Integrating species distribution model and interacting particle systems to predict the spread of an invasive alien plant

M.G. Smolik, S. Dullinger, F. Essl

Presentation by: Emmanuel, Andreas & Dietmar

Outline

- Introduction
- Materials & Methods
- Model evaluation
- Results
- Discussion
- Limitations of the modelling approach
- Conclusion

Introduction

Ambrosia artemisiifolia (ragweed)

- annual forb
- produces strong allergenic pollen
- impacts on human health: hay fever
- Native to central USA
- Invaded: Europe, Asia (India), northern & southern Africa, Australia, New Zealand and Hawaii

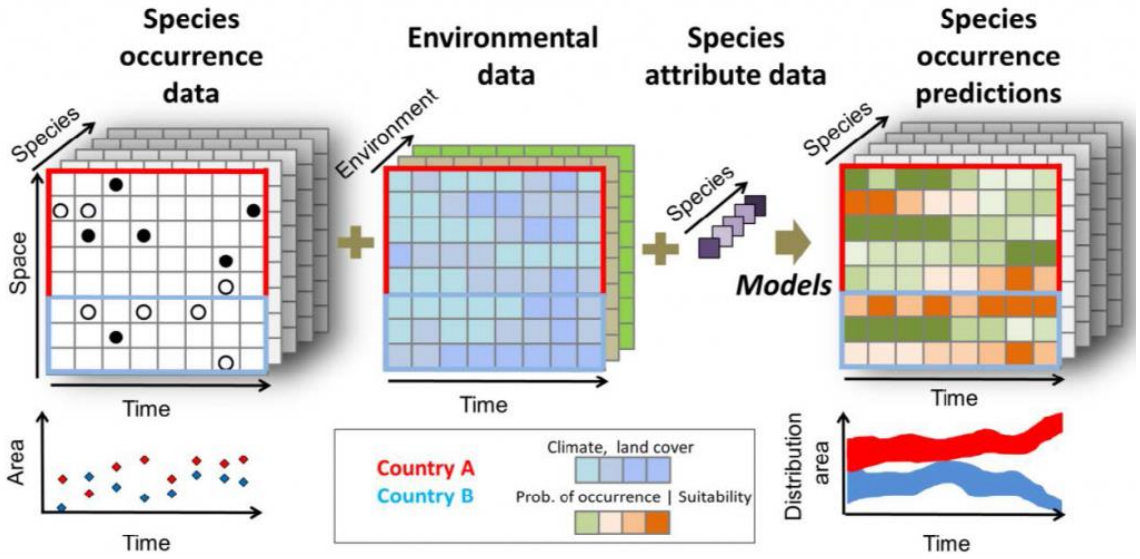
- Study location: Austria

Background: Modelling approach

- Previous knowledge gap:
  - the potential spatial spread of invasive species
- Previously used modelling approaches:
  - reaction-diffusion models
  - integro-difference frameworks
  - limitations:
    - cannot account for complex spatial patterns
    - operates more on landscape scales
    - habitat suitability is completely ignored!
    - e.g. IPS, cellular automata model
- IPS + SDMs = hybrid model
  - simulate invasive spread
  - driven by dispersal parameters and habitat suitability

Materials & Methods

1. Species distribution models (SDMs) → *HABITAT (H(x))*

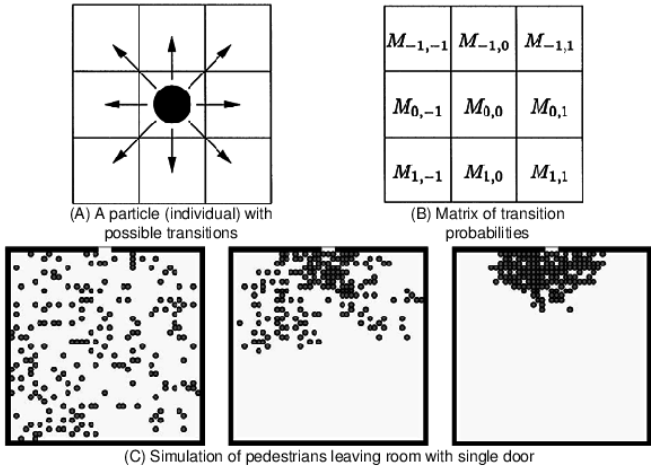


- Generalized linear models (GLMs)
- Generalized additive models (GAMs)

Materials & Methods

2. Interacting particle systems (IPSs) – SPREAD

- aka stochastic spatial models
- investigates biological spread processes
- each grid cells represents individual sites
- discrete time steps and states
- State transition rules, depending on
  - cell previous states
  - states of neighbouring cells



