Advances in
Science & Research
Open Access Proceedings

The verification of seasonal precipitation forecasts for early warning in Zambia and Malawi

O. Hyvärinen[1], L. Mtilatila[2], K. Pilli-Sihvola[1], A. Venäläinen[1], and H. Gregow[1]

[1]Finnish Meteorological Institute, Helsinki, Finland
[2]Department of Climate Change and Meteorological Services, Lilongwe, Malawi

Correspondence to: O. Hyvärinen (otto.hyvarinen@fmi.fi)

Abstract. We assess the probabilistic seasonal precipitation forecasts issued by Regional Climate Outlook Forum (RCOF) for the area of two southern African countries, Malawi and Zambia from 2002 to 2013. The forecasts, issued in August, are of rainy season rainfall accumulations in three categories (above normal, normal, and below normal), for early season (October–December) and late season (January–March). As observations we used in-situ observations and interpolated precipitation products from Global Precipitation Climatology Project (GPCP), Global Precipitation Climatology Centre (GPCC), and Climate Prediction Centre (CPC) Merged Analysis of Precipitation (CMAP). Differences between results from different data products are smaller than confidence intervals calculated by bootstrap.

We focus on below normal forecasts as they were deemed to be the most important for society. The well-known decomposition of Brier score into three terms (Reliability, Resolution, and Uncertainty) shows that the forecasts are rather reliable or well-calibrated, but have a very low resolution; that is, they are not able to discriminate different events. The forecasts also lack sharpness as forecasts for one category are rarely higher than 40 % or less than 25 %. However, these results might be unnecessarily pessimistic, because seasonal forecasts have gone through much development during the period when the forecasts verified in this paper were issued, and forecasts using current methodology might have performed better.

1   Introduction

Probabilistic seasonal precipitation forecasts in several parts of the world are issued by the Regional Climate Outlook Forums (RCOF, Ogallo et al., 2008). In these forums, national, regional and international climate experts meet to produce real-time regional climate outlooks based on input from National Meteorological Services, regional institutions, Regional Climate Centres, and global producers of climate predictions. The outlooks are consensus-based, implying that the forecast is made for the whole region and downscaling to national level is done afterwards. For the Southern African region, the seasonal outlooks are produced by the Southern Africa Regional Climate Outlook Forum (SARCOF, http://www.sadc.int/news-events/newsletters/climate-outlook/) that had the first meeting in 1997.

Malawi and Zambia are two southern African countries which face multiple challenges related to weather and climate, mainly due to their exposure and vulnerability to weather and climate shocks, particularly prolonged dry spells and floods. For instance, high reliance on rain-fed agriculture, poor disaster preparedness levels, and general lack of capacity in the communities expose the countries and people to a pertinent threat of food insecurity, malnutrition, and loss of lives. This hampers the general development efforts. The seasonal outlook issued by SARCOF and downscaled to Malawi and Zambia is disseminated widely to the disaster preparedness and response authorities, both on the government and major United Nations (UN) organisations, and it is used as one component behind the national contingency plans. To communities and farmers the forecast is issued by some non-governmental organisations (NGO) who work on

the community level; the potential of this, however, has not been fully harnessed.

The value of meteorological information to the end user is tightly connected to the accuracy and skill of the information. As seasonal forecasting is a relatively new endeavor in the field of meteorological forecasting, we aim to assess the performance of the seasonal forecasts issued in Malawi and Zambia. The verification results lay a basis for the discussion on the usability and value of seasonal forecasts for the early warning process in the region, as the forecast is widely used in the two countries and the potential benefits of the forecast are high. This study focuses on forecasts of "below normal" precipitation, as they are the most important forecasts for drought and if they prove to be skilful, they could be beneficial for farming practices and as an early warning sign.

The study forms a part of a two-year (2013–2014) research project "Study on risk management of extreme weather related disasters and climate change adaptation in Malawi and Zambia (SAFE-MET)", funded by the Academy of Finland and the Ministry for Foreign Affairs of Finland as a part of the Finnish Research Programme on Climate Change (FICCA). The goals of SAFE-MET were to examine, propose, and test ways to strengthen societies' resilience to climate and weather related hazards and to enhance multidisciplinary climate change research in Zambia and Malawi.

