The development and prudent application of climate-based forecasts of seasonal malaria in the Limpopo province in South Africa

Willem A Landman[a], Neville Sweijd[b], Nyakallo Masedi[c], and Noboru Minakawa[d]

a. Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa (Corresponding author: Willem.Landman@up.ac.za)

b. Alliance for Collaboration on Climate and Earth Systems Science, Council for Scientific and Industrial Research, Cape Town, South Africa

c. Department of Health, Limpopo Province, Malaria Control Program, Polokwane, South Africa

d. School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan

ABSTRACT

Seasonal Climate Forecasting (SCF) in South Africa has a history spanning several decades. During this period a number of SCF systems have been developed for the prediction of seasonal-to-interannual variability of rainfall and surface temperatures. Areas of highest predictability, albeit relatively modest, have also been identified. The north-eastern parts of South Africa that includes the Limpopo province has been demonstrated to be one of the areas of highest SCF skill in the country. Statistical post-processing techniques applied to global climate model output were part of this forecast system development, and were subsequently successfully used in the construction of forecasts systems for applications in sectors which are associated with ENSO-driven climate variability, such as dry-land crop yields and river flows. Here we follow a similar post-model processing approach to test SCF systems for application to the incidence of seasonal malaria in Limpopo. The malaria forecast system introduced here makes use of the seasonal rainfall output fields of one of the North American Multi-Model Ensemble (NMME) climate models, which is then linked statistically through multiple linear regression to observed malaria incidence. The verification results as calculated over a 20-year hindcast period show that the season of highest malaria incidence forecast skill is during the austral mid-summer time of

December to February. Moreover, the hindcasts based on the NMME model outscore those of statistical forecast models that separately use Indian and Pacific Ocean sea-surface temperatures as predictors, thus justifying the use of physical global climate models for this kind of application. Additional results indicate that model skill levels may include quasi-decadal variability, that the periods over which forecast verification is performed strongly influences forecast skill, and that poorly predicted malaria seasons may have serious financial implications on public health operations.

Key words: Malaria incidence; seasonal forecasts; South Africa; financial implications

1. Introduction

Advanced seasonal forecast modelling has been taking place globally and over a continuous period of time (e.g. Doblas-Reyes et al., 2013). Such modelling progress has also occurred in South Africa (e.g. Beraki et al., 2015) with an accompanying improvement in seasonal forecast skill. This improvement is a result of increased forecast model complexity (Landman, 2014). Modelling collaboration and predictability studies over southern Africa have also taken place recently in partnership with other nations (Ratnam et al., 2016, 2018; Yuan et al., 2014). In essence, modelling and associated predictability research have clearly demonstrated the existence of seasonal rainfall and temperature forecast skill over parts of the southern African region (Landman et al., 2012, 2014; Lazenby et al., 2014). Notwithstanding, seasonal forecast skill over southern Africa is modest compared to most regions globally (Landman et al., 2019), but within the region there are certain areas where predictability is relatively high. Such an area of relatively high seasonal predictability includes the northeastern parts of South Africa. However, the relatively high predictability is mainly during the austral summer months for both rainfall and temperatures (Landman et al., 2012; Lazenby et al., 2014). Areas where and seasons when forecasts have been found to be skillful were then considered in the development of statistical applications models. These models skillfully predicted inflows into major dams (Muchuru et al., 2016) and end-of-season crop yields (Malherbe et al., 2014).

The focus of this paper in on developing and testing malaria forecasts on the seasonal-to-interannual time scale as opposed to multi-decadal time scales. Notwithstanding, modelling on

the latter time scale has found a global net increase in climate suitability for malaria owing to climate change (Caminade et al., 2014; Ngarakana-Gwasira et al, 2016). Given that changing seasonal characteristics of climate are what comprise climate change at decadal scales, short-term seasonal prediction in near real-time assume greater importance and is relevant to the management of the disease which operates on seasonal to annual time scales. Both rainfall and temperature anomalies are considered to be drivers of seasonal-to-interannual variability of malaria incidence over parts of Africa (Thomson et al., 2005, 2018; Kovats, 2000). Southern African climate variability is influenced by the El Niño-Southern Oscillation (ENSO) phenomenon (e.g. Mason and Jury, 1997), and there is strong evidence that La Niña (El Niño) events are frequently associated with an increased (decreased) occurrence of malaria cases in the region (Mabaso et al., 2007). Moreover, tropical Indian Ocean sea-surface temperatures also play an apparent role in southern African rainfall variability (Washington and Preston, 2006) as well as influencing malaria prevalence (Behera et al., 2018). ENSO is nevertheless a strong forcing factor for climate variability over southern Africa and may be the main source of seasonal climate predictability over the region (Landman and Beraki, 2012). In addition, the strength of El Niño events matters since strong events usually result in deficit rainfalls during summer as opposed to that of moderate to weak events (Pomposi et al., 2018). Examples of this notion are the drought related impacts caused by the strong El Niño of 2015/16 (Archer et al., 2017). Similarly, the major flooding event of 2010/11 over the region is associated with a strong La Niña event (Muchuru et al., 2014). During La Niña (El Niño) years, southern Africa also tends to be cooler (warmer) than usual (Lazenby et al., 2014).

In areas of high or mountainous land malaria incidence may increase with higher temperatures associated with El Niño events, as occurred in association with heavy rainfall events in highland Rwanda and Uganda (Kovats, 2000). Since such atypical weather events (e.g. unseasonal rainfall inundation during predominantly hot and dry seasons) could cause epidemics, generalizing the effect of ENSO on epidemic malaria is onerous (Kovats et al., 2003). To further illustrate the notion of atypical weather influences, the unprecedented increased rainfall in Tanzania during the 1997 El Niño event (equatorial east Africa usually experience anomalously wet conditions during El Niño events) likely flushed away mosquito breeding sites and consequently lowered the

expected malaria incidence (Kovats et al., 2003). Similarly, over southern Africa the very wet austral summer period of 2000/01 resulted in malaria incidence lower than expected for wet seasons. Rainfall inundation could therefore occasionally result in less malaria incidence than expected (Thomson et al., 2005).

The use of climate models in SCF systems are our best chance of making skillful seasonal rainfall and temperature forecasts, and to some extent forecast production of commodities and ecosystem goods and services (and threats to these) associated with climate variations such as dry-land crops and riverflows. However, these climate models cannot include non-climatic influences such as, for example, the introduction of insecticide spray for malaria control (Thomson et al 2005) and therefore could have limited application for malaria prediction. Early warnings suffer from the inherent limitation in that a warning that evokes a proactive intervention may result in successful avoidance of the pertaining threat (e.g. a response intervention to the warning includes intensified anti-malarial Indoor Residual Spraying campaign which obviates the predicted epidemic) resulting in a seemingly "false positive" prediction. Moreover, if inadequately robust climate forecast models or signals are unavailable, as may often be the case, then observed statistics may actually provide the outputs for sufficiently early warning of an outbreak (Drake et al., 2019). Notwithstanding these difficulties, and although predicting malaria incidence using climate information is not new (Adeola et al., 2019; Thomson et al., 2006, 2018), here we ask to what extent the use of a contemporary global climate model can in fact predict seasonal malaria incidence; for which seasons and lead-times such forecasts work best; and what the cost-effectiveness of the forecasts may be.

2. Data

2.1. Malaria data

Provincial monthly malaria data of the total number of reported cases of human infection were obtained from the Department of Health of the Limpopo Province of South Africa. The period used in the analysis is the 20-year period from 1998/99 to 2017/18. Three- and four-month seasons of accumulated malaria cases (infections) are calculated as well as an accumulated total for all but the austral winter months of minimal malaria prevalence, i.e. June to August, in order

to obtain a 9-month accumulated incidence total for September to May (for each year). The result is a set of fourteen seasonally accumulated values for each chronological year. The fourteen malaria seasons considered are JFM, FMA, MAM, SON, OND, NDJ, DJF, JFMA, FMAM, SOND, ONDJ, NDJF, DJFM and Sep-to-May, since the austral winter months of June to August have minimal recorded malaria cases and are therefore excluded in the analysis. When the Lilliefors (Wilks, 2011) goodness-of-fit test is applied to the fourteen seasonal data values, they are found not to come from a distribution in the normal family, except for the NDJF 4-month season. In order to ensure that all of the seasonal malaria values are from normal distributions for optimal statistical modelling to be performed, the natural logarithms of the malaria values for each season are calculated. This approach to malaria case data transformation has been successfully applied before over southern Africa (Thomson et al., 2005). Subsequent Lilliefors tests show that the newly derived seasonal malaria incidence values are indeed from undefined normal distributions. Henceforth, seasonal malaria incidence refers to the data which have been transformed by calculating log values.

2.2. Climate model data

The North American Multi-Model Ensemble (NMME; Kirtman et al., 2014) experiment provided the output of the global climate prediction model used in the seasonal malaria forecast system. In particular, hindcasts from the GFDL-CM2p5-FLOR-B01 (referred to here as "GFDL") are available from March 1980 to present, for 12 ensemble members and for 11 lead-time months. In this study, only three lead-times are considered. Hence, the GFDL hindcasts for DJF (December-January-February) were respectively initialized in November (a 1-month lead-time), in October (a 2-month lead-time) and in September (a 3-month lead-time). Only a restricted area (30° to 15° South; 20° to 50° East) of the GFDL seasonal rainfall data is used as predictor to develop statistically-based malaria models since inclusion of the area over the southwestern Indian Ocean in downscaling models has already been proven useful (Landman and Goddard, 2005; Landman et al., 2014).

