

Developing Spatio-temporal Prediction Models for Arbovirus Activity in Northern Australia Based on Remotely Sensed Bioclimatic Variables

Bernhard Klingseisen, Robert J. Corner

Department of Spatial Sciences, Curtin University of Technology
GPO Box U1987, Perth WA 6845, Australia
bernhard.klingseisen@gmail.com, r.corner@curtin.edu.au

Mark Stevenson

EpiCentre, Massey University
Private Bag 11-222, Palmerston North 4442, New Zealand
m.stevenson@massey.ac.nz

ABSTRACT

Vector-borne diseases pose an ongoing threat to public and animal health in the north of Australia. A number of surveillance programs are in place to determine the extent of virus activity and control the risk, but these are labour- and cost intensive while producing data with large temporal and spatial gaps. Using the example of Bluetongue virus, the aim of this study was to investigate the potential of remotely sensed variables to facilitate the development of area-wide predictive models that complement traditional surveillance activities.

Bioclimatic variables were derived for the Northern Territory from MODIS and TRMM remote sensing data products covering a period of nine years. Spatial and temporal uncertainty in the surveillance data required the annual aggregation of environmental variables on a pastoral property level. Generalized Additive Models (GAM) were developed based on variables such as NDVI and land surface temperature to produce annual prediction maps of virus activity. External validation showed that the model correctly predicted 75% of the results from cattle stations tested for Bluetongue. Remaining uncertainty in the model can be mainly attributed to the spatio-temporal inconsistency of the available surveillance data.

This case study has developed a cost-effective approach based on a set of robust environmental predictors that facilitate the generation of arbovirus prediction maps soon after the peak of risk for infection. While this research focused on Bluetongue Virus, we see a large potential to expand the method to other areas and viruses particularly in view of the increasing populations in Northern Australia.

KEYWORDS: arbovirus, epidemiology, remote sensing, Northern Australia, spatio-temporal modelling

1 INTRODUCTION

Australia hosts more than 75 arthropod-borne (arbo-) viruses, which are a group of viruses maintained in nature by transmission between vertebrate hosts (e.g. humans, livestock, feral animals) and bloodsucking insects, such as mosquitoes, ticks or biting midges. Twelve of these viruses are of concern for human health, including Dengue, Ross River or Murray Valley Encephalitis. Others, such as Bluetongue, Akabane or Bovine Ephemeral Fever are affecting animal health and impact the economics of the livestock industry. Each component of the transmission cycle is influenced by the interplay of underpinning environmental variables such as climatic conditions, vegetation, terrain and soil properties, which characterise the vector and host habitat structure. This paper presents the results of a case study that investigated the utilisation of remotely sensed variables for the area wide prediction of Bluetongue Virus (BTV) to support

ground based monitoring efforts in the remote areas of Northern Australia. For full details see Klingseisen (2010).

The first evidence of Bluetongue virus in Australia was provided by the isolation of serotype 20 from *Culicoides* collected at Beatrice Hill, Northern Territory, in March 1975. Following this discovery, the existing sentinel herd system, established in the late 1960's in Northern Australia, was immediately expanded to other regions. In subsequent years, further serotypes, namely 1, 2, 3, 7, 9, 15, 16, 21 and 23, were isolated (Animal Health Australia 2001, 2008, 2009). Of those, serotypes 1 and 21 are widely distributed across northern and eastern coastal regions of Australia, including the northern parts of Western Australia (WA), the Northern Territory (NT) and Queensland (QLD) as well as eastern Queensland and north eastern New South Wales (NSW) (Kirkland 2004). This corresponds well with the observed limits of the biting midge *Culicoides brevitarsis* (Murray and Nix 1987) which, due to its abundance, is considered the most important vector species for BTV transmission in Australia, although it is not the most competent vector.

More competent vectors and pathogenic BTV serotypes have as yet been confined to the far North of the continent, far removed from susceptible commercial sheep populations. Although disease has been observed in small groups of susceptible sheep when they were moved to the tropical part of the NT (Kirkland 2004), no signs of clinical disease have been found in sheep raised in the area. However, due to factors such as climate change and increases in the frequency and distance of livestock movements, the risk of a southward spread of the virus cannot be ignored. The presence of BTV resulted in stringent export restrictions being applied to both cattle and sheep after its first isolation. Even if trade is not prevented, it becomes very expensive due to serological tests and other measures necessary to reduce the perceived BTV risk (Oliver 2004).

Since 1992, surveillance for BTV has been conducted as part of the National Arbovirus Monitoring Program (NAMP). The system consists of sentinel herds in areas where commercial livestock are raised with sites selected representatively to allow mapping of the distribution of infection (Melville 2004). Most herds are positioned along the border between expected infected and uninfected areas, or where infection occurs irregularly. As part of irregular opportunistic serological surveys, herds within the affected areas are also tested to assess the seasonal intensity of infection. Supplementary areas, expected to be uninfected, are monitored to verify their BTV-free status (Animal Health Australia 2009). Sentinel herds usually consist of 10 to 25 young cattle, which have initially been negatively tested for BTV antibodies. Cattle in sentinel herds are replaced annually or after seroconversion has occurred. Sentinel herds are bled at regular intervals, with the frequency being approximately proportional to the observed arbovirus activity. Seroconversion information is supplemented by vector trapping and the quantification of *Culicoides* species at sentinel herd sites and numerous other strategic locations (Cameron 2004).

Under the NAMP and its predecessor programs, many years of monitoring data have been collected and detailed knowledge of virus ecology and the linkages between environmental factors and virus presence has been accumulated. The ongoing operation of the NAMP is expensive and sampling is often impractical in the remote areas of Northern Australia, particularly during and after the wet season. Due to these constraints it is useful to investigate how accumulated data can be used to inform future surveillance and to determine the utility of alternative information sources, such as remotely sensed data, to predict both the place and time of seroconversion.

