

High-resolution vector-borne disease infection risk mapping with Area-to-Point Poisson Kriging and Species Distribution Modeling

Szymon Moliński¹

¹ DataverseLabs, Wrocław, Poland

Correspondence: simon@dataverselabs.com

SM: simon@dataverselabs.com

Abstract

Background: Disease infection data is usually aggregated and shared as a sum of infections in a given area over time. This data is presented as choropleth maps. The aggregation process protects privacy and simplifies decision-making but introduces visual bias for large areas and sparsely populated places. Moreover, aggregated areas of varying sizes cannot be simply used as the input for complex ecological models, which are based on data retrieved at higher resolution on regular grids. The issue is especially painful for vector-borne diseases, e.g. Lyme Disease, where infection risk is closely related to vector species and their ecological niche.

Methods: The paper presents the method of obtaining high-resolution risk maps using a pipeline with two components: (1) spatial disaggregation component, which transforms incidence rate aggregates into the point-support model using Area-to-Point Poisson Kriging, and (2) species distribution modeling component, which detects areas where ticks bite is more likely using MaxEnt model. The first component disaggregates Lyme Disease incidence rates summed over counties in Poland, Central Europe, in 2015. The second component uses ticks occurrence maps, Leaf Area Index, Normalized Difference Vegetation Index, Land Surface Temperature derived from Earth Observation satellites, and Digital Elevation Model.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

The final weighted population-at-risk map is a product of both components outputs. The pipeline is built upon open source and open science projects, and it is reusable.

Results: The presented pipeline creates high-resolution risk maps: vector occurrence probability map, population-at-risk map, and weighted population-at-risk map which includes information about local infections and about vector species. The final maps have much better resolution than aggregated incidence rates. Visual bias for population-at-risk maps is removed, and unpopulated areas are not presented on the map.

Conclusions: The pipeline might be used for other vector-borne diseases. The final weighted population-at-risk map might be used as an input for another analytical model requiring high-resolution data placed over a regular grid. The pipeline removes visual bias and transforms aggregated data into a high-resolution point-support layer.

Keywords: Species distribution modeling, Ticks, Lyme Disease, Ixodes, Poisson Kriging, Area-to-Point Kriging, Population at risk

Background

Lyme Disease incidence rates collected worldwide are aggregated over administrative units. Spatial aggregation protects patients' privacy and simplifies decision-making by governance bodies. However, the aggregation process has serious downsides. Administrative units have varying shapes and sizes (i). Larger areas on choropleth maps tend to get more attention from human viewers, and perceptual homogeneity of incidence rates across the area is misleading - especially for sparsely populated regions [1].

Size and extent of aggregated incidence rates are not comparable to other spatial data sources (ii) used for infection risk modeling. The examples are biotic and abiotic variables available at a higher resolution that correlate with ticks' abundance and questing behavior. Local weather is one of those factors. Temperature and humidity directly influence ticks activity [2, 3]. Climate and climate change variables are other variables worth considering. Milder winters and higher temperatures increase tick distribution worldwide. In Europe, ticks are sampled at higher altitudes and latitudes [4]. The number of ticks increases in the northern United States and Canada [5]. Land cover features are the next group of covariates indirectly informing about ticks' possible range [6, 7, 8]. Finally, there are multiple biotic factors affecting tick abundance and infection risk: vegetation type [9], host availability [10], and host behavior [11].

Using this data, creating the infection risk model with a fine-scale resolution is possible. The assumption is that the deconvoluted risk infection map might be weighted by the probability of the vector's occurrence derived from the species distribution model. The first step in the pipeline is the transformation of areal aggregates of incidence rates into high-resolution blocks and using those blocks as input for the spatial infection risk model (deconvolution). The second part is the estimation of the species distribution map. Tick spatial distribution is derived from the climate, spatial features, and vector occurrence

samples. The final model is an output from the merged deconvoluted infection risk map and species distribution map. It presents a high-resolution infection risk map, incorporating information about the vector abundance in a given area.

The study presents a methodology for retrieving the risk of vector-borne disease infection at a scale finer than the incidence rates aggregated over administrative regions. The system uses geostatistics, machine learning, remote sensing data, and information about hard-bodied ticks species to create a high-resolution Lyme disease infection risk map for Poland. Associated Python code and data are available in a Zenodo archive [12].

Methods

Data Sources

The system uses spatial (GIS), remote sensing, and public health datasets limited to the area of Central Europe (Poland and Germany). Processed datasets and metadata are available in the Zenodo repository.

The datasets used for the population-at-risk estimation are:

- The population density grid in Poland from the Central Statistical Office [13].
The baseline grid size was 1x1 kilometer. Transformed into 5x5 kilometers blocks.
- The incidence rates of Lyme Disease in each Polish county in 2015 (Figure 1).
Data was retrieved manually from each voivodship sanitary-epidemiological station.
- Administrative borders of Polish counties.

The ticks distribution model uses:

- Ticks occurrence locations in Germany [14],
- bioclimatic and environmental factors with spatial resolution 1x1 kilometer:
Land Surface Temperature, Leaf Area Index, Normalized Difference Vegetation Index, Digital Elevation Model (EU-DEM) for areas of Germany and Poland. Data was obtained from the Copernicus Land Monitoring Service [15].