Monthly hindcast data of the GFDL model are available at a 1° × 1° latitude–longitude resolution. This model resolution is sufficient for modelling climate variability over our relatively small

provincial region of interest, since there is only minimal benefit to be obtained by increasing atmospheric resolution in seasonal climate forecast models (Scaife et al., 2019).

2.3.   Sea-surface temperatures

Observed sea-surface temperature (SST) data are used for the development of baseline statistical malaria prediction models against which the GFDL-based models (see section 3.1) are to be compared. Extended reconstructed SST from NOAA NCDC ERSST version4 is used in this study (Huang et al., 2015) and converted into seasonal averages. Two SST regions are considered for the baseline models, one over the Pacific (10° South to 10° North; 170° to 80° West) and the other over the Indian Ocean (45° South to the equator; 30° to 120° East). These ocean regions have been found to be strongly related to seasonal rainfall and malaria incidence variations over southern Africa (Behera et al., 2018; Mabaso et al., 2007; Mason and Jury, 1997; Reason, 1999; Pomposi et al., 2018; Washington and Preston, 2006).

3. Prediction models

3.1.   GFDL-based downscaled model

Two GFDL output variables have initially been considered, namely seasonally averaged maximum temperatures and seasonal total rainfall. These variables have been predicted in hindcast mode by the climate model. Hindcasts are typically model integration over recent decades, and for the purpose of this study are used to investigate forecast model performance over these decades (Troccoli et al., 2008). The initial model testing proved that using maximum temperatures of the model as predictors show very little promise for Limpopo's seasonal malaria prediction as opposed to the use of the model's rainfall fields as predictors. Only climate model rainfall fields as specified over the domain that covers parts of southern Africa and the southwestern Indian Ocean are subsequently considered for the development of statistically-based malaria models. These malaria models are produced by statistically downscaling seasonal rainfall fields of the GFDL to concurrent seasonal malaria incidence (hereafter referred to as GFDL-based models). The canonical modes of the rainfall hindcasts as calculated by the Climate Predictability Tool (CPT; Mason and Tippett, 2016) are used in a multiple linear regression model as predictors. Three

lead-times are considered, and so typical configurations of malaria models may be designed as follows:

For a 1-month lead-time predicting JFM malaria incidence, the GFDL was initialized in December for a JFM rainfall hindcast set to be downscaled to JFM malaria incidence; for a 2-month lead-time predicting FMA malaria incidence, the GFDL was initialized in December for a FMA rainfall hindcast set, etc.

3.2.   Baseline models

Since SST anomalies over the central Pacific Ocean (i.e. Mabaso et al., 2007; Thompson et al., 2005) and the Indian Ocean have been found to be related to seasonal malaria over the region (e.g. Behera et al., 2018), the SST's of these ocean areas are used as predictors of seasonal malaria incidence in order to develop baseline models. As with the GFDL fields, the canonical modes of the respective SST fields are used as predictors in multiple linear regression models. However, the SST fields are not concurrent with the malaria seasons, but antecedent in order to constitute forecast models at various lead-times. For example, a 1-month lead-time forecast for DJF malaria incidence use observed ASO SST fields as predictors since ASO SST's only become available in early November.

4. Results
4.1.   Hindcast skill

First, the malaria models are tested over the 20 years of available data through a cross-validation design (Efron and Gong, 1983) for each of the 14 seasons as specified above. Figure 1 shows the correlation analysis between 1-year-out cross-validated (one year is withheld from the 20- year data and the remaining years are used to train the statistical models, after which a prediction is made for the omitted year; the procedure is repeated until a prediction has been made for each of the 20 years) malaria incidence predictions and observed seasonal malaria incidence, for each of the three lead-times considered. Three correlation values are calculated, including "ordinary" or Pearson correlation values, and two robust and resistant alternatives to Pearson, namely the Spearman rank correlation coefficient, and Kendall's tau which is also a measure of rank correlation (Wilks, 2011). Kendall's tau has close likeness to ROC scores (relative operating
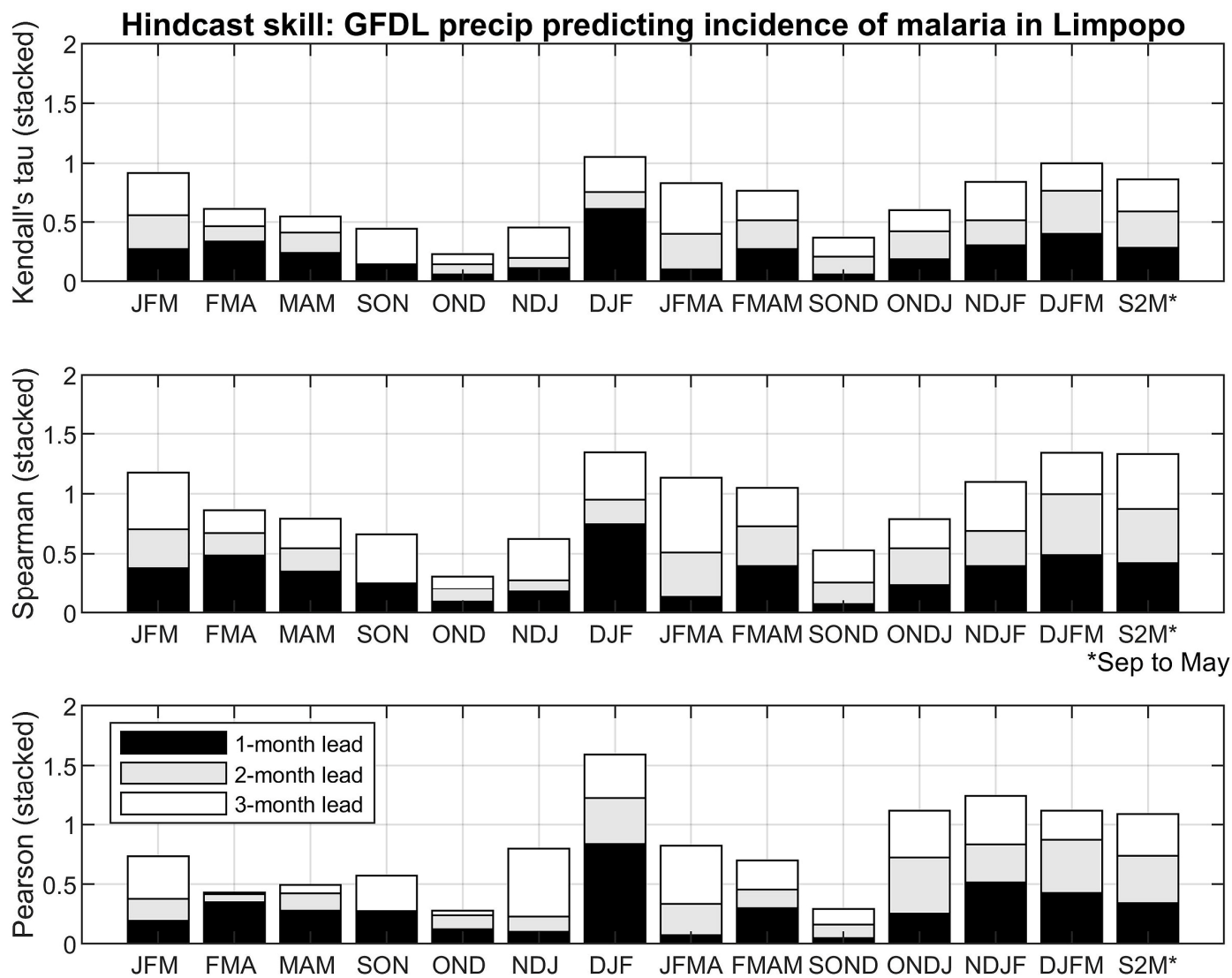
Fig. 1. Correlations between cross-validated hindcasts and observed malaria incidence for the seasons indicated. Forecast lead-times up to three months are presented.

characteristic; Mason and Graham, 2002) to be discussed later on in the paper on verification of probabilistic malaria forecasts. All three correlation coefficients show that the DJF season is the most predictable, and at a 1-month lead-time (i.e., forecasts to be issued in November for DJF).

Possible auto-correlation (correlation of the malaria data with itself) may lead to artificial skill. However, a multi-year-out cross-validation window may cater for this problem. Owing to the short malaria data record of only 20 years, large cross-validation windows may negatively impact on the robustness of the prediction equations and subsequently on model performance. Lag-1 auto-correlation values (Pearson) are subsequently calculated and found to be low enough to disregard (none are statistically significant at the 95% level), therefore justifying single year-out cross-validation models. Moreover, Figure 2 shows that the difference in correlation (skill) of most multi-year-out cross-validation models for DJF is insubstantial compared to the 1-year-out model, further justifying the use of this 1-year-out configuration.

