Project description

Lyme disease (a.k.a. Lyme borreliosis) is a vector-borne disease that is caused by tick bites. Ticks are spider-like animals that live in forests and meadows, mostly in the northern hemisphere. Ticks can live for three to four years, going through four life stages: egg, six-legged larva, eight-legged nymph, and adult. They need blood from an animal or human at every stage to survive. When biting, ticks suck themselves into the skin and consume blood for several hours or days. Ticks can be removed easily if identified in time. Yet, especially when remaining unnoticed, they can transfer many diseases including Lyme disease.

The abundance of ticks varies in space and time. Ticks are typically not active during winter because they require temperatures of at least 7 degrees Celsius to start questing. Questing ticks climb vegetation, for example, shrubs and high grasses, and wait for a host to attach to. Although ticks are mostly found in areas dominated by natural vegetation, they can also be found in campsites, gardens and other ”green” recreational areas.

As Lyme disease is a serious illness that can lead to disability, scientists work on mapping the abundance of ticks in the Netherlands by performing blanket dragging. At several sites, they pull a blanket over a stretch of natural vegetation and count the number of ticks on a monthly basis to monitor the number of ticks per life stage.

Tick bites are not only influenced by tick abundance, but human behavior is an important factor too. When performing activities with direct contact with vegetation, like climbing trees, berry picking or gardening, the chance of getting a tick bite increases. Other activities like cycling, walking and playing can also lead to tick bites. It is unlikely that all people are equally aware of the presence of ticks and that all take precautions to prevent tick bites.

Besides monitoring the abundance of ticks, the Netherlands also has a web portal called “Tekenradar” (https://www.tekenradar.nl/) where citizens can report their tick bites. Data from Tekenradar is available for your case study. You must realize, however, that like any other source of Volunteered Geographic Information (VGI), tick bite counts might be biased in space and time. Not all people are equally likely to report their tick bites. Hence, the number of reported tick bites is smaller than the total number of tick bites in the Netherlands. Moreover, when Tekenradar is in the news, the number of reports increases.

Theory and patterns

In the coming years, we would like to get more insight into the factors that determine the risk (R) of getting a tick bite. Risk is often defined as R = Hazard (H) x Exposure (E) x Vulnerability (V). The daily tick activity and abundance can approximate hazard at a certain location If there are no active ticks, for example, in winter, this factor will be zero, resulting in the absence of risk.The number of people at a certain location and the activities that these people undertake determines exposure.Not all activities have an equal risk. When you stay inside, the risk of getting a tick bite is negligible. Besides the activity and abundance of ticks and human presence and activities, there is a third factor that influences the risk of tick bites. Vulnerability refers to the characteristics of a person that make them more or less susceptible to getting a tick bite. By wearing protective clothes, a person can reduce the risk of getting a tick bite and, therefore, of getting Lyme disease.

When modelling, we often try to reproduce patterns. If we can create a model that simulates observed patterns, we know that the essential elements of the system under study have been captured. For tick bites, we can think of patterns in time. When plotting the yearly number of tick bites, we see that tick bites increase in spring, have a peak in late spring and summer, and that numbers decline in autumn. Different spatial locations have different tick bite curves over time (may peak at different moments). This is one of the aspects requiring further exploration.

Assignment

In this group project, you will create an Agent-Based Model (ABM) to investigate the impact of tick dynamics (i.e. activity and abundance) and human activity on the chances of getting a tick bite. You will also create a complementary data-driven (machine learning; ML) model that could either be integrated into the ABM model or be used to compare model outputs. The former option leads to a hybrid model where ML is used in one of the following three ways:

a. ML as a preprocessing step for ABM (e.g. to understand important features or help to parametrize the ABM model)
b. ML to steer the behavior of agents during the simulation (e.g. dynamically adapting their behavior)
c. ML for postprocessing output generated by the ABM (e.g. to capture output patterns or to understand the most important features of the ABM model)

As a full integration of ABM and ML is technically difficult (item b.) this will not be part of the assignment.