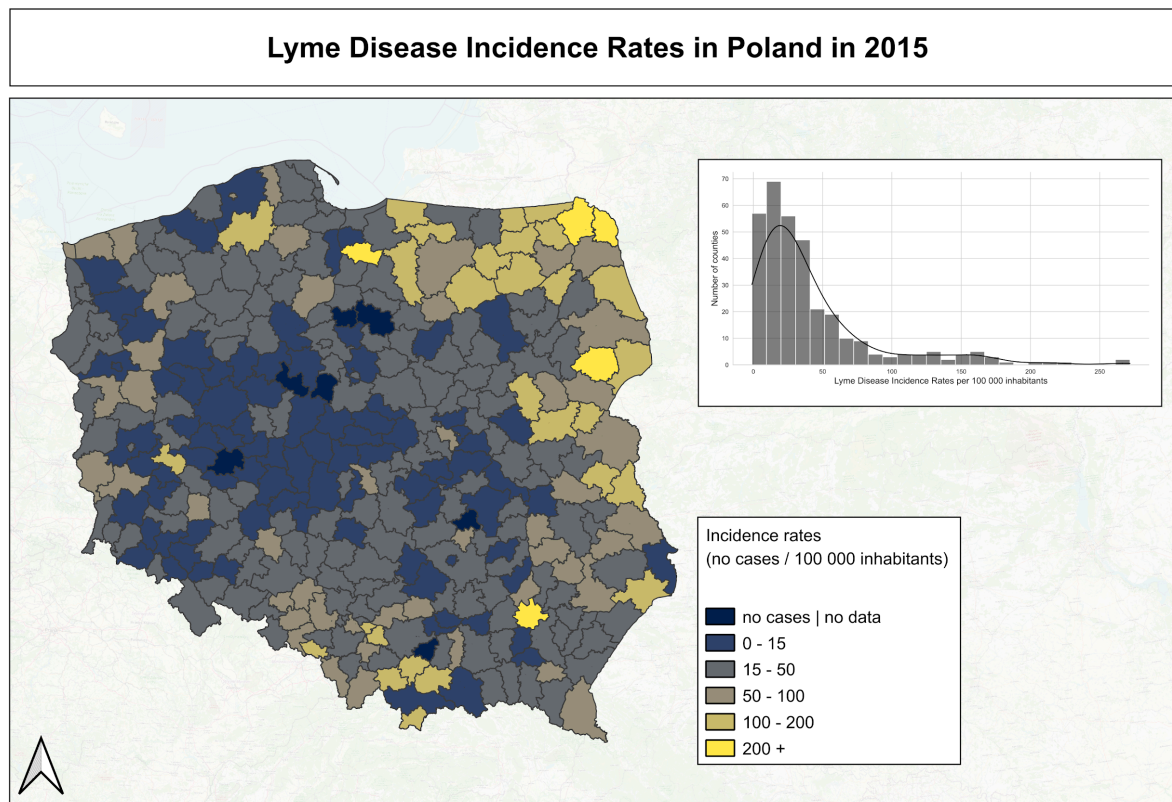


Figure 1: Spatial distribution of Lyme Disease incidence rates in Polish counties.

Poisson Kriging: spatial interpolation of population at risk

The population-at-risk map was developed using the Poisson Kriging technique. Poisson Kriging was used for smoothing and areal deconvolution in other public health studies related to cancer mortality, cholera, and dysentery [1, 16, 17].

The incidence rate is defined as $z(v_\alpha) = 10^5 * d(v_\alpha)/n(v_\alpha)$, where $d(v_\alpha)$ is the number of cases in a fixed time interval and $n(v_\alpha)$ is the total population in a region. The incidence rate is multiplied by 100,000, which indicates the number of cases for 100,000 inhabitants. The number of cases is interpreted as a random variable $D(v_\alpha)$ that follows the Poisson distribution with one parameter - the expected number of cases per year. It can be defined as a product of the local risk $r(v_\alpha)$ and population density on a given area $n(v_\alpha)$. The local risk

$r(v_\alpha)$ might be estimated as a linear combination of the local incidence rate $z(v_\alpha)$ and incidence rates observed along K neighboring areas v_i

$$\hat{r}(v_\alpha) = \sum_{i=1}^K \lambda_i z(v_i), \quad (1)$$

Weights λ_i assigned to K neighbors are obtained after solving the system of linear equations:

$$\sum_{j=1}^K \lambda_j \left[\bar{C}_R(v_i, v_j) + \delta_{ij} \frac{m_i}{n(v_i)} \right] + \mu(v_\alpha) = \bar{C}_R(v_i, v_\alpha), \quad (2)$$

$$i = 1, \dots, K$$

$$\sum_{j=1}^K \lambda_j = 1$$

The technique can be used for data where spatial autocorrelation is significant. The semivariogram analysis and modeling precede kriging. The semivariogram measures the dissimilarity between observations in the function of a distance. When semivariances at short distances are low and rise with increasing distance, the process is spatially correlated. (Compare it with Figure 3). Semivariance is the halved mean squared error between point observations and all other points in a specified range (bin).

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^N [z(x_i + h) - z(x_i)]^2, \quad (3)$$

Semivariance analysis, semivariogram modeling, and spatial interpolation steps were performed in the Python package Pyinterpolate [18].

Vector distribution model

Choropleth map of Lyme Disease incidence rates informs about the medical condition, but it tells nothing about the vector distribution. Thus, the spatial risk may be miscalculated - tick bites occur in different places where infected people live. The full information about the infection risk should include the vector density at a given area or the probability of the vector

occurrence. The occurrence probability is obtained from the species distribution model.

