dependent PD symptom development were based on parametrization of the results obtained after inoculation 36 grapevine varieties (194). In short, the density of vectors at a given site influences the number of new plants that will be inoculated with the bacterium, while the local temperature mediates the growth and survival processes of the in-plant bacterial population, leading to the initial inoculation to an infection or not. These temperature-driven growth and survival processes are described with the **accumulation** of two metrics denoted *Modified Growing Degree Days* (MGDD) and *Cold Degree Days* (CDD). The base function to compute the MGDD is proportional to the  $X_f$  temperature-dependent growth rate and is defined by,

$$f(T) = \begin{cases} 0 & \text{if } T < T_{\text{base}} \\ m_1 \cdot T - b_1 & \text{if } T_{\text{base}} \leq T < T_1 \\ m_2 \cdot T + b_2 & \text{if } T_1 \leq T < T_{\text{opt}} \\ m_3 + b_3 & \text{if } T_{\text{opt}} \leq T < T_2 \\ m_4 + b_4 & \text{if } T_2 \leq T \leq T_{\text{max}} \\ 0 & \text{if } T \geq T_{\text{max}} \end{cases}$$

where  $T_{\text{base}} = 12^\circ\text{C}$ ,  $T_1 = 18$ ,  $T_{\text{opt}} = 28^\circ\text{C}$ ,  $T_2 = 32$  and  $T_{\text{max}} = 35^\circ\text{C}$ ; the slopes are  $m_1 = 0.66$ ,  $m_2 = 1$ ,  $m_3 = -1.25$  and  $m_4 = -3$  and the intercepts are  $b_1 = -8$ ,  $b_2 = -14$ ,  $b_3 = 4$  and  $b_4 = 105$ . MGDD are then computed between 1<sup>st</sup> April and 31<sup>st</sup> October as

$$\text{MGDD}(t) = \frac{1}{24} \sum_{\tau \in t} f(T(\tau)), \quad (8.1)$$

where  $\tau$  is expressed in hours,  $t$  in years and we divide by 24 to obtain  $\text{MGDD}(t)$  in degree days. CDDs are computed between 1<sup>st</sup> November and 31<sup>st</sup> March as

$$\text{CDD}(t) = \frac{1}{24} \sum_{\tau \in t} (6 - T(\tau)) \quad \text{for } T_i \leq 6^\circ\text{C}. \quad (8.2)$$

Altogether, the number of infected hosts is described by the following recurrence relation

$$I(t) = I(t-1) e^{\gamma(R_0-1)} \mathcal{F}(\text{MGDD}(t)) \mathcal{G}(\text{CDD}(t)), \quad (8.3)$$

where  $\gamma$  is the death rate of infected vines,  $R_0$  is the basic reproduction number of the disease and  $\mathcal{F}(\cdot)$  and  $\mathcal{G}(\cdot)$  are sigmoidal-like functions that relate the MGDD and CDD metrics to the probability of developing an infection from a given inoculation. Following (194),  $R_0$  in each cell  $j$  is related to the climatic suitability of the vector such that

$$R_0^j = R_0^* \cdot s_j = 5 \cdot s_j, \quad (8.4)$$

$\gamma = 0.2$  and the specific form of  $\mathcal{F}(\cdot)$  and  $\mathcal{G}(\cdot)$  is given by

$$\mathcal{F}(x) = \frac{1}{1 + e^{-0.012(x-975)}} \quad (8.5)$$

$$\mathcal{G}(x) = \frac{2 \cdot 10^7}{2 \cdot 10^7 + x^3} \quad (8.6)$$

Finally, the risk index is derived as the effective growth rate of the infected population over the simulated time (194),

$$r_j = \max \left\{ \frac{\ln(I_j(T)/I(0))}{\gamma(R_0^j - 1) \cdot T}, -1 \right\}. \quad (8.7)$$

Because the typical time scale of the disease is 5 years ( $1/\gamma$ ), we simulate periods of 7 years. If more years are available to simulate, we perform a re-introduction of the disease as a single infected plant in each cell after each 7-year period (194).

### 8.2.3 Model adaptation to daily temperature data

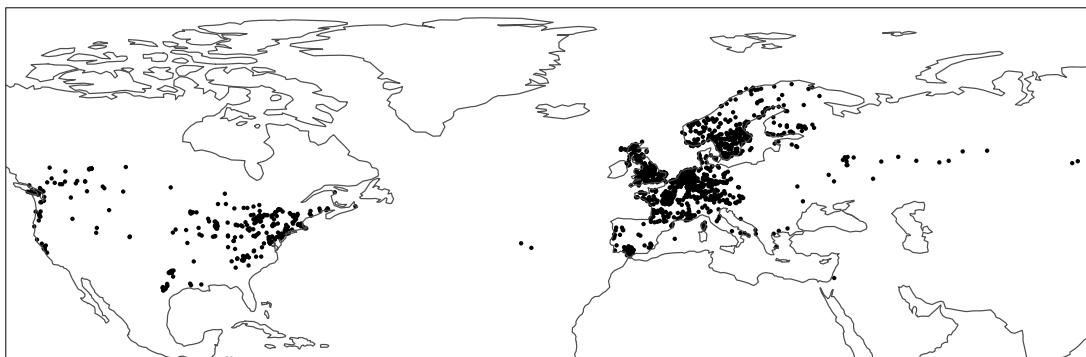
MGDD and CDD metrics were defined using hourly temperature data (194). However, the E-OBS and CORDEX datasets only provide daily granularity. To overcome this limitation, we use a basic sinusoidal extrapolation relating maximum and minimum daily temperature to hourly temperatures,

$$T_h = \frac{T_{max} + T_{min}}{2} + \frac{T_{max} - T_{min}}{2} \sin(w \cdot h), \quad (8.8)$$

with  $w = 2\pi/24$  and  $h$  ranging from 0 to 23. This approximation was validated with data from the national meteorological agency in Spain (AEMET). Basically, we used hourly temperature data obtained from 50 meteorological stations in the period 2010-2020 and computed both MGDD and CDD using the full hourly data and only the daily maximum and minimum temperatures, in the latter case using Eq. (8.8). The results showed no differences between hourly or daily temperatures computation to estimate MGDD and CDD (Online Supplementary Information). Because the temporal resolution of the E-OBS and ERA-5 land data sets are different and are acquired using different methodologies, we evaluated the possible divergence between the MGDD and CDD estimates (194). These metrics calculated with both data agreed, showing a mean difference of 54 and 17 units for MGDD and CDD, respectively, and a standard deviation of 200 units for both metrics (Online Supplementary Information).

### 8.2.4 Vector climatic suitability

Following (205), we used the MaxEnt (350) algorithm to calibrate the relationship of *P. spumarius* global occurrence (predictand) with moisture index and maximum temperatures during summer index (predictors) estimated from 2003 to 2022. Data of the presence records of *P. spumarius* were obtained from The Global Biodiversity Information Facility (GBIF) (180, 447) and different Spanish plant protection agencies and research institutions ("Instituto de Ciencias Agrarias" at CSIC, Madrid, Spain; "Servicio de Sanidad Vegetal de la Junta de Andalucía" based in Sevilla and Jaén, Andalucía, Spain; Sanidad Agrícola Econex S.L. based in Murcia, Spain), as reported in (205). A total of 1652 presence records were used (Fig. 8.1), ensuring that there were no duplicated records within each cell of the climate layer grid.



**Figure 8.1: Training presence records for modeling the distribution of *Phalaenus spumarius*.**

In addition, we randomly generated pseudo-absences, also known as background points, using "The Three-Step" method proposed in (237). This method incorporates a model performance criterion to determine the optimal sampling background extent, thereby ensuring that the model fitting was not adversely affected by the pseudo-absence sampling. Nevertheless, we accounted for the potential variability introduced by randomly selecting points from the background by performing 10 realizations of this sampling process. A total of 4956 pseudo-absences (three times the number of presences) were used in each realization.

Model evaluation was performed using a  $k$ -fold cross-validation approach (where  $k = 10$ ) and the resulting AUCs (Area Under the ROC Curve) consistently exceeded 0.9 within the range of 0 to 1, with a value of 1 indicating perfect prediction and 0.5 indicating no discriminatory power (i.e. random guessing). Finally, the calibrated models were used to predict the suitability of *P.*

*spumarius* in the reference historical period (2003-2022) and under increasing global warming scenarios (panels b, d, f, h in Fig. 8.2).

## 8.2.5 Risk velocity

To assess the dynamic nature of the risk index and its spatial propagation, we introduced the concept of risk velocity, a metric analogous to the recently proposed concept of climatic velocity (285). The risk velocity represents the rate at which the risk index changes over time and spreads across different locations. From an epidemiological perspective, risk velocity can be thought as the speed and direction the host would need to move to maintain its current risk conditions under climate change. Risk velocities were defined following the definition of climate velocity, as the ratio of the risk temporal trend and the risk spatial gradient in each cell. Thus, the units for the risk velocity correspond to kilometres per year (km/year). Risk velocities were computed using the VoCC R package (172, 173).

## 8.2.6 Maps

The maps of all Figures in the paper (Fig. 8.1 to Fig. 8.5) and in the Supplementary Information file were made with Python 3.11 (429) using the package Cartopy v0.21.1 (151, 306).

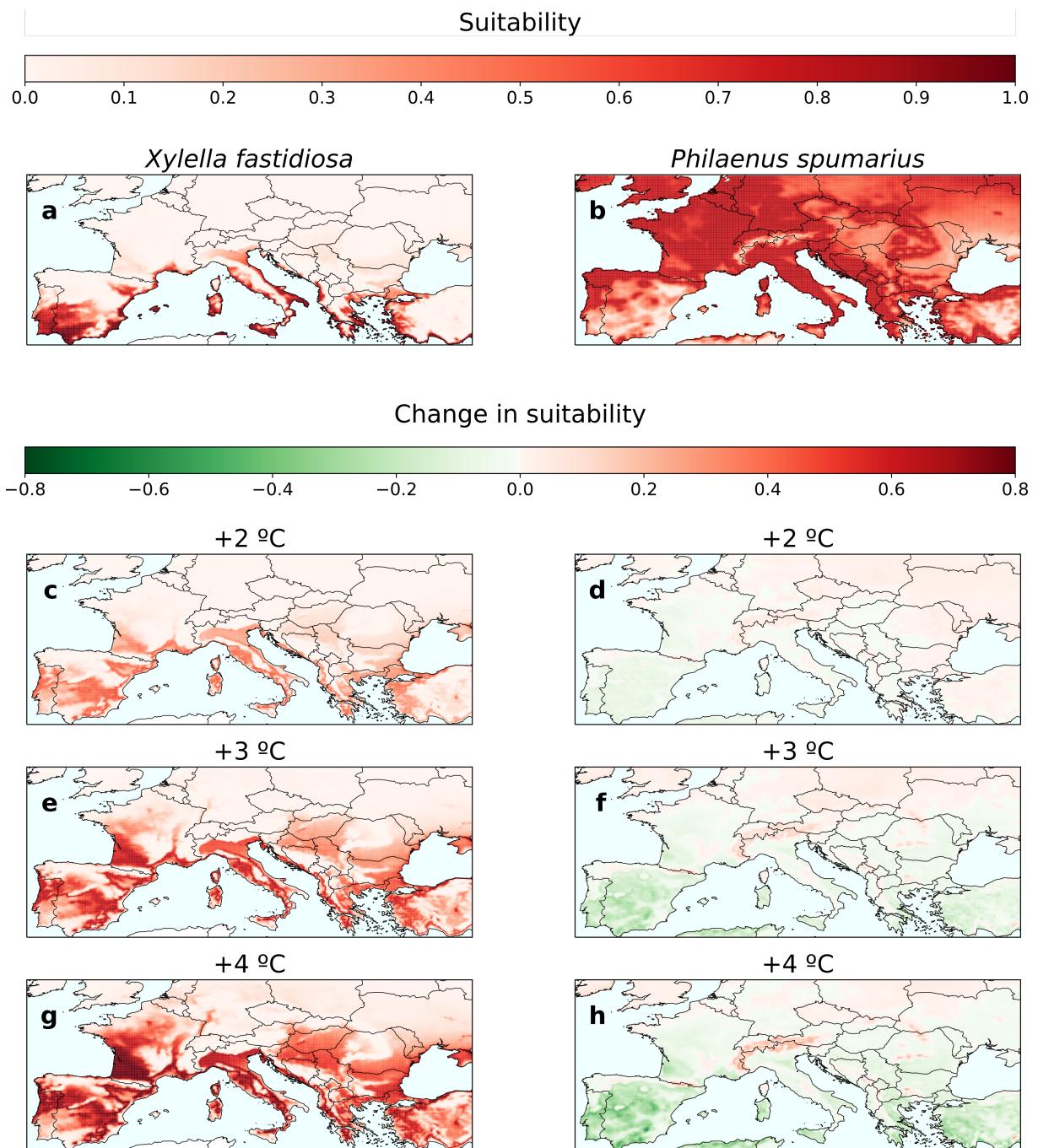
## 8.3 Results

### 8.3.1 Present and future climate suitability for *Xylella fastidiosa* (PD) and *Philaenus spumarius*

To gain a deeper understanding of how climate change affects each component of the pathosystem, we performed a separate analysis of climatic suitability conditions for  $Xf_{PD}$  and  $P. spumarius$ . The thermal dependence of  $Xf_{PD}$  growth and survival within the infected vine was mechanistically modeled by probability functions relating the accumulation of modified degree days ( $MGDD$ ) and cold degree days ( $CDD$ ) to symptom development and recovery,  $\mathcal{F}(MGDD)$  and  $\mathcal{G}(CDD)$ , respectively (see Methods). Climatic suitability for pathogen establishment was then determined by  $\mathcal{F}(MGDD) \cdot \mathcal{G}(CDD)$ , i.e. the overall probability of symptom development during the growing season and subsequent survival for overwintering infection (see Methods). For  $P. spumarius$ , climatic suitability was modeled using an SDM based on a previous study (205), with the climatic moisture index (452) and spring maximum temperatures as key predictors (see Methods). Both analyses were evaluated under current (2003-2022) and future climate conditions considering scenarios of increasing global warming (+1.5°C, +2°C, +3°C, and +4°C) based on the latest generation of regional climate projections covering Europe (240) (see Methods).

Progressive global warming increases the accumulation of MGDD during the growing season and reduces the recovery rate (i.e. CDD) during winter, thus favouring the geographic expansion of the pathogen (Fig. 8.2 and Online Supplementary Information). Conversely, increasing temperatures tend to reduce the climatic suitability of vectors in more arid areas of southern Europe leading to a progressive migration to higher areas and latitudes in continental regions in search of climatic refuge. These general trends hold for both organisms under the +2, +3 and +4 °C temperature increase scenarios (Fig. 8.2).

The mechanistic approach to modelling pathogen establishment risk enables each of the two opposing directional processes of growth and survival (MGDD vs. CDD) to be appropriately weighted in the final result. For example, the Bordeaux region in western France has not been at risk due to low cumulative MGDD and low winter protection effect. In the transition from the +1.5°C scenario to the +4°C scenario, this area will experience a spectacular increase in risk mainly due to the expected summer warming (Online Supplementary Information). Conversely, areas of Central Europe such as Hungary and Serbia already experience suitable conditions for pathogen growth in a +1.5°C scenario ( $\mathcal{F}(MGDD) > 0.6$ ); however, cold winters tend to eliminate any potential summer infection [ $\mathcal{G}(CDD) < 0.3$ ] (Online Supplementary Information). Climate change would further increase the growth of the pathogen and reduce the winter curing effect in Central Europe, ultimately exposing the region to  $Xf_{PD}$  (Online Supplementary Information).

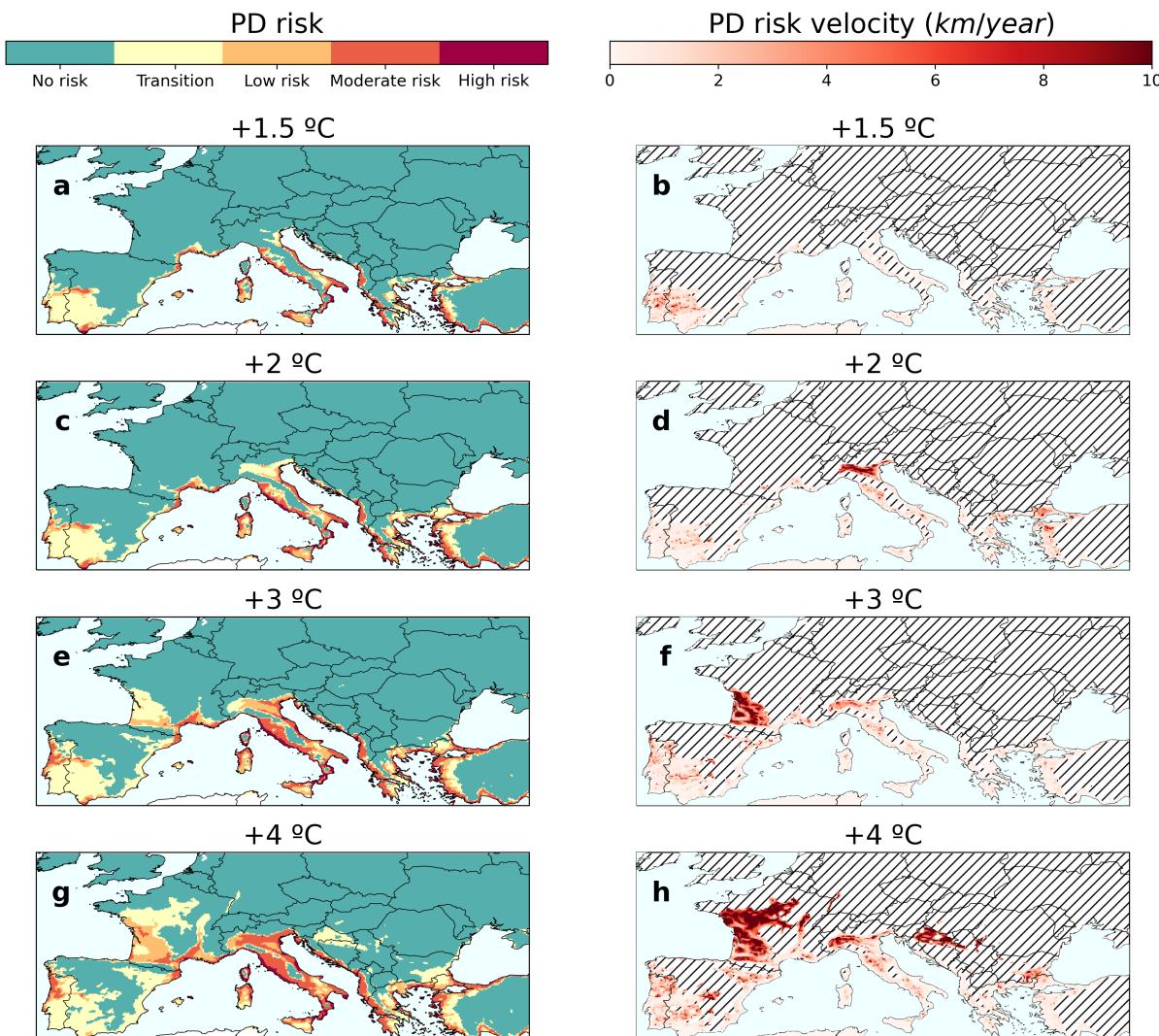


**Figure 8.2: Changes in *Xf<sub>PD</sub>* and *P. spumarius* climatic suitability (i.e. probability of occurrence) under different climate projections compared to the current scenario (2003-2022).** Current climatic suitability for the pathogen (a) and the vector (b). In an increasing temperature scenario (+2°C , +3°C and +4°C) the climatic suitability for the pathogen geographically expands in southern Europe and moves northwards (c, e and g) while the climatic suitability for the vector decreases (d, f and h). The suitability values for each scenario correspond to a 20-year average.

### 8.3.2 Pierce's disease risk projections under climate change

The limited intersection between the climatic suitability ranges for the pathogen and the vector (Fig. 8.2a and b) suggests a marginal risk of PD epidemics in Europe. Since disease transmission requires a vector, the climate-suitability maps for *P. spumarius* indicate a lower risk and potential economic impact of *X. fastidiosa*-induced diseases on any host in southern Europe, particularly Spain, than previously predicted (391). Realistic risk maps require a defined epidemiological

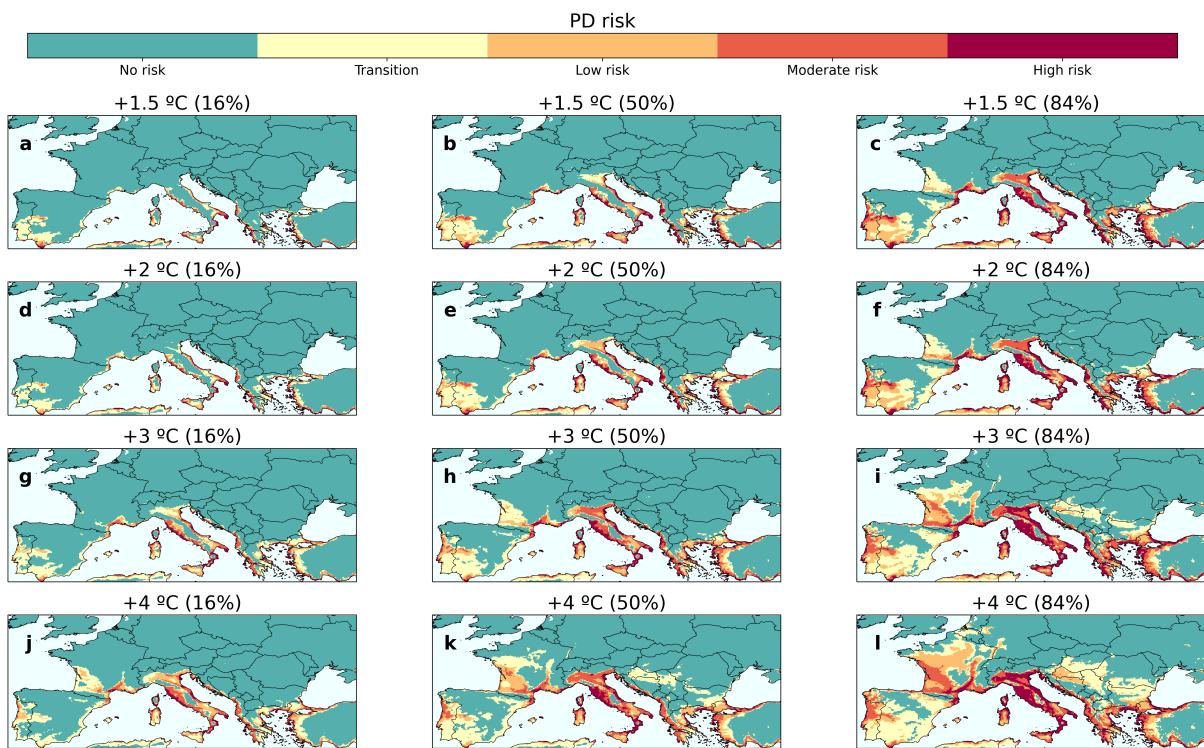
framework to account for inter-annual climate variation and transmission in disease dynamics, in addition to accounting for changes in the distribution of climatic conditions favorable to the pathogen and vector (i.e. climatic suitability). Our epidemic risk model focuses on delimiting the disease dynamics by simulating an epidemic process in which the emergence of newly exposed hosts is influenced by the climatic suitability of the vector, while the transition to the infectious state is driven by the climatic suitability for  $Xf_{PD}$  chronic infections. The effective growth rate of the infected host population over the simulated period is used to derive a risk index  $r$ , bounded between  $-1$  and  $1$ . Within this modeling framework, different risk categories naturally emerge: no risk ( $r < -0.1$ ), transition zone ( $-0.1 \leq r < 0.1$ ), low risk ( $0.1 \leq r < 0.33$ ), moderate risk ( $0.33 \leq r < 0.66$ ) and high risk ( $r \geq 0.66$ ). For further details see the Methods section and the original paper (194).



**Figure 8.3: PD risk maps and associated risk velocities under different climate projections.** (a,b)  $+1.5^{\circ}\text{C}$  climate projection. (c,d)  $+2^{\circ}\text{C}$  climate projection. (e,f)  $+3^{\circ}\text{C}$  climate projection. (g,h)  $+4^{\circ}\text{C}$  climate projection. Risk velocities have been calculated only in risk zones,  $r > 0$ , in each scenario. Hatched lines in panels (b,d,f,h) indicate no risk zones where risk velocities have not been calculated.

Global warming ( $+1.5^{\circ}\text{C}$ ,  $+2^{\circ}\text{C}$ ,  $+3^{\circ}\text{C}$ , and  $+4^{\circ}\text{C}$ ) is expected to increase the risk of PD epidemics in southern Europe, with France, Italy and Portugal being particularly affected (Fig. 8.3 and Online Supplementary Information). This general trend affects each of the risk categories under the different climate change scenarios (Online Supplementary Information). Furthermore, we observed that a global temperature increase above  $+3^{\circ}\text{C}$  represents a tipping point for the possible spread of PD beyond the Mediterranean (Fig. 8.3 and Online Supplementary

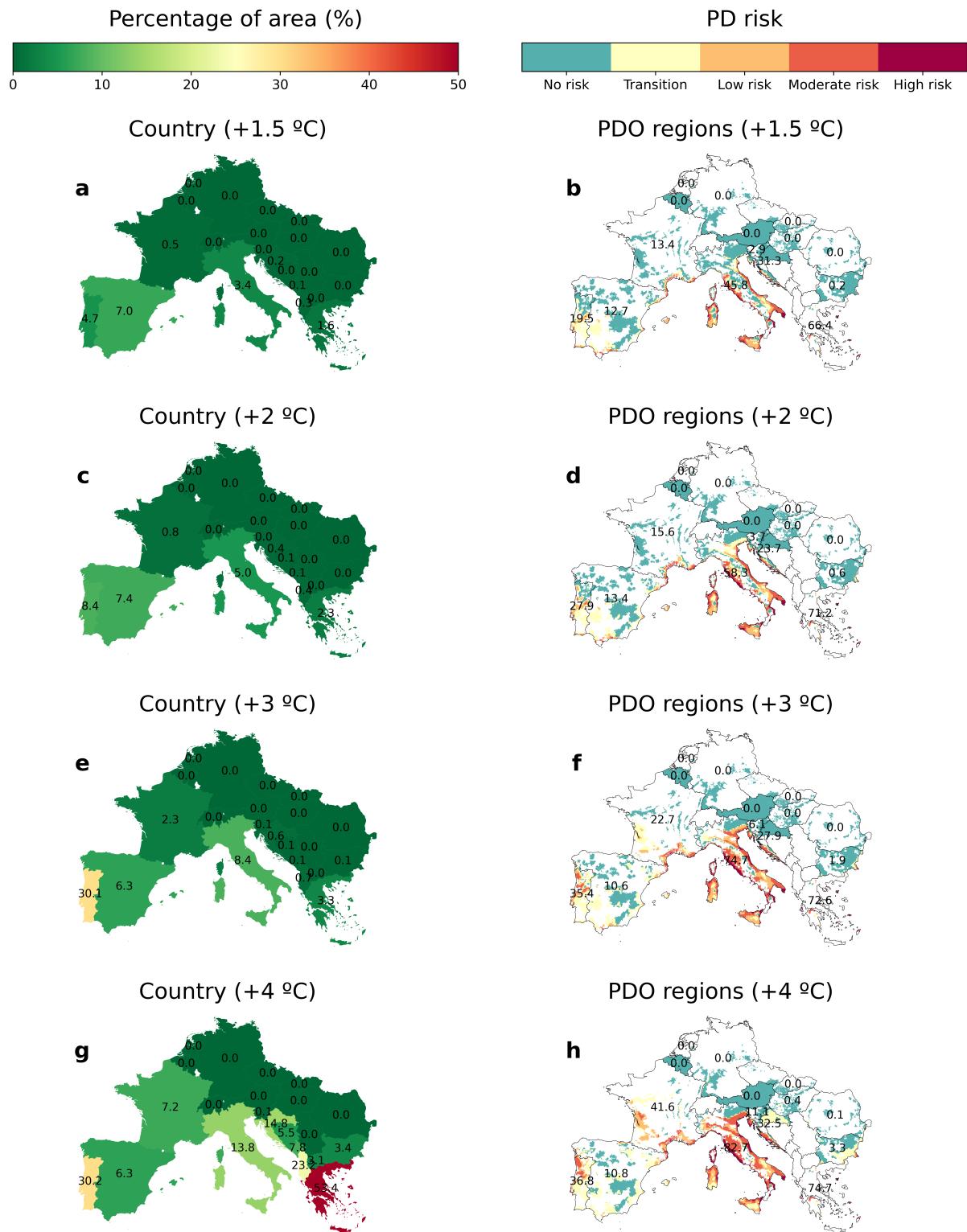
Information). To quantify the potential spread of PD, we calculated risk velocity, an index that allows us to identify areas where risk is changing or spreading rapidly (see Methods). We found a consistent and notable increase in the mean risk velocity within most of the identified risk zones (Fig. 8.3 and Online Supplementary Information), increasing from almost 1 km  $y^{-1}$  to 5 km  $y^{-1}$  as the temperature rises from a +1.5°C to a +4°C scenario (Online Supplementary Information). This acceleration is evident when we compare that in the +1.5°C scenario, approximately 6% of the grid cells have risk velocities greater than 5 km  $y^{-1}$ , while this value increases to 50% in the +4°C scenario (Online Supplementary Information). Furthermore, our estimates of PD risk velocity are broadly consistent with estimates of the velocity of temperature change (285), indicating that shifts in PD risk in our model adequately track climate change.



**Figure 8.4: Uncertainty in PD risk projections for climate change warming levels.** The maps show the result of the median risk values of the set of 40 regional climate model projections for each level of temperature increase (b, e, h, k) and the uncertainty of the projections considering the  $1\sigma$  deviations from the median, this is, the 16% (a, d, g, j) and 84% (c, f, i, l) percentiles. For each warming level (each row), the projected risk of PD is comprised between the results shown in the 16% percentile (first column) and those shown in the 84% percentile (last column) with 68% probability ( $1\sigma$ ).

Fig. 8.4 shows the uncertainty in the projections of the PD risk map, comparing the 16th and 84th percentiles ( $1\sigma$ ) from the set of 40 regional climate models to the median risk map. The spatial distribution of PD risk is robust across models, while the uncertainty in the level of warming is bounded by a  $\pm 1^\circ\text{C}$  increase (Fig. 8.4), e.g. the median spatial distribution of risk obtained for a +3°C warming level is expected to occur between the +2°C and the +4°C scenarios under a  $1\sigma$  confidence level. This means that, depending on the specific model, a given spatial distribution of risk (e.g., Fig. 8.4 h) may manifest under a +4°C warming scenario in more conservative projections or a +2°C warming scenario in others, while most models project it for a +3°C warming level. This indicates a high degree of confidence in the location and projected severity of future outbreaks, but greater uncertainty in their timing.

In order to improve the accuracy of the risk maps and the relative impact of PD, we carried out a comprehensive analysis at multiple scales, from the national level to the regions with Protected Designation of Origin (PDO) and finally taking into account the distribution of vineyards. This approach allows us to disaggregate the results at different administrative levels to facilitate the design of risk management and the implementation of appropriate phytosanitary measures.



**Figure 8.5: Multi-scale spatial analysis of PD future risk in Europe.** (a, c, e, g) Percentage of country areas at risk ( $r > 0.1$ ) for each climate projection. (b, d, f, h) PD risk zones in Protected Designation of Origin (PDO) wine regions for each climate projection. PDO data was obtained from [79]. The corresponding interactive analysis at the vineyard level can be found in [188]

Overall, our model simulations show a consistent increasing trend in the risk of PD in Europe under all climate change scenarios. The percentage of land area at risk in Europe increases

from 0.32% in the +1.5°C scenario to 1.87% in the +4°C, while the number of regions with PDO at risk increases from 18.17% to 47.32%. The vineyard area increases from 18.67% to 40.35% (Online Supplementary Information). At country level Portugal and Greece face the highest overall risk, escalating from 12% and 2% of their area in the +1.5°C scenario to a striking 47% and 63% respectively in a +4°C scenario. In contrast, countries such as France and Italy experience a smaller but still significant increase in risk area, never exceeding the 20% threshold, while Spain, the third largest wine producer, shows a decreasing trend in risk areas above the +2°C scenario (Fig. 8.5 and Online Supplementary Information). Such contrasting patterns in PD risk between countries only emerge when using our modelling framework.

A different picture emerges when looking at the spatial distribution of PDO regions and vineyards. For example, PD risk within French and Italian PDO regions increases significantly from 13.4% and 45.8% in a +1.5°C scenario to 41.6% and 82.7% in a +4°C scenario (Fig. 8.5 and Online Supplementary Information), while the percentage of vineyard area at risk rises from 24.21% and 57.49% in a +1.5°C scenario to an astonishing 80% in a +4°C scenario. Important European PDOs would be at risk from a warming of +2°C, such as parts of the southern Rhône Valley (Châteauneuf du Pape), Provence and Languedoc in France, Penedés in Spain, Bairrada in Portugal, and Chianti and Brunello di Montalcino among others in Italy (see Supplementary Information). A detailed interactive analysis of the impact of PD in European PDO regions and vineyards is available on our website (188).

## 8.4 Discussion

Previous research has attempted to assess the potential geographic distribution of  $Xf_{PD}$  subspecies, the insect vector *P. spumarius* and PD using species distribution models (SDMs) under future climates. While these climate-suitability-based predictions provide insights into the ecological niche of key disease players, bioclimatic correlative models neglect disease dynamics, a key factor in avoiding disease overestimation (194). Unlike previous attempts, our approach integrates the compound effect of climate change in the pathosystem using a mechanistic epidemiological model to overcome these limitations. Unlike dimensionless climatic suitability indexes or disease probabilities used in SDMs, the risk index,  $r$ , in our model provides information on the expected growth rate in the event of an outbreak. Furthermore, the risk index is not fixed but varies annually depending on the weather conditions of the previous years. Inter-annual climate variability thus has an impact on disease dynamics, especially in areas where the risk index is lower. Another feature of our risk predictions is their lack of ambiguity, which should not be confused with certainty. Risk estimates are based on  $R_0$ , which depends on the insect vector population (195), among other factors. Areas where  $r < -0.1$  permanently cannot theoretically support an outbreak, and the population of infected plants will decline over time. For example, our model clearly indicates that there is no risk of PD in the UK. This is not an arbitrary threshold; it is given by the epidemiological model. It is therefore very likely that the absence of PD in continental Europe is a consequence of low risk indices and that it has only become established in certain coastal areas since the late 1990s. On the contrary, the risk index in the Mediterranean islands has remained moderately high with little variation over the last 40 years (194).

Because Pierce's disease has only affected vineyards on the island of Mallorca (315), little attention has been paid to the risk of it reaching continental vineyards. Our risk model indicates why this possibility was very low until the mid-90s, and what the conditions were for it to occur on the Mediterranean islands (194). In this work, we clearly show that with increasing temperatures PD will become a serious threat to important wine-growing areas in southern Europe that were not previously at risk. A key finding of our study is the identification of a tipping point for the risk of PD establishment at a global mean temperature increase of +3°C. Beyond this threshold, the risk of PD spreading north of the Mediterranean region becomes remarkably higher, while the risk of PD epidemics in Portugal, Italy and France (Fig. 8.3) undergoes a significant quantitative leap. This suggests that as global temperatures continue to rise, the range of PD may expand into new territories. Indeed, the projected increase in risk velocities under higher warming scenarios further emphasizes the potential for rapid spread of PD into previously unaffected regions (Fig. 8.3).

Pest risk map projections are subject to uncertainties inherent in the variability of climate model predictions (436). While previous studies on pathogen and vector distributions have been based on a limited number of climate models, our risk maps are based on the most modern

set of regional climate projections produced by the EURO-CORDEX initiative, reflecting the state-of-the-art knowledge (Table 8.1). This allows us to adequately estimate the uncertainty of the resulting PD risk map projections for each temperature rise scenario. This confirms that although the spatial distribution of the risk of establishment is robust, there is an uncertainty of  $\pm 1^{\circ}\text{C}$  in the level of warming (Fig. 8.4). The models are therefore fairly good at pointing where the increased risk will occur, but it is more difficult to know when it will be reached.

Overall, our results highlight the contrasting effect of climate change on PD risk distribution in Europe, revealing it as a multi-factor and multi-scale process (??). Climate change has an opposite effect on each component of the pathosystem, enhancing areas of potential chronic PD infections while diminishing the suitable geographic range for the vector. At the same time, the characteristic spatial scale at which risk is assessed strongly influences conclusions. At the country level, there are significant variations in the extent of accumulated risk between different projections. However, when analyzed at a finer scale, such as at the level of PDO regions or vineyards, the results change completely. Countries that previously had marginal areas at risk now show a higher percentage of PDO regions and vineyard at risk. These results underlie the urgency of tailored mitigation and adaptation strategies to protect vineyards and PDOs, considering their specific spatial distribution and risk index, as well as the potential impacts of climate change.

Our results are influenced by the intrinsic uncertainty associated with the correlative models used to determine the spatial distribution of the vector, the epidemiological parameters, and the uncertainties in the climatic projections. Although the spatial resolution of our climate projections is considered to be high, it may not capture the complex micro-climate structure found in certain European wine-producing regions. Therefore, risk assessment results could differ locally with higher-resolution data. In addition, we have not considered the possible influence of climate change on latitudinal and altitudinal shifts in the distribution of European vineyards (218, 320), as this would only affect the calculation of the percentage of vineyard surface at risk but not the actual spatial distribution of risk. In any case, the risk estimates for the PDO regions include areas much larger than the areas of planted vines, which allows some margin in the adaptation and migration of the vineyards to different micro-climatic conditions. In addition, the PDO and vineyard databases used in this study are also have their own limitations. Future studies incorporating more refined modeling techniques, specific regional grapevine varieties, crop management and improved data resolution would enable a more nuanced understanding of PD risk and its potential impact at the local scale.

It is noteworthy that the mathematical framework employed in this study could be applied to other *Xylella fastidiosa* diseases, such as Almond Leaf Scorch Disease or Olive Quick Decline Syndrome and, more generally, to other vector-borne plant diseases. However, this requires the availability of some specific data and conducting some experiments. First, data for the temperature-dependent growth rate of the pathogen is needed to build the function that computes the MGDD. Then, symptom development experiments need to be carried out to build the  $\mathcal{F}(\text{MGDD})$  and  $\mathcal{G}(\text{CDD})$  functions that relate symptom development with temperature. Finally, the spatial distribution of the agent responsible for disease transmission is desired. Of course, using presence/absence data one can use SDM to obtain this spatial distribution.

Climate change is currently one of the biggest challenges for EU agricultural policy (156). Quantitative regional predictions of climate change on emerging diseases, such as this one, provide a valuable and unambiguous tool for decision-making. In our approach to the problem, risk indexes not only include information on where or where not PD may become established, but also reflect the exponential growth rate of potential epidemics, which are directly related to their potential economic impact. In addition, risk indices and velocities provide a dynamic framework for assessing the feasibility of eradication efforts when  $Xf_{\text{PD}}$  is detected in a new area, providing critical information for strategic crop protection. Our study evidences the need to selectively allocate more resources to surveillance and research on PD in southern European countries, considering the associated uncertainties. This strategic allocation of resources based on risk assessment can help to prioritise proactive measures and effectively manage the potential impact of PD in different European countries.

Our research highlights the complex dynamics of PD and its relationship with climate change. By adopting an interdisciplinary approach that integrates climate projections, epidemiological modelling, and spatial analysis, we provide valuable insights into the potential establishment and

spread of PD in European wine-growing regions from the country to the vineyard levels. Our study demonstrates that an accurate assessment of the risk of PD establishment requires a nuanced understanding of the vector-plant-pathogen-climate system and the explicit consideration of the vineyard spatial setting. These findings can inform decision-making processes and support the development of effective strategies to mitigate the risks posed by PD and safeguard the future of viticulture in the face of a changing climate.





