

LETTERS

Malaria early warnings based on seasonal climate forecasts from multi-model ensembles

M. C. Thomson¹, F. J. Doblas-Reyes², S. J. Mason¹, R. Hagedorn², S. J. Connor¹, T. Phindela³, A. P. Morse⁴ & T. N. Palmer²

The control of epidemic malaria is a priority for the international health community and specific targets for the early detection and effective control of epidemics have been agreed¹. Interannual climate variability is an important determinant of epidemics in parts of Africa² where climate drives both mosquito vector dynamics and parasite development rates³. Hence, skilful seasonal climate forecasts may provide early warning of changes of risk in epidemic-prone regions. Here we discuss the development of a system to forecast probabilities of anomalously high and low malaria incidence with dynamically based, seasonal-timescale, multi-model ensemble predictions of climate, using leading global coupled ocean–atmosphere climate models developed in Europe. This forecast system is successfully applied to the prediction of malaria risk in Botswana, where links between malaria and climate variability are well established⁴, adding up to four months lead time over malaria warnings issued with observed precipitation and having a comparably high level of probabilistic prediction skill. In years in which the forecast probability distribution is different from that of climatology, malaria decision-makers can use this information for improved resource allocation.

Malaria is the most important parasitic infection in people, accounting for an estimated 500 million clinical attacks worldwide and more than 1 million deaths a year, mostly in sub-Saharan Africa⁵. On this continent, epidemic malaria is associated with seasonally warm semi-arid and highland areas where approximately 124 million people are considered at risk of climate-related malaria epidemics⁶. For improved control in epidemic regions, the World Health Organisation (WHO) advocates the development of integrated malaria early warning systems based on vulnerability assessment, seasonal climate forecasts, weather and environmental monitoring and case surveillance^{2,7}.

In the semi-arid country of Botswana, malaria is ranked as one of the major public health problems and consequently there is a well developed National Malaria Control Programme that supports a set of key control activities: (1) vector control through annual indoor residual house spraying in endemic areas and in response to epidemics, (2) chemoprophylaxis for target groups during the malaria season, and (3) prompt and effective case management including drug therapy to affected individuals. After the devastating regional malaria epidemic of 1996, malaria control was reinvigorated through a range of initiatives including the development of an early warning system for epidemics, which peak in March and April following the bulk of the rainy season (November–February, NDJF)⁴. As a notifiable disease, malaria cases are reported routinely to the Ministry of Health and epidemic alerts are based on whether weekly numbers of cases exceed pre-defined thresholds. An earlier study⁴ has indicated that December–February seasonally averaged monitored rainfall and

also sea surface temperature (SST) for the same time period provide significant predictive skill for the malaria season one month in advance of its seasonal peak. Using data on yearly malaria incidence in Botswana, we discuss whether the lead time for providing reliable early warnings using observed precipitation data (for improved targeting of insecticides, drug stocks and other control measures) can be increased significantly using skilful seasonal climate forecasts.

The results in this Letter are based on re-forecasts (retrospective forecasts) using models that form the basis of a future real-time operational multi-model ensemble forecast system developed out of the original DEMETER (Development of a European Multi-model Ensemble Forecast System for Seasonal to Interannual Climate Prediction) project⁸. DEMETER used a multi-model ensemble forecast system for seasonal-to-interannual climate variability, comprising state-of-the-art coupled ocean–atmosphere global climate models developed in Europe.

An ensemble climate forecast system predicts not only the most likely evolution of climate, but also the uncertainty in such a prediction. More generally, an ensemble climate forecast system

