1+1=3
(finding the historical trend of an epidemic)

Gustaf Rydevik[1, 2, 3*], Mike Hutchings[2], Giles Innocent[1], Glenn Marion[1], Piran White[3]
1. Biomathematics and Statistics Scotland
2. Scottish Agricultural College
3. University of York
*Contact author: gustaf.rydevik@sac.ac.uk

Lawrence Wright, April 17th 2011, Creative Commons License

Rabbits may carry *Mycobacterium Paratuberculosis,* which can infect both cattle and sheep, manifesting as Johne's disease, thus causing animal suffering and large monetary losses . *M. Paratuberculosis* is one of 20 pathogens for which new tests are being developed within the WildTech project .
Conducting longitudinal studies on rabbits is difficult, and so any methods which can help understanding the epidemiology of their pathogens are much needed.

This poster describes ongoing work to produce methods for an R package that can draw inference of historlcal trends of epidemics of infectious diseases, based on cross-sectional, multiple test data.

A problem in wildlife disease surveillance

Consider a situation where one needs to find the historical distribution of the incidence of an infectious pathogen such as Salmonella, but is limited by monetary, practical or other constraints to only conducting a single study at one point in time.
Depending on the type of covariates available from that study, the difficulty of finding out what has happened before the study started can range from trivial to impossible. In this poster, we look at the scenario of using two different tests for the pathogen, each test having different response characteristics depending on time since event. By combining the information from these two tests, we get a better idea on when in the past the infection occurred, getting an indication of the general historical trend of the epidemic.
This work is motivated by the current development of new tests for infectious wildlife diseases within the FP7 WildTech project. These tests will use both the presence of antibodies as well as the presence of genetic material (nucleic acid/NA) to test for the presence of pathogen.

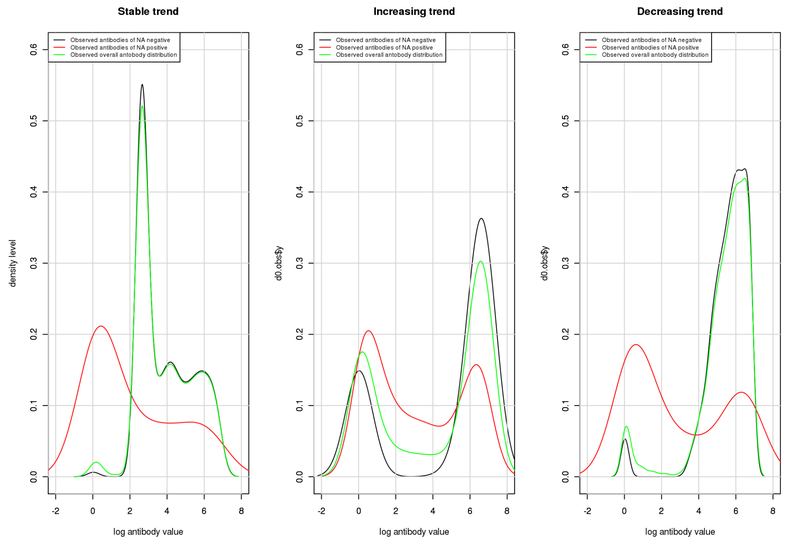
Generating input data

Since real multiple test data is not yet available, we simulated data using a three step process:

1. From a population of 10 000 individuals, select, with replacement, a number of individuals proportional to an incidence parameter (set to 1/10 per year) times an observation period parameter (set to ten years). Mark these as having been infected *n* times, equal to the number of times selected.
2. Among those infected, randomly assign an infection time according to some distribution of infection intensity over time. For those that have been selected several times, assign multiple infection times, and choose the time closest to the end of the observation period.
3. Generate expected antibody and pathogen growth levels at the end of the observation period according to a selected (deterministic or random) function for test levels after infection. We used a set of differential equations known as Lotka-Volterra equations, commonly used for Predator-Prey models, where here the pathogen was considered Prey and antibodies predators. The equations were solved using the *deSolve* package.

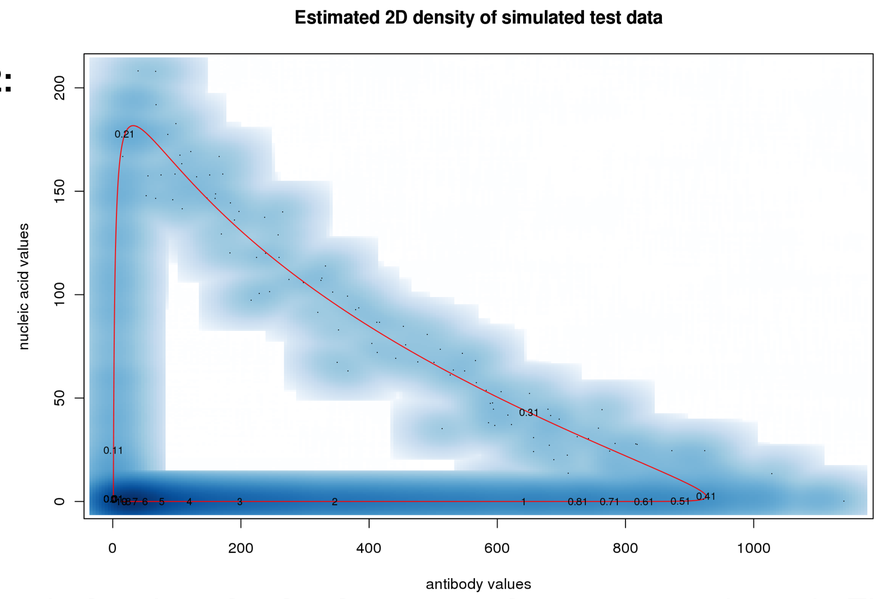
Investigating the historical trend

Fig.1:



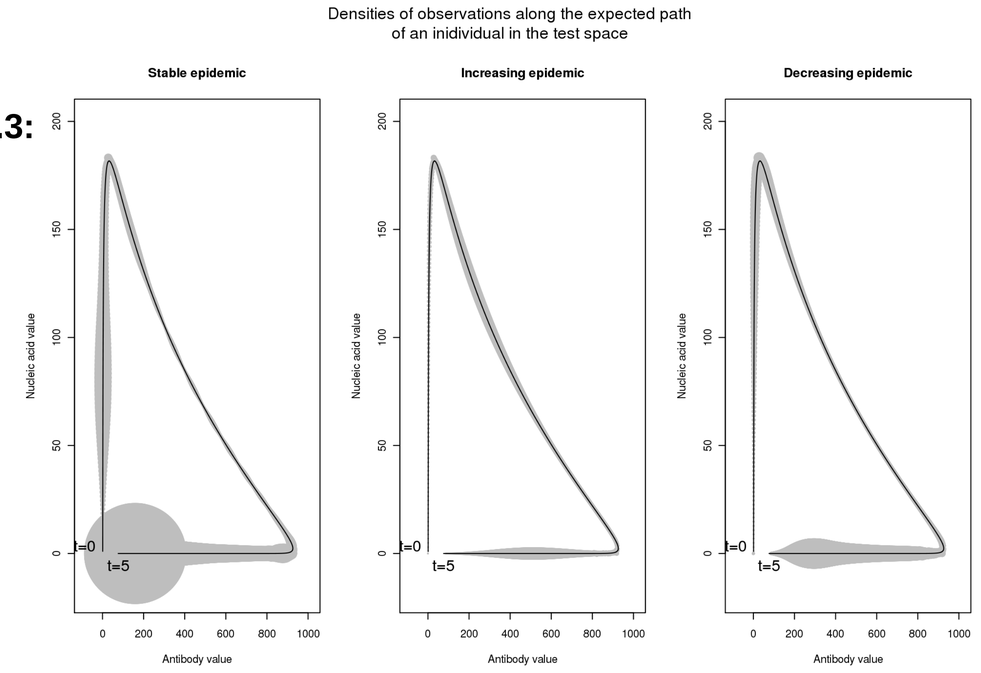
We investigated the relationship between historical incidence and observed test looking at both the marginal 1D densities of antibodies, and the joint 2D densities of antibodies/NA.
In the 1D case (seen above), we compared the estimated density function (using the standard *density* function of *stats*) of antibodies among those positive for the pathogen, with the estimated density function among those negative for antibodies.
As can be seen, clear differences showed up between data simulated from a stable epidemic, data simulated from an increasing epidemic, and data simulated from a decreasing epidemic.

Fig.2:



The 2D equivalent is under development, but the idea is to simulate a 2D density from a known joint distribution of the test over time under the null hypothesis of a stable epidemic, and compare this with the observed density.
The upper graph shows the expected path an individual would follow over time in the test space, together with an estimated density using the *smoothScatter* function in base R.
Figure 3 shows the density of observations along the expected path for the same scenarios as in Fig1. By aggregating the data in Fig2, you increase the sensitivity to differences between various trends. The shown graphs are based on known time points; we're still trying to figure out an unbiased way to calculate the density along the path.

Fig.3:



What next?

The work conducted so far is inspiring in that it shows that there is a potential of gaining additional knowledge by testing for pathogens using multiple different type of tests. However, the visualization techniques shown here only proof-of-concept. A lot of work remains; setting up a solid R framework to facilitate usage by epidemiologists and other non-statisticians, generating quantitative estimates, and clarifying the theoretical framework for the analyses.
We are currently in the process of implementing the data generating process using the *simecol* package for simulating ecological systems.
We are also hoping to build on the *mclust* package to generate numerical estimates of the composition of distributions in the 1D case, and from that getting quantitative estimates of the trend.

Ultimately, the developed methods will be put in an R package for the exploration and analysis of these kinds of test data, for use both by WildTech partners and by interested epidemiologists.

Acknowledgements

The work is being conducted under the auspices of three institutions; The University of York, Biomathematics and Statistics Scotland, and Scottish Agricultural College. The authors are indebted to all of them.
The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 222633, and is being conducted as part of the WildTech project.

BioSS
THE UNIVERSITY of York
WildTech
Novel Technologies for Surveillance of Emerging and Re-emerging Infections of Wildlife
SEVENTH FRAMEWORK PROGRAMME