## 9. Pierce's Disease risk with high-resolution climate data



Àlex Giménez-Romero<sup>1</sup>, Eduardo Moralejo<sup>2</sup>, Manuel A. Matias

1. Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC-CSIC-UIB), Palma de Mallorca 07122, Spain

2. Tragsa, Passatge Cala Figuera 6, 07009 Palma de Mallorca, Spain

### Published as

Àlex Giménez-Romero, Eduardo Moralejo, and Manuel A Matias. "High-resolution climate data reveals increased risk of Pierce's Disease for grapevines worldwide". In: *bioRxiv* (2024). DOI: [10.1101/2024.03.06.583743](https://doi.org/10.1101/2024.03.06.583743). eprint: [https://www.biorxiv.org/content/early/2024/03/11/2024.03.06.583743](https://www.biorxiv.org/content/early/2024/03/11/2024.03.06.583743.full.pdf). URL: <https://www.biorxiv.org/content/early/2024/03/11/2024.03.06.583743>

## 9.1 Introduction

Climate plays a pivotal role in shaping the distribution and dynamics of agricultural pests and pathogens (31, 34, 132, 219, 267), with implications for global food security (165, 368). As our climate undergoes unprecedented changes due to anthropogenic activities, agriculture faces multifaceted threats ranging from alterations in temperature and precipitation patterns to increased frequency of extreme weather events (402). Such shifts create novel environments that may favour the proliferation of certain pests or pathogens while posing challenges to the survival of others (31, 147). The consequences of these changes extend beyond immediate agricultural landscapes, reverberating through global food systems and posing significant challenges to the sustainability and resilience of food production (339).

Understanding the intricate relationships between climatic conditions, the pathosystem components, and the subsequent epidemiological dynamics is essential for developing effective strategies to mitigate and manage emerging agricultural challenges, especially in the face of changing environmental conditions. However, modelling disease epidemics is a complex task, as they are emergent phenomena resulting from non-linear interactions between disease components that also exhibit non-linear responses to changes in environmental variables (177, 245, 386). Thus, while climate primarily determines the potential geographic range of each organism in the pathosystem, the development of epidemic outbreaks depends on favourable host-pathogen-vector-climate interactions that drive transmission chains.

It has long been recognised that ecological phenomena typically depend on the scale of description, particularly with regard to the effects of climate (280). Climatic databases with finer spatial resolution are continuously being developed with the goal of allowing more accurate predictions (330). Some recent studies have shown that the local climate experienced by individuals might deviate substantially from regional averages, with implications for the population dynamics of a forest herb (102). Likewise, the choice of climate data affects the predictions of species distribution models (SDMs) (3). In particular, the spatial resolution of the data can influence the predictions of invasion risk for some species (146). It is therefore clear that the resolution of climate data will have a significant impact on predicting the risk of plant diseases and pests.

Among emerging pathogens *Xylella fastidiosa* (Xf) is considered one of the most dangerous phytopathogenic bacteria worldwide (1, 234). It is naturally transmitted by xylem sap-feeding insects, such as sharpshooters and spittlebugs, and exhibits a broad host range that encompasses economically important crops such as grapevines, citrus, almonds and olive trees (1, 364). The consequences of Xf diseases are devastating: about 200 million citrus trees are infected annually in Brazil (284), there are losses over \$100 million annually in the grape industry in California (424) and approximately 21 million olive trees have been killed by the bacterium on the Apulia region in Italy (375). Assuming massive spread throughout Europe, Xf has been projected to potentially contribute up to €5.2 billion of annual losses in the olive sector alone (391). Overall, Xf diseases pose a major threat to agrosystems worldwide, highlighting the need for precise and predictive models to guide effective management practices.

Previous research has provided insights into the potential geographic range of Xf subspecies through SDMs (55, 207). These models, however, have led to overestimates of risk by failing to account for the distribution and abundance of potential vectors necessary for disease transmission (205). A quite different approach to mapping PD risk has been developed based on climate-driven epidemiological models with the option to integrate vector's distribution information and the specificity of the Xf subsp. *fastidiosa* strain responsible for PD (hereafter Xf<sub>PD</sub>) (194). This model correctly identifies areas in the United States with recurrent PD outbreaks and forecasts increasing epidemic risk in Mediterranean islands and coastlines with ongoing climate change.

Although risk maps based on hourly temperature data from the ERA5 have allowed fine adjustments in the calibration of the thermal response to Xf infection, these achievements have entailed losses in spatial resolution (0.1° spatial resolution) (194, 322). Such limitation is particularly significant when dealing with vector-borne plant diseases like PD, where the interactions between the pathogen, vector, and host plants exhibit non-linear responses to climatic conditions. Subtle variations in temperature, humidity, or precipitation at the local scale thus can have profound effects on the reproduction and life cycles of the organisms involved

and, hence, on the dynamics of disease transmission.

Topographical heterogeneity is a recognised issue in invasion biology, but has received little attention in crop science. Vineyards are increasingly located in valleys, ridges, hillsides and riverbanks usually with altitudinal and microclimatic gradients in short transects. They are therefore a remarkable example of a crop subject to scaling problems when studying ecological or epidemiological processes at regional and global scales. In this work, we address this spatial resolution limitation by modelling the risk of PD using high-resolution climate data from the CHELSA dataset (253). The study period was deliberately chosen to include real data on temperature increases due to ongoing climate change. Our study shows a greater global risk of PD and a higher rate of risk increase, underscoring the urgency of reevaluating global strategies to prevent the spread of the pathogen with international trade in plant diseases.

## 9.2 Results

### 9.2.1 Global differences in PD risk between coarse and fine-grain climate data

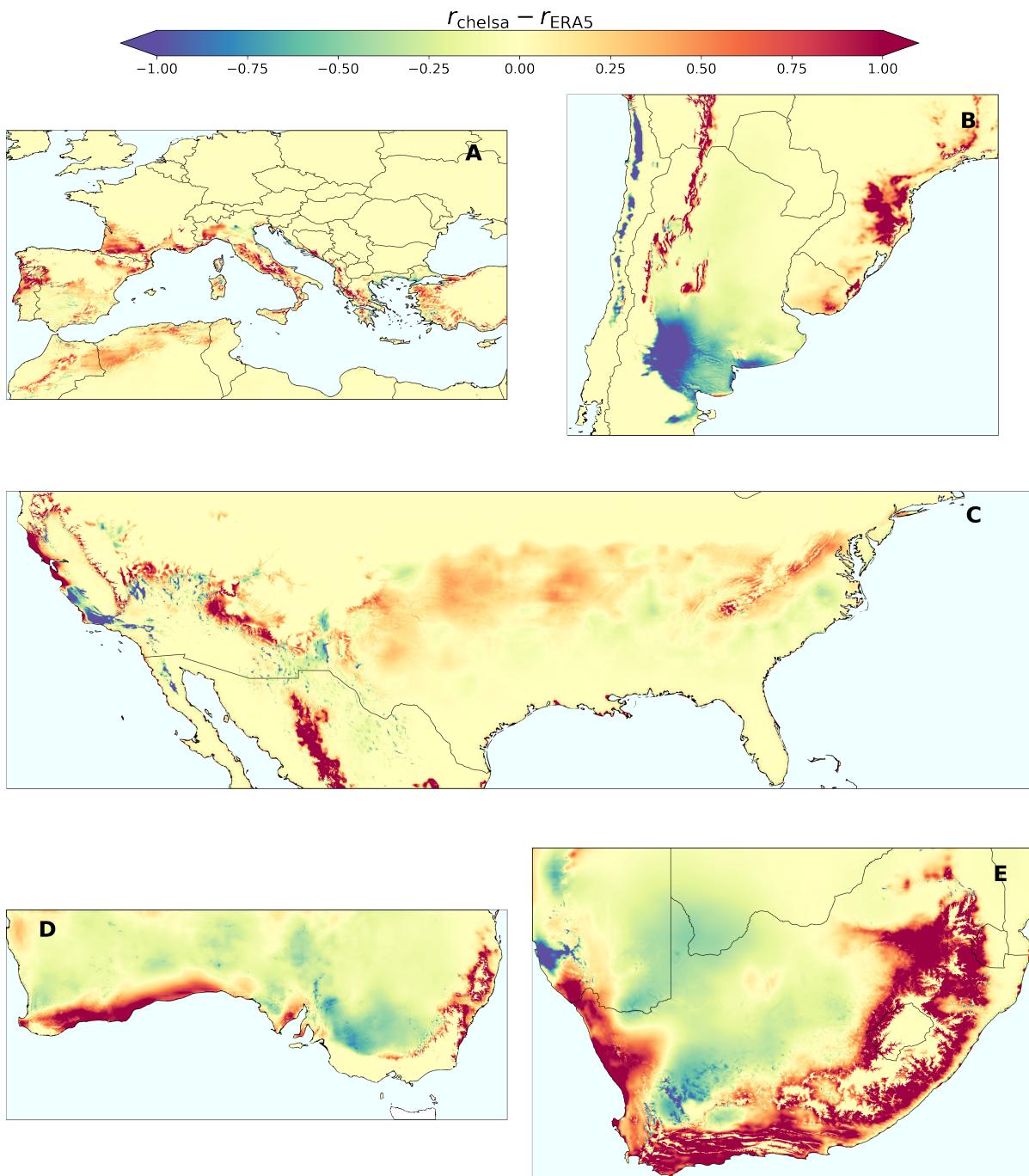
We computed the risk of PD using the previously developed climate-driven epidemiological model (194) coupled with the CHELSA dataset (253), which features key climate variables (e.g. temperature and precipitation) at a high spatial resolution of 1 km and daily temporal resolution covering the period 1979–2016. The resulting spatial and temporal patterns of disease risk in the main wine-growing regions were compared with previous risk projections derived from the ERA5 dataset (324), characterised by an intermediate spatial resolution of 10 km and hourly temporal resolution (194). Briefly, the model simulates the initial dynamics of the disease influenced by climatic variables and the presence of vectors, giving rise to a risk index,  $r$ , which represents the normalised growth rate of the infected population, where  $r = 1$  is the maximum rate achieved at optimal climatic conditions (see Methods). Negative risk indices project an exponential decrease of the infected population (no risk), whereas positive values give rise to an outbreak, with higher values accounting for major incidence and potential severity. Risk categories emerge naturally from this formalism as No Risk ( $r \leq -0.1$ ) Transition ( $-0.1 < r \leq 0.1$ ), Low Risk ( $0.1 < r \leq 0.33$ ), Moderate Risk ( $0.33 < r \leq 0.66$ ) and High Risk ( $r > 0.66$ ). Risk projections in Europe use the climatic suitability,  $s$ , of the main European vector, *P. spumarius* (see Methods), while for the rest of the world it is assumed that there are no risk-limiting effects due to the vector ( $s = 1$ ), but only due to climatic conditions.

**Table 9.1:** Changes in Pierce’s Disease Risk Zones in Different Viticulture Regions. The table illustrates transitions between risk and no-risk categories, as well as transitions among risk categories, highlighting the dynamic shifts in risk patterns across viticulture areas in Europe, the USA, South Africa, South America, and Australia. Risk increase refers to changes from low to moderate risk or from moderate to high risk. Likewise, risk decrease refers to changes from moderate to low risk or high to moderate risk.

	Europe	USA	South Africa	South America	Australia
<b>Risk to no-risk (km<sup>2</sup>)</b>	1.91e+04	3.93e+04	2.26e+04	2.49e+05	5.81e+04
<b>Transition to risk (km<sup>2</sup>)</b>	6.37e+04	1.37e+05	5.53e+04	3.79e+04	8.49e+04
<b>Risk decrease (km<sup>2</sup>)</b>	3.58e+04	1.46e+05	3.56e+05	1.76e+05	1.28e+06
<b>Risk increase (km<sup>2</sup>)</b>	6.28e+04	2.37e+05	1.05e+05	1.50e+05	1.55e+05
<b>No risk to risk (km<sup>2</sup>)</b>	2.04e+05	2.36e+05	2.77e+05	1.90e+05	2.15e+05
<b>Total changes (km<sup>2</sup>)</b>	3.85e+05	7.95e+05	8.15e+05	8.03e+05	1.79e+06

When contrasting model results derived from high- and medium-resolution data for the latest available time (2016), the disparity in risk projections extends beyond regional differences, showing a global increase in risk indices across wine-growing areas (Fig. 9.1 and Online Supplementary Information). Overall, these increases (Fig. 9.2) in the extension of PD risk areas ranged from 100,000 to 1 million km<sup>2</sup> across viticulture regions worldwide. Transitions from no-risk to risk zones covered an area one order of magnitude larger than those in the opposite direction –from risk to no-risk (Fig. 9.2 and Table 9.1). In total, a surface of 4.6 million km<sup>2</sup> changed its risk category with the CHELSA database, representing about a 16% of the land area studied. In contrast, the largest decreases in the risk indices occurred mainly in the Southern Hemisphere, although with

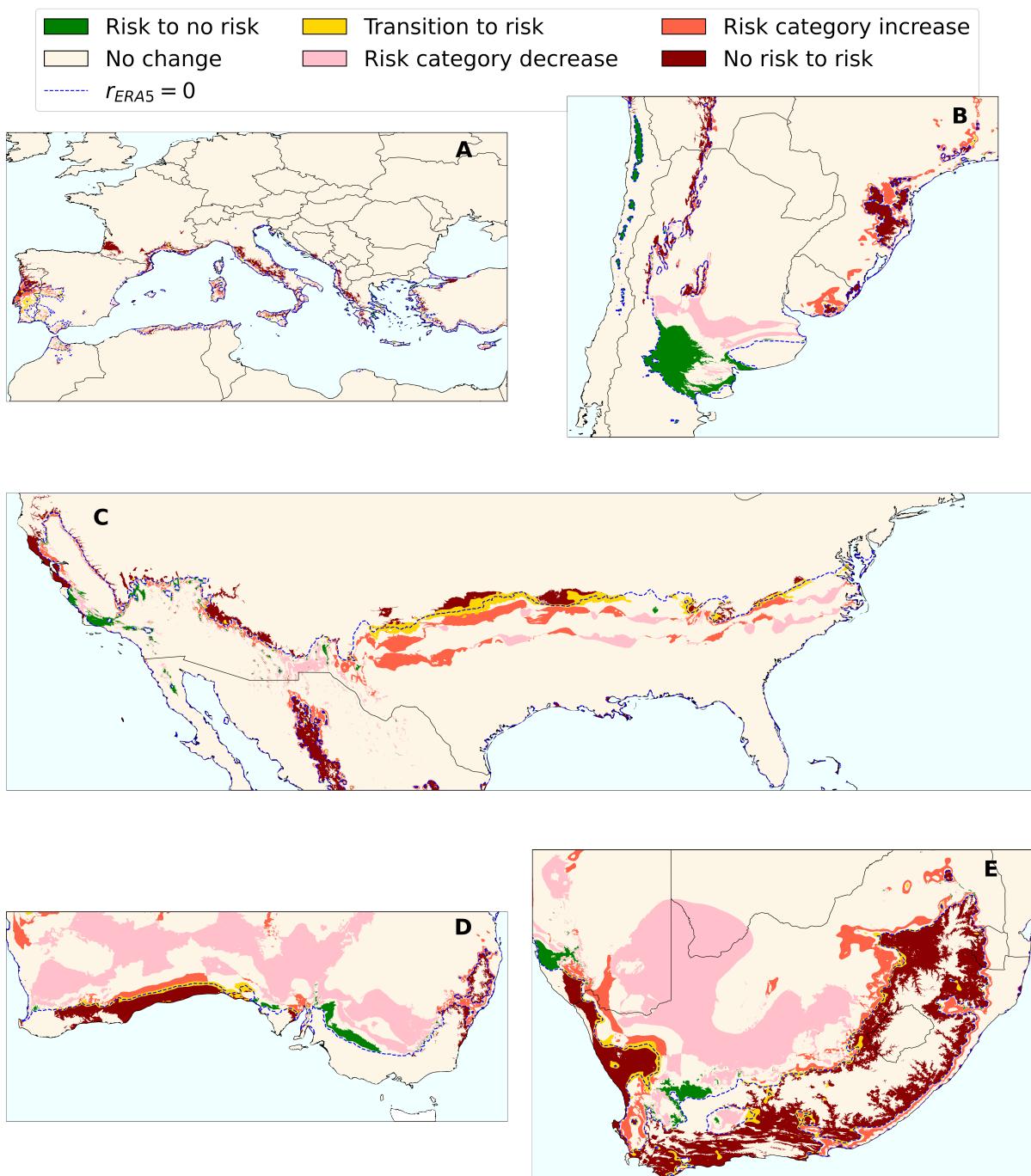
few exceptions most of these decreases remained within the risk zones (Fig. 9.1), while similar land expansions were observed to increase their risk category (low to moderate or moderate to high) (Fig. 9.2 and Table 9.1). The largest changes in risk indices occur in ecotones on both sides of the  $r = 0$  line, as is clearly seen in the south-eastern United States, in coastal areas (e.g., southern Australia and northern California) due to higher resolution that better distinguishes between land and coast, and finally in the river valleys and slopes of mountain systems (Fig. 9.2 and Table 9.1).



**Figure 9.1:** Difference in risk projections based on CHELSA (high-resolution, 1 km) and ERA5 (mid-resolution 10 km) datasets in global viticulture areas. (A) Europe (B) South America (C) United States (D) Australia (E) South Africa.

Next, we compared the temporal progression of the area at risk using both high and mid-

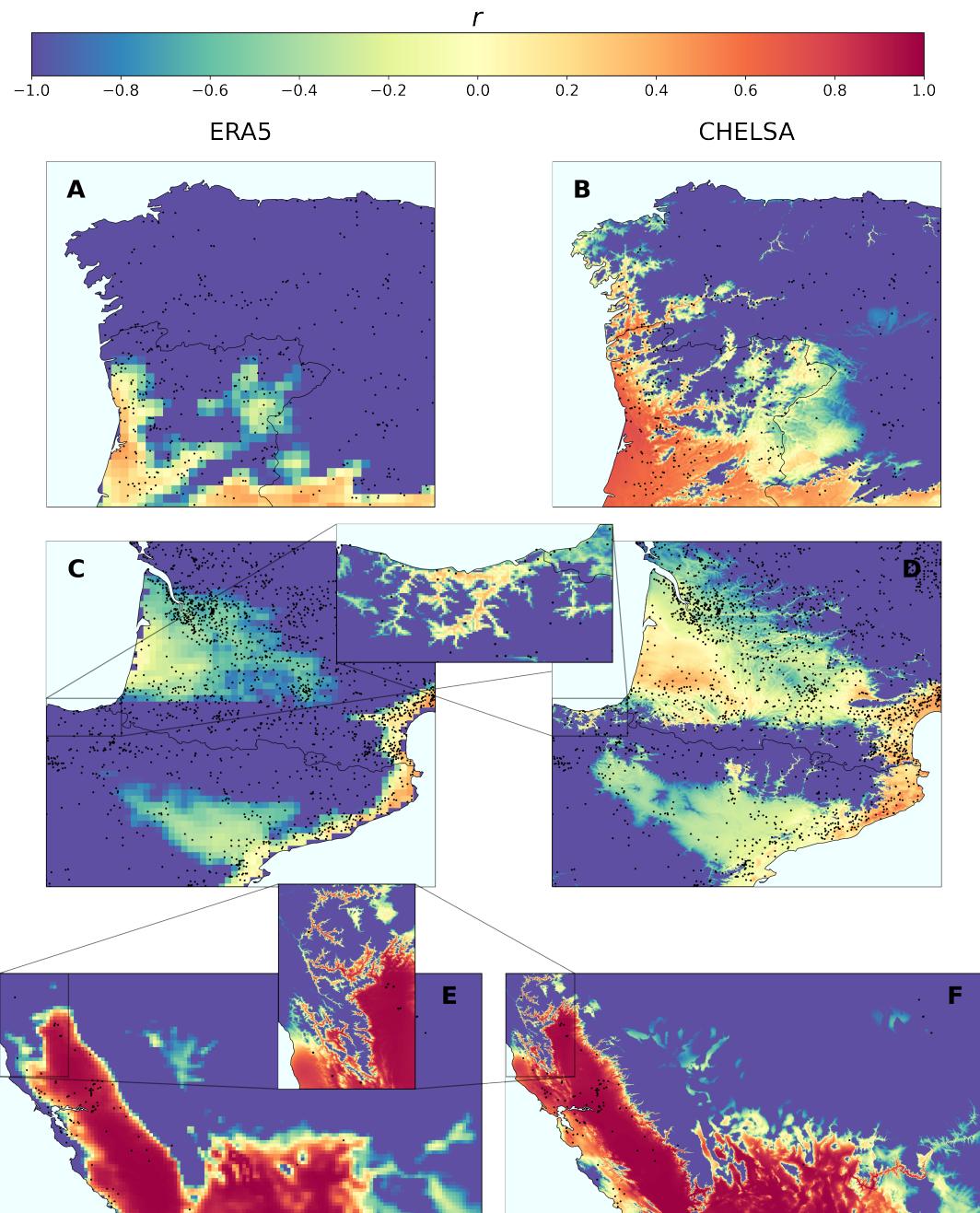
resolution data over the entire available time span (1986-2016), considering that the risk for each year is computed based on the preceding seven years. We found a notable global surge in the rate increase of the area at risk within viticulture zones worldwide, practically doubling previous estimates (Online Supplementary Information). These results point to an accelerated pace at which the risk of PD is growing, compatible with the predictions of different global warming scenarios (200).



**Figure 9.2:** Changes in risk categories between CHELSA (high-resolution, 1 km) and ERA5 (mid-resolution 10 km) projections in global viticulture areas. (A) Europe (B) South America (C) United States (D) Australia (E) South Africa. Risk category increase refers to changes from low to moderate risk or from moderate to high risk. Likewise, risk category decrease refers to changes from moderate to low risk or high to moderate risk.

## 9.2.2 Pierce's Disease risk surges in previously unresolved microclimates

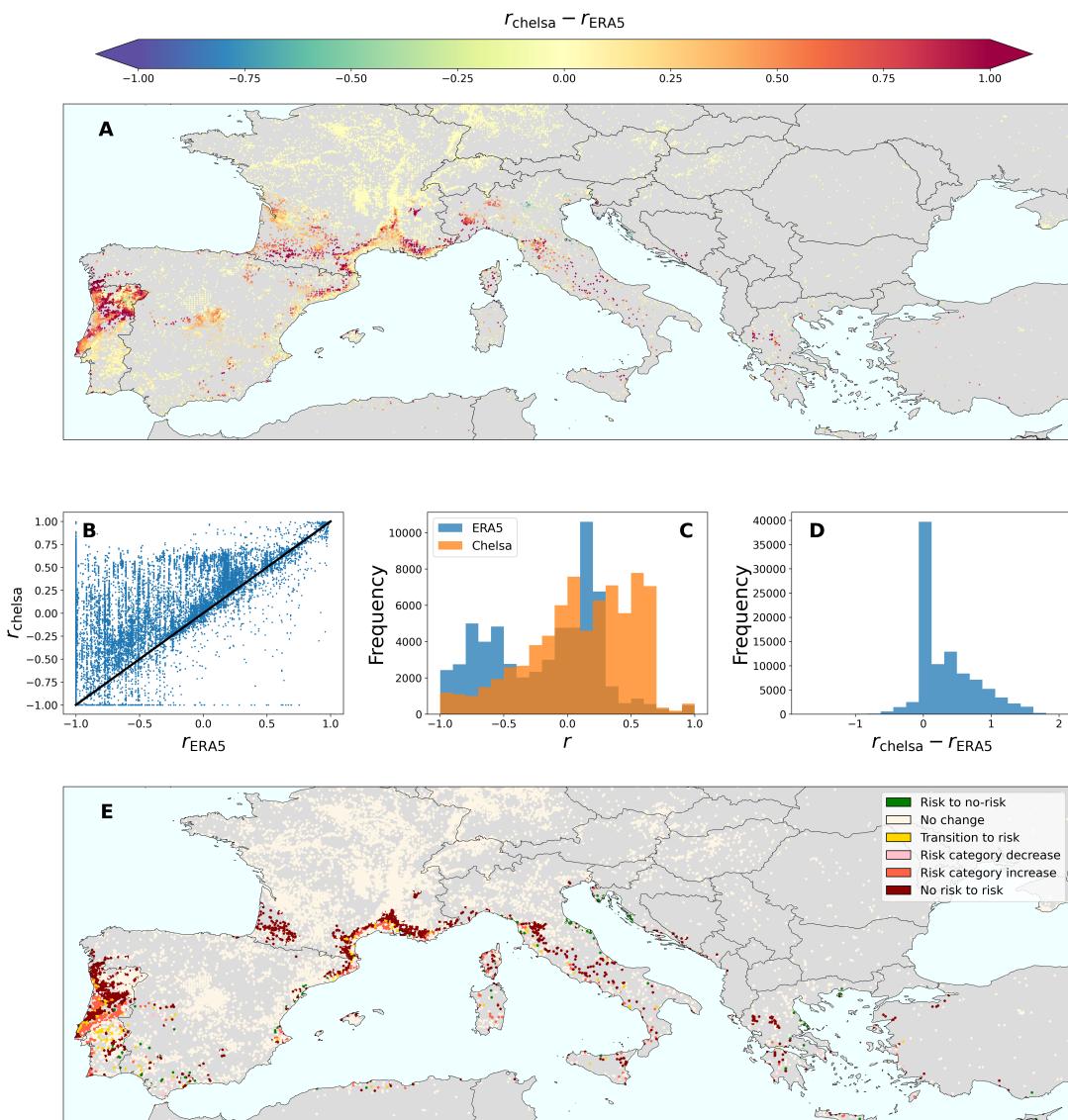
River valley vineyards are renowned for their high quality wines, such as Douro, Napa and Rhone and many others. It is therefore important to understand the risk of PD with climate change at a more detailed level. In our analysis, we have identified rivers and valleys as specific relief areas where a greater increase in PD risk is observed when employing CHELSA's finer- scale climate data (Fig. 9.3).



**Figure 9.3:** Effect of microclimatic conditions of rivers and valleys on Pierce's Disease of grapevines. Comparison of the risk predicted using ERA5 mid-resolution dataset (A,C,E) and CHELSA high-resolution dataset (B,D,F). (A-B) North-western Iberian Peninsula. (C-D) Sourthern France and North-eastern Spain. (E-F) Western United States. Black dots represent grapevines (*Vitis vinifera*) presence data obtained from GBIF (see Methods).

In some important wine-growing areas of southern Europe, we observed an abrupt emergence of risk zones previously classified as no-risk when using lower resolution climate data

(Fig. 9.3). Such pronounced differences in risk patterns are highlighted for example in the fairly steep valleys and hillsides along the Douro River in Portugal, where the specific microclimatic conditions were previously obscured by the coarser resolution of the ERA5 data. These findings are particularly significant for PD, as vineyards are often located in close proximity to rivers or valleys and their surroundings, creating microclimates that attenuate cold winters (black dots in (Fig. 9.3)). A gradual increase in the climatic suitability for PD in some river basins may thus favour the spread of the pathogen from coastal to interior areas of the continents, allowing interconnection between areas that would otherwise remain isolated. Coastal areas close to cool water masses may also undergo an increase in risk when using higher resolutions data, as exemplified in California (Fig. 9.3 e,f).



**Figure 9.4:** Impact of high-resolution climate data on the risk of Pierce’s Disease for grapevines worldwide. (A) Difference in risk indices in Europe, which accounts for the 96% of the points in the dataset. (B) Comparison of the risk indices derived from CHELSA and ERA5 datasets. Points with perfect agreement would lie in the solid black diagonal curve. (C) Histogram of risk indices derived from ERA5 (blue) and CHELSA (orange). (D) Histogram of the differences in risk indices between CHELSA and ERA5 datasets. (E) Changes in risk categories when using high-resolution climate data (CHELSA) with respect to mid-resolution data (ERA5). Risk category increase refers to changes from low to moderate risk or from moderate to high risk. Likewise, risk category decrease refers to changes from moderate to low risk or high to moderate risk.

Finally, to obtain a comprehensive assessment of the impact of microclimatic conditions

on the risk of PD establishment, we collated a dataset of over 100,000 *Vitis vinifera* presence locations worldwide from GBIF (179), with a predominant concentration of points from Europe (Online Supplementary Information). Each data point was assigned a risk index based on the ERA5 and CHELSA projections, respectively, using the nearest pixel from each database.

This approach revealed an increase in the risk indices associated with the vine locations (Fig. 9.4 A-D), mostly showing shifts towards higher risk indices (Fig. 9.4 A,E) from no risk to risk, or increases in risk category (low to moderate or moderate to high), while a negligible number of points decreased in risk category (Fig. 9.4 E). Such behaviour was common to all key viticulture regions studied, although the extent of increases differed between continents, with substantial expansion of vineyard areas at risk in Europe and South Africa (Table 9.2). Overall, our results emphasise the global relevance of microclimatic conditions in influencing the risk landscape for PD in viticultural areas (Table 9.2).

**Table 9.2:** Comparison of grapevine presence locations at risk in key viticulture regions using CHELSA and ERA5 datasets

	Nº points	risk CHELSA (%)	risk ERA5 (%)
<b>Europe</b>	96102	41.2	21.8
<b>USA</b>	792	69.8	66.3
<b>South Africa</b>	36	47.2	5.6
<b>South America</b>	112	77.7	74.1
<b>Australia</b>	186	51.6	45.7

## 9.3 Discussion

Our study sheds light on the relevance of the spatial scale of observation in the intricate interplay between microclimatic conditions and the risk of PD for grapevines on a global scale. The use of high-resolution climate data reveals previously unrecognised local areas with microclimates conducive to the establishment of PD worldwide. Contrary to the simplistic assumption that higher resolution data might yield only marginal distinctions at regional levels, our study demonstrates that slight variations in climate data at local scales can lead to a global surge in disease risk. These increases not only affect the spatial distribution of risk, but also its temporal dimension, as suggested by the rate of increase in the surface area at risk. In the case of PD, we show that this rate nearly doubles when high-resolution climate data is considered compared to previous estimates obtained with mid-resolution data. Thus, our findings indicate a critical need for the use of local or high-resolution climate data in the assessment of disease risk, especially in areas characterised by diverse topography and even when only attempting to global estimates.

Such observed differences arise from the non-linear nature of disease dynamics and the response of the pathosystem components to environmental shifts (147, 386). Therefore, models dependent on broader climate data may not capture the complexities of microclimates, resulting in an underestimation of disease risk. While this is not inherently negative, recognising these limitations helps to assume such risk estimates as a conservative lower bound until proven otherwise. Acknowledging these constraints is crucial for refining our understanding of disease dynamics and ensuring that our risk assessments are sufficiently cautious in the absence of more reliable data. Likewise, data coarsening procedures should be avoided, if possible, when modelling climate-driven disease dynamics, even in spite of computational efficiency. This recommendation applies not only to disease risk predictions but to all those in which non-linear functions depending on climate variables are present, such as species distribution models or phenological models (304).

Despite the valuable insights gained, our analysis heavily relies on the quality and resolution of the climate data from the CHELSA dataset (254). While this dataset offers information at a high spatial resolution, the temporal dimension is limited to a daily frequency, which forces

to apply an approximation to infer hourly data. Furthermore, the data may still be subject to biases or uncertainties inherent to the nature of the methodology employed in their construction. On the other hand, vector presence data is only accurately obtained for Europe, while an homogeneous presence is assumed in other viticulture areas. Additionally, the study primarily focuses on the effect of temperature conditions and the presence of potential vectors to determine the risk of *Xf* establishment, which may not encompass all possible contributing factors. Other variables, such as soil characteristics or vineyard management practices were not explicitly considered in this analysis, leaving room for additional complexities in the disease dynamics. Furthermore, the study predominantly examines the risk at a global scale, and the applicability of the findings to specific local contexts may vary.

Future research should aim to address the aforementioned limitations and provide a more comprehensive understanding of the multiple interactions influencing PD development in viticulture regions. Other factors influencing disease spread, such as human behaviour, land use changes, and ecological shifts, should also be explored, offering a more comprehensive and holistic view of the interplay between environmental conditions and disease vulnerability. The acceleration in the rate at which the risk of PD is growing calls for more research into control strategies to mitigate its impact on grapevine crops worldwide.

Although PD is currently restricted to North America and recently introduced in Taiwan (412), Mallorca (Balearic Islands, Spain) (210, 315) and Israel (457), since the mid-1990s climatic conditions are increasingly conducive to the establishment of PD in southern Europe (194). For example, with the increase in the resolution of climate data our model predicts the recent detection of PD in Portugal (288), which was not anticipated using the ERA5 data (194). In a short time, it is foreseeable that there will be more epidemic outbreaks in vineyards in southern Europe if the entry of infested plants is not controlled. This not necessarily have to be vines but can also include other plants such as almond trees or ornamental plants (316).

Overall, our study contributes to the growing body of knowledge on the impact of climate on agricultural pests and pathogens, emphasising the importance of considering microclimatic conditions for a more deep understanding of disease dynamics. Future research should focus on developing comprehensive models that integrate high-resolution climate data, considering both the global and local factors that influence disease dynamics. This holistic approach will enable a more accurate prediction of disease risk, allowing for the development of targeted management strategies and the enhancement of global food security.

## 9.4 Methods

### 9.4.1 Climate data

Climate data was downloaded from two datasets for our analysis: the ERA5 dataset (322, 324) and the CHELSA dataset (253, 254). ERA5 offers mid-resolution climate data with a spatial resolution of 10 km and hourly temporal resolution, while CHELSA provides high-resolution data with a spatial resolution of 1 km and daily temporal resolution. Both datasets exhibit global coverage and encompass crucial climate variables, such as temperature and precipitation. For our simulations, we used the mean hourly temperature data from ERA5 dataset and the maximum and minimum daily temperature data from CHELSA dataset.

### 9.4.2 Vector climatic suitability

Vector climatic suitability data was obtained from (207), in which a Generalised Additive Model (GAM) is employed to calibrate the relationship of *P. spumarius* global occurrence with moisture index and maximum temperatures during summer index estimated from 1979 to 2013 using the CHELSA dataset.

### 9.4.3 Vineyard data

To assess the risk of Pierce's Disease in locations where grapevines are present, we collected a comprehensive dataset of over 100,000 *Vitis vinifera* presence data records from the Global Biodiversity Information Facility (GBIF) (179, 447). We note that while the dataset spans the globe, 96% of the points are located in Europe (Online Supplementary Information).

#### 9.4.4 Climate-driven epidemiological model

We used the model developed in (194), which describes the initial exponential rise (or decrease) of infected plants at the onset of an epidemic based on the spatial distribution of the vector and the bacterial growth and survival processes mediated by temperature. The density of vectors at a given cell controls the number of new plants that will be inoculated with the bacterium, while the local temperature mediates the growth and survival processes of the in-plant bacterial population, leading the initial inoculation to an infection or not. These temperature-driven growth and survival processes are described with the accumulation of two metrics denoted *Modified Growing Degree Days* (MGDD) and *Cold Degree Days* (CDD). The base function to compute the MGDD is proportional to the  $X_f$  temperature-dependent growth rate and is defined by

$$f(T) = \begin{cases} 0 & \text{if } T < T_{\text{base}} \\ m_1 \cdot T - b_1 & \text{if } T_{\text{base}} \leq T < T_1 \\ m_2 \cdot T + b_2 & \text{if } T_1 \leq T < T_{\text{opt}} \\ m_3 + b_3 & \text{if } T_{\text{opt}} \leq T < T_2 \\ m_4 + b_4 & \text{if } T_2 \leq T < T_{\text{max}} \\ 0 & \text{if } T \geq T_{\text{max}} \end{cases}$$

where  $T_{\text{base}} = 12^\circ\text{C}$ ,  $T_1 = 18$ ,  $T_{\text{opt}} = 28^\circ\text{C}$ ,  $T_2 = 32$  and  $T_{\text{max}} = 35^\circ\text{C}$ ; the slopes are  $m_1 = 0.66$ ,  $m_2 = 1$ ,  $m_3 = -1.25$  and  $m_4 = -3$  and the intercepts are  $b_1 = -8$ ,  $b_2 = -14$ ,  $b_3 = 4$  and  $b_4 = 105$ . MGDD are then computed as

$$\text{MGDD}(t) = \frac{1}{24} \sum_{\tau \in t} f(T(\tau)), \quad (9.1)$$

where  $\tau$  is expressed in hours,  $t$  in years and we divide by 24 to obtain  $\text{MGDD}(t)$  in degree days. The accumulation period goes from the 1<sup>st</sup> of April to the 31<sup>st</sup> of October in the northern hemisphere and from the 1<sup>st</sup> of November to the 31<sup>st</sup> of March in the southern hemisphere.

CDD are computed between 1<sup>st</sup> November and 31<sup>st</sup> March in the northern hemisphere and between 1<sup>st</sup> April and 31<sup>st</sup> October in the southern hemisphere as

$$\text{CDD}(t) = \frac{1}{24} \sum_{\tau \in t} (6 - T(\tau)) \quad \text{for } T_i \leq 6^\circ\text{C}. \quad (9.2)$$

Altogether, the number of infected hosts is described by the following recurrence relation

$$I(t) = I(t-1) e^{\gamma(R_0-1)} \mathcal{F}(\text{MGDD}(t)) \mathcal{G}(\text{CDD}(t)), \quad (9.3)$$

where  $\gamma$  is the death rate of infected vines,  $R_0$  is the basic reproduction number of the disease and  $\mathcal{F}(\cdot)$  and  $\mathcal{G}(\cdot)$  are sigmoidal-like functions that relate the MGDD and CDD metrics to the probability of developing an infection from a given inoculation. Following (194),  $R_0$  in each cell  $j$  is related to the climatic suitability of the vector such that

$$R_0^j = R_0^* \cdot s_j = 5 \cdot s_j, \quad (9.4)$$

$\gamma = 0.2$  and the specific form of  $\mathcal{F}(\cdot)$  and  $\mathcal{G}(\cdot)$  is given by

$$\mathcal{F}(x) = \frac{1}{1 + e^{-0.012(x-975)}} \quad (9.5)$$

$$\mathcal{G}(x) = \frac{2 \cdot 10^7}{2 \cdot 10^7 + x^3} \quad (9.6)$$

Finally, the risk index is derived as the effective growth rate of the infected population over the simulated time (194),

$$r_j = \max \left\{ -1, \frac{\ln(I_j(T)/I(0))}{\gamma(R_0^j - 1) \cdot T} \right\}. \quad (9.7)$$

Because the typical time scale of the disease is 5 years ( $1/\gamma$ ) (9), we simulate periods of 7 years. If more years are available to simulate, we perform a re-introduction of the disease as a single infected plant in each cell after each 7-year period (194).

The code used to run the model is freely accessible at GitHub (189).

### 9.4.5 Model adaptation to daily temperature data

MGDD and CDD metrics were originally defined using hourly temperature data (Eqs. (9.1) and (9.2)) (194). However, the CHELSA dataset only provide daily granularity. To overcome this limitation, we use a basic sinusoidal extrapolation relating maximum and minimum daily temperature to hourly temperatures,

$$T_h = \frac{T_{max} + T_{min}}{2} + \frac{T_{max} - T_{min}}{2} \sin(w \cdot h), \quad (9.8)$$

with  $w = 2\pi/24$  and  $h$  ranging from 0 to 23. This approximation was validated in (200) with data from national meteorological stations in Spain (AEMET) using several locations and years, showing no differences between the use of hourly or daily temperatures to estimate MGDD and CDD. Similarly, the use of the approximation was validated across Europe using EURO-CORDEX data.

