Tellus (2005), 57A, 464–475
Printed in UK. All rights reserved

TELLUS

A forecast quality assessment of an end-to-end probabilistic multi-model seasonal forecast system using a malaria model

By ANDREW P. MORSE[1]*, FRANCISCO J. DOBLAS-REYES[2], MOSHE B. HOSHEN[3], RENATE HAGEDORN[2] and TIM N. PALMER[2], [1]*Department of Geography, University of Liverpool, Liverpool, L69 7ZT, UK;* [2]*European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK;* [3]*Department of Physics, University of Liverpool, Liverpool, UK*

(Manuscript received 2 April 2004; in final form 2 December 2004)

ABSTRACT

We discuss a novel three-tier hierarchical approach to the validation of an end-to-end seasonal climate forecast system. We present a malaria transmission simulation model (MTSM) driven with output from the DEMETER multi-model seasonal climate predictions, to produce probabilistic hindcasts of malaria prevalence. These prevalence hindcasts are second-tier validated against estimates from the MTSM driven with ERA-40 gridded analyses. The DEMETER–MTSM prevalence hindcasts are shown to be (tier-2) skilful for the one-month lead seasonal predictions as well as for the period covering the seasonal malaria peak with a 4–6 month forecast window for the event prevalence above the median. Interestingly, the tier-2 Brier skill score for the forecast window of the hindcasts starting in February, for the event prevalence above the median, is higher than for either the tier-1 precipitation or temperature forecasts, which were the MTSM driving variables.

1. Introduction

As exemplified in the seasonal prediction project DEMETER, probabilistic multi-model ensemble-based seasonal climate predictions have shown clear skill and reliability in a number of regions of the world for a number of meteorological variables (Palmer et al., 2004; Hagedorn et al., 2005). The level of skill in the tropics in particular has led to serious efforts to begin integrating quantitative application models to the climate prediction ensembles. In this paper we describe the integration of a malaria transmission simulation model (MTSM) in the DEMETER multi-model ensemble, to provide probabilistic predictions of simulated malaria prevalence scenarios. Based on the DEMETER hindcast data set, probabilistic hindcasts of simulated malaria prevalence scenarios for regions of tropical Africa have been produced.

In addition to the multidisciplinary approach to seasonal forecasting, in this paper we develop a novel conceptual framework for the forecast quality assessment of seasonal climate forecasts. The work described in Palmer et al. (2004) and Hagedorn et al. (2005) refers to the verification of individual meteorological fields, such as temperature and precipitation from the DEMETER hindcasts, using the corresponding meteorological fields from the global ERA-40 gridded meteorological reanalysis data as a reference data set. In this paper, we refer to this as tier-1 validation.

With the integration of a malaria model into the DEMETER ensemble system, we could also validate the resulting probabilistic forecasts of malaria prevalence against recorded malaria transmission data and other appropriate clinical data sets. We refer to this as tier-3 validation (Fig. 1). For end-user variables to be skilful in a tier-3 validation it is necessary not only that the seasonal forecasts and application models are skilful, but also that the downscaling methodologies (Yarnal et al., 2001), which take gridded climate forecast data on scales of hundreds of kilometres to more local scales, are accurate. Moreover, in a case such as that dealt with in this paper, for reliable tier-3 validation, it is necessary that adequate malaria clinical data of sufficiently high quality actually exist. Unfortunately, for most parts of Africa this is not the case.

From a scientific point of view, there is a vast gulf between tier-1 and tier-3 forecast quality assessment, and it is necessary to develop a more hierarchical approach in which the different elements of the full tier-3 validation can be isolated. In this paper, we introduce the notion of an intermediate tier-2 validation. In tier-2 validation, the essential reference data are meteorological (as in tier-1), but the data are integrated into the application

*Corresponding author.
e-mail: a.p.morse@liv.ac.uk

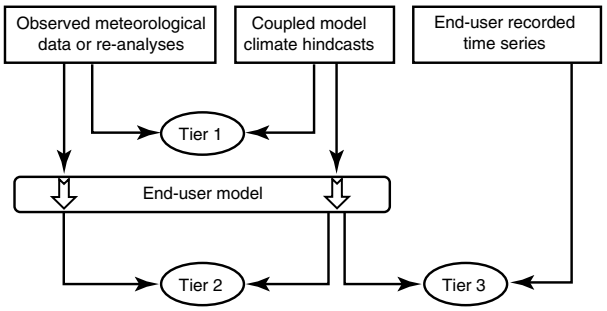

Fig 1. Schematic representation of the three-tier validation system. Rectangular boxes represent sources of data, while ovals indicate the different types of validation. A comparison of coupled model hindcasts with the corresponding observed/reanalysed variables is carried out in a tier-1 validation. In the example described in the paper, both observed/reanalysis and hindcast data are used in MTSM experiments (large arrows) to obtain simulated prevalence data. Simulated prevalence obtained with coupled model hindcasts can be compared with the MTSM output obtained using observed/reanalysis data (tier-2 validation) or with clinical cases (tier-3 validation).

model to produce an estimate of the application variable used in the tier-3 validation. For tier-2 validation to be meaningful, we require the end-user application model to be representative of the processes for that application field, e.g. crop yield or malaria (and hence to have undergone independent off-line validation); however, tier-2 does not validate the application model per se. By using a tier-2 validation, we will be validating the DEMETER output against ERA-40 gridded reanalysis, but in a situation where meteorological temperature and precipitation fields from both DEMETER output and ERA-40 have been integrated and synthesized to produce a single malaria-relevant field. As in a tier-3 validation, the tier-2 forecast quality assessment evaluates the skill of the predictions not as individual variables, but integrated through the end-user or application model, which is considered as a multivariate non-linear transfer function. Given that ERA-40 is a global field, the problem of regions with inadequate validation data (as in tier-3) is not an issue. The concept of a three-tier validation of an integrated probabilistic system is novel. As a consequence, the results reported in this paper represent the first attempt of a tier-2 validation using an application model integrated in a probabilistic seasonal forecast system.

Although malaria is present in many parts of the tropical world, the focus of our study is Africa, which has the greatest burden from the disease. Malaria is estimated to kill between 700 000 and 2 700 000 annually with over 75% of the victims being African children (see http://www.mim.su.se/ english/news/ newsrelease_080201.html; also see http://www.nature.com/ nature/outlook/malaria/ and http://www.nature.com/nature/focus/ malaria/). Including the annual number of acute cases probably in excess of 300 million, it is easy to start to understand the impact of this disease. In Africa, the temperature and rainfall regimes produce a spectrum of malaria transmission. This leads to areas with stable malaria transmission where disease rates are similar from year to year, and the non-pregnant adult population is largely immune to severe disease, through to areas of unstable malaria transmission where the disease is rare but epidemics may occur affecting all age groups. In unstable areas, epidemics are often the consequence of climate anomalies, which increase vector breeding and survivorship and parasite development rates. Poveda et al. (2004) show how malaria in Colombia is associated with the annual climatic cycle and how the anomalous climatic conditions during El Niño Southern Oscillation events lead to an increase in transmission. Snow et al. (1999) show the connection between climate and malaria in Africa. Further, Hay et al. (2002) show that, in complex topographic regions with a large amount of spatial–temporal variability, making a connection between existing monthly averaged climate data and clinical data can be difficult to achieve. The region used for the model runs in this paper does not have such complex topography. Malaria is caused by the *Plasmodium spp.* parasite that is passed between humans by species of *Anopheles spp.* mosquitoes, the vector. The disease only occurs in areas where environmental conditions are suitable for both the parasite and vector, and these conditions are sustained for a number of months. The temperature drives the development of the parasite within the vector and it also drives the developmental life cycle of the vector. For both developmental cycles, there are lower-temperature thresholds and, within certain upper bounds, higher temperatures lead to greater rates of development. Precipitation is important in providing breeding sites for mosquitoes and for increasing the humidity of the air, which increases the survivorship of the vectors.