Models predicting the spatial and temporal distribution of BTV and other diseases as well as their vectors have been developed for the Mediterranean and North Africa (Tatem et al. 2003), Spain (Calvete et al. 2009), Sicily (Purse et al. 2004) and Corsica (De La Rocque et al. 2004; Guis et al. 2007). In these models, environmental variables derived from remote sensing satellites have been used as surrogates for the bioclimatic factors that define vector and host habitats. Amongst the variables tested, temperature minima and maxima, and the Normalised Difference Vegetation Index (NDVI) have shown the most significant correlations with BTV vector occurrence (Baylis et al. 1998; Calvete et al. 2009; Purse et al. 2007; Wittmann, Mellor, and Baylis 2001).

Remote sensing satellites are able to provide area-wide coverage of environmental conditions over long periods and hence facilitate monitoring of vector presence as it changes over time. The Moderate Resolution Imaging Spectroradiometers (MODIS) on NASA's earth observation satellites Terra and Aqua are able to measure geophysical parameters at a spatial resolution between 250 and 1000 m four times daily. This configuration aids in obtaining cloud-free images in tropical areas, where cloud cover is a major issue and optical remote sensing data may

be only occasionally available during the wet season. A number of freely available pre-processed data products, including Land Surface Temperature, and Vegetation Indices, make MODIS a valuable resource for epidemiological applications (Tatem, Goetz, and Hay 2004). Data and derived products from the Tropical Rainfall Measuring Mission (TRMM), a joint mission by the US and Japanese Space Agencies NASA and JAXA, respectively, are also freely available as 3-hourly and daily rainfall estimates in tropical regions around the equator (Ebert, Janowiak, and Kidd 2007; Hay 2000). Both earth observation systems are operational at the time of writing and follow-up missions are planned to ensure the ongoing supply of data to build long term time series.

Despite the increasing availability and application of remote sensing data in spatial epidemiology (Ostfeld, Glass, and Keesing 2005) and although a number of models for BTV have been developed for Australia based on environmental variables (Cameron 2000; Murray 1995; Murray and Nix 1987; Ward and Thurmond 1995), no attempt has been made to use satellite images for the prediction of BTV seroconversion risk. In this study, we investigate the use of remotely sensed data as input for BTV distribution models as a means for providing better focus for BTV surveillance, i.e. identify areas of the country with a higher risk for BTV seroconversion. Using data for the seasons 2000/2001 to 2007/2008 we tested a range of environmental variables as predictors within a multivariate regression model. We then assessed the model's predictive capability using data from the 2008/2009 season. In contrast to other studies that compiled seasonality variables over a number of years to model general habitat suitability (Hay, Graham, and Rogers 2006; Scharlemann et al. 2008), prediction maps for virus presence are generated for each individual year. This approach facilitates the generation of annual prediction maps and the analysis of drivers for the inter-annual variations in virus distribution.

2 METHODOLOGY

2.1 Study Area

The analysis focused on the Northern Territory of Australia, where Bluetongue is endemic north of 13° latitude, and epidemic south of that line, with occurrences in Katherine and the Victoria River District in some years. The peak of virus activity is typically between January and May, but late seroconversions have occurred between August and September (Animal Health Australia 2001). The climate of the Northern Territory is under the influence of the north-west monsoon with a summer wet season and a winter dry-season. From north to south the climate changes from Köppen-Geiger zones Tropical Savannah to Hot Arid Steppe and Hot Arid Desert (Peel, Finlayson, and McMahon 2007), with annual rainfall ranging between 250-500 mm in the South and 1000-1700 mm in the North. Most rain falls between November and April. While it is warm in the North throughout the year, frost can occur in the arid interior as far north as Alice Springs (Wilson et al. 1990).

2.2 Arbovirus Monitoring Data

Surveillance data for BTV were obtained from the NAMP information system (Cameron 2004) on a pastoral property level for the years 2000 – 2009. This includes data on the seroconversion of animals within sentinel herds as well as data on serological status of animals derived from periodic cross sectional surveys. Due to the irregularity of sampling intervals at most localities, the true date of seroconversion is unknown. Therefore, data on BTV status were aggregated on an annual basis, using samples collected from 1 November to 31 October the following year. This period incorporates seroconversions at the peak of transmission during and shortly after the wet season between late summer and autumn (Animal Health Australia 2001) and the months thereafter to the onset of the new season. Figure 1 provides an overview of monitoring sites that were included in this study and a summary of the seroconversion data is provided in Table 1.