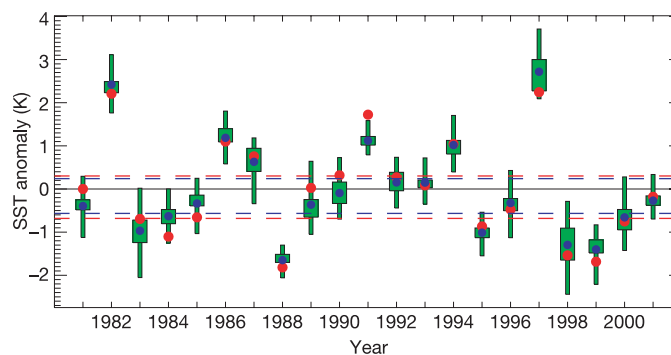


Figure 1 | Time series of the one-month lead December–February predicted SST averaged over the tropical Pacific region Niño3.4. The boundaries of Niño3.4 are 5° N–5° S and 170° W–120° W. The multi-model ensemble spread is depicted by the green box-and-whisker representation, the whiskers containing one-third of the ensemble members. The blue dots represent the ensemble mean and the larger red bullets show the ERA-40 anomalies. The horizontal dotted lines indicate the ERA-40 (red) and multi-model re-forecast (blue) climatological tercile boundaries. The ensemble-mean correlation with the reference is 0.97, while the ROC score (see text) for the three tercile categories (above normal, normal and below normal) ranges between 0.97 and 1.00. By comparison, the ROC scores for SST predictions based on the persistence of the August–October ERA-40 anomalies take values of the order of 0.85.

¹International Research Institute for Climate and Society (IRI), The Earth Institute at Columbia University, New York, New York 10964-8000, USA. ²European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, RG2 9AX, Reading, UK. ³National Malaria Control Programme, Epidemiology and Disease Control Unit, Ministry of Health, Private Bag 00269, Gaborone, Botswana. ⁴Department of Geography, University of Liverpool, Liverpool, L69 7ZT, UK.

Table 1 | ROC areas for annual standardized log malaria incidence predictions

Event	CMAP	DEMETER
Very high	0.93 (0.70–1.00)	0.80 (0.59–0.96)
High	0.36 (0.06–0.70)	0.48 (0.18–0.79)
Low	0.33 (0.11–0.58)	0.34 (0.12–0.60)
Very low	0.91 (0.73–1.00)	0.83 (0.61–0.99)

The predictions are carried out using observed CMAP and DEMETER multi-model ensemble precipitation over NDJF and have similar forecast quality. The four events for values are described in the Methods section. CMAP predictions are assessed as probabilistic forecasts with fixed, but unknown uncertainty²⁴. The 95% confidence limits obtained by bootstrapping with a sample size of 10,000 are shown in parentheses. CMAP-based predictions are available at the beginning of March, but the DEMETER multi-model predictions would be available to users four months earlier, at the beginning of November, a potentially crucial increase in lead time for the management of malaria epidemics.

predicts a probability distribution of climate. The uncertainties in climate prediction that generate such probability distributions arise from imprecise initial conditions, and numerical approximations to the underlying partial differential equations that govern climate⁹. In the DEMETER ensemble forecast system, model uncertainty is represented by the fact that the sub-grid parameterizations in the component models have, to a large extent, been developed independently. Here we analyse data from six-month integrations started on 1 November of every year (see Methods).

The physical basis for seasonal climate prediction lies, to great

extent, in the predictability of slowly varying interactions between the atmosphere and the oceans. The prototypical phenomenon illustrating such predictability is El Niño/Southern Oscillation (ENSO) which has been shown to be predictable on seasonal time-scales^{10,11}. It is known that interannual variability in the strength of ENSO can strongly influence the climate and maize yields of southern Africa¹² and is also a predictor for unusually high or low malaria anomaly years in Botswana⁴. To illustrate the ability of the DEMETER system to forecast atmosphere–ocean interactions, Fig. 1 shows probabilistic predictions of the SST during December–February averaged over the region in the tropical Pacific known as Niño3.4. These probabilistic forecasts are the most skilful dynamically based predictions reported to date, and substantially outperform deterministic predictions based on persistence of the SST anomalies of the month previous to the dynamical forecast start date¹³.