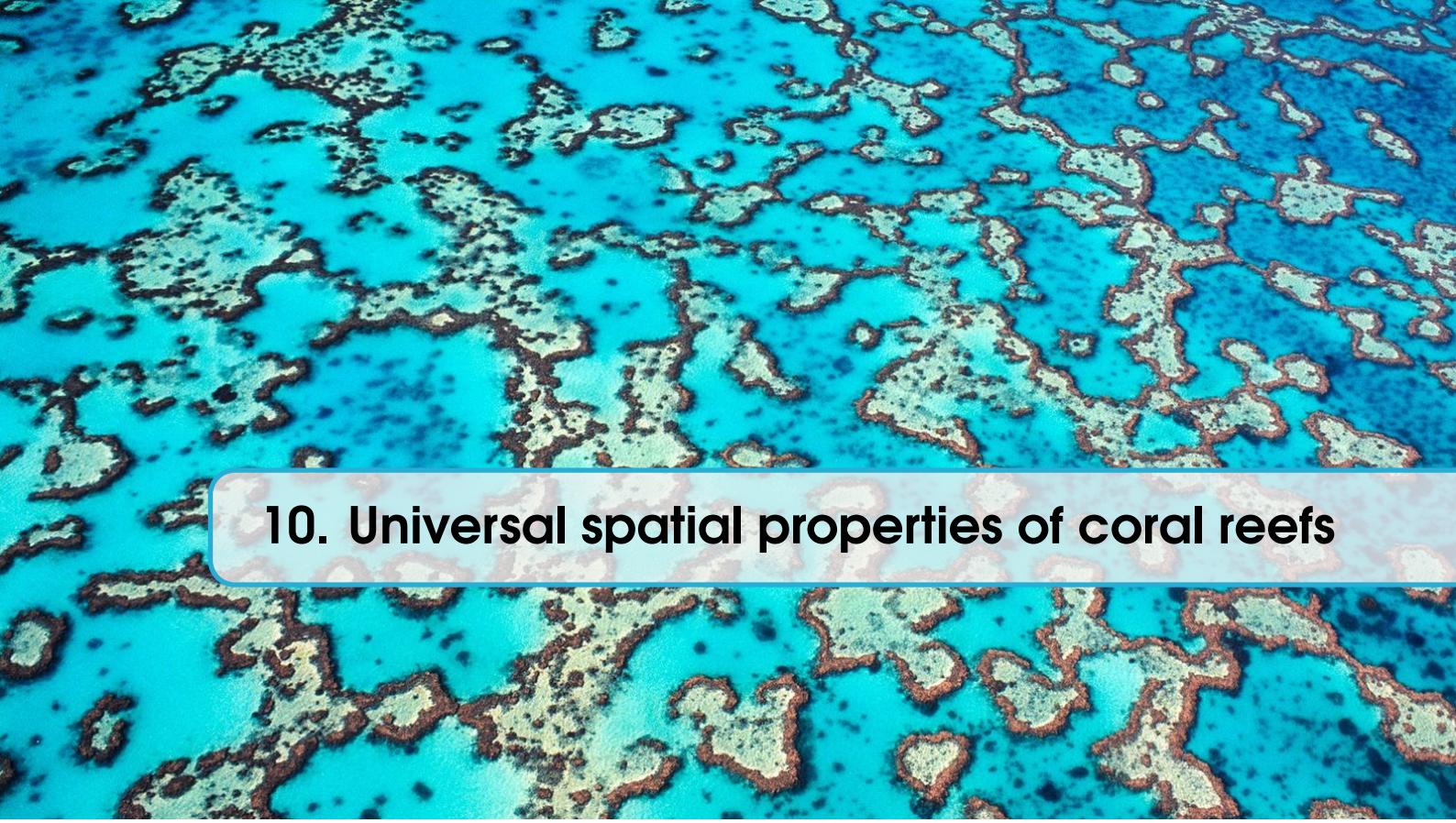




# Data-driven methods for global ecological problems

<b>10</b>	<b>Universal spatial properties of coral reefs . . . . .</b>	<b>123</b>
10.1	Introduction . . . . .	124
10.2	Results . . . . .	124
10.3	Discussion . . . . .	129
10.4	Methods . . . . .	130
<b>11</b>	<b>Reconstructing coastal pH time-series with machine learning . . . . .</b>	<b>133</b>
11.1	Introduction . . . . .	134
11.2	Results . . . . .	136
11.3	Discussion . . . . .	139
11.4	Methods . . . . .	141
<b>12</b>	<b>Mapping the distribution of seagrass meadows from space with deep convolutional neural networks . . . . .</b>	<b>145</b>





## 10. Universal spatial properties of coral reefs

Àlex Giménez-Romero<sup>1</sup>, Manuel A. Matías<sup>1</sup>, Carlos M. Duarte<sup>2</sup>

1. Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Palma de Mallorca 07122, Spain
2. Tragsa, Passatge Cala Figuera 6, 07009 Palma de Mallorca, Spain

Published as

## 10.1 Introduction

Coral reefs form some of the largest biogenic structures in the biosphere (451), and are prevalent structures across tropical coastal waters. Coral reefs often form complex, labyrinthine structures that render sailing in tropical waters challenging, but protect the shorelines of tropical coastal nations while supporting biodiversity and providing food supply supporting local communities (310). The formation of coral reefs has intrigued scientists for a long time, with Darwin formulating a model based on coral reefs accreting on volcanic structures (126). Darwin's model continues to stir discussion (144), as it would represent a class of coral reefs, at best, among the broad range of configurations and possible origins (393).

Coral reefs can form isolated oval or ring-shaped structures (e.g. coral atolls), linear structures parallel to the shoreline (fringing or barrier reefs) or convoluted or highly-branched structures (e.g. (361)), and often present nested structures, such as smaller reefs within large coral reef lagoons (58). Whereas previous studies have examined the geometry of coral reefs (16, 57, 58, 361, 408, 456), finding evidences of fractality both at individual reef (58, 408) or regional scales (361, 456), a global assessment of coral reef size and geometry was hitherto precluded by the lack of data on reef form and size at the global scale. Resolving the size and geometry of coral reefs has become, however, a matter of urgency, as coral reefs are rapidly declining, with an estimated 50% of coral cover lost globally from 1957-2007 (148). The IPCC projects that 70% to 99% of coral cover may be lost if climate change reaches 1.5 to 2.0 °C above pre-industrial levels, respectively (45). Yet, the Kunming-Montreal Biodiversity Framework (266) calls for stopping all biodiversity losses and restore 30% of degraded habitats, including coral reefs, by 2030, which requires action at individual reefs and, therefore, an understanding of coral reef form and size as an underpinning for the necessary conservation action.

The release of the Allen Coral Atlas (ACA) (6), a worldwide mapping initiative that provides benthic habitat data of shallow-water (above 10 m deep) tropical reefs, presents an unprecedented opportunity to characterize the form and size of coral reefs on a global scale. In contrast to previous efforts, such as NOAA's Coral Reef Information System (335), Khaled bin Sultan Living Oceans Foundation (85) or Millennium Coral Reefs (426), the ACA dataset combines an extensive global coverage of reef areas with the direct availability of processed habitat data obtained from recent high-resolution (3 m) satellite imagery using deep learning models. This invaluable resource offers a unique vantage point for researchers, scientists, and conservationists to explore and understand coral reef ecosystems in ways that were previously challenging or impossible. This requires processing the classified images to extract a canonical inventory of individual coral reefs that can be used to derive macroecological laws and principles from the patterns and trends observed across coral reefs. This, in turn, enables to advance and consolidate our understanding of these complex ecosystems and the scale of the effort required to conserve and restore them.

Here we report universal macroecological patterns (68) in shallow-water, tropical coral reef size and geometry using data from the 2022 release of the ACA (6). Specifically, we produced a global inventory of individual reefs (193), allowing the analysis of the size and inter-reef distance distributions of individual coral reef areas within each coral province of the Atlas. We then examined the relationship between reef size, shape and fractal geometry of coral reefs.

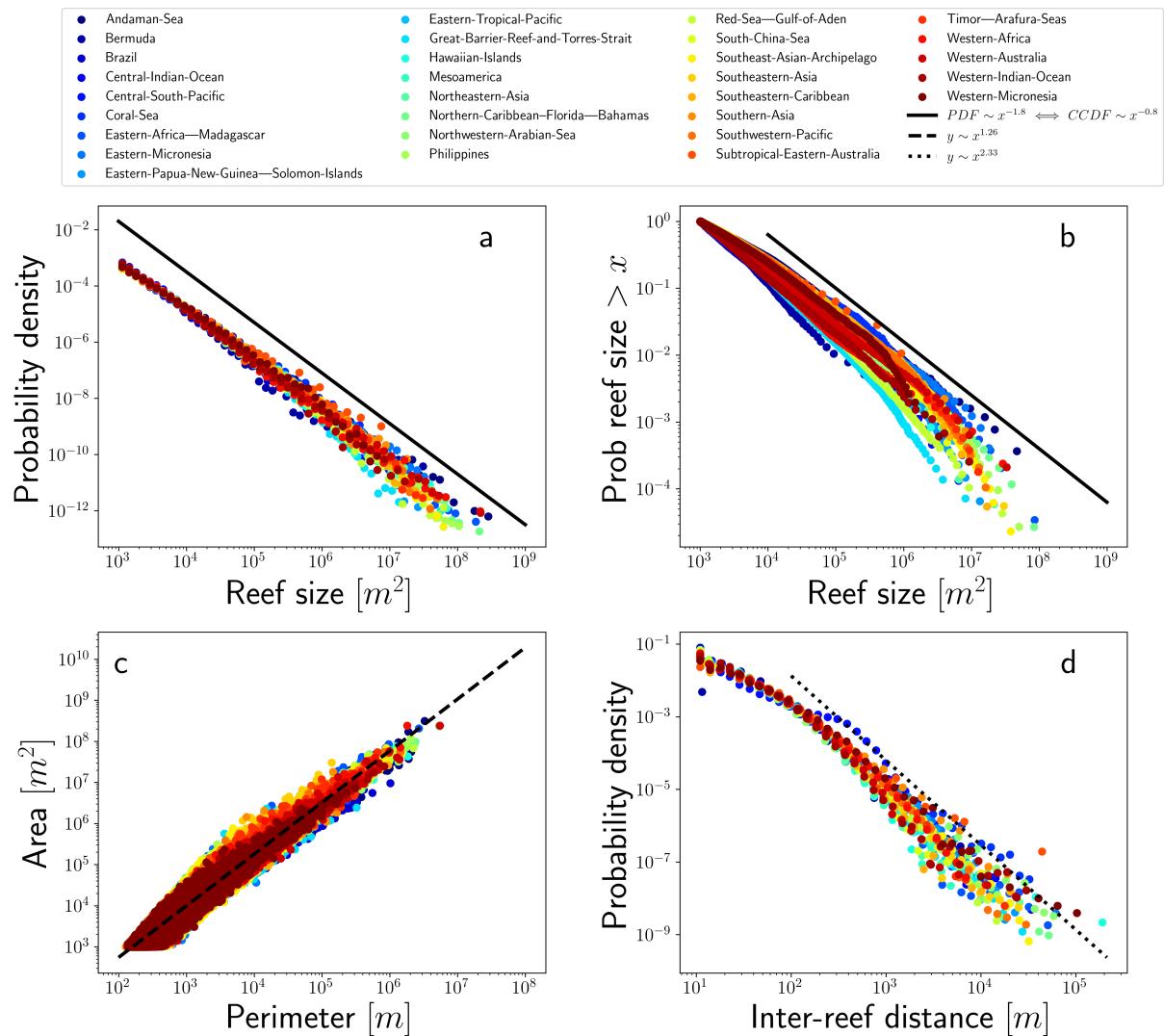
## 10.2 Results

### 10.2.1 Coral reefs macroecological patterns

The area of all coral reefs within each province of the ACA was computed after processing the data to identify individual reefs (see Methods). We identified a total of 1,579,772 individual shallow-water tropical reefs, extending over a total of 52,423 km<sup>2</sup> of ocean area. The canonical nature of this openly available dataset (see Methods), which includes all shallow-water tropical coral reefs worldwide, allows the mean and median size of individual reefs to be estimated at 3.32 ha and 0.3 ha, respectively, across all coral reef provinces in the Atlas (see Online Supplementary Information for statistics in each province).

Our analysis reveals that the size distribution of the area of coral reefs in all provinces converge to conform a power-law distribution,  $y \sim x^{-\alpha}$ , that holds over multiple scales, where  $y$  reflects the

probability of occurrence of reefs of area class  $x \text{ km}^2$  (Fig. 10.1 a-b). This behavior was consistent across all provinces within a range of 3 to 5 decades in coral reef area, yielding an average exponent of  $\langle \alpha \rangle = 1.84$  (95% CI: 1.55 to 2.12) (see Methods). The global size-frequency distribution of coral reefs, fitted to all data independently on the province, conforms to a power-law with an exponent of  $\alpha = 1.8$ , consistent with the mean exponent  $\langle \alpha \rangle$ , providing evidence for a universal scaling law governing the size distribution of coral reefs. This suggests that the observed distribution of reef areas arises mainly due to biological processes contributing to reef formation, common among the provinces. Local geophysical processes interacting with coral reef growth, which might vary among provinces, are therefore likely to have a limited role, provided the universality of the power-law scaling of the reef size distribution.



**Figure 10.1: Macroecological patterns of global coral reef size, geometry and spacing.** (a) Size distribution. The black line corresponds to a fitted power-law of exponent 1.8. (b) Complementary cumulative distribution function (CCDF) with corresponding exponent of 1.8. (c) Area-Perimeter relation. The black dashed line corresponds to a fitted power-law with exponent 1.23. (d) Inter-reef distance distribution. The black dotted line corresponds to a fitted power-law to the distribution tail with a exponent of 2.33.

The presence of a power-law size distribution also suggests that the object studied may be fractal in nature (319, 352, 394, 406, 437), at least along a certain range. A simple way to test whether coral reefs are fractals is using the area-perimeter relation (289, 292), a method of fractal analysis that characterizes the complexity of irregular shapes by examining the relationship between their area and perimeter (see Methods). The scaling of coral reef area to perimeter of

all individual coral reefs also converges into a single power-law with an exponent of  $\alpha = 1.2578$  (95% CI: 1.2573 to 1.2583), indicating again a universal behavior (Fig. 10.1 c). When the relationship is fitted for each province independently, we found a mean exponent of  $\langle \alpha \rangle = 1.2574$  (95% CI: 1.1757 to 1.3391), practically identical to the general exponent. The exponent for the area-perimeter relationship is significantly lower than the value of 2 that would be found for a smooth euclidean geometry. According to fractal analysis, the fractal dimensions of the perimeter and area of coral reefs are given by  $D_P = 1.2950$  and  $D_A = 1.6289$  (see Methods), respectively, which are clearly different than the putative euclidean dimensions of  $D_P = 1$  and  $D_A = 2$ . Taking each province independently, these fractal dimensions would vary between  $D_P^{\max} = 1.3506$ ,  $D_P^{\min} = 1.2468$  and  $D_A^{\max} = 1.6696$  and  $D_A^{\min} = 1.5879$  within a 95% confidence interval, further ensuring that the obtained fractal dimensions are significantly different from the euclidean geometry. Coral reefs develop fractal-like geometries, exhibiting complex, self-similar structures across different scales. This feature might arise from the complex physical and ecological processes that shape the distribution and growth of coral reefs. However, how this occurs remains largely unknown, as we lack mechanistic models able to generate coral reef landscapes across scales and time that can be challenged to reproduce these patterns.

The spatial distribution of coral reefs within each province was also investigated by means of the inter-reef distance, defined as the minimum distance between a reef and its nearest neighbor. We find a heavy-tailed relation where the tail conforms to a power-law with an exponent of 2.33 (Fig. 10.1 d). This reveals that most of the reefs are close to each other, while a non-negligible number of them are isolated. This finding is again mostly independent on the geographic location of the analyzed coral reefs, arising as a universal property of coral reef provinces.

## 10.2.2 The fractal nature of coral reefs

We computed the fractal dimension of the area and perimeter of each individual reef from all provinces using the well-known box-counting algorithm (see Methods). The mean values for the fractal dimension of the perimeter,  $D_P = 1.24$  (95% CI: 1.13 to 1.35) and the fractal dimension of the area,  $D_A = 1.60$  (95% CI: 1.39 to 1.81), are well defined and consistent with those obtained from the area-perimeter relationship (Fig. 10.2 a,b). The fractal dimension for the reef areas is quite stable around the mean, although it increases slightly for large coral reefs (Fig. 10.2 a). On the other hand, the fractal dimension for the perimeter shows a more pronounced increase as function of reef size. This indicates that as coral reefs grow, their contour gets more and more convoluted, increasing its complexity, while its surface remains geometrically more stable.

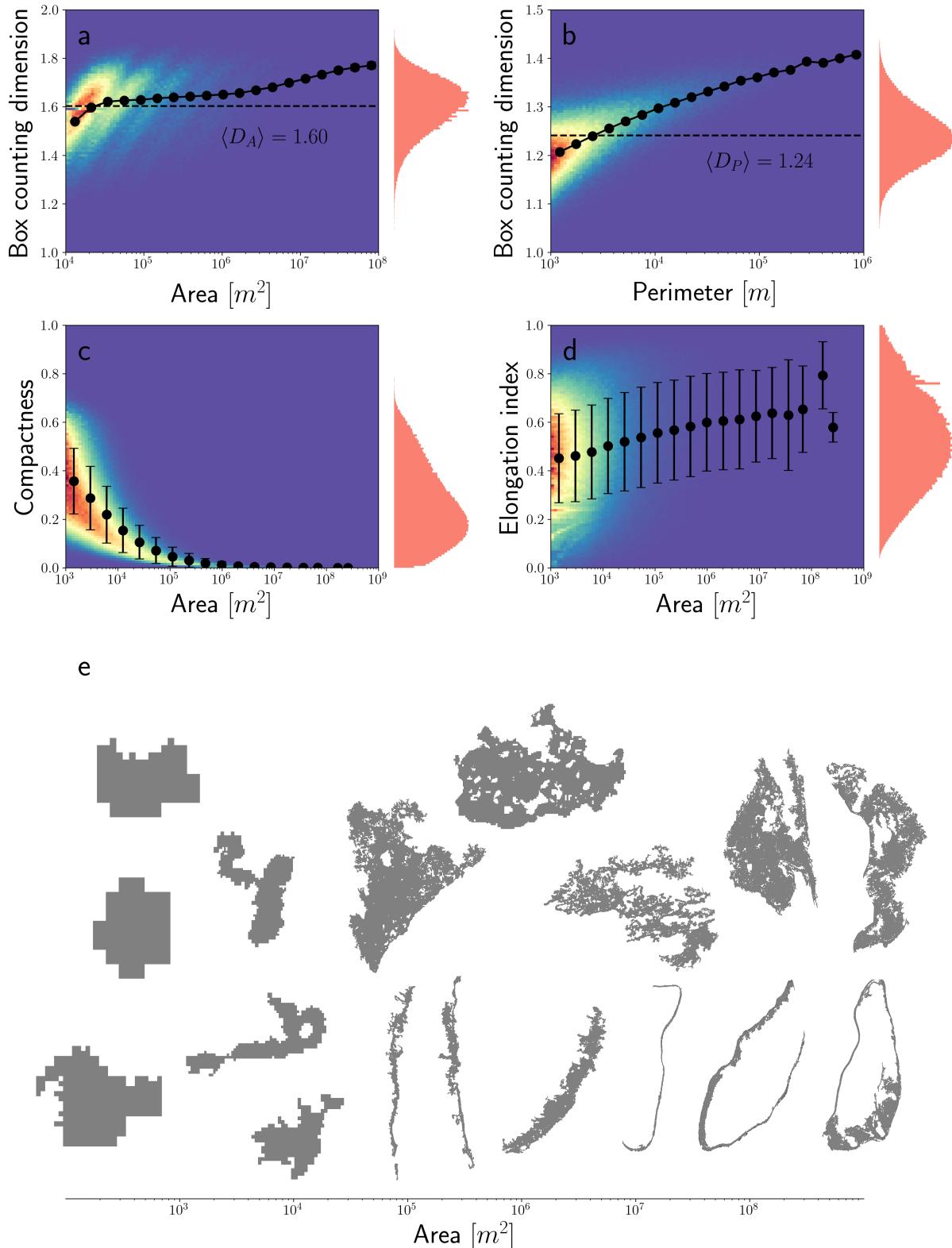
To further understand the reef formation process from a geometric perspective, we computed other shape measurements such as compactness and elongation indices (see Methods). We observe that the compactness of coral reefs decreases rapidly with increasing size (Fig. 10.2 c). Two changes of shape are consistent with this result: transitioning from round to elongated shapes or keeping the rounded shape while developing holes, reef lagoons, within their surface. These two processes are not necessarily mutually exclusive, as elongated shapes could appear from evolution of empty rounded shapes. The results obtained for the elongation index measurement show that both processes must occur, as reefs elongation increase with size while showing very high variance. This hypothesis can be contrasted by examining the different shapes that coral reefs have formed (Fig. 10.2 e).

Altogether, our analysis indicate that coral reefs evolve from simple rounded filled shapes (high compactness and low elongation index) to more complex elongated and less compact forms (low compactness and high elongation index), giving rise to fractal objects with stable surface fractal dimension and increasing perimeter fractal dimension as they grow (Fig. 10.2 e).

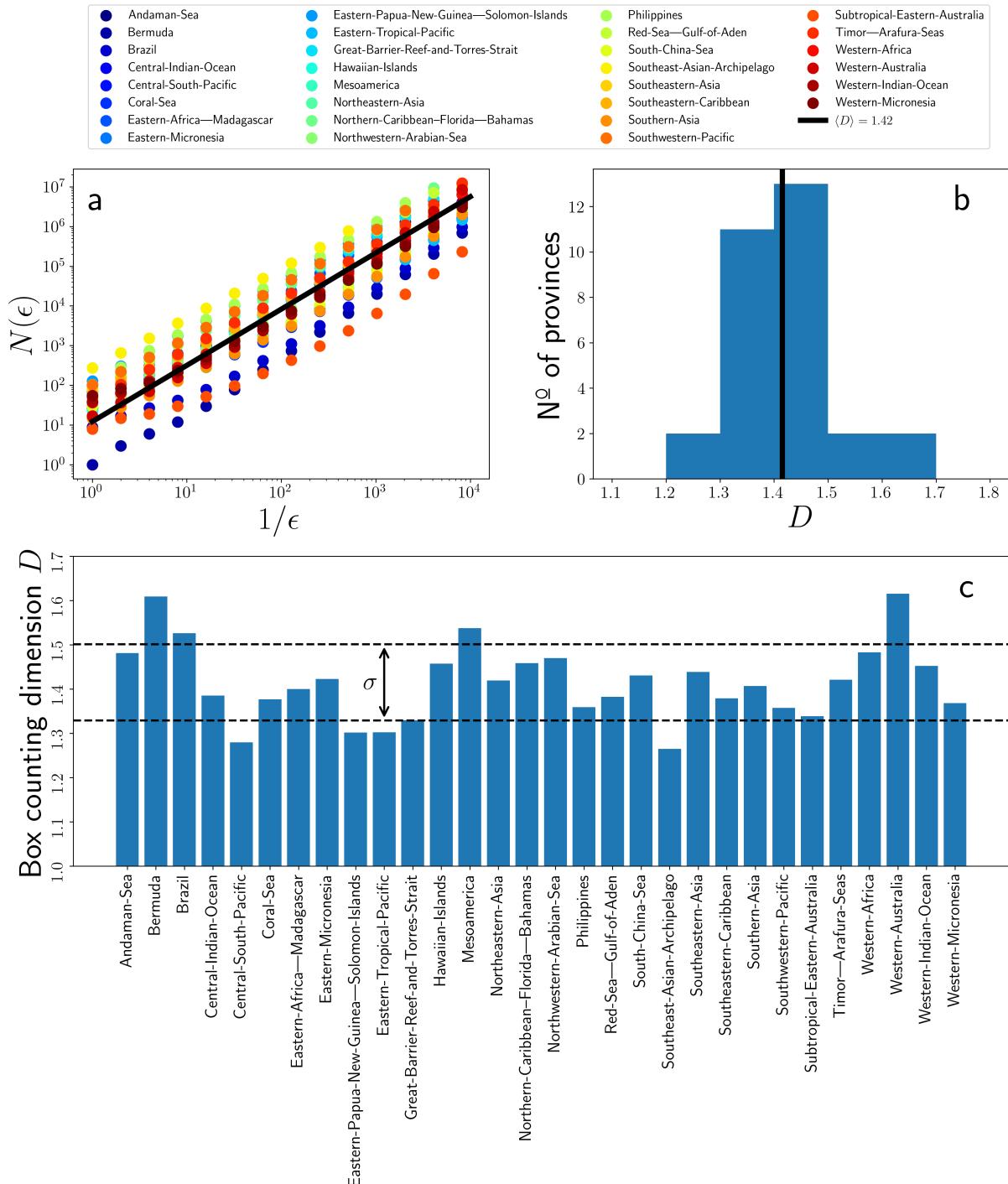
## 10.2.3 Fractality extends up to coral provinces

The fractal dimension of coral reefs varies across a range of spatial scales, spanning from individual colonies to entire reef systems (183). This variability arises from the different physical, biological, and ecological processes that come into play during the development of the different structures present at each organization level. The different coral provinces can be understood as the largest organizational level of these organisms, and the processes involved in maintaining such big structures should be different from that of individual reefs, giving rise to a

different fractal dimension.



**Figure 10.2: The fractal nature of global coral reefs.** 2D histograms from all shallow-water coral reefs worldwide for: (a) the surface fractal dimension and area; (b) the perimeter fractal dimension and perimeter; (c) the compactness and area and (d) the elongation index and area. The black line corresponds to the mean values of the Y axis measure as function of the X axis measure. The red histogram corresponds to the distribution of the Y axis measure. (e) Example of coral reefs shape as function of their surface.



**Figure 10.3:** Box Counting Dimension of coral reef provinces surface. (a) Scaling of the measure (number of boxes of length  $\epsilon$ ,  $N(\epsilon)$ ) as function of the ruler used (box length  $\epsilon$ ). The slope of the fit for each province corresponds to its Box-Counting surface fractal dimension,  $D$ . (b) Histogram of the obtained Box-Counting fractal dimensions. The solid black line represents the mean Box-Counting fractal dimension,  $\langle D \rangle$ . (c) Box-counting surface dimension for each coral province. Black dashed lines correspond to a  $1\sigma$  deviation from the mean.

To investigate this hypothesis, we computed the surface fractal dimension for each coral province as a whole using the well-known box-counting algorithm (Methods, Online Supplementary Information). We found a mean fractal dimension of  $\langle D \rangle = 1.42$  (95% CI: 1.24 to 1.59), consistent across the different coral provinces assuming a normal distribution (Fig. 10.3). The fractal dimension of coral reef landscapes is similar to those expected from sizes expanding along a Fibonacci series, which yield a fractal dimension of 1.44 (405). This suggests, again, that

coral reefs are self-organized systems that exhibit similar patterns of complexity and irregularity at different scales, and that the fractality is an intrinsic property of coral reefs that likely arise from the underlying biological, physical, and ecological processes involved in their growth dynamics.

## 10.3 Discussion

Coral reefs self-organize to form macroecological patterns that are largely independent of the geographical location of the reefs, suggesting that the specific physical conditions to which each coral province is subject have little influence on the development of these scaling laws. Coral reef geometries conform fractal structures that follow a power-law size distribution, with many reefs close to each other while others are completely isolated across coral reef provinces. Overall, our findings provide strong evidence that the power-law size distribution of coral reefs and their fractal geometries are fundamental features of these ecosystems, which reflect the underlying ecological and physical processes that shape them across scales. Furthermore, because the exponent of the power-law size distribution and the fractal dimensions are consistent across different regions, these features are not only fundamental, but universal laws reflecting coral reef growth processes.

These universal scaling properties must be used to challenge mechanistic models aimed at reflecting coral reef growth, which hitherto lacked such constraints, both as individual reefs as well as coral reef provinces. For example, at the scale of coral reef province, the characteristic spacing (inter-reef distance distribution) likely influences the interaction between coral reefs, hydrodynamic flows, and the dynamics of the limiting nutrients transported to support photosynthesis and calcification, further constrained by sea level change and available vertical accommodation space (53, 309, 329). Realistic models of coral reef growth and dynamics should be able to reproduce the universal features described here, such as the power law in coral reef size distribution, the fractal geometries and the changes in reef shape with size.

Power-law distributions have been also found in many natural systems (69, 96, 117, 296, 404, 446). Different mechanisms can produce such emergent power-laws (295): Self-Organized Criticality, in which the system evolves naturally towards a critical state (27); Highly Optimized Tolerance, in which the system evolves following a trade-off between yield, cost of resources and tolerance to risk (83, 84), or correlated noise models, in which external drives and internal dynamics compete on similar time scales, yielding a non-critical steady state characterized by heavy-tailed distributions (331). Ecological power-laws can arise through a combination of several ecological and physical processes, including competition for resources, random disturbances leading to failure or loss. In (258, 384) it was shown that the emergence of power-laws in vegetation patterns requires the interplay between global competitive and locally facilitative interactions (258, 384). These power-law distributions are different than those obtained in classical critical systems, where power laws occur exclusively at the transition point (46, 453). Actually, the ecological power-law reported in (258) only occurs far enough from the (true) critical transition to extinction. It has been conjectured that living systems exhibiting this behavior could draw important functional advantages from operating close to an emergent critical point, namely an optimal balance between robustness and flexibility (321). In the case of corals, this could reflect a balance among self-organization dynamics driven by competition for space and resources with other species, environmental factors like ocean currents and water quality, predation interactions within the reef ecosystem and local facilitative interactions, such as energy dissipation. The combination of these mechanisms should be explored in developing models of coral reef formation.

Fractal models have been also used to investigate universal principles that govern the structure and dynamics of complex ecological systems (69), such as the geometry of benthic ecosystems or power-law scaling (388). In coral reef ecosystems, fractal geometry might emerge as an efficient structure for nutrient acquisition (408). The fact that  $D_P > 1$  implies that the reef contour is highly convoluted, while  $D_A < 2$  is the result of the development of multiple holes in the reef surface. The combination of these mechanisms maximizes the chance that coral individuals living at the reef surface are in contact with an external flow, thus being able to obtain nutrients. According to (183), coral colonies with a larger space-filling surface and smaller perimeters increase energy gain while reducing the exposure to competitors. A similar argument could apply in the case of coral reefs, which can be included in modeling approaches.

The shape of coral reefs varies with size, and thereby during growth, from more compact, circular structures at small size, observed at the mean area of 3.32 ha, to increasingly elongated structures, which may break the closed shape to form long, linear, fringing reefs. The change in geometry from circular to fringing has been postulated to result mostly from reduced accommodation space as coral reefs grow, and the interaction with sea level (259), but has also been explained as resulting from the interaction between reefs and nutrients transported along hydrodynamic flows from a prevalent direction (309). Reefs inside reef structures can be also observed, suggesting that as coral reefs expand in size, smaller reefs appear inside them, starting as compact coral heads to then develop empty inner spaces. Overall, this phenomenology suggests that a Turing instability (119, 425) might be present, which indeed has been previously suggested (309), arising due to the interplay between diffusion of nutrient species and nutrient uptake and recycling processes within reefs. However the Turing mechanism would yield a normal distribution of inter-reef distances, not compatible with the heavy-tailed inter-reef distance distribution found in this study neither with the obtained power-law scaling.

Overall, our findings have important implications for understanding of the structure and function of coral reefs, as well as for their conservation and management. The macroecological characterization of universal laws in the geometry of coral reefs, along with the dataset of unique coral reefs globally, which are reported here for the first time, should help design effective coral reef restoration projects as well as optimize and quantify the effort and resources required.

## 10.4 Methods

### 10.4.1 Global coral reef data

Global-scale coral reef benthic data were obtained from the Allen Coral Atlas (ACA) (6), a publicly available dataset of high-resolution satellite imagery and machine learning-based coral reef classifications. We downloaded the data from the ACA website, which is already divided into the different coral reef provinces. The downloaded dataset consists of GeoJSON files for each coral province with several Polygons and Multipolygons forming the different benthic classes, from which we selected the “coral/algae” class (see (6)). Despite the data is already provided in vector format, the reefs are not identified as individual entities, i.e. a single reef can be formed by many polygons or multipolygon objects. Thus, we processed the dataset with the methods explained below to obtain a representation of individual reefs.

### 10.4.2 Coral reefs as clusters of connected coral/algae class polygons

A label assignment algorithm was developed to identify the different independent (not connected) components forming the coral reefs. Basically, we followed an iterative process in which connected components were assigned the same label, thus being identified as forming the same component. Polygons were considered to be connected if they intersected. To efficiently compute the intersections among polygons we used the Sort-Tile-Recursive algorithm (278) implemented in Python Shapely library (185). The implementation of the algorithm can be found in the Preprocessing.py module at (190).

Coral reefs of less than  $10^3 \text{ m}^2$  were considered as possible noise in the dataset, and thus disregarded. We made this choice based in the fact that the ACA is obtained from satellite imagery of about 3 m resolution. Thus, coral reefs of less than this area would represent less than 100 pixels.

### 10.4.3 Coral reefs area, perimeter and inter-reef distance

We computed coral reef area and perimeter using the `geopandas.GeoSeries.area` and `geopandas.GeoSeries.length` methods in geopandas Python’s library (52). For each coral reef, we defined the inter-reef distance as the distance to its nearest neighbor. Thus, the inter-reef distance distribution is obtained after obtaining the nearest neighbor to each reef and computing that distance. To make this computation efficient, we used the Sort-Tile-Recursive algorithm (278) implemented in Python Shapely library (185). The implementation of the algorithm can be found in (190).

We note that our estimates of the total area for each coral reef provinces (Online Supple-

mentary Information) slightly differ from that directly provided by the ACA because we removed “reefs” smaller than  $10^3 \text{ m}^2$ . Of course, if the area is computed before this data cleaning step the results are identical.

#### 10.4.4 Coral reef size distribution

We fitted the coral reef size data using the powerlaw package in Python (13, 103). We performed goodness-of-fit tests using a range of alternative distribution models, including log-normal, exponential, and stretched exponential distributions. We found that the power-law distribution (including its truncated form) provided a significantly better fit to the data than any of the alternative models with  $x_{\min}$  ranging from  $10^3 \text{ m}^2$  to  $10^4 \text{ m}^2$  (Online Supplementary Information).

#### 10.4.5 Fractal dimensions from area-perimeter relation

The area of regular objects such as squares or circles scale as the square of the perimeter  $A \sim P^2$ , while the area of irregular fractal objects scale more generally as  $A \sim P^\sigma$ , where  $\sigma = D_A/D_P$  with  $D_A$  and  $D_P$  being the fractal dimension of the area and the perimeter, respectively (97, 292). These fractals dimensions can be easily computed from the area-perimeter scaling exponent,  $\sigma$ , as  $D_P = (2 + \sigma)/2\sigma$  and  $D_A = (2 + \sigma)/2$  (97).

#### 10.4.6 Box-Counting fractal dimension

We computed the Box-Counting fractal dimension of all mapped areas following a box-counting algorithm (292). Briefly, the method computes the number of boxes of length  $\varepsilon$ ,  $N(\varepsilon)$ , needed to cover the underlying object. Then, the fractal dimension is simply defined as,

$$D = \lim_{\varepsilon \rightarrow 0} \frac{\ln N(\varepsilon)}{\ln 1/\varepsilon}. \quad (10.1)$$

In practice, the mathematical limit  $\varepsilon \rightarrow 0$  is unreachable and the fractal dimension is computed from the slope obtained in the plot of  $\ln N(\varepsilon)$  versus  $\ln(1/\varepsilon)$ . To efficiently compute the number of overlapping boxes we used the Sort-Tile-Recursive algorithm (278) implemented in Python Shapely library (185). The implementation of the algorithm can be found in (190).

#### 10.4.7 Compactness and elongation index

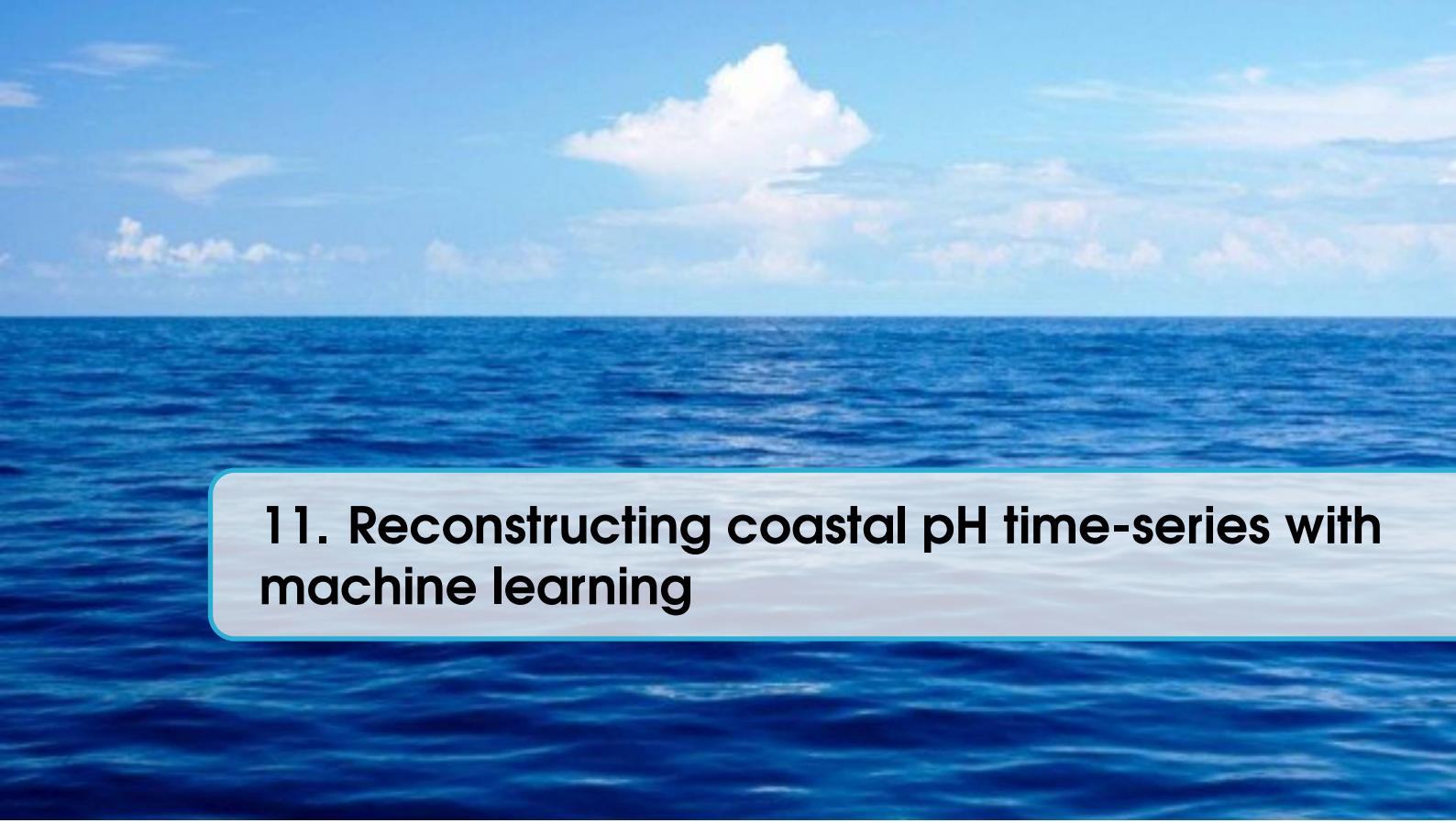
The compactness measurement is defined as the isoperimetric quotient,

$$C = \frac{4\pi A}{P^2}, \quad (10.2)$$

where  $A$  and  $P$  are the area and perimeter of the object under study, respectively.

The elongation index is defined as the Flaherty & Crumplin (1992) length-width measure, stated as measure LW\_7 in (14) and implemented in PySAL Python’s library (365).





## 11. Reconstructing coastal pH time-series with machine learning

Susana Flecha<sup>1,2</sup>, Àlex Giménez-Romero<sup>3</sup>, Joaquín Tintoré<sup>2,4</sup>, Fiz F. Pérez<sup>5</sup>, Eva Alou-Font<sup>1,4</sup>,  
Manuel A. Matías<sup>3</sup>, Iris E. Hendriks<sup>2</sup>

1. Instituto de Ciencias Marinas de Andalucía (ICMAN-CSIC), Polígono Río San Pedro s/n, 11519 Cádiz, Puerto Real, Spain.
2. Instituto Mediterráneo de Estudios Avanzados, IMEDEA (CSIC-UIB), E-07190 Esporles, Mallorca, Spain.
3. Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Palma de Mallorca 07122, Spain.
4. Balearic Islands Coastal Observing and Forecasting System (SOCIB), Parc Bit, Naorte, Bloc A 2op. pta. 3, 07121 Palma, Spain
5. Instituto de Investigaciones Marinas (IIM-CSIC), Eduardo Cabello 6, 36208 Vigo, Spain.

