Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

Deliverable Reference Number and Title

D6.21: Report on the forecast quality of seasonal predictions of malaria incidence in Botswana over the period 1982-2005

Due date of deliverable: Month 60 (August 2009)
Actual submission date: January 13, 2010

Start date of project: 1 September 2004 Duration: 60 Months

Organisation name of lead contractor for this deliverable: IRI
Stephen Connor, Ashley Curtis, Simon Mason, Barbara Platzer, Madeleine Thomson, Mike Tippett

Revision [draft 1]

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

Abstract

The ability of Stream 2 ENSEMBLES models to predict climate variability that is known to be related to annual variability in malaria over Botswana is tested. The ENSEMBLES multi-model is shown to have no clear difference in predictive skill compared to that for DEMETER models, but an ability to distinguish between high- and low-malaria years remains evident. More detailed analyses of the influence of observed climate variability on malaria incidence in Botswana were conducted. Evidence for district-scale climate signals is weak, as is an influence of seasonal temperature variability on malaria incidence, and so the ability of the ENSEMBLES models to predict these aspects was not considered. Real-time forecasts were communicated to the region during the duration of the project.

Background/Aims and Objectives

- To test the ability of the DEMETER and ENSEMBLES models to predict seasonal climate variability over Botswana as a means of estimating the annual incidence of malaria in the country.
- To investigate the possibility of producing downscaled malaria predictions at district level.
- To investigate the influence of temperature on malaria incidence in Botswana.

Data

Malaria incidence data

Botswana has compiled national annual data on cases of laboratory-confirmed malaria incidence for the period 1982–2007. 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 modelled to remove non-climate trends, largely attributed to chloroquine resistance 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 variability. For this analysis, the 2.5°-resolution monthly data from the Climate Prediction Center Merged Analysis of Precipitation (CMAP) 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 (DJF) 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 a quadratic model.

DEMETER and ENSEMBLES climate predictions

The three models used in this study comprise the operational DEMETER and ENSEMBLES 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 for the DEMETER models and 1959-2005 for the ENSEMBLES models. According to the set of available start dates, predictions for NDJ and DJF (months one to four of the simulations) were used. Multi-model ensemble predictions were averaged over Botswana in the same manner as the CMAP data.

Methods

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, which is highly correlated with CMAP. This requirement for a calibration period also limits the number of coupled models available from DEMETER and ENSEMBLES 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 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 very 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.

Testing the coupled model predictions

The maps below (Fig. 1) show DEMETER model forecasts of the averaged departure (in mm/day) of Dec-Jan-Feb rainfall from its normal value over central and southern Africa for years with highest (1988, 1989, 1993, 1996, 1997) and lowest (1982, 1983, 1987, 1992, 2002) malaria annual incidence in Botswana. These years were selected on the basis of observed malaria incidence adjusted for known vulnerability trends, which are climate-independent.

The forecasts, made on the first of the preceding November, come from the DEMETER and ENSEMBLES stream2 forecast models of ECMWF, Meteo France, and Met Office (nine ensemble members per model). The results for the ENSEMBLES multi-model are shown following (Fig. 2), as well as for the individual models (Fig. 3). Normal rainfall was taken as the average of the period 1982-2002. The key result is that below-normal rainfall occurs on average for the low malaria years, and that above-normal rainfall occurs on average for the high malaria years, but that the models are only successful in predicting dry conditions during low malaria years.