We now focus on the DEMETER prediction of the climate of southern Africa, a necessary condition for the accurate prediction of malaria incidence. Cumulative precipitation anomalies for NDJF using CMAP-estimated precipitation data (see Methods) are shown in Fig. 2a and b composited for the five highest and lowest malaria incidence anomaly years of the data sample (1982–2002). These figures illustrate that higher than expected malaria years are associated with above-average precipitation, while the lowest malaria years are associated with below-average precipitation. In addition, Fig. 2c

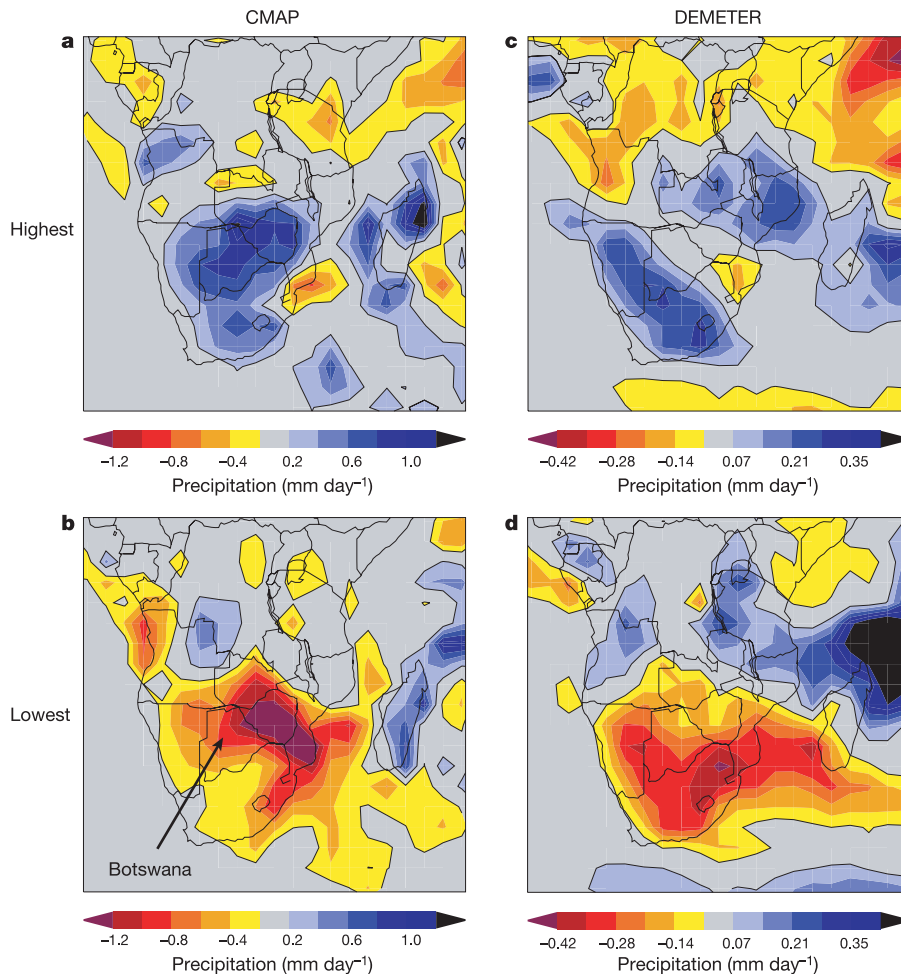


Figure 2 | Composites of austral summer precipitation for central and southern Africa as a function of the standardized malaria annual incidence for Botswana. Mean precipitation anomalies for the five years with the highest (a, c) and lowest (b, d) malaria annual incidence in Botswana for November–February CMAP (a, b) and DEMETER (c, d) ensemble-mean

precipitation. A different colour scale was used for the observed (CMAP) and re-forecast (DEMETER) precipitation because of the unavoidable reduction in amplitude of the predicted anomalies when averaging the 27 ensemble members.

and d shows that during the high and low malaria years, the one-to-four-month NDJF ensemble-mean precipitation from the DEMETER forecasts (available at the beginning of November) are above and below average over much of southern Africa, respectively.

