A climate-based model for malaria prediction in southeastern Africa

M.R. Jury†1 and A.D. Kanemba†

Malaria is a major health problem in southeastern Africa. In this study, we explore relationships between malaria and regional climate. Malaria incidence data from the eastern border of South Africa and Swaziland over a 33-year period were de-trended and used to identify epidemics. Composite weather maps were then constructed for seasons with high and low malaria incidence and evaluated. Surface-air temperature rose over the east coast during malaria epidemics and rainfall doubled over a large area including Swaziland, Zimbabwe and southern Mozambique. Remote climatic signals that anticipated malaria epidemics were found in composite analyses, and a statistical model was developed for prediction. Upper-level winds over the Western Pacific were found to predict 57% of malaria variance at a lead time of 6 months.

Introduction

Much of tropical Africa is prone to debilitating malaria epidemics; efforts to control the disease have had mixed success. In many regions where malaria was previously eradicated, the disease has made a comeback because of adaptation by the mosquito-borne parasite to drugs and insecticide. While sub-Saharan Africa carries the highest per capita burden of disease in the world, malaria is the single greatest cause of mortality. Of global deaths due to malaria, 90% occur in sub-Saharan Africa—one million annually, mostly children aged 2–6 years.2 In southeastern Africa, malaria is the cause of death of more than one in five young children.3 It places substantial demands on Africa’s fragile healthcare infrastructure. In malaria-endemic countries, as many as one-third of clinic visits and hospital admissions are for malaria. No country in Africa has shown a decline in malaria incidence since the 1980s.4 However, it should be noted that most health monitoring systems have difficulty in reliably tracking malaria incidence and mortality. The start of intensified mosquito control efforts in the 1990s coincided with rising malaria mortality due to drug resistance and HIV infection, meaning that without these efforts the situation might have been substantially worse.2

The length of the malaria transmission season is inversely related to the epidemic risk; thus, as the season gets shorter, the probability of epidemics increases. In Africa, the areas with a shorter transmission season tend to be located across the Sahel and, down through the Horn of Africa into east Africa and throughout southern Africa south of the Zambezi River.5–7 Estimates of the population at risk of malaria in Africa vary depending on the method and data used. Snow5 estimated 74 million in 1995, whereas a World Health Organization (WHO) study using United Nations population data for 2001 gave an estimate of 247.7 million, representing about 20% of the continent’s population.

Malaria epidemics are influenced by local and remote, climatic and non-climatic factors. Mosquito (vector) abundance and survival rate, the duration of parasite incubation, the density of infected human hosts and the coverage of wetlands determine the risk and distribution of malaria infection.8,9 Epidemics may be exacerbated by social upheaval, population migration and economic instability. A breakdown in healthcare systems and mosquito control activities allows malaria to recover for several years. The impact of malaria epidemics is greatest after a period of famine, when household income is depleted and communities are especially vulnerable.10

In semi-arid zones, malaria epidemics tend to follow warm, rainy spells that favour increased mosquito density and survival, and thus parasite transmission.5,11–13 Changes in local climate are themselves driven by remote, multi-year phenomena such as El Niño–Southern Oscillations (ENSO).14,15 Since malaria transmission is constrained by climate (see, for example, www.rbm.who.int/), epidemic forecasts over southeastern Africa are a possibility. In the study reported here, potentially useful predictors are revealed by consideration of temporal and spatial relationships between climate (cause) and malaria (effect). Thirty-three years of malaria data for South Africa and Swaziland give us a unique opportunity to examine these links.

Data and methods

The data used in this study derive from two sources. Monthly climate data were drawn from the U.S. National Center for Environmental Prediction (NCEP) website (www.cdc.noaa.gov/composites/). Annual malaria incidence data for South Africa and Swaziland over 33 years, and monthly malaria incidence data for Zimbabwe for 14 years, were provided by the South African Medical Research Council (MRC). These data are discussed by Craig et al.9

Meteorological data from NCEP are originally observed at the land surface, over the oceans by ships and in the upper air by balloons and aircraft. Satellites fill in the gaps by measuring widespread radiance patterns in a number of specific frequencies. All data were assimilated by a numerical weather forecast model that interpolated the observations to a ‘best fit’.1 The data used here comprise surface air temperature, wind speed, pressure, humidity and precipitation. Sea-surface temperature (SST) fields were analysed over the oceans,16 and upper air fields at 500-hPa (5 km) geopotential height and 200-hPa (12 km) wind were considered. Rainfall measurements from satellite and surface stations (CDC merged analysis precipitation) start in 1979, whereas the study considers 33 years of health data from 1970 to 2002. Rainfall estimates from the NCEP model were therefore included so as to fill the 1970–1978 gap. The two data sets closely correspond.