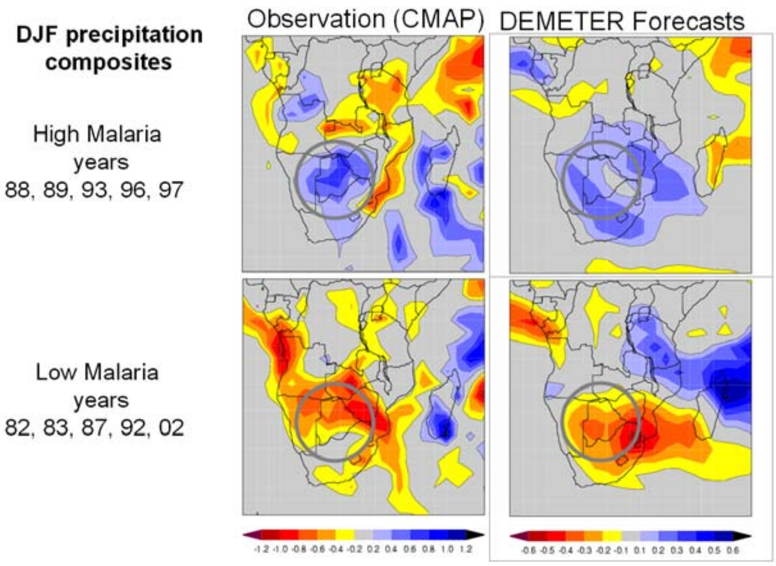

Fig. 1: Observed and multi-DEMETER-model average rainfall for high and low malaria years.

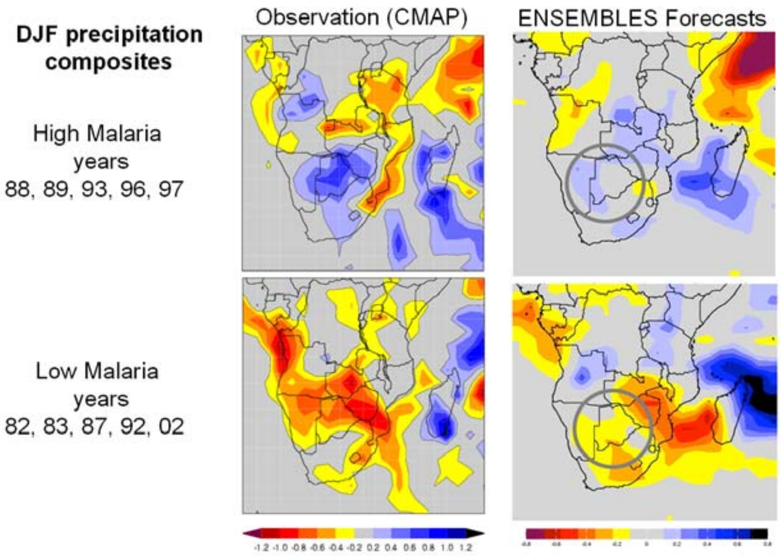

Fig. 2: Observed and multi-ENSEMBLES-model average rainfall for high and low malaria years.

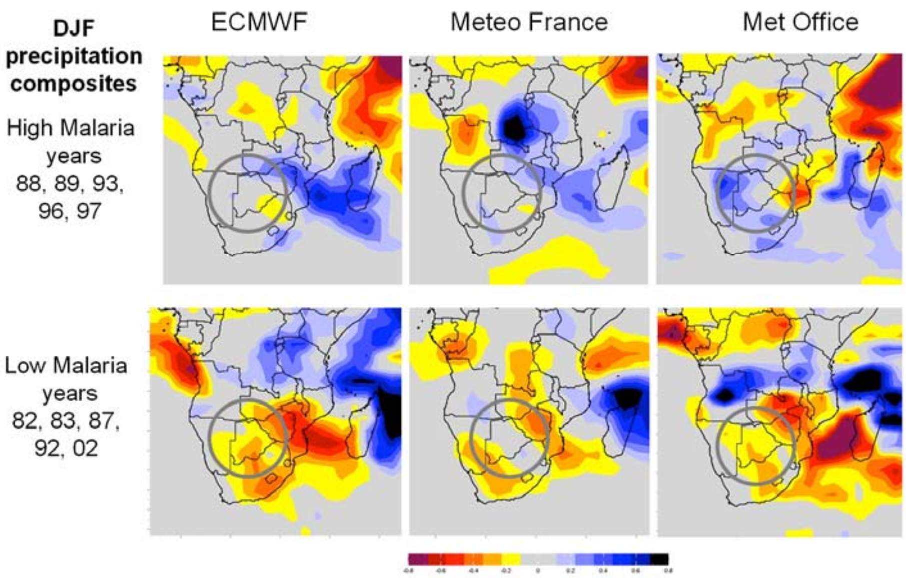

Fig. 3: Observed and individual ENSEMBLES-model average rainfall for high and low malaria years.

