Quantification, Prediction and Forecasting of Vibrio vulnificus in the Southern Baltic Sea

Using High-Resolution Data

Vibrio vulnificus is an opportunistic marine pathogen of growing public health concern in northern Europe. Climate-induced warming of coastal and estuarine waters, especially in the brackish Baltic Sea, has increased the frequency and geographic range of *Vibrio vulnificus* detections, with infections reported in countries surrounding the southern Baltic following periods of extreme summer heat. Given its potential to cause severe wound infections and septicemia, quantifying the abundance of *Vibrio vulnificus*, identifying its key environmental predictors, and developing reliable forecasting tools are critical steps toward managing its growing ecological and public health impact.

This thesis addresses these gaps by developing a high-resolution, multi-parameter dataset to explore the determinants of *Vibrio vulnificus* in the southern Baltic Sea. The central aim is to quantify the conditions that govern *Vibrio vulnificus* abundance and evaluate the effectiveness of machine learning approaches for predicting and forecasting its occurrence. Extensive field sampling was conducted over one year across 15 sites spanning a freshwater–brackish gradient. Environmental measurements included a wide array of biological, chemical, remote sensing, and meteorological parameters. To evaluate overall ecosystem health, we assessed anthropogenic pollution across the study area, revealing consistently high anthropogenic pressure throughout the year for aquatic organisms.

To quantify *Vibrio vulnificus*, a multi-method approach was employed: species-specific cultivation on thiosulfate-citrate-bile salts-sucrose agar and CHROMagar™ *Vibrio*, droplet digital polymerase chain reaction targeting the *Vibrio vulnificus* hemolysin A gene, and inference from sequencing data scaled to absolute abundance. These combined methodologies enabled robust cross-validation of *Vibrio vulnificus* occurrence.

After evaluating multiple machine learning approaches on complex environmental metabarcoding datasets, Random Forest was selected for its strong performance and robustness to high-dimensional, compositional data and applied to this dataset. The analysis found that elevated temperature and intermediate salinity remain important, with *Vibrio vulnificus* abundance strongly clustered in conditions with salinities between 12 and 18 and temperatures exceeding 18 °C. Importantly, the results indicate that changes in microbial communities, especially within prokaryotic composition, significantly influence the ecological patterns of *Vibrio vulnificus*.

Finally, advanced time-series modeling approaches were applied, using time-lagged Random Forests and Long Short-Term Memory neural networks, which leveraged historical dependencies within the dataset and enabled accurate forecasting of *Vibrio vulnificus* presence based on data from the previous 4–5 weeks. Results showed that *Vibrio vulnificus* forecasting was possible using prokaryotic, eukaryotic, or satellite datasets. This enabled development of a reliable early warning tool for detecting *Vibrio vulnificus* risk at bathing sites along the Baltic Sea coast.

Together, these findings demonstrate that high-resolution environmental and microbial monitoring can enhance our understanding of *Vibrio vulnificus* ecology. By integrating robust quantification methods with machine learning and time-series forecasting, this thesis presents a comprehensive framework for early detection and *Vibrio vulnificus* forecasting. The approach contributes valuable tools for public health preparedness amid climate-driven changes in coastal pathogen dynamics.