

# Fine-temporal forecasting of outbreak probability and severity: Ross River virus in Western Australia

I. S. KOOLHOF<sup>1</sup>\*, S. BETTIOL<sup>2</sup> AND S. CARVER<sup>1</sup>\*

<sup>1</sup>*School of Biological Sciences, University of Tasmania, Hobart, Tasmania, Australia*

<sup>2</sup>*School of Medicine, University of Tasmania, Hobart, Tasmania, Australia*

*Received 5 April 2017; Final revision 18 June 2017; Accepted 2 August 2017;  
first published online 4 September 2017*

## SUMMARY

Health warnings of mosquito-borne disease risk require forecasts that are accurate at fine-temporal resolutions (weekly scales); however, most forecasting is coarse (monthly). We use environmental and Ross River virus (RRV) surveillance to predict weekly outbreak probabilities and incidence spanning tropical, semi-arid, and Mediterranean regions of Western Australia (1991–2014). Hurdle and linear models were used to predict outbreak probabilities and incidence respectively, using time-lagged environmental variables. Forecast accuracy was assessed by model fit and cross-validation. Residual RRV notification data were also examined against mitigation expenditure for one site, Mandurah 2007–2014. Models were predictive of RRV activity, except at one site (Capel). Minimum temperature was an important predictor of RRV outbreaks and incidence at all predicted sites. Precipitation was more likely to cause outbreaks and greater incidence among tropical and semi-arid sites. While variable, mitigation expenditure coincided positively with increased RRV incidence ( $r^2 = 0.21$ ). Our research demonstrates capacity to accurately predict mosquito-borne disease outbreaks and incidence at fine-temporal resolutions. We apply our findings, developing a user-friendly tool enabling managers to easily adopt this research to forecast region-specific RRV outbreaks and incidence. Approaches here may be of value to fine-scale forecasting of RRV in other areas of Australia, and other mosquito-borne diseases.

**Key words:** Arboviruses, Early warning forecasting, Epidemiology, Mosquito-borne disease.

## INTRODUCTION

Arboviral diseases cause a significant impact upon human health worldwide, with multiple climatic zones (including tropical, semi-arid, and Mediterranean regions) experiencing increased burdens from a plethora of emerging and re-surfing forms [1, 2]. Owing to

dependence on environmental features, their dynamics (particularly mosquito-borne pathogens) are often uniquely suited for forecasting using readily available environmental data. Because of this, the timing of disease control can be matched effectively with vectorial and disease risk. However, for many mosquito-borne diseases forecast modelling of risk is often at a coarse temporal resolution, such as monthly scales [3–5]. While this helps to understand disease dynamics, forecasts made on coarse temporal resolutions are often limited in their applicability to ‘on the ground’ disease management, such as through vector reduction and well-

\* Authors for correspondence: I. S. Koolhof and S. Carver, School of Biological Sciences, University of Tasmania, Hobart, Tasmania, Australia and School of Medicine, University of Tasmania, Hobart, Tasmania, Australia.  
(Email: koolhofi@utas.edu.au and Scott.Carver@utas.edu.au)

timed public awareness campaigns. To enhance the impact of disease control strategies, models that forecast disease at temporal scales relevant to management (e.g. weekly) and are also easily accessible to managers are highly valuable.

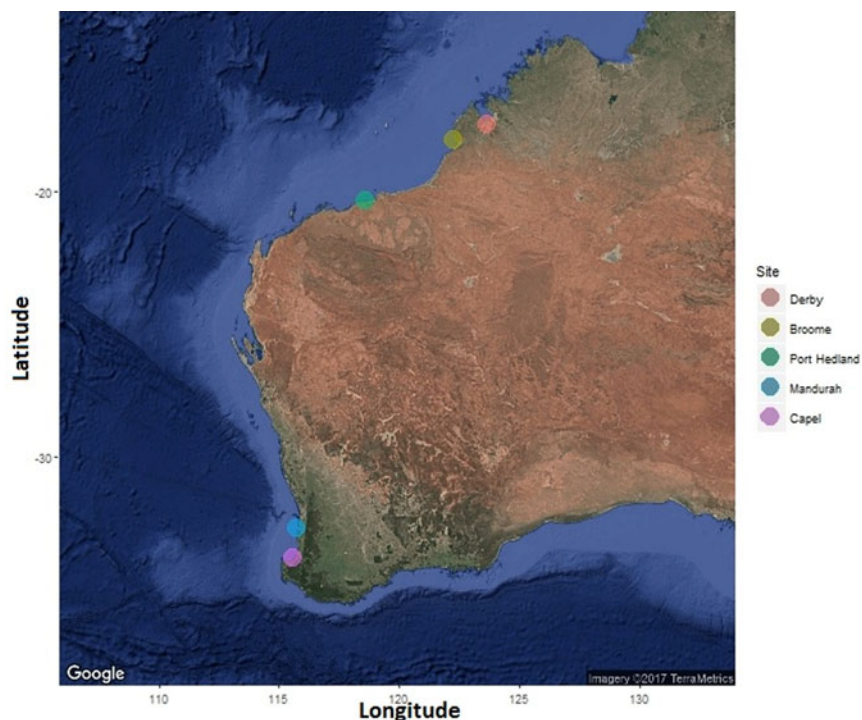
Epidemiologically Ross River virus (RRV), family *Togaviridae* genus *Alphavirus*, is Australia's most important vector-borne disease with 1451–9551 clinical notifications per year (per capita rate >40/100 000) at an estimated annual health care and lost productivity cost of \$15 million [6–8]. The epidemiology of RRV, like many mosquito-borne diseases, is fraught with complexity. The drivers involved in transmission span environmental conditions, vector and reservoir host dynamics, and existing human mitigation, all which vary across geographic and climatic regions [9, 10]. RRV epidemic areas of Australia occur across tropical, semi-arid, Mediterranean, and temperate climates. Thus, the processes underpinning RRV epidemics and their severity are complex [11–13]. Despite this complexity, studies have shown that disease forecasts can be achieved moderately well using environmental drivers in many instances, as this can be representative of many aspects of the biological mechanisms determining transmission.