The four best models for malaria cases predictions according to the rank correlation values of Figure 1 are for DJF, JFM, DJFM and S2M (the 9-month season of September to May). The 1-year-out cross-validation results of these four seasons at a 1-month forecast lead-time are shown in Figure 3. Along with the GFDL-based model results, the 1-year-out cross-validation Kendall's tau correlations from using respectively the equatorial Pacific and Indian Ocean SST as malaria season predictors are also presented in the figure. The models that use SST as predictors are considered here as the baseline models that need to be outscored by the GFDL-based models in order to justify the use of such physical global climate models for malaria prediction. The superiority of the GFDL-based malaria models is demonstrated by the fact that the GFDL-based models clearly outscore (higher correlations and statistical significance) the SST-based forecast models for the most part (except for JFM malaria predictions when similar skill levels are found when using either Pacific Ocean SST or the GFDL model). Take note that neither of the two modelling approaches could capture the excessively high malaria season of 2014/15. The consequence of such a forecast miss of an extreme season is discussed in more detail later on.

Further consideration of the cross-validation results of Figure 3 suggests that the best hindcasts appear to be during the middle section of the 20-year period. Such quasi-decadal variation in
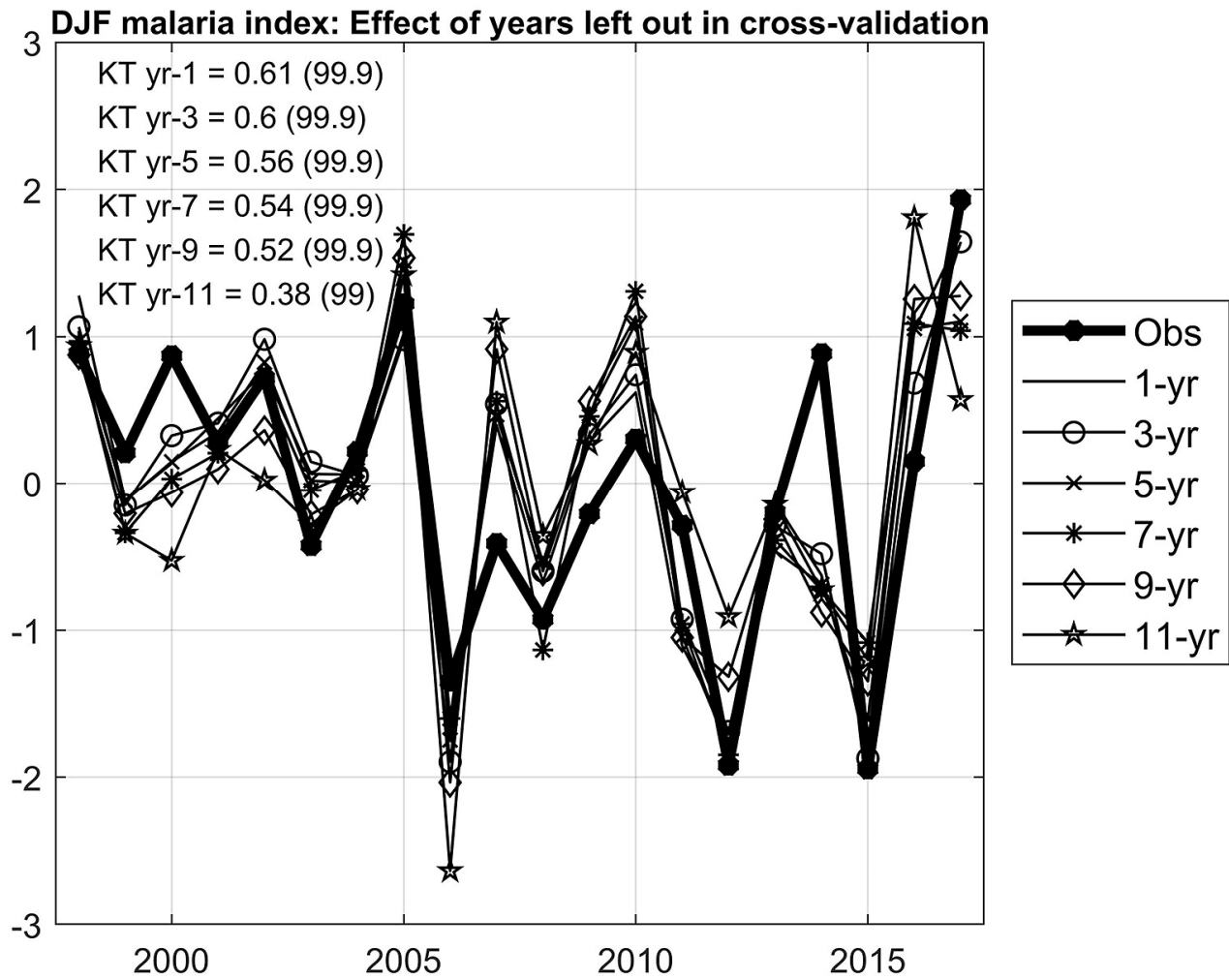
Fig. 2. Cross-validated hindcasts for DJF with various leave-years-out models. The Kendall's tau (KT) correlation for each statistical model is shown (e.g. yr-3 means a 3-year-out cross-validation window is used). Statistical significance is given in parentheses.
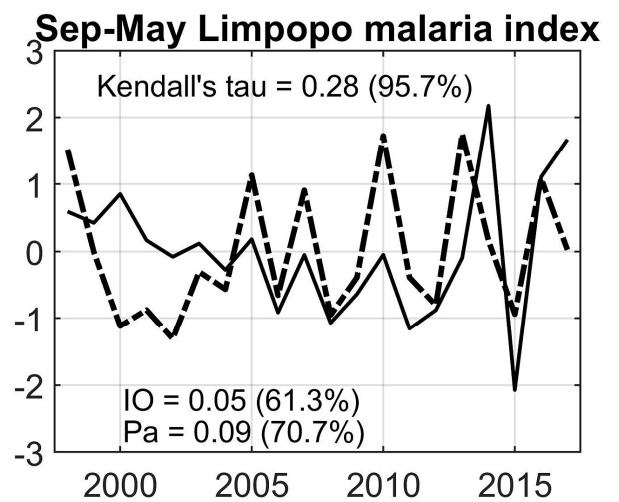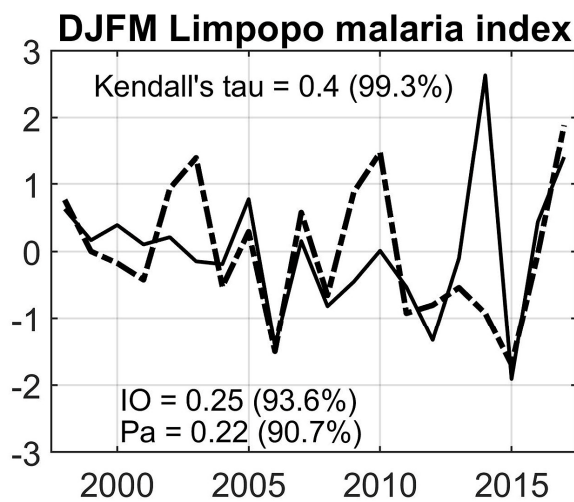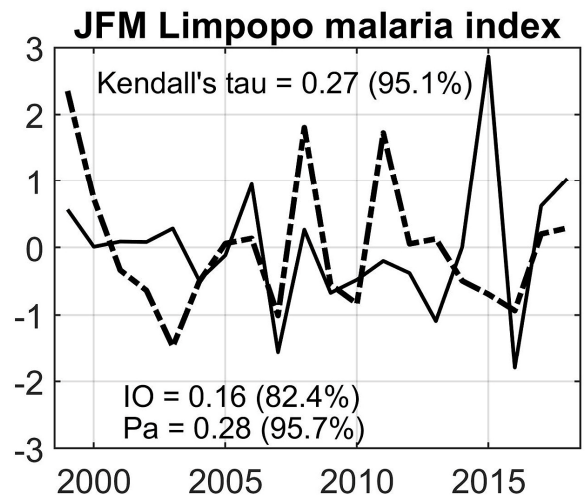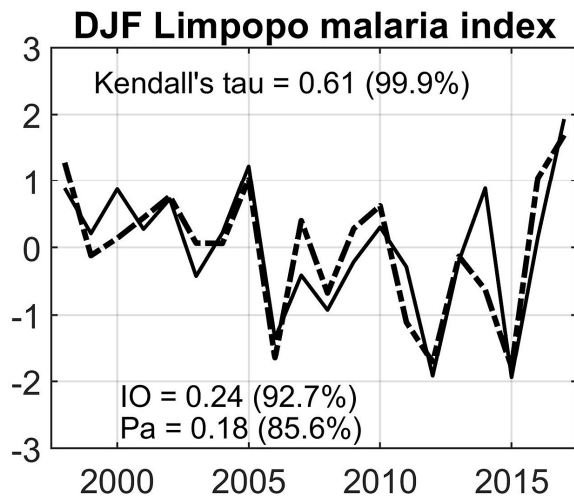
Fig. 3. Cross-validated malaria incidence hindcasts (1-month lead) for the four seasons selected using GFDL-based rainfall data as the predictor (chart). Time series have been normalized. The top value on each panel is the Kendall's tau (Kt) correlation between hindcast (dashed) and observed (black). Data provided below include IO: Kt when using Indian Ocean SST as predictor; Pa: Kt when using equatorial Pacific Ocean SST as predictor. Statistical significance is given as a % in parentheses.

skill, associated with the frequency of observed ENSO events, has been found before for southern African seasonal rainfall modelling (Landman and Goddard, 2002), and needs additional investigation here. Regarding the 20-year test period as 3 sub-periods of 6 earlier years, 7 middle years and 6 later years, it is noteworthy that the middle years are associated with 7 ENSO events, while the earlier and later years are associated with 2 and 1 ENSO-neutral years respectively (neither El Niño nor La Niña). Seasonal rainfall over southern Africa during ENSO-neutral years have been found to be poorly predicted (Landman and Beraki, 2012), and hence this notion explains the higher predictability of rainfall-driven malaria years when ENSO is most active.