Models of this type are estimated from the information about bio-climatic factors measured in places where species are sampled. The model analyses the correlation between bioclimatic variables and vector occurrences/absences. The popular approach is to use remote sensing data to study vector species [19].

The maximum entropy MaxEnt model is the most popular species distribution model architecture [20]. The risk-modeling pipeline uses an implementation of MaxEnt from the Elapid Python package [21]. The result of the calculations is a tick occurrence probability map (Figure 4). The pipeline model uses:

- points where Lyme Disease (*D. marginatus*, *D. reticulatus*, *I. hexagonus*, *I. ricinus*, *I. trianguliceps*) transmitting ticks were sampled, within Germany border [22]
- Leaf Area Index, Normalized Difference Vegetation Index (mean and standard deviation from years 2005-2015), Land Surface Temperature (mean and standard deviation from years 2017-2020), and Digital Elevation Model EU DEM 1.1 from Copernicus Land Monitoring Services [15].

Species Distribution Model input - training set - comes from Germany because tick observations for this country are available, and Germany has a climate and species similar to Poland.

Complex vector-borne disease infection risk model

Poisson Kriging model derives the population-at-risk of Lyme Disease infection, and the species distribution model generates a probability map of vector occurrence. Both parameters affect local infection risk, and might be related by following reasoning:

$$\frac{d}{n} = p_0 \frac{d}{b_{infected}} * p_1 \frac{b_{infected}}{t_{infected}} * p_2 \frac{t_{infected}}{t_{population}} * p_3 \frac{t_{population}}{n}, \quad (4)$$

where d - number of cases, n - population, $b_{infected}$ - number of bites of infected ticks,

$t_{infected}$ - number of infected ticks, $t_{population}$ - ticks total population. All variables are valid

for a given region and time. Variables p_0, p_1, p_2, p_3 represent unknown weighting factors for

each part of the equation. The first part $p_0 \frac{d}{b_{infected}}$ represents the probability of successful

disease transmission from vector to host. The second part $p_1 \frac{b_{infected}}{t_{infected}}$ is the probability of

being bitten by infected tick, the next part $p_2 \frac{t_{infected}}{t_{population}}$ is the probability of being a disease

vector, and the last part $p_3 \frac{t_{population}}{n}$ is the probability of spotting tick in the local area. Based

on the equation it can be said that the presented approach of merging Poisson Kriging model

with species distribution model estimates the last part of the first and the last of the equation

(Poisson Kriging represents risk of being infected, and species distribution model links host

and vector populations).

Merging both models allows for weighting local population at risk $\hat{r}(v_\alpha)$ with probability of

ticks (vectors) occurrence $P_*(v_\alpha)$. Species-probability weighting lowers local population-at-risk

indices when the probability of occurrence is below 0.5, and the risk increases for probabilities

greater than 0.5. The equation for normalized risk index is:

$$\omega(v_\alpha) = r(v_\alpha) * 2P_*(v_\alpha), \quad (5)$$

The high-level flow with input datasets and model outputs is presented in the diagram (Figure 2).

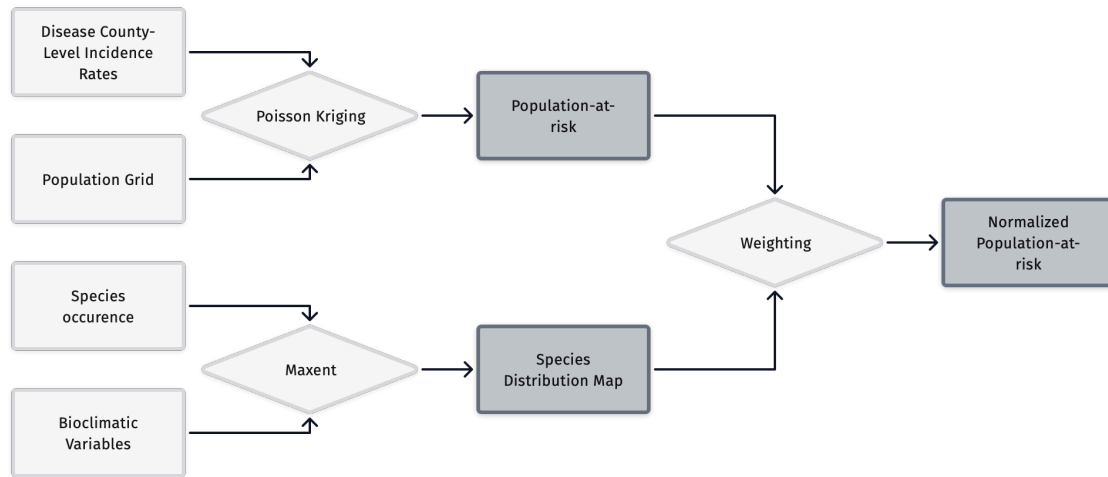


Figure 2: Modeling steps.

Results

The first step is to check if Lyme Disease incidence rates in neighboring areas are similar. It can be done by analyzing the semivariogram, the plot of squared error between neighboring areas as a function of the distance between them (Figure 3). Each county area was transformed to its centroid, and semivariance was measured between those centroids. The plot shows that neighboring counties have similar incidence rates.