Malaria incidence data were from the MRC for eastern South Africa and Swaziland for the period from 1970 to 2002. The system of data collection is stable, relatively accurate and recorded by municipal district. The data were found to be reliable in earlier studies.1 More general information on malaria comes from the WHO. Data were requested from the respective health departments in countries of the region; however, most time series were found to be short or intermittent.

To create the malaria time series, the annual number of cases of malaria was extracted to represent the environmentally sensitive component.
We employed a ‘targeted’ approach to reveal the climatic conditions that related to the malaria time series. High and low seasons of malaria incidence were identified, based on the combined de-trended malaria index. High years were taken to be 1978, 1985, 1988, 1989, 1999 and 2000; the corresponding low incidence years were 1973, 1982, 1992, 1993, 1995 and 2002 (years were labelled according to the end of the malaria season: January–April). Our search for climatic influences on malaria began with the mapping of average climate fields. Composite difference maps were calculated by subtracting means of atmospheric field data for the six ‘low’ seasons from the six ‘high’ seasons. The resultant maps point to key areas where significant differences occurred over large areas of about 1000 × 1000 km. Next, the time series of the field variable in the key area was extracted and the composite seasonal trends were plotted to see if ‘high’ and ‘low’ signals were maintained in a stable fashion at increasing lead times (for example, from the January–April malaria season back in time to the preceding July–September ‘predictor’ season). Temporal instabilities or a lack of coherence between cases caused the predictor to be rejected.

This method is a simple step-wise screening process as shown in Fig. 1. Input variables derive from composite differences with respect to the malaria season, whereas multivariate linear regression yields a hindcast fit that requires validation. A useful statistical tool is wavelet analysis to determine the degree to which cycles in the malaria, rainfall and climatic predictors reflect or lag each other. Continuous wavelet transform (CWT) was also used to filter the monthly time series to isolate particular oscillations. Here we filter the data to retain cycles with periods from 1.5 to 16 years.

Results

Temporal trends

The annual cycle was considered using monthly mean malaria cases and local climatic data averaged over the region (16–23°S, 22–35°E). Malaria incidence rose from December to March, followed by a sharp decline. The winter and spring months, June to October, exhibit a negligible number of cases. Surface temperatures <14°C and wind speeds >4 m s–1 hinder mosquito survival and incubation of the parasite then. While rainfall and temperature peak in summer (December to March), wind is minimal in autumn (March to May). Area-averaged rainfall in months with ‘high’ malaria cases reached 80 mm, whereas in ‘low’ months it was nearer 50 mm. The human pool of infection takes time to build each year. Hence, the malaria peak lags rainfall by about two months (Fig. 2). The de-trended malaria incidence time series that is the basis for subsequent analysis of climatic signals is illustrated in Fig. 3. It shows an alternation of low and high periods of malaria incidence with a cycle of 6–10 years. An epidemic typically occurs at the beginning of a cycle, followed by a slow downward trend to a minimum, followed by a sudden upswing. Although the malaria data come from a confined area, an analysis of regional climatic parameters for high and low malaria years indicates that a much wider area is influenced (for instance, 16–23°S, 22–35°E). Rainfall displays a large signal-to-noise ratio, whereas other local climatic elements show rather insignificant changes.

Composite spatial patterns: local impacts

The composite difference pattern for rainfall is shown in Fig. 4. There is almost twice as much rainfall over Zimbabwe and southern Mozambique in a high season than in a low season. The pattern aligns to the south of the Zambezi River valley and covers an area greater than 1000 × 1000 km. Of interest is a region of below-normal rainfall over Madagascar of reduced intensity. Both areas of above- and below-normal rainfall exhibit a north–south orientation.