Why apply seasonal forecasts to the field of malaria transmission prediction? Health planners would greatly benefit from prior knowledge of areas at risk of climate-related epidemics in the forthcoming season, and skilful seasonal climate forecasts may provide early warning to allow interventions to be in place before the start of the epidemic (Thomson et al., 2000). The malaria community has shown considerable interest in the use of seasonal climate forecasts for the development of malaria early warning systems (MEWS; World Health Organization 2001), as a direct consequence of numerous reports indicating that malaria incidence (including epidemics) in certain parts of the world can be shown to be correlated with sea surface temperatures (SSTs; Kovats et al., 2003). As seasonal climate forecasts are often most skilful for conditions when there are strong SST anomalies (Stockdale et al., 1998), it follows that seasonal climate forecasts could provide health planners with early warning of climate anomalies which predispose certain areas to malaria epidemics. Therefore, there is currently a need to develop a series of methodologies, including that contained in this paper, for the assessment of the actual value of seasonal climate forecasts and the development of useful products for the epidemiological community. This paper introduces the first stage to developing such a system as part of a MEWS.

The importance of models describing malaria transmission, taking into account the interannual variability and combined effect of both rainfall and temperature, has been highlighted by Zhou et al. (2004) in their work on malaria epidemics in the East African highlands. Statistical models can be developed to relate parameters, such as seasonal rainfall or degree–day totals, to seasonal crop yields or seasonal total malaria cases. The value of both applications can be limited as it is not only the seasonal total that controls the yield or cases but also the distribution of the meteorological parameters throughout the season. A dynamic approach running with probabilistic seasonal forecasts offers the possibility of capturing some of the variability through the season, such as the prediction of anomalous rainfall onset and cessation.

The paper is organized as follows. In Section 2 we outline the structure of the malaria model, data sets and data conditioning used, the process by which the malaria data sets were produced, the methods used to assess the forecast quality, and the forecast windows and categories that are used in these assessments. In Section 3 we describe the results for both the ERA-40 malaria prevalence, used as reference, and the multi-model hindcasts, followed by the description of detailed forecast assessments of the probabilistic forecasts of precipitation, temperature and malaria prevalence. In Sections 4 and 5, we present a discussion and the main conclusions, where we assess the possible reasons and implication of our findings and suggest where these findings may lead in the future.

2. Methods and data

2.1. Description of the malaria model

A detailed description of the MTSM used in this study is given in Hoshen and Morse (2004). The principles of malaria modelling can be found in Anderson and May (1991), Dietz (1988) and MacDonald (1957). A brief description of the model structure and its dynamics are given in this section. The malaria model used in this study is split into three sections covering (i) the larvae stage of the mosquito, (ii) the uninfected, infected and infectious stages of the adult mosquitoes and (iii) the human host also in uninfected, infected and infectious stages. The three sections of the model are briefly discussed in turn. The reader is referred to Hoshen and Morse (2004) for more specific details about the model and its performance.

The larvae stage is included in few models and therefore most have a near constant mosquito population. The reason for the lack of larvae stages is the paucity of empirical data relating this stage to climatic controls. The evidence of temperature-driven larvae development from Jepson et al. (1948) was used to deduce a larvae development rate, and a proxy breeding site availability is simulated through a precipitation-driven multiplication factor.

The adult mosquito stage follows in most models the same general governing values that relate to the two important thermally driven cycles. These cycles are the egg-laying cycle of the adult mosquito, which is in turn connected to adult mosquito survivorship and, secondly, the within vector development of the parasite. This section of the model has three states of the adult mosquito: uninfected, infected and infectious. Detinova (1962) showed that the gonotrophic (egg production) cycle takes 37 degree days above a threshold of 7.7°C in humid conditions. At typical tropical temperatures, this is about three, whole rounded up, days. The cycle is initiated by a blood meal and finishes with the laying of the eggs. If, during this blood meal, the mosquito bites an infectious human and ingests the malaria parasite, a second cycle is initiated within the vector, which is also driven by the daily temperature. This cycle is called the sporogonic cycle, which is the development of the parasite within the vector, and takes 111 degree days above the threshold of 18°C (Detinova, 1962); thus, at typical tropical temperatures about 12 d to complete. Therefore, it takes about four to five gonotrophic cycles to complete one sporogonic cycle, this ratio being dependent on temperature. When the daily temperature is close to the sporogonic threshold, the ratio increases to greater than 20. As a mosquito is prone to predation when taking its blood meal, literature values suggest a per gonotrophic cycle survivorship of 50%. Of a cohort of mosquitoes emerging simultaneously from the larvae stage, less than 1% would survive seven gonotrophic cycles. For malaria to be effectively transmitted to the human hosts, there needs to be a significant pool of infectious mosquitoes, and this is only going to occur when the sporogonic cycle is completed in a small number of gonotrophic cycles.

Once an infectious mosquito bites a human and parasites are injected, the parasites go through further stages of their life cycle development within the human host. It then takes about two weeks before the human host becomes infectious and is able to transmit the parasite when bitten by an uninfected mosquito. This stage is weather-independent. It is suggested that immunity may play a role in transmission of malaria in areas of stable transmission, as immune adults have fewer gametocytes, the sexual stage of the parasite cycle, that are ingested by a mosquito during a blood meal. There is however discussion of how this quantifiably affects transmission in the field and that other factors may affect the chance of transmission (see, for example, Taylor and Read, 1997; Piper et al., 1999; Collins et al., 2004). Therefore, the model simulations in this paper are used for epidemic simulation and not for simulation in stable areas. In areas prone to epidemics, there is no immunity and therefore this effect does not need to be included in this model.

At any daily time-step within the malaria model, information is available for both the vector and host population, giving the proportions of those populations that are uninfected, infected and infectious. The model does not contain any host mortality but does have a natural clear-up rate of infectious hosts at 3% per day. This leads to about a 90% probability of an infected host being clear of the disease after 80 d.

2.2. *Climatological data*

The DEMETER project (see http://www.ecmwf.int/research/demeter/) has been funded under the European Union (EU) Vth Framework to assess the skill and potential economic value of multi-model ensemble seasonal forecasts. The DEMETER multi-model prediction system comprises seven global coupled ocean–atmosphere models (for details, see Palmer et al., 2004). The DEMETER hindcasts were started four times a year from 1 February, 1 May, 1 August and 1 November. The hindcasts were integrated for 180 d and comprise an ensemble of nine members. Hindcasts have been produced over the period 1958–2001, although the period common to the seven models is 1980–2001. The seven models and nine ensemble members per model give a total of 63 hindcasts for each start date. The performance of the DEMETER system has been evaluated from this comprehensive set of hindcasts (Palmer et al., 2004; Hagedorn et al., 2005).

The ERA-40 reanalysis project (Uppala et al., 2004) was also funded under the EU Vth Framework to produce gridded global reanalyses for the period 1957–2002. Reanalysis is a global forecast model product and represents the initial conditions used in global model runs. It is obtained from a combination of observations from all available sources (e.g. meteorological stations, radiosondes, aircraft, ship measurements and satellite radiances) with short-range predictions carried out with the same atmospheric model. Unlike operational analysis that has to be competed by a cut-off time to allow the operational model products to be delivered on time, reanalyses have the luxury of being able to conduct comprehensive quality assurance on observational data and include more observations than is possible for an operational run. ERA-40 is the first of the second generation global reanalysis products building on previous work, especially ERA-15, which was produced in the 1990s. In DEMETER, it was used for all of the forecast models as the source of their initial fields. For the application modelling groups, it was used as a reference forecast to which the DEMETER hindcast application model runs were compared.