(A) A particle (individual) with possible transitions

(B) Matrix of transition probabilities

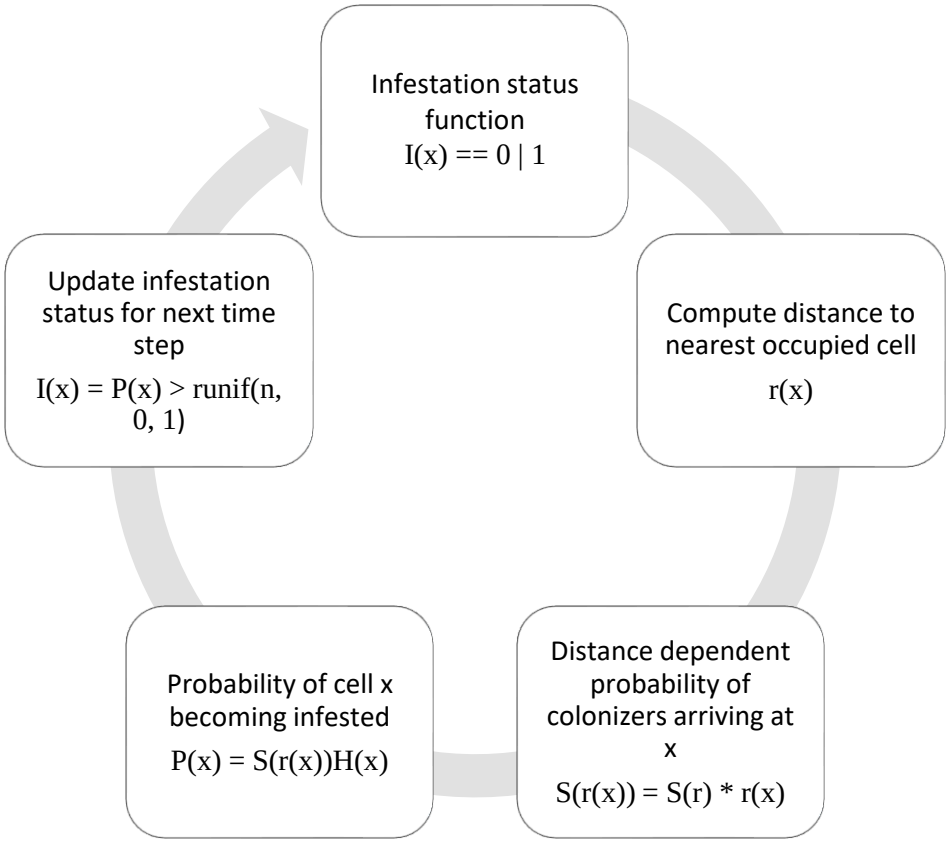
(C) Simulation of pedestrians leaving room with single door

-> Generalized in cellular automata model

Materials & Methods

3. Integration of IPS and SDM (COMBINED)

-> IPS accounts for spatial variation in state transitions with respect to *H(x)*



Distance dependent colonization of sites by population (Gaussian)

$S(r) == a \exp(-r^2/2d^2)$

Materials & Methods

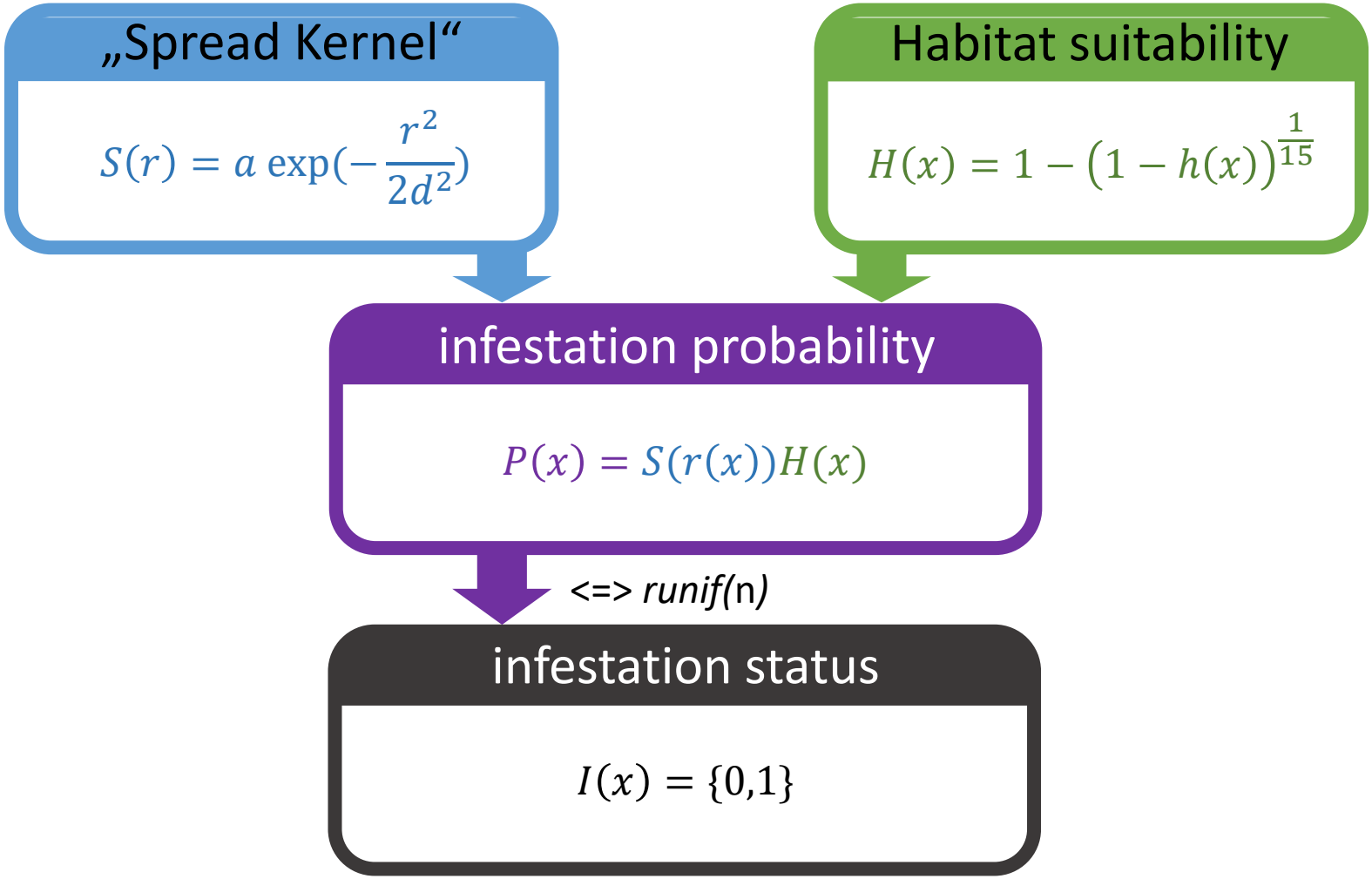
Species distribution data

- Datasets from Austria and Germany
  - Wider range of environmental conditions
  - estimate species limit more accurately
- Records from 1990 to 2005
- Data sources:
  - Floristic Mapping of Central Europe (FMC) database
  - Literature review
- Resolution: 3 x 5 min (30 km^2)