### Published as

Susana Flecha et al. "pH trends and seasonal cycle in the coastal Balearic Sea reconstructed through machine learning". In: *Scientific Reports* 12.1 (July 2022), page 12956. ISSN: 2045-2322. DOI: [10.1038/s41598-022-17253-5](https://doi.org/10.1038/s41598-022-17253-5). URL: <https://doi.org/10.1038/s41598-022-17253-5>

## 11.1 Introduction

Atmospheric carbon dioxide ( $\text{CO}_2$ ) emissions are exponentially increasing since the industrial revolution, principally due to fossil fuel use, industry and land-use change. Around a 46% of this  $\text{CO}_2$  remains in the atmosphere while the rest is captured by natural compartments: the terrestrial biosphere and the ocean(168). At present, the oceans have absorbed around an estimated 26% of the total anthropogenic  $\text{CO}_2$  released from 2011 to 2020 (168). Once  $\text{CO}_2$  dissolves in seawater, a sequence of chemical reactions occurs that derives in an increase of ( $\text{H}^+$ ) ions, which results in a decrease in seawater pH. This process, a consequence of increasing atmospheric  $\text{CO}_2$ , is termed Ocean Acidification (OA)(77). In addition to the pH decrease, ( $\text{H}^+$ ) ions react with carbonate ions ( $\text{CO}_3^{2-}$ ) to form ( $\text{HCO}_3^-$ ), leading to a reduction of the ( $\text{CO}_3^{2-}$ ) ion levels (142). Low carbonate levels affect the saturation state of calcium carbonate minerals, increasing difficulties in shell-forming for calcifying marine organisms (e.g., plankton, mollusks, echinoderms and corals). Consequences of OA are an important threat to marine ecosystems visible in higher levels of the trophic chain, with complex and wide-ranging impacts on the physiology of different species and therefore on their metabolism(265, 333). These metabolic effects will have numerous consequences at an organism scale, in particular, they can cause a decrease in growth, locomotion, reproductive capacity and homeostasis if they are not capable to control the conditions for calcification(226). Negative effects of this magnitude could cause an unexpected cascade effect impacting on the structure and functions of ecosystems and trophic networks(460) and cannot be easily generalized.

Also, ocean  $\text{CO}_2$  uptake and derived OA are not homogeneous at the global scale, with some areas more affected. For instance, the Mediterranean Basin is an area where effects are stronger compared to the global ocean(201). The Mediterranean Sea, constituting only a 0.82% of the surface and 0.32% of the volume of the global ocean, is cataloged as one of the most complex marine ecosystems, defined as a “miniature ocean”(39), inhabited by an extensive and diverse biota that represents between 4 and 18 % of the world’s total marine species (41) and serves as a model(39) to anticipate the responses of the global ocean to different types of pressures. It has been also defined as a climate change “hot spot” (201), with OA and its derived consequences characterized as one of the climatic threats with the greatest potential impact, followed by the temperature and UV radiation increase (307). The temperature rise in this semi-enclosed sea is expected to be two to four-fold times higher than that in the global ocean (432, 433). In addition, the sixth assessment report (AR6) of the IPPC7, places a high level of confidence on the increase in frequency of heatwaves and ongoing ocean acidification(298). Recent studies have confirmed that there is a trend of around 0.34 °C warming per decade in the Mediterranean Outflow Water (MOW) through the Strait of Gibraltar towards the Atlantic Ocean (174), associated with decreasing values of pH. Furthermore, in the Mediterranean Sea, due to its biogeochemical and hydrodynamic characteristics, such as the high alkalinity of its waters and the active thermohaline circulation (15), there is a larger absorption of atmospheric  $\text{CO}_2$  and an intense transport of this  $\text{CO}_2$  from the oceanic surface to deep areas(223, 343), already observed in the MOW (161, 162), with estimated OA trends of -0.0044 pH units per year in the Strait of Gibraltar (161) and ranging from -0.0017 to -0.003 in the Mediterranean Basin(251, 455). The Mediterranean Sea has an extensive coastline, which extends for 46 000 km and is shared by 21 countries (149). Coastal zones, as transitional areas, are inherently complex systems due to the strong biogeochemical-physical coupling, occurring relevant biogeochemical exchanges. Interactions in coastal areas involve terrestrial inputs of nutrients and particulate matter from river runoff and groundwater discharges, oceanic forcing (waves, tides, and currents), and atmospheric exchange of aerosols and trace gases, all of them which are influenced by the intense human activity in the coastline(121). Hence, processes related to the carbon system in coastal areas are more dynamic and complex than in the open ocean (50), and the range of pH change between -0.023 and 0.023 pH units per year(86) is therefore ~ 35 times larger than in the open ocean with -0.0013 to -0.0026 pH units yr<sup>-1</sup> (29). In particular, anthropogenic  $\text{CO}_2$  inputs appear to play a minor role compared to other sources of variability in coastal zones(232). Therefore, it is difficult to foresee how the pH conditions in the coastal areas in the year 2100 will differ from the present, due to the lack of knowledge on precise current pH values in the different coastal ecosystems and their variability obtained from long time series. Carbonate chemistry and in particular pH fluctuations are characterized by a wide spatial heterogeneity

and temporal variability (daily and seasonal oscillations) in coastal ecosystems (145, 232, 305). The variability of pH is determined by a wide range of physical and biogeochemical processes, from mesoscale hydrological processes to small-scale metabolic processes (264).

The primary production in the western Mediterranean Sea is characterized by a seasonal variability induced by the increase of the surface layer nutrients by the winter vertical mixing in the water column (208). In addition, the presence of macrophytes(325)in the coastal areas of the northern Mediterranean Sea, mainly the endemic *Posidonia oceanica* whose meadows extend from the surface to 30-40 m depth, are defined as highly productive habitats. In these ecosystems, variability tends to follow daily and seasonal cycles, since biological metabolism is responsible for variations in the concentrations of oxygen ( $O_2$ ) and  $CO_2$  (145, 225), increasing pH values are expected for autotrophic ecosystems (production > respiration) during daylight hours. Indeed, recent studies indicate that seagrass meadows can locally alleviate low pH conditions for extended periods of time with important implications for the conservation and management of coastal ecosystems (366).

Nevertheless, changes in pH can appear idiosyncratic and display a diversity of patterns depending on the coastal area under consideration, as many drivers of the carbon system can influence these variable ecosystems, including temperature variability, biological activity and terrestrial and open ocean inputs (86). Therefore the properties of the carbon system have to be evaluated while taking into account the different interactions in every area. To the present day, there is still a lack of understanding of how coastal areas behave and how they contribute to the global carbon budget, also in part due to the intensive effort necessary to obtain representative time series of the carbon system data according to standard practices. The Global Ocean Acidification Observing Network (GOA-ON) defined that the accuracy included in the “weather goal” should be better than 0.02 and for the “climate goal” <0.003 pH units (332). Instrumentation for autonomous pH measurements have improved in recent years and production costs have come down, but they remain complex and relatively expensive. For climate change studies commercial oceanographic instrumentation barely accomplish the GOA-ON “climate goal” accuracy recommendation, with only spectrophotometric devices and Ion Sensitive Field-Effect Transistors (ISFETs) based pH probes reaching the standards. In this sense, the SAMI-pH sensor (Submersible Autonomous Moored Instrument, Sunburst Sensors, LCC), based on spectrophotometric techniques has been denoted as an excellent pH sensor for OA studies (161). However, the maintenance of oceanographic time series stations entails several operational and non operational difficulties, involving financial costs, meteorological risks (i.e. bad conditions for navigation, instrumentation loss, etc.), deployment in areas with high transit, issues essentially related to the sensor itself (i.e. instrumental failure) and possible human errors. Therefore, the appearance of data gaps is common, implying the lack of pH data obtained using high-quality instrumentation for global carbon studies.

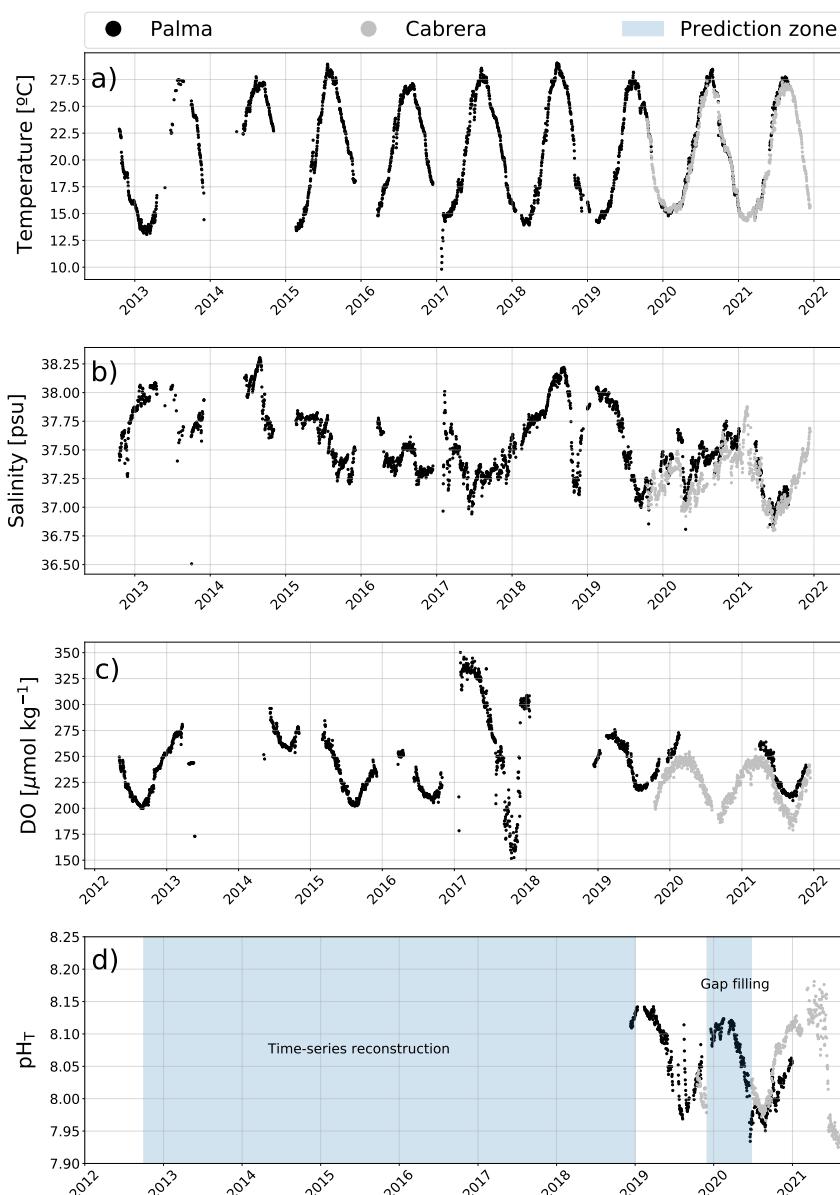
Currently, novel computational methods based on Machine Learning (ML) are allowing to tackle these data absence difficulties. Machine Learning is a part of Artificial Intelligence that has attained a mature status in the last decade or so, particularly through the so-called Deep Learning (DL) model (211), with major advances in solving problems that have resisted the best attempts of the artificial intelligence community for many years (275). In particular, some DL techniques are useful in time series forecasting (228) and also in the reconstruction of coupled time series (236), such as Recurrent Neural Network (RNN) architectures like Long Short-Term Memory (LSTM) (230) or Gated Recurrent Unit (GRU).

Nowadays, there is an increasing number of studies that use DL to understand the processes involved in the carbon system variability, but mainly focused on the open ocean (47, 66, 166, 169, 272), while relatively few studies focused on coastal seas (67, 108) and none specifically in the Mediterranean coastal Sea, perhaps because of the complexity and heterogeneity of the basin and its continental shelves. Therefore, the main objective of this study is to obtain the trend for pH decrease in the coastal Balearic Sea by applying Machine Learning techniques. In addition, this study aims to provide a useful tool to fill gaps in pH time series and to reconstruct pH data when additional environmental variables are available.

## 11.2 Results

### 11.2.1 Time series data

The collection of pH values, in total scale ( $\text{pH}_T$ ), started in December 2018 in the Bay of Palma, recording data almost continuously until the end of 2021. In the Cabrera station,  $\text{pH}_T$  was obtained from November 2019 to December 2021, with a relevant data gap from December 2019 to June 2020 (Fig. 11.1(d)) due to a sensor malfunction with a reparation prolonged for an extended period of time owing to the Covid-19 lockdown. Additional environmental parameters like temperature, salinity and dissolved oxygen (DO) concentration are available from the Bay of Palma station since 2012, while only a limited time series of these variables (since 2019) is available for Cabrera (Fig. 11.1(a-c)).



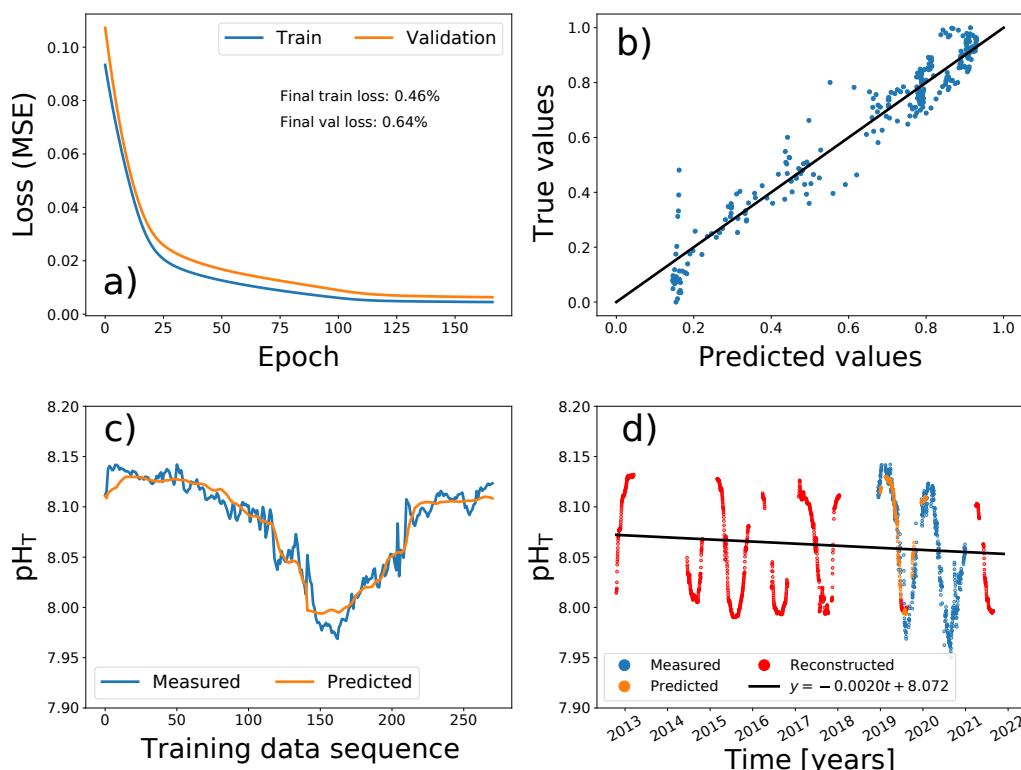
**Figure 11.1:** Daily averaged time series data from the Bay of Palma (black dots) and Cabrera stations (grey dots): a) Temperature ( $^{\circ}\text{C}$ ), b) Salinity (psu), c) Dissolved oxygen (DO) values ( $\mu\text{mol kg}^{-1}$ ) and d)  $\text{pH}_T$  in pH units. The  $\text{pH}$  time series of the Bay of Palma will be reconstructed in the period 2012-2021 while only gaps will be filled in Cabrera, as marked in blue in the figure.

In both stations, temperature ranged from a minimum of 12.99  $^{\circ}\text{C}$  to maximum values of

29.07°C from 2012 to 2021, with no observed differences between the stations in Cabrera and the Bay of Palma (Fig. 11.1(a)). The surface water temperatures are a clear representation of the typical Mediterranean climate seasonality with mild winters and warm to hot summers. Salinity did not show a repetitive seasonal pattern between years in either stations. However, in Cabrera salinity is slightly lower than in the Bay of Palma. During the data acquisition period, the lowest salinity value of 36.83 was found in Cabrera and highest of 38.30 in the Bay of Palma (Fig. 11.1(b)).

The surface water of the coastal sites in the Balearic Sea in the Palma Bay and the Cabrera stations was highly saturated with oxygen during all the seasons, with DO concentrations up to 348.94  $\mu\text{mol kg}^{-1}$  during winter and of 169.66  $\mu\text{mol kg}^{-1}$  during the summer and early autumn (Fig. 11.1(c)).  $\text{pH}_T$  values obtained starting in December, 2018 to December, 2021 increased during winter reaching up to 8.18 pH units at *in situ* temperature and decreasing to 7.91 pH units in summer, with the highest variability and maximum and minimum values measured in Cabrera (Fig. 11.1(d)).

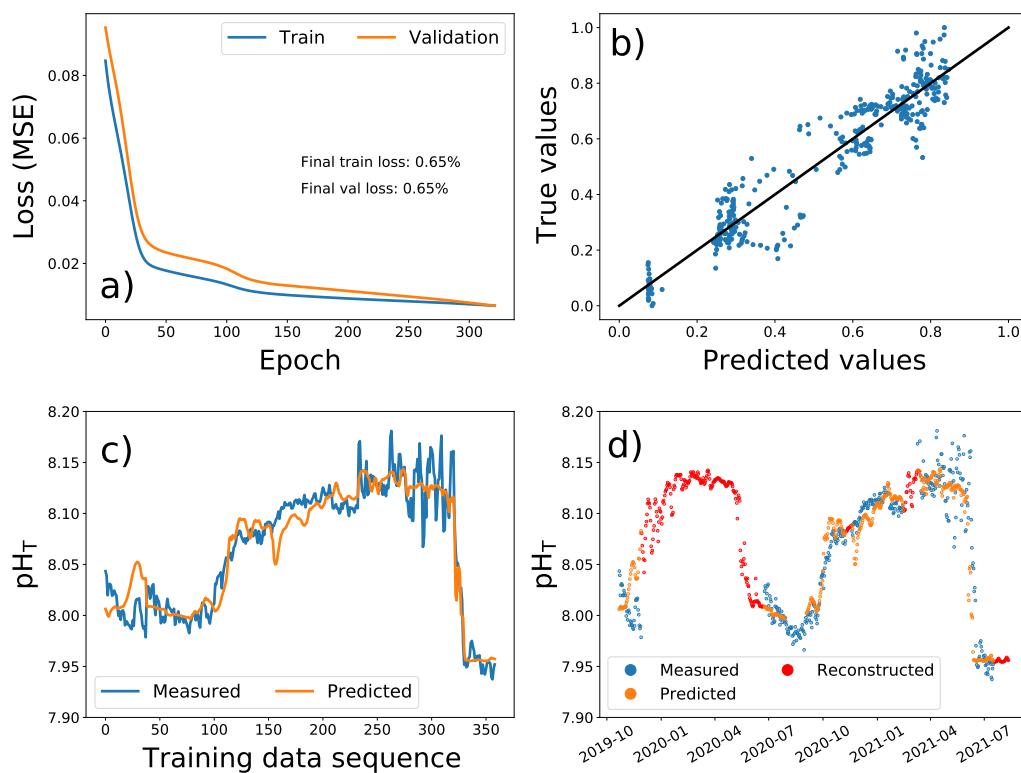
Considering sampling period of the additional (temperature and salinity) and calculated parameters (Total Alkalinity; TA) was larger in the Bay of Palma compared to Cabrera, we evaluate the linear tendencies with time for the Bay of Palma variables. The sea surface temperature in the Bay of Palma increased with a rate of  $0.05 \pm 0.03^\circ\text{C}$  per year ( $R^2=0.0008$ ,  $p\text{-value}=0.09$ ) from 2012 to 2021, whereas the salinity decreased significantly with  $-0.059 \pm 0.002$  psu per year ( $R^2=0.2456$ ,  $p\text{-value}<0.001$ ). The annual trend for TA, clearly related to the decrease in surface salinity, showed a relevant decrease of  $-4.0 \pm 0.4 \mu\text{mol kg}^{-1}$  ( $R^2=0.0379$ ,  $p\text{-value}<0.001$ ), supported by the discrete water samples for TA obtained during the period from 2019 to 2021 (??).



**Figure 11.2:** Bidirectional LSTM neural network model applied to assess the decadal  $\text{pH}_T$  trend in the Bay of Palma: a) Training process monitoring loss for both training and validation sets, b) Predicted  $\text{pH}_T$  values against their true values where the black line is the reference for a perfect prediction, c) Predicted  $\text{pH}_T$  time series in the training process (orange) and ground truth series (blue) and d) Final prediction for the decadal  $\text{pH}_T$  time series using the output data of the trained model and the measured data. Measured pH data shown in blue, predicted data in the training process is shown in orange and reconstructed data is shown in red. The black line represents the decadal pH trend.

## 11.2.2 Reconstruction pH time series with Deep Learning

The amount of available  $\text{pH}_T$  data from both Palma Bay and Cabrera stations is comparable and relatively short (mostly in Cabrera), but the length of the additional ambient data (temperature, salinity and DO) differs enormously between stations. Thus, there is a need to approach the time series prediction problem for both sites with different objectives. Common to both sites, a DL model with a RNN architecture will be developed to predict the  $\text{pH}_T$  time series from the accompanying ambient data (temperature, salinity and dissolved oxygen), which are expected to be correlated with  $\text{pH}_T$  (67, 166). To avoid the effect of site-specific correlations between ambient data and  $\text{pH}_T$  time series, the model will be trained independently with the dataset of each location. In this way a proper model calibration is ensured and the prediction power of the model is enhanced. In the Bay of Palma, the model will be used to reconstruct the  $\text{pH}_T$  time series from 2012, exclusively from the points for which the full set of ambient time series data are available (Fig. 11.1(d)). This is not possible in Cabrera, due to the fact that no temperature, salinity and DO concentration is available before 2019. Fortunately, these time series do not have the same gaps that the  $\text{pH}_T$  time series exhibits. Thus, we will use the model to fill the gaps in the  $\text{pH}_T$  time series from 2019 to present, as shown in (Fig. 11.1(d)).



**Figure 11.3:** Bidirectional LSTM Neural Network model applied to fill the gaps in the  $\text{pH}_T$  time series in Cabrera: a) Training process monitoring loss for both training and validation sets, b) Predicted  $\text{pH}_T$  values against their true values where the black line is the reference for perfect prediction, c) Predicted  $\text{pH}_T$  time series in the training process (orange) and ground truth series (blue) and d) Gaps in the  $\text{pH}_T$  time series filled with the trained model (red), while measured  $\text{pH}_T$  are shown in blue and predicted data in the training process shown in orange.

A BiDirectional Long Short-Term Memory (BD-LSTM) neural network (??) was selected as the best recurrent neural network architecture to reconstruct the  $\text{pH}_T$  time series in the Bay of Palma. The training process was successfully completed with no signs of overfitting achieving less than 1% error in both training and validation sets (Fig. 11.2(a)). The BD-LSTM neural network was able to fairly predict the majority of the individual pH data points in the time series, although there are some deviations (Fig. 11.2(b)). Furthermore, the time series pattern is perfectly captured by the neural network (Fig. 11.2(c)). Notice the gaps in the reconstructed pH points (in red) in (Fig. 11.2(d)), that are those for which the full ambient time-series is not available. Finally, the reconstructed pH data using the BD-LSTM model was used to assess the decadal trend of

acidification in Palma Bay, which yielded  $-0.0020$  units of pH per year (black line in (Fig. 11.2(d)). Indeed, to further characterize this decadal trend, 1000 independent training-prediction processes were carried out using a BD-LSTM neural network. The results showed a mean slope of  $-0.0020 \pm 0.00054$  for the decadal acidification trend (see methods).

Regarding the Cabrera data set, with the available ambient time series it is only possible to fill the data gaps, task for which a BD-LSTM neural network was also used. As for the Bay of Palma the training process was successfully completed with no signs of overfitting, yielding less than 1% error in both training and validation dataset (Fig. 11.3(a)). The model fairly predicts most of the individual  $\text{pH}_T$  data points in the training dataset, showing some deviations as usual (Fig. 11.3(b)). The tendency of the time series is perfectly captured by the model (Fig. 11.3(c)) and thus the gap can be filled with reliable data, red points in (Fig. 11.3(d)).

## 11.3 Discussion

The achievement of long time oceanographic data series suitable to evaluate the effects of climate change constitutes a great operational effort which is unequivocally accompanied by partial data loss due to multiple factors (human and instrumental). The advances in the development of pH sensors are enabling the acquisition of precise pH data without identified drift through highly accurate indicator-based spectrophotometric methods (395). However, in order to determine OA trends, several years of quality seawater pH data are needed, adding more difficulty to the vicissitudes inherent to field work. Recently, the application of computational methods based on Deep Learning (DL) is becoming a useful tool to fill the gaps due to data loss. Several studies have implemented the DL methodology and successfully predicted bio-optical and biogeochemical parameters (47, 66, 67, 108, 213, 277, 282, 383, 435).

Here, the application of a BiDirccional Long Short-Term Memory (BD-LSTM) neural network to predict  $\text{pH}_T$  from physical data, namely temperature, salinity, and dissolved oxygen, the latter as a key indicator of biological activity, permitted the reconstruction of gaps in the time series of  $\text{pH}_T$  and allowed the reconstruction of nine years of  $\text{pH}_T$  data. The BD-LSTM architecture has been proved extremely effective in predicting sequence data, such as time series, as they combine the information for both front and back directions of time (Online Supplementary Information) (212), and is more effective (accurate and stable) compared to unidirectional Long Short-Term Memory neural networks. Therefore, in this study the BD-LSTM offered better estimation results over the other neural networks considered to reconstruct time series but also in the completion of missing data.

In the Cabrera station, the BD-LSTM applied permitted a reliable reconstruction of the gaps in  $\text{pH}_T$  data from December 2019 to June 2020 (Fig. 11.1(d)), constituting an advantageous methodology to support the acquisition of long time series data without loosing accuracy, as the model can reproduce  $\text{pH}_T$  data with an error lower than 1% (Fig. 11.3(b)), closely following the annual variability of the observations (Fig. 11.3(d)).

The ability of the BD-LSTM to reconstruct time series was observed through the reconstruction of nine years of  $\text{pH}_T$  data in the Bay of Palma station (Fig. 11.2(d)). The modeled  $\text{pH}_T$  data combined with the observations allowed the accomplishment of a long pH time series in order to estimate a pH trend, seasonally adjusted through a sinusoidal fitting, with a rate of decrease of  $0.0020 \pm 0.00054$  pH units per year ( $R^2 = 0.1$ ,  $p\text{-value} < 0.001$ , ??), and represents the first estimate of pH trend obtained in the Balearic coastal Sea. Additionally, we applied a linear fit on the reconstructed pH time series obtaining trend of  $-0.0025 \pm 0.00053 \text{ yr}^{-1}$  ( $R^2 = 0.01$ ,  $p\text{-value} < 0.001$ ). This fit was discarded, because it was shown to introduce a bias in the pH decrease trend.

The observed pH decrease in the Balearic Sea coastal area is well aligned with OA trends reported for open ocean areas, from  $-0.0013$  pH units  $\text{yr}^{-1}$  in the Munida station (New Zealand) to the high trend found in the Cariaco Basin station up to  $-0.0026$  pH units per year(29). The processes associated with the increased pH decline in the Cariaco Basin where related to the upwelling of Subtropical Underwater, rich in dissolved inorganic carbon, thus lowering the pH.

In the Mediterranean Sea, previous annual estimates in open ocean areas ranged from -0.003 to -0.0044 (161, 455), reflecting the effect of the hydrodynamical and biogeochemical characteristics of the basin on the seawater pH variability(276, 343, 390). However, it can be assumed that differences in physical oceanography and ecological processes between areas may modulate local changes of pH. In a coastal Mediterranean area located in the

northwestern basin, close to Villefranche-sur-Mer, a rate of pH change of  $-0.0028 \pm 0.0003$  pH units  $\text{yr}^{-1}$  was observed(251) and attributed principally to atmospheric forcing and secondly to increased warming. The calculated trend of pH decrease due to the atmospheric CO<sub>2</sub> growth during the period of this study, from 2013 to 2021, was of  $0.0025 \pm 0.0002$  pH units per year ( $R^2=0.95$ ,  $p\text{-value}<0.001$ ), consistently related to the seawater pH decline. Therefore these analyses suggest that the atmospheric forcing is the main driver responsible for the pH decreasing trend found in the surface coastal Balearic Sea. Subsequently, the difference between the seawater pH decreasing trend obtained and the pH trend calculated from the atmospheric levels could be related to natural biogeochemical processes, not distinctly quantifiable with the available length of the Bay of Palma pH time series.

In addition, the effect of temperature on surface ocean pH, occurring directly through the temperature dependence of the seawater CO<sub>2</sub> chemistry, as changes in temperature and salinity influence the equilibrium constants of the oceanic CO<sub>2</sub> system and indirectly through air-sea exchange of CO<sub>2</sub>, can be considered. The influences of these two temperature processes on surface ocean pH has been found responsible of a 50% of the increase in (H<sup>+</sup>) ions, thus a pH decrease, in the surface layers of the Iceland and Irmingen Seas(346). In the Mediterranean Sea northwestern basin, a temperature increase of  $0.072 \pm 0.022$  °C  $\text{yr}^{-1}$  was estimated to be responsible for a 40 percent of the pH decrease(251). The obtained temperature variability in the Balearic Sea coastal area during this study was of  $0.035 \pm 0.008$  ( $R^2 = 0.008$ ,  $p\text{-value} < 0.001$ , ??), indicating that temperature-driven changes could also be assumed to affect the pH trend. The observed seasonal variability of the data, presented a pH<sub>T</sub> increase from 7.91 during summer up to 8.18 pH units (Fig. 11.1(d)) in winter seasons, clearly followed by the TA values (??). Seasonal changes in TA levels in the study area are ranging from around 2350 to 2550  $\mu\text{mol kg}^{-1}$  (??), largely overtaking the seasonal differences reported previously in the Balearic Sea of up to 50  $\mu\text{mol kg}^{-1}$  in total(118). This discrepancy in variability could be explained by the intense metabolic processes at the coastal location of the Bay of Palma station. This shallow area has a strong coverage of *Posidonia oceanica*, which due to its high ecosystem production(263) could be triggering an increase of pH and TA levels, as seen in salinity normalized TA values (NTA, not shown) during winter-spring, due to the uptake of nitrate and phosphate and the calcium carbonate dissolution(28, 118), and during summer, related to the lower community production(95) an NTA-pH decrease(118).

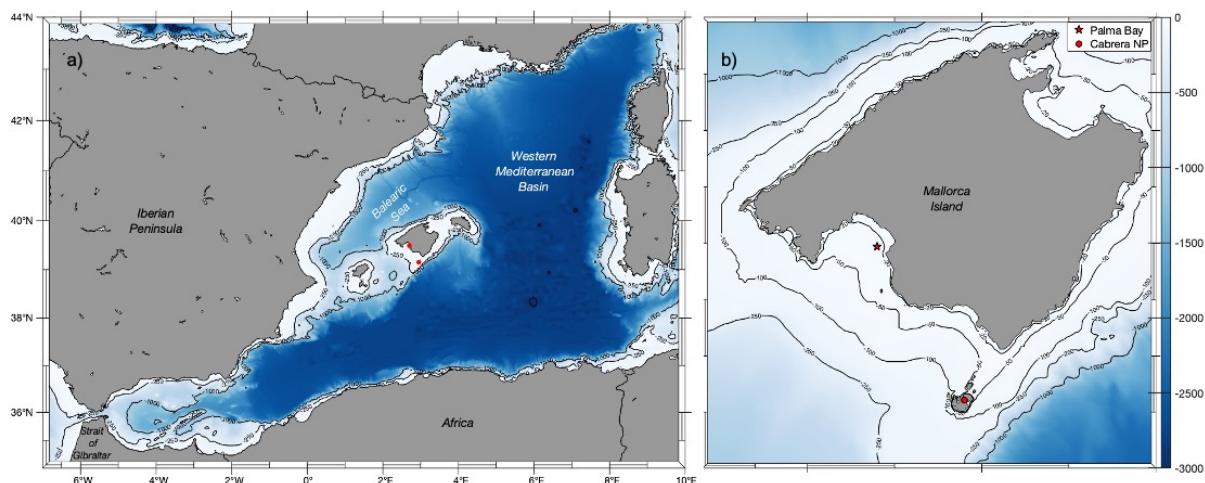
Another result from this study worth to mentioning is the obtained decreasing TA trend in the Bay of Palma of  $-4.0 \pm 0.4$   $\mu\text{mol kg}^{-1}$  per year. Although the Western Mediterranean is characterized with lower total alkalinity values in relation to the rest of the basin, resulted from the nearby influence of Atlantic waters, less salty with low-alkalinity water(224, 369), was not expected to influence decreasing decadal TA values. In the northwestern basin, TA values increased over time at a rate of  $2.08 \pm 0.19$   $\mu\text{mol kg}^{-1} \text{ yr}^{-1}$ . In the Balearic Sea, the decreasing TA confirm the Atlantic forcing on the alkalinity values and the negligible TA discharges due to rivers in the Balearic Islands. There is a marked south-to-north surface gradient in the western region coupled with the west-to-east gradient of alkalinity in the Mediterranean Sea related to the Atlantic influence(118, 182). Due to a well established linear relationship of TA and salinity(389) and the calculated origin of our values(182) we cannot neglect the strong TA related to the salinity decrease in the study area of  $-0.059 \pm 0.002$  psu per year ( $R^2=0.25$ ,  $p\text{-value}<0.001$ ). This rate is in agreement with the salinity decrease found at the coastal site at Villefranche-sur-Mer ( $-0.0017 \pm 0.0044$  psu  $\text{yr}^{-1}$ ).(251). Notwithstanding, the intense salinity decrease observed in the Bay of Palma can be linked to a decrease of the intensity of the southern spreading of the Balearic Current trough the Ibiza channel (located between Ibiza and Mallorca Islands) driven by mesoscale processes, and the prevalence of new Atlantic Water coming from the Strait of Gibraltar(308). Although, this observation is out of the scope of this study and therefore further investigation is needed.

In summary, this work pointed out the useful use of DL techniques, specifically the BD-LSTM architecture, to reconstruct pH data relevant to evaluate seasonal pH variability and to elucidate the climate change consequences, as the OA effect, in a coastal area of the Balearic Sea, which can be extended to the coastal areas of the Western Mediterranean Sea Basin. Nevertheless, future research is necessary to assess and confirm these regional trends, which highlights the importance of maintaining the time series monitoring networks whose data are the base of this study.

## 11.4 Methods

### Study area

We monitored two coastal stations located in the Archipelago of the Balearic Islands in the Western Mediterranean Sea (Fig. 11.4). One site was positioned within the Bay of Palma ( $39^{\circ}29.57088'N$ ;  $2^{\circ}42.02430'E$  Fig. 11.4(b)) at a fixed station consisting of an oceanographic buoy managed by the Balearic Islands Coastal Ocean Observing and Forecasting System (SOCIB; <https://www.socib.eu/>). Here meteorological, hydrological and hydrodynamic data are collected with an hourly frequency since October 2012. The buoy is located at the surface over 20m bottom depth. The Bay of Palma is a large bay with a surface area of  $217 \text{ km}^2$  and approximately 30% seagrass cover(178). The second station was located at 4m depth on a mooring line over 8m bottom depth deployed in the Marine and Terrestrial National Park of the Archipelago of Cabrera ( $39^{\circ}9.08217'N$ ;  $2^{\circ}57.04767'E$ ; Fig. 11.4(b)). The mooring line is in a small bay of just under  $1 \text{ km}^2$  and full protection with the largest meadow of the archipelago, covering 89.1% of the surface area between 0-10m depth(293). Neither site has important freshwater inputs. Both stations are part of the Balearic Ocean Acidification Time Series (BOATS) network included in the Interdisciplinary Thematic Platform: Water:iOS (<https://pti-waterios.csic.es/>).



**Figure 11.4:** a) Map of the stations location in the Western Mediterranean Sea Basin (red dots) and b) detailed location of the Bay of Palma (red star) and the Cabrera National Park (Cabrera NP, red dot) study sites. Maps were developed with the MATLAB® R2010b software (<https://mathworks.com>) by using the M\_Map toolbox[345].

#### 11.4.1 Data collection

In both stations a SAMI-pH (Sunburst Sensors LCC) was attached, at 1 m in the Bay of Palma and at 4 m depth in Cabrera. The pH sensors were measuring pH, in the total scale ( $\text{pH}_T$ ), hourly since December 2018 in the Bay of Palma and since November 2019 in Cabrera. The sensor precision and accuracy are  $<0.001 \text{ pH}$  and  $\pm 0.003 \text{ pH}$  units, respectively. Monthly maintenance of the sensors was performed including data download and surface cleaning.

Temperature and salinity from the Bay of Palma oceanographic buoy was obtained from October 2012 and for the Cabrera mooring line from November 2019 with a CT SBE37 (Sea-Bird Scientific©) in both stations. Accuracy of the CT is  $\pm 0.002^{\circ}\text{C}$  for temperature and  $\pm 0.003 \text{ mS cm}^{-1}$  for conductivity. Additionally, oxygen data from a SBE 63 (Sea-Bird Scientific ©) sensor attached to the CT in Cabrera and from a YSI 6600V2-4 Multiparameter Water Quality Sonde with a 6450 ROX DO sensor (Yellow Spring Instruments Inc. ©)(419) and a miniDot (PME, Inc. ©) in the Bay of Palma were used. Accuracy of oxygen sensors is  $\pm 2\%$ ,  $\pm 1\%$  and  $\pm 5\%$  for the SBE 63, the YSI and the miniDot, respectively.

Periodically water samplings for dissolved oxygen (DO), pH in total scale at  $25^{\circ}\text{C}$  ( $\text{pH}_{T25}$ ) and total alkalinity (TA) were obtained during the sensor maintenance campaigns. DO and ( $\text{pH}_{T25}$

samples were collected in order to validate the data obtained by the sensors.

DO concentrations were evaluated with the Winkler method modified by Benson and Krause (1983) (35) by potentiometric titration with a Metrohm 808 Titrand with an accuracy of the method of  $\pm 2.9 \mu\text{mol kg}^{-1}$  and with an obtained standard deviation from the sensors data and the water samples collected of  $\pm 5.9 \mu\text{mol kg}^{-1}$ .

$\text{pH}_{T25}$  data was obtained by the spectrophotometric method with a Shimadzu UV-2501 spectrophotometer containing a 25°C-thermostated cells with unpurified *m*-cresol purple as indicator following the methodology established by Clayton and Byrne (1993)(104) by using Certified Reference Material (CRM Batch #176 supplied by Prof. Andrew Dickson, Scripps Institution of Oceanography, La Jolla, CA, USA). The accuracy obtained from the CRM Batch was of  $\pm 0.0051$  pH units and the precision of the method of  $\pm 0.0034$  pH units. The mean difference between the SAMI-pH and discrete samples was of 0.0017 pH units.

TA samples were collected in 50 mL Falcom vials and poisoned with 20  $\mu\text{L}$  of  $\text{HgCl}_2$  and determined by open cell potentiometric titration with a Titrand 808 (Metrohm) following the Standard Operation Procedure (SOP) 3b(137). TA values were also calculated from the temperature and salinity values obtained in the Bay of Palma from 2012 by using a second-order polynomial model for TA specifically described for the Mediterranean Basin (182).

pH values due to the atmospheric  $\text{CO}_2$  levels were estimated by using the CO2SYSv3 program(399), with the most internally consistent and preferred carbon (135, 303) and sulfate dissociation constants(136) for current surface ocean studies(454), with the Bay of Palma *in situ* temperature and salinity, the calculated TA values and the atmospheric  $\text{CO}_2$  levels converted from dry air to wet(443) as inputs. Carbon dioxide ( $\text{CO}_2$ ) atmospheric molar fraction used was obtained from the monitoring station of Lampedusa (LMP), Italy of the NOAA (National Oceanic and Atmospheric Administration, USA) monitoring network(141).

## 11.4.2 Data processing

Once data was validated, several processing steps were performed to ensure an optimal training process for the neural network models. First, all the data of the time series were re-sampled by averaging the data points obtaining a daily frequency. Afterwards, a standard feature-scaling procedure (min-max normalization) was applied to every feature (Temperature, Salinity and Oxygen) and to  $\text{pH}_T$ . Finally, we built our training and validations sets as tensors with dimensions ( $\text{batch\_size}, \text{window\_size}, N_{\text{features}}$ ), where  $\text{batch\_size}$  is the number of examples to train per iteration,  $\text{batch\_size}$  is the number of past and future points considered and  $N_{\text{features}}$  is the number of features used to predict the target series. Temperature values below  $T = 12.5^\circ\text{C}$  were discarded as they are considered outliers in sensor data outside the normal range in the study area.

## 11.4.3 Computing the trend of seasonal data

The trend of seasonal time-series is often computed by means of statistical methods based on moving averages or more advanced techniques such as the Seasonal Trend Decomposition Loess (105). Nevertheless, these procedures do not work with gappy time series, so that a different approach is needed. In this work we fitted the following oscillatory function with trend to our data:

$$y(t) = A \sin(\omega t + \phi) + Bt + C , \quad (11.1)$$

where the parameter  $B$  corresponds to the trend of the data.