The association between malaria incidence and precipitation depicted in Fig. 2a and b can be made more precise if we use a quadratic relationship similar to that described in an earlier study⁴ between seasonal precipitation averaged over Botswana and the logarithm of standardized (detrended) confirmed malaria incidence (per 1,000 of population), as illustrated in Fig. 3. This nonlinear relationship provides the quantitative link that allows the DEMETER probability forecasts of precipitation to be transformed into probability forecasts of malaria incidence. The quadratic relationship can also be used to issue malaria warnings based on monitored precipitation⁴, such as NDJF CMAP precipitation, although this would be available only at the beginning of March, four months later than the multi-model forecast issue date. As an illustration, Fig. 4 shows DEMETER forecast probability distributions of malaria incidence associated with a year of very high malaria incidence, and another year of very low malaria incidence. For decision-makers, that these probability distributions are not sharp quantifies the inevitable uncertainties that exist in forecasting a variable that is strongly influenced by partially chaotic climatic processes. On the other hand, forecast probability distributions for the high malaria years are distinctly different from the probability distributions for the low malaria years: the null-hypothesis that the distributions for DEMETER-based predictions in the two categories are not distinct can be rejected using a Kolmogorov–Smirnov test with a *P* value smaller than 0.0001.

The quality of these forecast probability distributions of malaria incidence has been assessed using the area under the Receiver Operating Characteristic (ROC) curve¹⁴ for the events ‘values higher than the top quartile’ (very high), ‘between the top quartile and the median’ (high), ‘between the median and the lowest quartile’ (low) and ‘lower than the lowest quartile’ (very low). The ROC score is an appropriate measure for assessing the value of probability climate forecasts for decision-making^{14,15}. Suppose it is decided to take precautionary action (targeting drug or insecticide supplies to a particular region) if the probability of malaria in the highest-quartile category exceeds some chosen threshold probability. Over the full forecast data sample, the number of correct and incorrect decisions

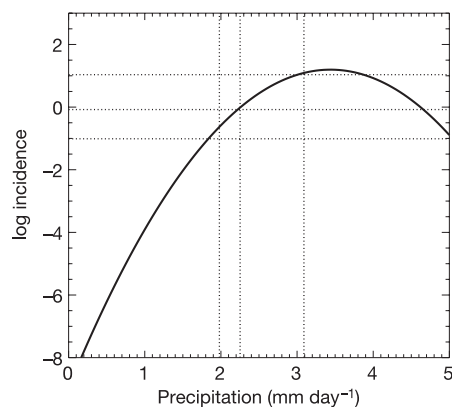


Figure 3 | Relationship between standardized log malaria annual incidence and summer precipitation for Botswana. Standardized log malaria annual incidence (per 1,000 population) versus November–February CMAP precipitation for the period 1982–2002 as obtained from fitting a quadratic function to observations⁴. The quadratic function is $y = -8.9 + 5.9x - 0.8x^2$, where x denotes precipitation (measured in mm day^{-1}) and y the standardized log malaria incidence index. The horizontal and vertical dotted lines denote the quartiles of the standardized log malaria incidence index, and precipitation, respectively.

can be assessed for this particular threshold probability: these are referred to in terms of ‘hit rate’ and ‘false-alarm rate’, respectively. The ROC score is a measure of the mean hit rate to false-alarm rate, averaging over all threshold probabilities. A ROC score larger than 0.5 indicates a forecast system more skilful than a forecast based on climatological probabilities; a ROC score of 1.0 indicates a perfect forecast system.

Table 1 shows the ROC scores for predictions of standardized malaria incidence, first using the CMAP precipitation and second with the DEMETER multi-model ensemble forecasts initialized on 1 November. For forecasts of both very low and very high malaria incidence, the multi-model DEMETER forecasts have significantly positive ROC scores. The DEMETER scores are slightly lower than the CMAP scores, but we point out that the DEMETER forecasts are available at least four months before the CMAP data becomes available in early March.