Additional evidence for predictive skill is provided by the correlation between the ensemble means and observed (CMAP) DJF rainfall as shown in the Table below.

	Pearsons	Spearmans
ECMWF	0.356	0.430
Meteo France	0.229	0.199
Met Office UK	0.503	0.507
Multi-model average	0.420	0.418

A Poisson regression was developed between the number of confirmed Botswana malaria cases 1982-2006 and the following covariates (predictors):

- ENSEMBLES DJF rainfall forecast Nov 1;
- the square of ENSEMBLES DJF rainfall Nov 1 forecast;
- a binary variable which is zero for years up to and including 1996;
- the year;
- the product of the last two variables.

The ENSEMBLES rainfall forecast is an area-average of the multimodel ensemble mean. Annual population is treated as an exposure. Unlike the regression analysis using observed (CMAP) rainfall, we find that neither the forecast rainfall, nor its square, is a significant covariate.

The figure below shows the observed number of confirmed cases / 1000 population and the regression fit values (along with 95% confidence intervals) for the regression model with and without rainfall as a covariate. The confidence intervals of the model with ENSEMBLES rainfall as a predictor are generally wider than those of the model without rainfall as a predictor; the means are similar. We note that the Poisson regression implicitly corrects systematic rainfall forecast errors. A likely explanation for the poor performance of the regression model based on ENSEMBLES forecast rainfall is the fact that the correlation between observed and forecast rainfall is modest, 0.44, though typical of the skill of seasonal precipitation forecasts. Also shown on the figure below are results from a similar analysis with DEMETER forecasts. The correlation of the DEMETER rainfall forecasts with observed rainfall (1982- 2002) is 0.28 which is not significant at the 95% level; the correlation between ENSEMBLES rainfall forecasts and observations during the same period is 0.32, a slight improvement, but also not significant.

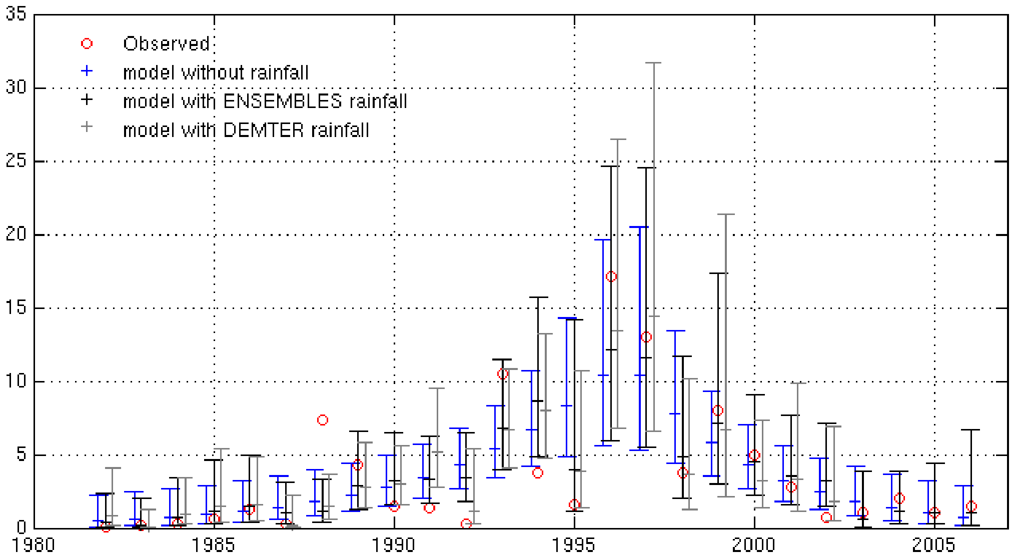

Fig. 4: Observed and modeled malaria incidence.