Evironmental variables

Variable	Source	Original scale
1. Temperature	ATEAM Project	10 x 10
2. Precipitation	ATEAM Project	10 x 10
3. River network	Digital Austrian Map	1:50k to 1:200k
4. Topography: Elevation	DEM	250m
5. Land cover/habitat quality	CORINE land cover	100m
6. Human traffic	Austria: Tele Atlas N.V. Germany: Digital Basic Landscape	1:50k to 1:200k

Model


„Spread Kernel“
$S(r) = a \exp(-\frac{r^2}{2d^2})$
Habitat suitability
$H(x) = 1 - (1 - h(x))^{\frac{1}{15}}$
infestation probability
$P(x) = S(r(x))H(x)$
<=> runif(n)
infestation status
$I(x) = \{0,1\}$

Model evaluation

- Maximum Likelihood

The probability density of observing a single data point *x*, that is generated from a Gaussian distribution is given by:

$$P(x;\mu,\sigma) = \frac{1}{\sigma * \sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

infestation probability

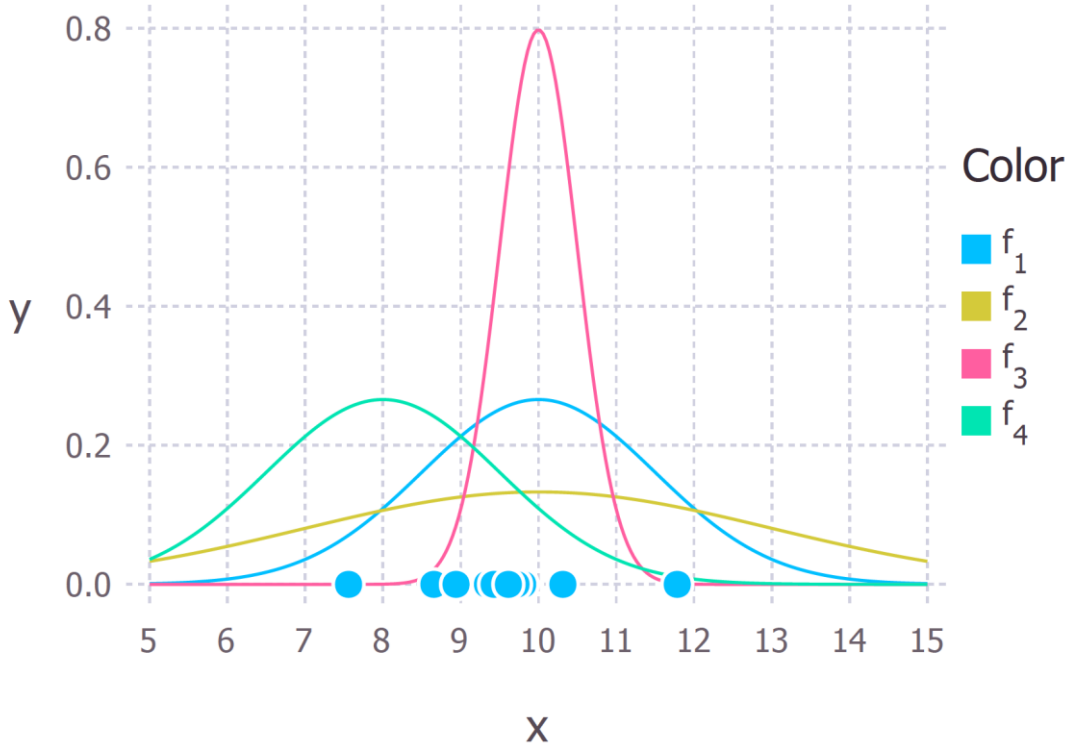
$$P(x) = S(r(x))H(x)$$

„Spread Kernel“

$$S(r) = a\exp(-\frac{r^2}{2d^2})$$

Model evaluation

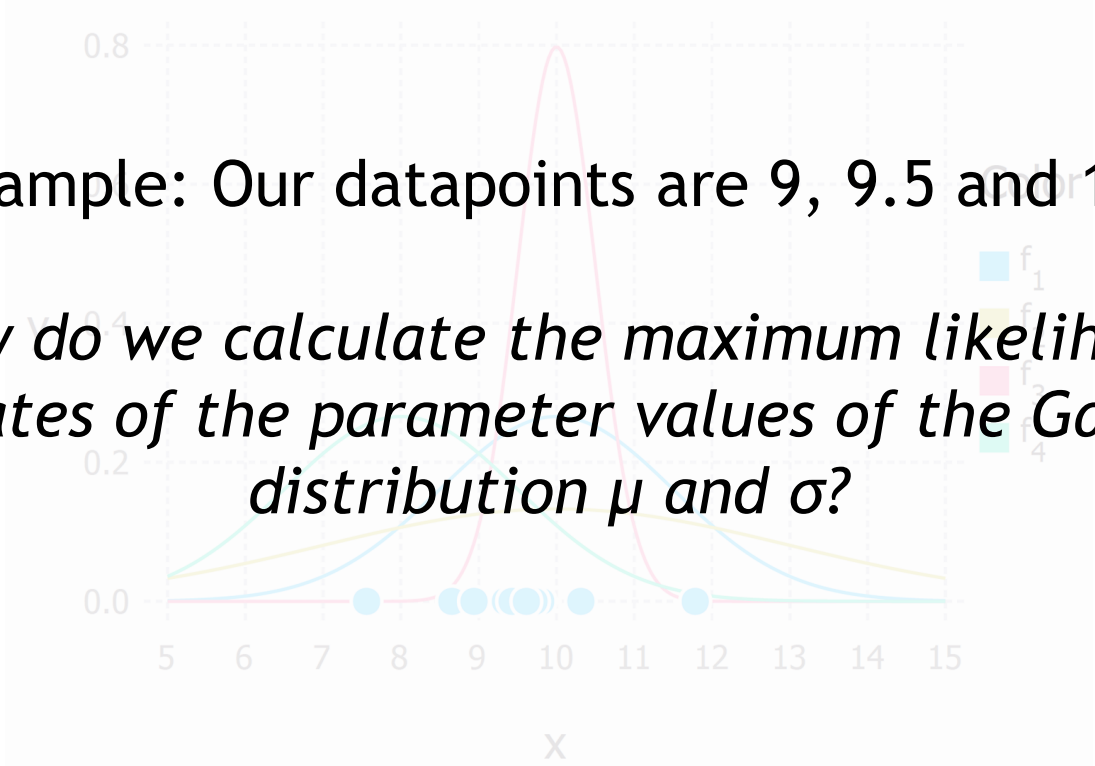
- Maximum Likelihood



The 10 data points and possible Gaussian distributions from which the data were drawn. The goal of maximum likelihood is to find the parameter values that give the distribution that maximize the probability of observing the data.

Model evaluation

- Maximum Likelihood



Example: Our datapoints are 9, 9.5 and 11.