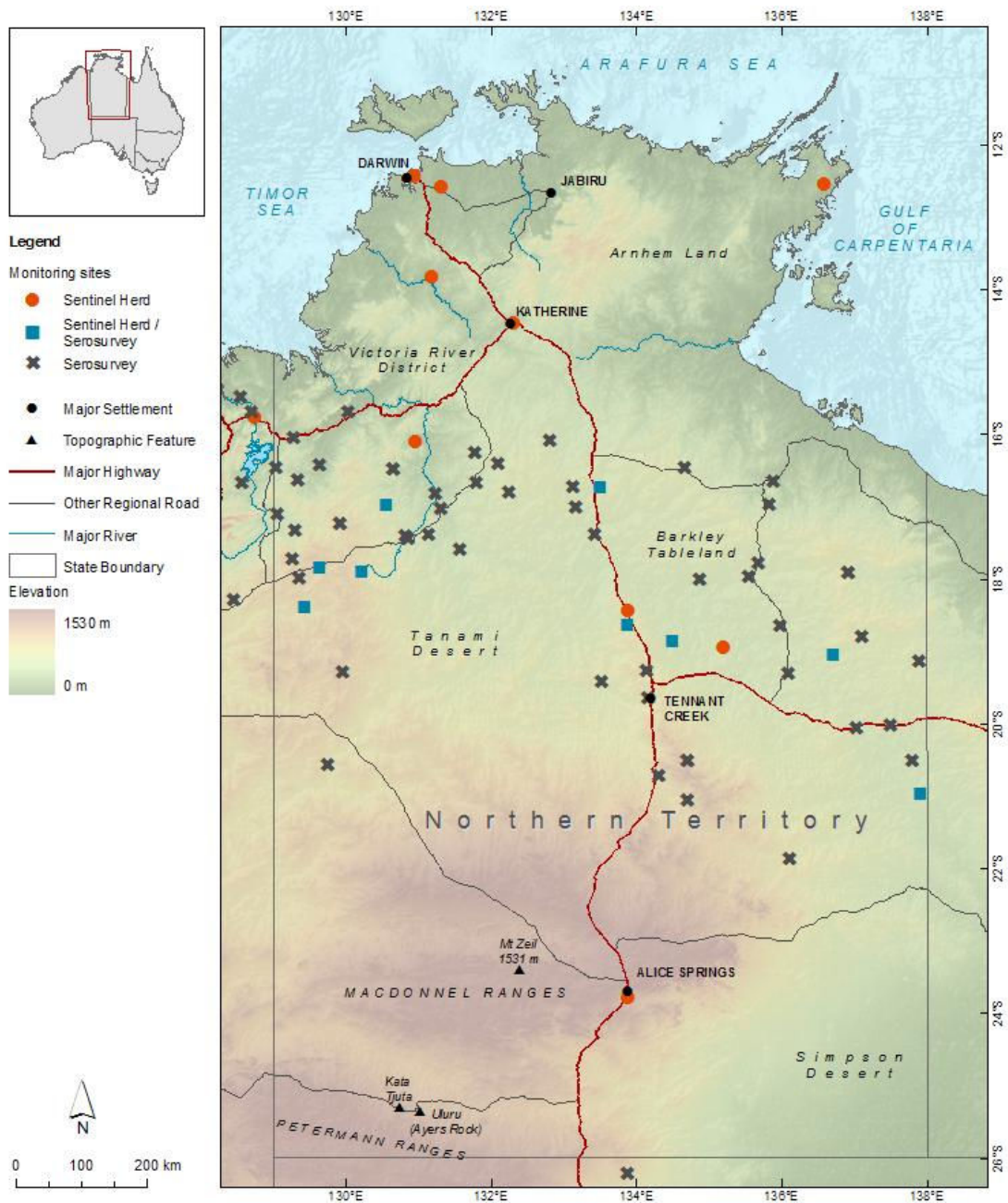


Figure 1: Map of NAMP sites in the NT tested between November 2000 and October 2009

Table 1: Summary of BTV seroconversion data used in this study

Year	No. of tested properties	Prevalence (95% confidence interval)
2000/2001	14	0.36 (0.13-0.37)
2001/2002	18	0.67 (0.41-0.87)
2002/2003	25	0.68 (0.46-0.85)
2003/2004	23	0.74 (0.52-0.90)
2004/2005	18	0.28 (0.10-0.53)
2005/2006	10	0.80 (0.44-0.97)
2006/2007	17	0.70 (0.44-0.90)
2007/2008	37	0.54 (0.37-0.71)
2008/2009	38	0.71 (0.54-0.85)

2.3 Remotely sensed environmental variables

Based on a review of BTV host and vector ecology in Australia, land surface temperature, precipitation, and vegetation indices as well as phenological characteristics of the growing season were identified as most relevant for the presence of the virus. These factors are related to the survival and abundance of the main BTV vector *C. brevitarsis*, which maintains a strong relationship with cattle as the preferred hosts for blood feeding, oviposition and larval development in bovine dung pads. Considering the large area covered in this study, part of which is characterised by a tropical climate, moderate to low spatial resolution satellite data products from MODIS and TRMM delivered at a high temporal resolution were considered most appropriate to monitor environmental conditions associated with vector habitat dynamics.

To derive vegetation indices and land surface temperature, the MODIS MOD11A2 Land Surface Temperature product (LST) as well as the MODIS MCD43A4 Nadir Bi-directional reflectance Adjusted Surface Reflectance data (NBAR), both from Collection 5, were acquired from the Land Processes Distributed Active Archive Center (LP DAAC) using the NASA Warehouse Inventory Search Tool (WIST). The LST product is produced every 8 days at 1 km spatial resolution for day and night time temperatures and achieves an accuracy of better than 1 K (Wan 2008). The NBAR product is produced every 8 days at 500 m spatial resolution from data collected over a 16-day period (Schaaf et al. 2002). Using the MODIS reprojection tool MRT (Dwyer and Schmidt 2006) the single tiles were reprojected from the original sinusoidal projection to a geographical coordinate system, resampled to 0.0023° and mosaiced into a seamless dataset for each 8-day epoch from 2000 to 2009. Low quality pixels, such as those affected by clouds, were rejected using the data quality information supplied. Spatial gaps present in the data were interpolated with a mean filter if the gaps were not wider than five pixels. Temporal gaps were interpolated linearly, if no more than two successive 8-day periods were missing. The NBAR data were further processed to calculate the NDVI as well as the Enhanced Vegetation Index (EVI) (Huete et al. 2002).

Daily rainfall was accumulated at a 0.25° (about 25 km) grid resolution for an intercomparison study of precipitation time series from TRMM and interpolated surfaces from the Australian Bureau of Meteorology rain gauge observations (Renzullo 2008).