Agent-based modeling

The main structure of the ABM will be created during the exercises. This model will be further developed with your group members. Each group will create a model that includes both tick dynamics and tick bites. To increase the diversity in models, we selected three study areas: de Bilt, Schiermonnikoog and Ede. However, you can also select another study area

The fact that you have been assigned a study area does not mean that you will only run your model for this area or that your model should be restricted to a single municipality. We encourage you to think about scaling up your project area, or running for a sample of different municipalities selected based on transparent selection criteria.

Machine learning

Several ML tasks and algorithms will be presented to you during the course. Each group will select a specific ML task to be used in their model(s). Group members, in agreement with the practical supervisor, will have to decide the role of ML in their assignment (i.e. using ML as a pre- or post-processing task,). All groups will perform exploratory (spatial) data analysis. To provide students with some degree of flexibility, we provide a list of options (and restrictions) when defining your ML task. At the end, each group should decide on a specific and unique ML task. See the list below:

- Your target feature can be either “tick dynamics” or “tick bites”.
- You can choose whether to perform E(S)DA on the input or the output data of your model.
  - However, you shall all do this type of analysis.
- You can perform a clustering, classification, or regression task.
  - However, there can be only two groups working on the same task.
    - groups working on the same tasks will apply different algorithms
- You can apply the ML task in one of the following scenarios:
  - As a pre-processing task
  - As a post processing task

Validation and reflection

At the end of the course, a reflection will take place, including a comparison of project results. This reflection must be part of the final report (discussion of results and conclusions). It must be included in the group presentation and, whenever possible, in the individual peer review (critical reflection of the reviewed work).

Different types of comparison are possible in this project. You can compare your model with ground truth data, but you can also compare models developed for different areas (ABM part) and different ML algorithms (ML part). Ground truth will be made available to calibrate/validate your model.

To be submitted

For this assignment you will submit the following elements via Canvas:

- Project report
    - A document that explains describes your ABM model (use the ODD protocol to structure your description), the results of your ABM model and the role and results of the ML elements designed to study ticks in the Netherlands. This document will also include a detailed list of the tasks performed by each of the team members.
- Code:
    - A zipped folder with your ABM and ML models (Netlogo and python files/scrips, including all the necessary input files to (re)run the models)
- Presentation:
    - The slides your prepared for the presentation

Datasets

The available data for this assignment can be downloaded via Canvas.

Assessment

The rubric for this assignment is included in appendix 1.

Possible further reading

- Alkishe, A. A., et al. (2017). "Climate change influences on the potential geographic distribution of the disease vector tick Ixodes ricinus." PLOS ONE 12(12): e0189092.
- Garcia-Martí, I. et al. (2017a). "Identifying Environmental and Human Factors Associated With Tick Bites using Volunteered Reports and Frequent Pattern Mining," Trans. GIS, 21(2), 277–299.
- Garcia-Martí, I. et al. (2017b). Modelling and mapping tick dynamics using volunteered observations," Int. J. Health Geogr., 16(1), 41.
- Garcia-Marti, I., et al. (2018). "Using volunteered observations to map human exposure to ticks." Scientific Reports 8(1): 15435.
- Hofmeester, T. R., et al. (2017). "Deer presence rather than abundance determines the population density of the sheep tick, Ixodes ricinus, in Dutch forests." Parasit Vectors, 10(1): 433.
- Kjær, L. J., et al. (2019). "Predicting and mapping human risk of exposure to Ixodes ricinus nymphs using climatic and environmental data, Denmark, Norway and Sweden, 2016." Eurosurveillance, 24(9): 1800101.
- Sprong, H., et al. (2018). "Control of Lyme borreliosis and other Ixodes ricinus-borne diseases." Parasit Vectors, 11(1): 145.

- Swart, A., et al. (2014). "Predicting Tick Presence by Environmental Risk Mapping." Front Public Health. , 2(238).