How do we calculate the maximum likelihood estimates of the parameter values of the Gaussian distribution μ and σ?

Model evaluation

- Maximum Likelihood

The probability density of observing a single data point x, that is generated from a Gaussian distribution is given by:

$$P(x;\mu,\sigma) = \frac{1}{\sigma * \sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

$$P(x) = S(r(x))H(x) \qquad S(r) = a\exp(-\frac{r^2}{2d^2})$$

Model evaluation

- Log Likelihood
- Maximum Likelihood Estimates

$$\frac{\partial \ln(P(x;\mu,\sigma))}{\partial \mu} = \frac{1}{\sigma^2}[9 + 9.5 + 11 - 3\mu]$$

$$\mu = \frac{9 + 9.5 + 11}{3} = 9.833$$

Find the MLE of the mean, μ, by taking the partial derivative of the function with respect to μ

Model evaluation

- Spread kernel parameter estimates

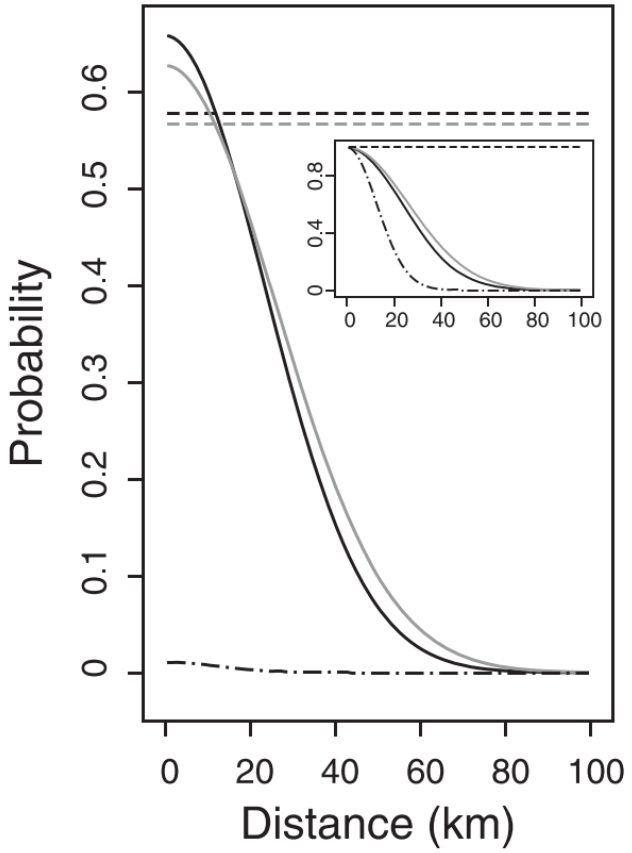
Model	Kernel height (a)	Kernel width (d; km)
1. COMBINED$_{GLM}$	0.659	23.43
2. HABITAT$_{GLM}$	0.579	Inf
3. COMBINED$_{GAM}$	0.628	26.05
4. HABITAT$_{GAM}$	0.567	Inf
5. SPREAD	0.011	12.58

$$S(r) = a \exp(-\frac{r^2}{2d^2}) \qquad H(x) = 1 - \left(1 - h(x)\right)^{\frac{1}{15}}$$

$$P(x) = S(r(x))H(x)$$

Assuming gaussian distribution and independent datapoints.

Model evaluation



Combined GLM

Combined GAM

Habitat GLM

Habitat GAM

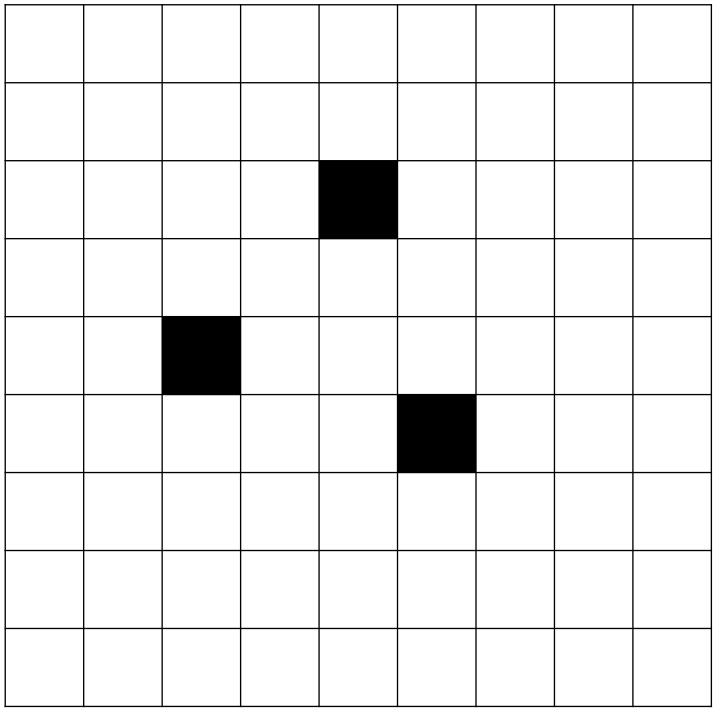
Spread

Optimised spread kernels *S(r)*

Assuming Gaussian distribution and independent datapoints.

