

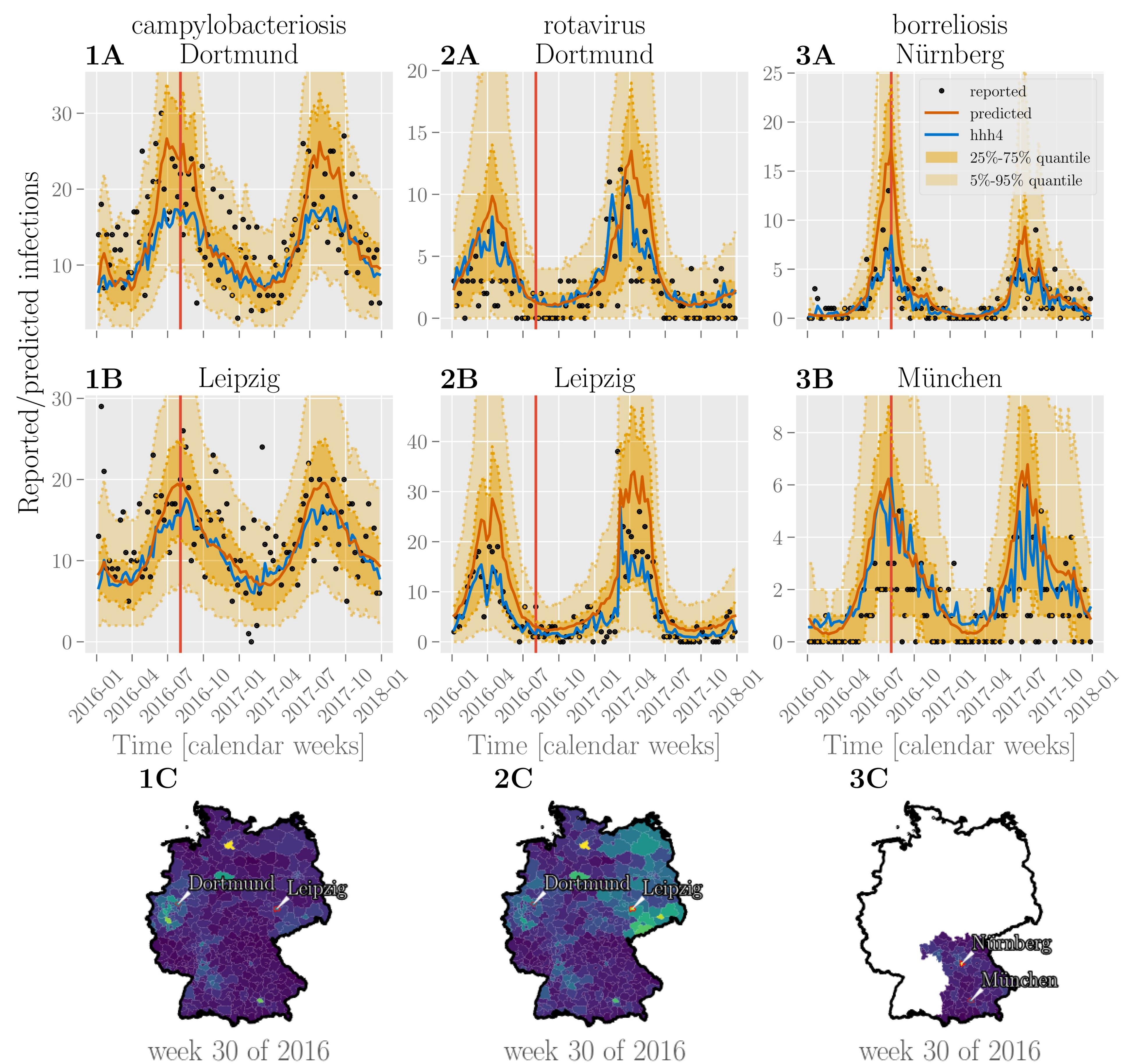
Visualizing the spread of infectious diseases using public health data

Public-health agencies have the responsibility to detect, prevent and control infections in the population. The Robert Koch Institute [1] in Germany collects a wide range of factors, such as location, age, gender, pathogen, and further specifics, of laboratory confirmed cases for approximately 80 infectious diseases through a mandatory surveillance system [2]. This data is publicly available, but in order to be useful for a broader public, it should be processed and presented in an interpretable form, using data visualizations and interactive tools.