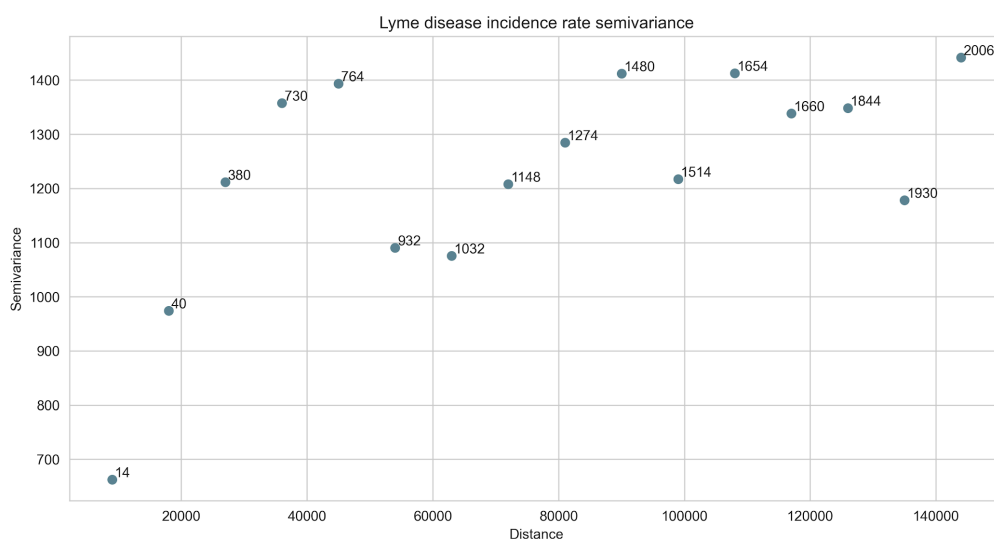


Figure 3: Dissimilarity of point pairs representing county centroids over a rising distance.

Numbers show how many point pairs are averaged to obtain a result.

Semivariances increase up to 60 kilometers, then oscillate around the maximum variance threshold. This means that the spatial correlation between neighboring areas is noticeable up to 60 kilometers, thus the Poisson Kriging model's maximum range is set to this distance. The input for the population-at-risk are Lyme Disease incidence rates in Polish counties (Figure 1) and the Population Density grid (Figure 4).

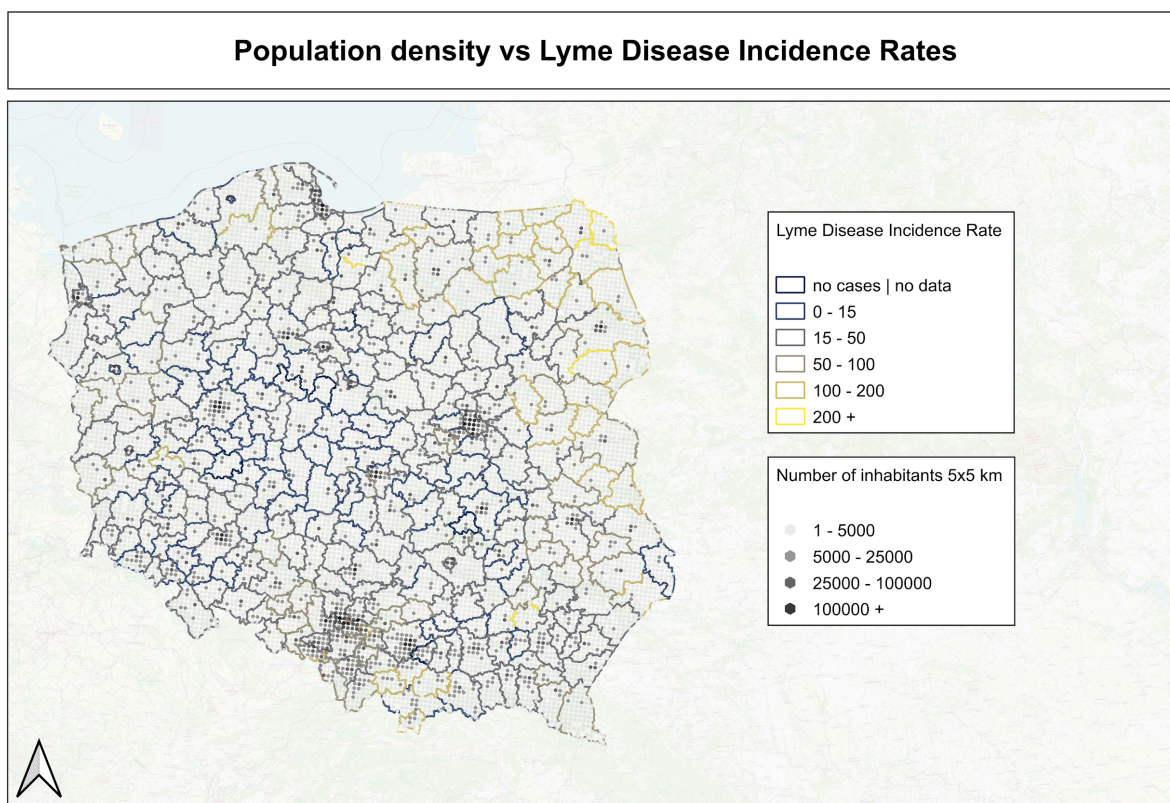


Figure 4: Number of inhabitants in Poland on 5x5 kilometers grid.