Appendix 1 - Rubric

Group Project

Criteria	Ratings					Pts
Innovation: Ambition (breadth and/or depth)	4 Pts **Perfect**	3 Pts **Good**	2 Pts **Sufficient**	1 Pts **Low**	0 Pts **Insufficient**	4 pts
Data: Critical attitude on data quality	4 Pts **Perfect**	3 Pts **Good**	2 Pts **Sufficient**	1 Pts **Low**	0 Pts **Insufficient**	4 pts
Data: Exploratory data analysis contributes to solving the problem	4 Pts **Perfect**	3 Pts **Good**	2 Pts **Sufficient**	1 Pts **Low**	0 Pts **Insufficient**	4 pts
Method: ABM methods (implementation/Practical Part)	4 Pts **Perfect** ABM has been considerably improved from the baseline exercises. The quality of the implementation is perfect. The group show creativity in their work.	3 Pts **Good** ABM has been improved from the baseline exercise and the quality of the implementation is good	2 Pts **Sufficient** Improvement of the ABM is sufficient	1 Pts **Low** Severe shortcomings in the implementation of the ABM	0 Pts **Insufficient** Implementation of the ABM is Insufficient	4 pts
Method: ML methods (implementation/ Practical Part)	4 Pts **Perfect**	3 Pts **Good**	2 Pts **Sufficient**	1 Pts **Low**	0 Pts **Insufficient**	4 pts
Method: Integration of ML and ABM (implementation/ Practical Part)	4 Pts **Perfect**	3 Pts **Good**	2 Pts **Sufficient**	1 Pts **Low**	0 Pts **Insufficient**	4 pts
Method: Critical attitude on results	4 Pts **Perfect**	3 Pts **Good**	2 Pts **Sufficient**	1 Pts **Low**	0 Pts **Insufficient**	4 pts
Communication: ODD protocol	4 Pts **Perfect** The ODD protocol is complete, the given information is correct, and illustrated with supporting diagrams (UML, flowcharts). Based on this ODD it is clear which functionality this model contains	3 Pts **Good** Most (relevant) elements of the ODD protocol are correctly filled out, some supporting diagrams are provided and the ODD provides sufficient information to understand the basic setup of the model.	2 Pts **Sufficient** The ODD protocol has been filled out yet some information is missing and/or incorrect (or in the wrong part of the protocol). The ODD provides insufficient information to reconstruct the model.	1 Pts **Low** The ODD protocol is incomplete, or larger parts of the information are in an incorrect section. It contains partially incorrect information. Diagrams that are essential are missing.	0 Pts **Insufficient** The ODD protocol is not included in the documentation.	4 pts
Communication: Oral presentation	4 Pts **Perfect** Slides provide a good overview of the model, the results etc. The presentation is well structured and presented Group members responds correctly to questions asked by the audience	3 Pts **Good** Slides are of sufficient quality and structure but both content and structure could have been improved. Presentation and answering of questions is sufficient yet, not of very high quality	2 Pts **Sufficient** Slides are incomplete, contains spelling errors, missing pictures, or are not well structured. Presentation or answering of the questions should have been improved.	1 Pts **Low** Slides are poor, multiple categories (layout, structure, presented elements). Both presentation and answering of questions was poor.	0 Pts **Insufficient** No presentation given	4 pts
Communication: Reproducibility	4 Pts **Perfect** All necessary elements were present and of good quality. It was possible to run the model and reproduce the provided output files.	3 Pts **Good** Most necessary elements were present and of reasonable quality. It was possible to run the model although the provided output files could not be immediately reproduced.	2 Pts **Sufficient** All files were present yet, of poorer quality or elements were missing. It was difficult to get the model to run.	1 Pts **Low** Either part of the documentation was missing completely, or it was impossible to run the model.	0 Pts **Insufficient** No reproducibility materials were provided	4 pts
						Total points: 40