Model evaluation

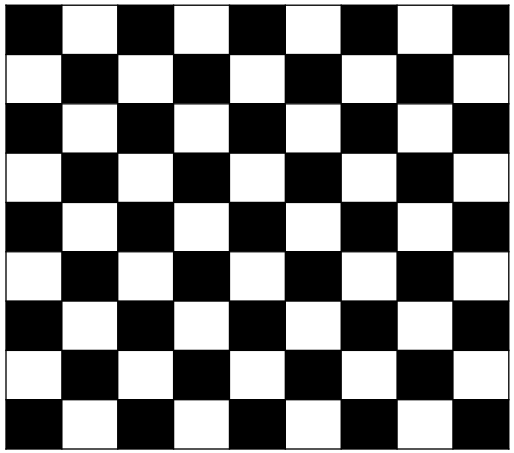
- Moran's *I*



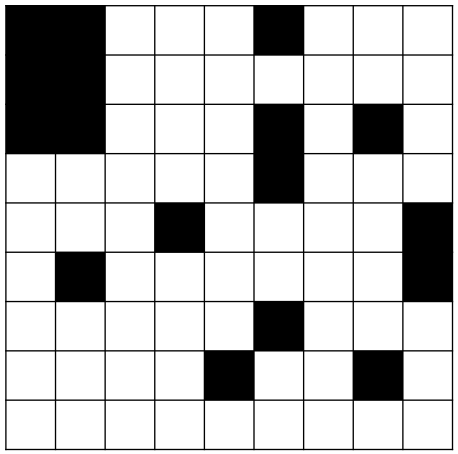
… independent datapoints?