Downscaling

The numbers of confirmed malaria cases in four districts (Boteti, Chobe, Gantsi and Tutume) for the period 1982-2003 were fit to a Poisson regression model including the following covariates:

- observed DJF rainfall;
- the square of observed DJF rainfall;
- a binary variable which is zero for years up to and including 1996;
- the year;
- the product of the last two variables.

The regression model's parameters represent climate factors as well as trends and policy interventions. Estimated annual population is treated as an exposure. For each district, the model has the form:

number of cases = population Exp (a + b rainfall + c rainfall^2 + d (year<1996) + e year + f (year<1996) year)

and contains parameters that quantify the sensitivity of the number of cases to the covariates. The coefficient estimates are:

	a	b	c	d	e	f
Boteti	-271.78	6.64	-0.89	247.75	0.13	-0.12
Chobe	-227.21	2.10	-0.26	476.15	0.11	-0.24
Gantsi	-603.41	3.73	-0.48	1874.36	0.30	-0.94
Tutume	-631.69	5.41	-0.76	1058.23	0.31	-0.53

with standard errors:

Boteti	7.3	0.15	0.022	27	0.004	0.013
Chobe	4.1	0.06	0.009	13	0.002	0.007
Gantsi	29.0	0.32	0.048	72	0.015	0.036
Tutume	7.1	0.09	0.014	18	0.004	0.009

The magnitude of the standard errors would suggest that the districts display significantly different sensitivities to both the climate and non-climate factors. However, such a conclusion is not strictly correct since the covariates display considerable colinearity. To explore the idea that the four districts can be characterized by a single model, we fit a Poisson regression model to the aggregated district data. This aggregate model has coefficients

```
-357.04  3.61   -0.48   703.16   0.17   -0.35
```

with standard errors:

```
3.1     0.04   0.007    10   0.002  0.005
```

We characterize the enhancement in number of cases due to climate factors by the ratio

total number of cases / number of cases due to non-climate factors

The number of cases due to non-climate factors is obtained from the Poisson regression model with the climate variables (rainfall and rainfall^2) replaced by their climatological (long-term average) values. Therefore the number of cases due to non-climate factors reflects the trend and intervention policy. The figure below shows the base-10 logarithm of the observed, district modeled, aggregate modeled climate enhancement ratios for the four districts. Both the district and aggregate modeled enhancement ratios are well correlated (r values in figure titles) with observed ones indicating responses to a common climate signal. However, in two of the districts, Boteti and Tutume, the district models showed an advantage in squared error over the aggregate model, suggesting that improved downscaled predictions may be made for these districts.