A multitude of studies incorporating various designs have been used over the past three decades to define environmental determinants of RRV notifications [e.g. 5, 14, 15], and indeed many important mosquito-borne diseases. A common theme among these studies is the importance of precipitation, temperature, tidal or flood variation, and humidity. These environmental determinants directly or indirectly shape mechanisms underscoring disease dynamics, including vector mosquito activity, reservoir hosts movement, breeding interactions, and viral activity [16]. For instance, the inundation of wetlands from precipitation and high tides (where vectors are halotolerant) may cause the movement of reservoir hosts from these areas closer to the boundaries of urban areas while also providing an ideal breeding site for the mosquitos [10, 16, 17]. The outcome of these environmental drivers can enhance vector–host interactions, viral transmission and spillover to humans [11, 18, 19]. Vectors in particular are highly responsive to changes in environmental conditions, having the greatest potential for rapid translation into increased abundance and infections [15, 20, 21].

While environmental factors can predict RRV notifications, it should be acknowledged that predicting RRV disease is best achieved using detailed

knowledge of vector abundance [13] and, where available, host information [5]. However, collection of such detailed vector–host information is for many areas infeasible, particularly over long durations. In contrast, environmental information is often much more readily available and modelling approaches that can best exploit widely available data sources to predict and understand diseases are of significant value. However, not all environmental determinants are equal in all climates, and thus an appreciation of their relative importance for predicting RRV across geographic and climatic gradients is of value. Further, while previous forecast modelling has aided in the epidemiological understanding of RRV, very few provide an accessible format by which health departments may easily adopt modelling approaches to public health and disease reduction.

Here we tackle the essential challenge of predicting RRV at fine-temporal scales across multiple epidemic centres using environmental determinants of disease; and we produce a user friendly early warning system based on these findings. We use RRV notification data and environmental data over 23 years in five epidemiologically important centres in Western Australia, spanning tropical, semi-arid, and Mediterranean climates. We construct predictive models for RRV outbreaks and their severity/incidence (the probability of notifications above long-term averages and the number of disease notifications, respectively), and for one site, where data are available we determine how well timing of expenditure on mitigation coincides with RRV incidence. At a broad scale, this research fills the long-standing void in improving our understanding of environmental drivers on a fine-temporal scale to forecast RRV transmission. Building from previous literature, we provide insight into the distinct differences between Mediterranean, semi-arid and tropical areas for RRV transmission. Knowledge of the fundamental differences in RRV predictors between these climatic regions is critical for environmental and public health managers when considering region-specific approaches to vector mitigation and public education programmes. Furthermore, we provide a means to improve predictive capabilities for RRV outbreaks where mosquito surveillance is not currently being conducted. We apply this research to create a user-friendly predictive tool meant to assist public health managers with timing of vector control, decision making, and public health disease risk awareness campaigns.



**Fig. 1** Study sites. Map location of study sites. Showing Derby and Broome from a tropical region, Port Hedland from a semi-arid region, and Mandurah and Capel from a Mediterranean region.

## METHODS

### Study sites

Five sites in Western Australia were selected due to their high mosquito-borne disease attack rates and being epidemiologically important outbreak areas (Fig. 1) [6, 22–24]. The northern sites of Derby and Broome are situated in the Kimberly region and experience tropical monsoonal rainfall. Port Hedland, south of these two sites is in the Pilbara region which experiences a semi-arid climate. These northern sites commonly have *Culex annulirostris* (a freshwater breeding mosquito) and *Aedes vigilax* (a saltmarsh breeding mosquito) implicated in RRV transmission [10]. All northern sites have salt marshes neighbouring their town or shire. The southern sites of Mandurah and Capel are in the Peel and southwest geographic regions, respectively, a Mediterranean climatic area. At these sites two saltmarsh breeding mosquitoes, *Aedes camptorhynchus* and *Ae. vigilax*, are the primary vectors for RRV transmission [25]. Mandurah also plays host to one of the biggest estuarine systems in Southern Western Australia, with a multitude of canals and river systems stretching inland. All sites have mosquito vector dynamics locally driven by both tides and precipitation, except

one site, the Shire of Capel, which is non-tidal and was selected to contrast the difference in drivers in RRV transmission for non-tidal and tidal sites of the Southern Mediterranean region.

### Data sources

Data of RRV notifications was provided by the Western Australian Department of Health and the Western Australian Notifiable Infectious Diseases Database for the period of January 1991–December 2014 containing the number of notifications and the week of notification. Region-specific annual population data were collected from current and historical records from the Australian Bureau of Statistics [26]. RRV incidence for each region was then calculated using the respective annual population size with the number of RRV notifications per week to create incidence (per 1000 individuals) per week. A fixed incidence threshold was used as the definition of an outbreak, where RRV incidence above the long term average of the mean for each individual site resulted in an outbreak, excluding weeks where incidence were equal to zero [27]. Environmental data were collected for the entire study period and gaps in missing data recorded (Supplementary Table S1).

Daily tidal data were collected from the Western Australian Department of Transport and the Australia Bureau of Meteorology for all sites excluding Capel (which does not have a locally occurring tidally driven saltmarsh). Tide data were then summarised into the recorded observed maximum and minimum weekly tide heights. Derby experiences very high tides, presenting some gauging problems, where the minimum tide measurement could only be recorded down to 375 cm (above sea level) where the tide gauge is placed. As tidal measure below 375 cm were not recorded in Derby, weekly minimum tide was omitted from the Derby analysis. Furthermore, it was rare for tidal data to be continuous for the complete period (1991–2014) for any site. Where this occurred, data from additional sites that were spatially close to our primary study areas were compared using cross-correlations to account for any time lags and then linear regression models used for interpolating missing values. Spatially matched sites included: Derby matched with Broome (from 1991 to 1995), and Port Hedland matched with Cape Lambert (years 1995, 1997, and 2013–2014). All tidal data were standardised to account for any alterations in instrument measurements. In Eqn. (1),  $P_i$  is the adjusted environmental observation where values were missing,  $\mu$  represents the observed environmental variable in the comparative data set,  $\beta_1$  is the regression slope, and  $\beta_0$  is the regression intercept.

$$P_i = \mu \times \beta_1 + \beta_0. \quad (1)$$

Precipitation data were collected from the Bureau of Meteorology containing the amount of precipitation on a daily scale. The daily precipitation was summarised into weekly total precipitation. For sites which did not have complete continuous data, a relatively close comparative site was chosen, cross correlations and linear regressions were again used on to make interpolations, using the same methodology as described for tides. Comparative matched sites included: Derby matched with Yeeda (from 1991 to 1994), Mandurah matched with Karnet (from 2001 to 2014), and Capel matched with Thirlmere (from 2003 to 2014). All precipitation measurements are reported in millimetres.