Model evaluation

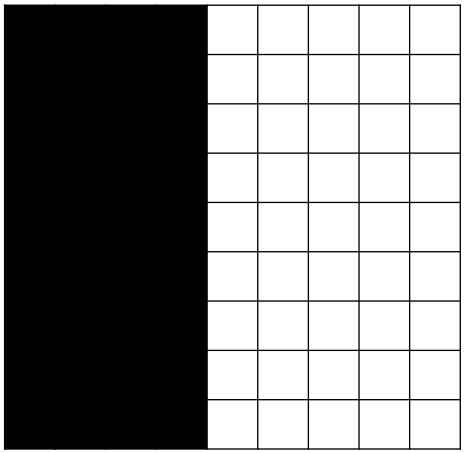
- Moran‘s *I*



Perfect dispersion:
Moran‘s *I* equals -1

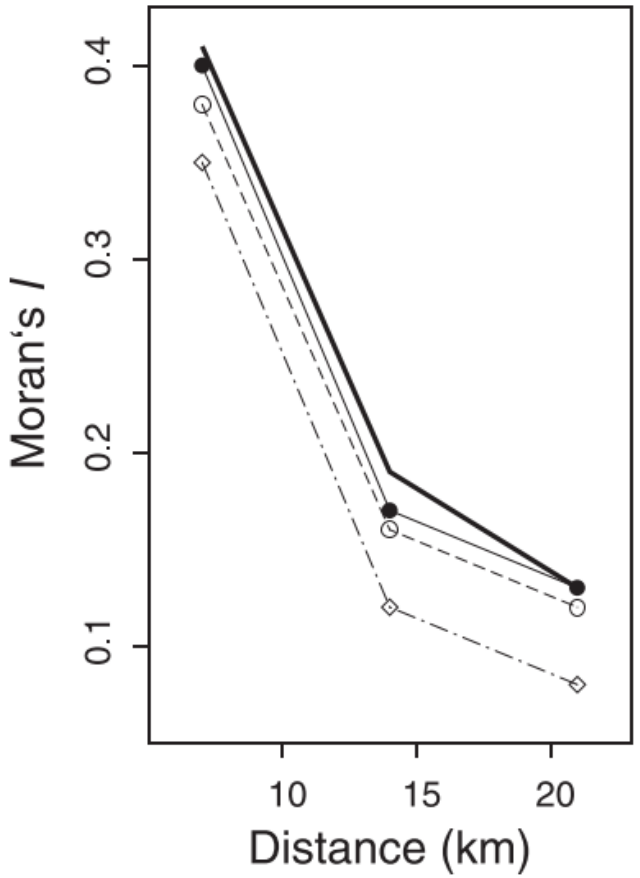


Perfect Randomness:
Moran‘s *I* equals 0



Perfect Clustering:
Moran‘s *I* equals +1

Model evaluation



- Moran's *I*

Combined GLM

Habitat GLM

Spread

Real

Autocorrelograms illustrating the spatial autocorrelation

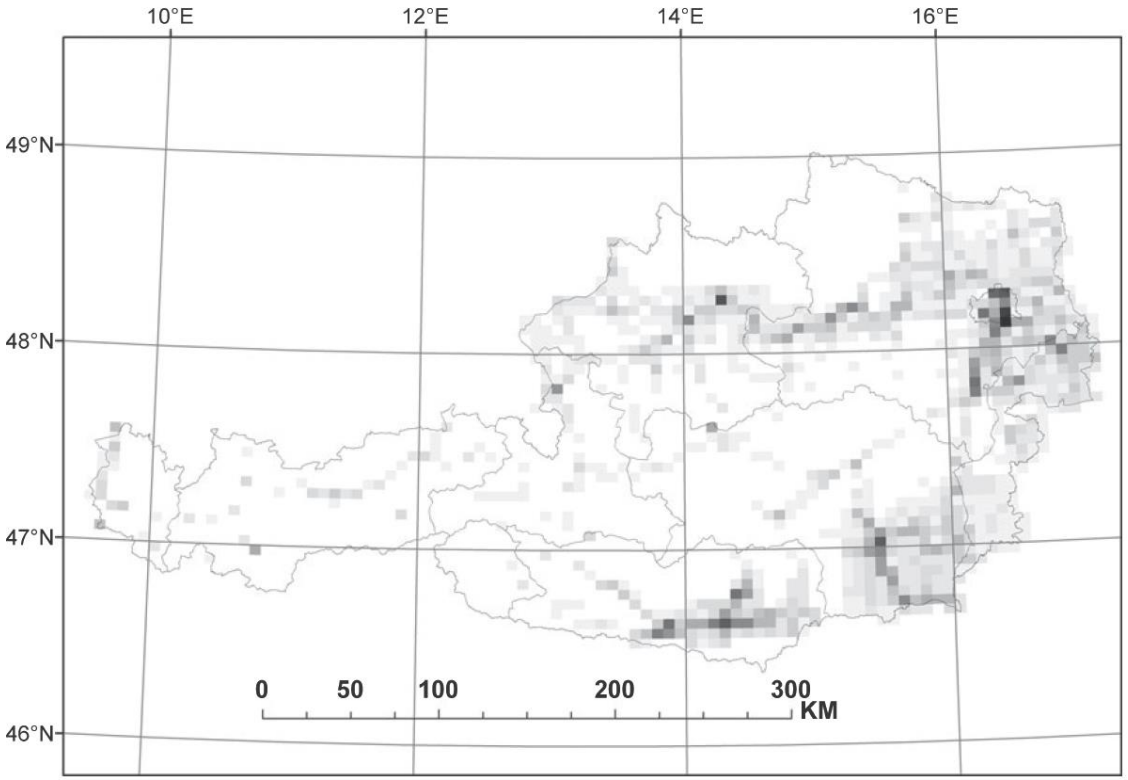
Model evaluation

- AUC = Area Under the Curve.
- AUROC = Area Under the Receiver Operating Characteristic curve.

Model	AUC
1. $\text{COMBINED}_{\text{GLM}}$	0.83
2. $\text{HABITAT}_{\text{GLM}}$	0.82
3. $\text{COMBINED}_{\text{GAM}}$	0.83
4. $\text{HABITAT}_{\text{GAM}}$	0.83
5. SPREAD	0.60

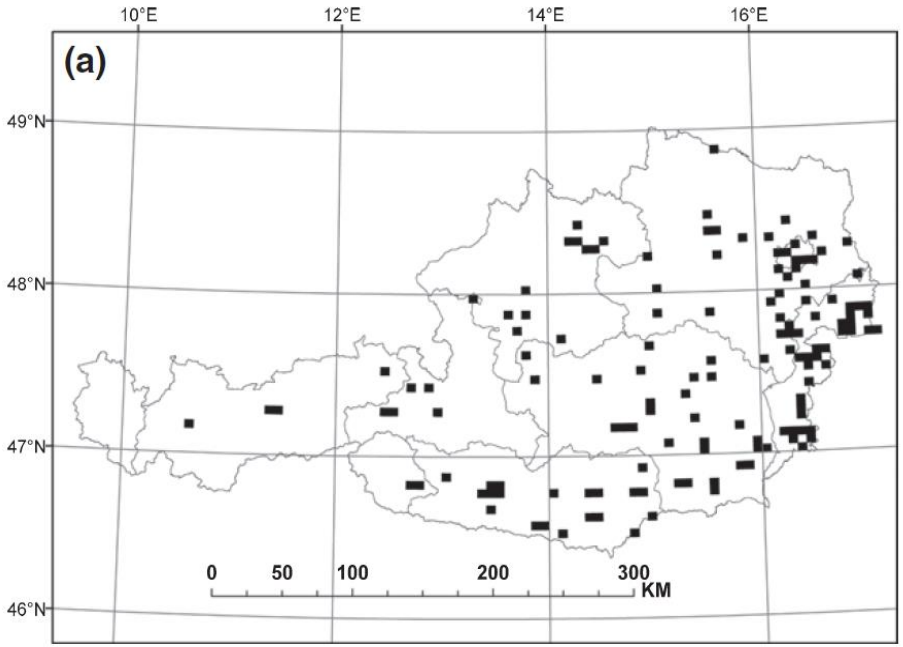
AUC values ranges between 0 and 1	
AUC = 0	Perfect inverse discrimination
AUC = 0.5	random predictions
0.7 < AUC < 0.8	Fair predictions
0.8 < AUC < 0.9	Good predictions
AUC > 0.9	Excellent predictions