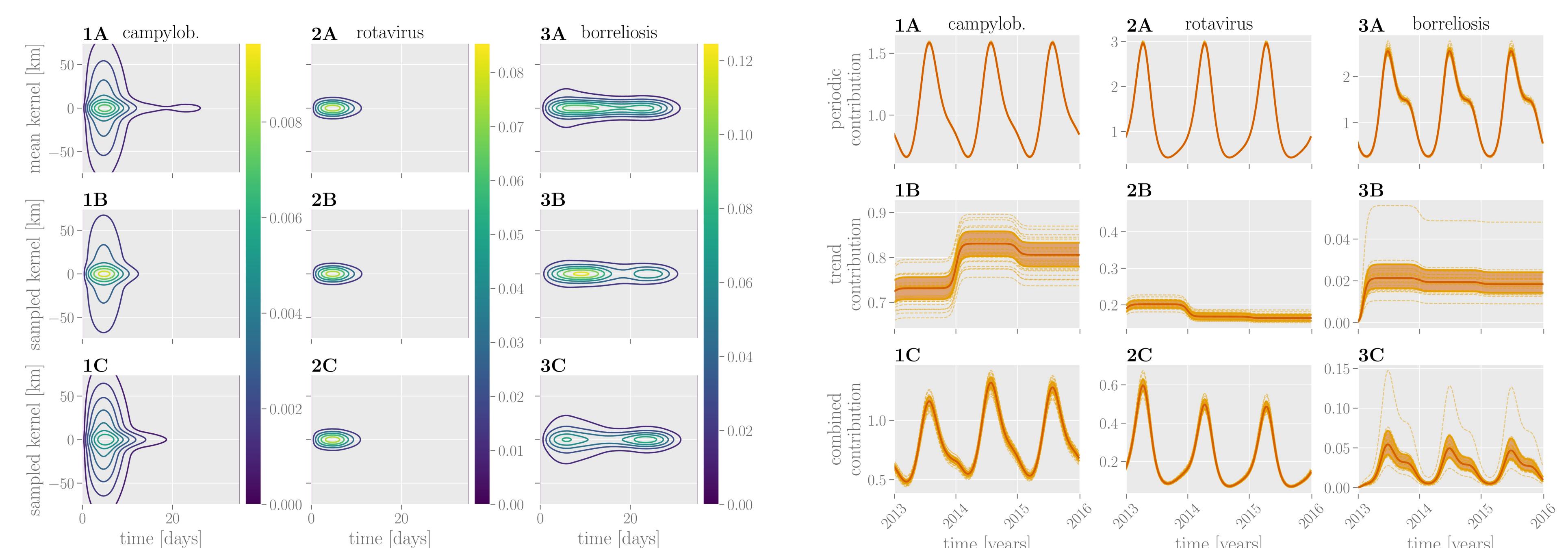
We develop a single model for predictions of infectious disease cases in all counties in Germany. To present the results of the model and to communicate potential risks of infection, **disease prediction maps** show a broad overview of the disease development in space, while plotting **prediction curves** together with collected data points show the disease development in time.

Bayesian spatio-temporal interaction model

The presented Bayesian spatio-temporal interaction model (BSTIM) [3] is a probabilistic model which predicts aggregated case counts within counties and calendar weeks. To this end, **publicly available health data**, region-specific and demographic data are used. We evaluated the BSTIM on a **one-week-ahead prediction** task for two diseases (campylobacteriosis [4] and rotaviral enteritis [5]) across Germany and for Lyme borreliosis [6] across the federal state of Bavaria. The BSTIM model predicts how many people are expected to become infected during the next week, in each county (shown in the upper figure). In addition, it provides **uncertainty estimates**, which give a sense of how confident the model is. Domain experts, in addition to model predictions, can access the dynamics and evolution of diseases. Visualizing learned model components (shown in the lower left figure) allows to inspect how diseases spread in time and space, while visualizing learned trends and seasonality (shown in the lower right figure) allows to see temporal evolution of diseases over the years. This **transparency and interpretability of machine learning models** increase scientific understanding and safety [7].



Predictions of case counts for various diseases by county. Reported infections (black dots), predictions of case counts by BSTIM (orange line) and the hh4 reference model (blue line) for campylobacteriosis (column 1), rotavirus (column 2) and borreliosis (column 3) for two counties in Germany (for campylobacteriosis and rotavirus) or Bavaria (borreliosis), are shown in rows A and B. The shaded areas show the inner 25%-75% and 5%-95% percentile. Row C shows predictions of the respective disease for each county in Germany or the federal state of Bavaria in week 30 of 2016 (indicated by a vertical red line in rows A and B).



Key points:

- We develop and present a probabilistic model for prediction of infectious disease cases
- Prediction maps or risk awareness maps are beneficial for communicating a message to the public
- Visualizations of components of a prediction model are useful for domain experts for assessing the dynamics of diseases
- The machine learning system for prediction of infectious diseases should at the same time optimize for two aspects:
 - the prediction accuracy,
 - interpretability

References:

- [1] https://www.rki.de/DE/Home/homepage_node.html
- [2] Faensen D, Claus H, Benzler J, Ammon A, Pfösch T, Breuer T, et al (2006) **SurvNet@RKI – a multistate electronic reporting system for communicable diseases.** Euro surveillance: bulletin européen sur les maladies transmissibles=European communicable disease bulletin. 11(4):100–103.
- [3] Stojanovic O, Leugering J, Pipa G, Ghozzi S, Ullrich A. (2019). **A Bayesian Monte Carlo approach for predicting the spread of infectious diseases.** *Biorxiv*; under review (PLOS ONE).
- [4] <https://www.who.int/news-room/fact-sheets/detail/campylobacter>
- [5] <https://www.who.int/immunization/diseases/rotavirus/en/>
- [6] <https://www.who.int/ith/diseases/lyme/en/>
- [7] Molnar C. (2019) **Interpretable machine learning. A Guide for Making Black Box Models Explainable** (Chapter 2).