2.4 Seasonal bioclimatic variables

To investigate the relationship between the annual serological status of a pastoral property and environmental conditions, a number of seasonal variables were generated for each year. This step also reduced the environmental data layers to a manageable number before analysis, whilst retaining as much information about seasonal characteristics as possible (Robinson, Rogers, and Williams 1997). The aggregation periods that have been selected based on ecological principles and the compositing periods of remotely sensed data are annual (1.11.-31.10.), virus transmission season (1.11.-1.6.) summer (3.12.-25.2.), autumn (26.2.-1.6.) and the previous winter (2.6.-28.8.). Aggregated over these periods, maps of minimum, mean, maximum day and night land surface temperatures, maximum NDVI and EVI, as well as accumulated rainfall were produced. Furthermore, the time integrated NDVI, as an important phenological characteristic of the growing season, was extracted from a smoothed NDVI time series using TIMESAT (Jönsson and Eklundh 2004). Prior to statistical analysis, the 17 environmental variables were averaged over an entire pastoral property to be linked with the seroconversion data using the zonal statistics tool in ArcGIS (ESRI Inc. 2008). Due to the lack of additional data, the whole property was considered of being at equal risk for virus transmission, regardless of any factors that favour or exclude the presence of host and vector habitats. Only pixels with NDVI values below 0.1, representing surface water and bare ground, as well as built-up areas, mines, road reserves, perennial water courses and lakes, and swamps extracted from 1:250,000 topographic base data were excluded from the analysis.

2.5 Statistical model development and validation

Univariate analyses were conducted using the R statistical environment (R Development Core Team 2010) to identify those variables able to discriminate between BTV positive and

negative properties. The Kruskal-Wallis test was used to identify variables significantly associated with property BTV status. Subsequently, variables were assessed pairwise for collinearity to avoid using variables with redundant information in the same model.

A non-parametric Generalised Additive Model (GAM) approach (Hastie and Tibshirani 1990) was selected due to its flexibility and the ability to explore the often non-linear shape of the response curve for each variable efficiently. Responses to changes in environmental variables that cannot be described by linear or higher order functions are frequently found in ecological and epidemiological modelling, where e.g. rising temperature might initially increase vector activity, but after reaching an optimum temperature range, an additional increase in temperature might have an adverse effect (Gubbins et al. 2008). GAMs were fitted using the `gam()` function implemented within the `mgcv` package (Wood 2004) in R. The model building process started with a model that included all predictor variables with a Kruskal Wallis $p < 0.2$. A stepwise exclusion approach was then applied, removing the least significant variable in model until all variables were significant ($p < 0.05$). The p-value provided by the `gam()` function for each smooth term is approximate and based on the test statistic motivated by Nychka's (1988) analysis of the frequentist properties of Bayesian confidence intervals for smoothers (Wood 2004). Where variables had the same or a similar p-value, those variables that are known to be important for the BTV cycle were retained in the model. Hence, the model was developed utilising expert knowledge as well as statistics. This minimises the risk of potentially using relationships which may be statistically significant but have no grounding in the ecological process being modelled. Several models were developed using this hybrid approach and the best model was chosen by maximising the discriminatory ability, given by the area under the receiver operator characteristic (ROC) curve (Pearce and Ferrier 2000).

The effects of spatial autocorrelation have not been considered when building the fixed effects GAM described in this study. A typical approach to identify unaccounted for second-order spatial effects, is to examine the model residuals for evidence of spatial autocorrelation. If there is evidence for spatial autocorrelation, the model can be extended to account for the spatial dependence between sites. The peculiarities of the data that are dealt with in this study made application of this approach difficult. Firstly, due to the large distances between sample sites it is unlikely that spatial autocorrelation was present. Secondly, the samples that have been used for model building describe the BTV status over several years and some stations have multiple representations. Consequently, residuals would have to be analysed for each year separately. In the case of evidence for autocorrelation and considering that the effects will vary between years, distribution models would be needed to be developed on an annual basis. Given that the fixed effects model has originally been built from eight years of data comprising only 162 samples, reducing the sample size to between ten (season 2005/2006) and 25 samples annually (season 2002/2003) would lead to instability in the model. It was therefore concluded that the data available in this study would not be sufficient to support an investigation of spatial autocorrelation. The previously developed stable and ecologically meaningful model is therefore used to generate annual prediction maps.

In contrast to the station average variables used for model development, prediction maps for the seasons 2000/2001 through to 2008/2009 were generated on a pixel basis, using the annual bioclimatic variable raster maps. In order to process data for the whole study area, the nominal spatial resolution was reduced to 1 km to avoid computer memory limitations. It should be noted that areas outside the bounds of pastoral properties are unlikely to support virus survival due to the lack of cattle as the major group of hosts. Nevertheless, area-wide prediction maps have been generated to delineate areas where the environmental conditions might in principle provide suitable host and vector habitats. Although there is no evidence that native animals are involved in BTV transmission, recent laboratory experiments have demonstrated that camels, which are plentiful in central Australia, may act as hosts for BTV (Batten et al. 2011). Knowing about areas with a higher probability for virus transmission also aids in logistic planning of livestock movement and if necessary, precautionary measures can be taken to avoid contact between hosts and potentially infected vectors.

External validation of the model was conducted with data from the season 2008/2009, which were not used for model development. Model validation was performed in two ways. Firstly, we predicted the probability for observing BTV seroconversion in 2008/2009 based on station average variables (which weren't included in model development) and then assessed the model predictions

against the serological test results from that season. The second method took the following approach. Since model output was expressed in terms of BTV probability for each 1 km × 1km pixel that comprised the study area, it was of interest to determine what cut point should be used to classify locations as BTV positive. For the model developed using data for 2000/2001 to 2007/2008, the sensitivity and specificity of model predictions as a function of a series of candidate cut points ranging from 0 to 1 was plotted and the cut point that maximised both sensitivity and specificity identified. Predictions of BTV probability were then made using the 2008/2009 data. All pixels with a predicted BTV probability greater than the identified cut point of 0.6 were then classified as positive and compared with the observed BTV status for each station for 2008/2009.

