Predictive modeling as a tool for global foot-and-mouth disease surveillance

Garabed, R.B.[1]; Perez, A.M.[1, 2]; Johnson, W.O.[3]; Thurmond, M.C.[1]
[1] FMD Modeling and Surveillance Laboratory, University of California, Davis; [2] CONICET/INTA Balcarce, Argentina; [3] Department of Statistics University of California, Irvine.

Abstract

Determining the presence of foot-and-mouth disease (FMD) on a global scale is difficult due to the limitations of reporting in remote or underdeveloped areas. Voluntary reports from those areas typically provide annual information at the country level, but specific temporal and spatial data exist only in certain areas. The purpose of this study is to create a high-resolution model to predict FMD presence worldwide. FMD data and hypothesized predictors for 187 countries from 1996 to 2003 were used to construct a Bayesian hierarchical logistic regression model to predict FMD presence at a resolution of 2500 km^2. FMD reports with a town or city of the outbreak recorded, which was considered to be a resolution of 2500 km^2 or better, were considered cases (n = 2272). Controls were all areas specified as "free" or "free with vaccination" by the World Organization for Animal Health (OIE). Hypothesized risk factors for FMD in a given 2500 km^2 cell were monthly precipitation, temperature, and adjacency to FMD-positive cells; fixed variables were ruminant, monogastric, and human population density. Annual country-specific variables were gross domestic product (GDP) per *capita*, political voice, adjacency to FMD-positive countries, and historical FMD presence. The model was validated with data from 2004 and was used to predict FMD in areas with unknown FMD information. Preliminary data maps match the existing low-granularity OIE maps, but show finer distinctions in FMD risk within countries. With further validation, these maps could be used to help concentrate surveillance and control efforts in local high risk areas.

Introduction

The global control and eradication of FMD is complicated by its political, social, and economic consequences, as well as by its biology. Experts continually debate strategies for outbreak control, methods of diagnosis and plans for trade restriction, but most control strategies require distribution of international funds, testing equipment, personnel, vaccine, and carcass disposal and disinfection supplies (or some subset of these) to the disease locations. While individual countries and areas of interest within countries may have efficient methods of disease reporting, in some areas, especially remote and underdeveloped areas, recognizing and reporting of an FMD case is challenging. In addition, there is a lag between the time a case is recognized and the time that supplies and resources can be deployed to control its spread. By using available data on FMD occurrence to predict the likelihood of FMD cases in time and space for the whole world, future surveillance and control efforts can be focused on areas predisposed to FMD, independent of the area's ability to report FMD.

Objective

The objective is to predict the probability of FMD occurrence for the entire world from 1997-2004, using all available data on FMD occurrence and hypothetical predictors of FMD.

Methods

General Approach

Data on 2272 occurrences of FMD between 1997 and 2004 and on 17,206 to 18,660 areas free of FMD (depending on the month) were used to fit a Bayesian hierarchical logistic regression model, using semi-informative prior information and data on hypothetical predictor variables at a spatial resolution of 2500 km^2 and a monthly temporal resolution.

Data

The response data used to fit the model were all reported cases of FMD based on reports to OIE from 1996-2004, searches of online popular sources, and some data-sharing agreements with specific countries. Because the definition of “case” and “outbreak” varied for different sources of data, the definition used here is the report of at least one FMD-positive animal (clinical case, virus identification or FMD-positive NSP serology test) during the specified month. Only cases with a reported town, city, or geographic location (latitude and longitude) were included in the study. Cases were assigned to 2500 km^2 cells by finding the geographic coordinates of the town or city and matching them to a reference grid in ArcGIS™ (ESRI®, Redlands, CA). Similarly, all areas specified by OIE as “free of FMD without vaccination” or “free of FMD with vaccination” were assigned to grid cells for the time that they were considered free.

Predictor data were collected from a variety of sources and assigned to the same reference grid. The predictive variables can be placed into three categories based on the type of information they primarily represented – animal density, health care, and movement and exposure to FMD. Animal density is associated with human population (Landscan Global Population 2004 Database, Oak Ridge National Laboratory, 0.008° resolution), density of ruminants, and density of monogastrics (Environmental Research Group Oxford, 0.05° resolution). The quality of animal health care is represented by country-level predictors of people’s voice (an indicator of overall governance quality; Governance and Anti-Corruption Indicators, World Bank Institute) and gross domestic product *per capita* (CIA World Factbook). Animal movement and exposure to FMD are represented by prior official FMD status (positive, free, or unknown, OIE), proportion of country’s land borders with countries not free of FMD (OIE and CIA Factbook), an inverse distance weighting of all case cells, and indicators for high precipitation (top 5^{th} percentile for monthly average precipitable water, NOAA-ESRL, 2.5° resolution), high or low temperatures (over 312°K or under 273°K, NOAA-ESRL, 2.5° resolution), and the month in the Gregorian calendar of the Islamic feast of sacrifice in countries with primarily Muslim populations (CIA Factbook). The feast of sacrifice is celebrated at the 10^{th} day of the 12^{th} month of the Islamic calendar.