Figure 4 respectively shows 5- and 10-year moving windows of the cross-validated Pearson correlations of the results presented in Figure 3. The DJF season has not only been found to be associated with the highest skill (Figure 3), but in Figure 4 we see that DJF is also the season during which forecast skill remains the most consistent as opposed to the other three seasons that show big variation in forecast skill. Since predicting DJF malaria incidence at a 1-month lead-time using GFDL forecasts in the downscaling model has been shown here to be the best case for successful malaria prediction, the remainder of the paper will focus on this seasonal malaria forecast model. Although skillful DJF forecasts have also been found here at longer lead-times (Figure 1), we henceforth focus on 1-month lead-time DJF forecasts in order to demonstrate operational forecast utility and to identify potential caveats in a skillful forecast system.

4.2.   Probabilistic forecast skill

Seasonal climate forecasts are best expressed probabilistically due to the inherent variability of the atmosphere (weather noise) and the hitherto inability to perfectly replicate all the components adequately in climate models. This requirement is especially true for southern Africa that ranks poorly relative to other parts of the Globe in terms of seasonal rainfall forecast skill (Landman et al., 2019). Nevertheless, probabilistic forecasts exhibit considerably higher reliability in comparison with those achieved by corresponding deterministic forecasts (Murphy 1998). As with other climate services products, probabilistic forecasts must also be assessed in terms of the value of the forecast from a decision-makers perspective to determine its true utility. The value of the interpretation may be best achieved through comprehensive verification of probabilistic
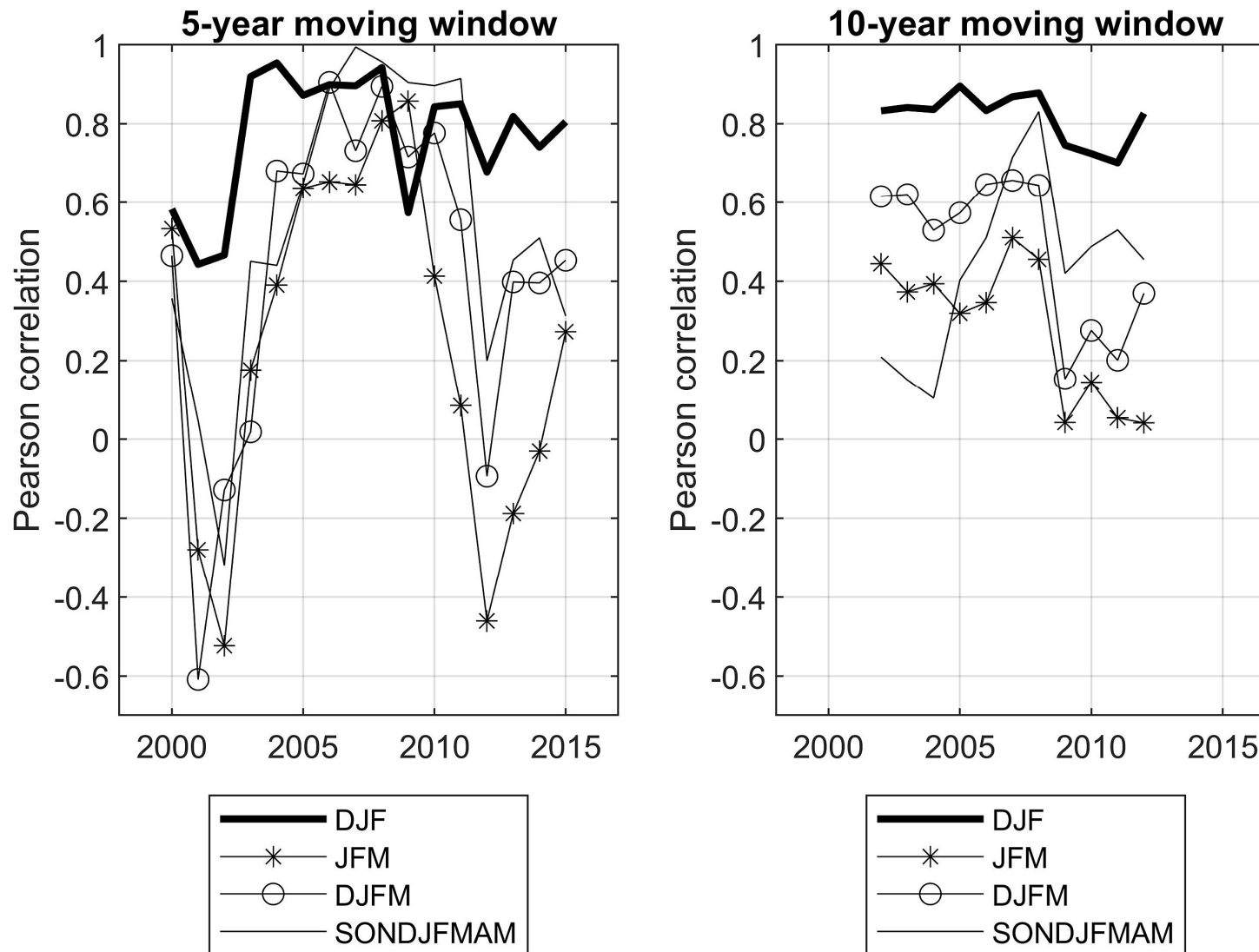
Fig. 4. Respectively 5- (left hand panel) and 10-year (right hand panel) moving windows are used to calculate cross-validated correlations between observed and hindcasts for the four seasons indicated.

forecasts in order to guide decision-makers on which forecast systems are likely to be informative and, more importantly, establish under which circumstances forecast systems are likely to provide either optimal or unreliable information.

The manner in which probabilistic seasonal forecasts are produced here is through the calculation of error variances of cross-validated GFDL ensemble mean downscaled values (Troccoli et al., 2008). The error variances are then used in the creation of so-called retro-active probabilistic forecasts of malaria incidence: a model is trained using only the first part of the data after which a prediction is made for the year immediately after the end of the training period. The model is then re-trained by adding the year just predicted to the training period in order to predict for the following year. This procedure is repeated until a prediction for the last year of the data has been made. Such an approach has been used successfully over southern Africa for a range of probabilistic forecast applications (Archer et al., 2019; Landman et al., 2012; Malherbe et al., 2014; Muchuru et al., 2016).

For the malaria data provided for this paper, the earlier years of the 20-year period are used to train the statistical downscaling model through cross-validation in order to produce probabilistic malaria forecasts for the later years in a stepwise fashion, e.g., the first 10 years are used to train the model in order to predict year 11, followed by 11 years to train the model to predict year 12, etc., until 19 years are used to train the model to predict for year 20. However, this approach only allows for the creation of 10 years of probabilistic forecasts. In order to obtain 20 years of probabilistic forecasts, the 20 years of data are divided into two chronological 10-year periods. Probabilistic forecasts for the second 10-year period are obtained as described above, but for the first 10-year period the two 10-year periods are swapped around, with the result that the second 10-year period is used to assist in the production of probabilistic forecasts for the first 10-year period, again in a stepwise fashion. This process results in the production of 20 years of probabilistic forecasts that can be verified and also compared with the verification results of the two 10-year periods. Figure 5 shows both the original cross-validated hindcasts of the DJF malaria predictions (top panel) and the cross-validated hindcasts obtained by placing the second 10 years of the 20-year period ahead of the first 10 years (bottom panel). Take note that he cross-validation Pearson correlations are identical for both cases presented in the figure.