The data for this paper are taken from four grid points in southern Africa. The grid points are at 2.5° resolution along latitudes of 17.5°S, from 22.5°E, which is on the eastern edge of Angola, to 30°E, which is in Zimbabwe. The data have been run through the model at the 2.5° resolution and no downscaling was undertaken. The results are based on 15 yr of daily data from bias-corrected hindcasts for the period 1987–2001. The corresponding ERA-40 data end in April 2002. The fields used are daily accumulated precipitation and 2-m maximum temperature. Data from ERA-40 and the 63 hindcasts were used to drive the MTSM runs. The temperature has an offset of −5°C to represent the daily mean temperature.

Seasonal hindcasts generated using coupled models are prone to large biases. In order to remove biases from daily predictions of temperature and precipitation, an estimate of the seasonal cycle at each grid point was obtained by averaging daily data. This estimate was smoothed out by retaining the three first harmonics in a Fourier decomposition of the time series. The same method was used to estimate the seasonal cycle with the ERA-40 data. The bias was defined as the difference between the model and the ERA-40 seasonal cycles and this bias is removed from the hindcasts.

2.3. *Production of malaria transmission hindcasts*

The hindcasts are produced as 180-d long integrations starting on 1 February, 1 May, 1 August and 1 November. The MTSM is initialized with the previous 2 yr of ERA data for each of the start dates. The 180-d hindcasts, along with the ERA-40 data from the equivalent time period, are run as separate integrations making 64 simulations in total for each start date. Therefore, some 3840 malaria model runs per grid point for the time period were investigated in this paper. Outputs from the MSTM are stored at a daily time-step through the 180-d integration. For the analysis in this paper, the output values were accumulated over 90-d blocks within the integration, producing an average through the two to four month and four to six month forecast windows.

The main outputs from the MTSM are incidence, which is the number of new cases of infection in a period, and prevalence, which is the total number of cases at any one point in time. Prevalence may be calculated by the integral of daily incidence minus the daily number of cases that clear up. In this paper we concentrate on simulated prevalence. Although the term prevalence is used, it represents a simulated prevalence and cannot be taken as a direct prediction of actual clinical cases. The three-month seasonal prevalence averages from the malaria model when driven by the ERA-40 forecast, which in a tier-2 validation framework is taken to be reality, and the 63 DEMETER hindcast ensemble members were run through standard forecast quality analysis routines (see Section 2.4) to produce a series of standardized validation plots.

2.4. *Assessment of forecast quality*

Forecast quality assessment is an essential component of the forecast formulation process (Jolliffe and Stephenson, 2003). Forecast quality is a complex concept described through a number of different attributes that provide useful information about the performance of a forecasting system. Thus, no single measure is sufficient for judging and comparing forecast quality. Forecast quality has been evaluated in this paper using measures of bias, reliability and accuracy. The data are presented as box-and-whisker plots, Brier skill scores and relative operating characteristic (ROC) skill scores. The box-and-whisker plots represent the three terciles of the hindcast ensemble probability distribution function (PDF), the box is the middle tercile and the whiskers the upper and lower terciles with the ensemble mean value as a solid circle and ERA-40 reference forecast value as a hollow diamond (as in Figs 5 and 6, discussed below). The

Brier score is a scoring measure that estimates the quality of probabilistic forecasts for dichotomous events. It is defined as

$$\mathrm{BS} = 1/N \sum_{i=1}^{N} (p_i - o_i)^2,$$

where the summation is over all cases and/or grid points, p_i represents the probability of predicting a dichotomous event (of the type yes/no) and o_i is the corresponding observed probability, which usually is taken as 1 if the event occurs and 0 otherwise. The relative quality of this score is measured against a trivial reference forecast, which in this case has been the climatological frequency of the event. The relative change of the Brier score against this reference is known as the Brier skill score. Positive values of the skill score imply an increase of the forecast quality against a climatological forecast. The ROC skill score indicates the performance of dichotomous predictions in terms of hit and false alarm rate stratified by the verification. The hit rate is the relative number of times an event was forecast when it occurred, while the false alarm rate is the relative number of times the event was forecast and it did not occur. Ideally, the hit rate will always exceed the false alarm rate (Richardson, 2001). Therefore, in the case of the ROC skill score, any positive value indicates a skilful system. A system with no skill (made by either random or constant forecasts) will show a skill score of zero. In order to give a measure of the usefulness of the forecasts in some objective way, a simple decision model whose inputs are the hit and false alarm rate can be used. This simple conceptual model allows the estimate of the potential value of a set of forecasts. Consider a potential user who can take some specific precautionary action depending on the probability of the event. The action incurs a cost, C, regardless of whether the event occurs or not. However, if the event occurs and no action has been taken, a loss L is incurred. The expense associated with each combination of action/inaction and occurrence/non-occurrence can be expressed as a function of the cost–loss ratio C/L (Richardson, 2000). If only climatological information is available, two basic options remain: either always or never take precautionary action. The cost–loss model estimates the reduction in expenses beyond what could be achieved using climatological information alone. Further details can be found in Wilks (1995), Doblas-Reyes et al. (2005), Thornes and Stephenson (2001), Jolliffe and Stephenson (2003), Doblas-Reyes et al. (2003) and Mason (2004).

2.5. *Forecast windows and forecast categories*

Three forecast windows, which represent the average forecast over a three-month window, were chosen to coincide with stages in the annual malaria cycle found in the averaged MTSM runs for this region. Two forecast windows from the February start date of forecast months 2 to 4 (referred to as Feb 2–4, MAM) and months 4 to 6 (referred to as Feb 4–6, MJJ) were chosen to represent forecasts at the start and peak of the model simulated malaria season. The third forecast window was the 4 to 6 month window from the November start date (referred to as Nov 4–6, FMA) to coincide with a longer forecast window running across the start of the rise in malaria prevalence. As the rainy season starts in October to November, for precipitation only, additional forecast windows were evaluated. These forecasts are the 2 to 4 month forecast from the November start date (referred to as Nov 2–4, DJF) and the 4 to 6 month forecast from the August start date (referred to as Aug 4–6, NDJ) to investigate if the rainfall 'onset' has a skilful forecast.

Depending on the forecast assessment undertaken, data are either examined across the 15 yr of model output for each of the grid points or the values are computed for all years and grid points. The former is the case for the box-and-whisker plots and the latter is the case for the Brier skill scores and ROC skill scores. Brier and ROC skill scores are calculated for the three forecast windows and for three forecast event categories: (i) the prediction of an anomaly within the lower tercile event, (ii) the prediction of an anomaly above the median and (iii) the prediction of an anomaly within the upper tercile event. The upper tercile indicates the ability to forecast events in the upper third of the observations; the lower being the same for the lower third and above the median the ability of the system to forecast events in the upper half of the hindcasts. These categories are the standard ones used in the DEMETER verification system and are commonly used in other forecasting systems. In addition, these categories are of relevance for a MEWS. Individual applications may develop tailored appropriate event categories.

3. Results

3.1. *ERA-40 climate and malaria model average performance*

The mean seasonal cycles have a fair correspondence albeit a little lagged (see below) with respect to the modelled malaria transmission seasons from the Mapping Malaria Risk in Africa (MARA) project (http://www.mara.org.za/). This prevents the results being used in a tier-3 validation scheme. However, as we are only attempting a tier-2 validation in this paper, there is no requirement for the simulation in this modelling system to correspond exactly with those observed or, in the case of MARA, modelled seasonal transmissions patterns for the objectives of this paper to be accomplished.