Higher surface temperatures (Fig. 5a) and northeasterly winds were found over the east coast and Mozambique Channel with respect to high malaria seasons (Fig. 5b). The onshore flux of tropical air is drawn into a cyclonic low pressure cell situated over southern Mozambique, Zimbabwe and northeast South Africa. The warm moist air rises over the eastern escarpment, cools and precipitates. When a season is dominated by such a circulation, communities in southeastern Africa experience an increase in malaria. The lack of a temperature signal over the malaria impact zone (warmer to the east and cooler to the west) may explain why some researchers fail to find the expected relationship.

Remote effects: the oceans

When searching for remote influences on African climate, oceanic signals have proved useful due to atmospheric coupling that shifts the rainfall regimes and thus malaria incidence. Mapping the composite SST differences before the season, we found significant areas in the east and southwest Atlantic and Central Pacific. We also considered sea-level pressure (SLP) and
found significant differences before the season over the South Atlantic and South Pacific. Following the designation of these signals in key areas, we considered the composite time series to assess stability in the previous year, and the wavelet filtered time series for spectral (cyclical) matching. From this work, five potential predictors were identified: East Atlantic SST, South-West Atlantic SST, equatorial Pacific SST, New Zealand SLP, and South Atlantic SLP.

Warm SST over the equatorial east Atlantic (Fig. 6a) are driven by a zonal overturning circulation. This SST index acts as a stable indicator in the precursor year (Fig. 6b). Warming of the Atlantic Ocean near Angola and associated westerly winds anticipate a Pacific La Niña. A relaxation of the St Helena anticyclone as a result of changes in SLP over the South Atlantic, may help shift rainy weather from Madagascar to southeastern Africa.

Remote effects: the upper atmosphere

The westerly jet stream is located at the boundary between the warm tropics and the cold polar air on either side of the equator near 40° latitude. At times, the jet stream develops bulges known as standing waves, in response to heating anomalies in the tropics. These waves can channel areas of high rainfall toward certain longitudes and may thus have an influence on malaria. To reveal the influence of the upper circulation, we analysed the 200-hPa (~12 km) wind and 500-hPa (~5 km) geopotential height differences.

Upper-level winds that flow east–west near the equator are indicative of the Walker circulation and associated ENSO phase. Winds over the West Pacific (5°N–10°S, 120°–170°E) ‘connect’ with the circulation of the Indian Ocean. In our malaria composite map (Fig. 7), the upper wind is from east (west) over the Atlantic (Pacific), hence opposed prior to high malaria seasons. This upper level circulation induces wet (dry) weather over Africa (South America). An asymmetry of geopotential height over the poles indicates a shift in the polar jet axis away from the SW Pacific towards the South Atlantic prior to years of high malaria incidence (Fig. 8). The southern hemisphere circulation therefore provides a further indication of changes in malaria risk. Seven predictors were derived from composite mapping of the upper air and include: 500-hPa geopotential height and 200-hPa zonal wind speed from the West Pacific, Australia, New Zealand, Central and South Atlantic. Added to these are the SST predictors.

The filtered time series for two candidate predictors, zonal winds over the Pacific and Atlantic, reveal significant covariance with rainfall over southeastern Africa (Figs 9 and 10). Both exhibit cyclical matching in the ranges 3–4 years and 6–10 years. While the West Pacific and equatorial Atlantic zonal winds are tropical and may be expected to exhibit ‘memory’, other variables derive from the southern mid-latitudes and may be ‘noisy’.

To test this, the signal-to-noise ratio was computed (Fig. 11, e.g. ‘high season’ and ‘low season’ divided by the respective means). The equatorial Atlantic SST demonstrated the largest difference in high seasons (>20%), whereas the West Pacific zonal wind exhibited significant differences (>15%) in both high and low seasons. The high latitude predictors show changes that are
>10%, but some are biased to either low or high years and may fail to indicate both.

**Model development**

The composite mapping and time series analysis uncovered numerous potential predictors that could be used to develop a statistical model. Predictors from the preceding July–September season would provide the necessary advance warning for mitigating action. As the malaria data were manipulated by de-trending, we also considered area-averaged rainfall as a less manipulated proxy for malaria. Statistical inference is an important step in data analysis and model building. A principle applied here was to use the minimum number of predictors to reduce artificial skill from co-linearity. The form of the multivariate linear regression model was assumed to be 

$$Y = a + bx_1 + cx_2,$$

where $Y$ is a dependent variable; $a$, $b$ and $c$ are constants, and $x_1$ and $x_2$ are predictors.