3 RESULTS

All of the 17 bioclimatic variables initially considered for model development showed significant difference between positive and negative sites with a Kruskal Wallis $p < 0.05$. Although only observable at the least significant digits of the p-value, the strongest difference was found for the vegetation and rainfall variables. Strong correlations were generally found between vegetation and rainfall variables and also between seasonal and summer rain, indicating that most rainfall occurs during the summer months. The strongest correlation was present between maximum NDVI and maximum EVI variables for any aggregation period, which meant that only one of these two variables should be included in the model. In relation to temperature, the highest correlations were found between the summer and autumn day temperature variables.

The GAM with the best ability to discriminate between BTV positive and negative stations, with an area under the ROC curve of 0.8644, is summarised in Table 2. In this model, maximum seasonal NDVI (*maxndviss*), mean day temperature of the warmest 8-day period in autumn (*maxlstdau*), mean day temperature in summer (*meanlstdsu*) and mean night temperature of the previous winter months (*meanlstdnpw*) are the predictor variables.

Table 2: Significance of smooth (*maxndviss*, *maxlstdau*, *meanlstdnpw*) and parametric (*meanlstdsu*) model terms in the BTV distribution model

Model terms	Df/edf	Chi-square	p-value
s(maxndviss)	3.001	11.001	0.0117*
s(maxlstdau)	3.254	9.551	0.0282*
s(meanlstdnpw)	7.015	7.699	0.0241*
meanlstdsu	1	5.087	0.0416*

*: The variable is statistically significant ($p \leq 0.05$)

The GAM plot (Figure 2) shows the smoothers for these four covariates and the 95% confidence interval. The horizontal axis of each plot shows the density of samples for each value of each covariate in the model. The plots show that the effect of one of the four variables on the outcome can be expressed by a linear function (*meanlstdsu*), while the other three variables were best described by a unimodal smoothing function (*maxndviss*, *maxlstdau*, and *meanlstdnpw*). Model coefficients and the shape of the smooth functions indicate that the probability of BTV presence is generally higher in areas with high NDVI, although this relationship is not linear. High average summer day temperatures on the other hand decrease the probability of BTV presence, which could be explained by the risk of desiccation, particularly in the southern arid part of the Northern Territory. Maximum autumn day temperatures, or more precisely the average day temperature of the warmest 8-day period in autumn, had a positive effect on BTV presence up to about 40°C, but higher temperatures reduced the likelihood of virus activity. Higher mean day temperatures in winter were associated with BTV detection.

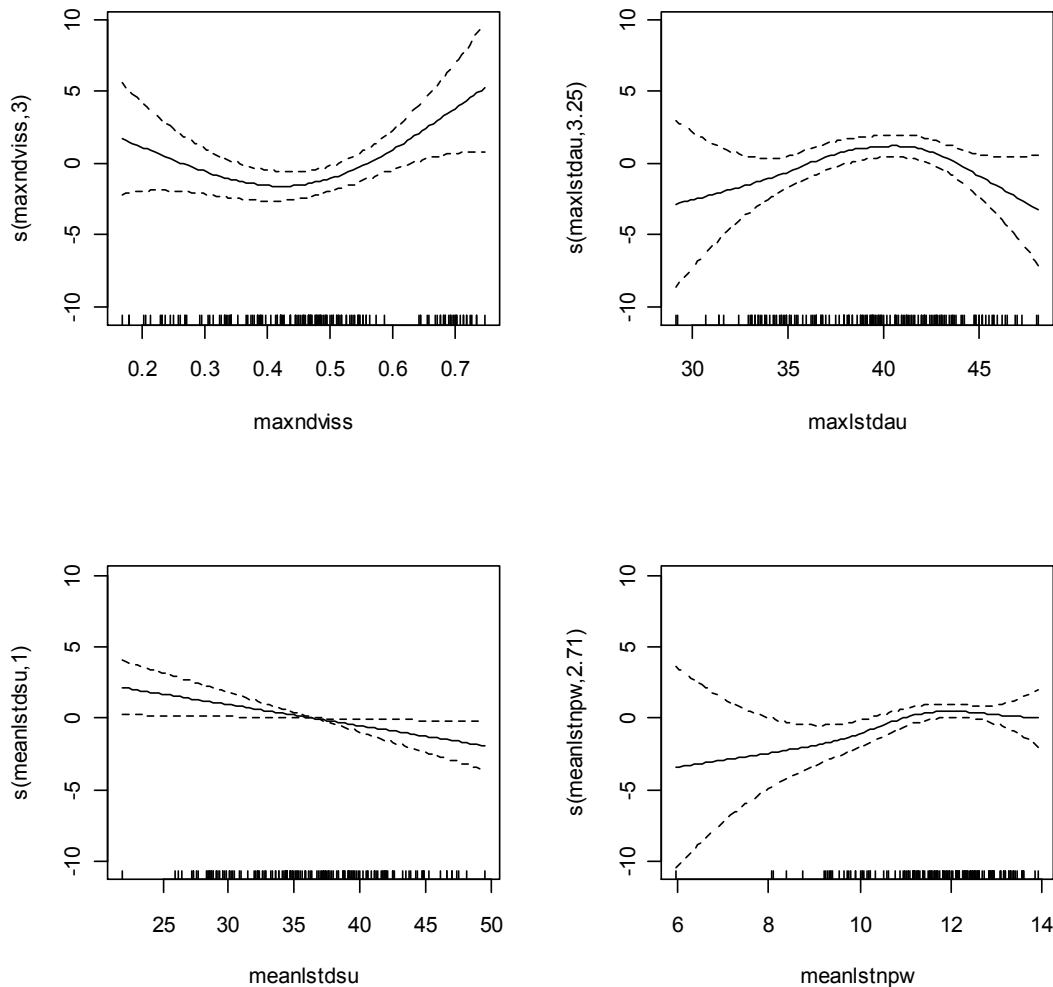


Figure 2: Smoothed fits to the variables *maxndviss*, *maxlst dau*, *meanlst dsu* and *meanlst npw* used in the final BTV distribution model for the Northern Territory (solid line). Dashed lines give 95% confidence intervals.