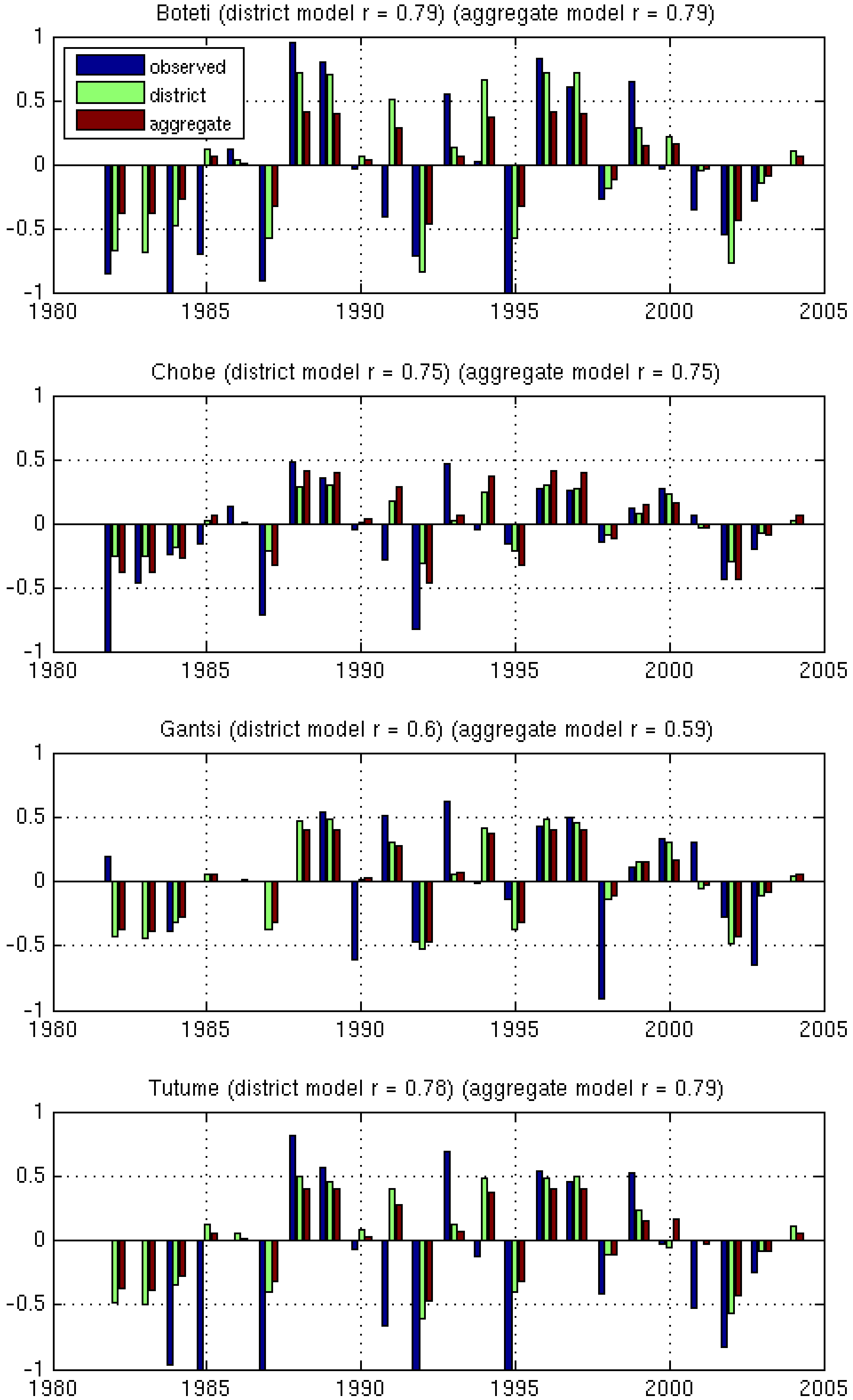

Fig. 5: Observed and modeled log malaria incidence anomalies.

Temperature as an additional predictor

A statistical model using both observed temperature and rainfall data over Botswana, 1982-2002, was used to assess the impact of temperature on the malaria model. Adding JFM temperature (observed) as a predictor did not result in a more skillful model. This is seen using Akaike's information criterion (AIC), which essentially indicates that adding temperature as a predictor does improve the fit, but that the improvement is due to over-fitting. Also, the prediction performance in the independent period (2003 on) is poor. Since observed temperature was a poor predictor, there is no reason to consider forecast temperatures from the coupled models.

A Poisson regression was developed between the number of confirmed Botswana malaria cases 1982-2007 and the following covariates (predictors): observed DJF rainfall, the square of observed DJF rainfall, a binary variable which is zero for years up to and including 1996, the year, and the product of the last two variables. Annual population is treated as an exposure. The resulting Poisson regression model is:

number of cases = population Exp(-353.31 + 3.60 rainfall - 0.48 rainfall^2 + 743.6391 (year>1996) + 0.17 year - 0.37 year (year>1996))

Adding observed JFM temperature as a covariate yields the Poisson regression model:

number of cases = population Exp(-385.26 + 3.83 rainfall - 0.50 rainfall^2 + 736.33 (year>1996) + 0.17 year - 0.37 year (year>1996) + 0.12 temp)

Adding temperature as a covariate reduces the Akaike information criterion (AIC), an approximation of the predictive error of the model. The reduction of AIC would seem to suggest that adding temperature improves the regression model. However, temperature and rainfall have a strong negative correlation (-0.84) and the resulting colinearity may make AIC inappropriate for model selection. Therefore leave-one-out cross-validation was used to estimate the predictive error of the two models more directly. Including temperature as covariate increases the cross-validated error, and therefore temperature is not a useful factor in the model.