Our evidence supports the use of multi-model ensemble climate predictions initialized at least five months before the peak malaria season in Botswana and four months earlier than a prediction based on monitored precipitation for early warning of epidemic malaria risks. The significant gain in lead time with small reduction in prediction skill suggests that probabilistic seasonal climate forecasts can be used to predict malaria incidence, not only in Botswana, but also in neighbouring epidemic-prone areas, as shown in Fig. 2. These results are not only of importance to the malaria control managers in the region who are actively involved in using climate information to achieve malaria-reduction targets¹⁶, but are also relevant to other resource managers (health, hydrology, agriculture, and so on) faced with climate-sensitive decisions. Unlike the products that would be derived from single forecast models, the malaria incidence forecasts described here allow the user to obtain a probability distribution of disease risk, thus indicating to the user the uncertainty of the information available in the forecast and the extent to which it differs from climatology.

Although the greatest burden of malaria in Africa is suffered by those living in endemic regions, epidemics pose a serious threat to many millions of people⁶ and their prevention remains a priority¹. The ability of a malaria early-warning system to improve resource allocation and assist in the reduction of malaria morbidity and mortality depend on decision-makers’ capacity to make effective use of new information within their own control paradigm. To this end we have focused our efforts on the integration of the multi-model DEMETER forecast system into the routine epidemic malaria control activities currently promoted by the WHO-AFRO Southern Africa Inter-Country Malaria Team (SAMC) in Zimbabwe and undertaken by the National Malaria Control Unit in Botswana. It

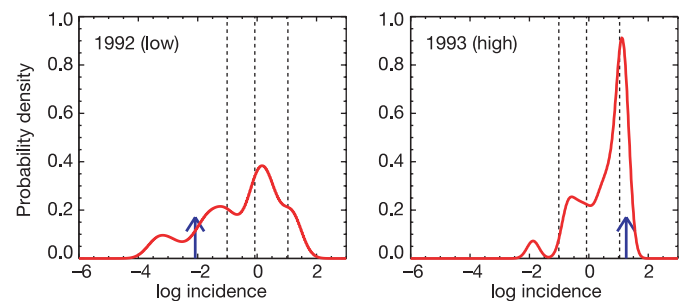


Figure 4 | Forecast probability distribution function of standardized log malaria annual incidence for Botswana. The probability distribution functions of predicted standardized log malaria annual incidence for the years 1992 (anomalously low incidence, left) and 1993 (anomalously high incidence, right) computed with the DEMETER multi-model ensemble forecast system are depicted in red. The vertical dashed lines denote the quartiles of the distribution of the standardized log malaria incidence index and the vertical blue arrows indicate the values recorded by the Botswana Ministry of Health.

is here that the final arbiters of the utility of the forecasts will be found.

METHODS

Malaria incidence data. Botswana has compiled national annual data on cases of laboratory-confirmed malaria incidence for the period 1982–2002. This data set offers a unique opportunity to analyse and predict malaria incidence in a desert-fringe area. Confirmed incidence per 1,000 population was modelled⁴ to remove non-climate trends, largely attributed to chloroquine resistance¹⁷ and the impact of a major policy intervention, which included the change of the first-line drug. The residual, a standardized log malaria incidence index, is the variable used in this study. Epidemiological and population data were obtained from the Ministry of Health's Epidemiology and Disease Control Unit, and the Central Statistics Office in the capital city, Gaborone.

Precipitation data. Precipitation in Botswana is concentrated in the period November to March and is subject to high interannual variability¹⁸. For this analysis, the 2.5°-resolution monthly data from the Climate Prediction Center Merged Analysis of Precipitation (CMAP)¹⁹ were averaged across the 20 grid points between 17.5°–27.5° S and 17.5°–30.0° E for the rainy season before the peak malaria season (NDJF) from 1979 to date. Calendar years for precipitation are referred to by the year of the January used in the seasonal average. The best relationship between seasonally averaged precipitation is described by the quadratic model depicted in Fig. 3 (ref. 4).