The MTSM prevalence and ERA-40 precipitation and temperature average seasonal cycles are discussed below. The ERA-40 data show a bimodal temperature regime (Fig. 2) with a secondary peak in April and the highest temperatures in October. All the sites follow the same seasonal temperature pattern, preserving the rank order throughout the year. The temperature for the coolest grid point is consistently between 3°C and 4°C lower than the warmest grid point, with temperatures decreasing from west to east across the four grid points. All of the grid points have a minimum temperature in July. This minimum at one grid

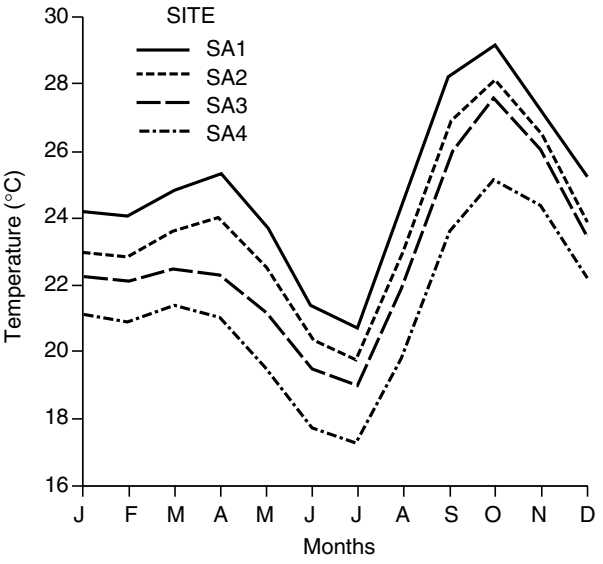

Fig 2. ERA-40 monthly average temperature for 15 yr from 1987 to 2002, where SA1, SA2, SA3 and SA4 are the grid points 17.5°S 22.5°E, 17.5°S 25.0°E, 17.5°S 27.5°E and 17.5°S 30.0°E, respectively.

point is below the 18°C threshold for the sporogonic cycle and the annual minima at the other grid points are close to this threshold. The ERA-40 precipitation (Fig. 3) has a unimodal distribution for all four grid points. The rainy season starts just as the temperature peaks, and as the precipitation increases, the temperatures fall. The precipitation peaks in February and ceases in June. The ERA-40 precipitation data at each grid point have different seasonal peak values and seasonal totals. Once the rains fully commence, the mosquito population, initially uninfected, starts to increase rapidly through January and reaching its peak in March with no new mosquitoes emerging after May. This limit is due to the cessation of precipitation; as the model has no land surface or hydrological features, there is no possibility of year-round breeding sites. The inclusion of a more realistic land surface and the requirements of downscaling the hindcast data to more appropriate scales is a topic currently in development. The emergence of the infectious mosquitoes lags behind the emergence of the uninfected mosquitoes due to the biological process outlined in Section 2.1 with the peak occurring between April and May depending on the grid point, which may already indicate the impact of differences in temperature between the grid points. Therefore, it is no surprise that the host prevalence peaks in May (Fig. 4), the prevalence curve starting to rise in February and returning to its pre-season low levels in September, reflecting in part the natural clear-up rate in the human host. The peak prevalence is different in each grid point, as is the total number of cases over the season. This is a reflection of the differences in the temperature and precipitation amplitude across the four grid points.

3.2. *Precipitation forecast quality*

Box-and-whisker plots (not shown) for the periods Nov 2–4 (DJF), Feb 2–4 (MAM), Aug 4–6 (NDJ), Nov 4–6 (FMA) and Feb 4–6 (MJJ) show, in most years across the four grid points, that the ERA-40 value is captured within the hindcast PDF. This is a desirable feature because it indicates that the hindcast values belong to the same population as the reference values. The ensemble spread is the largest for Nov 2–4 (DJF), but this is

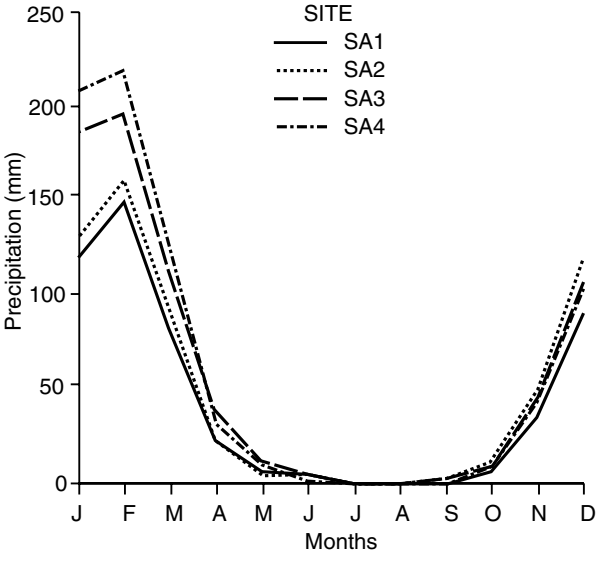

Fig 3. ERA-40 monthly average precipitation for 15 yr from 1987 to 2002, where SA1, SA2, SA3 and SA4 are the grid points 17.5°S 22.5°E, 17.5°S 25.0°E, 17.5°S 27.5°E and 17.5°S 30.0°E, respectively.

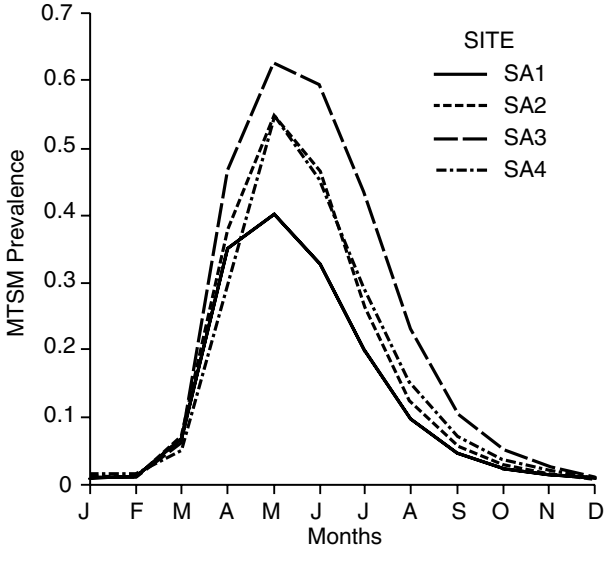

Fig 4. MTSM monthly mean proportional prevalence for 15 yr from 1987 to 2002 using model runs driven by ERA-40 temperature and precipitation, where SA1, SA2, SA3 and SA4 are the grid points 17.5°S 22.5°E, 17.5°S 25.0°E, 17.5°S 27.5°E and 17.5°S 30.0°E, respectively.