Original DJF malaria hindcasts

Corr=0.84

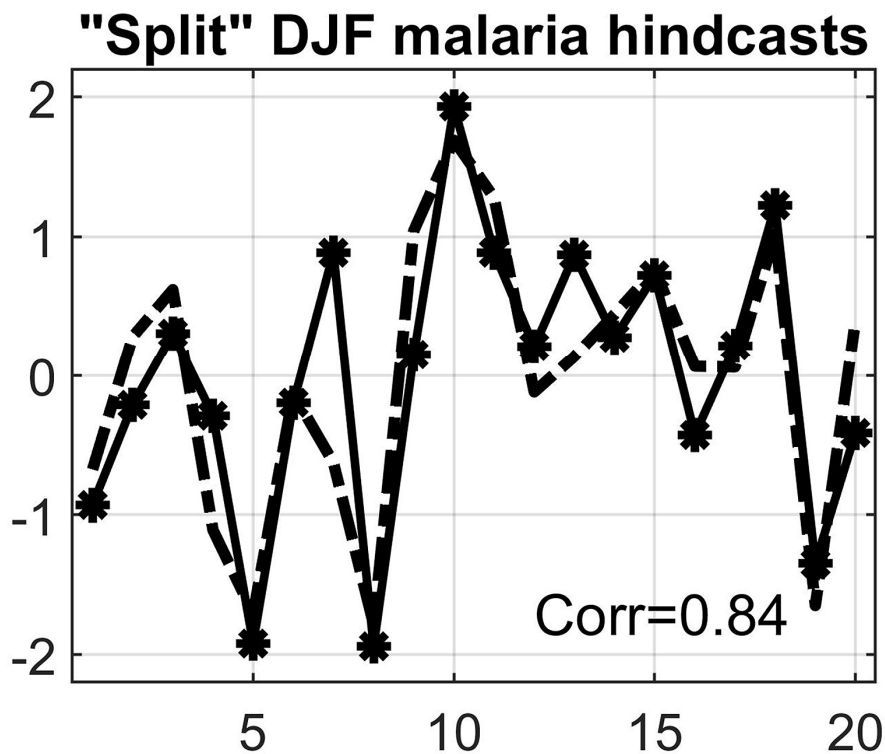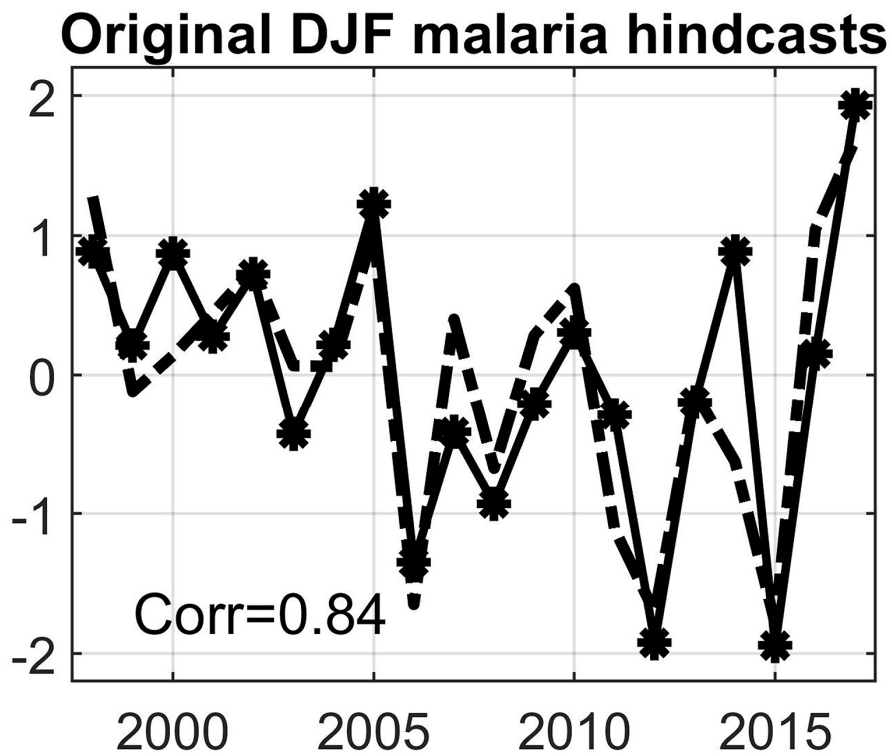## "Split" DJF malaria hindcasts

Corr=0.84

Fig. 5. Top panel shows the DJF hindcasts from Fig. 3, whereas the bottom panel shows the cross-validated hindcasts by placing the second ten years ahead of the first ten years of the 20-year period. The black asterisked lines represent the observed values, while the dashed lines represent the hindcasts.

Health practitioners as decision-makers should be able to derive benefit from knowing whether or not a forecast system has the attribute of discrimination when a season of anomalously high or low malaria incidence is predicted. By discrimination we mean that the forecast system is able to discriminate seasons of high malaria incidence from the rest of the seasons, or is able to discriminate seasons of low malaria incidence from the rest of the seasons. For this attribute, we calculate ROC scores over the two 10-year periods and over the 20-year period, for three equi-probable (33.3% each) categories of predicted above-normal, near-normal and below-normal malaria incidence. ROC scores represent the areas beneath the ROC curve that are produced by plotting the forecast hit rates against the false alarm rates for each forecast category. If a forecast model has no skill, the ROC area would be ≤0.5; for a maximum ROC score of 1.0, perfect discrimination is the result. We also calculate the Brier skill score (BSS; when forecasts are compared with a reference forecast such as a forecast of climatology) for each category, since the Brier score (BS; Wilks, 2011) is analogous to the better known mean-squared error value (although the BS is defined in terms of the "error" in the probabilities of the forecasts). To conclude the verification for the categories separately, we also present an indication on the reliability of the forecasts. The forecasts are considered reliable if there is consistency between the predicted probabilities of the defined malaria incidence categories (below-, near-, and above-normal, i.e. low, near-average and high) and the observed relative frequencies of the observed malaria incidence being assigned to these categories. Here, for the sake of brevity, we only present the resolution slopes of the reliability diagramme (Hamill, 1997): for perfectly reliable probability forecasts, the resolution slopes need to have a value of one. Finally, we present scores that encompass all of the three forecast categories collectively, namely the generalized ROC score (GROC; Mason and Weigel, 2009) that also shows the degree of correct forecast discrimination. The advantage of this score is that it was designed to indicate forecast quality to the general public: the expected score of a no-skill set of forecasts (random guessing or perpetually identical forecasts) is 50%. The GROC score can be interpreted as an indication of how often the forecasts are correct.

Figure 6 presents the probabilistic forecast verification statistics for DJF at a 1-month lead-time as determined over the whole 20-year period, as well as separately for the two 10-year periods.
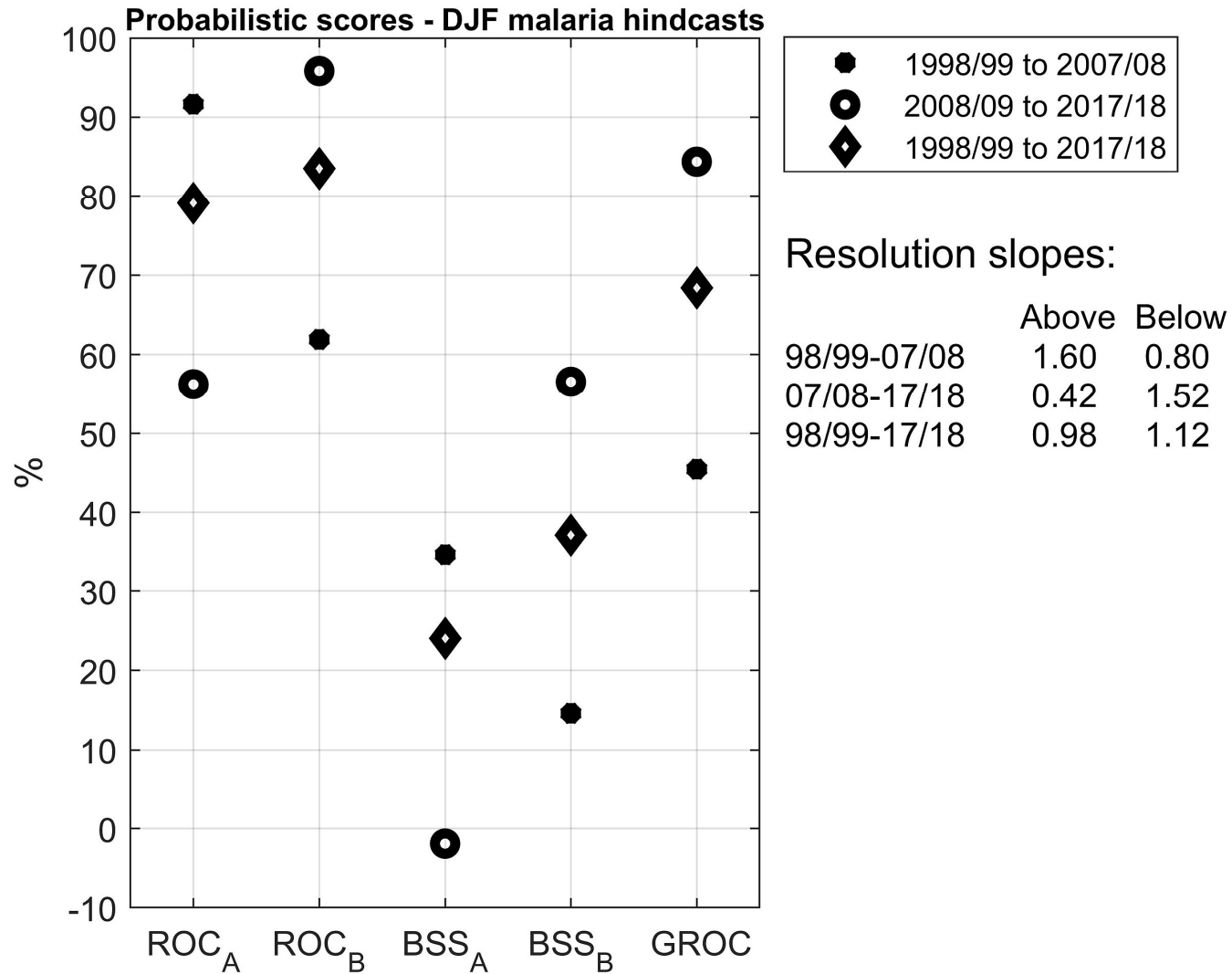
Fig. 6. Verification scores (ROC: relative operating characteristics; BSS: Brier skill score; GROC: generalized ROC) for the two 10-year and full 20-year periods. The "A" ("B") symbol refers to the above- (below-) normal category.