Daily maximum and minimum temperatures (°C) were also provided by the Australia Bureau of Meteorology with temperature being on a daily scale. Temperatures were summarised into weekly maximum and minimum mean temperatures. For non-continuous data, comparative sites were used, as

described above. Comparative matched sites included: Derby matched with a station at the Royal Australian Air Force Base ‘Curtin’ (from 1991 to 1995), and Mandurah matched with Karnet (from 2001 to 2014). The Bureau of Meteorology has no temperature monitoring stations located in and around Capel. Temperature data for Capel were instead collected from the Shire of Donnybrook 24 km East.

## Statistical analyses

### *Environmental predictors of RRV*

We utilised a structured approach to link environmental predictors with RRV notifications. Firstly, we determined the optimal time-lags between environmental predictors and RRV incidence. To determine time lags between environmental variables and weekly RRV incidence we utilised cross-correlation analysis, a lagged effect was introduced for each environmental variable at each site [28, 29]. Temporal lags account for effects of environmental predictors on the lifecycle of the virus and mosquitoes, human incubation prior to disease notification, and host dynamics [5, 16, 30]. Cut-off points for maximum lag periods included biological considerations, visual observation of relative peaks in RRV incidence in relation to the environmental variable, and relative comparisons with other lags presented in the literature [5, 29, 31]. This meant that the cut-off periods for each environmental variable were equal across all the sites. The given cut-off time periods included 5 weeks for tidal heights, 10 weeks for precipitation, and 15 weeks for temperature measurements. Before modelling, environmental variables were checked for multi-collinearity to ensure they were not correlated with one another (>0.85 Pearson’s correlation; Supplementary Tables S2–S6).

Secondly, we applied two modelling approaches to link time-lagged environmental factors with weekly RRV: hurdle models (with a binomial logit function) for the probability of an outbreak and linear models for the severity (RRV incidence) of an outbreak. All predictor variable combinations were undertaken for each modelling approach and Akaike’s Information Criterion correction corrected for sample sizes (AICc) were determined [32]. Rank of models was assessed by delta AICc [32].

Model-averaged coefficients of the model combinations were constructed, averaged variable estimates were derived from each model where the variable was included (see the ‘Results’ section). The estimates



from the hurdle models were then back-transformed, exponentiating the model coefficients and used as predictors in forecasting the probability of a RRV outbreak (Eqn. (2)), where the severity of an outbreak (in incidence) was calculated based on the model-averaged coefficients from the linear regression models (Eqn. (3)). Where  $Y_i$  represents your predicted outcome odd ratio and incidence (Eqns (2) and (3), respectively),  $\beta_0$  represents the regression intercept, and  $\beta_{1,2,p}x_{1,2,p_i}$  represents the beta coefficients from the regression models.

$$(Y_i) = \left( \frac{\exp + (\beta_0 + (\beta_1 x_{1_i} + \beta_2 x_{2_i} + \dots + \beta_p x_{p_i}))}{1 + \exp + (\beta_0 + (\beta_1 x_{1_i} + \beta_2 x_{2_i} + \dots + \beta_p x_{p_i}))} \right), \quad (2)$$

$$(Y_i) = \beta_0 + (\beta_1 x_{1_i} + \beta_2 x_{2_i} + \dots + \beta_p x_{p_i}). \quad (3)$$

In all cases, variable importance weights were also calculated representing the sum of the *Akaike weights* over all the models in which the variable appears [32].

The fit of models to the data were assessed using a variety of techniques. For outbreak probability, we calculated sensitivity, specificity, and negative and positive predicted values based on the model-averaged coefficients. Cross-validation was also conducted using a *K*-fold validation technique with the data being split into ten equal folds and the model being tested giving both a prediction error for the leave-one-out cross-validation and an adjusted design to compensate for the leave-one-out cross-validation [33, 34]. The *K*-fold cross-validation prediction error is used to determine the amount of variation in the hurdle model predictions where we are testing predictions against the entire data available [35]. *K*-fold cross-validations were run ten times to obtain a standardised error to correct for possible errors produced in the data partitioning process. For outbreak severity we assessed model fit via *K*-fold cross-validations, again based on model-averaged coefficient effects.

#### *Relationship between RRV incidence and expenditure*

Following modelling of RRV incidence based on environmental factors, we examined the relationship between timing of expenditure on RRV control and RRV incidence. This was undertaken for Mandurah, where monthly expenditure data were available from 2007 to 2012. Data were aggregated on a monthly basis. Monthly opposed to weekly data were utilised to account for infrequent mosquito control activity throughout each year. The model-averaged

coefficients of RRV incidence in Mandurah were used to predict RRV incidence using a linear regression model (Supplementary Table S7). The residuals of the predicted incidence to predict true incidence were then used as the outcome variable predicted by the monthly expenditure. To examine model robustness, a cross-validation of the linear regression was conducted. A 4-fold cross-validation was used due to the limited size of the data. Data partitioning were repeated ten times, allowing for equal amounts of observation in each partition for each model construction and comparison.

All statistical analyses described were performed in R (Version 3.2.5, [www.r-project.org](http://www.r-project.org)) using the packages ‘pscl’, ‘MuMIn’, ‘pROC’, and ‘boot’ in RStudio (Version 0.99.489).

Using the modelled average coefficients from predicting RRV outbreaks (Eqn. (2)) and incidence (Eqn. (3)), we constructed a user friendly Microsoft Excel-based tool capable of predicting a minimum of 1 week into the future for each site studied (Supplementary Table S8). All supplementary materials are available on the Cambridge Core website.