Moreover, after this fit, the seasonal component ( $A \sin(\omega t + \phi)$ ) can be removed from the original time-series and a standard linear regression can be performed to the transformed data to obtain the trend (which is exactly  $B$ ) with the  $R^2$  and  $p$ -value estimates given by the linear regression (????).

## 11.4.4 Selecting the best neural network architecture

Several recurrent neural network (RNN) architectures were considered as candidates to reconstruct the pH time series, including a Simple Recurrent neural network (SRNN), Long-Short Term Memory (LSTM), BiDirectional LSTM (BD-LSTM, ??) and BiDirectional Gated Recurrent Unit (BD-GRU).

Initially, manual tests were performed on each architecture to determine the optimal set of parameters that yielded the best possible results. These tests were based on minimizing the errors in both training and validation set while avoiding overfitting. To avoid overfitting we implemented automated callbacks to stop the training process whenever the validation loss increased or crossed the training loss. During this test we determined the minimum number of nodes, which helps in avoiding overfitting and the minimum window size, which allows to use the most possible number of data points for training and prediction. All the RNNs were trained in batches of size 32. To enhance clarity and accessibility, the optimal values obtained for the more relevant parameters are summarized in [Table 11.1](#).

**Table 11.1:** Optimal parameters used for the different RNN architectures

	Hidden layers	Nodes/Cells	Window size	Activation function	Output function	Loss	Learning rate	Optimizer
<b>SRNN</b>	1	3	6	Tanh	Sigmoid	MSE	0.01	Adam
<b>LSTM</b>	1	3	6	Tanh	Sigmoid	MSE	0.01	Adam
<b>BD-LSTM</b>	1	3	6	Tanh	Sigmoid	MSE	0.01	Adam
<b>BD-GRU</b>	1	1	6	Tanh	Sigmoid	MSE	0.01	Adam

In order to identify the best-performing architecture an automated procedure was developed to statistically compare the outputs of each model. Each architecture was trained in 1000 independent processes, ensuring a final training mean-squared error of less than 0.8% while avoiding overfitting implementing the previously mentioned callbacks. The code used for the analysis can be found in [\(192\)](#).

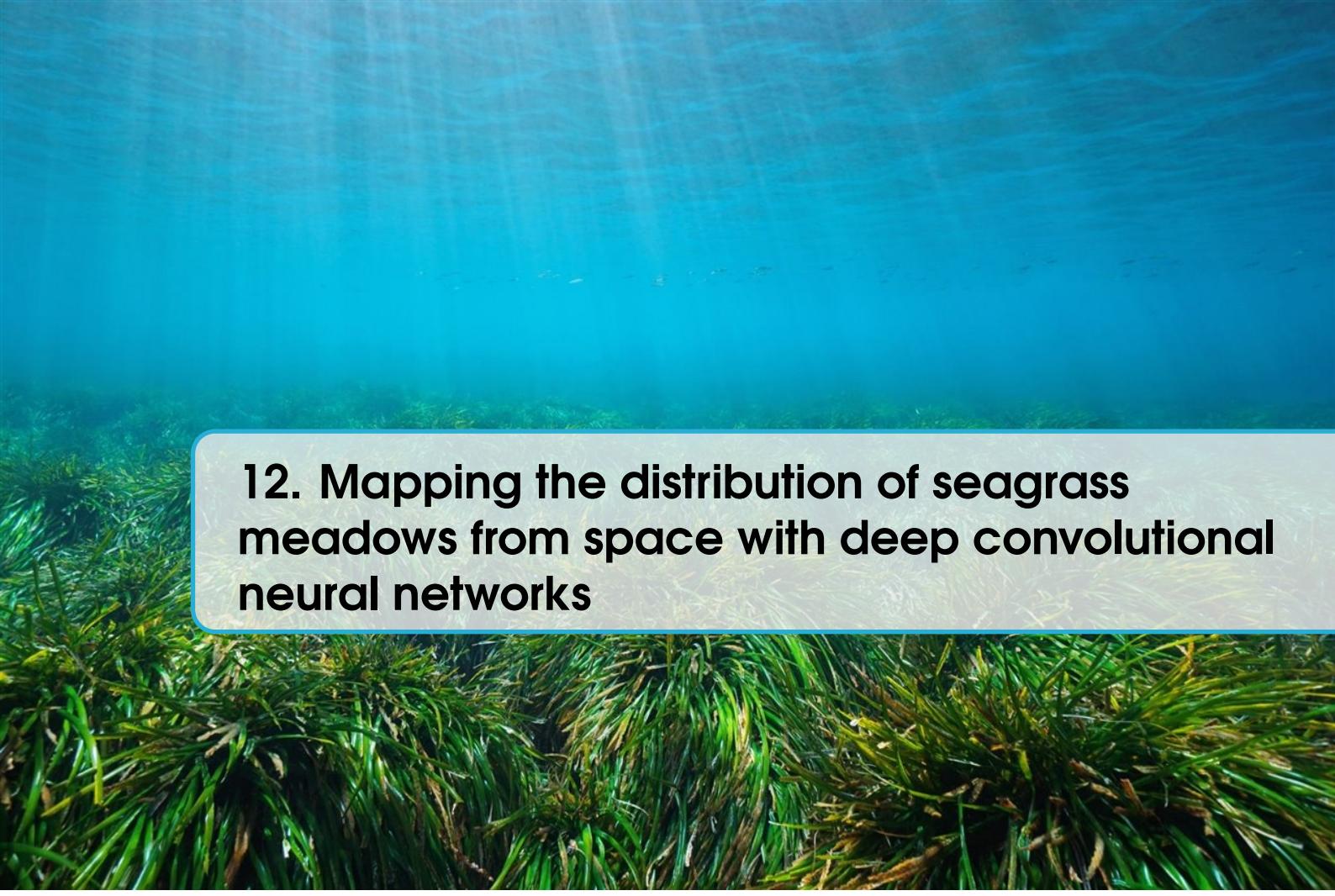
In [Table 11.2](#) a summary of the statistical results obtained for each architecture is presented. All architectures provide similar training and validation errors and provide similar results for the decadal pH<sub>T</sub> trend, predicting a slope of around  $-0.0030$  pH units per year with an intercept of 8.07 pH units. However, the BD-LSTM turns out to be the architecture providing most accurate (smallest training and validation errors) and precise (smallest statistical error) results (??). Thus, we selected the BD-LSTM neural network as the best architecture to reconstruct the pH<sub>T</sub> time series. Data corresponding to the Bay of Palma were used in the selection of the best neural network architecture.

**Table 11.2:** Statistical comparison between different RNN architectures

	Slope	Intercept	Training error	Validation error	Training epochs	Training time
<b>RNN</b>	$-0.0021 \pm 0.00077$	$8.07 \pm 0.006$	$0.54 \pm 0.08$	$0.72 \pm 0.12$	$293 \pm 95$	$15.52 \pm 4.75$
<b>LSTM</b>	$-0.0018 \pm 0.00067$	$8.06 \pm 0.005$	$0.49 \pm 0.03$	$0.68 \pm 0.05$	$245 \pm 68$	$17.55 \pm 4.21$
<b>BD-LSTM</b>	$-0.0020 \pm 0.00054$	$8.07 \pm 0.004$	$0.46 \pm 0.03$	$0.64 \pm 0.04$	$167 \pm 45$	$15.13 \pm 3.00$
<b>BD-GRU</b>	$-0.0020 \pm 0.00066$	$8.07 \pm 0.005$	$0.51 \pm 0.07$	$0.74 \pm 0.10$	$347 \pm 95$	$27.68 \pm 6.84$

The code and data used to determine the best neural network architecture can be found in a GitHub repository [\(191\)](#).



A photograph of a seagrass meadow underwater. The dense, green, blade-like leaves of the seagrass grow in patches across the sandy ocean floor. The water is clear, allowing light to penetrate and illuminate the plants. The background shows more of the seagrass extending towards the horizon.

## **12. Mapping the distribution of seagrass meadows from space with deep convolutional neural networks**



## **Bibliography**



# Bibliography

- [1] (EFSA) et al. “Update of the *Xylella* spp. host plant database – systematic literature search up to 31 December 2022”. In: *EFSA Journal* 21.6 (2023), e08061. DOI: <https://doi.org/10.2903/j.efsa.2023.8061> (cited on page 110).
- [2] C. Abboud et al. “Dating and localizing an invasion from post-introduction data and a coupled reaction-diffusion-absorption model”. In: *Journal of Mathematical Biology* 79.2 (July 2019), pages 765–789. ISSN: 1432-1416. DOI: <10.1007/s00285-019-01376-x>. URL: <https://doi.org/10.1007/s00285-019-01376-x> (cited on page 60).
- [3] Umarfarooq A. Abdulwahab, Edd Hammill, and Charles P. Hawkins. “Choice of climate data affects the performance and interpretation of species distribution models”. In: *Ecological Modelling* 471 (2022), page 110042. ISSN: 0304-3800. DOI: <https://doi.org/10.1016/j.ecolmodel.2022.110042>. URL: <https://www.sciencedirect.com/science/article/pii/S0304380022001521> (cited on page 110).
- [4] A. Alonso Aguirre and Gary M. Tabor. “Global Factors Driving Emerging Infectious Diseases”. In: *Annals of the New York Academy of Sciences* 1149.1 (2008), pages 1–3. DOI: <https://doi.org/10.1196/annals.1428.052> (cited on page 30).
- [5] R. Albibi et al. “RAPD fingerprinting *Xylella fastidiosa* Pierce’s disease strains isolated from a vineyard in North Florida”. In: *FEMS Microbiology Letters* 165 (1998), pages 347–352 (cited on page 92).
- [6] Allen Coral Atlas. *Imagery, maps and monitoring of the world’s tropical coral reefs*. 2022. DOI: <10.5281/zenodo.3833242>. URL: <https://zenodo.org/record/6622015> (cited on pages 124, 130).
- [7] R. P. P. Almeida. “*Xylella fastidiosa* Vector Transmission Biology”. In: *Vector-Mediated Transmission of Plant Pathogens*. Edited by Judith K. Brown. St Paul, MN, USA: APS Publications, 2016. Chapter 12, pages 165–173. ISBN: 978-0-89054-535-5. DOI: <10.1094/9780890545355.012> (cited on page 78).
- [8] R. P. P. Almeida and L. Nunney. “How do plant diseases caused by *Xylella fastidiosa* emerge?” In: *Plant Disease* 99 (11 1987), pages 1457–1467. DOI: <10.1094/PDIS-02-15-0159-FE> (cited on pages 60, 78).
- [9] R. P. P. Almeida and A. H. Purcell. “Biological Traits of *Xylella fastidiosa* Strains from Grapes and Almonds”. In: *Applied and Environmental Microbiology* 69.12 (12 2003), pages 7447–7452. DOI: <10.1128/AEM.69.12.7447-7452.2003>. eprint: <https://journals.asm.org/doi/pdf/10.1128/AEM.69.12.7447-7452.2003>. URL: <https://journals.asm.org/doi/abs/10.1128/AEM.69.12.7447-7452.2003> (cited on pages 60, 78, 89, 91, 118).
- [10] R. P. P. Almeida et al. “Addressing the New Global Threat of *Xylella fastidiosa*”. In: *Phytopathology* 109.2 (2019), pages 172–174. DOI: <10.1094/PHYTO-12-18-0488-FI> (cited on page 78).
- [11] Rodrigo P.P. Almeida and Alexander H. Purcell. “Biological traits of *Xylella fastidiosa* strains from grapes and almonds”. In: *Applied and Environmental Microbiology* 69.12 (2003), pages 7447–7452. DOI: <https://doi.org/10.1128/AEM.69.12.7447-7452.2003> (cited on page 96).
- [12] Rodrigo PP Almeida and Leonard Nunney. “How do plant diseases caused by *Xylella fastidiosa* emerge?” In: *Plant Disease* 99.11 (2015), pages 1457–1467. DOI: <https://doi.org/10.1094/PDIS-02-15-0159-FE> (cited on page 96).
- [13] J. Alstott, E. Bullmore, and D. Plenz. “powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions”. In: *PLoS ONE* 9.1 (Jan. 2014), pages 1–11. DOI: <10.1371/journal.pone.0085777>. URL: <https://doi.org/10.1371/journal.pone.0085777> (cited on page 131).
- [14] M. Altman. “Traditional Districting Principles: Judicial Myths vs. Reality”. In: *Social Science History* 22.2 (1998), pages 159–200. DOI: <10.2307/1171534>. URL: <https://doi.org/10.2307/1171534> (cited on page 131).
- [15] M. Álvarez et al. “The CO<sub>2</sub> system in the Mediterranean Sea: A basin wide perspective”. In: *Ocean Science* 10.1 (2014), pages 69–92. ISSN: 18120784. DOI: <10.5194/os-10-69-2014> (cited on page 134).

- [16] Lorenzo Alvarez-Filip et al. “Flattening of Caribbean coral reefs: region-wide declines in architectural complexity”. In: *Proceedings of the Royal Society B: Biological Sciences* 276.1669 (2009), pages 3019–3025. DOI: [10.1098/rspb.2009.0339](https://doi.org/10.1098/rspb.2009.0339). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.2009.0339>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2009.0339> (cited on page 124).
- [17] O. Anas et al. “The effect of warming winter temperatures on the severity of Pierce’s disease in the Appalachian mountains and Piedmont of the southeastern United States”. In: *Plant Health Progress* 9 (1 2008), pages 1–17. DOI: [10.1094/PHP-2008-0718-01-RS](https://doi.org/10.1094/PHP-2008-0718-01-RS) (cited on pages 78, 81, 90, 92).
- [18] Roy M. Anderson. “Discussion: The Kermack-McKendrick epidemic threshold theorem”. In: *Bulletin of Mathematical Biology* 53 (1991), pages 3–32. URL: <https://www.sciencedirect.com/science/article/pii/S0092824005800394> (cited on pages 8, 30, 46).
- [19] J. Andrews. “Epizootiology of diseases of oysters (*Crassostrea virginica*), and parasites of associated organisms in eastern North America”. In: *Helgolander Meeresuntersuchungen* 37 (Jan. 1984), pages 149–166. URL: <https://scholarworks.wm.edu/vimsarticles/1447> (cited on page 20).
- [20] S. Antonatos et al. “Seasonal Appearance, Abundance, and Host Preference of *Philaenus spumarius* and *Neophilaenus campestris* (Hemiptera: Aphrophoridae) in Olive Groves in Greece”. In: *Environmental Entomology* 50.6 (Sept. 2021), pages 1474–1482. ISSN: 0046-225X. DOI: [10.1093/ee/nvab093](https://doi.org/10.1093/ee/nvab093). eprint: <https://academic.oup.com/ee/article-pdf/50/6/1474/41745778/nvab093.pdf>. URL: <https://doi.org/10.1093/ee/nvab093> (cited on page 62).
- [21] Apache Software Foundation. *GRIB.jl*. Version 0.3.0. Jan. 4, 2021. URL: <https://github.com/weech/GRIB.jl> (cited on page 91).
- [22] Isabelle Arzul and Ryan B. Carnegie. “New perspective on the haplosporidian parasites of molluscs”. In: *Journal of Invertebrate Pathology* 131 (2015), pages 32–42. DOI: [10.1016/j.jip.2015.07.014](https://doi.org/10.1016/j.jip.2015.07.014) (cited on pages 9, 24).
- [23] Tejas Athni et al. *How vector-borne disease shaped the course of human history*. June 2020. DOI: [10.22541/au.159135318.82566392](https://doi.org/10.22541/au.159135318.82566392) (cited on page 46).
- [24] Corinne Audemard et al. “*Bonamia exitiosa* transmission among, and incidence in, Asian oyster *Crassostrea ariakensis* under warm euhaline conditions”. In: *Diseases of Aquatic Organisms* 110 (July 2014), pages 143–50. DOI: [10.3354/dao02648](https://doi.org/10.3354/dao02648) (cited on page 20).
- [25] Nicolas Bacaër. “Approximation of the Basic Reproduction Number R0 for Vector-Borne Diseases with a Periodic Vector Population”. In: *Bulletin of Mathematical Biology* 69.3 (Apr. 2007), pages 1067–1091. ISSN: 1522-9602. DOI: [10.1007/s11538-006-9166-9](https://doi.org/10.1007/s11538-006-9166-9) (cited on page 62).
- [26] Nicolas Bacaër and Souad Guernaoui. “The epidemic threshold of vector-borne diseases with seasonality”. In: *Journal of Mathematical Biology* 53.3 (Sept. 2006), pages 421–436. ISSN: 1432-1416. DOI: [10.1007/s00285-006-0015-0](https://doi.org/10.1007/s00285-006-0015-0) (cited on page 53).
- [27] Per Bak, Chao Tang, and Kurt Wiesenfeld. “Self-organized criticality”. In: *Phys. Rev. A* 38 (1 1988), pages 364–374. DOI: [10.1103/PhysRevA.38.364](https://doi.org/10.1103/PhysRevA.38.364). URL: <https://link.aps.org/doi/10.1103/PhysRevA.38.364> (cited on page 129).
- [28] Cristina Barrón et al. “Organic carbon metabolism and carbonate dynamics in a Mediterranean seagrass (*Posidonia oceanica*), meadow”. In: *Estuaries and Coasts* 29.3 (2006), pages 417–426. DOI: [10.1007/BF02784990](https://doi.org/10.1007/BF02784990) (cited on page 140).
- [29] Nicholas R Bates et al. “A Time-Series View of Changing Surface Ocean Chemistry Due to Ocean Uptake of Anthropogenic CO<sub>2</sub> and Ocean Acidification”. In: *Oceanography* 27.1 (2014), pages 126–141. DOI: [10.5670/oceanog.2014.16](https://doi.org/10.5670/oceanog.2014.16) (cited on pages 134, 139).
- [30] D. J. Beal et al. “Seasonal Abundance and Infectivity of *Philaenus spumarius* (Hemiptera: Aphrophoridae), a Vector of *Xylella fastidiosa* in California Vineyards”. In: *Environmental Entomology* 50.2 (2 Jan. 2021), pages 467–476. ISSN: 0046-225X. DOI: [10.1093/ee/nvaa178](https://doi.org/10.1093/ee/nvaa178). eprint: <https://academic.oup.com/ee/article-pdf/50/2/467/37313684/nvaa178.pdf>. URL: <https://doi.org/10.1093/ee/nvaa178> (cited on pages 62, 78, 89).
- [31] D. P. Bebber, M. A. T. Ramotowski, and S. J. Gurr. “Crop pests and pathogens move polewards in a warming world”. In: *Nature Climate Change* 3.11 (Nov. 2013), pages 985–988. ISSN: 1758-6798. DOI: [10.1038/nclimate1990](https://doi.org/10.1038/nclimate1990) (cited on pages 79, 110).

- [32] D. P. Bebber et al. “Modelling coffee leaf rust risk in Colombia with climate reanalysis data”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1709 (2016), page 20150458. DOI: [10.1098/rstb.2015.0458](https://doi.org/10.1098/rstb.2015.0458) (cited on page 79).
- [33] Daniel P Bebber. “Climate change effects on Black Sigatoka disease of banana”. In: *Philosophical Transactions of the Royal Society B* 374.1775 (2019), page 20180269. DOI: <https://doi.org/10.1098/rstb.2018.0269> (cited on page 96).
- [34] Daniel P. Bebber, Timothy Holmes, and Sarah J. Gurr. “The global spread of crop pests and pathogens”. In: *Global Ecology and Biogeography* 23.12 (2014), pages 1398–1407. DOI: <https://doi.org/10.1111/geb.12214> (cited on page 110).
- [35] Bruce B. Benson and Daniel Krause. “The concentration and isotopic fractionation of oxygen dissolved in freshwater and seawater in equilibrium with the atmosphere”. In: *Limnology and Oceanography* 29.3 (1984), pages 620–632. ISSN: 19395590. DOI: [10.4319/lo.1984.29.3.0620](https://doi.org/10.4319/lo.1984.29.3.0620) (cited on page 142).
- [36] Magali Bergot et al. “Simulation of potential range expansion of oak disease caused by *Phytophthora cinnamomi* under climate change”. In: *Global Change Biology* 10.9 (2004), pages 1539–1552. DOI: <https://doi.org/10.1111/j.1365-2486.2004.00824.x> (cited on page 96).
- [37] B. Berisha et al. “Isolation of Peirce’s disease bacteria from grapevines in Europe”. In: *European Journal of Plant Pathology* 104 (8 1998), pages 427–433. DOI: [10.1023/A:1008655621235](https://doi.org/10.1023/A:1008655621235) (cited on page 88).
- [38] E. Bertuzzo et al. “On spatially explicit models of cholera epidemics”. In: *Journal of The Royal Society Interface* 7.43 (2010), pages 321–333. DOI: [10.1098/rsif.2009.0204](https://doi.org/10.1098/rsif.2009.0204) (cited on page 38).
- [39] Jean Pierre Bethoux et al. “The Mediterranean Sea: a miniature ocean for climatic and environmental studies and a key for the climatic functioning of the North Atlantic”. In: *Progress in Oceanography* 44.1-3 (1999), pages 131–146. DOI: [10.1016/S0079-6611\(99\)00023-3](https://doi.org/10.1016/S0079-6611(99)00023-3) (cited on page 134).
- [40] Jeff Bezanson et al. “Julia: A fresh approach to numerical computing”. In: *SIAM Review* 59.1 (2017), pages 65–98. DOI: [10.1137/141000671](https://doi.org/10.1137/141000671) (cited on pages 20, 26, 63, 64, 91).
- [41] C. Nike Bianchi and Carla Morri. “Marine biodiversity of the Mediterranean Sea: Situation, problems and prospects for future research”. In: *Marine Pollution Bulletin* 40.5 (2000), pages 367–376. ISSN: 0025326X. DOI: [10.1016/S0025-326X\(00\)00027-8](https://doi.org/10.1016/S0025-326X(00)00027-8) (cited on page 134).
- [42] G. Bidegain et al. “Microparasitic disease dynamics in benthic suspension feeders: Infective dose, non-focal hosts, and particle diffusion”. In: *Ecological Modelling* 328 (2016), pages 44–61. ISSN: 0304-3800. DOI: [10.1016/j.ecolmodel.2016.02.008](https://doi.org/10.1016/j.ecolmodel.2016.02.008) (cited on pages 8, 11, 30).
- [43] G. Bidegain et al. “Modeling the transmission of *Perkinsus marinus* in the Eastern oyster *Crassostrea virginica*”. In: *Fisheries Research* 186 (2017), pages 82–93. ISSN: 0165-7836. DOI: [10.1016/j.fishres.2016.08.006](https://doi.org/10.1016/j.fishres.2016.08.006) (cited on pages 8, 30).
- [44] Gorka Bidegain et al. “Marine infectious disease dynamics and outbreak thresholds: contact transmission, pandemic infection, and the potential role of filter feeders”. In: *Ecosphere* 7.4 (2016), e01286. DOI: [10.1002/ecs2.1286](https://doi.org/10.1002/ecs2.1286) (cited on pages 8, 10, 11, 14, 19, 23, 24).
- [45] N. L. Bindoff et al. “Changing Ocean, Marine Ecosystems, and Dependent Communities”. In: *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. Edited by H.-O. Pörtner et al. Cambridge, UK: Cambridge University Press, 2019. Chapter 5, pages 447–587. DOI: [10.1017/9781009157964.007](https://doi.org/10.1017/9781009157964.007) (cited on page 124).
- [46] J. J. Binney et al. *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. Oxford (UK): Clarendon Press, 1992. ISBN: 9780198513933. URL: <https://global.oup.com/academic/product/the-theory-of-critical-phenomena-9780198513933> (cited on page 129).
- [47] Henry C Bitting et al. “An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO<sub>2</sub> Variables and Nutrient Concentrations From T, S, and O<sub>2</sub> Data Using Bayesian Neural Networks”. In: *Frontiers in Marine Science* 5 (2018), page 328. ISSN: 2296-7745. DOI: [10.3389/fmars.2018.00328](https://doi.org/10.3389/fmars.2018.00328). URL: <https://www.frontiersin.org/article/10.3389/fmars.2018.00328> (cited on pages 135, 139).
- [48] N. Bodino et al. “Phenology, seasonal abundance and stage-structure of spittlebug (Hemiptera: Aphrophoridae) populations in olive groves in Italy”. In: *Scientific reports* 9.1 (2019), pages 1–17. DOI: [10.1038/s41598-019-54279-8](https://doi.org/10.1038/s41598-019-54279-8) (cited on page 60).

- [49] N. Bodino et al. “Temporal dynamics of the transmission of *Xylella fastidiosa* subsp. *pauca* by *Philaenus spumarius* to olive plants”. In: *Entomologia Generalis* 41.5 (Oct. 2021), pages 463–480. DOI: [10.1127/entomologia/2021/1294](https://doi.org/10.1127/entomologia/2021/1294). URL: <http://dx.doi.org/10.1127/entomologia/2021/1294> (cited on pages 60, 68).
- [50] Alberto V. Borges and Nathalie Gypens. “Carbonate chemistry in the coastal zone responds more strongly to eutrophication than to ocean acidification”. In: *Limnology and Oceanography* 55.1 (2010), pages 346–353. ISSN: 00243590. DOI: [10.4319/lo.2010.55.1.0346](https://doi.org/10.4319/lo.2010.55.1.0346) (cited on page 134).
- [51] S. G. Borkar. *History of Plant Pathology*. Boca Raton, FL, USA: WPI Publishing, 2017. DOI: [10.1201/b22437](https://doi.org/10.1201/b22437) (cited on pages 78, 89).
- [52] Joris Van den Bossche et al. *geopandas/geopandas: v0.13.2*. Version v0.13.2. June 2023. DOI: [10.5281/zenodo.8009629](https://doi.org/10.5281/zenodo.8009629). URL: <https://doi.org/10.5281/zenodo.8009629> (cited on page 130).
- [53] Hemmo Bosscher and Wolfgang Schlager. “Computer simulation of reef growth”. In: *Sedimentology* 39.3 (1992), pages 503–512. DOI: [10.1111/j.1365-3091.1992.tb02130.x](https://doi.org/10.1111/j.1365-3091.1992.tb02130.x). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-3091.1992.tb02130.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3091.1992.tb02130.x> (cited on page 129).
- [54] L. Bosso et al. “Shedding light on the effects of climate change on the potential distribution of *Xylella fastidiosa* in the Mediterranean basin”. In: *Biological Invasions* 23 (8 2016), pages 1759–1768. DOI: [10.1007/s10530-016-1118-1](https://doi.org/10.1007/s10530-016-1118-1) (cited on page 78).
- [55] Luciano Bosso et al. “Shedding light on the effects of climate change on the potential distribution of *Xylella fastidiosa* in the Mediterranean basin”. In: *Biological Invasions* 18.6 (June 2016), pages 1759–1768. ISSN: 1573-1464. DOI: <https://doi.org/10.1007/s10530-016-1118-1> (cited on pages 96, 110).
- [56] Antonio Box et al. “Reduced Antioxidant Response of the Fan Mussel *Pinna nobilis* Related to the Presence of *Haplosporidium pinnae*”. In: *Pathogens* 9.11 (Nov. 2020), page 932. ISSN: 2076-0817. DOI: [10.3390/pathogens9110932](https://doi.org/10.3390/pathogens9110932) (cited on page 9).
- [57] Yves-Marie Bozec, Lorenzo Alvarez-Filip, and Peter J. Mumby. “The dynamics of architectural complexity on coral reefs under climate change”. In: *Global Change Biology* 21.1 (2015), pages 223–235. DOI: <https://doi.org/10.1111/gcb.12698>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.12698>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.12698> (cited on page 124).
- [58] R. H. Bradbury and R. E. Reichelt. “Fractal dimension of a coral reef at ecological scales”. In: *Marine Ecology Progress Series* 10 (1983), pages 169–171. URL: <https://www.int-res.com/articles/meps/10/m010p169.pdf> (cited on page 124).
- [59] C. Bragard et al. “Update of the Scientific Opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory”. In: *EFSA Journal* 17 (5 2019), page 5655. DOI: [10.2903/j.efsa.2019.5665](https://doi.org/10.2903/j.efsa.2019.5665) (cited on pages 78, 79, 86, 88, 90).
- [60] C. Bragard et al. “Status and Prospects of Plant Virus Control Through Interference with Vector Transmission”. In: *Annual Review of Phytopathology* 51.1 (2013), pages 177–201. DOI: [10.1146/annurev-phyto-082712-102346](https://doi.org/10.1146/annurev-phyto-082712-102346) (cited on page 46).
- [61] Fred Brauer. “Models for the spread of universally fatal diseases”. In: *Journal of Mathematical Biology* 28 (1990), pages 451–462. DOI: [10.1007/BF00178328](https://doi.org/10.1007/BF00178328) (cited on page 11).
- [62] Fred Brauer. “Compartmental Models in Epidemiology”. In: *Mathematical Epidemiology*. Edited by Fred Brauer, Pauline van den Driessche, and Jianhong Wu. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pages 19–79. ISBN: 978-3-540-78911-6. DOI: [10.1007/978-3-540-78911-6\\_2](https://doi.org/10.1007/978-3-540-78911-6_2) (cited on pages 36, 46).
- [63] Fred Brauer et al. “Some models for epidemics of vector-transmitted diseases”. In: *Infectious Disease Modelling* 1.1 (2016), pages 79–87. ISSN: 2468-0427. DOI: [10.1016/j.idm.2016.08.001](https://doi.org/10.1016/j.idm.2016.08.001) (cited on pages 46, 48, 49, 72).
- [64] Broder Breckling. “Individual-Based Modelling Potentials and Limitations”. In: *The Scientific World Journal* 2 (Apr. 2002), page 684985. ISSN: 2356-6140. DOI: [10.1100/tsw.2002.179](https://doi.org/10.1100/tsw.2002.179) (cited on page 31).
- [65] M. T. Brewer and M. G. Milgroom. “Phylogeography and population structure of the grape powdery mildew fungus, *Erysiphe necator*, from diverse *Vitis* species”. In: *BMC Evolutionary Biology* 10 (2010), page 268. DOI: [10.1186/1471-2148-10-268](https://doi.org/10.1186/1471-2148-10-268) (cited on pages 78, 89).

- [66] D. Broullón et al. “A global monthly climatology of total alkalinity: a neural network approach”. In: *Earth System Science Data* 11.3 (2019), pages 1109–1127. DOI: [10.5194/essd-11-1109-2019](https://doi.org/10.5194/essd-11-1109-2019). URL: <https://essd.copernicus.org/articles/11/1109/2019/> (cited on pages 135, 139).
- [67] Daniel Broullón, Fiz Pérez, and María Dolores Doval. “Weekly reconstruction of pH and total alkalinity in an upwelling-dominated coastal ecosystem through neural networks (AtpH – NN): The case of Ría de Vigo (NW Spain) between 1992 and 2019”. In: *Biogeosciences Discussions* February (2021), pages 1–36. ISSN: 1726-4170. DOI: [10.5194/bg-2021-33](https://doi.org/10.5194/bg-2021-33) (cited on pages 135, 138, 139).
- [68] J. H. Brown. *Macroecology*. Macroecology. University of Chicago Press, 1995. ISBN: 9780226076157. URL: <https://press.uchicago.edu/ucp/books/book/chicago/M/bo3632297.html> (cited on page 124).
- [69] James H. Brown et al. “The fractal nature of nature: power laws, ecological complexity and biodiversity”. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357.1421 (2002), pages 619–626. DOI: [10.1098/rstb.2001.0993](https://doi.org/10.1098/rstb.2001.0993). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2001.0993>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2001.0993> (cited on page 129).
- [70] James H. Brown et al. “Toward a Metabolic Theory of Ecology”. In: *Ecology* 85.5 (2004), pages 1771–1789. DOI: [10.1890/03-9000](https://doi.org/10.1890/03-9000) (cited on page 23).
- [71] M. Brunetti et al. “A mathematical model for *Xylella fastidiosa* epidemics in the Mediterranean regions. Promoting good agronomic practices for their effective control.” In: *Ecological Modelling* 432 (2020), page 109204. ISSN: 0304-3800. DOI: [10.1016/j.ecolmodel.2020.109204](https://doi.org/10.1016/j.ecolmodel.2020.109204). URL: <https://www.sciencedirect.com/science/article/pii/S030438002030274X> (cited on page 60).
- [72] Colleen A. Burge et al. “Climate change influences on marine infectious diseases: Implications for management and society”. In: *Annual Review of Marine Science* 6 (2014), pages 249–277. ISSN: 19411405. DOI: [10.1146/annurev-marine-010213-135029](https://doi.org/10.1146/annurev-marine-010213-135029) (cited on page 8).
- [73] Eugene M. Burreson and Susan E. Ford. “A review of recent information on the Haplosporidia, with special reference to *Haplosporidium nelsoni* (MSX disease)”. In: *Aquatic Living Resources* 17.4 (Oct. 2004), pages 499–517. ISSN: 0990-7440. DOI: [10.1051/alr:2004056](https://doi.org/10.1051/alr:2004056) (cited on page 9).
- [74] Luca Bütkofer et al. “The problem of scale in predicting biological responses to climate”. In: *Global Change Biology* 26.12 (2020), pages 6657–6666 (cited on page 80).
- [75] A Butler, Nardo Vicente, and Béatrice De Gaulejac. “Ecology of the pterioid bivalves *Pinna bicolor Gmelin* and *Pinna nobilis L.*”. In: *Marine Life* 3.1-2 (1993), pages 37–45. URL: <https://www.semanticscholar.org/paper/Ecology-of-the-pterioid-bivalves-Pinna-bicolor-and-Butler-Vicente-ead2e17a9578a52c274c058bef27084566e12f33> (cited on page 9).
- [76] Miguel Cabanellas-Reboredo et al. “Tracking a mass mortality outbreak of pen shell *Pinna nobilis* populations: A collaborative effort of scientists and citizens”. In: *Scientific Reports* 9.1 (Sept. 2019), page 13355. ISSN: 2045-2322. DOI: [10.1038/s41598-019-49808-4](https://doi.org/10.1038/s41598-019-49808-4) (cited on pages 9, 20, 30).
- [77] Ken Caldeira and Michael E Wickett. “Anthropogenic carbon and ocean pH”. In: *Nature* 425.6956 (2003), pages 365–365. DOI: [10.1038/425365a](https://doi.org/10.1038/425365a) (cited on page 134).
- [78] C. Caminade et al. “Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015”. In: *Proceedings of the National Academy of the USA* 114 (1 2017), pages 119–124. DOI: [10.1073/pnas.1614303114](https://doi.org/10.1073/pnas.1614303114) (cited on page 87).
- [79] Sebastian Candiago et al. “A geospatial inventory of regulatory information for wine protected designations of origin in Europe”. In: *Scientific Data* 9.1 (July 2022), page 394. ISSN: 2052-4463. DOI: <https://doi.org/10.1038/s41597-022-01513-0> (cited on page 104).
- [80] Danielle L. Cantrell et al. “Modeling Pathogen Dispersal in Marine Fish and Shellfish”. In: *Trends in Parasitology* 36.3 (2020), pages 239–249. ISSN: 1471-4922. DOI: [10.1016/j.pt.2019.12.013](https://doi.org/10.1016/j.pt.2019.12.013) (cited on page 8).
- [81] F. Carella et al. “A mycobacterial disease is associated with the silent mass mortality of the pen shell *Pinna nobilis* along the Tyrrhenian coastline of Italy”. In: *Scientific Reports* 9.1 (Feb. 2019), page 2725. ISSN: 2045-2322. DOI: [10.1038/s41598-018-37217-y](https://doi.org/10.1038/s41598-018-37217-y) (cited on page 9).
- [82] Jessica Cariboni et al. “The Role of Sensitivity Analysis in Ecological Modelling”. In: *Ecological Modelling* 203 (Apr. 2007), pages 167–182. DOI: [10.1016/j.ecolmodel.2005.10.045](https://doi.org/10.1016/j.ecolmodel.2005.10.045) (cited on page 26).

- [83] J. M. Carlson and John Doyle. "Highly optimized tolerance: A mechanism for power laws in designed systems". In: *Phys. Rev. E* 60 (2 1999), pages 1412–1427. DOI: [10.1103/PhysRevE.60.1412](https://doi.org/10.1103/PhysRevE.60.1412). URL: <https://link.aps.org/doi/10.1103/PhysRevE.60.1412> (cited on page 129).
- [84] J. M. Carlson and John Doyle. "Highly Optimized Tolerance: Robustness and Design in Complex Systems". In: *Phys. Rev. Lett.* 84 (11 2000), pages 2529–2532. DOI: [10.1103/PhysRevLett.84.2529](https://doi.org/10.1103/PhysRevLett.84.2529). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.84.2529> (cited on page 129).
- [85] R. Carlton et al. *Global Reef Expedition Final Report*. Volume 15. Annapolis, MD: Khaled bin Sultan Living Oceans Foundation, 2021 (cited on page 124).
- [86] Jacob Carstensen and Carlos M. Duarte. "Drivers of pH Variability in Coastal Ecosystems". In: *Environmental Science and Technology* 53.8 (2019), pages 4020–4029. ISSN: 15205851. DOI: [10.1021/acs.est.8b03655](https://doi.org/10.1021/acs.est.8b03655) (cited on pages 134, 135).
- [87] M. Carvajal-Yepes et al. "A global surveillance system for crop diseases: Global preparedness minimizes the risk to food supplies". In: *Science* 364 (6447 2019), pages 1237–1239. DOI: [10.1126/science.aaw1572](https://doi.org/10.1126/science.aaw1572) (cited on page 78).
- [88] Carla Carvalho-Luis, José Manuel Rodrigues, and Luís M Martins. "Dispersion of the bacterium *Xylella fastidiosa* in Portugal". In: *Journal of Agricultural Science and Technology A* 12 (2022), pages 35–41. DOI: <https://doi.org/10.17265/2161-6256/2022.01.005> (cited on page 96).
- [89] A.I. Castillo et al. "Allopatric Plant Pathogen Population Divergence following Disease Emergence". In: *Applied and Environmental Microbiology* 87 (7 2021). DOI: <https://doi.org/10.1128/AEM.02095-203> (cited on page 78).
- [90] C. Castillo-Chavez and H. R. Thieme. "Asymptotically autonomous epidemic models". In: *Mathematical Population Dynamics: Analysis of Heterogeneity, Vol. I, Theory of Epidemics*. Edited by D. Arino O. Axelrod, M. Kimmel, and M. Langlais. Wuerz publishing (Winnipeg), 1995, pages 33–50. URL: <https://ecommons.cornell.edu/bitstream/handle/1813/31834/BU-1248-M.pdf> (cited on page 49).
- [91] Mario Castro et al. "The turning point and end of an expanding epidemic cannot be precisely forecast". In: *Proceedings of the National Academy of Sciences* 117.42 (2020), pages 26190–26196. ISSN: 0027-8424. DOI: [10.1073/pnas.2007868117](https://doi.org/10.1073/pnas.2007868117) (cited on pages 13, 23).
- [92] Gaetano Catanese et al. "*Haplosporidium pinnae* sp. nov., a haplosporidan parasite associated with mass mortalities of the fan mussel, *Pinna nobilis*, in the Western Mediterranean Sea". In: *Journal of Invertebrate Pathology* 157 (2018), pages 9–24. ISSN: 0022-2011. DOI: [10.1016/j.jip.2018.07.006](https://doi.org/10.1016/j.jip.2018.07.006) (cited on pages 9, 20).
- [93] V. Cavalieri et al. "Transmission of *Xylella fastidiosa* Subspecies *Pauca* Sequence Type 53 by Different Insect Species". In: *Insects* 10.10 (2019), page 324. ISSN: 2075-4450. DOI: [10.3390/insects10100324](https://doi.org/10.3390/insects10100324) (cited on pages 61, 67).
- [94] Thomas M. Chaloner, Sarah J. Gurr, and Daniel P. Bebb. "Plant pathogen infection risk tracks global crop yields under climate change". In: *Nature Climate Change* 11.8 (Aug. 2021), pages 710–715. ISSN: 1758-6798. DOI: <https://doi.org/10.1038/s41558-021-01104-8>. URL: <https://doi.org/10.1038/s41558-021-01104-8> (cited on page 96).
- [95] Willy Champenois and A. V. Borges. "Seasonal and interannual variations of community metabolism rates of a *Posidonia oceanica* seagrass meadow". In: *Limnology and Oceanography* 57.1 (2012), pages 347–361. DOI: [10.4319/lo.2012.57.1.0347](https://doi.org/10.4319/lo.2012.57.1.0347) (cited on page 140).
- [96] Jérôme Chave and Simon Levin. "Scale and Scaling in Ecological and Economic Systems". In: *Environmental and Resource Economics* 26.4 (2003), pages 527–557. ISSN: 1573-1502. DOI: [10.1023/B:EARE.0000007348.42742.49](https://doi.org/10.1023/B:EARE.0000007348.42742.49). URL: <https://doi.org/10.1023/B:EARE.0000007348.42742.49> (cited on page 129).
- [97] Y. Chen. "A set of formulae on fractal dimension relations and its application to urban form". In: *Chaos, Solitons & Fractals* 54 (2013), pages 150–158. ISSN: 0960-0779. DOI: [10.1016/j.chaos.2013.07.010](https://doi.org/10.1016/j.chaos.2013.07.010). URL: <https://www.sciencedirect.com/science/article/pii/S0960077913001343> (cited on page 131).
- [98] Y. H. Chew et al. "Mathematical Models Light Up Plant Signaling". In: *The Plant Cell* 26.1 (Jan. 2014), pages 5–20. ISSN: 1040-4651. DOI: [10.1105/tpc.113.120006](https://doi.org/10.1105/tpc.113.120006). eprint: [https://academic.oup.com/plcell/article-pdf/26/1/5/35892793/plcell\\_v26\\_1\\_5.pdf](https://academic.oup.com/plcell/article-pdf/26/1/5/35892793/plcell_v26_1_5.pdf). URL: <https://doi.org/10.1105/tpc.113.120006> (cited on page 60).

