Applications of ecological niche modelling for mapping the risk of Rift Valley fever in Kenya

Purity Kiunga¹,², Philip Kitala¹, K.A. Kipronoh³, Jusper Kiplimo², Gladys Mosomtai⁴ and Bernard Bett²

¹University of Nairobi; ²International Livestock Research Institute; ³Kenya Agricultural and Livestock Research Organization; ⁴icipe

The first Sub-Saharan Conference on Spatial and Spatiotemporal Statistics
Johannesburg, South Africa, 17-21 November 2014
Geographical area of study: Kenya
Outline

• Background and objectives
• Methodology
• Outputs
• Discussion
• Conclusion
Rift Valley fever (RVF) is an acute febrile arthropod-borne zoonotic disease

**Aetiology**: RVFV, family *Bunyaviridae*, genus *Phlebovirus*.

**RVF History in Kenya**:

- 1912- first report of RVF-like disease in sheep
- 1931-Virus isolation and confirmation (Daurbney *et al.*, 1931)
- 2006/2007-last outbreak in Kenya
Background

RVF niche

✓ *El Niño/Southern Oscillation* (ENSO) – causing flooding
✓ soil types - solonetz, solanchaks, planosols
✓ Elevation - less than 1100m asl
✓ Natural Difference Vegetation Index (NDVI) - 0.1 units more than 3 months
✓ Vector - *Aedes*, *Culicine*, and others
✓ Temperature
Objective

Map RVF potential distribution

 ✓ Disease occurrence maps

 ✓ This study used ecological niche modelling:
   • Uses presence data
   • Shows potential areas where RVF can occur
Methodology

Two way
> ENM
> Logit
Methodology

>ENM

Strategy for estimating the actual or potential geographic distribution of a species; is to characterize the environmental conditions that are suitable for the species and then identify where suitable environments are distributed in space
Methodology

1. **Collate GIS database of environmental layers** (e.g., temperature, precipitation, soil type)

2. **Process environmental layers to generate predictor variables that are important in defining species’ distributions** (e.g., maximum daily temperature, frost days, soil water balance)

3. **Map the known species’ distribution** (localities where the species has been observed, and sometimes also localities where the species is known to be absent)

4. **Apply modeling algorithm** (e.g., Maxent, artificial neural network, general linear model, boosted regression tree)

5. **Model calibration** (select suitable parameters, test importance of alternative predictor variables)

6. **Create map of current distribution**

7. **Predict species’ distribution in a different region** (e.g., for an invasive species) or for a different time period (e.g., under future climate change)

8. **Test predictive performance through additional fieldwork or data-splitting approach** (statistical assessment using test such as AUC or Kappa)

9. **If possible, test prediction against observed data, such as occurrence records in an invaded region, or distribution shifts over recent decades**
ENM

✓ Environmental layers
  – Land use and land cover maps
  – Precipitation
  – NDVI
  – Temperature
  – Elevation
  – Soil types

✓ Occurrence data
  - Data describing the known distribution of a species (RVF) exist in a GIS format – point data (lat long)
✓ Algorithms (GARP)

Genetic Algorithm for Rule set Production (GARP); an open modeller software creates ecological niche models for species

GARP algorithm was used to map the actual and potential distribution of Rift Valley fever distribution in Kenya and result compared to random forest cover

Uses rules of selection, evaluation, testing and incorporation or rejection in modeling
ENM

✓ Evaluation

Assess the accuracy (Confusion matrix)

• Area Under Cover (AUC)

> Defined by plotting sensitivity against 1 specificity across the range of possible thresholds of 0.82

Swets (1988) and Manel et al. (2001) AUC of 0.5 – 0.7= poor, 0.7 – 0.9= moderate and >0.9 is high performance

• A partial receiver operating characteristic (ROC) analyses AUC prediction with a value of 1.77(0= not good, at 1.0= very good and 2.0= excellent)
• ENM output was compared with random forest (covers more spatial areas and shows consistency)

• Jackknife analysis = Variable analysis
• 2-year data (2006-07)

• Case-control design cases (grid 25 by 25 km +ve (20%) control)

• Done to rank variables contributing to output

• Input (soil, rain, NDVI, elevation, temperature)
Jackknife Output rainfall and temperature
Discussion

• ENM only shows spatial distribution areas of the disease but doesn’t show variable contribution to output (correlation)

• Shows potential and high-risk areas where disease can occur

• Both models important shows consistency

• Logit done = ranks variable contribution to output and shows relationship between variables
Conclusion

This will help policymakers to know which areas to focus their attention and put plans in place when the outbreak occurs again.
THANK YOU
better lives through livestock

ilri.org