Table 1. Brier skill scores (BSS) and ROC skill scores (ROCSS) for the forecast windows Nov 4–6, Feb 2–4 and Feb 4–6 for precipitation, temperature and MTSM prevalence, where LT is the lower tercile event, AM is the above-the-median event and UT is the upper tercile event

	BSS			ROCSS		
	Nov 4–6 (FMA)	Feb 2–4 (MAM)	Feb 4–6 (MJJ)	Nov 4–6 (FMA)	Feb 2–4 (MAM)	Feb 4–6 (MJJ)
Precipitation						
UT	−0.052	−0.020	−0.049	−0.202	0.078	−0.062
AM	−0.005	−0.009	−0.039	0.071	0.087	0.048
LT	−0.038	−0.094	−0.012	−0.077	−0.126	0.193
Temperature						
UT	0.256	0.230	0.314	0.694	0.538	0.861
AM	0.231	0.148	0.210	0.616	0.399	0.603
LT	0.176	0.080	0.104	0.556	0.316	0.479
Prevalence						
UT	−0.067	0.046	0.178	−0.065	0.537	0.501
AM	−0.034	0.461	0.289	0.008	0.773	0.642
LT	0.017	0.396	0.167	0.174	0.720	0.501

also the forecast window with the highest precipitation values. Interestingly, the ensemble spread on the three longer-range forecast windows in general is not larger than for the 2 to 4 month forecasts, apart from the occasional year in the Aug 4–6 (NDJ) forecast window. The Feb 4–6 (MJJ) period, which covers the start of the dry season, shows no anomalous ensemble members, i.e. no grossly late cessation of the rainy season showing up in any of the ensemble members.

Brier skill scores (see Table 1) for the multi-model show that there is no skill for Feb 2–4 (MAM), Nov 4–6 (FMA) and Feb 4–6 (MJJ) for all forecast categories. Skill can be gained through ensemble refinement, but the skill remains very low. Here, the term ensemble refinement means the process of improving the predictions by removing the worst performing individual model. This can be repeated to see the impact of the removal of the lowest skilled two or three models within the ensemble. This least skilful model is not easily identified due to the small sample and can change between forecast categories within the same forecast window. Furthermore, the removal of the least skilful model needs to be done in cross-validation mode and this shows that the selection of such a model is not always a robust decision, mainly due to the small sample size. These difficulties illustrate the advantages of using a simple multi-model approach in which all the models are given the same weight in the multi-model ensemble, as shown by Hagedorn et al. (2005). The additional precipitation forecast windows are not included in Table 1. For Nov 2–4 (DJF), the multi-model has skill in all categories but it is very low for the lower tercile. Ensemble refinement increases the skill in all categories. The Aug 4–6 (NDJ) multi-model has large negative scores, i.e. the forecast is less skilful than using climatology and does not gain skill through ensemble refinement. This forecast window has the worst performance out of the five that were examined. These last two forecast windows represent, in part, the forecast of the onset of the rainy season with two different lead times.

ROC skill scores (Table 1) are positive in the above-the-median category but are negative for the majority of forecast windows in the other two categories. Negative scores (i.e. worse than climatology) are found for half of these forecast windows. However, through ensemble refinement they can all become positive. For the Nov 2–4 (DJF) forecast (not shown), there are positive scores in each category with the best score of almost 0.4 for the above the median category. For the Aug 4–6 (NDJ) forecast (not shown), the multi-model ROC skill score is either zero or negative in all categories but ensemble refinement can lead to slightly positive scores for the lower and upper tercile categories.

The precipitation predictions are poor overall when compared with the other two variables under review in this paper. The ROC scores tend to be better than the Brier skill scores, which points to a problem with the reliability of the predictions. There is also evidence that there is low skill at the start of the rainy season.

3.3. Temperature

The forecast windows Feb 2–4 (MAM), Nov 4–6 (FMA) and Feb 4–6 (MJJ) allow the investigation of the temperature hindcasts through the seasonal mosquito cycle of January to May, and its control within vector parasite development, through the use of the MTSM. The MJJ hindcast captures the coolest part of the year when the temperature becomes marginal for the development of the parasite within the mosquito, at some of the grid points during June and July. Box-and-whisker plots show for all the forecast

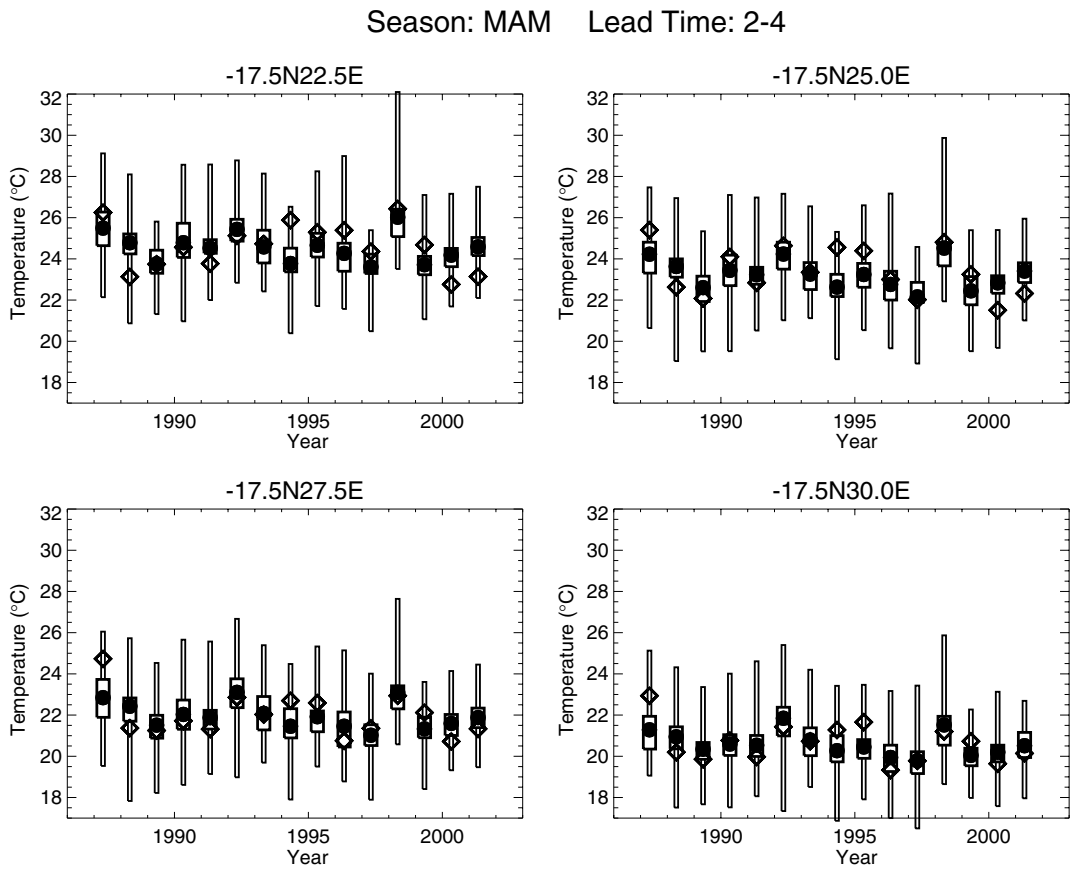

Fig 5. Temperature as a box-and-whisker plot for the four grid points for the Feb 2–4 (MAM) forecast window, showing the ERA value (hollow diamond) and ensemble mean (solid circle) where the range of the box is the middle tercile and the upper and lower whiskers the upper and lower terciles of the ensemble distribution, respectively.

windows the capture of the ERA-40 temperature by the hindcast PDF for all years and all grid points. The ensemble spread is of a similar magnitude for the Feb 2–4 (MAM) and Nov 4–6 (FMA) hindcast windows. The box-and-whisker plot for Feb 2–4 (MAM) is shown as an example in Fig. 5. However, the spread is larger for the Feb 4–6 (MJJ) data, particularly for two grid points that have lower average temperatures. To investigate the tercile asymmetry an additional hindcast, the May 2–4 (JJA) forecast, was examined. This forecast does not show the asymmetry in the tercile spread for these grid points. Therefore, this increase in the range of the lower tercile may be due to the extended Feb 4–6 (MJJ) forecast window.