DEMETER climate predictions. The DEMETER multi-model ensemble system for climate prediction comprises seven global coupled ocean–atmosphere models^{8,20}. Re-forecasts have been produced over the period 1958–2001, although the period common to the seven models is 1980–2001. The three models used in this study comprise the operational DEMETER system and are from the following institutions: the ECMWF (European Centre for Medium-Range Weather Forecasts), Centre National de Recherches Météorologiques (Météo-France, France), and The Met Office (UK). The corresponding re-forecasts have a common period of 1959–2001. The uncertainty in the initial conditions derives from the fact that each DEMETER model is integrated from nine different initial conditions, each with different estimates of initial-condition uncertainty. Atmospheric and land-surface initial conditions are taken directly from the ECMWF re-analysis ERA-40 (ref. 21). The nine-member ensemble re-forecasts from each model were integrated for 180 days. Because of the computational cost of running multi-model ensembles, DEMETER re-forecast start dates were quarterly rather than monthly: 1 February, 1 May, 1 August and 1 November. More detailed information on the models and the initialization procedures can be found on <http://www.ecmwf.int/research/demeter/general/docmodel/index.html>. According to the set of available start dates, predictions for NDJF (months one to four of the simulations) were used. DEMETER multi-model ensemble predictions were averaged over Botswana in the same manner as the CMAP data.

Climate prediction calibration. Given the typical systematic errors of coupled model simulations, each prediction is corrected to have statistical properties similar to those of the observed precipitation. To allow robust estimates of the correction, the CMAP precipitation time series for Botswana starting in 1979 has been extended back to 1959 using the Climate Research Unit of the University of East Anglia (CRU) precipitation data^{22,23}, which is highly correlated with CMAP. This also limits the number of coupled models available from DEMETER to the three mentioned above. A scheme that mimics an operational prediction system has been used to correct the precipitation ensemble predictions. Mean and variance estimates for a specific year are obtained for each single model and for CMAP/CRU for the period from 1960 to the year before the target year. The difference in means and ratio of variances between each model and the reference is then computed and applied to the predictions of the target year. The same procedure is used for subsequent years. This scheme makes available each year a longer sample to compare the CMAP/CRU data with each single model. The corrected DEMETER-predicted precipitation was used to create ensemble predictions of standardized log malaria incidence for the period 1982–2002. The quadratic relationship between CMAP precipitation and malaria incidence is applied to the corrected predicted precipitation of each ensemble member to obtain the malaria incidence predictions.

Formulation of probability forecasts. Probabilistic predictions have been formulated for four different categories defined using quartiles: very high, high, low, and very low (see text). The most useful for assessing malaria epidemic risk are the very low and high categories, defined by the values below and above the first and third quartile, respectively. Probabilities have been estimated as the relative number of members of a given ensemble of predicted standardized log malaria incidence falling within a given category, a simple method that assigns the same weight to every single model.

Received 6 May; accepted 1 December 2005.