Of the possible predictors arising from composite mapping, 10 were deemed suitable for model development. Another seven exhibited unstable behaviour, poor spectral matching, or low signal-to-noise ratios and were rejected. The predictors retained are listed in Table 1. Using step-wise regression in a standard statistical package, predictive equations were developed for the period 1970 to 2002:

- **Malaria ($Y_1$):**
  $$Y_1 = 0.43(W_{Pac\_ZoWin}) - 0.50(S_{Au\_Zo})$$
  $$r^2 = 0.55$$

- **Rainfall ($Y_2$):**
  $$Y_2 = 0.67(W_{Pac\_ZoWin})$$
  $$r^2 = 0.59$$

Out of the predictor pool, the West Pacific and Australia zonal wind gave the best fit (Fig. 12), an improvement over previous attempts that used Atlantic SST. A contingency table indicates useful tercile forecasts (Table 2) in all cases except for a tendency to predict normal seasons as high (false alarm).

The reliability of a forecast model was assessed by comparing predictions against independent observations. Forecast errors can be calculated, and model ‘drift’ detected. The most recent observations can be used to update the forecasts. Historically based models assume replication, but future conditions will change in space and time. A climatic feature that dominated during a past decade may be less influential in future. To test for this, the first and last decades in the 33-year record were removed and the model re-fitted. In the earlier period, 1970 to 1991, there is a stronger dependence on predictors from the Atlantic and a lower degree of fit ($r^2 = 0.28$). This may be due to less accurate climate data before 1980 or to a shift of climate regime.

**Discussion and conclusion**

Recent improvements in the availability of climate and health data and increased use of geographical information systems and remote sensing make a climate-based warning system increasingly feasible from a technical point of view. Our research sought to filter out the non-climatic factors affecting malaria incidence through creation of a de-trended record. This essential step was taken prior to searching for local impact or remote predictive associations. We thereby demonstrated that local climate anomalies affect malaria incidence in southeastern Africa. This contrasts with the work of Craig et al., who found no relationship between climate and malaria data. Their work provides useful short-term predictions based on persistence.

We have indicated the value of using indices of the upper atmospheric circulation as predictors, in addition to tropical SST. The analytical steps involved in quantifying the climate–malaria link included: 1) De-trending malaria data to remove increases associated with drug resistance. From the stable record, cyclicity was evident and high and low malaria years were identified. 2) Searching for climate predictors from global ocean and atmospheric data sets; followed by the testing of predictors for signal-to-noise ratio, phase lag and co-variance with respect to de-trended malaria and area-averaged rainfall. 3) Developing
model equations and performing independent validation tests. Model accuracy was found to be such that over half the variance of malaria incidence and its proxy could be explained using only one or two variables at a six-month lead time.

Shifts in tropical rainfall are influenced by ‘meandering’ of the jet stream in the higher latitudes. Composite upper winds for high minus low malaria seasons revealed a pattern of alternating northward and southward flow every 3500 km around the southern hemisphere (wave 5). This adjustment by the southern jet stream to ENSO (starting in the west Pacific) provides a basis for malaria forecasts in southeastern Africa. Changes in sea temperatures and winds over the Atlantic were similarly related to the ENSO signal, and may be seen as a downstream continuation of the same forcing.

The northeasterly winds around Madagascar are part of the Indian Ocean Walker circulation. In seasons with high malaria incidence, they transfer warm moist air from Madagascar towards the east coast of southern Africa. These changes often bring widespread rain over southeastern Africa, the most recent example being the floods of 2000, when malaria incidence reached its highest level ever.

Of several possible needs in health control, the strengthening of systems for reporting the incidence of epidemic diseases is seen as a high priority. More commonly, disease–climate modelling has been restricted to short data sets representing small areas. Here, the South African/Swaziland data point to a much more extensive signal affecting Africa southeast of the Zambezi River valley. The implication because of a shortage of suitable epidemiological data. More commonly, disease–climate modelling has been restricted to short data sets representing small areas. Here, the South African/Swaziland data point to a much more extensive signal affecting Africa southeast of the Zambezi River valley. The implication