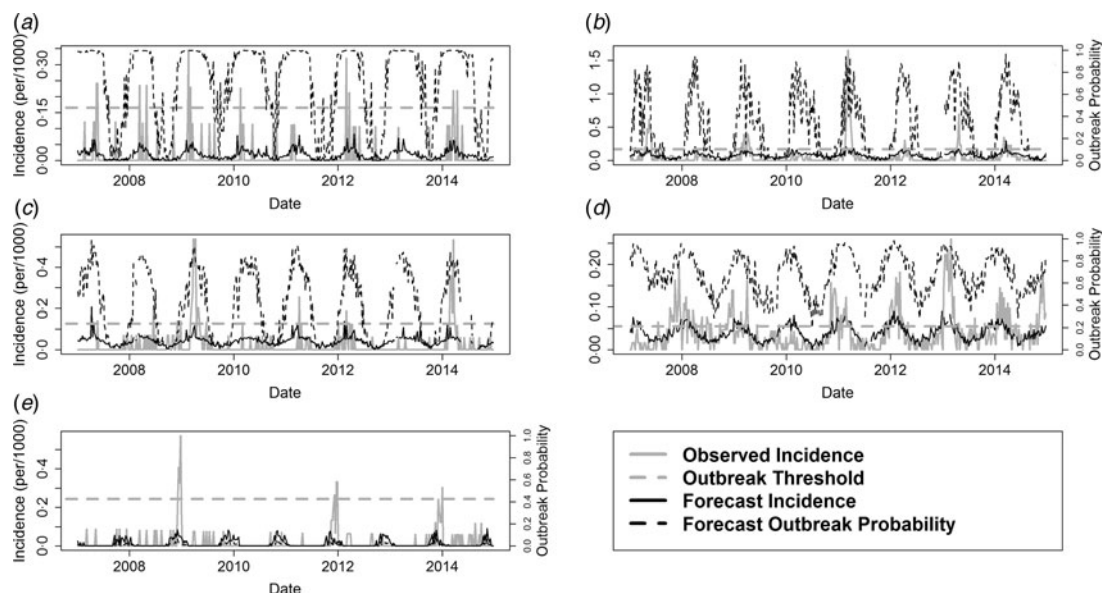
## RESULTS

There were 3567 notified cases of RRV across all sites for the study period; these included Derby with 134, Broome with 635, Port Hedland with 349, Mandurah with 2167, and Capel with 282 notified cases. The various numbers of cases across the sites predominantly reflect the population differences between the sites.

### **The weekly probability of an outbreak**

An outbreak was defined by the mean weekly incidence (weeks of zero incidence were omitted), whereby values above the mean were classified as an outbreak for each site. The incidence (per/1000 individuals) which characterised an outbreak were 0.165 in Derby, 0.169 in Broome, 0.128 in Port Hedland, 0.055 in Mandurah, and 0.244 in Capel and represented as a threshold in Figure 2. The occurrence of outbreaks differed between the northern and southern sites (Supplementary Fig. S1).

RRV outbreaks in semi-arid and tropical northern sites were positively predicted by weekly precipitation and minimum temperatures (Table 1). Weekly maximum tide heights were further positive predictors for outbreak probabilities in tropical Derby and



**Fig. 2** Outbreak probabilities and incidence across all sites. Observed incidence (left axis) plotted against the threshold, which defines and outbreak (left axis), predicted outbreak probabilities (right axis), and predicted incidence (per/1000 individuals, left axis) for a 7-year timespan from 2007 to 2014 for five sites in Western Australia: a, Derby; b, Broome; c, Port Hedland; d, Mandurah; e, Capel.

Broome, with weekly maximum temperature also becoming an important predictor in Derby (Table 1). For the Southern Mediterranean climate site of Mandurah, in comparison, RRV outbreaks were predicted by weekly minimum and maximum temperatures and maximum tide heights (Table 1). With the exception of the semi-arid Port Hedland, maximum tide heights had similar importance in model selection between the northern and southern sites. The weekly lag between precipitation and RRV activity was greater in the southern sites, relative to the northern sites (Table 1). Variable importance suggested that weekly minimum temperatures in the northern sites had a greater relative influence on the probability of an outbreak than in the southern sites. Furthermore, weekly maximum temperature had a greater variable importance weight in the southern site of Mandurah than that of the northern sites (Table 1).

Outbreak predictions of observed RRV outbreaks made from the modelled averaged coefficients had high sensitivity (Fig. 2, Table 2). Specificity on the other hand was lower (Table 2). Models for the sites Broome and Port Hedland both had the greatest balance between high model sensitivity and specificity (Table 2). This was reflected when visually plotted, where weekly outbreak probabilities and predicted incidence increase when observed incidence rose (Fig. 2). The error from the *K*-fold cross-validation

suggests that the models were consistently predicting outcomes of similar accuracy across the entire study period for all sites (Table 2). The exception to this was Capel, where outbreaks were not well predicted for.

### The weekly severity of an outbreak

Similar to predicting outbreaks, weekly RRV incidence across the northern sites were driven by weekly minimum temperatures and precipitation (Table 3). There were some differences in predictors of RRV for the northern sites. For instance, weekly maximum temperatures and tidal heights were associated with RRV incidence in Derby, and maximum tide heights in Broome (Table 3). Weekly RRV incidence in the southern site of Mandurah was predicted by weekly minimum temperatures and maximum tide heights (Table 3). For Capel minimum temperatures exhibited a trending relationship to RRV incidence, and incidence was negatively associated with maximum temperatures.

While predicted incidence were generally conservative, prediction trends matched that of weekly observed incidence (Fig. 2). *K*-fold cross-validation of the linear regression models, with the exception of Capel, found very little variation in cross-validation error suggesting the model accuracy and agreement

Table 1 Environmental hurdle model characteristics of Ross River virus across all sites.

Site & variable	Lag (weeks)	Hurdle Model			
		OR	SE (logit)	Variable importance	P-value
<b>Derby</b>					
$T_{\min}^*$	10	1.640	0.151	0.99	0.001
$T_{\max}^*$	18	1.296	0.110	0.47	0.018
$TH_{\max}^*$	2	1.017	0.003	0.84	<0.001
$P_{\text{tot}}^*$	6	1.004	0.003	0.49	0.215
<b>Broome</b>					
$T_{\min}^*$	13	1.447	0.070	>0.99	<0.001
$T_{\max}$	18	1.067	0.075	0.22	>0.999
$TH_{\min}$	16	0.999	0.002	0.09	0.527
$TH_{\max}^*$	0	1.012	0.004	>0.99	<0.001
$P_{\text{tot}}^*$	8	1.011	0.002	>0.99	<0.001
<b>Port Hedland</b>					
$T_{\min}^*$	10	1.384	0.056	>0.99	<0.001
$T_{\max}$	12	1.057	0.062	0.16	0.372
$TH_{\min}$	8	1.002	0.003	0.15	0.440
$TH_{\max}$	5	1.002	0.004	0.12	0.617
$P_{\text{tot}}^*$	5	1.012	0.003	0.98	<0.001
<b>Mandurah</b>					
$T_{\min}^*$	0	1.166	0.062	0.64	0.014
$T_{\max}^*$	0	1.115	0.043	0.64	0.011
$TH_{\min}$	6	1.007	0.010	0.17	0.470
$TH_{\max}^*$	0	1.020	0.007	0.90	0.003
$P_{\text{tot}}$	10	1.004	0.007	0.13	0.525
<b>Capel</b>					
$T_{\min}^*$	18	0.853	0.080	0.47	0.488
$T_{\max}^*$	18	0.699	0.129	0.90	0.006
$P_{\text{tot}}$	11	1.009	0.009	0.18	0.297

Environmental variables predicting RRV outbreaks: OR, odds ratio; SE, standard error. Where the values:  $P_{\text{tot}}$ , is for weekly precipitation;  $T_{\min}$ , minimum temperature;  $T_{\max}$ , maximum temperature;  $TH_{\min}$ , minimum tide height; and  $TH_{\max}$ , maximum tide height.