Brier skill scores (Table 1) are positive for all forecast categories and all forecast windows. The skill has higher values, surprisingly, for the 4–6 month forecasts than the 2–4 month forecast. The highest Brier skill scores for all of the forecast windows are found in the upper tercile category for each start date. In this region of Africa, temperature as well as precipitation has control of the development of malaria so the ability of the DEMETER system to produce skilful forecasts for the upper tercile is particularly encouraging. The higher skill score found in the upper tercile category is repeated for the ROC skill scores (Table 1) with a ROC skill score of 0.86 in the upper tercile for the Feb 4–6 forecast.

3.4. Prevalence

A box-and-whisker plot for the four grid points from the Feb 2–4 (MAM) forecast window is shown in Fig. 6. It can be seen that, as in the case of precipitation and temperature, the ERA-40 reference forecast is captured by the DEMETER hindcast driven PDF for almost all of the years at each of the four grid points. The spread of the hindcast members varies from grid point to grid point, with the largest spread generally found mainly in the lower tercile. The largest spread is mostly seen for two of the grid points (17.5°S 27.5°E and 17.5°S 30.0°E) and is probably due to the lower average temperature (the temperature box-and-whisker plot for the same forecast window is shown in Fig. 5). For these grid points, some of the ensemble members have temperatures that range below the 18°C threshold for the development of the malaria parasite. Besides, most of the lower tercile values have a temperature during the critical stages of the parasite development

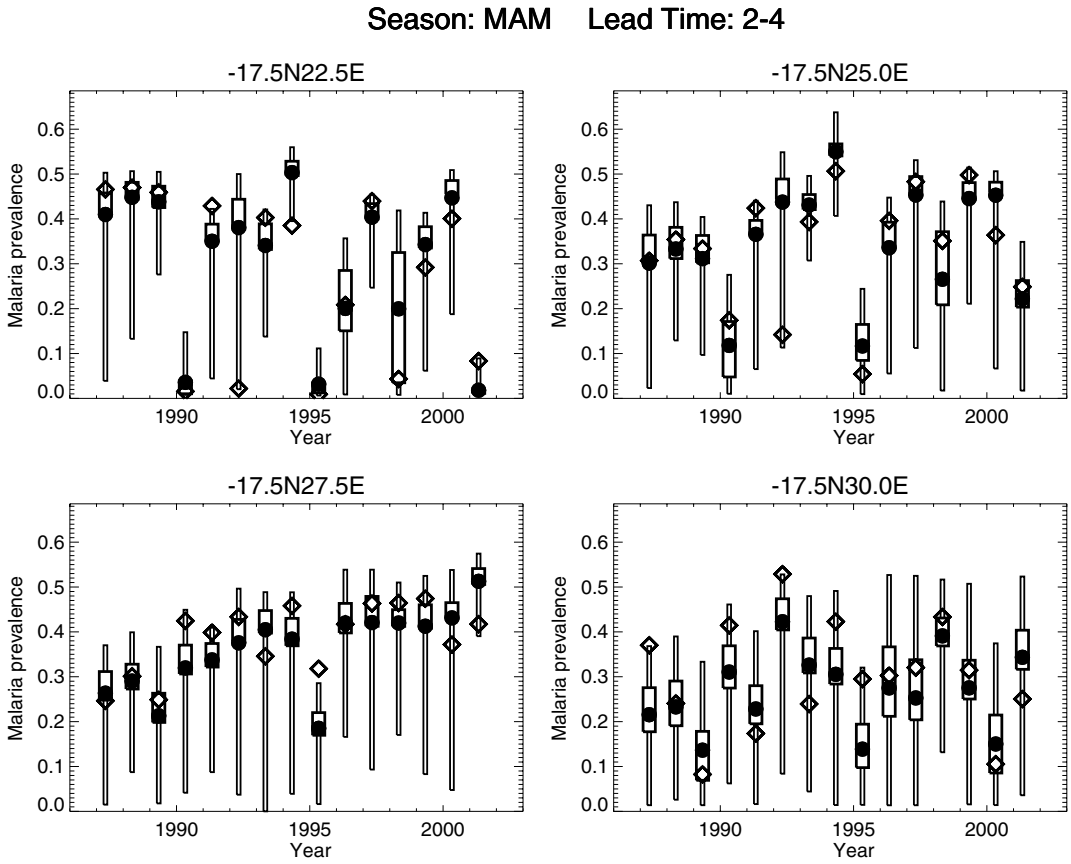

Fig 6. Prevalence from the MTSM as a box-and-whisker plot for the four grid points for the Feb 2–4 (MAM) forecast window, showing the ERA value (hollow diamond) and ensemble mean (solid circle) where the range of the box is the middle tercile and the upper and lower whiskers the upper and lower terciles of the ensemble distribution, respectively.

close to 18°C, which would lead to very slow development rates. Similar plots are seen for the other forecast windows in as far as the ERA-40 run is almost always captured by the hindcast PDF. The largest spread is seen on the Feb 4–6 (MJJ) plots (not shown).

For the Feb 2–4 (MAM) forecast window, the multi-model Brier skill score was positive (Table 1) with the highest score gained in the above the median category. The skill is retained, perhaps surprisingly, for the longer lead time Feb 4–6 (MJJ) forecast window. The multi-model DEMETER system is skilful in all prediction categories with the greatest skill in the above-the-median category. However, the Nov 4–6 (FMA) has little, if any, skill.

ROC skill scores for the Feb 2–4 (MAM) and Feb 4–6 (MJJ) forecasts (Table 1) show that the skill scores are positive in all three categories, with the above-the-median event gaining the highest scores for both forecast windows. This shows that there would be potential to make a forecast with four-month lead time. The scores are higher in the 2–4 month window compared to the 4–6 month window, suggesting a positive impact of the initial conditions. In the Nov 4–6 (FMA) forecast window, the DEMETER multi-model system has a positive skill only for the lower tercile event. The upper tercile event, ideally, should have the maximum forecast skill for potential use as part of a MEWS, but here it has no skill. However, for many uses, including malaria early warning, a reliable and skilful forecast of a non-event is also important for planning purposes. The good agreement found between the ROC and Brier score results indicates that the prevalence predictions are reliable, in spite of the lack of reliability of the precipitation predictions.

Cost–loss ratio curves for the prevalence forecasts across the three tercile categories have been examined for the Feb 2–4 (MAM) and Feb 4–6 (MJJ) events. These forecast windows were chosen as they have positive skill scores as discussed above. Zhu et al. (2002) and Palmer (2002) discuss the use of this type of diagnostic for assessing potential economic value within a forecasting system. With this diagnostic, the actual economic value of the simulated malaria prevalence forecast within an ensemble prediction system is not assessed, but the potential value of such a system. This would also apply to a tier-3 validation framework. As such, the multi-model result has a positive potential economic value for a range of cost–loss ratios for the three forecast

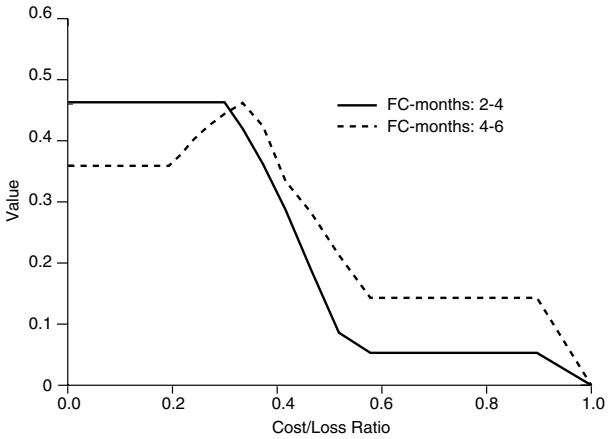

Fig 7. Potential economic value curve for a range of cost–loss ratios for the MTSM for the upper tercile event category for the Feb 2–4 (MAM) and Feb 4–6 (MJJ) forecast windows.

categories for both forecast windows. In both forecast windows, the greatest potential economic value was found in the above-the-median category with potential economic value across the full range of cost–loss ratios. The upper tercile event (Fig. 7) has lower potential economic values (at peak) than the other forecast categories (not shown) but the result is still positive. It can be seen that there is little value beyond a cost–loss ratio of about 0.6. Actual cost–loss ratios will depend on the application and the area of its operation. Although this paper does not address an actual situation, it is generally accepted that prevention schemes are much less expensive than curing the disease and even more cost effective if associated economic losses are taken into account.