### Table 1. Details of candidate predictors considered in regression modelling.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Abbreviation</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>West Pacific 200-hPa zonal wind</td>
<td>W_Pac_ZoWin</td>
<td>4°N–8°S, 130–160°E</td>
</tr>
<tr>
<td>Surface pressure New Zealand</td>
<td>SLP_New</td>
<td>37–50°S, 160°E–170°W</td>
</tr>
<tr>
<td>New Zealand 500-hPa geopotential</td>
<td>Ge_New</td>
<td>35–50°S, 160°E–170°W</td>
</tr>
<tr>
<td>Central Pacific SST</td>
<td>Pac_SST</td>
<td>5°N–5°S, 170–130°W</td>
</tr>
<tr>
<td>South Atlantic 200-hPa zonal wind</td>
<td>S_Alt_Zo</td>
<td>22–32°S, 40–20°W</td>
</tr>
<tr>
<td>South of New Zealand 200-hPa zonal wind</td>
<td>S_Ne_Zo</td>
<td>55–65°S, 140–180°E</td>
</tr>
<tr>
<td>South Australia 200-hPa zonal wind</td>
<td>S_Au_Zo</td>
<td>30–40°S, 130–150°E</td>
</tr>
<tr>
<td>Equatorial Atlantic 200-hPa zonal wind</td>
<td>Eq_Alt_Zo</td>
<td>2–12°S, 50–30°W</td>
</tr>
<tr>
<td>Equatorial East Atlantic SST</td>
<td>Eq_Alt_SST</td>
<td>2°N–12°S, 2°W–12°E</td>
</tr>
<tr>
<td>South Atlantic SST</td>
<td>S_Alt_SST</td>
<td>30–45°S, 40–20°W</td>
</tr>
<tr>
<td>South Atlantic 500-hPa geopotential</td>
<td>S_Alt_Ge</td>
<td>32–45°S, 50–10°W</td>
</tr>
</tbody>
</table>

### Table 2. Contingency table for model predicted and de-trended (observed) malaria incidence in tercile categories.

<table>
<thead>
<tr>
<th>Pred</th>
<th>Pred_Low</th>
<th>Pred_Norm</th>
<th>Pred_High</th>
<th>Obs_Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs_Low</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Obs_Norm</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Obs_High</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Pred_Total</td>
<td>10</td>
<td>3</td>
<td>17</td>
<td>30</td>
</tr>
</tbody>
</table>

Pred, predicted; Obs, observed.
is that our model should be tested against health data from
countries sharing the same climate signal (for example, Zimba-
bwe and Mozambique). In some cases, viable surveillance
systems may exist but require modification to ensure timely
transfer of data from the point of collection to the point of
analysis. For diseases such as malaria, which often are diagnosed
clinically, further work needs to be carried out to determine the
extent to which the quality of diagnosis affects our ability to
recognize (and predict) epidemics.

We thank the Medical Research Council’s Malaria Research Programme for
funding, with sponsorship also from the South African Department of Environ-
mental Affairs and Tourism, and the NRF.

Fig. 10. Interannual filtered time series for rainfall (red) and predictor (blue). The
co-variance spectra and lag are analysed in the middle and lower panels. Refer to
Table 1 for predictor locations.

Fig. 11. Ratio (%) of changes for candidate predictors, comparing their composite
responses in high and low malaria incidence years to the long-term mean. Refer to
Table 1 for predictor details.

Fig. 12. De-trended malaria observations (continuous line) and model fit at
six-month lead time (dashed line).

Received 14 October 2005. Accepted 6 February 2007.

amr2003/amr_toc.htm
4. Byass P. et al. (2002). The role of demographic surveillance systems in assessing
the health of communities: an example from rural Ethiopia. Public Health 116,
145–150.
Pasteur 86, 97–130.
Exploring 30 years of malaria case data in KZ-Natal South Africa: Part 1. The
Med. Parasitol. 45, 74–79.
6, 55–62.
model of malaria transmission in sub-Saharan Africa. Parasitology Today 15,
105–111.
Met. Soc. 77, 437–471.
analysis using optimum interpolation. J. Climate 7, 929–948.
19. Leovisinhi M.E. (1994). Climate warming and increased malaria incidence in
Rwanda. The Lancet 343, 714–718.
on malaria transmission in Africa. The Lancet 362, 1792–1798.
severe malaria years in Zimbabwe, not South Africa. S. Afr. J. Sci. 100, 262.
equatorial and meridional modes of climatic variability in the tropical Atlantic.
and Indo-Pacific ENSO: statistical analysis and numerical simulation. Global