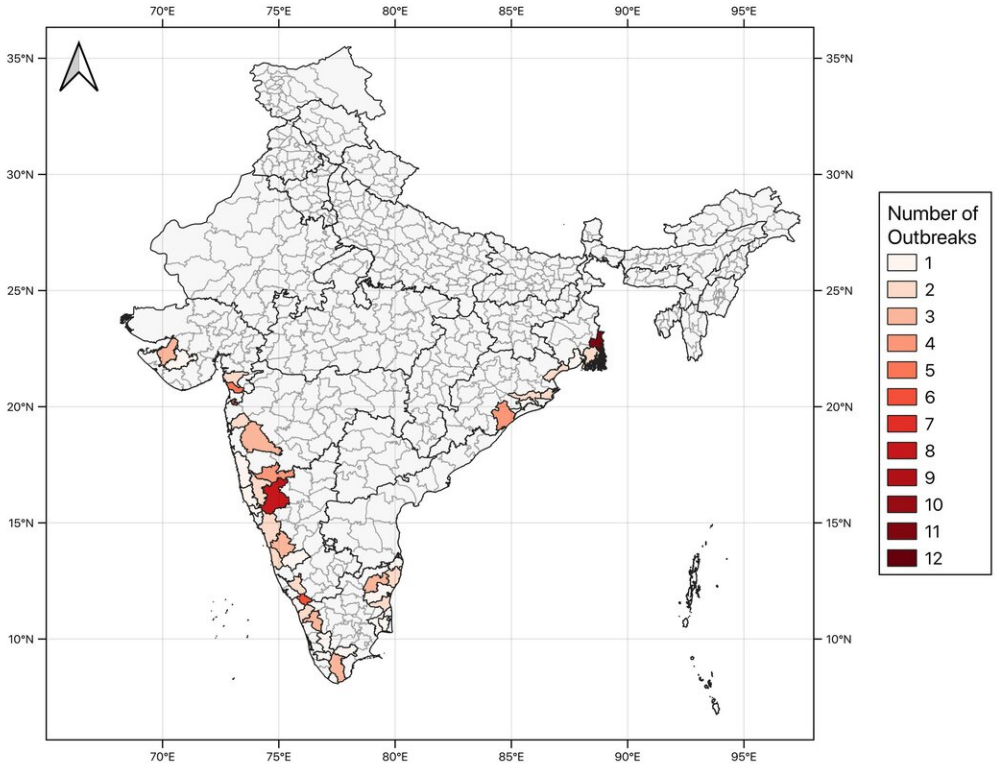
Cholera outbreaks predicted using climate data and AI

December 17 2020



Number of cholera outbreaks reported in the weekly epidemiological reports published by the Integrated Disease Surveillance Programme of India (IDSP) during the period January 2010 to December 2018 for the 40 coastal districts of India selected in the study. Only the districts reporting cholera incidence data for which all seven Essential Climate Variable (ECV) datasets were available are shown. Credit: Campbell et al., 2020

Climate data taken from Earth orbiting satellites, combined with machine learning techniques, are helping to better predict outbreaks of cholera and potentially save lives.

Cholera is a waterborne disease caused by the ingestion of water or food contaminated with the bacterium Vibrio cholerae, which can be found in many coastal regions around the world, especially in densely populated tropical areas. The responsible pathogen generally lives under warm temperatures, moderate salinity and turbidity, and can be harbored by plankton and detritus in the water.

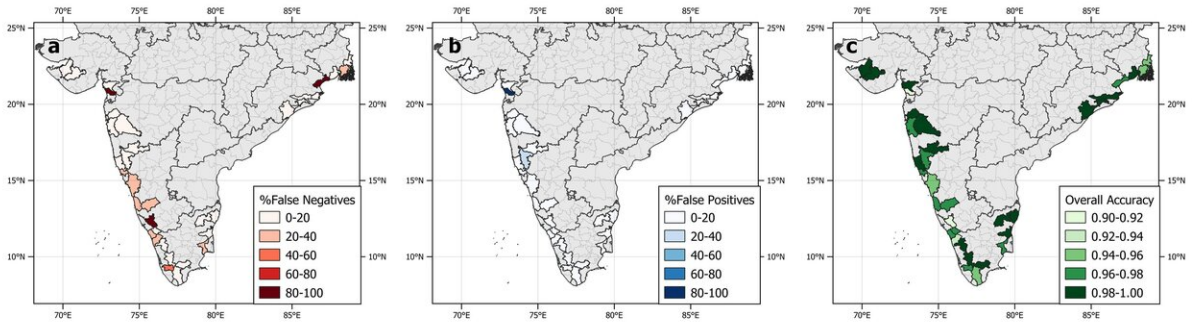
Global warming and an increase in extreme weather events are driving outbreaks of cholera—a disease that affects 1.3 to 4 million people each year worldwide and causes up to 143 000 fatalities. A new study shows how cholera outbreaks in coastal regions of India can be predicted with an 89% success rate, in the first demonstration of using sea surface salinity for forecasting cholera.

The research published yesterday in the *International Journal of Environmental Research and Public Health* focuses on predicting outbreaks of cholera around the northern Indian Ocean, where more than half of global cases of the disease were reported in the 2010-16 period.

The relationship between the environmental drivers of cholera incidence are complex, and vary seasonally, with different lagged effects, for example from the monsoon season. Machine learning algorithms can help to overcome these issues by learning to recognize patterns across large datasets in order to make testable predictions.

The study was led by Amy Campbell during a year-long graduate traineeship with the ESA Climate Office. Amy, along with her co-authors at the Plymouth Marine Laboratory (PML), used a machine learning algorithm popular in environmental science applications—the

random forest classifier—which can recognize patterns across long datasets and make testable predictions.



Performance metrics‘ results of the Random Forest Model when applied to unseen test data for individual districts in coastal India that reported cholera outbreaks. Coastal districts with no cholera outbreaks reported in the study period and non-coastal districts are shown in grey colour. Credit: Campbell et al., 2020

The algorithm was trained on disease outbreaks reported in coastal districts in India between 2010 and 2018, and learned the relationships with six satellite-based climate records generated by ESA's Climate Change Initiative (CCI).

By including or removing environmental variables and sub-setting for different seasons, the algorithm identified key variables for predicting cholera outbreaks as land surface temperature, sea surface salinity, chlorophyll-a concentration and sea level difference from average (sea level anomaly).

Amy Campbell said, "The model showed promising results, and there's a lot of scope for developing this work using different cholera surveillance

datasets or in different locations. In our study, we tested different machine learning techniques and found the random forest classifier to be the best, but there are far more techniques that could be investigated.

"It would be interesting to test the impact of including socio-economic datasets; remote sensing data could be used to develop records to account for human factors that are important for cholera incidence, such as access to water resources."

The study and its new insights have contributed to the UKRI-NERC Pathways Of Dispersal for Cholera And Solution Tools (PODCAST) Project led by co-author Marie-Fanny Racault at PML, which is assessing the impact of climate warming and climate extremes on habitats suitable for Vibrio cholerae.

The results from the study will be demonstrated at the UNFCCC's COP26 meeting in 2021 via a web-based forecasting tool as part of the PODCAST-DEMO project.

More information: Amy Marie Campbell et al. Cholera Risk: A Machine Learning Approach Applied to Essential Climate Variables, *International Journal of Environmental Research and Public Health* (2020). DOI: 10.3390/ijerph17249378

Provided by European Space Agency