2   Data

Data used consisted of RCOF forecasts and gridded and in-situ observations. The use of more than one observational data set can help in determining the uncertainty of the results.

2.1   Forecasts

The SARCOF forecasts, issued in August, are of summer rainfall accumulations in three categories (above normal, normal, and below normal), for early season (October–December, OND) and late season (January–March, JFM). The forecasts from 2002 to 2013 were available for verification. The forecasts are disseminated only as pictures with forecast probabilities shown as filled contour lines. Therefore, these pictures had to be reverse-engineered into data before they can be compared with the gridded observations.

We divided an area from -8.5 to $-18°$ latitude and from 20 to 37° longitude into $1.0° \times 1.0°$ and $2.5° \times 2.5°$ grids of forecasts (Fig. 1). These grids correspond to the two grids of gridded observations below. The grids were then filled with values from the pictures. In most cases, it was unambiguous what value should be used but when the contour line between two forecast categories crossed a grid square in the picture, the value of forecast category covering the largest area of the square was used. This subjective part of digitization might add some noise to the results. However, the number of ambiguous grid values was small, they amounted to less than
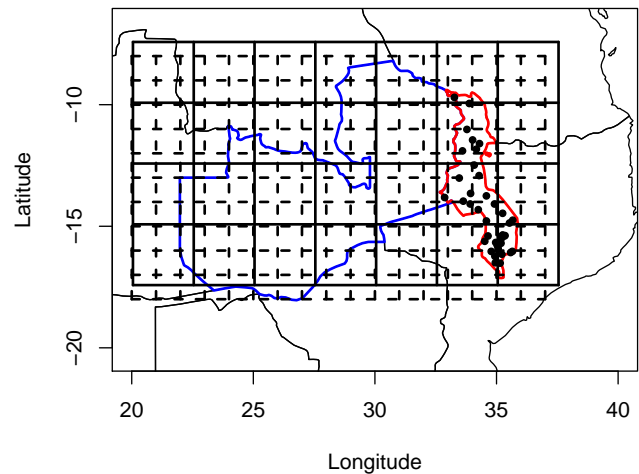

Figure 1. The map of the study area, where Malawi (red lines) and Zambia (blue lines) are emphasized. The grids used for subjective digitalization of forecasts are also shown, the $1.0° \times 1.0°$ grid in dashed line and the $2.5° \times 2.5°$ grid in solid line. Black dots show the stations with in-situ observations in Malawi.

2 % of grid points of the $1.0° \times 1.0°$ grid and 3 % of grid points of the $2.5° \times 2.5°$ grid.

The distribution of forecasts (Fig. 2) shows that forecasts lack sharpness in a verification sense (e.g., Wilks, 2011) as forecasts for one category do not differ much from the climatological values (33 %). Probabilities of forecasts for above normal are usually somewhat smaller than the climatological values, and probabilities of forecasts for normal are usually somewhat greater than the climatological values. Moreover, distributions of forecasts of both grids are very similar, even though grids have different resolutions and grid points are in somewhat different locations. This gives us confidence in our digitalization of forecasts, and noise added by our subjective digitization is probably not significant.