Take note that the verification scores for individual categories (i.e., ROC and BSS) are presented here only for above- and for below-normal malaria incidence, since predicting the near-normal category usually results in low skill values (Landman et al., 2012; Van den Dool and Toth, 1991). The first 10-year period (1998/99-2007/08) has more above-normal incidence years and the malaria model gets higher ROC and BSS values for above-normal forecasts as opposed to below-normal forecasts. Conversely, the second 10-year period (2008/09-2017/18) has more below-normal incidence years and the malaria model gets higher ROC and BSS values for below-normal forecasts as opposed to above-normal forecasts. The verification period may therefore strongly influence perceived forecast skill. Although about equal ROC scores are found for these two outer categories when considering the whole 20-year period, the result shows the importance of verification over a long as possible period as opposed to short periods where particular outcomes may be directed as a consequence of, for example, a high frequency of ENSO-related low incidence years. The instability in forecast skill is also found for the GROC scores (encompassing all three forecast categories) that show higher values for the second 10-year period. The resolution slopes for the two categories are close to unity (0.98 and 1.12 respectively) only for the 20-year period, indicating a strong association between forecast probabilities and observed frequencies of malaria cases for each category when considering the whole period.

4.3.   Operational utility

Verification results presented above are for three categories of incidence as determined by pre-defined thresholds, since for the most part the general practice is to produce seasonal forecasts for these three categories. However, users (in this case, for example, malaria control officials working on malaria in the Limpopo province) may require additional detail about the forecast probability distribution that can only be obtained by considering the entire probability distribution. With the entire distribution, users may choose intersects of particular interest to them, and are therefore not limited to pre-determined tercile thresholds. A tool with which to examine the complete probability distribution is the probability of exceedance (PoE) forecast graph, and these are presented in Figure 7 for six DJF seasons (2012/13-2017/18). There are two types of curves presented on the figure. The think black curve represents the climatological probability distribution of DJF malaria incidence, and the remaining curves represent predicted
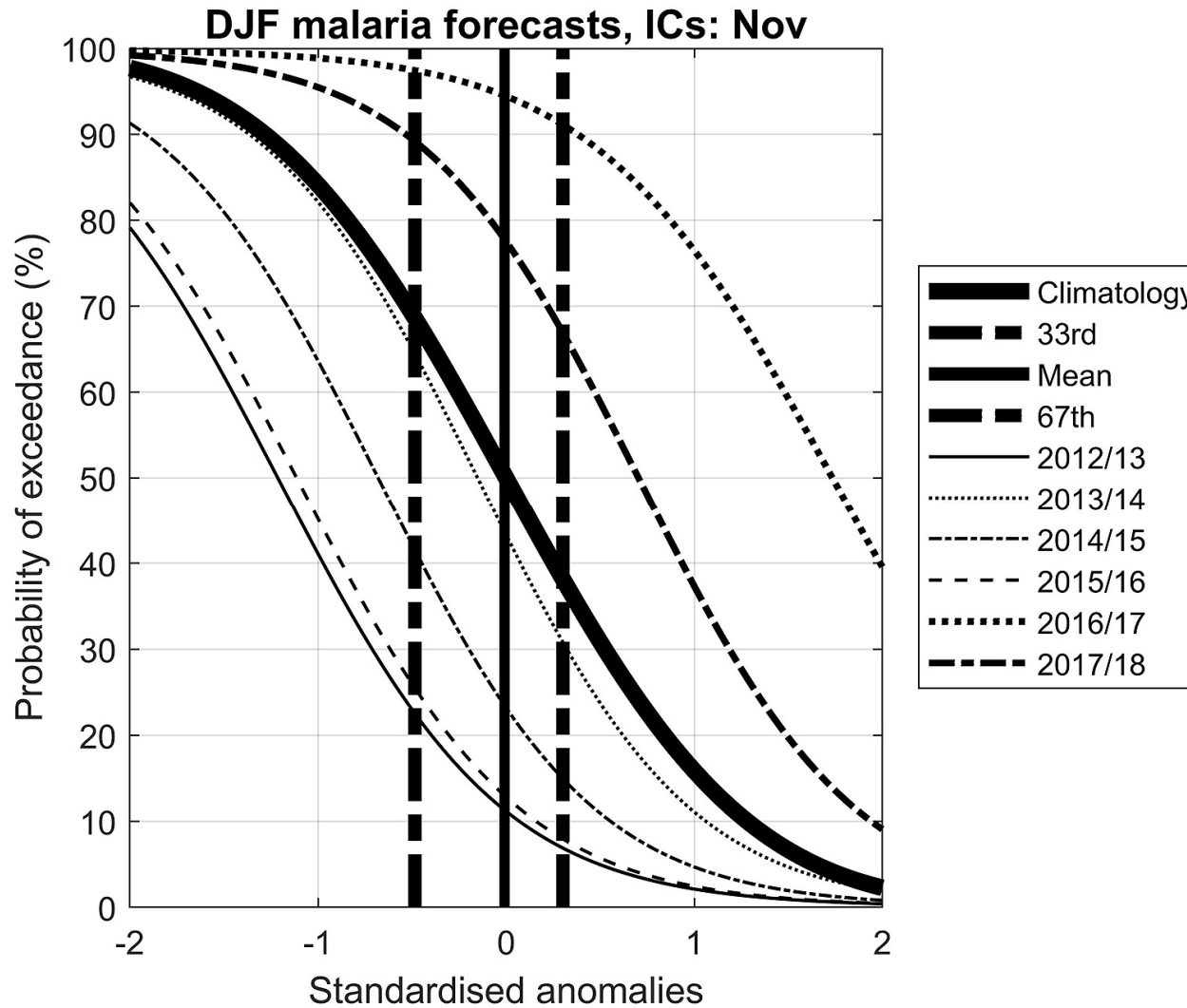
Fig. 7. Probability of exceedance malaria forecasts for the most recent six DJF seasons of the 20 year period. The vertical dashed lines respectively represent below- and above-normal thresholds. The thick black curve is a fitted theoretical distribution (Students t) based on climatology, while the remaining curves are fitted distributions of the forecasts.

probability distributions for the DJF seasons indicated on the figure. The vertical dashed lines represent the category thresholds. The forecasts curves are generated by training the DJF malaria model with 14 years of data (1998/99-2011/12) and then produce forecasts for the remaining six years of the data (2012/13-2017/18).

For the purpose of the discussion to follow, the easiest way to interpret the forecast curves of Figure 7 would be to consider those curves below the thick black climatological distribution curve to be probabilistic forecasts for anomalously low seasonal malaria incidence (2012/13, 2013/14, 2014/15 and 2015/16). Forecast curves above the climatological distribution curve are probabilistic forecasts for anomalously high seasonal malaria cases (2016/17 and 2017/18). Refer to Figure 3 (top left panel) to see the cross-validated deterministic forecasts and observed values for this 6-year period. In general, the forecasts seem to have been useful, except for 2014/15 when its forecast curve is predicted to be below the climatological curve, while the observed outcome is for malaria cases to be close to one standard deviation above average. Also, the 2016/17 DJF season turned out to be only slightly above the average, even though the forecast shows a large anomaly (Figure 3) and a forecast curve (Figure 7) associated with high probabilities of anomalously high malaria incidence.

4.4.   On financial implications for "bad" malaria forecasts

Of the six mid-summer malaria forecasts presented in section 4.3., two of them failed: the 2014/15 forecast underestimated the close to one standard deviation above-average incidence, and for 2016/17 the forecast overestimated actual malaria incidence. If these two predictions were used operationally, they may have had operational and financial consequences which would render the system untrustworthy from the point of view of users. Furthermore it can be argued that there may have been financial costs associated with these. Here we seek to demonstrate how these two failed forecasts could potentially have significant financial implications for the health industry, notwithstanding the fact that the malaria model used for these forecasts bear respectable verification statistics.

The verification scoring metrics presented above primarily focus on the general skill of the malaria forecast system. Because malaria incidence has associated cost implications, it may be of

value to health practitioners to reflect on the potential economic value of a malaria forecast system (Hagedorn and Smith, 2009). In order to investigate this further we make use of the calculation of cumulative profit (CP) values, which evaluate probabilistic forecasts by means of quantifying the skill of the forecast using an effective daily financial interest rate. In order to prepare for a malaria control response, some capital is invested into the first of a series of consecutive probabilistic forecasts (e.g. purchase of insecticides, employment of staff). Depending on the outcome of how well the forecast performs, a return is obtained on the investment. Here we calculate the CP values over the 20 years of probabilistic forecasts based on an arbitrary USD100 investment. The results shown in Figure 8 can be interpreted as follows: for the CP value of about 20 found for 2011/12 on the figure, means that an initial investment of USD100 in 1997 would be worth USD(100x20=)2000 in 2011/12. One would subsequently invest all USD2000 on the next year's forecast, and so forth. Although the profits from using the probabilistic forecasts are initially small, a steep rise in profits is observed from 2005/06 towards 2012/13. Take note of the significant drop in the accumulated profit in 2014/15 when the forecast of below average malaria incidence missed the (close to) one standard deviation above average outcome. Also, another significant decline in profit occurred during the 2016/17 DJF season when the forecast shows a strong positive anomaly while the observation turned out to only be a small anomaly, albeit positive. However, the skillful forecasts of 2015/16 and 2017/18 contributed to a recovery in profits based on forecast use. These simplistic examples where the forecasts provided poor guidance to a health practitioner that potentially could have led to the reduction of financial benefits demonstrate the danger that could be associated with using forecasts for decision-making in isolation. Take note that we did not take into consideration how the "false alarm" of the 2016/17 season may be the result of early warning intervention that could have taken place. Given this information, a careful cost-benefit analysis (accounting for more than merely the profit/saving's gains and losses, but also for the long term benefits of reduction of morbidity and mortality despite a 20% failure rate) should be considered.