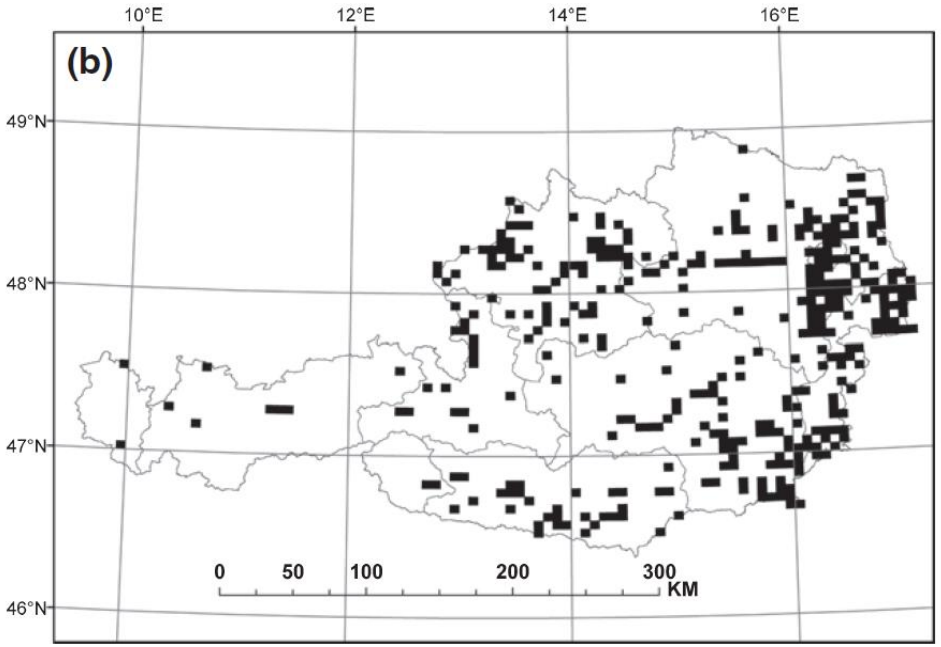
Results



- Rescaled habitat suitability map for Ambrosia artemisiifolia
- GLM with observed data till 2005
- squares represent the cells of the 3· 5 min grid
- Suitability increases from white to black

Results





(a) Observed distribution in 1990

(b) observed distribution in 2005

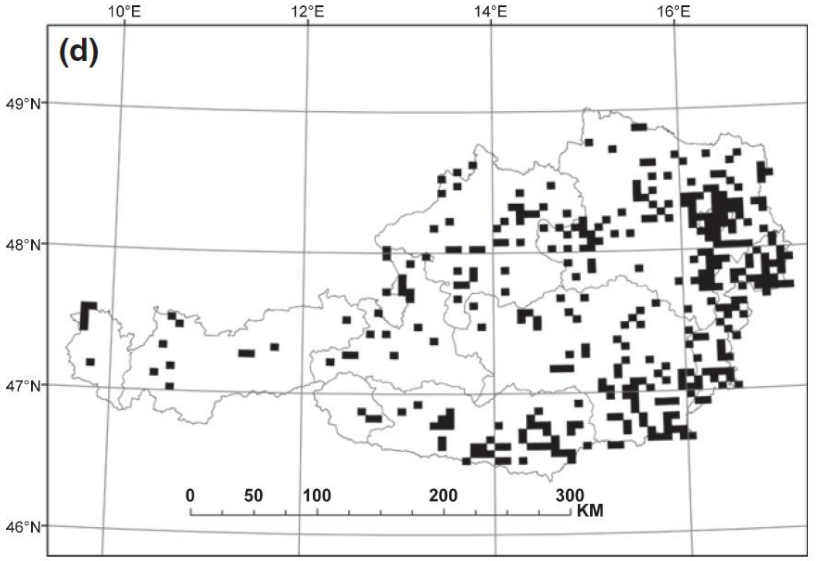
Results

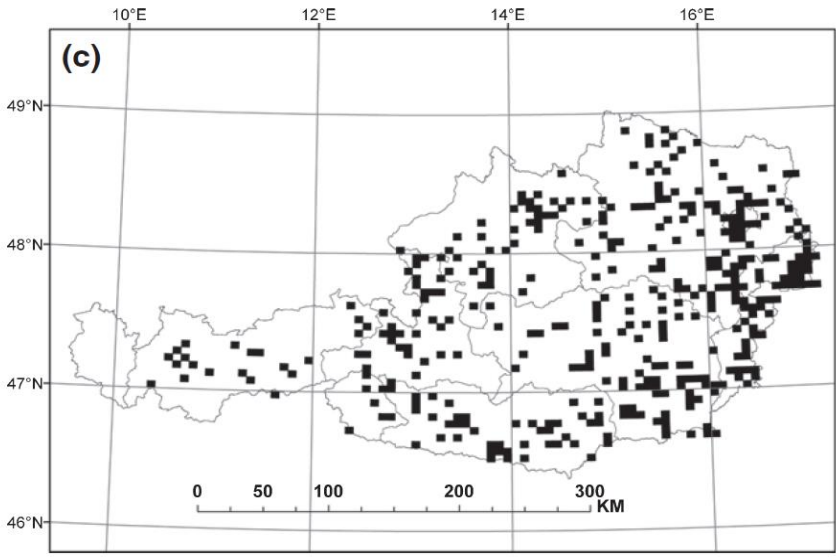
(c–e) distribution in 2005 as predicted by simulation runs of