2.2   Gridded observations

As gridded observations, we used interpolated precipitation products from Global Precipitation Climatology Project (GPCP) (Adler et al., 2003), Global Precipitation Climatology Centre (GPCC) (Schneider et al., 2013), and Climate Prediction Centre (CPC) Merged Analysis of Precipitation (CMAP) (Xie and Arkin, 1997). All interpolated products used measurements from rain gauge stations and information from different satellite instruments. For GPCP and CMAP, the resolution was $2.5° \times 2.5°$ and for GPCC, the resolution was $1.0° \times 1.0°$. Years from 1980 to 2001 were used for climatology, so quantiles of 33 and 66% were calculated using those years for each grid point and grid points for years with forecasts (2002–2013) were then classified using those quantiles. Different data sets can have differences in precipitation levels but for our purposes, absolute values need not be exact, if the relative values are consistent.
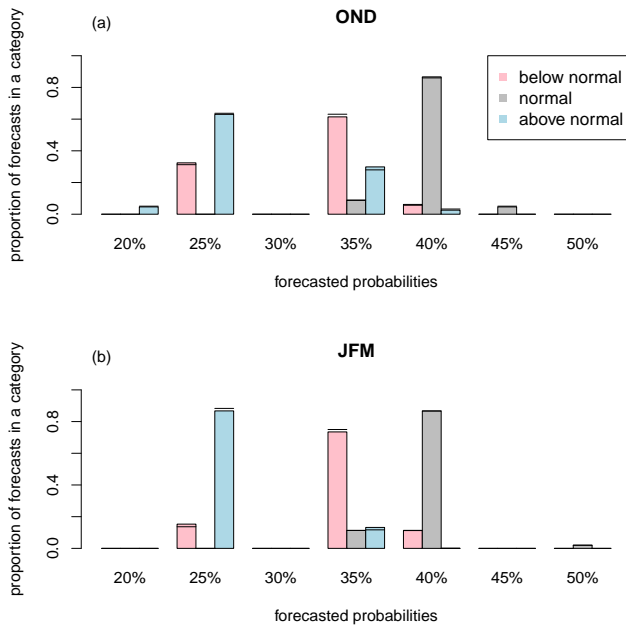
Figure 2. The distribution of forecasted probabilities of early (OND) and late (JFM) season forecasts. Proportions of different forecast probabilities for each category are shown, so bars of the same color add to one. Bars are calculated using the $1.0° \times 1.0°$ grid, results using the $2.5° \times 2.5°$ grid are shown as vertical lines over the bars.

In addition to the interpolated precipitation products, precipitation from the ERA-Interim reanalysis was tested. Unfortunately, there was a clear trend in the precipitation in the area under study, not present in the other products, and the results from ERA-Interim were not used. This non-optimal quality of African precipitation in ERA-Interim is a known problem (Dee et al., 2011).

2.3   In-situ observations

In Malawi, 39 stations were used for the period from 2007 to 2012. Stations covered the whole Malawi for both early season and late season forecasts. The stations used in this analysis are those that are also used to generate the operational forecasts. In Malawi during the study period, below-normal forecasts of 25 and 35 % were issued for OND, while only 25 % forecasts were issued for JFM.

3   Methods

As we focused on below normal forecasts, the methodology for binary forecasts can be used (e.g., Wilks, 2011). Results were assessed graphically using the attributes diagram (which is a refinement of the reliability diagram) and the Receiver Operating Characteristic (ROC) curve, and more quantitatively using the Brier Skill Score (BSS). BSS can be

written as

$$BSS = (\text{Resolution} - \text{Reliability})/\text{Uncertainty}, \qquad (1)$$

where the three terms come from the decomposition of Brier Score. Skill score values larger than zero indicate skillful forecasts compared with climatology.

For gridded data sets, one verification measure was calculated from the whole data set, as point-wise calculations would be based on only 11 data points. The climatology of BSS is also for the whole data set. The block-bootstrapping, as in Hamill (1999), was used to calculate confidence intervals (CIs) in order to take into account the high spatial correlation of grid points. In the standard bootstrap, all data points are sampled with replacement and the correlation of data points is therefore lost, resulting in too narrow CIs. Here we sampled the whole grids (that is, years) so the spatial correlation of the data is not lost, resulting in more realistic CIs. The number of bootstrap samples was 15 000.

For in-situ measurements, block-bootstrapping was not pursued due to limited resources, and only standard bootstrap ($n = 1000$) was used. Because of the proximity of the stations, there is strong autocorrelation, and therefore the standard bootstrap intervals are probably too narrow.