5. Discussion and conclusions

The province in South Africa most strongly affected by malaria outbreaks is Limpopo. This province is located over the northeastern parts of South Africa and is in an area of relatively high
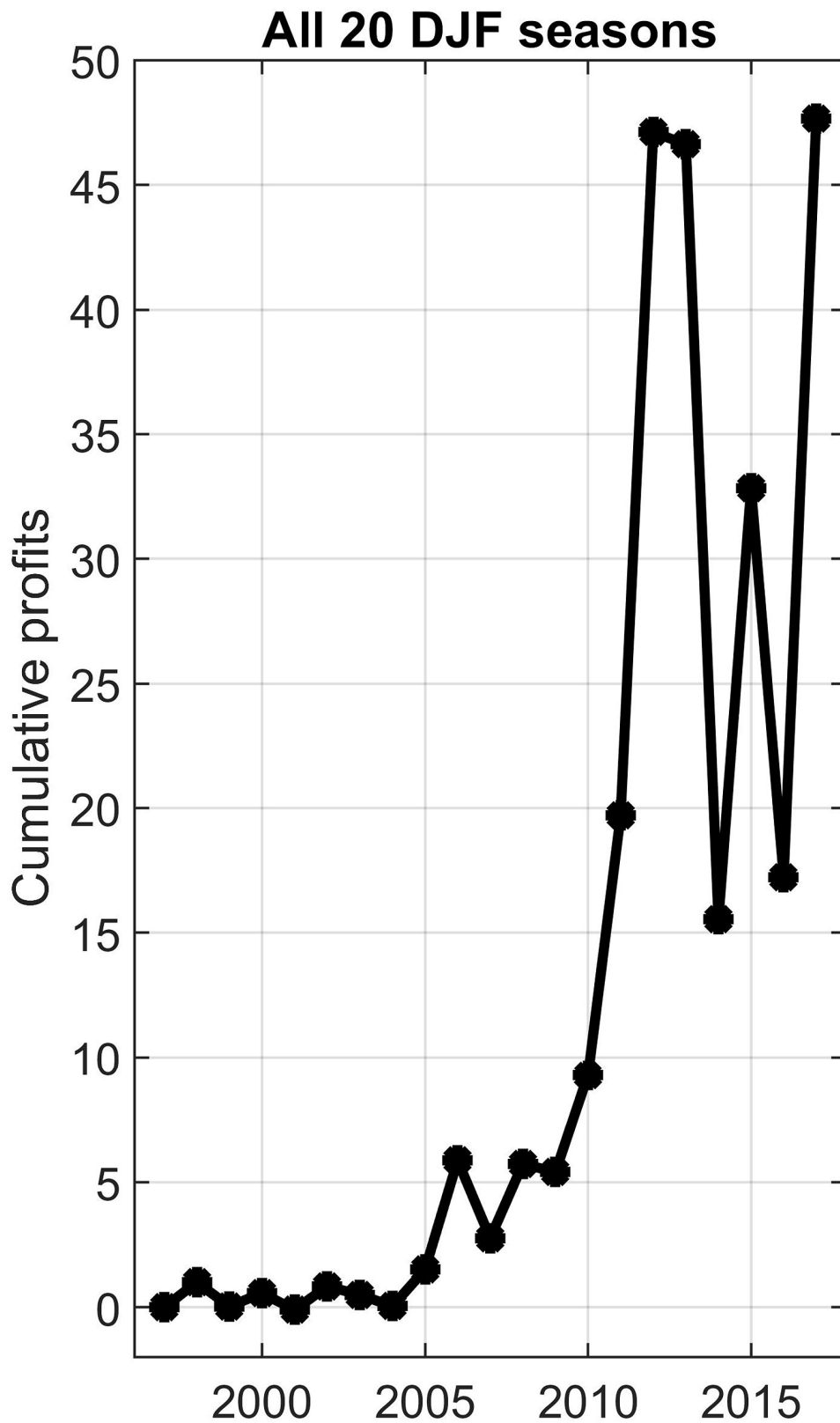
Fig. 8. Cumulative profits graph based on 20-years of probabilistic DJF malaria forecasts to communicate the monetary value of probabilistic forecasts. The higher the amount of capital placed on a forecast that is correct, the higher the profit or return will be.

seasonal rainfall and temperature forecast predictability when compared to the rest of the country (i.e. Landman et al., 2012; Lazenby et al., 2014). Since malaria incidence has a strong link with climate variables such as rainfall and temperature, using knowledge of how these variables may range among summer seasons when malaria is most prevalent, can help to predict the likelihood of a particular summer season being associated with high or low risk of a malaria outbreak in Limpopo. Here we proposed the development of a statistical interface that downscales large-scale forecast fields from a contemporary global climate model to observed seasonal malaria incidence in order to develop objective malaria forecast systems of which forecast output can be analyzed and verified.

Previous work has introduced the notion of using statistical means to relate physical global climate models to commodities such as dry-land crop yields (Malherbe et al., 2014) and river flows (Muchuru et al., 2016) over southern Africa. In order for this kind of application development to have been successful, phenomena to be predicted should contain a climate signal in the data (e.g. Muchuru et al., 2016), consistent time series have to have been archived over sufficiently long enough periods so that robust statistical relationship could be developed between the climate model output and what was observed on the ground over an extended period (Wilks, 2011), and some form of quality control of the observed data had taken place (Troccoli et al., 2008). Seasonal malaria incidence over southern Africa is linked to ENSO (e.g. Mabaso et al., 2007), therefore, the observed malaria data used here contain the required climate signal. The 20 years of monthly malaria data that were provided by the Limpopo Department of Health were proven here to be of sufficient duration for our modelling purpose. It was assumed that the malaria data have been adequately quality checked before dissemination. We selected two output variables, seasonal rainfall totals and seasonal mean maximum temperatures as predicted by a climate model, as the only two possible predictors of seasonal malaria incidence. This selection is based on the established link between these two variables and malaria incidence in Africa (Thomson et al., 2017). We found that the climate model's rainfall fields are a far superior predictor of Limpopo's seasonal malaria incidence and so all results presented here are based on predicted rainfall as a predictor of malaria cases (and not temperature). We used concurrently observed DJF rainfall as "predictor" of malaria in a multiple

linear regression model, and obtained a cross-validation Pearson correlation of 0.6, therefore further justifying the use of rainfall for malaria prediction.

Altogether fourteen multiple month annual periods or "seasons" are considered for malaria forecast model development, and the resulting cross-validation analysis shows that DJF is the season of highest skill of seasonal malaria incidence. The DJF season is found to be the most predictable since correlation analysis shows the best fit for this season between predicted and observed malaria (highest correlation values, especially at a 1-month lead-time), and DJF is the season least subject to quasi-decadal changes in forecast skill. The relatively high malaria predictability skill found for DJF may be related to contemporary coupled climate models' ability to best capture seasonal rainfall variability during DJF (e.g. Landman et al., 2014), and not necessarily that DJF malaria seasons are the most predictable in a general sense – hence model performance improvement may allow this system to be extended operationally to other seasons. We also compared the use of climate model output as predictor against the use of Indian and Pacific Ocean SST as predictors, and the climate model presented the highest skill levels. This result justifies the use of global climate models for seasonal malaria incidence prediction for Limpopo, which needs further exploration especially since models are constantly improving and collaborative international programs make model output data available for possible multi-model forecast system development which should further improve forecast skill.

Since operational seasonal forecasts have to be conducted probabilistically, owing to various sources of uncertainty and to the benefits to users when applying such forecasts, it is required that the DJF forecast system is evaluated in a probabilistic sense. We subsequently created 20-years of three equi-probable category probabilistic hindcast data for the DJF malaria season and performed verification analysis on the forecasts. The dangers of performing verification over very short periods are presented by showing how different conclusions from the same model can be drawn when verifying for different test periods. However, the DJF malaria forecast model is found to be skillful in general, as was the case for the cross-validated deterministic DJF forecasts. Then the DJF malaria model was used to predict the last six DJF malaria seasons of the 20-year period by using the first 14 years to train the model. The six forecasts are presented as full probability distributions (shown as a probability of exceedance curves in Figure 7). This format has the

advantage over categorized forecasts since the forecast curves show the probability of exceeding (or not exceeding) all possible thresholds across a range of values. The forecasts over these six seasons are then compared with the observed outcomes, and four of the six years are predicted accurately. The two seasons which fared poorly are respectively underestimating (2014/15) and overestimating (2016/17) the observed occurrences. Although a success rate of four out of six is at present a reasonable approximation of truest estimate for seasonal forecast success rates over southern Africa (see Landman et al., 2012), the two poorly predicted seasons, if utilised operationally, may have potentially resulted in some negative financial and material losses over the 20 year period. Hence we recommend that these model outputs are operationalized after a cost-benefit analysis.