\* Represents environmental variables included in the best-fit model.

in predictability when validated against the entire data set in their repeated predictability (Table 3).

### Timing of mitigation expenditure to RRV incidence

To examine the relationship between RRV incidence and expenditure on mitigation in Mandurah from 2007 to 2014 (while taking into account the environmental determinates of incidence) we undertook a linear regression analysis on residual RRV incidence from the model-averaged full model. Residual RRV incidence was positively related to expenditure ( $\beta^{(\text{cost})} = 7.60 \times 10^{-06}$ ,  $F = 19.7$ ,  $df = 70$ ,  $P = < 0.001$ ),

suggesting proportionality between the two (Fig. 3). Model accuracy assessed using a  $K$ -fold cross-validation with 4-folds repeated ten times, showed model agreement indicated consistency in this relationship (overall cross-validation error = 0.014).

### Forecasting tool

We utilised our weekly modelling of outbreaks and incidence to create a user-friendly forecasting tool for each site (Supplementary Table S8). Utilising the environmental lags, predictions into the future could be made. Across the sites, future predictions could be made for: 2 weeks into the future for Derby, 1 week into the future for Broome, 5 weeks into the future for Port Hedland, and 1 week into the future for Mandurah, Capel was omitted from the forecasting tool due to poor model performance.

### DISCUSSION

Transmission of mosquito-borne disease is controlled by a combination of environmental, and vector and reservoir host dynamics, which vary across geographical and climatic regions. In forecasting the transmission of RRV, previous studies have used monthly environmental data and have highlighted the importance of having mosquito trapping data to supplement environmental predictors [13, 30, 36]. While these studies have been vital in developing our understanding of RRV epidemics, their course temporal resolution and recruitment for labour intensive surveillance [13, 24, 29, 37–42] limit practical application for timely forecasts to inform mitigation activities and public awareness campaigns.

Our weekly temporal scale forecast models suggest that in Western Australia, the environmental determinants of RRV activity differs among climatic regions. Southern Mediterranean regions are less dependent on precipitation compared with tropical and semi-arid environments in the northern areas of the state. Precipitation did not significantly predict RRV outbreaks or incidence in the Mediterranean sites compared to the semi-arid and tropical sites. The fine-temporal drivers of RRV across climatic regions (Mediterranean, semi-arid, and tropical) are poorly studied and critical in creating effective region-specific mosquito-borne disease management.

Differences in environmental drivers of RRV in this study may be reflective of geographical differences in landscape between these regions. The semi-arid and

Table 2 Sensitivity, specificity, and cross-validation of model performance and predictions

	Sensitivity	Specificity	PPV	NPV	K-fold cross-validation error (range)
Derby	1	0.304	0.034	1	0.0224 (0.0221–0.0225)
Broome	0.798	0.831	0.263	0.982	0.0545 (0.0542–0.0549)
Port Hedland	0.914	0.656	0.123	0.993	0.0454 (0.0450–0.0465)
Mandurah	0.975	0.179	0.192	0.973	0.1814 (0.1801–0.1824)
Capel	0	1	NA	0.973	0.0463 (0.0459–0.0466)

Sensitivity and specificity of the predictions produced from the hurdle models using a standard 0.5 cut-off point and the cross-validation error. PPV and NPV represent positive and negative predictive values, respectively.

Table 3 Environmental linear regression model characteristics of Ross River virus across all sites

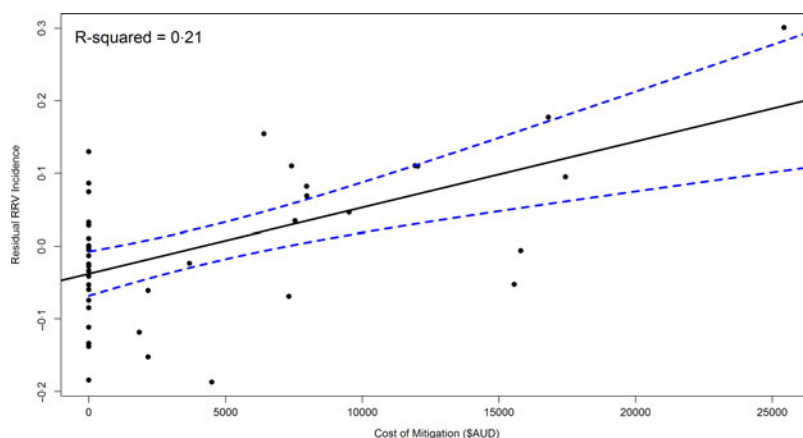
	$\beta$	SE	Variable importance	P-value	K-fold cross-validation error (range)
<b>Derby</b>					
$T_{\min}$	$9.302 \times 10^{-04}$	$3.355 \times 10^{-04}$	0.95	0.006	0.002 (0.001–0.006)
$T_{\max}$	$1.603 \times 10^{-03}$	$4.666 \times 10^{-04}$	0.99	<0.001	
TH <sub>min</sub>	$1.153 \times 10^{-04}$	$3.311 \times 10^{-05}$	0.99	<0.001	
$P_{\text{tot}}$	$2.291 \times 10^{-04}$	$3.904 \times 10^{-05}$	>0.99	<0.001	
<b>Broome</b>					
$T_{\min}$	$4.159 \times 10^{-03}$	$6.535 \times 10^{-04}$	>0.99	<0.001	0.010 (0.004–0.018)
$T_{\max}$	$1.904 \times 10^{-03}$	$1.343 \times 10^{-03}$	0.50	0.157	
TH <sub>min</sub>	$5.495 \times 10^{-05}$	$3.699 \times 10^{-05}$	0.52	0.138	
TH <sub>max</sub>	$2.160 \times 10^{-04}$	$5.034 \times 10^{-05}$	>0.99	<0.001	
$P_{\text{tot}}$	$7.036 \times 10^{-04}$	$7.850 \times 10^{-05}$	>0.99	<0.001	
<b>Port Hedland</b>					
$T_{\min}$	$2.873 \times 10^{-3}$	$4.567 \times 10^{-4}$	>0.99	<0.001	0.004 (0.002–0.007)
$T_{\max}$	$-2.255 \times 10^{-4}$	$7.272 \times 10^{-4}$	0.28	0.757	
TH <sub>min</sub>	$4.063 \times 10^{-5}$	$3.911 \times 10^{-5}$	0.38	0.309	
TH <sub>max</sub>	$5.553 \times 10^{-5}$	$4.890 \times 10^{-5}$	0.41	0.257	
$P_{\text{tot}}$	$6.136 \times 10^{-4}$	$8.485 \times 10^{-8}$	>0.99	<0.001	
<b>Mandurah</b>					
$T_{\min}$	$4.027 \times 10^{-3}$	$1.459 \times 10^{-3}$	0.89	0.006	0.004 (0.001–0.004)
$T_{\max}$	$1.800 \times 10^{-3}$	$1.204 \times 10^{-3}$	0.55	0.135	
TH <sub>min</sub>	$1.043 \times 10^{-4}$	$2.419 \times 10^{-4}$	0.30	0.667	
TH <sub>max</sub>	$5.676 \times 10^{-4}$	$1.793 \times 10^{-4}$	0.98	0.002	
$P_{\text{tot}}$	$2.044 \times 10^{-4}$	$1.689 \times 10^{-4}$	0.43	0.227	
<b>Capel</b>					
$T_{\min}$	$-5.292 \times 10^{-3}$	$3.149 \times 10^{-3}$	0.60	0.093	0.033 (0.003–0.047)
$T_{\max}$	$-1.193 \times 10^{-2}$	$3.925 \times 10^{-3}$	0.97	0.002	
$P_{\text{tot}}$	$4.166 \times 10^{-4}$	$3.553 \times 10^{-4}$	0.42	0.242	