- [99] C. Chiyaka et al. “Modeling huanglongbing transmission within a citrus tree”. In: *Proceedings of the National Academy of Sciences* 109.30 (2012), pages 12213–12218. DOI: [10.1073/pnas.1208326109](https://doi.org/10.1073/pnas.1208326109) (cited on page 60).
- [100] S. M. Chmiel and M. C. Wilson. “Estimation of the Lower and Upper Developmental Threshold Temperatures and Duration of the Nymphal Stages of the Meadow Spittlebug, *Philaenus spumarius*”. In: *Environmental Entomology* 8.4 (Aug. 1979), pages 682–685. ISSN: 0046-225X. DOI: [10.1093/ee/8.4.682](https://doi.org/10.1093/ee/8.4.682). eprint: <https://academic.oup.com/ee/article-pdf/8/4/682/18278470/ee8-0682.pdf>. URL: <https://doi.org/10.1093/ee/8.4.682> (cited on page 60).
- [101] Gerardo Chowell. “Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts”. In: *Infectious Disease Modelling* 2.3 (2017), pages 379–398. ISSN: 2468-0427. DOI: [10.1016/j.idm.2017.08.001](https://doi.org/10.1016/j.idm.2017.08.001) (cited on pages 46, 70).
- [102] Ditte M. Christiansen et al. “High-resolution data are necessary to understand the effects of climate on plant population dynamics of a forest herb”. In: *Ecology* 105.1 (2024), e4191. DOI: <https://doi.org/10.1002/ecy.4191>. eprint: <https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.4191>. URL: <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.4191> (cited on page 110).
- [103] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. “Power-Law Distributions in Empirical Data”. In: *SIAM Review* 51.4 (2009), pages 661–703. DOI: [10.1137/070710111](https://doi.org/10.1137/070710111) (cited on page 131).
- [104] Tonya D. Clayton and Robert H. Byrne. “Spectrophotometric seawater pH measurements: total hydrogen ion concentration scale calibration of m-cresol purple and at-sea results”. In: *Deep-Sea Research Part I* 40.10 (1993), pages 2115–2129. ISSN: 09670637. DOI: [10.1016/0967-0637\(93\)90048-8](https://doi.org/10.1016/0967-0637(93)90048-8) (cited on page 142).
- [105] Robert B Cleveland et al. “STL: A seasonal-trend decomposition”. In: *J. Off. Stat* 6.1 (1990), pages 3–73 (cited on page 142).
- [106] S. M. Coakley, H. Scherm, and S. Chakraborty. “Climate change and plant disease management”. In: *Annual Review of Phytopathology* 37 (1999), pages 399–426. DOI: [10.1146/annurev.phyto.37.1.399](https://doi.org/10.1146/annurev.phyto.37.1.399) (cited on pages 79, 87).
- [107] Joao R Coelho and Fernando SM Bezerra. “The effects of temperature change on the infection rate of *Biomphalaria glabrata* with *Schistosoma mansoni*”. en. In: *Memórias do Instituto Oswaldo Cruz* 101 (Mar. 2006), pages 223–224. ISSN: 0074-0276. DOI: [10.1590/S0074-02762006000200016](https://doi.org/10.1590/S0074-02762006000200016) (cited on page 27).
- [108] Steefan Contractor and Moninya Roughan. “Efficacy of Feedforward and LSTM Neural Networks at Predicting and Gap Filling Coastal Ocean Timeseries: Oxygen, Nutrients, and Temperature”. In: *Frontiers in Marine Science* 8 (2021), page 368. ISSN: 2296-7745. DOI: [10.3389/fmars.2021.637759](https://doi.org/10.3389/fmars.2021.637759). URL: <https://www.frontiersin.org/article/10.3389/fmars.2021.637759> (cited on pages 135, 139).
- [109] D. Cornara, D. Bosco, and A. Fereres. “*Philaenus spumarius*: when an old acquaintance becomes a new threat to European agriculture”. In: *Journal of Pest Science* 91.3 (2018), pages 957–972. DOI: [10.1007/s10340-018-0966-0](https://doi.org/10.1007/s10340-018-0966-0) (cited on pages 60, 96).
- [110] D. Cornara et al. “Transmission of *Xylella fastidiosa* to Grapevine by the Meadow Spittlebug”. In: *Phytopathology* 106.11 (2016), pages 1285–1290. DOI: [10.1094/PHYTO-05-16-0202-R](https://doi.org/10.1094/PHYTO-05-16-0202-R) (cited on page 78).
- [111] D. Cornara et al. “Spittlebugs as vectors of *Xylella fastidiosa* in olive orchards in Italy”. In: *Journal of Pest Science* 90.2 (Mar. 2017), pages 521–530. ISSN: 1612-4766. DOI: [10.1007/s10340-016-0793-0](https://doi.org/10.1007/s10340-016-0793-0). URL: <https://doi.org/10.1007/s10340-016-0793-0> (cited on pages 60, 62).
- [112] D. Cornara et al. “Transmission of *Xylella fastidiosa* by naturally infected *Philaenus spumarius* (Hemiptera, Aphrophoridae) to different host plants”. In: *Journal of Applied Entomology* 141.1-2 (2017), pages 80–87. DOI: [10.1111/jen.12365](https://doi.org/10.1111/jen.12365) (cited on page 67).
- [113] D. Cornara et al. “EPG combined with micro-CT and video recording reveals new insights on the feeding behavior of *Philaenus spumarius*”. In: *PLoS One* 13.7 (July 2018), pages 1–20. DOI: [10.1371/journal.pone.0199154](https://doi.org/10.1371/journal.pone.0199154) (cited on pages 60, 71, 78).
- [114] D. Cornara et al. “An overview on the worldwide vectors of *Xylella fastidiosa*”. In: *Entomologia Generalis* 39 (2019), pages 157–181. DOI: [10.1127/entomologica/2019/0811](https://doi.org/10.1127/entomologica/2019/0811) (cited on page 78).
- [115] D. Cornara et al. “Natural areas as reservoir of candidate vectors of *Xylella fastidiosa*”. In: *Bulletin of Insectology* 74 (2021), pages 173–180 (cited on page 60).

- [116] Richard C. Cornes et al. “An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets”. In: *Journal of Geophysical Research: Atmospheres* 123.17 (2018), pages 9391–9409. ISSN: 2169-8996. DOI: <https://doi.org/10.1029/2017JD028200>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2017JD028200> (visited on 06/19/2023) (cited on page 97).
- [117] Á. Corral and Á. González. “Power Law Size Distributions in Geoscience Revisited”. In: *Earth and Space Science* 6.5 (2019), pages 673–697. DOI: [10.1029/2018EA000479](https://doi.org/10.1029/2018EA000479). eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018EA000479>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018EA000479> (cited on page 129).
- [118] G Cossarini, P Lazzari, and C Solidoro. “Spatiotemporal variability of alkalinity in the Mediterranean Sea”. In: *Biogeosciences* 12.6 (2015), pages 1647–1658. DOI: [10.5194/bg-12-1647-2015](https://doi.org/10.5194/bg-12-1647-2015) (cited on page 140).
- [119] Michael Cross and Henry Greenside. *Pattern formation and Dynamics in Nonequilibrium Systems*. Cambridge, UK: Cambridge University Press, 2009. ISBN: 9780511627200. DOI: [10.1017/CBO9780511627200](https://doi.org/10.1017/CBO9780511627200). URL: <https://doi.org/10.1017/CBO9780511627200> (cited on page 130).
- [120] Paul C Cross et al. “Utility of  $R_0$  as a predictor of disease invasion in structured populations”. In: *Journal of The Royal Society Interface* 4.13 (2007), pages 315–324. DOI: [10.1098/rsif.2006.0185](https://doi.org/10.1098/rsif.2006.0185) (cited on page 30).
- [121] Christopher J Crossland et al. “The coastal zone—a domain of global interactions”. In: *Coastal fluxes in the Anthropocene*. Springer, 2005, pages 1–37 (cited on page 134).
- [122] S Culloty and M Mulcahy. “*Bonamia ostreae* in the Native Oyster *Ostrea edulis*”. In: *Marine Environment and Health Series* 29.1649 (2007), pages 1–36. ISSN: 1649-0053. URL: <https://oar.marine.ie/bitstream/handle/10793/269/No%202029%20Marine%20Environment%20and%20Health%20Series.pdf?sequence=1&isAllowed=y> (cited on page 20).
- [123] N. J. Cunniffe et al. “Thirteen challenges in modelling plant diseases”. In: *Epidemics* 10 (2015), pages 6–10. DOI: [10.1016/j.epidem.2014.06.002](https://doi.org/10.1016/j.epidem.2014.06.002) (cited on page 71).
- [124] Andrew A. Cunningham, Peter Daszak, and James L. N. Wood. “One Health, emerging infectious diseases and wildlife: two decades of progress?” In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1725 (2017), page 20160167. DOI: [10.1098/rstb.2016.0167](https://doi.org/10.1098/rstb.2016.0167) (cited on page 30).
- [125] Susana Darriba. “First haplosporidan parasite reported infecting a member of the Superfamily Pinoidea (*Pinna nobilis*) during a mortality event in Alicante (Spain, Western Mediterranean)”. In: *Journal of Invertebrate Pathology* 148 (2017), pages 14–19. ISSN: 0022-2011. DOI: [10.1016/j.jip.2017.05.006](https://doi.org/10.1016/j.jip.2017.05.006) (cited on page 9).
- [126] Charles Darwin. *The Structure and Distribution of Coral Reefs*. London: Smith, Elder & Co., 1874 (cited on page 124).
- [127] Peter Daszak, Andrew A. Cunningham, and Alex D. Hyatt. “Emerging Infectious Diseases of Wildlife—Threats to Biodiversity and Human Health”. In: *Science* 287.5452 (2000), pages 443–449. ISSN: 0036-8075. DOI: [10.1126/science.287.5452.443](https://doi.org/10.1126/science.287.5452.443) (cited on page 30).
- [128] M. P. Daugherty and R. P. P. Almeida. “Understanding How an Invasive Vector Drives Pierce’s Disease Epidemics: Seasonality and Vine-to-Vine Spread”. In: *Phytopathology* 109.2 (2019), pages 277–285. DOI: [10.1094/PHYTO-07-18-0217-FI](https://doi.org/10.1094/PHYTO-07-18-0217-FI). URL: <https://doi.org/10.1094/PHYTO-07-18-0217-FI> (cited on pages 60, 78, 88).
- [129] M. J. Davis, A. H. Purcell, and S. V. Thomson. “Pierce’s disease of grapevines: Isolation of the causal bacterium”. In: *Science* 199 (4324 1978), pages 75–77. DOI: [10.1126/science.199.4324.75](https://doi.org/10.1126/science.199.4324.75) (cited on page 78).
- [130] A. Delbianco et al. “A new resource for research and risk analysis: the updated European food safety authority database of *Xylella* spp. host plant species”. In: *Phytopathology* 109 (2019), pages 213–215. DOI: <https://doi.org/10.1094/PHYTO-09-18-0343-A> (cited on page 78).
- [131] Alice Delbianco et al. “A New Resource for Research and Risk Analysis: The Updated European Food Safety Authority Database of *Xylella* spp. Host Plant Species”. In: *Phytopathology* 109.2 (2019), pages 213–215. DOI: <https://doi.org/10.1094/PHYTO-09-18-0343-A> (cited on page 96).
- [132] Manuel Delgado-Baquerizo et al. “The proportion of soil-borne pathogens increases with warming at the global scale”. In: *Nature Climate Change* 10.6 (June 2020), pages 550–554. DOI: <https://doi.org/10.1038/s41558-020-0759-3> (cited on pages 96, 110).

- [133] N. Denancé et al. “Several subspecies and sequence types are associated with the emergence of *Xylella fastidiosa* in natural settings in France”. In: *Plant Pathology* 66 (2017), pages 1054–1064. DOI: [10.1111/ppa.12695](https://doi.org/10.1111/ppa.12695) (cited on page 78).
- [134] Nicolas Denancé et al. “Several subspecies and sequence types are associated with the emergence of *Xylella fastidiosa* in natural settings in France”. In: *Plant Pathology* 66.7 (2017), pages 1054–1064. DOI: <https://doi.org/10.1111/ppa.12695> (cited on page 96).
- [135] A.G. Dickson and Frank J Millero. “A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media”. In: *Deep Sea Research Part A. Oceanographic Research Papers* 34.10 (1987), pages 1733–1743. DOI: [10.1016/0198-0149\(87\)90021-5](https://doi.org/10.1016/0198-0149(87)90021-5) (cited on page 142).
- [136] Andrew G Dickson. “Standard potential of the reaction:  $\text{AgCl}(\text{s}) + 12\text{H}_2(\text{g}) \rightleftharpoons \text{Ag}(\text{s}) + \text{HCl}(\text{aq})$ , and the standard acidity constant of the ion  $\text{HSO}_4^-$  in synthetic sea water from 273.15 to 318.15 K”. In: *The Journal of Chemical Thermodynamics* 22.2 (1990), pages 113–127. DOI: [10.1016/0021-9614\(90\)90074-Z](https://doi.org/10.1016/0021-9614(90)90074-Z) (cited on page 142).
- [137] Andrew Gilmore Dickson, Christopher L Sabine, and James Robert Christian. *Guide to best practices for ocean CO<sub>2</sub> measurements*. North Pacific Marine Science Organization, 2007. eprint: [https://www.ncei.noaa.gov/access/ocean-carbon-data-system/oceans/Handbook\\_2007.html](https://www.ncei.noaa.gov/access/ocean-carbon-data-system/oceans/Handbook_2007.html) (cited on page 142).
- [138] O. Diekmann, J. A. P. Heesterbeek, and M. G. Roberts. “The construction of next-generation matrices for compartmental epidemic models”. In: *Journal of The Royal Society Interface* 7.47 (2010), pages 873–885. DOI: [10.1098/rsif.2009.0386](https://doi.org/10.1098/rsif.2009.0386) (cited on pages 13, 26, 30, 33, 46, 49, 56, 57, 62, 70, 72).
- [139] Odo Diekmann, Hans Heesterbeek, and Tom Britton. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton: Princeton U.P., 2013. ISBN: 978-0-691-15539-5 (cited on page 10).
- [140] Javier Diez-Sierra et al. “Consistency of the regional response to global warming levels from CMIP5 and CORDEX projections”. In: *Climate Dynamics* (Apr. 25, 2023). ISSN: 1432-0894. DOI: <https://doi.org/10.1007/s00382-023-06790-y>. URL: <https://doi.org/10.1007/s00382-023-06790-y> (visited on 06/12/2023) (cited on page 97).
- [141] E.J. Dlugokencky et al. *Atmospheric Nitrous Oxide Dry Air Mole Fractions from the NOAA GML Carbon Cycle Cooperative Global Air Sampling Network, 1997-2020*. Technical report. NOAA, July 2021. DOI: [10.15138/53g1-x417](https://doi.org/10.15138/53g1-x417) (cited on page 142).
- [142] Scott C Doney et al. “Ocean acidification: the other CO<sub>2</sub> problem”. In: *Annual Review of Marine Science* 1 (2009), pages 169–192. DOI: [10.1146/annurev.marine.010908.163834](https://doi.org/10.1146/annurev.marine.010908.163834) (cited on page 134).
- [143] P. van den Driessche and James Watmough. “Further Notes on the Basic Reproduction Number”. In: *Mathematical Epidemiology*. Edited by Fred Brauer, Pauline van den Driessche, and Jianhong Wu. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pages 159–178. ISBN: 978-3-540-78911-6. DOI: [10.1007/978-3-540-78911-6\\_6](https://doi.org/10.1007/978-3-540-78911-6_6) (cited on page 13).
- [144] André W. Droxler and Stéphan J. Jorry. “The Origin of Modern Atolls: Challenging Darwin’s Deeply Ingrained Theory”. In: *Annual Review of Marine Science* 13.1 (2021), pages 537–573. DOI: [10.1146/annurev-marine-122414-034137](https://doi.org/10.1146/annurev-marine-122414-034137). eprint: <https://doi.org/10.1146/annurev-marine-122414-034137> (cited on page 124).
- [145] Carlos M. Duarte et al. “Is Ocean Acidification an Open-Ocean Syndrome? Understanding Anthropogenic Impacts on Seawater pH”. In: *Estuaries and Coasts* 36.2 (2013), pages 221–236. ISSN: 15592723. DOI: [10.1007/s12237-013-9594-3](https://doi.org/10.1007/s12237-013-9594-3) (cited on page 135).
- [146] Nicolas Dubos et al. “Choice of climate data influences predictions for current and future global invasion risks for two *Phelsuma* geckos”. In: *Biological Invasions* 25 (2023), pages 2929–2948. DOI: <https://doi.org/10.1007/s10530-023-03082-8> (cited on page 110).
- [147] Joan Dudney et al. “Nonlinear shifts in infectious rust disease due to climate change”. In: *Nature Communications* 12.1 (Aug. 2021), page 5102. ISSN: 2041-1723. DOI: <https://doi.org/10.1038/s41467-021-25182-6>. URL: <https://doi.org/10.1038/s41467-021-25182-6> (cited on pages 96, 110, 116).
- [148] Tyler D. Eddy et al. “Global Decline in Capacity of Coral Reefs to Provide Ecosystem Services”. In: *One Earth* 4.9 (2021), pages 1278–1285. DOI: [10.1016/j.oneear.2021.08.016](https://doi.org/10.1016/j.oneear.2021.08.016) (cited on page 124).
- [149] EEA. “State and pressures of the marine and coastal Mediterranean environment”. In: *European Environment Agency* 5 (1999), pages 1–44. eprint: <https://www.eea.europa.eu/publications/medsea/download> (cited on page 134).

- [150] Morgan E. Eisenlord et al. “Ochre star mortality during the 2014 wasting disease epizootic: role of population size structure and temperature”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1689 (2016), page 20150212. DOI: [10.1098/rstb.2015.0212](https://doi.org/10.1098/rstb.2015.0212) (cited on page 30).
- [151] Phil Elson et al. *SciTools/cartopy: v0.21.1*. Version v0.21.1. Dec. 2022. DOI: [10.5281/zenodo.7430317](https://doi.org/10.5281/zenodo.7430317). URL: <https://doi.org/10.5281/zenodo.7430317> (cited on page 100).
- [152] Lourdes Esteva and Cristobal Vargas. “Analysis of a dengue disease transmission model”. In: *Mathematical Biosciences* 150.2 (1998), pages 131–151. ISSN: 0025-5564. DOI: [10.1016/S0025-5564\(98\)10003-2](https://doi.org/10.1016/S0025-5564(98)10003-2) (cited on page 49).
- [153] European Union, Copernicus Land Monitoring Service 2018, European Environment Agency (EEA). <https://land.copernicus.eu/pan-european/corine-land-cover>. 2018 (cited on page 93).
- [154] H. Feil, W. S. Feil, and A. H. Purcell. “Effects of Date of Inoculation on the Within-Plant Movement of *Xylella fastidiosa* and Persistence of Pierce’s Disease Within Field Grapevines”. In: *Phytopathology* 93 (2003), pages 244–251. DOI: [10.1094/PHYTO.2003.93.2.244](https://doi.org/10.1094/PHYTO.2003.93.2.244) (cited on pages 78, 80, 81, 90).
- [155] H. Feil and A. H. Purcell. “Temperature-Dependent Growth and Survival of *Xylella fastidiosa* in Vitro and in Potted Grapevines”. In: *Plant Disease* 85 (12 2001), pages 1230–1234. DOI: [10.1094/PDIS.2001.85.12.1230](https://doi.org/10.1094/PDIS.2001.85.12.1230) (cited on pages 78–80, 87, 89, 90).
- [156] Thomas Fellmann et al. “Major challenges of integrating agriculture into climate change mitigation policy frameworks”. In: *Mitigation and Adaptation Strategies for Global Change* 23 (2018), pages 451–468. DOI: <https://doi.org/10.1007/s11027-017-9743-2> (cited on page 106).
- [157] H. Feng and M. Zhang. “Global land moisture trends: drier in dry and wetter in wet over land”. In: *Scientific Reports* 5 (2016), page 18018. DOI: <https://doi.org/10.1038/srep18018> (cited on page 88).
- [158] A. Fierro, A. Liccardo, and F. Porcelli. “A lattice model to manage the vector and the infection of the *Xylella fastidiosa* on olive trees”. In: *Scientific Reports* 9.1 (June 2019), page 8723. ISSN: 2045-2322. DOI: [10.1038/s41598-019-44997-4](https://doi.org/10.1038/s41598-019-44997-4). URL: <https://doi.org/10.1038/s41598-019-44997-4> (cited on pages 61, 63).
- [159] J. A. N. Filipe and M. M. Maule. “Analytical methods for predicting the behaviour of population models with general spatial interactions”. In: *Mathematical Biosciences* 183 (2003), pages 15–35. DOI: [10.1016/S0025-5564\(02\)00224-9](https://doi.org/10.1016/S0025-5564(02)00224-9) (cited on page 36).
- [160] J. A. N. Filipe and M. M. Maule. “Effects of dispersal mechanisms on spatio-temporal development of epidemics”. In: *Journal of Theoretical Biology* 226 (2004), pages 125–141. DOI: [10.1016/S0022-5193\(03\)00278-9](https://doi.org/10.1016/S0022-5193(03)00278-9) (cited on page 36).
- [161] Susana Flecha et al. “Trends of pH decrease in the Mediterranean Sea through high frequency observational data: Indication of ocean acidification in the basin”. In: *Scientific Reports* 5.April (2015), pages 1–8. ISSN: 20452322. DOI: [10.1038/srep16770](https://doi.org/10.1038/srep16770). URL: <http://dx.doi.org/10.1038/srep16770> (cited on pages 134, 135, 139).
- [162] Susana Flecha et al. “Decadal acidification in Atlantic and Mediterranean water masses exchanging at the Strait of Gibraltar”. In: *Scientific Reports* 9.1 (2019), pages 1–11. ISSN: 20452322. DOI: [10.1038/s41598-019-52084-x](https://doi.org/10.1038/s41598-019-52084-x) (cited on page 134).
- [163] Susana Flecha et al. “pH trends and seasonal cycle in the coastal Balearic Sea reconstructed through machine learning”. In: *Scientific Reports* 12.1 (July 2022), page 12956. ISSN: 2045-2322. DOI: [10.1038/s41598-022-17253-5](https://doi.org/10.1038/s41598-022-17253-5). URL: <https://doi.org/10.1038/s41598-022-17253-5> (cited on pages 17, 133).
- [164] R. Fletcher. “Newton-Like Methods”. In: *Practical Methods of Optimization*. John Wiley & Sons, Ltd, 2013. Chapter 3, pages 44–79. ISBN: 9781118723203. DOI: [10.1002/9781118723203.ch3](https://doi.org/10.1002/9781118723203.ch3) (cited on page 20).
- [165] Helen N. Fones et al. “Threats to global food security from emerging fungal and oomycete crop pathogens”. In: *Nature Food* 1.6 (June 2020), pages 332–342. ISSN: 2662-1355. DOI: [10.1038/s43016-020-0075-0](https://doi.org/10.1038/s43016-020-0075-0) (cited on page 110).
- [166] Marine Fourrier et al. “A Regional Neural Network Approach to Estimate Water-Column Nutrient Concentrations and Carbonate System Variables in the Mediterranean Sea: CANYON-MED”. In: *Frontiers in Marine Science* 7 (2020), page 620. ISSN: 2296-7745. DOI: [10.3389/fmars.2020.00620](https://doi.org/10.3389/fmars.2020.00620). URL: <https://www.frontiersin.org/article/10.3389/fmars.2020.00620> (cited on pages 135, 138).
- [167] J. H. Freitag. “Host range of the Pierce’s disease virus of grapes as determined by insect transmission”. In: *Phytopathology* 41.10 (1951), pages 920–934 (cited on pages 60, 61).

- [168] P. Friedlingstein et al. “Global Carbon Budget 2021”. In: *Earth System Science Data Discussions* 2021 (2021), pages 1–191. DOI: [10.5194/essd-2021-386](https://doi.org/10.5194/essd-2021-386). URL: <https://essd.copernicus.org/preprints/essd-2021-386/> (cited on page 134).
- [169] T. Friedrich and A. Oschlies. “Basin-scale pCO<sub>2</sub> maps estimated from ARGO gfloat data: A model study”. In: *Journal of Geophysical Research: Oceans* 114.10 (2009), pages 1–9. ISSN: 21699291. DOI: [10.1029/2009JC005322](https://doi.org/10.1029/2009JC005322) (cited on page 135).
- [170] SM Fry, RD Milholland, et al. “Multiplication and translocation of *Xylella fastidiosa* in petioles and stems of grapevine resistant, tolerant, and susceptible to Pierce’s disease.” In: *Phytopathology* 80.1 (1990), pages 61–65 (cited on pages 80, 87).
- [171] L. C. Galvez et al. *The Threat of Pierce’s disease to Midwest Wine and Table Grapes*. 2010. DOI: [10.1094/APSnetFeature-2010-1015](https://doi.org/10.1094/APSnetFeature-2010-1015) (cited on page 81).
- [172] Jorge Garcia Molinos and Chris Brown. *JorGarMol/VoCC: First release VoCC*. Version v1.0.0. Aug. 2019. DOI: <https://doi.org/10.5281/zenodo.3382092> (cited on page 100).
- [173] Jorge García Molinos et al. “VoCC: An R package for calculating the velocity of climate change and related climatic metrics”. In: *Methods in Ecology and Evolution* 10.12 (2019), pages 2195–2202. DOI: <https://doi.org/10.1111/2041-210X.13295> (cited on page 100).
- [174] Jesús García-Lafuente et al. “Hotter and Weaker Mediterranean Outflow as a Response to Basin-Wide Alterations”. In: *Frontiers in Marine Science* 8 (2021). DOI: [10.3389/fmars.2021.613444](https://doi.org/10.3389/fmars.2021.613444) (cited on page 134).
- [175] J.R. García-March et al. “Can we save a marine species affected by a highly infective, highly lethal, waterborne disease from extinction?” In: *Biological Conservation* 243 (2020), page 108498. ISSN: 0006-3207. DOI: [10.1016/j.biocon.2020.108498](https://doi.org/10.1016/j.biocon.2020.108498) (cited on pages 9, 20).
- [176] R Garms, JF Walsh, and JB Davies. “Studies on the reinvansion of the *Onchocerciasis* Control Programme in the Volta River Basin by *Simulium damnosum* sI with emphasis on the south-western areas”. In: *Tropenmedizin und Parasitologie* 30.3 (1979), pages 345–362 (cited on page 46).
- [177] K. A. Garrett et al. “Complexity in climate-change impacts: an analytical framework for effects mediated by plant disease”. In: *Plant Pathology* 60.1 (2011), pages 15–30. DOI: <https://doi.org/10.1111/j.1365-3059.2010.02409.x> (cited on pages 96, 110).
- [178] Frédéric Gazeau et al. “Whole-system metabolism and CO<sub>2</sub> fluxes in a Mediterranean Bay dominated by seagrass beds (Palma Bay, NW Mediterranean)”. In: *Biogeosciences* 2.1 (2005), pages 43–60. DOI: [10.5194/bg-2-43-2005](https://doi.org/10.5194/bg-2-43-2005) (cited on page 141).
- [179] GBIF.Org. (13 July 2023) *GBIF Occurrence Download*. 2023. DOI: [10.15468/DL.SQNJ3T](https://doi.org/10.15468/DL.SQNJ3T). URL: <https://www.gbif.org/occurrence/download/0031995-231120084113126> (cited on pages 116, 117).
- [180] GBIF.org. (13 July 2022) *GBIF Occurrence Download*. DOI: <https://doi.org/10.15468/dl.k5ezjg> (cited on page 99).
- [181] H. Ge, K. Xu, and Z. Ghahramani. “Turing: a language for flexible probabilistic inference”. In: *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9–11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*. 2018, pages 1682–1690. URL: <http://proceedings.mlr.press/v84/ge18b.html> (cited on page 63).
- [182] E. Gemayel et al. “Climatological variations of total alkalinity and total dissolved inorganic carbon in the Mediterranean Sea surface waters”. In: *Earth System Dynamics* 6.2 (2015), pages 789–800. DOI: [10.5194/esd-6-789-2015](https://doi.org/10.5194/esd-6-789-2015). URL: <https://esd.copernicus.org/articles/6/789/2015/> (cited on pages 140, 142).
- [183] E. E. George et al. “Space-filling and benthic competition on coral reefs”. In: *PeerJ* 9 (2021), e11213. DOI: [10.7717/peerj.11213](https://doi.org/10.7717/peerj.11213) (cited on pages 126, 129).
- [184] Daniel T. Gillespie. “Exact stochastic simulation of coupled chemical reactions”. In: *The Journal of Physical Chemistry* 81.25 (Dec. 1977), pages 2340–2361. ISSN: 0022-3654. DOI: [10.1021/j100540a008](https://doi.org/10.1021/j100540a008) (cited on page 32).
- [185] Sean Gillies et al. *Shapely: manipulation and analysis of geometric objects*. 2007–. URL: <https://github.com/shapely/shapely> (cited on pages 130, 131).

- [186] Christopher A. Gilligan and Frank van den Bosch. “Epidemiological Models for Invasion and Persistence of Pathogens”. In: *Annual Review of Phytopathology* 46 (2008), pages 385–418. DOI: [10.1146/annurev.phyto.45.062806.094357](https://doi.org/10.1146/annurev.phyto.45.062806.094357) (cited on page 30).
- [187] James F. Gillooly et al. “Effects of Size and Temperature on Metabolic Rate”. In: *Science* 293.5538 (2001), pages 2248–2251. ISSN: 0036-8075. DOI: [10.1126/science.1061967](https://doi.org/10.1126/science.1061967) (cited on page 27).
- [188] A. Giménez-Romero. <http://pdrisk.ifisc.uib-csic.es/>. 2021 (cited on pages 79, 83, 85, 87–89, 104, 105).
- [189] A. Giménez-Romero. <https://github.com/agimenezromero/PierceDisease-GlobalRisk-Predictions>. 2022 (cited on page 118).
- [190] A. Giménez-Romero. “Coral Reef Analysis”. In: *GitHub Repo* (2022). URL: <https://github.com/agimenezromero/Coral-Reef-Analysis> (cited on pages 130, 131).
- [191] Alex Giménez-Romero. “Coastal pH variability reconstructed through neural networks the coastal Balearic Sea case study”. In: *GitHub Repo* (2021). eprint: <https://github.com/agimenezromero/Coastal-pH-variability-reconstructed-through-neural-networks-the-coastal-Balearic-Sea-case-study>. URL: <https://github.com/agimenezromero/Coastal-pH-variability-reconstructed-through-neural-networks-the-coastal-Balearic-Sea-case-study> (cited on page 143).
- [192] Alex Giménez-Romero. “A compartmental model for *Xylella fastidiosa* related diseases”. In: *GitHub Repo* (2022). URL: <https://github.com/agimenezromero/Spatial-effects-in-parasite-induced-marine-diseases-of-immobile-hosts> (cited on pages 32, 63, 143).
- [193] Alex Giménez-Romero, Manuel A. Matias, and Carlos M. Duarte. *Universal spatial properties of coral reefs*. Zenodo, Oct. 2023 (cited on page 124).
- [194] Alex Giménez-Romero et al. “Global predictions for the risk of establishment of Pierce’s disease of grapevines”. In: *Communications Biology* 5.1 (Dec. 2022), page 1389. ISSN: 2399-3642. DOI: [10.1038/s42003-022-04358-w](https://doi.org/10.1038/s42003-022-04358-w). URL: <https://doi.org/10.1038/s42003-022-04358-w> (cited on pages 17, 60, 77, 96–99, 102, 105, 110, 111, 117–119).
- [195] Àlex Giménez-Romero, Rosa Flaquer-Galmés, and Manuel A. Matías. “Vector-borne diseases with nonstationary vector populations: The case of growing and decaying populations”. In: *Phys. Rev. E* 106 (5 Nov. 2022), page 054402. DOI: [10.1103/PhysRevE.106.054402](https://doi.org/10.1103/PhysRevE.106.054402). URL: <https://link.aps.org/doi/10.1103/PhysRevE.106.054402> (cited on pages 17, 45, 62, 70, 72, 105).
- [196] Àlex Giménez-Romero, Eduardo Moralejo, and Manuel A. Matías. “High-resolution climate data reveals increased risk of Pierce’s Disease for grapevines worldwide”. In: *bioRxiv* (2024). DOI: [10.1101/2024.03.06.583743](https://doi.org/10.1101/2024.03.06.583743). eprint: <https://www.biorxiv.org/content/early/2024/03/11/2024.03.06.583743.full.pdf>. URL: <https://www.biorxiv.org/content/early/2024/03/11/2024.03.06.583743> (cited on pages 18, 109).
- [197] Àlex Giménez-Romero, Eduardo Moralejo, and Manuel A. Matías. “A Compartmental Model for *Xylella fastidiosa* Diseases with Explicit Vector Seasonal Dynamics”. In: *Phytopathology®* 113.9 (2023). PMID: 36774557, pages 1686–1696. DOI: [10.1094/PHYTO-11-22-0428-V](https://doi.org/10.1094/PHYTO-11-22-0428-V). eprint: <https://doi.org/10.1094/PHYTO-11-22-0428-V>. URL: <https://doi.org/10.1094/PHYTO-11-22-0428-V> (cited on pages 17, 59).
- [198] Àlex Giménez-Romero et al. “Modelling parasite-produced marine diseases: The case of the mass mortality event of *Pinna nobilis*”. In: *Ecological Modelling* 459 (2021), page 109705. ISSN: 0304-3800. DOI: <https://doi.org/10.1016/j.ecolmodel.2021.109705>. URL: <https://www.sciencedirect.com/science/article/pii/S030438002100260X> (cited on pages 17, 7, 30, 31, 33–35, 40, 47).
- [199] Àlex Giménez-Romero et al. “Spatial effects in parasite-induced marine diseases of immobile hosts”. In: *Royal Society Open Science* 9.8 (2022), page 212023. DOI: [10.1098/rsos.212023](https://doi.org/10.1098/rsos.212023). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.212023>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsos.212023> (cited on pages 17, 29).
- [200] Àlex Giménez-Romero et al. “Contrasting Patterns of Pierce’s Disease Risk in European Vineyards Under Global Warming”. In: *bioRxiv* (2023). DOI: [10.1101/2023.07.17.549293](https://doi.org/10.1101/2023.07.17.549293). eprint: <https://www.biorxiv.org/content/early/2023/07/18/2023.07.17.549293.full.pdf>. URL: <https://www.biorxiv.org/content/early/2023/07/18/2023.07.17.549293> (cited on pages 17, 95, 113, 119).

- [201] F Giorgi. “Climate change hot-spots”. In: *Geophysical Research Letters* 33.8 (2006), page L08707. DOI: [10.1029/2006GL025734](https://doi.org/10.1029/2006GL025734) (cited on page 134).
- [202] Filippo Giorgi and William J. Gutowski. “Regional dynamical downscaling and the CORDEX initiative”. In: *Annual Review of Environment and Resources* 40.1 (2015), pages 467–490. DOI: <https://doi.org/10.1146/annurev-environ-102014-021217>. URL: <http://dx.doi.org/10.1146/annurev-environ-102014-021217> (visited on 06/07/2016) (cited on page 97).
- [203] M. Godefroid et al. “*Xylella fastidiosa*: climate suitability of European continent”. In: *Scientific Reports* 9 (2019), page 8844. DOI: [10.1038/s41598-019-45365-y](https://doi.org/10.1038/s41598-019-45365-y) (cited on pages 78, 86).
- [204] M. Godefroid et al. “Climate tolerances of *Philaenus spumarius* should be considered in risk assessment of disease outbreaks related to *Xylella fastidiosa*”. In: *Journal of Pest Science* 1 (2021), page 1. DOI: [10.1007/s10340-021-01413-z](https://doi.org/10.1007/s10340-021-01413-z) (cited on pages 78, 79, 86, 88, 89, 93).
- [205] M. Godefroid et al. “Climate tolerances of *Philaenus spumarius* should be considered in risk assessment of disease outbreaks related to *Xylella fastidiosa*”. In: *Journal of Pest Science* 95.2 (Mar. 2022), pages 855–868. ISSN: 1612-4766. DOI: <https://doi.org/10.1007/s10340-021-01413-z> (cited on pages 96, 99, 100, 110).
- [206] Martin Godefroid et al. “*Xylella fastidiosa*: climate suitability of European continent”. In: *Scientific Reports* 9.1 (2019), page 8844. DOI: <https://doi.org/10.1038/s41598-019-45365-y> (cited on page 96).
- [207] Martin Godefroid et al. “Forecasting future range shifts of *Xylella fastidiosa* under climate change”. In: *Plant Pathology* 71.9 (2022), pages 1839–1848. DOI: <https://doi.org/10.1111/ppa.13637> (cited on pages 96, 110, 117).
- [208] Stefano Goffredo and Zvy Dubinsky. *The Mediterranean Sea: Its history and present challenges*. Springer Science & Business Media, 2013 (cited on page 135).
- [209] M. Gomila et al. “Draft genome resources of two strains of *Xylella fastidiosa* XYL1732/17 and XYL2055/17 isolated from Mallorca vineyards”. In: *Phytopathology* 109 (2019), pages 222–224. DOI: [10.1094/PHYTO-08-18-0298-A](https://doi.org/10.1094/PHYTO-08-18-0298-A) (cited on page 78).
- [210] Margarita Gomila et al. “Draft genome resources of two strains of *Xylella fastidiosa* XYL1732/17 and XYL2055/17 isolated from Mallorca vineyards”. In: *Phytopathology* 109.2 (2019), pages 222–224. DOI: <https://doi.org/10.1094/PHYTO-08-18-0298-A> (cited on page 117).
- [211] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016 (cited on page 135).
- [212] Alex Graves and Jürgen Schmidhuber. “Framewise phoneme classification with bidirectional LSTM and other neural network architectures”. In: *Neural networks* 18.5-6 (2005), pages 602–610. DOI: [10.1016/j.neunet.2005.06.042](https://doi.org/10.1016/j.neunet.2005.06.042) (cited on page 139).
- [213] Luke Gregor et al. “A comparative assessment of the uncertainties of global surface ocean CO<sub>2</sub> estimates using a machine-learning ensemble (CSIR-ML6 version 2019a)—have we hit the wall?” In: *Geoscientific Model Development* 12.12 (2019), pages 5113–5136. DOI: [10.5194/gmd-12-5113-2019](https://doi.org/10.5194/gmd-12-5113-2019) (cited on page 139).
- [214] Volker Grimm and Steven F Railsback. *Individual-based Modeling and Ecology*. Princeton (NJ): Princeton University Press, 2005. ISBN: 978-0691096650 (cited on pages 30, 71).
- [215] B. R. Gruber and M. P. Daugherty. “The biology of xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology”. In: *Plant Pathology* 62 (1 2012), pages 194–204. DOI: [10.1111/j.1365-3059.2012.02611.x](https://doi.org/10.1111/j.1365-3059.2012.02611.x) (cited on pages 78, 88).
- [216] Ximing Guo and Susan E. Ford. “Infectious diseases of marine molluscs and host responses as revealed by genomic tools”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1689 (2016), page 20150206. DOI: [10.1098/rstb.2015.0206](https://doi.org/10.1098/rstb.2015.0206) (cited on page 30).
- [217] D. Hail et al. “Detection and analysis of the bacterium, *Xylella fastidiosa*, in glassy-winged sharpshooter, *Homalodisca vitripennis*, populations in Texas”. In: *Journal of Insect Science* 10.168 (2008), pages 1–11 (cited on page 92).
- [218] Lee Hannah et al. “Climate change, wine, and conservation”. In: *Proceedings of the National Academy of Sciences* 110.17 (2013), pages 6907–6912. DOI: <https://doi.org/10.1073/pnas.1210127110> (cited on page 106).
- [219] C. Drew Harvell et al. “Climate Warming and Disease Risks for Terrestrial and Marine Biota”. In: *Science* 296.5576 (2002), pages 2158–2162. DOI: [10.1126/science.1063699](https://doi.org/10.1126/science.1063699) (cited on page 110).