Population density creates point support for the deconvolution of areal incidence rate data into small and regularly spaced units. The final population-at-risk map is presented in Figure 5.

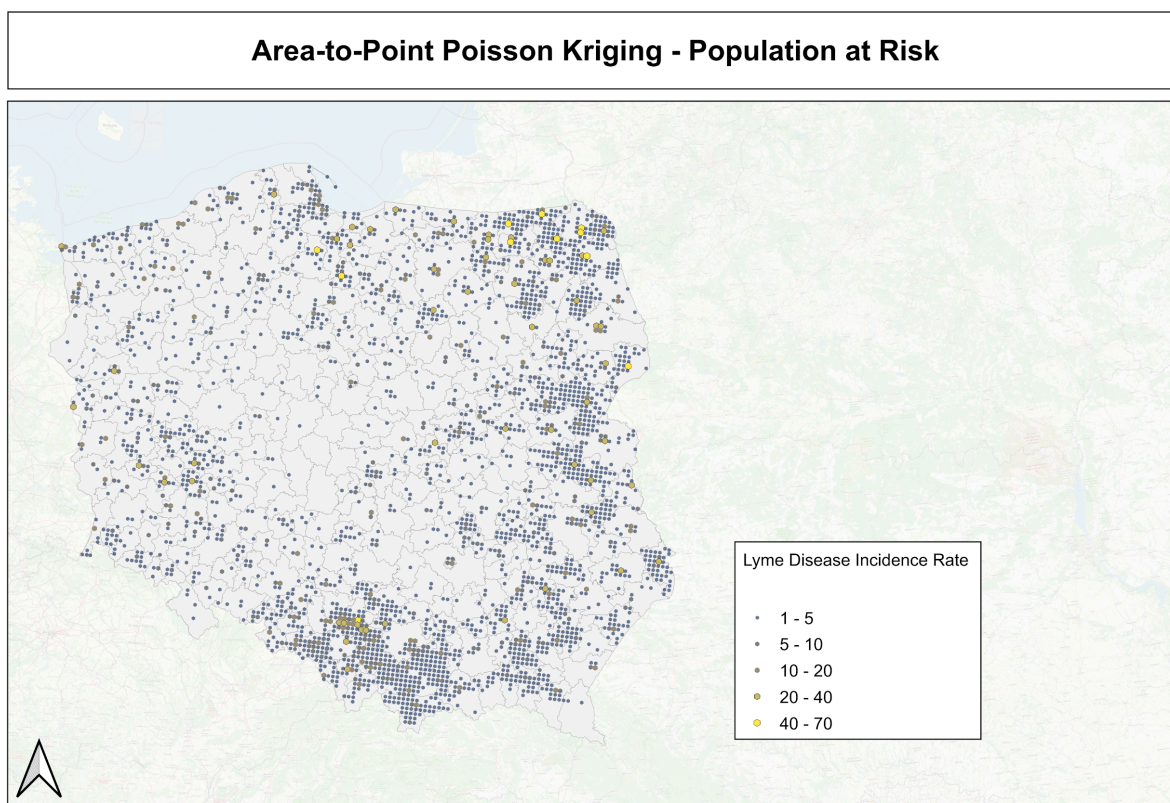


Figure 5: Population-at-risk after Area-to-Point Poisson Kriging transformation.

The next step is creating a species distribution map with tick occurrence probability based on the MaxEnt model and remotely sensed covariates. The map is presented in Figure 6 and shows blocks with tick occurrence probability.