We have developed and presented a robust and skillful seasonal forecast system for mid-summer malaria incidence over the Limpopo province. Although the malaria prediction model has been shown to be skillful, we have demonstrated how occasional false-positive and false-negative forecasts can result in material financial losses. Therefore we urge caution in relying on these forecast models exclusively for disease management to obviate loss of confidence in the forecast systems. We therefore suggest, as others have (Thomson and Mason, 2019), to make sure that good climate monitoring systems are in place to supplement forecasts from models. Moreover, we propose that disease surveillance and control activities should not be replaced by forecasts owing to its caveats (Holmes et al., 2018), but that forecasts should only be used to supplement health practices that are currently going on in the region.

NOAA/NCEP, NOAA/CTB, and NOAA/CPO jointly provided coordinating support and led development of the NMME-Phase II system.

References

Adeola, A.M., Botai, J.O., Olwoch, J.M., Rautenbach, H.C.J.deW., Adisa, O.M., de Jager, C., Botai, C.M., Aaron, M., 2019. Predicting malaria cases using remotely sensed environmental variables in Nkomazi, South Africa. Geospatial Health, 14:676, 81-91. doi:10.4081/gh.2019.676.

Archer, E., Landman W.A., Malherbe, J., Tadross, M. and Pretorius, S., 2019. South Africa's winter rainfall region drought: a region in transition? Climate Risk Management, accepted for publication. DOI: 10.1016/j.crm.2019.100188.

Behera, S.K., Morioka, Y., Ikeda, T., Doi, T., Ratnam, J.V., Nonaka, M., Tsuzuki, A., Imai, C., Kim, Y., Hashizume, M., Iwami, S., Kruger, P., Maharaj, R., Sweijd, N., Minakawa, N., 2018. Malaria incidences in South Africa linked to a climate mode in southwestern Indian Ocean. Environmental Development. 27, 47-57. https://doi.org/10.1016/j.envdev.2018.07.002.

Caminadea, C., Kovats, S., Rocklov, J., Tompkins, A.M., Morse, A.P., Colón-González, F.J., Stenlund, H., Martens, P., Lloyd, S.J., 2014. Impact of climate change on global malaria distribution. Proceedings of the National Academy of Sciences, 111, 3286–3291, https://doi.org/10.1073/pnas.1302089111

Efron, B., Gong, G., 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. The American Statistician, 37, 36–48.

Hamill, T. M., 1997. Reliability diagrams for multicategory probabilistic forecasts. Weather and Forecasting, 12, 736–741.

Holmes, E.C., Rambaut, A., Andersen, K.G., 2018. Pandemics: spend on surveillance, not prediction. Nature, 558, 180-182.

Huang, B., Banzon, V.F., Freeman, E., Lawrimore, J., Liu, W., Peterson, T.C., Smith, T.M., Thorne, P.W., Woodruff, S.D., and Zhang, H.-M., 2015. Extended Reconstructed Sea Surface Temperature

Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons. Journal of Climate, 28, 911–930. doi: http://dx.doi.org/10.1175/JCLI-D-14-00006.1

Kovats, R.S., 2000. El Niño and human health. Bulletin of the World Health Organization. 78, 1127—1135.

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

Landman, W.A. and Goddard, L., 2002. Statistical recalibration of GCM forecasts over southern Africa using model output statistics. Journal of Climate, 15, 2038-2055.

Landman, W.A. and Goddard, L., 2005. Predicting southern African summer rainfall using a combination of MOS and perfect prognosis. Geophysical Research Letters, 32, L15809, DOI: 10.1029/2005GL022910.

Landman, W.A., Barnston, A.G., Vogel, C. and Savy, J., 2019. Use of El Niño-Southern Oscillation related seasonal precipitation predictability in developing regions for potential societal benefit. International Journal of Climatology, DOI: 10.1002/JOC.6157.

Landman, W.A., Mason, S.J., Tyson, P.D. and Tennant, W.J., 2001. Retro-active skill of multi-tiered forecasts of summer rainfall over southern Africa. International Journal of Climatology, 21, 1-19 https://doi.org/10.1002/joc.592.

Landman, W.A., Beraki, A., DeWitt, D., Lötter, D. 2014. SST prediction methodologies and verification considerations for dynamical mid-summer rainfall forecasts for South Africa, Water SA, 40(4), 615-622, DOI: 10.4314/wsa.v40i4.6.

Landman, W.A., DeWitt, D. Lee, D.-E., Beraki, A. and Lötter, D., 2012. Seasonal rainfall prediction skill over South Africa: 1- vs. 2-tiered forecasting systems. Weather and Forecasting, 27, 489-501. DOI: 10.1175/WAF-D-11-00078.1.

Mabaso, M.L., Kleinschmidt, I., Sharp, B., Smith, T., 2007. El Nino Southern Oscillation (ENSO) and annual malaria incidence in Southern Africa. Transactions of the Royal Society of Tropical Medicine and Hygiene. 101 (4), 326–330.

Malherbe, J., Landman, W.A., Olivier, C., Sakuma, H. and Luo, J.-J., 2014. Seasonal forecasts of the SINTEX-F coupled model applied to maize yield and streamflow estimates over north-eastern South Africa. Meteorological Applications, 21, 733-742, DOI: 10.1002/met.1402.

Mason, S.J., Jury, M.R., 1997. Climatic variability and change over southern Africa: a reflection on underlying processes. Progress in Physical Geography. 21, 23-50.

Mason, S.J., Graham, N.E., 2002. Areas beneath the relative operating characteristics (ROC) and levels (ROL) curves: Statistical significance and interpretation. Quarterly Journal of the Royal Meteorological Society, 128, 2145–2166.

Mason, S.J. and Tippett, M.K., 2016. Climate predictability tool version 15.3. New York, NY: Columbia University Academic Commons. https://doi.org/10.7916/D8NS0TQ6.

Mason, S.J., Weigel, A.P., 2009. A generic forecast verification framework for administrative purposes. Monthly Weather Review, 137, 331-349. DOI: 10.1175/2008MWR2553.1

Muchuru, S., Landman, W.A. and DeWitt, D., 2016. Prediction of inflows into Lake Kariba using a combination of physical and empirical models. International Journal of Climatology, 36, 2570–2581, DOI: 10.1002/joc.4513.

Murphy, A.H., 1998. The early history of probability forecasts: Some extensions and clarification. Weather and Forecasting, 13, 5–15.

Ngarakana-Gwasira, E.T., Bhunu, C.P., Masocha, M. and Mashonjowa, E. 2016. Assessing the role of climate change in malaria transmission in Africa. Malaria Research and Treatment, 2016. Article ID 7104291 http://dx.doi.org/10.1155/2016/7104291

Pomposi, C., Funk, C., Shukla, S., Harrison, L. and Magadzire, T., 2018. Distinguishing southern Africa precipitation response by strength of El Niño events and implications for decision-making. Environmental Research Letters. 13(7), 074015, doi: 10.1088/1748-9326/aacc4c.

Reason, C. J. C., 1999. Interannual warm and cool events in the subtropical/mid-latitude south Indian Ocean. Geophysical Research Letters, 26, 215–218.

Scaife, A.A., Camp, J., Comer, R., Davis, P., Dunstone, N., Gordon, M., MacLachlan, C., Martin, N., Nie, Y., Ren, H.-L., Roberts, M., Robinson, W., Smith, D. and Vidale, P.L., 2019. Does increased atmospheric resolution improve seasonal climate predictions? Atmospheric Science Letters. DOI: 10.1002/asl.922.

Thomson, M.C., Doblas-Reyes, F.J., Mason, S.J., Hagedorn, R., Connor, S.J, Phindela, T., Morse, A.P. Palmer, T.N., 2006. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature. 439, 576–579.

Thomson, M.C., Ukawuba, I., Hershey, C.L., Bennett, A., Ceccato, P., Lyon, B., Dinku, T., 2017. Using rainfall and temperature data in the evaluation of national malaria control programs in Africa. The American Journal of Tropical Medicine and Hygiene, 97(Suppl 3), 32–45, doi:10.4269/ajtmh.16-0696

Troccoli, A., Harrison, M., Anderson, D.L.T., Mason, S.J., 2008: Seasonal Climate: Forecasting and Managing Risk. NATO Science Series on Earth and Environmental Sciences, Vol. 82, Springer, 467 pp.

Van den Dool, H. M., Toth, Z., 1991. Why do forecasts for ''near normal'' often fail? Weather and Forecasting, 6, 76–85.

Washington, R., Preston, A., 2006. Extreme wet years over southern Africa: Role of the Indian Ocean sea surface temperatures. Journal of Geophysical Research. 111, D15104. doi:10.1029/2005JD006724.

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd edition. Academic Press: Amsterdam, pp. 676.