Model

A Bayesian hierarchical logistic regression model was fit using WinBUGS© (Imperial College & MRC, UK). The model included a random effect for country (centered on the country-level variables) to account for correlation in repeated observations from the same country and to account for unknown country-specific predictors. The expert opinions of our research group and collaborators based on over 20 years of work with these variables and on FMD in general also were incorporated in the model as informative prior distributions. These prior distributions were made for the probability of FMD occurring in a cell with a specific combination of covariates. Because these estimates were made for the probability of FMD independent of the ability of authorities to find and report a case with the given covariates, these opinions helped to weight the model towards prediction of all cases rather than just those in the data (reported cases). Each year was fit sequentially with the next year’s fit being influenced by all previous estimates. This output was then manipulated using R© (The R Foundation for Statistical Computing) to produce predictions for the whole world that were then processed for display in an ArcGIS™ interface.

Results

The model was fit and predictions were made and displayed for 1997-2004. Initially the 2004 data were excluded and used to validate the model by making predictions for this year based on the previous year’s covariate values. A preliminary example of the predictions excluding the country-level random effects is shown in Figure 1. These are the predicted probability of at least one FMD-positive animal in each 2500 km^2 cell based on the predictor variables alone. The countries shown in grey were excluded here due to a lack of predictor data (Greenland, Antarctica, and Western Sahara) or a problem in the data assembly (Bosnia and Herzegovina, Uzbekistan, French Guiana, North Korea, and Taiwan). Figure 2A shows the specificity and sensitivity of this model (without random effects), using the cutoff of 0.50 as the divider for positive *versus* negative cells for each year. Figure 2B shows the specificity and sensitivity of the same model, using a cutoff of 0.01. The poor sensitivity in 2001 is because of the cases that occurred in the UK, France, Ireland, and the Netherlands, which were not predicted by the model. Aside from this dip in sensitivity, the model generally improves over time, as would be expected.

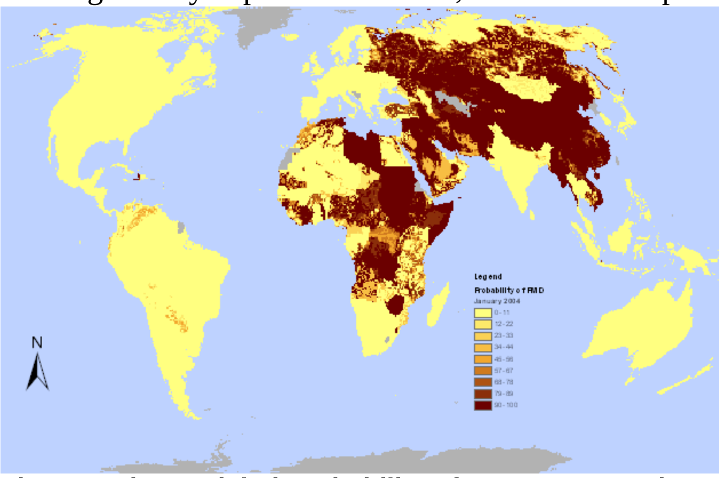
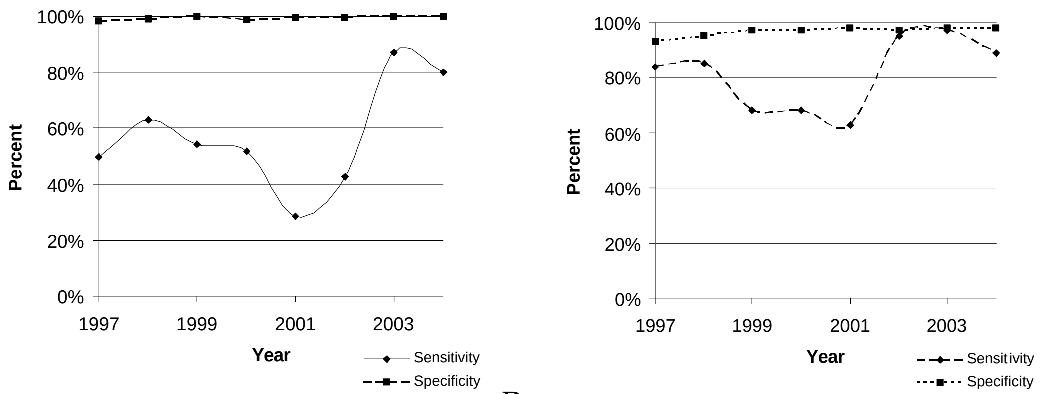


Figure 1 The modeled probability of FMD presence in January 2004 (multiplied by 100)



A. B.

Figure 2 The specificity and sensitivity of the model to detecting actual cases and controls based on a cutoff of 0.50 (A) versus 0.01 (B) as the probability specifying a case of FMD

Discussion

Overall, this model performs well in that it reaches 97% sensitivity and 98% specificity in 2003, when a cutoff value of only 0.01 is used. The addition of the random country-level effects accounts for a large number of incorrect predictions, thus, improving the model further. Because, this model uses priors that are more informed by the data each successive year, it is expected that the model will continue improving as more years of data are added to the historical basis. However, each addition of data overpowers the initial expert-based prior information. Therefore, as the model uses more data of reported outbreaks, it is less likely to model unreported outbreaks and uncommon outbreaks. Still, the model provides good estimates for the probability of finding FMD in 2500 km^2 cells with widely available predictor information. This model could be used to assess effects of different predictors on FMD presence as well. Such analyses could prove useful in helping determine courses of action in controlling future outbreaks. This method also provides an alternative tool to compare to and use in conjunction with simulation models.