The maps in Figure 3 show, for each year, the estimated probability that cattle tested in an area will be BTV seropositive. Superimposed on each map are the points positioned at the centroid of each station that took part in testing, colour coded according to BTV status. Predictions from the GAM model consistently spatially replicated the distribution of BTV in cattle tested between 2000 and 2009. In particular, the season 2004/2005, for which dry and hot conditions were experienced throughout the study area, with associated low virus activity, demonstrates the predictive capabilities of the model. The general pattern of decreasing probability of BTV presence from North to South is present throughout the years analysed, but the divide between high and low probability varies between the years and largely follows the observed BTV presence. Also, smaller patches of high and low probability are depicted well (e.g. season 2003/2004) and are confirmed by the test results. Comparison of predicted and observed BTV occurrence also reveals that some positive test results are found in areas of low predicted probability. However, often there is a small area within a cattle station with a high probability. Without the knowledge of the true geographical origin of the tested animals, one can assume that they could have come from these areas. It is therefore recommended that the maximum observed probability be used as a guideline, when deciding if a property is to be declared BTV positive or negative.

External validation based on data from 38 pastoral stations tested between November 2008 and October 2009 resulted in an area under the curve (AUC) of 0.8215, which again demonstrated that the model had good discriminatory capabilities based on the classification of AUC values by Hosmer and Lemeshow (2000). A number of measures have been derived from the classified presence/absence maps using both the station average environmental variables as well as the

maximum predicted probability per station. As can be seen in Table 3, using the station average variables for predictions on a station level resulted in the highest overall accuracy (measured by Cohen's Kappa). However, using the maximum predicted probability per station decreased the likelihood of false negative classifications, which is a crucial factor when the objectives of a surveillance program (as in this case) was to detect seroconversion, particularly in areas that have previously been free of virus activity.

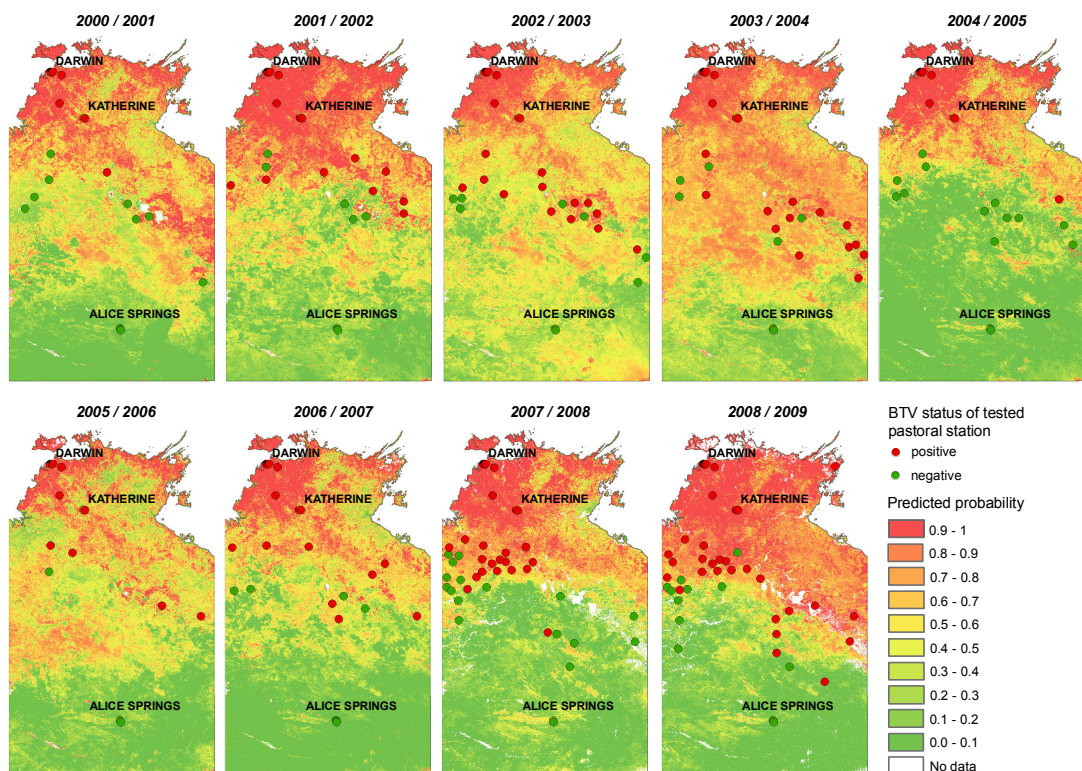


Figure 3: Predicted BTV seropositivity in the NT for the seasons 2000/2001 to 2008/2009 in relation to the surveyed presence/absence data

Table 3: Accuracy measures for the classified predicted presence/absence maps

Measure	Predictions from station average variables	Maximum predicted probability per station
Correct classification rate	78.9%	76.3%
Sensitivity	81.5%	100%
Specificity	72.7%	18.2%
False positive rate	27.3%	81.8%
False negative rate	18.5%	0%
Positive predictive power	88%	75%
Negative predictive power	61.5%	100%
Cohen's Kappa	0.514	0.24

4 CONCLUSION

This study presents a novel application of remotely sensed environmental variables used in a generalised additive model to predict BTV seropositivity in Northern Australia. The selected environmental data and the modelling approach are well suited to reproduce ground observations. However, the model can only be as good as the input data, and as such, the results of this study are seen as complimentary to the data produced by conventional surveillance. What this study has demonstrated well, is that the spatial gaps between the sampled pastoral properties may be filled

to provide a more complete picture of potential virus activity. To a limited extent, the model can also be used to forecast BTV seropositivity. The fact that the model uses data collected during the previous winter as well as the summer and autumn months means that after data acquisition and processing, a prediction can be generated at the end of June each year. Depending on the timing of the annual muster, during which cattle are accessible for blood sampling, most data collected as part of serological surveys are not available in the NAMP database earlier than the predictions. The model therefore has the capability of acting as an early indicator for potential virus activity hotspots to indicate where monitoring should be intensified.