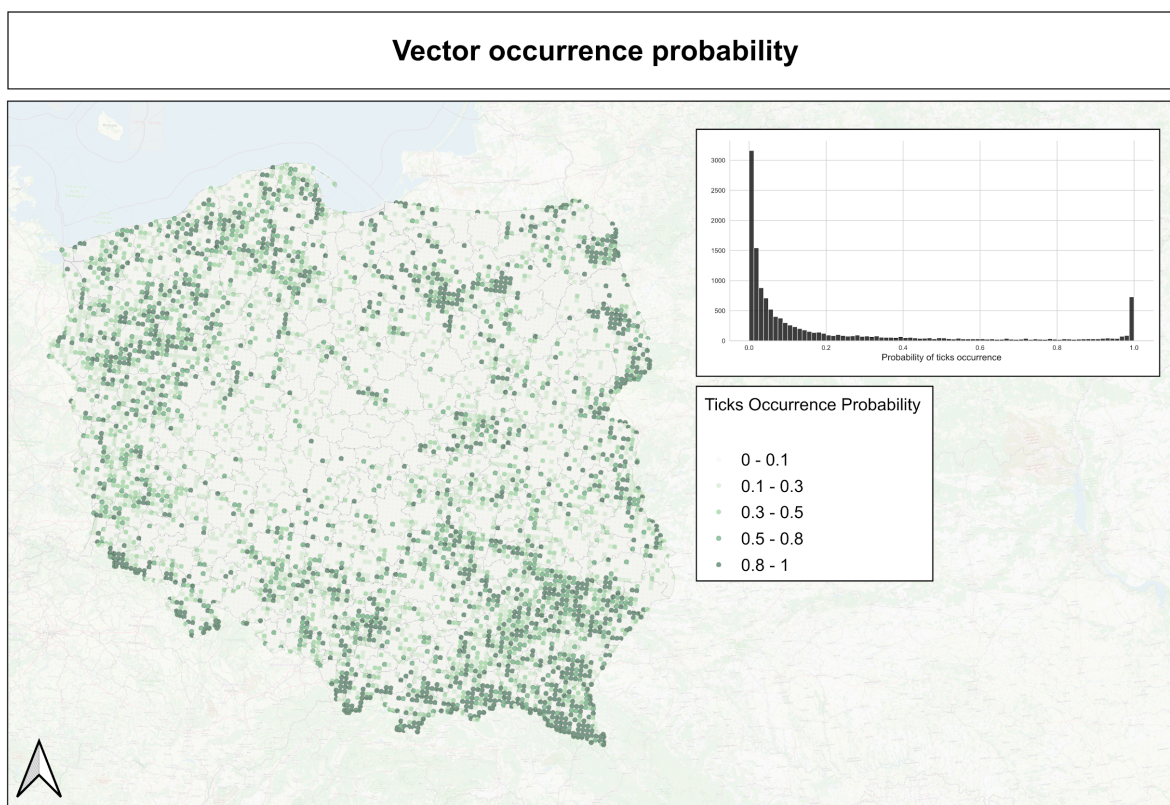


Figure 6: Hard-bodied ticks occurrence probability in Poland.

Both models are weighted (see equation 5), and we obtain the final weighted Lyme Disease infection risk map. Data at this stage is smoothed and ready for further analysis and decision-making (Figure 7).

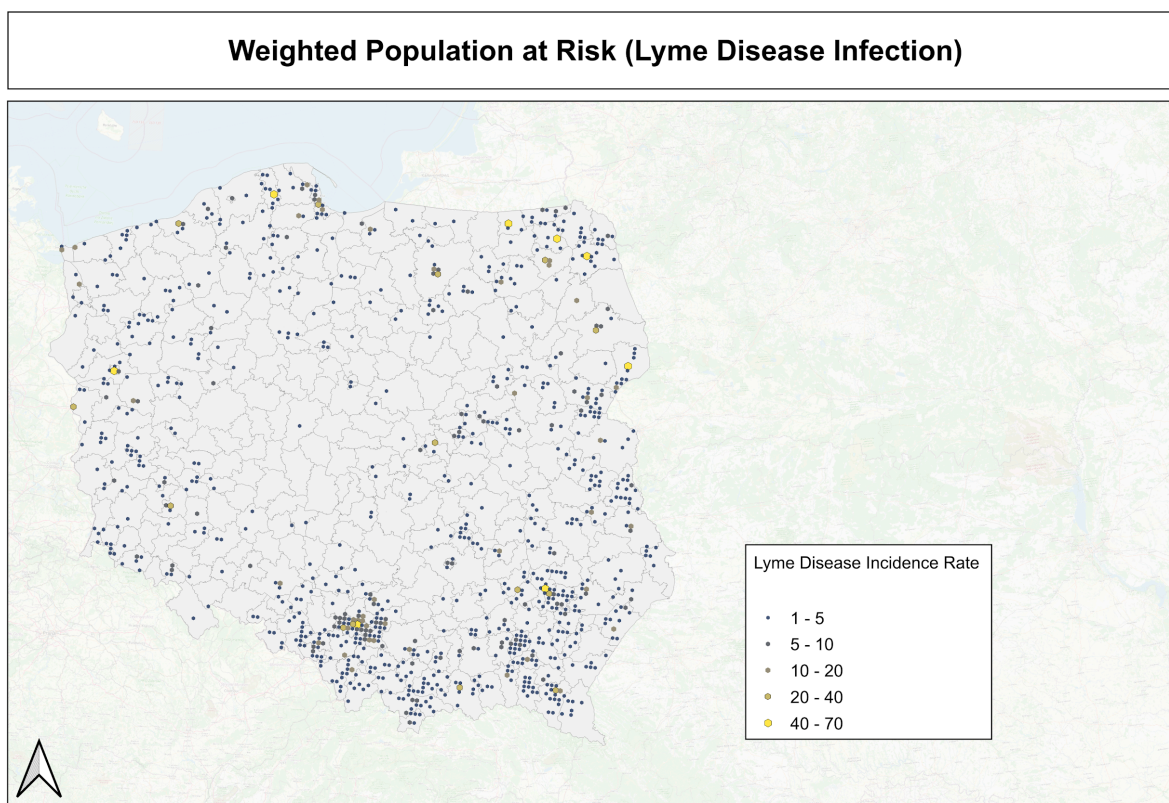


Figure 7: Weighted Population-at-Risk model output.

The pipeline of risk modeling is especially useful for representation of different kinds of risks. There are three outputs: population-at-risk map (1), species distribution model (2), and weighted infection-risk map (3). Every single output might be used for decision-making purposes. Figure 8 shows the approach where a researcher may focus on different aspects of disease monitoring (with filtering results that are below some risk or probability threshold).