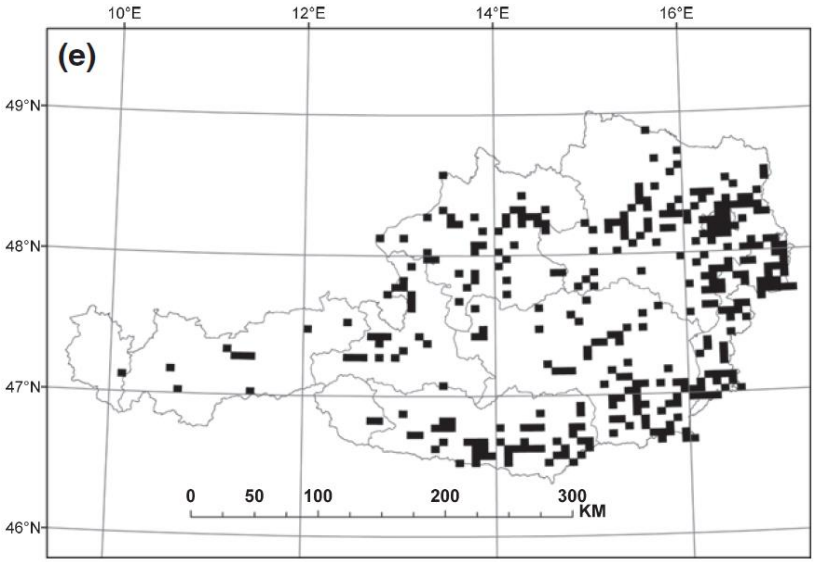
(c) Spread

(d) Habitat GLM

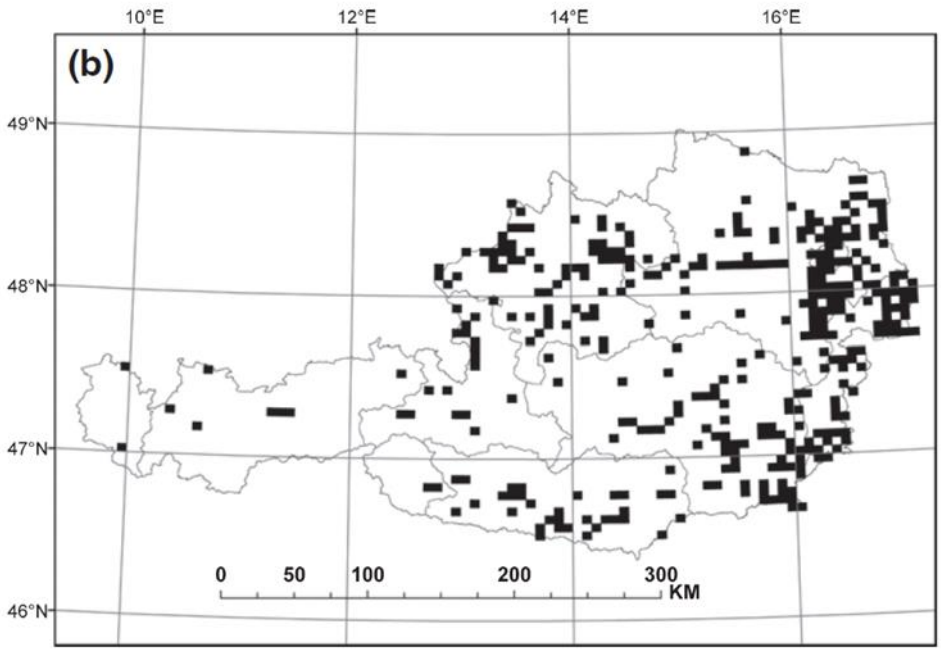
(e) Combined GLM



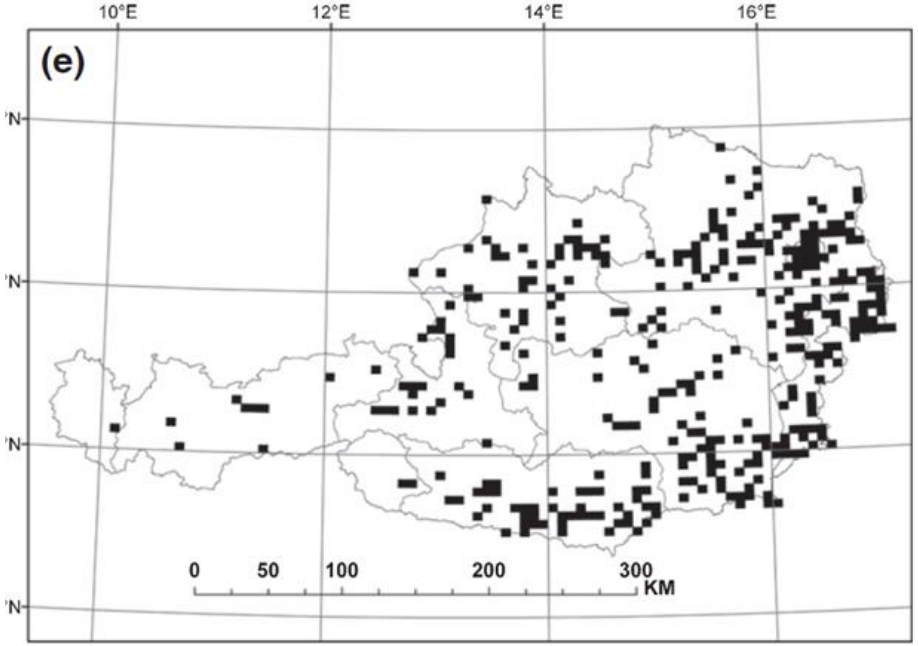




Results



(b) observed distribution in 2005



(e) Combined GLM

Discussions

- results of our model comparison confirm the combined effect of habitat and dispersal constraints
- Combined slightly better
- Combined could work better in less autocorrelated or larger areas
- Combined model has the advantage that environmental changes can be consistently taken into account
- Many invasive plant species thought to be currently limited or at least slowed by climatic constraints

Limitations of the modelling approach

- Model only takes the nearest occupied cell into account, not the distances to all potential seed sources
- Model does not account for possible extinction in cells
- Simple Gaussian instead of a leptokurtic function
- Dispersal events are influenced by a multitude of factors
- SDM assumes species' current distribution is in equilibrium with its environment
- spread probabilities for ragweed were derived from patterns of empirical data on a purely phenomenological basis
- Generic validity of parameter estimates questionable
- local population growth together with dispersal kernels, is not explicitly represented
- Combined model only takes the nearest occupied cell into account, not the distances to all potential seed sources

Conclusion

- Smolik et al. use a time series of *Ambrosia* spread from 1990 to 2005 to fit three models, one based solely on the time pattern of spread, one based solely on the pattern of habitat associations, and one model that combined the time pattern of spread with habitat associations.

- Spread model < Habitat model < Combined model

- In such hybrid models colonizations can be captured more appropriately as, for example, less suitable sites may occasionally function as stepping stones of migration when propagule input is high

- Huge amount of data and calculations necessary (question of practicality)

B4: Spatial Ecology 2019/2020