Model building should not be considered a static process. As seen in Europe, where distribution models for BTV vectors have been built using long term climate records (e.g. Tatem et al., 2003), the predictions are soon outdated and therefore less likely to predict incursion of new virus strains and vectors and their vast expansion under a changing climate. This happened recently with BTV 8 (Mellor et al. 2009). A similar scenario is also likely to happen in Australia, where the windborne incursion of vectors infected with a new BTV strain from Southeast Asia pose an ongoing threat to livestock. While surveillance in northern Australia is able to detect these incursions, it is important to constantly monitor the environmental conditions that may create favourable habitats for BTV vectors in regions beyond the current endemic zone. We suggest using the increasing archive of virus and environmental data to gradually recalibrate the model on an annual basis. It is anticipated that the significance of single parameters may change with the availability of a broader data base, e.g. due to the effects of climate change.

This research has demonstrated the limitations of using data from an operational surveillance system that has been designed predominantly to fulfil international biosecurity requirements, such as NAMP. In order to facilitate more in-depth spatio-temporal studies of virus environmental relationships based on NAMP data, the following improvements are recommended:

- Record spatial information more accurately, including the location of vector traps, bleeding sites and, if known, the location(s) at which tested cattle have been held prior to bleeding.
- Maintain a core network of sentinel herds in strategic locations which are tested at regular intervals. This will provide a continuous record of virus activity data and facilitate comparison of inter-annual variations. Bluetongue surveillance in the Northern Territory, based on government research stations, provides an example of such a network.
- Select sites for serological surveys, where possible, based on perceived risk (e.g. informed by previous survey results and modelled probability).

Presently, the study focuses on environmental variables that can be derived in a timely and efficient manner using largely pre-processed and validated datasets. These are not available for all factors known to be associated with BTV. Further research is necessary to investigate other factors that might be associated with virus presence. Of particular interest are the effects of humidity, air temperature, landscape structure, and cattle density. Potential data sources need to be identified and assessed. Potential for future research has been identified in the 2.5 decades of data from the Advanced Very High Resolution Radiometer (AVHRR). These could be analysed to make full use of the long records of epidemiological data from more than 30 years of arbovirus surveillance on Australia. Also the definition of alternative aggregation periods for the bioclimatic variables is worth investigating.

ACKNOWLEDGEMENTS

This research was funded by the Australian Biosecurity Cooperative Research Centre for Emerging Infectious Disease (AB-CRC), Project 3.036R, represented by the project partners Curtin University of Technology, Department of Agriculture and Food Western Australia, Berrimah Veterinary Laboratories, Northern Territory Department of Primary Industries, Fisheries and Mines (DPIFM), University of Western Australia (Arbovirus Surveillance and Research Laboratory) and the Mosquito Borne Disease Control Unit, WA Department of Health. We want to acknowledge all contributors of the National Arbovirus Monitoring Program for access to the surveillance data and scientific advice. We also want to thank Luigi Renzullo from CSIRO Land and Water, Canberra for providing the TRMM 3B42 daily rainfall accumulates for the entire study period.

REFERENCES

- Animal Health Australia. 2001. *The history of bluetongue, Akabane and ephemeral fever viruses and their vectors in Australia*. Deakin, ACT, Australia: Animal Health Australia.
- Animal Health Australia. 2008. *National Arbovirus Monitoring Program - 2007-2008 Report*. Australian Animal Health Council.
- Animal Health Australia. 2009. *National Arbovirus Monitoring Program - 2008-2009 Report*. Australian Animal Health Council.
- Batten, C. A., B. Harif, et al. 2011. Experimental infection of camels with bluetongue virus. *Research in Veterinary Science* 90 (3): 533-535.
- Baylis, Bouayoune, et al. 1998. Use of climatic data and satellite imagery to model the abundance of *Culicoides imicola*, the vector of African horse sickness virus, in Morocco. *Medical and Veterinary Entomology* 12 (3): 255-266.
- Calvete, C., R. Estrada, et al. 2009. Ecological correlates of bluetongue virus in Spain: Predicted spatial occurrence and its relationship with the observed abundance of the potential *Culicoides* spp. vector. *The Veterinary Journal* 182 (2): 235-243.
- Cameron, A. R. 2000. Development of a tempero-spatial disease distribution model for arboviruses and their vectors in Australia. *Proceedings of the IXth Symposium of the International Society for Veterinary Epidemiology and Economics (ISVEE)*, Breckenridge, Colorado, USA, 6-11 August, 2000.
- Cameron, A. R. 2004. Data management and analysis systems for bluetongue virus zoning in Australia. *Veterinaria Italiana* 40 (3): 365-368.
- De La Rocque, S., V. Michel, et al. 2004. Remote sensing and epidemiology: examples of applications for two vector-borne diseases. *Comparative Immunology, Microbiology and Infectious Diseases* 27 (5): 331-341.
- Dwyer, J., and G. Schmidt. 2006. The MODIS Reprojection Tool. In *Earth Science Satellite Remote Sensing*, ed. J. J. Qu, W. Gao, et al., 162-177. Springer Berlin Heidelberg.
- Ebert, E. E., J. E. Janowiak, et al. 2007. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society* 88 (1): 47-64.
- ESRI Inc. 2008. *ArcGIS, version 9.3* (program).
- Gubbins, S., S. Carpenter, et al. 2008. Assessing the risk of bluetongue to UK livestock: Uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *Journal of the Royal Society Interface* 5 (20): 363-371.
- Guis, H., A. Tran, et al. 2007. Use of high spatial resolution satellite imagery to characterize landscapes at risk for bluetongue. *Veterinary Research* 38: 669-683.
- Hastie, T. J., and R. J. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.
- Hay, S. I. 2000. An overview of remote sensing and geodesy for epidemiology and public health application. In *Remote Sensing and Geographical Information Systems in Epidemiology, Advances in Epidemiology* 47, ed. S. I. Hay, S. E. Randolph, et al., 1-35. London: Academic Press.
- Hay, S. I., A. Graham, et al., eds. 2006. *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications, Advances in Parasitology* 62. London: Academic Press.
- Hosmer, D. W., and S. Lemeshow. 2000. *Applied logistic regression*. 2nd ed, *Wiley series in probability and statistics Texts and references section*. New York: Wiley.
- Huete, A., K. Didan, et al. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* 83 (1-2): 195-213.