Environmental variables predicting RRV incidence:  $\beta$ , beta coefficient; SE, standard error; and the K-fold cross-validation error. Where the values:  $P_{\text{tot}}$ , is for weekly precipitation;  $T_{\min}$ , minimum temperature;  $T_{\max}$ , maximum temperature; TH<sub>min</sub>, minimum tide height; and TH<sub>max</sub>, maximum tide height.

tropical sites have extensive mangrove and saltmarsh areas surrounding them, becoming inundated by precipitation leading to mosquito breeding. While the Mediterranean site of Mandurah has an extensive waterway system resulting in a lower impact of precipitation on RRV transmission. Similar to previous studies in the south-western parts of Western Australia [13, 43]

we found that in the Mediterranean site of Mandurah, weekly maximum and minimum temperatures and maximum tide heights were important contributors to RRV epidemics. In the semi-arid and tropical sites the weekly environmental factors minimum temperature, maximum tide height, and precipitation generally predicted RRV notifications and transmission with





**Fig. 3** Mitigation expenditure on residual Ross River virus (RRV) incidence. Mitigation expenditure influence on RRV incidence for Mandurah. The best-fit model combination residuals of a linear model being predicted by the cost of mitigation on a monthly basis (95% confidence intervals in blue).

minor differences between these regions. This may demonstrate the role geographical and environmental characteristics have among Mediterranean, semi-arid, and tropical regions. The models constructed for Mandurah were the only ones to consistently have predicted incidence exceeding the threshold that we classified as an outbreak of RRV. The remaining sites' predictions may be conservative due to the relative small human populations and low cases of RRV being reported. While incidence prediction may be conservative for Derby, Broome, and Port Hedland, periods of sustained RRV activity and peaks still provide useful information during decision-making processes when undertaking mitigation activities.

This research also supports previous studies on mosquito–host interactions, their role in RRV transmission and outbreaks, owing to maximum tidal heights in estuarine, flood plains, and wetland areas [42]. The tropical regions of Derby and Broome experience some of the highest tidal variation in Australia, and in the southern hemisphere [44, 45]. In both Derby and Broome the weekly maximum tidal height was an important driver in the probability of an outbreak and also incidence, with the inundation of mangroves being known to force animal reservoirs closer to urban areas [16]. The Southern Mediterranean site of Mandurah has an extensive tidal estuarine salt marsh area coupled with inflowing rivers creating ideal conditions for mosquito breeding [46, 47]. In Mandurah, weekly maximum tide height had similar odds of increasing the probability of an outbreak as the tropical sites. However in Port Hedland, RRV outbreaks were not significantly predicted by maximum tide height. Instead, precipitation

in Port Hedland had a greater influence on the odds of a RRV outbreak than that of any other study site. Port Hedland, in contrast to our other coastal sites, is industrialised and frequently used as a major shipping transport hub, has three surrounding suburbs, which are further inland, and has a deeper port within the estuary.

Following a precipitation event in coastal areas, the northern tropical regions of Western Australia experienced RRV activity sooner than that of Mediterranean southern regions. However, the southern site of Mandurah experienced much shorter environmental lags in RRV activity from changes in temperature than that of the northern sites. Both these findings may reflect the environmental differences among Mediterranean, semi-arid, and tropical regions and are of importance when tackling mosquito-borne diseases at a regional level. These lagged effects of RRV activity occurring after an environmental event vary greatly between regions across Australia, reflecting the biotic and abiotic processes involved in the transmission [5, 10, 16, 24, 48]. Our findings help to illustrate the latitudinal responses in RRV activity in relation to environmental factors among climates [49].

While the economic impact of RRV infections have been investigated [50, 51], there is often little knowledge of the role mitigation expenditure has in association with RRV activity [50, 51]. Our findings indicate that while variable, expenditure increases proportional to that of residual RRV notifications (after having controlled for the effect of environmental predictors). This finding helps to support and justify current mitigation practices already in place, and variability in the association illustrates the complex nature of managing

this mosquito-borne disease. It could be argued that the relationship between mitigation expenditure and RRV incidence should be negative. However, it is important to recognise that mosquito mitigation is only carried out during times when mosquito control officers suspect mosquito breeding is most likely, and potential for disease outbreak is greater. Thus, when RRV incidence is likely to be at its greatest, so too should mitigation expenditure. As a consequence, the strength of the positive association becomes a proxy for the correct timing for mitigation activities and expenditure.