4. Discussion

The importance of the user community for the application of climate forecasts (Pfaff et al., 1999; Archer, 2003) for probabilistic seasonal forecasts (Hartmann et al., 2003; Zhu et al., 2002; Palmer, 2002) is recognized. The literature is limited on probabilistic application forecasts that apply some form of seasonal scale forecasts (Franz et al., 2003; Potgieter et al., 2003) with few reports of probabilistic application models embedded within an ensemble prediction system (e.g. Cantelaube and Terres, 2005; Marletto et al., 2005).

For a seasonal climate forecasting system to be successfully used for malaria prediction in Africa, it should be expected to forecast the seasonal cycles of both precipitation and temperature. It is particularly important to get the timing of the rainy season, both the start and end, well forecast. The rains start well before the onset of the malaria season, and an anomalous extension of the rainy season is likely to prolong the season of malaria transmission, increasing the risk of epidemic. However, in areas where for parts of the year the temperature is close or marginal to the threshold for the transmission of malaria, such as the area analysed in this paper, a skilful seasonal forecast of temperature becomes even more important.

As the grid points come from a region that has a distinct cool period following the rains, it is worth reviewing the performance of the forecasting system first as a tier-1 approach for the meteorological variables. However, given the non-linear interaction between precipitation and temperature, a tier-2 approach is required for the MTSM prevalence. The area studied in this paper is on the verge of the extratropical band, where seasonal precipitation predictability tends to be lower than in tropical areas, and when dealing with summer precipitation, it must be remembered that it is mainly convective in origin, which makes it more difficult to forecast. Temperature was skilfully forecast for all the forecast windows, for the seasons, associated with the thermally driven parts of the disease cycle, thus from the rainfall onset to after the cessation of the rains. Interestingly, the upper tercile category gained the highest level of forecast skill. Prevalence was skilfully predicted for both the Feb 2–4 and Feb 4–6 forecast windows, with the highest skill found in the above-the-median category. These two windows represent the rise of malaria prevalence in the MTSM and the peak of the prevalence in the MTSM. However, the prevalence forecast has no skill for Nov 4–6 (FMA), which would be the long-range forecast for the start of the season, while the temperature for the same forecast window is skilfully forecast. This indicates that, depending on the season, precipitation may play a more important role in prevalence prediction than temperature. However, the initial conditions may have a significant impact on the forecast and this needs further investigation in the future.

It is difficult to deconvolute the reasons for the model's ability to work over one 4–6 month forecast window and not another (e.g. Feb 4–6 and Nov 4–6). It is important to remember the non-linear nature of the MSTM and the possible role of initial conditions. Further, the biological development in the model lags behind the precipitation cycle and the precipitation from the preceding two or three months probably has a larger impact on the peak malaria prevalence than the concurrent precipitation. This is obviously a question for further examination. A further question that needs to be addressed is the ability to make skilful MTSM prevalence predictions, even though there was no skill in the precipitation forecasts. This may be due to the marginal temperatures for malaria transmission seen in the ERA-40 data for some of the grid points used in this paper and the ability of the forecasting system to skilfully predict upper tercile temperature for a range of forecast windows. It is possible that the MTSM model is oversensitive to temperature and undersensitive to precipitation. Further, the greater level of skill in certain forecast window categories found for the prevalence when compared with either of the driving variables is an interesting and important finding and will have important implications for the comprehensive validation of seasonal forecasting systems. In addition, it illustrates the relevance of a tier-2 verification of a

forecast system. A comprehensive forecast quality assessment requires an end-to-end approach to the forecasting problem and enlightens aspects of the forecasting system that are not obvious from a tier-1 perspective.

The results presented in this paper cannot be assumed to be universal, as different results may be found in different African regions or other parts of the world, or, obviously, other application models. Importantly, from the results in this paper there is no claim that there is a skilful forecast of actual malaria prevalence for this part of Africa, but rather that, should a more realistic MTSM be developed, there would be the possibility of issuing useful early warning of malaria epidemics. Both the malaria model and its integration within the DEMETER seasonal forecasting system are at the research and development stage.

5. Conclusions

The notion of a three-tier hierarchical approach for the validation of seasonal climate forecasts has been developed. The results discussed in this paper focus on tier-2 validation based on a realistic dynamic malaria model forced by global reanalysis data. The challenges of producing a realistic dynamic malaria model, which can be driven by seasonal forecasts, are substantial. The malaria model developed for the task needs to respond to a set of meteorological drivers changing through the season in a manner consistent with the known transmission dynamics and epidemiology of the disease. The MTSM is a first step in the direction towards this perfect malaria model and provides many insights to working in a probabilistic forecasting system. The seasonal forecasts of both precipitation and temperature that drive the application model needs to be skilful (i.e. more informative than climatology). The forecasts need to capture the onsets and cessation of seasonal cycles, particularly precipitation. The forecasts need to capture the interannual variability of these cycles, and attempt to capture gross features of the intrannual variability. The results presented in this paper are at 2.5° resolution, and thus the underlying land surface and its complexities are not clearly represented. The MSTM could be run in the future at higher spatial resolutions and modified to take the underlying land surface into account. In such a framework, the issue highlighted by Zhou et al. (2004), that not only did the climate variability strongly influence the malaria transmission but land use, topography and local microclimate all play a role, would have to be taken into account. This is not part of this study due to the low spatial resolution of the seasonal hindcast data.

These results show that there is potential for the skilful prediction of MTSM prevalence when driven by probabilistic seasonal forecasts with 2 to 4 and 4 to 6 month lead times through the start and peak of the simulated prevalence curve. The future ability to make predictions at these long lead times would have a substantial impact of planning activities and would allow the focusing of disease prevention activities. The above-the-median category has the greatest MTSM prevalence skill, and this represents a potentially useful forecast of conditions of above average malaria risk. The cost–loss curves for the upper tercile category indicate potential economic value over a wide range of cost–loss ratios.

A malaria model with further development and testing against actual epidemic transmission data sets, coupled to an ever-improving seasonal weather forecasting system, could eventually contribute to an operational seasonal malaria forecast that would be used as part of a MEWS, but it would be a tool to be used in conjunction with observational evidence and local knowledge. Furthermore, such a model would allow for a tier-3 validation.

Realistically, a 'perfect' seasonal forecast system will not be fully realized. Instead, future application models should assess the current 'skill-in-hand' for the forecasting system into which they are integrating their application model using either a tier-2 or tier-3 approach. Where possible, the application models should take maximum advantage of variables that are skilfully forecast for the region and forecast windows of interest. The forecast skill requirements of the application groups will help set the forecast targets for the meteorological seasonal forecast modelling community. It is likely that the most successful integrated modelling systems will emerge where there is a close working relationship between both seasonal forecasting and application groups, as has been attempted in the EU-funded DEMETER project and is to be continued in the EU-funded ENSEMBLES project (www.ensembles-eu.org).