- Jönsson, P., and L. Eklundh. 2004. TIMESAT--a program for analyzing time-series of satellite sensor data. *Computers & Geosciences* 30 (8): 833-845.
- Kirkland, P. D. 2004. Bluetongue viruses, vectors and surveillance in Australia – the current situation and unique features. *Veterinaria Italiana* 40 (3): 47-50.
- Klingseisen, B. 2010. Spatio-temporal modelling of bluetongue virus distribution in Northern Australia based on remotely sensed bioclimatic variables. PhD Thesis, Department of Spatial Sciences, Curtin University of Technology, Perth, Western Australia
- Mellor, P. S., S. Carpenter, et al. 2009. Bluetongue in Europe and the Mediterranean Basin. In *Bluetongue*, 235-264. London: Academic Press.
- Melville, L. F. 2004. Bluetongue surveillance methods in an endemic area: Australia. *Veterinaria Italiana* 40 (3): 184-187.
- Murray, M. D. 1995. Influences of vector biology on transmission of arboviruses and outbreaks of disease: the *Culicoides brevitarsis* model. *Veterinary Microbiology* 46 (1-3): 91-99.
- Murray, M. D., and H. A. Nix. 1987. Southern Limits of Distribution and Abundance of the Biting-Midge *Culicoides-brevitarsis* Kieffer (Diptera, Ceratopogonidae) in Southeastern Australia - an Application of the Growest Model. *Australian Journal of Zoology* 35 (6): 575-585.
- Nychka, D. 1988. Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association* 83: 1134-1143.
- Oliver, G. R. 2004. Bluetongue trade issues – an Australian perspective. *Veterinaria Italiana* 40 (4): 688-690.
- Ostfeld, R. S., G. E. Glass, et al. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution* 20 (6): 328-336.
- Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133 (3): 225-245.
- Peel, M. C., B. L. Finlayson, et al. 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences* 11 (5): 1633-1644.
- Purse, B. V., B. J. J. McCormick, et al. 2007. Incriminating bluetongue virus vectors with climate envelope models. *Journal of Applied Ecology* 44 (6): 1231-1242.
- Purse, B. V., A. J. Tatem, et al. 2004. Modelling the distributions of *Culicoides* bluetongue virus vectors in Sicily in relation to satellite-derived climate variables. *Medical and Veterinary Entomology* 18 (2): 90-101.
- R Development Core Team. 2010. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
- Renzullo, L. 2008. Considerations for the blending of multiple precipitation data sets for hydrological applications. *Proceedings of the 4th International Precipitation Working Group Workshop*, Beijing, China, 13-17 October 2008.
- Robinson, T., D. Rogers, et al. 1997. Univariate analysis of tsetse habitat in the common fly belt of Southern Africa using climate and remotely sensed vegetation data. *Medical and Veterinary Entomology* 11 (3): 223-234.
- Schaaf, C. B., F. Gao, et al. 2002. First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment* 83 (1-2): 135-148.
- Scharlemann, J. P. W., D. Benz, et al. 2008. Global Data for Ecology and Epidemiology: A Novel Algorithm for Temporal Fourier Processing MODIS Data. *PLoS ONE* 3 (1): e1408.
- Tatem, A. J., M. Baylis, et al. 2003. Prediction of bluetongue vector distribution in Europe and north Africa using satellite imagery. *Veterinary Microbiology* 97 (1-2): 13-29.
- Tatem, A. J., S. J. Goetz, et al. 2004. Terra and Aqua: new data for epidemiology and public health. *International Journal of Applied Earth Observation and Geoinformation* 6 (1): 33-46.

- Wan, Z. 2008. New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. *Remote Sensing of Environment* 112 (1): 59-74.
- Ward, M. P., and M. C. Thurmond. 1995. Climatic factors associated with risk of seroconversion of cattle to bluetongue viruses in Queensland. *Preventive Veterinary Medicine* 24 (2): 129-136.
- Wilson, B. A., P. S. Brocklehurst, et al. 1990. *Vegetation Survey of the Northern Territory, Australia. Technical Report No. 49*. Darwin: Conservation Commission of the Northern Territory. http://nt.gov.au/nreta/natres/natveg/vegmapping/pdf/ntveg_rpt.pdf (accessed
- Wittmann, E. J., P. S. Mellor, et al. 2001. Using climate data to map the potential distribution of *Culicoides imicola* (Diptera: Ceratopogonidae) in Europe. *OIE Revue Scientifique et Technique* 20 (3): 731-740.
- Wood, S. N. 2004. Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. *Journal of the American Statistical Association* 99 (467): 673-686.