Another way of seeing the negative impact of using temperature in the model is seen when the two models are fit to data up to 2002 and then applied to 2003-2007 data. The figure below shows the observed and modeled incidence rates per 1000. The period in the shaded region is independent of that used to train the model, and the performance of the model with temperature as a covariate is poor there.

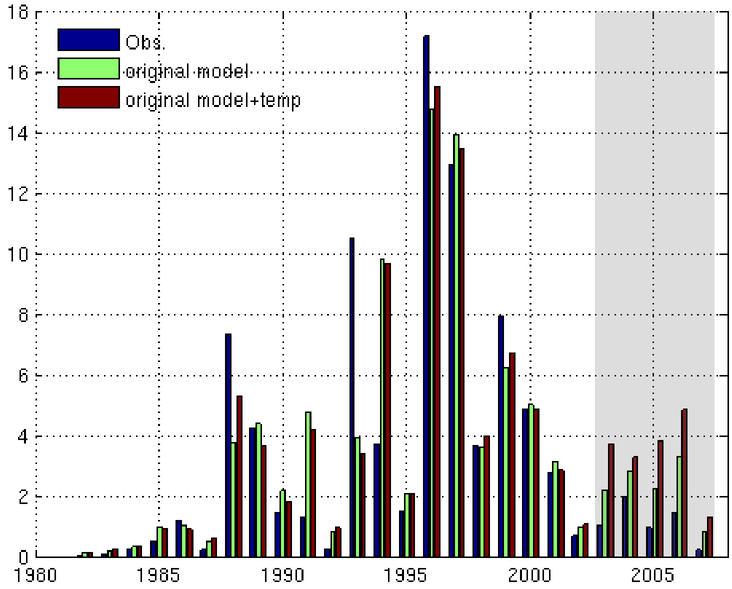

Fig. 5: Observed and modeled malaria incidence.

Example of a Real-time Forecast

In November 2008 a consensus seasonal forecast for Southern Africa was created, based largely on forecasts from the EuroSIP suite, and drawing upon research results initially conducted under the DEMETER project. The forecast was targeted specifically at advising the health community in the region on the prospects of above-normal rainfall over the 2008/09 rainfall season, largely in response to the cholera epidemic affecting Zimbabwe, and threatening surrounding countries. The forecast was targeted also at malaria control. The La Nina-type conditions at the time, and the canonical teleconnection over Southern Africa of above-normal rainfall was leading to concerns about a climate-induced exacerbation of disease outbreaks in the region. The forecasts were forwarded directly to the SADC Drought Monitoring Centre (DMC), and to the Eastern and Southern Africa Malaria Control (ESAMC).

Summary

The ENSEMBLES multi-model was shown to have no clear difference in predictive skill compared to that for DEMETER models, but an ability to distinguish between high- and low-malaria years remains evident. The correlation between the multi-model means and observed rainfall is slightly stronger for the ENSEMBLES models than for the DEMETER models, but the differences in the composite rainfall for high- and low-malaria years are less. Evidence for district-scale climate signals on malaria incidence in Botswana is weak: two out of four districts showed some evidence for local modification of large scale climate effects. There is no evidence for an influence of seasonal temperature variability on malaria incidence in Botswana. The co-variance between temperature and rainfall in the country is very strong and so temperatures were found not to provide significant additional predictive skill. ENSEMBLES models continue to be useful tools for predicting climate related variability in malaria in Southern Africa and steps should be taken to ensure ongoing inputs to real-time forecasting initiatives in the region.