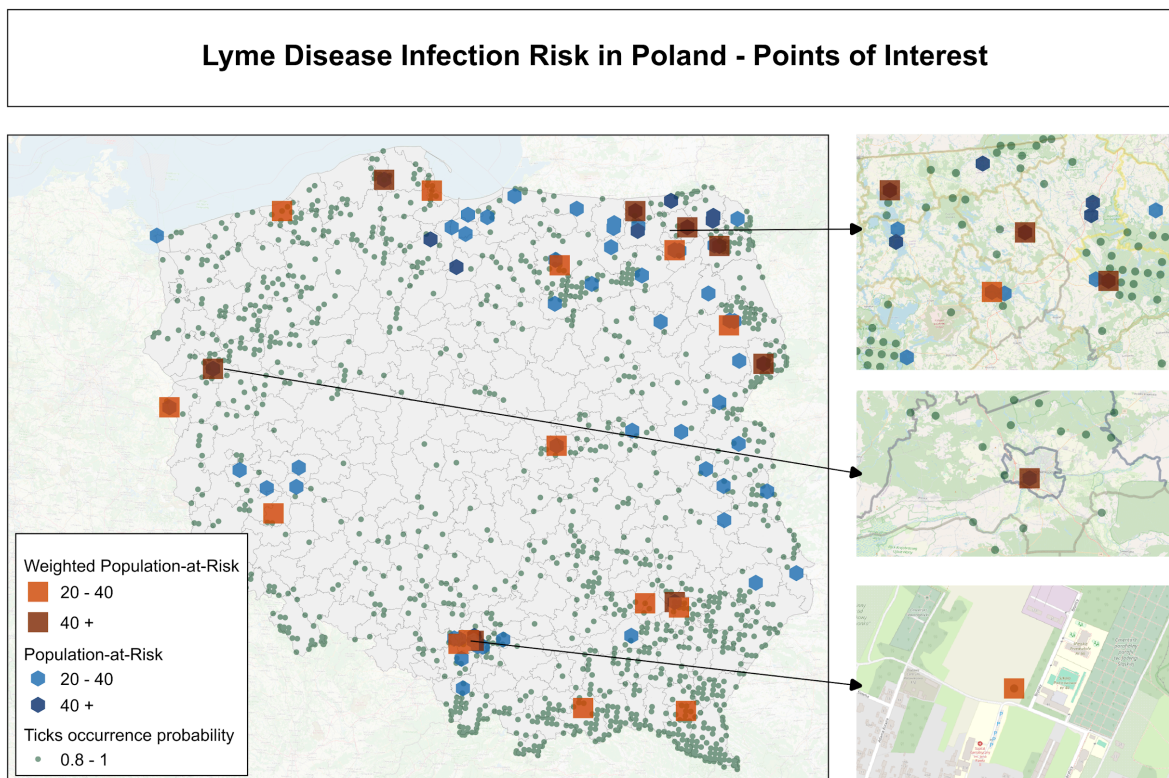


Figure 8: Areas that need special attention. The most dangerous regions are those where the Weighted Population at Risk is highest and are close to the environment supporting ticks' activity.

Discussion

The Presented vector-borne infection risk modeling pipeline is a cost-effective method. It does not require large-scale sampling or sharing of exact location data about patients. Poisson Kriging smooths aggregated incidence rates maps and transforms areal counts into population-unit blocks. The weighting step by species distribution map puts more importance on populations where the probability of vector occurrence is high, making final results more reliable for decision-making. The most important result of the analysis is when the infection risk is high and environmental factors support ticks' existence, those areas and their population require additional monitoring.

The model has downsides. Tick bites may occur in a place different from the population block. People are traveling and hiking long distances from their homes. Thus, some infections might be “imported”, which is the model's source of bias. The model does not estimate risk levels using weather variables that are affecting tick questing behavior, so it is only a spatial representation of the infection risk. It might be expanded into the spatio-temporal model using additional weighting factors related to weather patterns.

Using Poisson Kriging model the researcher might analyze spatial patterns and the distance at which infections are correlated. Kriging property is the generation of variance error map, which shows undersampled areas where additional sampling might be performed.

It is worth noticing that output risk map might be used as an input for complex models using high resolution remotely sensed layers and spatial information for better risk prediction. The modeling pipeline is designed for vector-borne diseases, but Poisson Kriging model alone might work with other diseases and conditions aggregated over areas as a ratio of cases per population.

Summary

The paper presents the analytical pipeline for the interpolation of local vector-borne disease risk. The pipeline uses Area-to-Point Poisson Kriging with aggregated incidence rates and population density support for the development of the population-at-risk map and species distribution modeling for the probability of vector occurrence. The product of both components is a weighted population-at-risk map, which includes information about local infections and vector species.

References

- [1] Goovaerts P. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. *International Journal of Health Geographics* [Internet]. 2006 Jan 1;5(1):52. Available from: <https://doi.org/10.1186/1476-072x-5-52>
- [2] Lees AD. The water balance in *Ixodes ricinus* L. and certain other species of ticks. *Parasitology* [Internet]. 1946 Jan 1;37(1–2):1–20. Available from: <https://doi.org/10.1017/s0031182000013093>
- [3] Kiewra D, Kryza M, Szymanowski M. Influence of selected meteorological variables on the questing activity of *Ixodes ricinus* ticks in Lower Silesia, SW Poland. *Journal of Vector Ecology* [Internet]. 2014 May 12;39(1):138–45. Available from: <https://doi.org/10.1111/j.1948-7134.2014.12080.x>
- [4] Gray J, Dautel H, Estrada-Peña A, Kahl O, Lindgren E. Effects of climate change on ticks and Tick-Borne diseases in Europe. *Interdisciplinary Perspectives on Infectious Diseases* [Internet]. 2009 Jan 1;2009:1–12. Available from: <https://doi.org/10.1155/2009/593232>
- [5] Eisen RJ, Eisen L, Ogden NH, Beard CB. Linkages of Weather and Climate With *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae), *Enzootic Transmission of Borrelia burgdorferi*, and Lyme Disease in North America. *Journal of Medical Entomology* [Internet]. 2015 Dec 17;53(2):250–61. Available from: <https://doi.org/10.1093/jme/tjv199>
- [6] Perez G, Bastian S, Agoulon A, Bouju A, Durand A, Faille F, et al. Effect of landscape features on the relationship between *Ixodes ricinus* ticks and their small mammal hosts. *Parasites & Vectors* [Internet]. 2016 Jan 15;9(1). Available from: <https://doi.org/10.1186/s13071-016-1296-9>