- [220] C. Drew Harvell et al. “Climate warming and disease risks for terrestrial and marine biota”. In: *Science* 296.5576 (2002), pages 2158–2162. DOI: <https://doi.org/10.1126/science.1063699> (cited on page 96).
- [221] Drew Harvell et al. “The rising tide of ocean diseases: Unsolved problems and research priorities”. In: *Frontiers in Ecology and the Environment* 2.7 (2004), pages 375–382. ISSN: 15409309. DOI: [10.1890/1540-9295\(2004\)002\[0375:TRTOOD\]2.0.CO;2](10.1890/1540-9295(2004)002[0375:TRTOOD]2.0.CO;2) (cited on pages 8, 10).
- [222] Harold H Haskin and Jay D Andrews. “Uncertainties and speculations about the life cycle of the eastern oyster pathogen *Haplosporidium nelsoni* (MSX)”. In: *Special Publication (American Fisheries Society)* 18 (1988). URL: <https://ir.library.oregonstate.edu/downloads/pn89df056?locale=en> (cited on page 20).
- [223] Abed El Rahman Hassoun et al. “Acidification of the Mediterranean Sea from anthropogenic carbon penetration”. In: *Deep-Sea Research Part I: Oceanographic Research Papers* 102 (2015), pages 1–15. ISSN: 09670637. DOI: <10.1016/j.dsr.2015.04.005>. URL: <http://dx.doi.org/10.1016/j.dsr.2015.04.005> (cited on page 134).
- [224] Abed El Rahman Hassoun et al. “Modeling of the total alkalinity and the total inorganic carbon in the Mediterranean Sea”. In: *Journal of Water Resources and Ocean Science* (2015), pages 24–32. DOI: <10.11648/j.wros.20150401.14> (cited on page 140).
- [225] I E Hendriks et al. “Photosynthetic activity buffers ocean acidification in seagrass meadows”. In: *Biogeosciences* 11.2 (2014), pages 333–346. DOI: <10.5194/bg-11-333-2014>. URL: <https://bg.copernicus.org/articles/11/333/2014/> (cited on page 135).
- [226] Iris E. Hendriks et al. “Biological mechanisms supporting adaptation to ocean acidification in coastal ecosystems”. In: *Estuarine, Coastal and Shelf Science* 152 (2015), A1–A8. DOI: <10.1016/j.ecss.2014.07.019> (cited on page 134).
- [227] Iris E. Hendriks et al. “Seagrass Meadows Modify Drag Forces on the Shell of the Fan Mussel *Pinna nobilis*”. In: *Estuaries and Coasts* 34.1 (Jan. 2011), pages 60–67. ISSN: 1559-2731. DOI: <10.1007/s12237-010-9309-y> (cited on page 9).
- [228] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. “Recurrent Neural Networks for Time Series Forecasting: Current status and future directions”. In: *International Journal of Forecasting* 37.1 (2021), pages 388–427. ISSN: 0169-2070. DOI: <j.ijforecast.2020.06.008> (cited on page 135).
- [229] P.M. Hine. “The ecology of Bonamia and decline of bivalve molluscs”. In: *New Zealand Journal of Ecology* 20.1 (1996), pages 109–116. URL: <https://newzealandecology.org/nzje/1993> (cited on page 20).
- [230] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (Nov. 1997), pages 1735–1780. ISSN: 0899-7667. DOI: <10.1162/neco.1997.9.8.1735> (cited on page 135).
- [231] M. S. Hoddle. “The potential adventive geographic range of glassy-winged sharpshooter, *Homalodisca coagulata* and the grape pathogen *Xylella fastidiosa*: implications for California and other grape growing regions of the world”. In: *Crop Protection* 23 (8 2004), pages 691–699. DOI: <10.1016/j.cropro.2003.11.017> (cited on page 78).
- [232] Gretchen E. Hofmann et al. “High-frequency dynamics of ocean pH: A multi-ecosystem comparison”. In: *PLoS ONE* 6.12 (2011). ISSN: 19326203. DOI: <10.1371/journal.pone.0028983> (cited on pages 134, 135).
- [233] M. D. Homan and A. Gelman. “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo”. In: *J. Mach. Learn. Res.* 15.1 (Jan. 2014), pages 1593–1623. ISSN: 1532-4435. URL: <https://dl.acm.org/doi/10.5555/2627435.2638586> (cited on page 63).
- [234] D. L. Hopkins and A. H. Purcell. “*Xylella fastidiosa*: Cause of Pierce’s Disease of Grapevine and Other Emergent Diseases”. In: *Plant Disease* 86.10 (10 Oct. 2002), pages 1056–1066. DOI: <10.1094/PDIS.2002.86.10.1056> (cited on pages 60, 78, 79, 81, 110).
- [235] Weijie Huang et al. “Bacterial Vector-Borne Plant Diseases: Unanswered Questions and Future Directions”. In: *Molecular Plant* 13.10 (2020), pages 1379–1393. ISSN: 1674-2052. DOI: <10.1016/j.molp.2020.08.010> (cited on page 46).
- [236] Y. Huang, L. Yang, and Z. Fu. “Reconstructing coupled time series in climate systems using three kinds of machine-learning methods”. In: *Earth System Dynamics* 11.3 (2020), pages 835–853. DOI: <10.5194/esd-11-835-2020> (cited on page 135).

- [237] Maialen Iturbide et al. “A framework for species distribution modelling with improved pseudo-absence generation”. In: *Ecological Modelling* 312 (Sept. 24, 2015), pages 166–174. ISSN: 0304-3800. DOI: <https://doi.org/10.1016/j.ecolmodel.2015.05.018>. URL: <https://www.sciencedirect.com/science/article/pii/S030438001500215X> (visited on 08/01/2022) (cited on page 99).
- [238] Maialen Iturbide et al. *Repository supporting the implementation of FAIR principles in the IPCC-WGI Atlas*. Aug. 9, 2021. DOI: <https://doi.org/10.5281/zenodo.5171760>. URL: <https://zenodo.org/record/5171760> (visited on 10/27/2021) (cited on page 97).
- [239] IUCN. “IUCN. The IUCN Red List of Threatened Species. Version 2019-3”. In: (2019). URL: <http://www.iucnredlist.org> (cited on page 9).
- [240] Daniela Jacob et al. “Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community”. In: *Regional Environmental Change* 20.2 (Apr. 23, 2020), page 51. ISSN: 1436-378X. DOI: <https://doi.org/10.1007/s10113-020-01606-9>. URL: <https://doi.org/10.1007/s10113-020-01606-9> (visited on 03/28/2022) (cited on pages 97, 100).
- [241] M. J. Jeger and C. Bragard. “The Epidemiology of *Xylella fastidiosa*; A Perspective on Current Knowledge and Framework to Investigate Plant Host–Vector–Pathogen Interactions”. In: *Phytopathology* 109.2 (2019), pages 200–209. DOI: [10.1094/PHYTO-07-18-0239-FI](https://doi.org/10.1094/PHYTO-07-18-0239-FI) (cited on pages 60, 79, 87).
- [242] M. J. Jeger, L. V. Madden, and F Van Den Bosch. “Plant virus epidemiology: Applications and prospects for mathematical modeling and analysis to improve understanding and disease control”. In: *Plant Disease* 102.5 (2018), pages 837–854. DOI: [10.1094/PDIS-04-17-0612-FE](https://doi.org/10.1094/PDIS-04-17-0612-FE) (cited on page 71).
- [243] M. J. Jeger et al. “A model for analysing plant-virus transmission characteristics and epidemic development”. In: *Mathematical Medicine and Biology: A Journal of the IMA* 15.1 (Mar. 1998), pages 1–18. ISSN: 1477-8599. DOI: [10.1093/imammb/15.1.1](https://doi.org/10.1093/imammb/15.1.1). eprint: <https://academic.oup.com/imammb/article-pdf/15/1/1/1995834/15-1-1.pdf>. URL: <https://doi.org/10.1093/imammb/15.1.1> (cited on page 60).
- [244] M. J. Jeger et al. “Epidemiology of insect-transmitted plant viruses: modelling disease dynamics and control interventions”. In: *Physiological Entomology* 29.3 (2004), pages 291–304. DOI: [10.1111/j.0307-6962.2004.00394.x](https://doi.org/10.1111/j.0307-6962.2004.00394.x). eprint: <https://resjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/j.0307-6962.2004.00394.x>. URL: <https://resjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.0307-6962.2004.00394.x> (cited on page 60).
- [245] Michael Jeger and Claude Bragard. “The Epidemiology of *Xylella fastidiosa*; A Perspective on Current Knowledge and Framework to Investigate Plant Host–Vector–Pathogen Interactions”. In: *Phytopathology* 109.2 (2019), pages 200–209. DOI: <https://doi.org/10.1094/PHYTO-07-18-0239-FI> (cited on pages 96, 110).
- [246] Alberto Jiménez-Valverde. “Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling”. In: *Global Ecology and Biogeography* 21.4 (2012), pages 498–507 (cited on page 92).
- [247] K. Jones et al. “A review of fibropapillomatosis in Green turtles (*Chelonia mydas*)”. In: *The Veterinary Journal* 212 (2016), pages 48–57. ISSN: 1090-0233. DOI: <https://doi.org/10.1016/j.tvjl.2015.10.041> (cited on page 30).
- [248] Peter Juroszek and Andreas von Tiedemann. “Linking plant disease models to climate change scenarios to project future risks of crop diseases: a review”. In: *Journal of Plant Diseases and Protection* 122 (2015), pages 3–15. DOI: <https://doi.org/10.1007/BF03356525> (cited on page 96).
- [249] Jean Claude Kamgang and Gauthier Sallet. “Computation of threshold conditions for epidemiological models and global stability of the disease-free equilibrium (DFE)”. In: *Mathematical Biosciences* 213.1 (2008), pages 1–12. ISSN: 0025-5564. DOI: [10.1016/j.mbs.2008.02.005](https://doi.org/10.1016/j.mbs.2008.02.005) (cited on page 46).
- [250] Yu-Han Kao and Marisa C. Eisenberg. “Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment”. In: *Epidemics* 25 (2018), pages 89–100. ISSN: 1755-4365. DOI: [10.1016/j.epidem.2018.05.010](https://doi.org/10.1016/j.epidem.2018.05.010) (cited on page 46).
- [251] L. Kapsenberg et al. “Coastal ocean acidification and increasing total alkalinity in the northwestern Mediterranean Sea”. In: *Ocean Science* 13.3 (2017), pages 411–426. DOI: [10.5194/os-13-411-2017](https://doi.org/10.5194/os-13-411-2017). URL: <https://os.copernicus.org/articles/13/411/2017/> (cited on pages 134, 140).

- [252] R. Karban and M. Huntzinger. “Decline of meadow spittlebugs, a previously abundant insect, along the California coast”. In: *Bull Ecol Soc Am* 8 (2018), pages 1–3. DOI: <https://doi.org/10.1002/bes2.1448> (cited on page 88).
- [253] Dirk Nikolaus Karger et al. “Climatologies at high resolution for the earth’s land surface areas”. In: *Scientific Data* 4.1 (Sept. 2017), page 170122. ISSN: 2052-4463. DOI: [10.1038/sdata.2017.122](https://doi.org/10.1038/sdata.2017.122) (cited on pages 111, 117).
- [254] Dirk Nikolaus Karger et al. “Climatologies at high resolution for the earth’s land surface areas”. In: *EnviDat* (2021). DOI: [10.16904/envidat.228](https://doi.org/10.16904/envidat.228) (cited on pages 116, 117).
- [255] Stelios Katsanevakis. “Transplantation as a conservation action to protect the Mediterranean fan mussel *Pinna nobilis*”. In: *Marine Ecology Progress Series* 546 (2016), pages 113–122. DOI: [10.3354/meps11658](https://doi.org/10.3354/meps11658) (cited on page 9).
- [256] M. Kearney and W. Porter. “Mechanistic niche modelling: combining physiological and spatial data to predict species’ ranges”. In: *Ecology Letters* 12.4 (2009), pages 334–350. DOI: [10.1111/j.1461-0248.2008.01277.x](https://doi.org/10.1111/j.1461-0248.2008.01277.x). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2008.01277.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2008.01277.x> (cited on page 79).
- [257] Matt J. Keeling and Bryan T. Grenfell. “Individual-based Perspectives on R<sub>0</sub>”. In: *Journal of Theoretical Biology* 203.1 (2000), pages 51–61. ISSN: 0022-5193. DOI: <https://doi.org/10.1006/jtbi.1999.1064> (cited on page 36).
- [258] Sonia Kéfi et al. “Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems”. In: *Nature* 449.7159 (2007), pages 213–217. ISSN: 1476-4687. DOI: [10.1038/nature06111](https://doi.org/10.1038/nature06111). URL: <https://doi.org/10.1038/nature06111> (cited on page 129).
- [259] D. M. Kennedy and C. D. Woodroffe. “Fringing reef growth and morphology: a review”. In: *Earth-Science Reviews* 57.3 (2002), pages 255–277. ISSN: 0012-8252. DOI: [10.1016/S0012-8252\(01\)00077-0](https://doi.org/10.1016/S0012-8252(01)00077-0). URL: <https://www.sciencedirect.com/science/article/pii/S0012825201000770> (cited on page 130).
- [260] W. O. Kermack and A. G. McKendrick. “Contributions to the mathematical theory of epidemics-III. Further studies of the problem of endemicity”. In: *Bulletin of Mathematical Biology* 53.1 (1933), pages 89–118. ISSN: 0092-8240. URL: [http://www.sciencedirect.com/science/article/pii/S0092824005800424](https://www.sciencedirect.com/science/article/pii/S0092824005800424) (cited on page 8).
- [261] William Ogilvy Kermack and A. G. McKendrick. “A contribution to the mathematical theory of epidemics”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (1927), pages 700–721. DOI: [10.1098/rspa.1927.0118](https://doi.org/10.1098/rspa.1927.0118) (cited on pages 8, 10, 46).
- [262] William Ogilvy Kermack and A. G. McKendrick. “Contributions to the mathematical theory of epidemics-II. The problem of endemicity”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 138.834 (1932), pages 55–83. DOI: [10.1098/rspa.1932.0171](https://doi.org/10.1098/rspa.1932.0171) (cited on page 8).
- [263] Dirk Koopmans et al. “High net primary production of Mediterranean seagrass (*Posidonia oceanica*) meadows determined with aquatic eddy covariance”. In: *Frontiers in Marine Science* 7 (2020), page 118. DOI: [10.3389/fmars.2020.00118](https://doi.org/10.3389/fmars.2020.00118) (cited on page 140).
- [264] D Krause-Jensen et al. “Macroalgae contribute to nested mosaics of pH variability in a subarctic fjord”. In: *Biogeosciences* 12.16 (2015), pages 4895–4911. DOI: [10.5194/bg-12-4895-2015](https://doi.org/10.5194/bg-12-4895-2015). URL: <https://bg.copernicus.org/articles/12/4895/2015/> (cited on page 135).
- [265] Kristy J Kroeker et al. “Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming”. In: *Global Change Biology* 19.6 (2013), pages 1884–1896. DOI: [10.1111/gcb.12179](https://doi.org/10.1111/gcb.12179) (cited on page 134).
- [266] *Kunming-Montreal Global Biodiversity Framework*. Convention on Biological Diversity. 2022. URL: <https://www.cbd.int/gbf/> (visited on 10/31/2023) (cited on page 124).
- [267] Kevin D. Lafferty. “The ecology of climate change and infectious diseases”. In: *Ecology* 90.4 (2009), pages 888–900. DOI: <https://doi.org/10.1890/08-0079.1> (cited on page 110).
- [268] Kevin D. Lafferty, James W. Porter, and Susan E. Ford. “Are Diseases Increasing in the Ocean?” In: *Annual Review of Ecology, Evolution, and Systematics* 35 (2004), pages 31–54. ISSN: 1543592X, 15452069. DOI: [10.1146/annurev.ecolsys.35.021103.105704](https://doi.org/10.1146/annurev.ecolsys.35.021103.105704) (cited on pages 8, 30).

- [269] Kevin D. Lafferty et al. “Infectious diseases affect marine fisheries and aquaculture economics”. In: *Annual Review of Marine Science* 7 (2015), pages 471–496. ISSN: 19410611. DOI: [10.1146/annurev-marine-010814-015646](https://doi.org/10.1146/annurev-marine-010814-015646) (cited on page 8).
- [270] C. Lago et al. “Degree-day-based model to predict egg hatching of *Philaenus spumarius* (Hemiptera: Aphrophoridae), the main vector of *Xylella fastidiosa* in Europe”. In: *bioRxiv* (2022). DOI: [10.1101/2022.11.10.515963](https://doi.org/10.1101/2022.11.10.515963). eprint: <https://www.biorxiv.org/content/early/2022/11/13/2022.11.10.515963.full.pdf>. URL: <https://www.biorxiv.org/content/early/2022/11/13/2022.11.10.515963> (cited on pages 69, 71).
- [271] Clara Lago et al. “Degree-day-based model to predict egg hatching of *Philaenus spumarius* (Hemiptera: Aphrophoridae), the main vector of *Xylella fastidiosa* in Europe”. In: *Environmental Entomology* 52.3 (Apr. 2023), pages 350–359. ISSN: 0046-225X. DOI: [10.1093/ee/nvad013](https://doi.org/10.1093/ee/nvad013). eprint: <https://academic.oup.com/ee/article-pdf/52/3/350/50615564/nvad013.pdf>. URL: <https://doi.org/10.1093/ee/nvad013> (cited on page 19).
- [272] P Landschützer et al. “A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink”. In: *Biogeosciences* 10.11 (2013), pages 7793–7815. DOI: [10.5194/bg-10-7793-2013](https://doi.org/10.5194/bg-10-7793-2013). URL: <https://bg.copernicus.org/articles/10/7793/2013/> (cited on page 135).
- [273] Abid Ali Lashari and Gul Zaman. “Global dynamics of vector-borne diseases with horizontal transmission in host population”. In: *Computers & Mathematics with Applications* 61.4 (2011), pages 745–754. ISSN: 0898-1221. DOI: [10.1016/j.camwa.2010.12.018](https://doi.org/10.1016/j.camwa.2010.12.018) (cited on page 49).
- [274] István G. Laukó. “Stability of disease free sets in epidemic models”. In: *Mathematical and Computer Modelling* 43.11 (2006), pages 1357–1366. ISSN: 0895-7177. DOI: [10.1016/j.mcm.2005.06.011](https://doi.org/10.1016/j.mcm.2005.06.011) (cited on page 46).
- [275] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *Nature* 521 (2015), pages 436–444. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539) (cited on page 135).
- [276] Kitack Lee et al. “Roles of marginal seas in absorbing and storing fossil fuel CO<sub>2</sub>”. In: *Energy & Environmental Science* 4.4 (2011), pages 1133–1146. DOI: [10.1039/C0EE00663G](https://doi.org/10.1039/C0EE00663G) (cited on page 139).
- [277] Nathalie Lefévre, Andrew J Watson, and Adam R Watson. “A comparison of multiple regression and neural network techniques for mapping in situ pCO<sub>2</sub> data”. In: *Tellus B: Chemical and Physical Meteorology* 57.5 (2005), pages 375–384. DOI: [10.3402/tellusb.v57i5.16565](https://doi.org/10.3402/tellusb.v57i5.16565) (cited on page 139).
- [278] S. T. Leutenegger, M. A. Lopez, and J. Edgington. “STR: A simple and efficient algorithm for R-tree packing”. In: *Proceedings 13th international conference on data engineering*. IEEE. 1997, pages 497–506 (cited on pages 130, 131).
- [279] S. A. Levin, editor. *Mathematics and Biology: The interface, challenges and opportunities*. Lawrence Berkeley Lab., CA (USA), 1992. eprint: <https://escholarship.org/uc/item/870966k3> (cited on page 60).
- [280] Simon A. Levin. “The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture”. In: *Ecology* 73.6 (1992), pages 1943–1967. DOI: <https://doi.org/10.2307/1941447>. eprint: <https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.2307/1941447>. URL: <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1941447> (cited on page 110).
- [281] Jing Li, Daniel Blakeley, and Robert J. Smith. “The failure of  $R_0$ ”. In: *Computational and Mathematical Methods in Medicine* 2011 (2011), page 527610. ISSN: 1748-6718. DOI: [10.1155/2011/527610](https://doi.org/10.1155/2011/527610) (cited on page 30).
- [282] Xiaoshuang Li et al. “A Neural Network-Based Analysis of the Seasonal Variability of Surface Total Alkalinity on the East China Sea Shelf”. In: *Frontiers in Marine Science* 7 (2020), page 219. ISSN: 2296-7745. DOI: [10.3389/fmars.2020.00219](https://doi.org/10.3389/fmars.2020.00219). URL: <https://www.frontiersin.org/article/10.3389/fmars.2020.00219> (cited on page 139).
- [283] J. H. Lieth et al. “Modeling cold curing of Pierce’s disease in *Vitis vinifera* ‘Pinot Noir’ and ‘Cabernet Sauvignon’ grapevines in California”. In: *Phytopathology* 101 (4 2011), pages 1492–1500. DOI: [10.1094/PHYTO-08-10-0207](https://doi.org/10.1094/PHYTO-08-10-0207) (cited on pages 78, 80, 81, 90–92).
- [284] Steven Lindow. “Money Matters: Fueling Rapid Recent Insight Into *Xylella fastidiosa*—An Important and Expanding Global Pathogen”. In: *Phytopathology* 109.2 (2019), pages 210–212. DOI: [10.1094/PHYTO-09-18-0325-PER](https://doi.org/10.1094/PHYTO-09-18-0325-PER) (cited on page 110).

- [285] Scott R. Loarie et al. “The velocity of climate change”. In: *Nature* 462.7276 (Dec. 2009), pages 1052–1055. ISSN: 1476-4687. DOI: <https://doi.org/10.1038/nature08649> (cited on pages 100, 103).
- [286] J. López-Mercadal, P. Mercadal-Frontera, and M. Á. Miranda. “Mechanical management of weeds drops nymphal density of *Xylella fastidiosa* vectors”. In: *bioRxiv* (2022). DOI: [10.1101/2022.10.18.512680](https://doi.org/10.1101/2022.10.18.512680) (cited on pages 60, 71).
- [287] J. López-Mercadal et al. “Collection of data and information in Balearic Islands on biology of vectors and potential vectors of *Xylella fastidiosa* (GP/EFSA/ALPHA/017/01)”. In: *EFSA Supporting Publications* 18.10 (2021), 6925E. DOI: [10.2903/sp.efsa.2021.EN-6925](https://doi.org/10.2903/sp.efsa.2021.EN-6925) (cited on pages 62, 70, 72).
- [288] Talita Loureiro et al. “*Xylella fastidiosa*: A Glimpse of the Portuguese Situation”. In: *Microbiology Research* 14.4 (2023), pages 1568–1588. DOI: [10.3390/microbiolres14040108](https://doi.org/10.3390/microbiolres14040108) (cited on page 117).
- [289] S. Lovejoy. “Area-Perimeter Relation for Rain and Cloud Areas”. In: *Science* 216.4542 (1982), pages 185–187. DOI: [10.1126/science.216.4542.185](https://doi.org/10.1126/science.216.4542.185). eprint: <https://www.science.org/doi/pdf/10.1126/science.216.4542.185>. URL: <https://www.science.org/doi/abs/10.1126/science.216.4542.185> (cited on page 125).
- [290] George Macdonald et al. “The epidemiology and control of malaria.” In: *The Epidemiology and Control of Malaria*. (1957) (cited on page 46).
- [291] L. V. Madden, M. J. Jeger, and F. van den Bosch. “A Theoretical Assessment of the Effects of Vector-Virus Transmission Mechanism on Plant Virus Disease Epidemics”. In: *Phytopathology* 90.6 (2000), pages 576–594. DOI: [10.1094/PHYTO.2000.90.6.576](https://doi.org/10.1094/PHYTO.2000.90.6.576). eprint: <https://doi.org/10.1094/PHYTO.2000.90.6.576>. URL: <https://doi.org/10.1094/PHYTO.2000.90.6.576> (cited on page 60).
- [292] Benoit B. Mandelbrot. *The fractal geometry of nature*. 3rd edition. New York: W. H. Freeman and Comp., 1983 (cited on pages 125, 131).
- [293] Núria Marbà et al. “Effectiveness of protection of seagrass (*Posidonia oceanica*) populations in Cabrera National Park (Spain)”. In: *Environmental Conservation* 29.4 (2002), pages 509–518 (cited on page 141).
- [294] Ester Marco-Noales et al. “Evidence that *Xylella fastidiosa* is the causal agent of almond leaf scorch disease in Alicante, mainland Spain (Iberian Peninsula)”. In: *Plant Disease* 105.11 (2021), pages 3349–3352. DOI: <https://doi.org/10.1094/PDIS-03-21-0625-SC> (cited on page 96).
- [295] Dimitrije Marković and Claudio Gros. “Power laws and self-organized criticality in theory and nature”. In: *Physics Reports* 536.2 (2014), pages 41–74. ISSN: 0370-1573. DOI: [10.1016/j.physrep.2013.11.002](https://doi.org/10.1016/j.physrep.2013.11.002). URL: <https://www.sciencedirect.com/science/article/pii/S0370157313004298> (cited on page 129).
- [296] Pablo A. Marquet et al. “Scaling and power-laws in ecological systems”. In: *Journal of Experimental Biology* 208.9 (May 2005), pages 1749–1769. ISSN: 0022-0949. DOI: [10.1242/jeb.01588](https://doi.org/10.1242/jeb.01588). eprint: <https://journals.biologists.com/jeb/article-pdf/208/9/1749/1254844/1749.pdf>. URL: <https://doi.org/10.1242/jeb.01588> (cited on page 129).
- [297] Maia Martcheva. *An Introduction to Mathematical Epidemiology*. New York: Springer, 2015. ISBN: 978-1-4899-7611-6. DOI: [10.1007/978-1-4899-7612-3](https://doi.org/10.1007/978-1-4899-7612-3) (cited on pages 11, 47, 62).
- [298] V. Masson-Delmotte et al. “Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change”. In: *Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2021 (cited on page 134).
- [299] Robert M. May and Roy M. Anderson. “Population biology of infectious diseases: Part II”. In: *Nature* 280.5722 (Aug. 1979), pages 455–461. ISSN: 1476-4687. DOI: [10.1038/280455a0](https://doi.org/10.1038/280455a0). URL: <https://doi.org/10.1038/280455a0> (cited on pages 10, 23).
- [300] Hamish I. McCallum et al. “Does terrestrial epidemiology apply to marine systems?” In: *Trends in Ecology & Evolution* 19.11 (2004), pages 585–591. ISSN: 0169-5347. DOI: [10.1016/j.tree.2004.08.009](https://doi.org/10.1016/j.tree.2004.08.009) (cited on pages 8, 10, 23).
- [301] Joseph B. McLaughlin et al. “Outbreak of *Vibrio parahaemolyticus* Gastroenteritis Associated with Alaskan Oysters”. In: *New England Journal of Medicine* 353.14 (2005). PMID: 16207848, pages 1463–1470. DOI: [10.1056/NEJMoa051594](https://doi.org/10.1056/NEJMoa051594) (cited on page 30).
- [302] Gregory S. McMaster and W.W. Wilhelm. “Growing degree-days: one equation, two interpretations”. In: *Agricultural and Forest Meteorology* 87.4 (1997), pages 291–300. ISSN: 0168-1923. DOI: [10.1016/S0168-1923\(97\)00027-0](https://doi.org/10.1016/S0168-1923(97)00027-0) (cited on pages 79, 89).

- [303] Carl Mehrbach et al. “Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure”. In: *Limnology and oceanography* 18.6 (1973), pages 897–907. DOI: [10.4319/lo.1973.18.6.0897](https://doi.org/10.4319/lo.1973.18.6.0897) (cited on page 142).
- [304] Annette Menzel et al. “European phenological response to climate change matches the warming pattern”. In: *Global Change Biology* 12.10 (2006), pages 1969–1976. DOI: [10.1111/j.1365-2486.2006.01193.x](https://doi.org/10.1111/j.1365-2486.2006.01193.x) (cited on page 116).
- [305] Jesús M. Mercado and F. J. L. Gordillo. “Inorganic carbon acquisition in algal communities: are the laboratory data relevant to the natural ecosystems?” In: *Photosynthesis Research* 109.1 (2011), page 257. ISSN: 1573-5079. DOI: [10.1007/s11120-011-9646-0](https://doi.org/10.1007/s11120-011-9646-0). URL: <https://doi.org/10.1007/s11120-011-9646-0> (cited on page 135).
- [306] Met Office. *Cartopy: a cartographic python library with a Matplotlib interface*. Exeter, Devon, 2010 - 2015. URL: <https://scitools.org.uk/cartopy> (cited on page 100).
- [307] Fiorenza Micheli et al. “Cumulative human impacts on Mediterranean and Black Sea marine ecosystems: assessing current pressures and opportunities”. In: *PLoS One* 8.12 (2013), e79889. DOI: [10.1371/journal.pone.0079889](https://doi.org/10.1371/journal.pone.0079889) (cited on page 134).
- [308] Claude Millot. “Circulation in the Western Mediterranean Sea”. In: *Journal of Marine Systems* 20.1-4 (1999), pages 423–442. DOI: [10.1016/S0924-7963\(98\)00078-5](https://doi.org/10.1016/S0924-7963(98)00078-5) (cited on page 140).
- [309] Susannah Mistr and David Bercovici. “A Theoretical Model of Pattern Formation in Coral Reefs”. In: *Ecosystems* 6.1 (2003), pages 0061–0074. ISSN: 1432-9840. DOI: [10.1007/s10021-002-0199-0](https://doi.org/10.1007/s10021-002-0199-0). URL: <https://doi.org/10.1007/s10021-002-0199-0> (cited on pages 129, 130).
- [310] Fredrik Moberg and Carl Folke. “Ecological goods and services of coral reef ecosystems”. In: *Ecological Economics* 29.2 (1999), pages 215–233. ISSN: 0921-8009. DOI: [https://doi.org/10.1016/S0921-8009\(99\)00009-9](https://doi.org/10.1016/S0921-8009(99)00009-9). URL: <https://www.sciencedirect.com/science/article/pii/S0921800999000099> (cited on page 124).
- [311] Péter K. Molnár et al. “Thermal Performance Curves and the Metabolic Theory of Ecology—A Practical Guide to Models and Experiments for Parasitologists”. In: *Journal of Parasitology* 103.5 (2017), pages 423–439. DOI: [10.1645/16-148](https://doi.org/10.1645/16-148) (cited on page 23).
- [312] H. A. Mooney and E. E. Cleland. “The evolutionary impact of invasive species”. In: *Proceedings of the National Academy of the USA* 98.10 (2001), pages 5446–5451. DOI: [10.1073/pnas.091093398](https://doi.org/10.1073/pnas.091093398) (cited on page 78).
- [313] E. Moralejo et al. “Insights into the epidemiology of Pierce’s disease in vineyards of Mallorca, Spain”. In: *Plant Pathology* 68.8 (2019), pages 1458–1471. DOI: <https://doi.org/10.1111/ppa.13076>. eprint: <https://bsppjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/ppa.13076>. URL: <https://bsppjournals.onlinelibrary.wiley.com/doi/abs/10.1111/ppa.13076> (cited on pages 60, 78, 79, 83, 88, 89).
- [314] E. Moralejo et al. “Phylogenetic inference enables reconstruction of a long-overlooked outbreak of almond leaf scorch disease (*Xylella fastidiosa*) in Europe”. In: *Communications Biology* 3.1 (Oct. 2020), page 560. ISSN: 2399-3642. DOI: [10.1038/s42003-020-01284-7](https://doi.org/10.1038/s42003-020-01284-7). URL: <https://doi.org/10.1038/s42003-020-01284-7> (cited on pages 60, 63, 78).
- [315] Eduardo Moralejo et al. “Insights into the epidemiology of Pierce’s disease in vineyards of Mallorca, Spain”. In: *Plant Pathology* 68.8 (2019), pages 1458–1471. DOI: <https://doi.org/10.1111/ppa.13076> (cited on pages 105, 117).
- [316] Eduardo Moralejo et al. “Phylogenetic inference enables reconstruction of a long-overlooked outbreak of almond leaf scorch disease (*Xylella fastidiosa*) in Europe”. In: *Communications Biology* 3.1 (2020), page 560. DOI: <https://doi.org/10.1038/s42003-020-01284-7> (cited on page 117).
- [317] David M. Morens, Gregory K. Folkers, and Anthony S. Fauci. “The challenge of emerging and re-emerging infectious diseases”. In: *Nature* 430.6996 (July 2004), pages 242–249. ISSN: 1476-4687. DOI: [10.1038/nature02759](https://doi.org/10.1038/nature02759) (cited on page 30).
- [318] M. Morente et al. “Distribution and relative abundance of insect vectors of *Xylella fastidiosa* in olive groves of the Iberian Peninsula”. In: *Insects* 9.4 (2018), page 175. DOI: [10.3390/insects9040175](https://doi.org/10.3390/insects9040175) (cited on page 60).

- [319] T. Mori, T. E. Smith, and W.-T. Hsu. “Common power laws for cities and spatial fractal structures”. In: *Proceedings of the National Academy of Sciences* 117.12 (2020), pages 6469–6475. DOI: [10.1073/pnas.1913014117](https://doi.org/10.1073/pnas.1913014117). eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.1913014117>. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1913014117> (cited on page 125).
- [320] Marco Moriondo et al. “Projected shifts of wine regions in response to climate change”. In: *Climatic change* 119 (2013), pages 825–839. DOI: <https://doi.org/10.1007/s10584-013-0739-y> (cited on page 106).
- [321] Miguel A. Muñoz. “Colloquium: Criticality and dynamical scaling in living systems”. In: *Review of Modern Physics* 90 (3 2018), page 031001. DOI: [10.1103/RevModPhys.90.031001](https://doi.org/10.1103/RevModPhys.90.031001). URL: <https://link.aps.org/doi/10.1103/RevModPhys.90.031001> (cited on page 129).
- [322] J. Muñoz Sabater. *ERA5-Land hourly data from 1950 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). 2019. DOI: [10.24381/cds.e2161bac](https://doi.org/10.24381/cds.e2161bac) (cited on pages 110, 117).
- [323] J. Muñoz-Sabater. *ERA5-Land hourly data from 1981 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). July 2019. DOI: [10.24381/cds.e2161bac](https://doi.org/10.24381/cds.e2161bac) (cited on page 91).
- [324] Joaquín Muñoz-Sabater et al. “ERA5-Land: a state-of-the-art global reanalysis dataset for land applications”. In: *Earth System Science Data* 13.9 (Sept. 7, 2021), pages 4349–4383. ISSN: 1866-3508. DOI: <https://doi.org/10.5194/essd-13-4349-2021>. URL: <https://essd.copernicus.org/articles/13/4349/2021/> (visited on 06/19/2023) (cited on pages 97, 111, 117).
- [325] Kevin Murphy et al. “World distribution, diversity and endemism of aquatic macrophytes”. In: *Aquatic Botany* 158 (2019), page 103127. DOI: [10.1016/j.aquabot.2019.06.006](https://doi.org/10.1016/j.aquabot.2019.06.006) (cited on page 135).
- [326] J. D. Murray. *Mathematical Biology*. Berlin (Germany): Springer, 1989 (cited on page 60).
- [327] J. D. Murray. *Mathematical Biology. I. An Introduction*. 3rd edition. New York, NY: Springer, 2002. ISBN: 978-0-387-22437-4 (cited on pages 13, 25).
- [328] A. L Myers et al. “Pierce’s disease of grapevines: Identification of the primary vectors in North Carolina”. In: *Phytopathology* 97 (2007), pages 1440–1450. DOI: [10.1094/PHYTO-97-11-1440](https://doi.org/10.1094/PHYTO-97-11-1440) (cited on page 92).
- [329] T. Nakamura and T. Nakamori. “A geochemical model for coral reef formation”. In: *Coral Reefs* 26.4 (2007), pages 741–755. ISSN: 1432-0975. DOI: [10.1007/s00338-007-0262-6](https://doi.org/10.1007/s00338-007-0262-6). URL: <https://doi.org/10.1007/s00338-007-0262-6> (cited on page 129).
- [330] Carlos Navarro-Racines et al. “High-resolution and bias-corrected CMIP5 projections for climate change impact assessments”. In: *Scientific Data* 7 (2020), page 7. DOI: <https://doi.org/10.1038/s41597-019-0343-8> (cited on page 110).
- [331] M. E. J. Newman and Kim Sneppen. “Avalanches, scaling, and coherent noise”. In: *Phys. Rev. E* 54 (6 1996), pages 6226–6231. DOI: [10.1103/PhysRevE.54.6226](https://doi.org/10.1103/PhysRevE.54.6226). URL: <https://link.aps.org/doi/10.1103/PhysRevE.54.6226> (cited on page 129).
- [332] JA Newton et al. *Global ocean acidification observing network: requirements and governance plan*. 2015. eprint: <https://archimer.ifremer.fr/doc/00651/76343/> (cited on page 135).
- [333] Göran E Nilsson et al. “Near-future carbon dioxide levels alter fish behaviour by interfering with neurotransmitter function”. In: *Nature Climate Change* 2.3 (2012), pages 201–204. DOI: [10.1038/nclimate1352](https://doi.org/10.1038/nclimate1352) (cited on page 134).
- [334] L. Nunney, H. Azad, and R. Stouthamer. “An Experimental Test of the Host-Plant Range of Nonrecombinant Strains of North American *Xylella fastidiosa* subsp. *multiplex*”. In: *Phytopathology* 109 (2019), pages 294–300. DOI: [10.1094/PHYTO-07-18-0252-FI](https://doi.org/10.1094/PHYTO-07-18-0252-FI) (cited on page 78).
- [335] S O’Connor et al. *NOAA Coral Reef Information System, National Centers for Environmental Information*. Silver Spring, MD, 2p. 2020 (cited on page 124).
- [336] D. Olmo et al. “First Detection of *Xylella fastidiosa* Infecting Cherry (*Prunus avium*) and *Polygala myrtifolia* Plants, in Mallorca Island, Spain”. In: *Plant Disease* 101.10 (10 2017), pages 1820–1820. DOI: [10.1094/PDIS-04-17-0590-PDN](https://doi.org/10.1094/PDIS-04-17-0590-PDN) (cited on page 78).
- [337] D. Olmo et al. “Landscape Epidemiology of *Xylella fastidiosa* in the Balearic Islands”. In: *Agronomy* 11.3 (2021), page 473. ISSN: 2073-4395. DOI: [10.3390/agronomy11030473](https://doi.org/10.3390/agronomy11030473) (cited on page 60).
- [338] Diego Olmo et al. “Landscape epidemiology of *Xylella fastidiosa* in the Balearic Islands”. In: *Agronomy* 11.3 (2021), page 473. DOI: <https://doi.org/10.3390/agronomy11030473> (cited on page 96).