6. Acknowledgment

The DEMETER project has been funded by the EU under the contract EVK2–1999-00197. The ECMWF staff and consultants are acknowledged for their invaluable cooperation during the constant development of the system. Madeleine Thomson is acknowledged for her support to the DEMETER project. Sandra Mather is acknowledged for the production of the figures. The authors wish to thank the three anonymous reviewers for their constructive comments on an earlier draft.

References

Anderson, R. M. and May, R. M. 1991. *Infectious Diseases of Humans – Dynamics and Control*. Oxford University Press, Oxford, 757.

Archer, E. R. M. 2003. Identifing underserved end-user groups in the provisions of climate information. *Bull. Am. Meteorol. Soc.* **84**, 1525–1532.

Cantelaube, P. and Terres, J. M. 2005. Seasonal weather forecasts for crop yield modelling in Europe. *Tellus* **57A**, 476–487.

Collins, W. E., Jeffery, G. M. and Roberts, J. M. 2004. A retrospective examination of the effect of fever and microgametocyte count on mosquito infection on humans infected with plasmodium vivax. *Am. J. Trop. Med. Hyg.* **70**, 638–641.

Detinova, T. S. 1962. *Age grouping methods in Diptera of medical importance*. World Health Organization Monograph 47, Geneva.

Dietz, K. 1988. Matematical models for transmission and control of malaria. In: *Malaria, Principles and Practise of Malariology*, Volume 2 (eds.W. H. Wernsdorfer, and I. McGregor). Churchill Livingstone, London, 1091–1133.

Doblas-Reyes, F. J., Pavan, V. and Stephenson, D. B. 2003. Multi-model seasonal hindcasts of the NAO. *Clim. Dyn.* **21**, 501–514.

Doblas-Reyes, F. J., Hagedorn, R. and Palmer, T. N. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. *Tellus* **57A**, 234–252.

Franz, K. J., Hartmann, H. C., Sorooshian, S. and Bales, R. 2003. Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado River Basin. *J. Hydrometeorol.* **4**, 1105–1118.

Hagedorn, R., Doblas-Reyes, F. J. and Palmer T. N. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concepts. *Tellus* **57A**, 219–233.

Hartmann, H. C., Pagano, C., Sorooshian, S. and Bales, R. 2003. Confidence builders – evaluating seasonal climate forecasts from user perspectives. *Bull. Am. Meteorol. Soc.* **83**, 683–698.

Hay, S. I., Cox, J., Roger, D. J., Randolph, S., Stern, D. I. and co-authors. 2002. Climate change and the resurgence of malaria in the East African highlands. *Nature* **415**, 905–909.

Hoshen, M. B. and Morse, A. P. 2004. A weather-driven model of malaria transmission. *Malaria Journal* **3**, 32, (doi:10.1186/1475-2875-3-32; http://www.malariajournal.com/content/3/1/32).

Jepson, W. F., Moutia, A and Courtois, C. 1948. The malaria problem in Mauritius: the bionomics of Mauritian anophelines. *Bull. Entomolo. Res.* **38**, 177–208.

Jolliffe, I. T. and Stephenson, D. B. (eds) 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, New York, 240 pp.

Kovats, R. S., Bouma, M. J., Hajat, S., Worrall, E. and Haines, A. 2003. El Niño and health. *Lancet* **362**, 1481–1489.

MacDonald, G. 1957. *The Epidemiology and Control of Malaria*. Oxford University Press, Oxford.

Marletto, V., Zinoni, F., Criscuolo, L., Fontana, G., Marchesi, S. and co-authors. 2005. Evaluation of downscaled DEMETER multi-model ensemble seasonal hindcasts in a northern Italy location by means of a model of wheat growth and soil water balance. *Tellus* **57A**, 488–497.

Mason, S. J. 2004. On using climatology as a reference strategy in the Brier and ranked probability skill scores. *Mon. Wea. Rev.* **132**, 1891–1895.

Palmer, T. N. 2002. The economic value of ensemble forecasts as a tool for risk assessment: from days to decades. *Q. J. R. Meteorol. Soc.* **128**, 747–774.

Palmer, T. N., Alessandri, A, Andersen, U., Cantelaube, P., Davey, M. and co-authors. 2004. Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER). *Bull. Am. Meteorol. Soc.* **85**, 853–872.

Pfaff, A., Broad, K. and Glantz, M. 1999. Who benefits from climate forecasts? *Nature* **397**, 645–646.

Piper, K. P., Hayward, R. H., Cox, M. J. and Day, K. P. 1999. Malaria transmission and naturally acquired immunity to PfEMP-1. *Infection and Immunity* **67**, 6369–6374, (http://iai.asm.org/cgi/content/full/67/12/6369).

Potgieter, A. B., Everingham, Y. L. and Hammer, G. L. 2003. On measuring quality of a probabilistic commodity forecast for a system that incorporates seasonal climate forecasts. *Int. J. Climatol.* **23**, 1195–1210.

Poveda, G., Rojas, W., Quiñones, M. L., Vélez, I. D., Mantilla, R. I. and co-authors. 2004. Coupling between annual and ENSO time-scales in the Malaria-Climate Association in Colombia. *Environmental Health Perspectives* **109**, 489–493, (http://ehp.niehs.nih.gov/members/2001/109p489-493poveda/poveda.pdf),.

Richardson, D. S. 2000. Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **126**, 649–667.

Richardson, D. S. 2001. Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Q. J. R. Meteorol. Soc.* **127**, 2473–2489.

Snow, R. W., Craig, M., Deichmann, U. and Marsh, K. 1999. Estimating mortality and morbidity and disability due to malaria among Africa's non-pregnant population. *World Health Organization Bulletin* **77**, 624–604, (http://whqlibdoc.who.int/bulletin/1999/Vol77-No8/bulletin_1999_77(8)_624-640.pdf),.

Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S. and Balmaseda, M. A. 1998. Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature* **392**, 370–373.

Taylor, L. H. and Read, A. F. 1997. Why so few transmission stages? Reproductive restraint by malaria parasites. *Parasitol. Today* **13**, 135–140.

Thomson, M. C., Palmer, T., Morse, A. P., Cresswell, M. and Connor, S. J. 2000. Forecasting disease risk using seasonal climate predictions. *Lancet*, **355**, 1559–1560.

Thornes, J. E. and Stephenson, D. B. 2001. How to judge the quality and value of weather forecast products. *Meteorol. Appl.* **8**, 307–314.

Uppala, S., Kållberg, P., Hernandez, A., Saarinen, S., Fiorino, M. and co-authors. 2004. ERA-40: ECMWF 45-yr reanalysis of the global atmosphere and surface conditions 1957–2002. *ECMWF Newsletter* **101**, 2–21, (http://www.ecmwf.int/publications/newsletters/list.html).

Wilks, D. 1995. *Statistical Methods in the Atmospheric Sciences.* Academic, London.

World Health Organization. 2001. *Malaria Early Warning Systems: Concepts, Indication and Partners*. World Health Organization, Geneva.

Yarnal, B., Comrie, A. C., Frakes, B. and Brown, D. P. 2001. Developments and prospects in synoptic climatology. *Int. J. Climatology* **21**, 1923–1950.

Zhou, G., Minakawa, N., Githeko, A. and Guiyun, Y. 2004. Association between climate variability and malaria epidemics in the East African highlands. *PNAS* **101**, 2375–2380, (http://www.pnas.org/cgi/doi/10.1073/pnas.0308714100),.

Zhu, Y., Toth, Z., Wobus, R., Richardson, D. and Mylne, K. 2002. The economic value of ENSEMBLES-based weather forecasts. *Bull. Am. Meteorol. Soc.* **83**, 73–83.