A considerable strength to this study is the extensive history of RRV notifications and environmental data in both constructing forecasts and validating the models. This is the first study, to our knowledge, to employ hurdle models to predict longitudinal RRV outbreak dynamics [52]. While our approach does not necessarily imply our models are superior to previous methods, our models do allow for alternative approaches when modelling mosquito-borne disease data at fine-temporal resolutions. Previous studies forecasting RRV outbreaks and their associated predictors have used logistic and negative binomial regression, and seasonal autoregressive integrated moving average (SARIMA) modelling techniques [5, 13, 14, 24, 49]. Hurdle models offer some distinct attributes that may make them well suited to predicting outbreaks. From our models, we found strong agreement among our model validation analyse. Model sensitivity and specificity in predicting outbreaks were generally greater than other studies, and these sensitivities rivalled those that explicitly include mosquito surveillance in modelling [13].

Resulting from our findings was the construction of early warning forecasting tools for RRV outbreaks and incidence among our study sites. Previous epidemiological studies of RRV often suggest that early warning forecasting systems can be developed and would be of benefit for public health and mosquito control programmes [13, 31, 48]. Yet, to our knowledge, very few studies take the much needed step of translating their findings into a user friendly tool. Here we have applied our findings into a practical user friendly forecasting tool, which has been delivered to the Western Australian department of health and local councils, and is available in the supplementary material to this article.

It is notable that for our study site of Capel, the models predicting outbreaks and incidence were poor fits to the RRV notifications comparative to

that of the other sites. One reason for this could be the source of temperature data, which were unavailable for Capel and had to be gathered from a neighbouring town. Furthermore, the dynamics of RRV notifications at Capel were quite sporadic, and the low population size of the town meant that high sporadic numbers of RRV notifications had large influences of levels of incidence. These limitations are unfortunately difficult to overcome at this location, and accordingly we are tentative in our interpretations of results from this site. Moreover, the variation of mosquito-control programmes and its relative effect to RRV notifications was not assessed here and would provide valuable insight into the role it plays in relation to environmental conditions and RRV outbreaks.

## CONCLUSION

We present models capable of filling the void in forecasting RRV outbreaks and disease incidence on a weekly temporal scale and across multiple climates. The use of these models is particularly valuable for forecasting RRV without the need for mosquito surveillance. We also demonstrate the complex nature of RRV transmission between climatically difference regions, and the necessity to understand these differences when creating effective forecasting tools. Forecasting weekly RRV outbreaks and incidence may help with timing of mitigation activities and public health notifications at critical times when disease risk is greatest. The early warning predictive tool developed here helps translate epidemiological modelling studies into easy to use tools that can be adopted by regional and local health bodies, which may inform local decision maker's processes. The coupling of the predictive modelling with mitigation expenditure is also informative of the extent to which mitigation activities are well timed with RRV incidence. While this study bridges previous knowledge gaps in fine-temporal transmission dynamics of RRV, and the role mitigation has during outbreaks, we highlight the need for further research extending these analyses across multiple sites in similar climatic regions, so that broader generalisations may be made to assist the public health and vector management community.

## SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at <https://doi.org/10.1017/S095026881700190X>

## ACKNOWLEDGEMENTS

The authors would like to thank Peter Neville and the Western Australia Department of Health for access to the notification data and for the 'on the ground' aspect of this study. Financial support for this project was provided through the Funding Initiative for Mosquito Management In Western Australia Research Grants, administered by the Department of Health, Western Australia.

## REFERENCES

1. **Gratz NG.** Emerging and resurging vector-borne diseases. *Annual Review of Entomology* 1999; **44**: 51–75.
2. **Mackenzie JS, Gubler DJ, Petersen LR.** Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. *Nature Medicine* 2004; **10**: 98–109.
3. **Tipayamongkhogul M, et al.** Effects of the El Niño-Southern Oscillation on dengue epidemics in Thailand, 1996–2005. *BioMed Central Public Health*. 2009; **9**: 1–15.
4. **Naish S, Hu W, et al.** Weather variability, tides, and barmah forest virus disease in the gladstone region, Australia. *Environmental Health Perspectives* 2006; **114**: 678–683.
5. **Yu W, Dale P, Turner L, Tong S.** Projecting the impact of climate change on the transmission of Ross River virus: methodological challenges and research needs. *Epidemiology and Infection* 2014; **142**: 2013–2023.
6. **Harley D, Sleight A, Ritchie S.** Ross River virus transmission, infection, and disease: a cross-disciplinary review. *Clinical Microbiology Reviews* 2001; **14**: 909–932.
7. **Aaskov J, Fokine A, Liu W.** Ross River virus evolution: implications for vaccine development. *Future Virology* 2012; **7**: 173–178.
8. **Australian Government Department of Health.** Notifications for all diseases by State & Territory and year ([http://www9.health.gov.au/cda/source/rpt\\_2\\_sel.cfm](http://www9.health.gov.au/cda/source/rpt_2_sel.cfm)). Accessed 16 September 2016.
9. **Kelly-Hope LA, Purdie DM, Kay BH.** Ross River virus disease in Australia, 1886–1998, with analysis of risk factors associated with outbreaks. *Journal of Medical Entomology* 2004; **41**: 133–150.
10. **Russell RC.** Ross river virus: ecology and distribution. *Annual Review of Entomology* 2002; **47**: 1–31.
11. **Carver S, et al.** Influence of hosts on the ecology of arboviral transmission: potential mechanisms influencing dengue, Murray Valley encephalitis, and Ross River virus in Australia. *Vector-Borne and Zoonotic Diseases* 2009; **9**: 51–64.
12. **Claffin SB, Webb CE.** Ross River virus: many vectors and unusual hosts make for an unpredictable pathogen. *PLoS Pathogens* 2015; **11**: 1–5.
13. **Woodruff RE, et al.** Early warning of Ross River virus epidemics – combining surveillance data on climate and mosquitoes. *Epidemiology* 2006; **17**: 569–575.
14. **Tong S, et al.** Climate variability, social and environmental factors, and Ross River virus transmission: research development and future research needs. *Environmental Health Perspectives* 2008; **116**: 1591–1597.
15. **Tall JA, Gatton ML, Tong SL.** Ross River virus disease activity associated with naturally occurring nontidal flood events in Australia: a systematic review. *Journal of Medical Entomology* 2014; **51**: 1097–1108.
16. **Jacups SP, Whelan PI, Currie BJ.** Ross River virus and Barmah Forest virus infections: a review of history, ecology, and predictive models, with implications for tropical northern Australia. *Vector-Borne and Zoonotic Diseases* 2008; **8**: 283–297.
17. **Haynes CD, Ridpath MG, Williams MAJ.** *Monsoonal Australia: Landscape, Ecology and Man in the Northern Lowlands*. Rotterdam: A.A. Balkema, 1991.
18. **Carver S, et al.** Environmental monitoring to enhance comprehension and control of infectious diseases. *Journal of Environmental Monitoring* 2010; **12**: 2048–2055.
19. **Woodruff RE, et al.** Predicting Ross River virus epidemics from regional weather data. *Epidemiology* 2002; **13**: 384–393.
20. **Russell RC.** Seasonal activity and abundance of the arbovirus vector *Culex annulirostris* skuse near echuca, Victoria, in the Murray Valley of southeastern Australia 1979–1985. *Australian Journal of Experimental Biology & Medical Science* 1986; **64**: 97–103.
21. **Rowbottom R, et al.** Resource limitation, controphic ostracod density and larval mosquito development. *PLoS ONE* 2015; **10**: 1–13.
22. **Lindsay M, et al.** Ross River and Barmah Forest viruses in Western Australia, 2000/01–2003/04: contrasting patterns of disease activity. *Arbovirus Research in Australia* 2005; **9**: 194–201.
23. **Lindsay M, et al.** Western Australian arbovirus surveillance and research program, Annual Report: 1997–1998. Arbovirus Surveillance and Research Laboratory and the Western Australian Centre for Pathology and Medical Research. Western Australian Government, 1998.
24. **McIver L, et al.** A climate-based early warning system to predict outbreaks of Ross River virus disease in the Broome region of western Australia. *Australian and New Zealand Journal of Public Health* 2010; **34**: 89–90.
25. **Severn S, Brendan I.** 2013/14 Mosquito Management Annual Report. Mandurah, 2014.
26. **Australia Bureau of Statistics.** 3218.0 - Regional Population Growth, Australia, 2013–14 (<http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3218.02013-14?OpenDocument>). Accessed 2 February 2016.
27. **Brady OJ, et al.** Dengue disease outbreak definitions are implicitly variable. *Epidemics* 2015; **11**: 92–102.
28. **Abraham B, Ledolter J.** *Statistical Methods for Forecasting*. New York: Wiley, 1983.
29. **Tong S, Hu W.** Different responses of Ross River virus to climate variability between coastline and inland cities in Queensland, Australia. *Occupational and Environmental Medicine* 2002; **59**: 739–744.