- [339] Ariel Ortiz-Bobea et al. “Anthropogenic climate change has slowed global agricultural productivity growth”. In: *Nature Climate Change* 11.4 (Apr. 2021), pages 306–312. ISSN: 1758-6798. DOI: [10.1038/s41558-021-01000-1](https://doi.org/10.1038/s41558-021-01000-1) (cited on page 110).
- [340] L. M. Overall and E. J. Rebek. “Seasonal Abundance and Natural Inoculativity of Insect Vectors of *Xylella fastidiosa* in Oklahoma Tree Nurseries and Vineyards”. In: *Journal of Economic Entomology Pest Management* (1 2015), pages 1–10. DOI: [10.1093/jee/tov261](https://doi.org/10.1093/jee/tov261) (cited on page 92).
- [341] L. M. Overall and E. J. Rebek. “Insect vectors and current management strategies for diseases caused by *Xylella fastidiosa* in the Southern United States”. In: *Journal of Integrated Pest Management* 8 (1 2017), pages 1–12. DOI: [10.1093/jipm/pmx005](https://doi.org/10.1093/jipm/pmx005) (cited on page 78).
- [342] Ivane Lilian Pairaud et al. “Impacts of climate change on coastal benthic ecosystems: assessing the current risk of mortality outbreaks associated with thermal stress in NW Mediterranean coastal areas”. In: *Ocean Dynamics* 64.1 (Jan. 2014), pages 103–115. ISSN: 1616-7228. DOI: [10.1007/s10236-013-0661-x](https://doi.org/10.1007/s10236-013-0661-x) (cited on page 8).
- [343] J. Palmiéri et al. “Simulated anthropogenic CO<sub>2</sub> storage and acidification of the Mediterranean Sea”. In: *Biogeosciences* 12.3 (2015), pages 781–802. ISSN: 17264189. DOI: [10.5194/bg-12-781-2015](https://doi.org/10.5194/bg-12-781-2015) (cited on pages 134, 139).
- [344] Ireneo B Pangga, Jim Hanan, and Sukumar Chakraborty. “Pathogen dynamics in a crop canopy and their evolution under changing climate”. In: *Plant Pathology* 60.1 (2011), pages 70–81. DOI: <https://doi.org/10.1111/j.1365-3059.2010.02408.x> (cited on page 96).
- [345] R Pawlowicz. “M\_Map: A mapping package for MATLAB, version 1.4 m”. In: *Computer software, UBC EOAS* (2020). eprint: <https://www.eoas.ubc.ca/rich/map.html> (cited on page 141).
- [346] Fiz F Pérez et al. “Contrasting drivers and trends of ocean acidification in the subarctic Atlantic”. In: *Scientific Reports* 11.1 (2021), pages 1–16. DOI: [10.1038/s41598-021-93324-3](https://doi.org/10.1038/s41598-021-93324-3) (cited on page 140).
- [347] Fabrice Pernet et al. “Infectious diseases in oyster aquaculture require a new integrated approach”. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371.1689 (Mar. 2016), page 20150213. DOI: [10.1098/rstb.2015.0213](https://doi.org/10.1098/rstb.2015.0213) (cited on page 30).
- [348] Fabrice Pernet et al. “Determination of risk factors for herpesvirus outbreak in oysters using a broad-scale spatial epidemiology framework”. In: *Scientific Reports* 8.1 (July 2018), page 10869. ISSN: 2045-2322. DOI: [10.1038/s41598-018-29238-4](https://doi.org/10.1038/s41598-018-29238-4) (cited on page 30).
- [349] Bruno Petton et al. “The Pacific Oyster Mortality Syndrome, a Polymicrobial and Multifactorial Disease: State of Knowledge and Future Directions”. In: *Frontiers in Immunology* 12 (2021), page 52. ISSN: 1664-3224. DOI: [10.3389/fimmu.2021.630343](https://doi.org/10.3389/fimmu.2021.630343) (cited on page 30).
- [350] Steven J. Phillips, Robert P. Anderson, and Robert E. Schapire. “Maximum entropy modeling of species geographic distributions”. In: *Ecological Modelling* 190.3 (2006), pages 231–259. ISSN: 0304-3800. DOI: <https://doi.org/10.1016/j.ecolmodel.2005.03.026>. URL: <http://www.sciencedirect.com/science/article/pii/S030438000500267X> (visited on 12/13/2013) (cited on page 99).
- [351] D. Pimentel et al. “Environmental and Economic Costs of Nonindigenous Species in the United States”. In: *BioScience* 50.1 (2000), pages 53–65. DOI: [10.1641/0006-3568\(2000\)050\[0053:EAECON\]2.3.CO;2](https://doi.org/10.1641/0006-3568(2000)050[0053:EAECON]2.3.CO;2) (cited on page 78).
- [352] C. M. A. Pinto, A. Mendes Lopes, and J.A. Tenreiro Machado. “Double power laws, fractals and self-similarity”. In: *Applied Mathematical Modelling* 38.15 (2014), pages 4019–4026. ISSN: 0307-904X. DOI: [10.1016/j.apm.2014.01.012](https://doi.org/10.1016/j.apm.2014.01.012). URL: <https://www.sciencedirect.com/science/article/pii/S0307904X1400047X> (cited on page 125).
- [353] Eric N Powell et al. “Modeling the MSX parasite in eastern oyster (*Crassostrea virginica*) populations. III. Regional application and the problem of transmission”. In: *Journal of Shellfish Research* 18.2 (1999), pages 517–537. URL: [https://digitalcommons.odu.edu/ccpo\\_pubs/76/](https://digitalcommons.odu.edu/ccpo_pubs/76/) (cited on pages 20, 30).
- [354] Eric N. Powell and Eileen E. Hofmann. “Models of marine molluscan diseases: Trends and challenges”. In: *Journal of Invertebrate Pathology* 131 (2015), pages 212–225. ISSN: 0022-2011. DOI: [10.1016/j.jip.2015.07.017](https://doi.org/10.1016/j.jip.2015.07.017) (cited on pages 8, 10, 23, 31).
- [355] Patricia Prado et al. “*Pinna nobilis* in suboptimal environments are more tolerant to disease but more vulnerable to severe weather phenomena”. In: *Marine Environmental Research* (2020), page 105220. ISSN: 0141-1136. DOI: [10.1016/j.marenvres.2020.105220](https://doi.org/10.1016/j.marenvres.2020.105220) (cited on page 9).

- [356] A. H. Purcell. “Almond Leaf Scorch: Leafhopper and Spittlebug Vectors12”. In: *Journal of Economic Entomology* 73.6 (4 Dec. 1980), pages 834–838. ISSN: 0022-0493. DOI: [10.1093/jee/73.6.834](https://doi.org/10.1093/jee/73.6.834). eprint: <https://academic.oup.com/jee/article-pdf/73/6/834/19227706/jee73-0834.pdf>. URL: <https://doi.org/10.1093/jee/73.6.834> (cited on page 78).
- [357] A. H. Purcell. “Paradigms: Examples from the Bacterium *Xylella fastidiosa*”. In: *Annual Review of Phytopathology* 51 (2013), pages 339–356. DOI: [10.1146/annurev-phyto-082712-102325](https://doi.org/10.1146/annurev-phyto-082712-102325) (cited on pages 78, 79).
- [358] A. H. Purcell and A. Finlay. “Evidence for noncirculative transmission of Pierce’s disease bacterium by sharpshooter leafhoppers”. In: *Phytopathology* 69.4 (1979), pages 393–395. DOI: [10.1094/Phyto-69-393](https://doi.org/10.1094/Phyto-69-393) (cited on pages 60, 61).
- [359] AH Purcell. “Spatial patterns of Pierce’s disease in the Napa Valley”. In: *American Journal of Enology and Viticulture* 25.3 (1974), pages 162–167 (cited on page 90).
- [360] Alexander H Purcell et al. “Vector preference and inoculation efficiency as components of resistance to Pierce’s disease in European grape cultivars”. In: *Phytopathology* 71.4 (1981), pages 429–435 (cited on page 90).
- [361] S. J. Purkis et al. “The Statistics of Natural Shapes in Modern Coral Reef Landscapes”. In: *The Journal of Geology* 115 (5 2007), pages 493–508. DOI: [10.1086/519774](https://doi.org/10.1086/519774) (cited on page 124).
- [362] Christopher Rackauckas and Qing Nie. “DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia”. In: *The Journal of Open Research Software* 5.1 (2017), page 15. DOI: [10.5334/jors.151](https://doi.org/10.5334/jors.151) (cited on pages 20, 26, 64).
- [363] R. A. Redak et al. “The biology of xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology”. In: *Annual Review of Entomology* 49.1 (2004), pages 243–270. DOI: [10.1146/annurev.ento.49.061802.123403](https://doi.org/10.1146/annurev.ento.49.061802.123403). eprint: <https://doi.org/10.1146/annurev.ento.49.061802.123403> (cited on pages 60, 78, 79, 87, 88).
- [364] Richard A. Redak et al. “The biology of xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology”. In: *Annual Review of Entomology* 49.1 (2004), pages 243–270. DOI: <https://doi.org/10.1146/annurev.ento.49.061802.123403> (cited on pages 96, 110).
- [365] Sergio J. Rey and Luc Anselin. “PySAL: A Python Library of Spatial Analytical Methods”. In: *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Edited by Manfred M. Fischer and Arthur Getis. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pages 175–193. ISBN: 978-3-642-03647-7. DOI: [10.1007/978-3-642-03647-7\\_11](https://doi.org/10.1007/978-3-642-03647-7_11). URL: [https://doi.org/10.1007/978-3-642-03647-7\\_11](https://doi.org/10.1007/978-3-642-03647-7_11) (cited on page 131).
- [366] Aurora M Ricart et al. “Coast-wide evidence of low pH amelioration by seagrass ecosystems”. In: *Global Change Biology* 27.11 (2021), pages 2580–2591. DOI: [10.1111/gcb.15594](https://doi.org/10.1111/gcb.15594) (cited on page 135).
- [367] Steven Riley et al. “Five challenges for spatial epidemic models”. In: *Epidemics* 10 (2015). Challenges in Modelling Infectious Disease Dynamics, pages 68–71. ISSN: 1755-4365. DOI: [10.1016/j.epidem.2014.07.001](https://doi.org/10.1016/j.epidem.2014.07.001) (cited on page 30).
- [368] Jean B. Ristaino et al. “The persistent threat of emerging plant disease pandemics to global food security”. In: *Proceedings of the National Academy of Sciences* 118.23 (2021), e2022239118. DOI: [10.1073/pnas.2022239118](https://doi.org/10.1073/pnas.2022239118) (cited on page 110).
- [369] Paola Rivaro et al. “Distributions of carbonate properties along the water column in the Mediterranean Sea: Spatial and temporal variations”. In: *Marine Chemistry* 121.1 (2010), pages 236–245. ISSN: 0304-4203. DOI: <https://doi.org/10.1016/j.marchem.2010.05.003>. URL: <https://www.sciencedirect.com/science/article/pii/S030442031000068X> (cited on page 140).
- [370] Joacim Rocklöv and Robert Dubrow. “Climate change: an enduring challenge for vector-borne disease prevention and control”. In: *Nature Immunology* 21.5 (May 2020), pages 479–483. ISSN: 1529-2916. DOI: [10.1038/s41590-020-0648-y](https://doi.org/10.1038/s41590-020-0648-y) (cited on pages 46, 96).
- [371] Jason R. Rohr and Jeremy M. Cohen. “Understanding how temperature shifts could impact infectious diseases”. In: *PLoS Biology* 18.11 (2020), e3000938. DOI: [10.1371/journal.pbio.3000938](https://doi.org/10.1371/journal.pbio.3000938) (cited on page 23).
- [372] Kimberlyn Roosa and Gerardo Chowell. “Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models”. In: *Theoretical Biology and Medical Modelling* 16.1 (Jan. 2019), page 1. ISSN: 1742-4682. DOI: [10.1186/s12976-018-0097-6](https://doi.org/10.1186/s12976-018-0097-6) (cited on pages 46, 70).

- [373] M. Rouxel et al. “Geographic distribution of cryptic species of *Plasmopara viticola* causing downy mildew on wild and cultivated grape in Eastern North America”. In: *Phytopathology* 104 (7 2014), pages 694–701. DOI: [10.1094/PHYTO-08-13-0225-R](https://doi.org/10.1094/PHYTO-08-13-0225-R) (cited on pages 78, 89).
- [374] Edward P. Rybicki. “A Top Ten list for economically important plant viruses”. In: *Archives of Virology* 160 (1 2015), pages 17–20. DOI: [10.1007/s00705-014-2295-9](https://doi.org/10.1007/s00705-014-2295-9) (cited on page 46).
- [375] Chiara Sabelli. “Deadly olive tree pathogen came by road and rail”. In: *Nature Italy* (Aug. 2023). DOI: [10.1038/d43978-023-00118-4](https://doi.org/10.1038/d43978-023-00118-4) (cited on page 110).
- [376] Cristiano Salata et al. “Coronaviruses: a paradigm of new emerging zoonotic diseases”. In: *Pathogens and Disease* 77.9 (Feb. 2020), ftaa006. DOI: [10.1093/femspd/ftaa006](https://doi.org/10.1093/femspd/ftaa006) (cited on page 30).
- [377] Andrea Saltelli et al. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. USA: Halsted Press, 2004. ISBN: 0470870931 (cited on pages 26, 63).
- [378] M. Saponari et al. “Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (Southern Italy)”. In: *Journal of Plant Pathology* 95.3 (2013). DOI: [10.4454/JPP.V95I3.035](https://doi.org/10.4454/JPP.V95I3.035) (cited on pages 60, 96).
- [379] M. Saponari et al. “Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Southern Italy”. In: *Journal of Plant Pathology* 55 (13 2013), page 668. DOI: [10.4454/JPP.V95I3.035](https://doi.org/10.4454/JPP.V95I3.035) (cited on page 78).
- [380] M. Saponari et al. “*Xylella fastidiosa* in olive in Apulia: Where we stand”. In: *Phytopathology* 109.2 (2019), pages 175–186. DOI: [10.1094/PHYTO-08-18-0319-FI](https://doi.org/10.1094/PHYTO-08-18-0319-FI) (cited on page 60).
- [381] Tomislav Šarić et al. “Epidemiology of Noble Pen Shell (*Pinna nobilis* L. 1758) Mass Mortality Events in Adriatic Sea Is Characterised with Rapid Spreading and Acute Disease Progression”. In: *Pathogens* 9.10 (Sept. 2020), page 776. ISSN: 2076-0817. DOI: [10.3390/pathogens9100776](https://doi.org/10.3390/pathogens9100776) (cited on page 9).
- [382] S. Sarkar et al. “Biodiversity conservation planning tools”. In: *Annual Review of Environment and Resources* 31 (2006), pages 123–59. DOI: [10.1146/annurev.energy.31.042606.085844](https://doi.org/10.1146/annurev.energy.31.042606.085844) (cited on page 60).
- [383] Raphaëlle Sauzède et al. “Estimates of water-column nutrient concentrations and carbonate system parameters in the global ocean: a novel approach based on neural networks”. In: *Frontiers in Marine Science* (2017), page 128. DOI: [10.3389/fmars.2017.00128](https://doi.org/10.3389/fmars.2017.00128) (cited on page 139).
- [384] Todd M. Scanlon et al. “Positive feedbacks promote power-law clustering of Kalahari vegetation”. In: *Nature* 449.7159 (2007), pages 209–212. ISSN: 1476-4687. DOI: [10.1038/nature06060](https://doi.org/10.1038/nature06060). URL: [10.1038/nature06060](https://doi.org/10.1038/nature06060) (cited on page 129).
- [385] Fabio Scarpa et al. “Multiple Non-Species-Specific Pathogens Possibly Triggered the Mass Mortality in *Pinna nobilis*”. In: *Life* 10.10 (Oct. 2020), page 238. ISSN: 2075-1729. DOI: [10.3390/life10100238](https://doi.org/10.3390/life10100238) (cited on page 9).
- [386] H Scherm and AHC Van Bruggen. “Global warming and nonlinear growth: how important are changes in average temperature?” In: *Phytopathology* 84.12 (1994), pages 1380–1384. URL: [https://www.apsnet.org/publications/phytopathology/backissues/Documents/1994Abstracts/Phyto\\_84\\_1380.htm](https://www.apsnet.org/publications/phytopathology/backissues/Documents/1994Abstracts/Phyto_84_1380.htm) (cited on pages 96, 110, 116).
- [387] H. Scherm and A. H. C. van Bruggen. “Global Warming and Nonlinear Growth: How Important are Changes in Average Temperature?” In: *Phytopathology* 84 (1994), pages 1380–1384. DOI: [10.1094/phyto-84-1380](https://doi.org/10.1094/phyto-84-1380) (cited on pages 79, 87).
- [388] P. E. Schmid. “Fractal Properties of Habitat and Patch Structure in Benthic Ecosystems”. In: edited by A.H. Fitter and D.G. Raffaelli. Volume 30. *Advances in Ecological Research*. Academic Press, 1999, pages 339–401. DOI: [10.1016/S0065-2504\(08\)60021-5](https://doi.org/10.1016/S0065-2504(08)60021-5). URL: <https://www.sciencedirect.com/science/article/pii/S0065250408600215> (cited on page 129).
- [389] Anke Schneider, Douglas WR Wallace, and Arne Körtzinger. “Alkalinity of the Mediterranean Sea”. In: *Geophysical Research Letters* 34.15 (2007). DOI: [10.1029/2006GL028842](https://doi.org/10.1029/2006GL028842) (cited on page 140).
- [390] Anke Schneider et al. “High anthropogenic carbon content in the eastern Mediterranean”. In: *Journal of Geophysical Research: Oceans* 115.C12 (2010). DOI: [10.1029/2010JC006171](https://doi.org/10.1029/2010JC006171) (cited on page 139).
- [391] Kevin Schneider et al. “Impact of *Xylella fastidiosa* subspecies *pauca* in European olives”. In: *Proceedings of the National Academy of Sciences* 117.17 (17 2020), pages 9250–9259. ISSN: 0027-8424. DOI: [10.1073/pnas.1912206117](https://doi.org/10.1073/pnas.1912206117). URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1912206117> (cited on pages 46, 78, 79, 87, 96, 101, 110).

- [392] Sandra K. Schumacher and Jeffrey I. Campbell. “Travel Medicine”. In: *Urgent Care Medicine Secrets*. Edited by Robert P. Olympia, Rory M. O’Neill, and Matthew L. Silvis. Elsevier, 2018. Chapter 56, pages 352–357. ISBN: 978-0-323-46215-0. DOI: [10.1016/B978-0-323-46215-0.00056-2](https://doi.org/10.1016/B978-0-323-46215-0.00056-2) (cited on page 46).
- [393] T. P. Scoffin and J. E. Dixon. “The distribution and structure of coral reefs: one hundred years since Darwin”. In: *Biological Journal of the Linnean Society* 20.1 (1983), pages 11–38. DOI: [10.1111/j.1095-8312.1983.tb01587.x](https://doi.org/10.1111/j.1095-8312.1983.tb01587.x). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1095-8312.1983.tb01587.x> (cited on page 124).
- [394] D. A. Seekell et al. “A fractal-based approach to lake size-distributions”. In: *Geophysical Research Letters* 40.3 (2013), pages 517–521. DOI: <https://doi.org/10.1002/grl.50139>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/grl.50139>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50139> (cited on page 125).
- [395] Matthew P Seidel, Michael D DeGrandpre, and Andrew G Dickson. “A sensor for in situ indicator-based measurements of seawater pH”. In: *Marine Chemistry* 109.1-2 (2008), pages 18–28. DOI: [j.marchem.2007.11.013](https://doi.org/10.1016/j.marchem.2007.11.013) (cited on page 139).
- [396] H. H. P. Severin. “Spittle-insect vectors of Pierce’s disease virus: II. Life history and virus transmission”. In: *Hilgardia* 19 (11 1950), pages 357–382. DOI: [10.3733/hilg.v19n11p357](https://doi.org/10.3733/hilg.v19n11p357) (cited on page 78).
- [397] Nita Shah and Gupta Jyoti. “SEIR Model and Simulation for Vector Borne Diseases”. In: *Applied Mathematics* 4 (Jan. 2013), pages 13–17. DOI: [10.4236/am.2013.48A003](https://doi.org/10.4236/am.2013.48A003) (cited on page 49).
- [398] Lillian L. M. Shapiro, Shelley A. Whitehead, and Matthew B. Thomas. “Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria”. In: *PLOS Biology* 15.10 (10 Oct. 2017), e2003489. DOI: [10.1371/journal.pbio.2003489](https://doi.org/10.1371/journal.pbio.2003489) (cited on page 27).
- [399] Jonathan D Sharp et al. *CO2SYSv3 for MATLAB*. 2020. DOI: [10.5281/zenodo.3950563](https://doi.org/10.5281/zenodo.3950563). URL: <https://zenodo.org/record/3950563> (cited on page 142).
- [400] A. Sicard et al. “*Xylella fastidiosa*: Insights into an Emerging Plant Pathogen”. In: *Annual Review of Phytopathology* 56 (2018), pages 181–202. DOI: [10.1146/annurev-phyto-080417-045849](https://doi.org/10.1146/annurev-phyto-080417-045849) (cited on page 78).
- [401] Brajesh K. Singh et al. “Climate change impacts on plant pathogens, food security and paths forward”. In: *Nature Reviews Microbiology* (May 2023). ISSN: 1740-1534. DOI: <https://doi.org/10.1038/s41579-023-00900-7>. URL: <https://doi.org/10.1038/s41579-023-00900-7> (cited on page 96).
- [402] Sandra Skendžić et al. “The Impact of Climate Change on Agricultural Insect Pests”. In: *Insects* 12.5 (May 2021), page 440. DOI: [10.3390/insects12050440](https://doi.org/10.3390/insects12050440) (cited on page 110).
- [403] I.M Sobol. “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates”. In: *Mathematics and Computers in Simulation* 55.1 (2001), pages 271–280. ISSN: 0378-4754. DOI: [10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6) (cited on pages 26, 64).
- [404] R. V. Solé and S. C. Manrubia. “Are rainforests self-organized in a critical state?” In: *Journal of Theoretical Biology* 173.1 (1995), pages 31–40. DOI: <https://doi.org/10.1006/jtbi.1995.0040>. URL: <https://www.sciencedirect.com/science/article/pii/S0022519385700409> (cited on page 129).
- [405] C. M. Sorensen and C. Oh. “Divine proportion shape preservation and the fractal nature of cluster-cluster aggregates”. In: *Phys. Rev. E* 58 (6 1998), pages 7545–7548. DOI: [10.1103/PhysRevE.58.7545](https://doi.org/10.1103/PhysRevE.58.7545). URL: <https://link.aps.org/doi/10.1103/PhysRevE.58.7545> (cited on page 128).
- [406] C. M. Sorensen and G. M. Wang. “Size distribution effect on the power law regime of the structure factor of fractal aggregates”. In: *Phys. Rev. E* 60 (6 1999), pages 7143–7148. DOI: [10.1103/PhysRevE.60.7143](https://doi.org/10.1103/PhysRevE.60.7143). URL: <https://link.aps.org/doi/10.1103/PhysRevE.60.7143> (cited on page 125).
- [407] S. Soubeirand et al. “Inferring pathogen dynamics from temporal count data: the emergence of *Xylella fastidiosa* in France is probably not recent”. In: *New Phytologist* 219.2 (2018), pages 824–836. DOI: <https://doi.org/10.1111/nph.15177>. eprint: <https://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/nph.15177>. URL: <https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.15177> (cited on page 60).
- [408] Damien Sous et al. “On the small-scale fractal geometrical structure of a living coral reef barrier”. In: *Earth Surface Processes and Landforms* 45.12 (2020), pages 3042–3054. DOI: <https://doi.org/10.1002/esp.4950>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/esp.4950>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/esp.4950> (cited on pages 124, 129).

- [409] N. Spence, L. Hill, and J. Morris. “How the global threat of pests and diseases impacts plants, people, and the planet”. In: *Plants People Planet* 2 (2020), pages 5–13. DOI: [10.1002/ppp3.10088](https://doi.org/10.1002/ppp3.10088) (cited on page 78).
- [410] J. F. Stevenson et al. “Grapevine Susceptibility to Pierce’s Disease II: Progression of Anatomical Symptoms”. In: *American Journal of Enology and Viticulture* 55.3 (2004), pages 238–245. ISSN: 0002-9254. DOI: [10.5344/ajev.2004.55.3.238](https://doi.org/10.5344/ajev.2004.55.3.238). eprint: <https://www.ajevonline.org/content/55/3/238.full.pdf>. URL: <https://www.ajevonline.org/content/55/3/238> (cited on page 61).
- [411] C. C. Su et al. “Pierce’s disease of Grapevines in Taiwan: Isolation, Cultivation and Pathogenicity of *Xylella fastidiosa*”. In: *Journal of Phytopathology* 161 (6 2013), pages 389–396. DOI: [10.1111/jph.12075](https://doi.org/10.1111/jph.12075) (cited on pages 78, 83).
- [412] Chiou-Chu Su et al. “Pierce’s Disease of Grapevines in Taiwan: Isolation, Cultivation and Pathogenicity of *Xylella fastidiosa*”. In: *Journal of Phytopathology* 161.6 (2013), pages 389–396. DOI: <https://doi.org/10.1111/jph.12075> (cited on page 117).
- [413] Yevhen F. Suprunenko, Stephen J. Cornell, and Christopher A. Gilligan. “Analytical approximation for invasion and endemic thresholds, and the optimal control of epidemics in spatially explicit individual-based models”. In: *Journal of The Royal Society Interface* 18.176 (2021), page 20200966. DOI: [10.1098/rsif.2020.0966](https://doi.org/10.1098/rsif.2020.0966) (cited on page 36).
- [414] Karl E. Taylor, Ronald J. Stouffer, and Gerald A. Meehl. “An overview of CMIP5 and the experiment design”. In: *Bulletin of the American Meteorological Society* 93.4 (Oct. 7, 2011), pages 485–498. ISSN: 0003-0007. DOI: <https://doi.org/10.1175/BAMS-D-11-00094.1>. URL: <http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-11-00094.1> (visited on 03/28/2016) (cited on page 97).
- [415] J. Tello et al. “Major outbreaks in the nineteenth century shaped grape *Phylloxera* contemporary genetic structure in Europe”. In: *Scientific Reports* 9 (2019), page 17540. DOI: [10.1038/s41598-019-54122-0](https://doi.org/10.1038/s41598-019-54122-0) (cited on pages 78, 89).
- [416] B. L. Teviotdale and J. H. Connell. “Almond Leaf Scorch”. In: *ANR University of California* (2003), page 8106. DOI: [10.3733/ucanr.8106](https://doi.org/10.3733/ucanr.8106) (cited on pages 61, 63).
- [417] Terry M Therneau. *A Package for Survival Analysis in R*. R package version 3.4-0. 2022. URL: <https://CRAN.R-project.org/package=survival> (cited on page 90).
- [418] Horst R. Thieme. “Convergence results and a Poincaré-Bendixson trichotomy for asymptotically autonomous differential equations”. In: *Journal of Mathematical Biology* 30.4 (1992), pages 755–763. DOI: [10.1007/BF00173267](https://doi.org/10.1007/BF00173267) (cited on page 49).
- [419] J. Tintoré and B. Casas Pérez. *Buoy Bahía de Palma Physicochemical parameters of sea water data [Data set]*. Balearic Islands Coastal Observing and Forecasting System, SOCIB. 2022. DOI: [doi.org/10.25704/S6JB-CK61](https://doi.org/10.25704/S6JB-CK61). URL: [https://apps.socib.es/data-catalog/#/data-products/buoy\\_bahiadepalma\\_physicochemical\\_parameters](https://apps.socib.es/data-catalog/#/data-products/buoy_bahiadepalma_physicochemical_parameters) (cited on page 141).
- [420] Fiona M. Tomley and Martin W. Shirley. “Livestock infectious diseases and zoonoses”. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364.1530 (Sept. 2009), pages 2637–2642. ISSN: 1471-2970. DOI: [10.1098/rstb.2009.0133](https://doi.org/10.1098/rstb.2009.0133) (cited on page 30).
- [421] S. Trigos et al. “Utilization of muddy detritus as organic matter source by the fan mussel *Pinna nobilis*”. In: *Mediterranean Marine Science* 15.3 (2014), pages 667–674. ISSN: 1791-6763. DOI: [10.12681/mms.836](https://doi.org/10.12681/mms.836) (cited on page 9).
- [422] J. E. Truscott and C. A. Gilligan. “Response of a deterministic epidemiological system to a stochastically varying environment”. In: *Proceedings of the National Academy of the USA* 100 (15 2003), pages 9067–9072. DOI: [10.1073/pnas.1436273100](https://doi.org/10.1073/pnas.1436273100) (cited on page 79).
- [423] K. P. Tumber, J. M. Alston, and K. B. Fuller. “Pierce’s disease costs California 104 million per year”. In: *California Agriculture* 68 (1 2014), pages 20–29. DOI: [10.3733/ca.v068n01p20](https://doi.org/10.3733/ca.v068n01p20) (cited on page 78).
- [424] Kabir B. Tumber, Julian M. Alston, and Kate B. Fuller. “Pierce’s disease costs California \$104 million per year”. In: *California Agriculture* 68.1 (2014), pages 20–29. DOI: [10.3733/ca.v068n01p20](https://doi.org/10.3733/ca.v068n01p20). URL: <https://ucanr.edu/repository/fileAccessPublic.cfm?fn=cav6801p20-133590.pdf> (cited on pages 46, 96, 110).
- [425] Alan Mathison Turing. “The Chemical Basis of Morphogenesis”. In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 237 (1952), pages 37–72. DOI: [10.1098/rstb.1952.0012](https://doi.org/10.1098/rstb.1952.0012) (cited on page 130).

- [426] UNEP-WCMC et al. *Global distribution of warm-water coral reefs, compiled from multiple sources (listed in "Coral\_Source.mdb"), and including IMaRS-USF and IRD (2005), IMaRS-USF (2005) and Spalding et al. (2001)*. Cambridge (UK): UNEP World Conservation Monitoring Centre. 2010. URL: <http://data.unep-wcmc.org/datasets/13> (cited on page 124).
- [427] Frank Van den Bosch and Michael J. Jeger. "The basic reproduction number of vector-borne plant virus epidemics". In: *Virus Research* 241 (2017), pages 196–202. ISSN: 0168-1702. DOI: [10.1016/j.virusres.2017.06.014](https://doi.org/10.1016/j.virusres.2017.06.014) (cited on page 46).
- [428] Pauline van den Driessche. "Reproduction numbers of infectious disease models". In: *Infectious Disease Modelling* 2.3 (2017), pages 288–303. ISSN: 2468-0427. DOI: [10.1016/j.idm.2017.06.002](https://doi.org/10.1016/j.idm.2017.06.002) (cited on page 46).
- [429] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN: 1441412697 (cited on page 100).
- [430] M. Vanhove et al. "Genomic Diversity and Recombination among *Xylella fastidiosa* Subspecies". In: *Applied and Environmental Microbiology* 85 (13 2019), e02972–18. DOI: [10.1128/AEM.02972-18](https://doi.org/10.1128/AEM.02972-18) (cited on pages 78, 88, 92).
- [431] M. Vanhove et al. "Genomic diversity and recombination among *Xylella fastidiosa* subspecies". In: *Applied and Environmental Microbiology* 85.13 (2019), e02972–18. DOI: [10.1128/AEM.02972-18](https://doi.org/10.1128/AEM.02972-18) (cited on page 60).
- [432] Manuel Vargas-Yáñez et al. "Warming trends and decadal variability in the Western Mediterranean shelf". In: *Global and Planetary Change* 63.2-3 (2008), pages 177–184. DOI: [10.1016/j.gloplacha.2007.09.001](https://doi.org/10.1016/j.gloplacha.2007.09.001) (cited on page 134).
- [433] Manuel Vargas-Yáñez et al. "Climate change in the Western Mediterranean sea 1900-2008". In: *Journal of Marine Systems* 82.3 (2010), pages 171–176. DOI: [10.1016/j.jmarsys.2010.04.013](https://doi.org/10.1016/j.jmarsys.2010.04.013) (cited on page 134).
- [434] Maite Vázquez-Luis et al. "S.O.S. *Pinna nobilis*: A Mass Mortality Event in Western Mediterranean Sea". In: *Frontiers in Marine Science* 4 (2017), page 220. ISSN: 2296-7745. DOI: [10.3389/fmars.2017.00220](https://doi.org/10.3389/fmars.2017.00220) (cited on pages 9, 30).
- [435] A Velo et al. "Total alkalinity estimation using MLR and neural network techniques". In: *Journal of Marine Systems* 111 (2013), pages 11–18. DOI: [10.1016/j.jmarsys.2012.09.002](https://doi.org/10.1016/j.jmarsys.2012.09.002) (cited on page 139).
- [436] Robert C. Venette et al. "Pest Risk Maps for Invasive Alien Species: A Roadmap for Improvement". In: *BioScience* 60.5 (May 2010), pages 349–362. ISSN: 0006-3568. DOI: <https://doi.org/10.1525/bio.2010.60.5.5> (cited on page 105).
- [437] B. Vidondo et al. "Some aspects of the analysis of size spectra in aquatic ecology". In: *Limnology and Oceanography* 42.1 (1997), pages 184–192. DOI: [10.4319/lo.1997.42.1.0184](https://doi.org/10.4319/lo.1997.42.1.0184). eprint: <https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1997.42.1.0184>. URL: <https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lo.1997.42.1.0184> (cited on page 125).
- [438] P. Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pages 261–272. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2) (cited on page 63).
- [439] Maurizio Vurro, Barbara Bonciani, and Giovanni Vannacci. "Emerging infectious diseases of crop plants in developing countries: impact on agriculture and socio-economic consequences". In: *Food Security* 2.2 (June 2010), pages 113–132. ISSN: 1876-4525. DOI: [10.1007/s12571-010-0062-7](https://doi.org/10.1007/s12571-010-0062-7) (cited on page 30).
- [440] A. K. Wallingford et al. "Expansion of the range of Pierce's disease in Virginia". In: *Plant Health Progress* (2007). DOI: [10.1094/PHP-2007-1004-01-BR](https://doi.org/10.1094/PHP-2007-1004-01-BR) (cited on page 92).
- [441] Jessica R Ward and Kevin D Lafferty. "The Elusive Baseline of Marine Disease: Are Diseases in Ocean Ecosystems Increasing?" In: *PLoS Biology* 2.4 (Apr. 2004). Edited by Larry Crowder, e120. ISSN: 1545-7885. DOI: [10.1371/journal.pbio.0020120](https://doi.org/10.1371/journal.pbio.0020120) (cited on page 8).
- [442] Hui-Ming Wei, Xue-Zhi Li, and Maia Martcheva. "An epidemic model of a vector-borne disease with direct transmission and time delay". In: *Journal of Mathematical Analysis and Applications* 342.2 (2008), pages 895–908. ISSN: 0022-247X. DOI: [10.1016/j.jmaa.2007.12.058](https://doi.org/10.1016/j.jmaa.2007.12.058) (cited on page 49).
- [443] RF Weiss and BA Price. "Nitrous oxide solubility in water and seawater". In: *Marine chemistry* 8.4 (1980), pages 347–359 (cited on page 142).

- [444] J. M. Wells et al. “*Xylella fastidiosa* gen. nov., sp. nov: Gram-Negative, Xylem-Limited, Fastidious Plant Bacteria Related to *Xanthomonas* spp.” In: *International Journal of Systematic and Evolutionary Microbiology* 37 (2 1987), pages 136–143. DOI: [10.1099/00207713-37-2-136](https://doi.org/10.1099/00207713-37-2-136) (cited on page 78).
- [445] Curtis L. Wesley and Linda J.S. Allen. “The basic reproduction number in epidemic models with periodic demographics”. In: *Journal of Biological Dynamics* 3.2-3 (2009), pages 116–129. DOI: [10.1080/17513750802304893](https://doi.org/10.1080/17513750802304893) (cited on page 53).
- [446] Geoffrey B. West, J. H. Brown, and B. J Enquist. “Scaling in biology: patterns and processes, causes and consequences”. In: *Scaling in biology*. Edited by J. H. Brown and Geoffrey B. West. Oxford, UK: Oxford University Press, 2000, pages 1–24. ISBN: 9780195131420 (cited on page 129).
- [447] What is GBIF? <https://www.gbif.org/what-is-gbif>. URL: <https://www.gbif.org/what-is-gbif> (visited on 07/03/2023) (cited on pages 99, 117).
- [448] S. M. White et al. “Modelling the spread and control of *Xylella fastidiosa* in the early stages of invasion in Apulia, Italy”. In: *Biological Invasions* 19.6 (June 2017), pages 1825–1837. ISSN: 1573-1464. DOI: [10.1007/s10530-017-1393-5](https://doi.org/10.1007/s10530-017-1393-5). URL: <https://doi.org/10.1007/s10530-017-1393-5> (cited on page 60).
- [449] S. M. White et al. “Estimating the epidemiology of emerging *Xylella fastidiosa* outbreaks in olives”. In: *Plant Pathology* 69.8 (2020), pages 1403–1413. DOI: <https://doi.org/10.1111/ppa.13238>. URL: <https://bsppjournals.onlinelibrary.wiley.com/doi/abs/10.1111/ppa.13238> (cited on pages 60, 63, 91).
- [450] Fact sheets of vector-borne diseases. 2018. URL: [https://www.hiobs.org/lib/file/manager/Vector-Borne\\_Disease\\_General\\_Fact\\_Sheet.pdf](https://www.hiobs.org/lib/file/manager/Vector-Borne_Disease_General_Fact_Sheet.pdf) (cited on page 46).
- [451] Carlie Wiener and Amara Davis. “Exploration Down Under”. In: *Oceanography* 34.1 (2021), pages 58–67. URL: <https://www.jstor.org/stable/10.2307/27218318> (cited on page 124).
- [452] Cort J. Willmott and Johannes J. Feddema. “A More Rational Climatic Moisture Index”. In: *The Professional Geographer* 44.1 (Feb. 1, 1992), pages 84–88. ISSN: 0033-0124. DOI: <https://doi.org/10.1111/j.0033-0124.1992.00084.x>. URL: <https://doi.org/10.1111/j.0033-0124.1992.00084.x> (visited on 06/19/2023) (cited on page 100).
- [453] Kenneth G. Wilson. “Problems in Physics with Many Scales of Length”. In: *Scientific American* 241 (1979), pages 158–179. DOI: [10.1038/scientificamerican0879-158](https://doi.org/10.1038/scientificamerican0879-158) (cited on page 129).
- [454] Ryan J Woosley. “Evaluation of the temperature dependence of dissociation constants for the marine carbon system using pH and certified reference materials”. In: *Marine Chemistry* 229 (2021), page 103914. DOI: [10.1016/j.marchem.2020.103914](https://doi.org/10.1016/j.marchem.2020.103914) (cited on page 142).
- [455] Koffi Marcellin Yao et al. “Time variability of the north-western Mediterranean Sea pH over 1995–2011”. In: *Marine Environmental Research* 116 (2016), pages 51–60. DOI: [10.1016/j.marenvres.2016.02.016](https://doi.org/10.1016/j.marenvres.2016.02.016) (cited on pages 134, 139).
- [456] D. G. Zawada and J. C. Brock. “A Multiscale Analysis of Coral Reef Topographic Complexity Using Lidar-Derived Bathymetry”. In: *Journal of Coastal Research* 2009.10053 (2009), pages 6–15. DOI: [10.2112/SI53-002.1](https://doi.org/10.2112/SI53-002.1). URL: <https://doi.org/10.2112/SI53-002.1> (cited on page 124).
- [457] Noa Zecharia et al. “*Xylella fastidiosa* Outbreak in Israel: Population Genetics, Host Range, and Temporal and Spatial Distribution Analysis”. In: *Phytopathology* 112.11 (2022), pages 2296–2309. DOI: <https://doi.org/10.1094/PHYTO-03-22-0105-R> (cited on pages 96, 117).
- [458] Shi Zhao et al. “Modelling the effective reproduction number of vector-borne diseases: the yellow fever outbreak in Luanda, Angola 2015–2016 as an example”. In: *PeerJ* 8 (Feb. 2020), e8601–e8601. ISSN: 2167-8359. DOI: [10.7717/peerj.8601](https://doi.org/10.7717/peerj.8601) (cited on page 49).
- [459] Maria Zotou et al. “*Pinna nobilis* in the Greek seas (NE Mediterranean): on the brink of extinction?” In: *Mediterranean Marine Science* 21.3 (July 2020), pages 575–591. ISSN: 1791-6763. DOI: [10.12681/mms.23777](https://doi.org/10.12681/mms.23777) (cited on page 9).
- [460] Serena Zunino et al. “Impact of ocean acidification on ecosystem functioning and services in habitat-forming species and marine ecosystems”. In: *Ecosystems* 24.7 (2021), pages 1561–1575. DOI: [10.1007/s10021-021-00601-3](https://doi.org/10.1007/s10021-021-00601-3) (cited on page 134).