- [7] Ehrmann S, Ruyts SC, Scherer-Lorenzen M, Bauhus J, Brunet J, Cousins S a. O, et al. Habitat properties are key drivers of *Borrelia burgdorferi* (s.l.) prevalence in *Ixodes ricinus* populations of deciduous forest fragments. *Habitat Properties Are Key Drivers of Borrelia Burgdorferi* (s.l.) Prevalence in *Ixodes Ricinus* Populations of Deciduous Forest Fragments [Internet]. 2018 Jan 8;11(1). Available from: <https://doi.org/10.1186/s13071-017-2590-x>
- [8] Kowalec M, Szewczyk T, Welc-Falęciak R, Siński E, Karbowski G, Bajer A. Ticks and the city - are there any differences between city parks and natural forests in terms of tick abundance and prevalence of spirochaetes? *Parasites & Vectors* [Internet]. 2017 Nov 21;10(1). Available from: <https://doi.org/10.1186/s13071-017-2391-2>
- [9] Obsomer V, Wirtgen M, Lindén A, Claerebout E, Heyman P, Heylen D, et al. Spatial disaggregation of tick occurrence and ecology at a local scale as a preliminary step for spatial surveillance of tick-borne diseases: general framework and health implications in Belgium. *Parasites & Vectors* [Internet]. 2013 Jun 22;6(1). Available from: <https://doi.org/10.1186/1756-3305-6-190>
- [10] Perkins SE, Cattadori IM, Tagliapietra V, Rizzoli A, Hudson PJ. Localized deer absence leads to tick amplification. *Ecology* [Internet]. 2006 Aug 1;87(8):1981–6. Available from: [https://doi.org/10.1890/0012-9658\(2006\)87\[1981:ldaltt\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[1981:ldaltt]2.0.co;2)
- [11] Donohoe H, Pennington-Gray L, Omodior O. Lyme disease: Current issues, implications, and recommendations for tourism management. *Tourism Management* [Internet]. 2015 Feb 1;46:408–18. Available from: <https://doi.org/10.1016/j.tourman.2014.07.006>
- [12] Moliński S. High-Resolution Vector-borne Disease Infection Risk Mapping with Area-to-Point Kriging and Species Distribution Modeling - Datasets. Zenodo; 2024.
- [13] Główny Urząd Statystyczny (eng.: Statistics Poland). INSPIRE [Internet]. c2024 [cited 2024 May 03]. Available from: <https://portal.geo.stat.gov.pl/en/home/inspire-en/>

[14] Rubel F, Brugger K, Monazahian M, Habedank B, Dautel H, Leverenz S, et al. The first German map of georeferenced ixodid tick locations. *Parasites & Vectors* [Internet]. 2014 Oct 10;7(1). Available from: <https://doi.org/10.1186/s13071-014-0477-7>

[15] <https://land.copernicus.eu/en>

[16] Goovaerts P, Gebreab SY. How does Poisson kriging compare to the popular BYM model for mapping disease risks? *International Journal of Health Geographics* [Internet]. 2008 Jan 1;7(1):6. Available from: <https://doi.org/10.1186/1476-072x-7-6>

[17] Ali M, Goovaerts P, Nazia N, Haq MZ, Yunus M, Emch M. Application of poisson kriging to the mapping of cholera and dysentery incidence in an endemic area of Bangladesh. *International Journal of Health Geographics* [Internet]. 2006 Jan 1;5(1):45. Available from: <https://doi.org/10.1186/1476-072x-5-45>

[18] Moliński S. Pyinterpolate: Spatial interpolation in Python for point measurements and aggregated datasets. *Journal of Open Source Software* [Internet]. 2022 Feb 23;7(70):2869. Available from: <https://doi.org/10.21105/joss.02869>

[19] Estrada-Peña A, Estrada-Sánchez A, De La Fuente J. A global set of Fourier-transformed remotely sensed covariates for the description of abiotic niche in epidemiological studies of tick vector species. *Parasites & Vectors* [Internet]. 2014 Jan 1;7(1):302. Available from: <https://doi.org/10.1186/1756-3305-7-302>

[20] Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* [Internet]. 2006 Jan 1;190(3–4):231–59. Available from: <https://doi.org/10.1016/j.ecolmodel.2005.03.026>

[21] Anderson CB. elapid: Species distribution modeling tools for Python. *Journal of Open Source Software* [Internet]. 2023 Apr 19;8(84):4930. Available from: <https://doi.org/10.21105/joss.04930>

[22] Rubel F, Brugger K, Monazahian M, Habedank B, Dautel H, Leverenz S, et al. The first German map of georeferenced ixodid tick locations. *Parasites & Vectors* [Internet]. 2014 Oct 10;7(1). Available from: <https://doi.org/10.1186/s13071-014-0477-7>