30. **Jacups SP, et al.** Determining meteorological drivers of salt marsh mosquito peaks in tropical northern Australia. *Journal of Vector Ecology* 2015; **40**: 277–281.
31. **Williams CR, Fricker SR, Kokkinn MJ.** Environmental and entomological factors determining Ross River virus activity in the River Murray Valley of South Australia. *Australian and New Zealand Journal of Public Health* 2009; **33**: 284–288.
32. **Burnham KP, Anderson DR.** *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer, 2002.
33. **Wong TT.** Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*. 2015; **48**: 2839–2846.
34. **Meijer RJ, Goeman JJ.** Efficient approximate k-fold and leave-one-out cross-validation for ridge regression. *Biometrical Journal* 2013; **55**: 141–155.
35. **Rodriguez JD, Perez A, Lozano JA.** Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2010; **32**: 569–575.
36. **Jacups SP, et al.** Predictive indicators for Ross River virus infection in the Darwin area of tropical northern Australia, using long-term mosquito trapping data. *Tropical Medicine & International Health* 2008; **13**: 943–952.
37. **Bi P, Parton KA.** Climate variations and the transmission of Ross River virus infection in coastal and inland region of Queensland: an analysis from Townsville and Toowoomba. *Environmental Health* 2003; **3**: 73–80.
38. **Hu W, et al.** Difference in mosquito species (Diptera: Culicidae) and the transmission of Ross River virus between coastline and inland areas in Brisbane, Australia. *Environmental Entomology* 2010; **39**: 88–97.
39. **Hu W, et al.** Development of a predictive model for Ross river virus disease in Brisbane, Australia. *The American Journal of Tropical Medicine and Hygiene* 2004; **71**: 129–137.
40. **Hu W, et al.** Mosquito species (Diptera: Culicidae) and the transmission of Ross River virus in Brisbane, Australia. *Journal of Medical Entomology* 2006; **43**: 375–381.
41. **Hu WB, et al.** Rainfall, mosquito density and the transmission of Ross River virus: a time-series forecasting model. *Ecological Modelling* 2006; **196**: 505–514.
42. **Tong S, et al.** Climatic, high tide and vector variables and the transmission of Ross River virus. *Internal Medicine Journal* 2005; **35**: 677–680.
43. **Mackenzie JS, et al.** Arboviruses causing human-disease in the Australasian zoogeographic region. *Archives of Virology* 1994; **136**: 447–467.
44. **Wright AE, et al.** A preliminary investigation of the ecology of arboviruses in the derby area of the Kimberley region, western Australia. *Australian Journal of Experimental Biology and Medical Science* 1981; **59**: 357–367.
45. **Shire of Derby Council.** Shire of Derby/West Kimberly: Shire of Derby (<http://www.sdwk.wa.gov.au/>). Accessed 5 May 2016.
46. **Jardine A, Neville PJ, Lindsay MD.** Proximity to mosquito breeding habitat and Ross River virus risk in the Peel region of Western Australia. *Vector-Borne and Zoonotic Diseases* 2015; **15**: 141–146.
47. **Lindsay M, et al.** An outbreak of Ross River virus disease in the south-west of Western Australia. *Communicable Diseases Intelligence* 1996; **20**: 136–139.
48. **Werner AK, et al.** Environmental drivers of Ross River virus in southeastern Tasmania, Australia: towards strengthening public health interventions. *Epidemiology and Infection*. 2012; **140**: 359–371.
49. **Gatton ML, Kay BH, Ryan PA.** Environmental predictors of Ross River virus disease outbreaks in Queensland, Australia. *The American Journal of Tropical Medicine and Hygiene* 2005; **72**: 792–799.
50. **Geelhoed E.** *The Economic Costs of Ross River Virus Infection*. Western Australia: Health Department of Western Australia, 1995.
51. **Ratnayake JTB.** The Valuation of Social and Economic Costs of Mosquito-transmitted Ross River Virus (Dissertation). Nathan, Queensland: Griffith University, 2005.
52. **Mullahy J.** Specification and testing of some modified count data models. *Journal of Econometrics* 1986; **33**: 341–365.