1. WHO/UNICEF. *The World Malaria Report*. 1–120, Report no. WHO/HTM/MAL/2005.1102 (WHO/UNICEF, Geneva, 2005); (<http://rbm.who.int/wmr2005/index.html>).
2. Najera, J. A., Kouznetzov, R. L. & Delacollette, C. *Malaria Epidemics: Detection And Control, Forecasting And Prevention*. 1–81, Report no. WHO/MAL/98.1084 (WHO, Geneva, 1998); (http://www.who.int/malaria/docs/najera_epidemics/naj_toc.htm).
3. Thomson, M. C., Connor, S. J., Milligan, P. J. M. & Flasse, S. P. The ecology of malaria—as seen from Earth-observation satellites. *Ann. Trop. Med. Parasitol.* **90**, 243–264 (1996).
4. Thomson, M. C., Connor, S. J., Phindela, T. & Mason, S. J. Rainfall and sea-surface temperature monitoring for malaria early warning in Botswana. *Am. J. Trop. Med. Hyg.* **73**, 214–221 (2005).
5. Greenwood, B. M., Bojang, K., Whitty, C. & Targett, G. A. T. Malaria. *Lancet* **365**, 1487–1498 (2005).
6. Worrall, E., Rietveld, A. & Delacollette, C. The burden of malaria epidemics and cost-effectiveness of interventions in epidemic situations in Africa. *Am. J. Trop. Med. Hyg.* **71** (Suppl. 2), 136–140 (2004).
7. WHO. *Malaria Early Warning Systems, Concepts, Indicators and Partners. A framework for Field Research in Africa*. 1–84, Report no. WHO/CDS/RBM/2001.32 (WHO/Roll Back Malaria/Technical Support Network for Prevention and Control of Malaria, Geneva, 2001); (http://mosquito.who.int/cmc_upload/0/000/014/807/mews2.pdf).
8. Palmer, T. N. *et al.* Development of a European ensemble system for seasonal to inter-annual prediction. *Bull. Am. Meteorol. Soc.* **85**, 853–872 (2004).
9. Palmer, T. N. *et al.* Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.* **33**, 163–193 (2005).
10. Goddard, L. *et al.* Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.* **21**, 1111–1152 (2001).
11. Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S. & Balmaseda, M. A. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature* **392**, 370–373 (1998).
12. Cane, M., Eshel, G. & Buckland, R. Forecasting Zimbabwean maize yield using eastern equatorial Pacific sea surface temperature. *Nature* **16**, 3059–3071 (1994).
13. Coelho, C. A. S., Pezzulli, S., Balmaseda, M., Doblas-Reyes, F. J. & Stephenson, D. B. Forecast calibration and combination: A simple Bayesian approach for ENSO. *J. Clim.* **17**, 1504–1516 (2004).
14. Jolliffe, I. T. & Stephenson, D. B. In *A Practitioner's Guide in Atmospheric Science* (eds Jolliffe, I. T. & Stephenson, D. B.) 1–12 (I.T. Wiley, Chichester, 2003).
15. Richardson, D. S. Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J.R. Meteorol. Soc.* **126**, 649–667 (2000).
16. DaSilva, J. *et al.* Improving epidemic malaria planning, preparedness and response in southern Africa. *Malaria J.* **3**, doi:10.1186/1475-2875-3-37 (2004).
17. Brinkmann, U. & Brinkmann, A. Malaria and health in Africa—the present situation and epidemiologic trends. *Trop. Med. Parasitol.* **42**, 204–213 (1991).
18. Mason, S. J. & Tyson, P. D. in *Droughts Vol. 1 A Global Assessment* (ed. Wilhite, D. A.) 113–134 (Routledge, New York, 2000).
19. Xie, P. & Arkin, P. A. Global monthly precipitation from satellite-observed outgoing longwave radiation. *J. Clim.* **11**, 137–164 (1998).
20. Hagedorn, R., Doblas-Reyes, F. J. & Palmer, T. N. The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concepts. *Tellus A* **57**, 219–233 (2005).
21. Uppala, S. *et al.* ERA-40: ECMWF 45-year reanalysis of the global atmosphere and surface conditions 1957–2002. *ECMWF Newsl.* **101**, 2–21 (2004).
22. New, M., Hulme, M. & Jones, P. D. Representing twentieth century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Clim.* **12**, 829–856 (1999).
23. New, M., Hulme, M. & Jones, P. D. Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.* **13**, 2217–2238 (2000).
24. Mason, S. J. & Graham, N. E. Areas beneath the relative operating characteristics (ROC) and levels (ROL) curves: statistical significance and interpretation. *Q. J.R. Meteorol. Soc.* **128**, 2145–2166 (2002).

Acknowledgements We thank the WHO-AFRO Southern Africa Inter-Country Malaria Team (SAMC) in Zimbabwe and the National Malaria Control Unit in Botswana for enabling this study. The work reported here is part of the EU-funded DEMETER and ENSEMBLES projects. It was additionally supported financially by the UK Department for International Development through the Malaria Knowledge Programme, Liverpool School of Tropical Medicine, UK, and by a cooperative agreement from the US National Oceanic and Atmospheric Administration. The authors acknowledge considerable technical support from ECMWF staff and consultants, in particular from the Seasonal Forecast Section. The views herein contained are those of the authors and do not necessarily reflect the views of WHO, DfID, EU, NOAA or any of their sub-agencies.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.N.P. (tim.palmer@ecmwf.int).