4   Results

4.1   Gridded observations

In Figs. 3a, c, e and 4a, c, e, points of the attribute diagrams should be as near the diagonal as possible, but, in reality, they are rather near the horizontal "No resolution" line, while some data points are in the grey area of skillful forecasts. Similarly, in ROC curves (Figs. 3b, d, f and 4b, d, f), points should be as far as possible from the diagonal line towards the top left corner, but they are very near the diagonal and even below it. Taken at face value, results suggest not very skillful forecasts. However, CIs are very large and cover both skillful and not skillful areas of diagrams and curves. Also, values of BSS (Table 1) indicate that forecasts do not have much skill, as all values are around zero, and no CIs cover only positive values. From BSS decomposition terms, we can conclude that forecasts are rather reliable or well-calibrated (Reliability terms are rather small), but have very low resolution (Resolution terms are almost zero); that is, they are not able to discriminate different events. All in all, differences between data products are smaller than confidence intervals calculated by block-bootstrap. Furthermore, our subjective digitization of forecasts from pictures might add some noise to the results, but small sensitivity tests, by moving the position of forecasts slightly, did not change the results substantially.

From the attributes diagrams (Figs. 3a, c, e and 4a, c, e) the base rate of below normal in different data sets can be estimated, which shows how consistent the different products
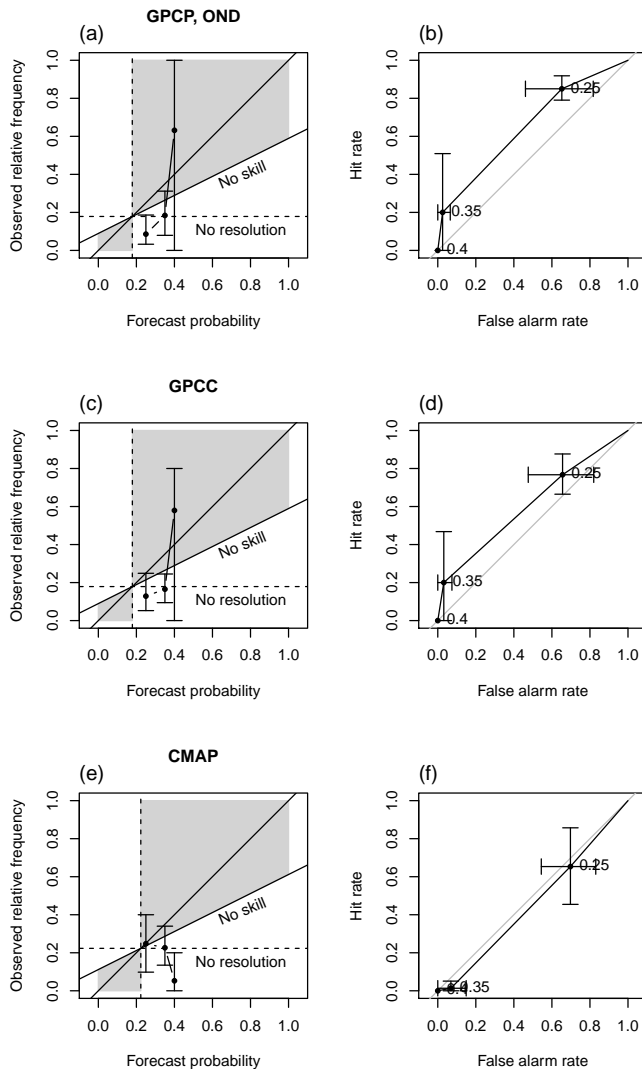
Figure 3. The attributes diagrams and the ROC curves for early season forecasts, based on different data sets for verification. The confidence intervals are calculated using the block bootstrap ($n = 15\,000$).

Figure 4. The attributes diagrams and the ROC curves for late season forecasts, based on different data sets for verification. The confidence intervals are calculated using the block bootstrap ($n = 15\,000$).

are with each other. For early season, the below-normal conditions occur about 20 % in all three products, showing good consistency. In late season, GPCC has a somewhat smaller base rate than others, which then produces a slightly worse BSS, as observations diverge more from forecasts.

4.2 In-situ observations

For in-situ observations, the results are very similar to the gridded results. In the attribute diagrams, all CIs cover the vertical line of "no skill", while points in the ROC curve are very near the diagonal line (Fig. 5). For more quantitative results, BSS is, for practical purposes, zero (Table 2) for both early and late season, and therefore Resolution terms, as well as Reliability terms, are also almost zero. The Resolution of
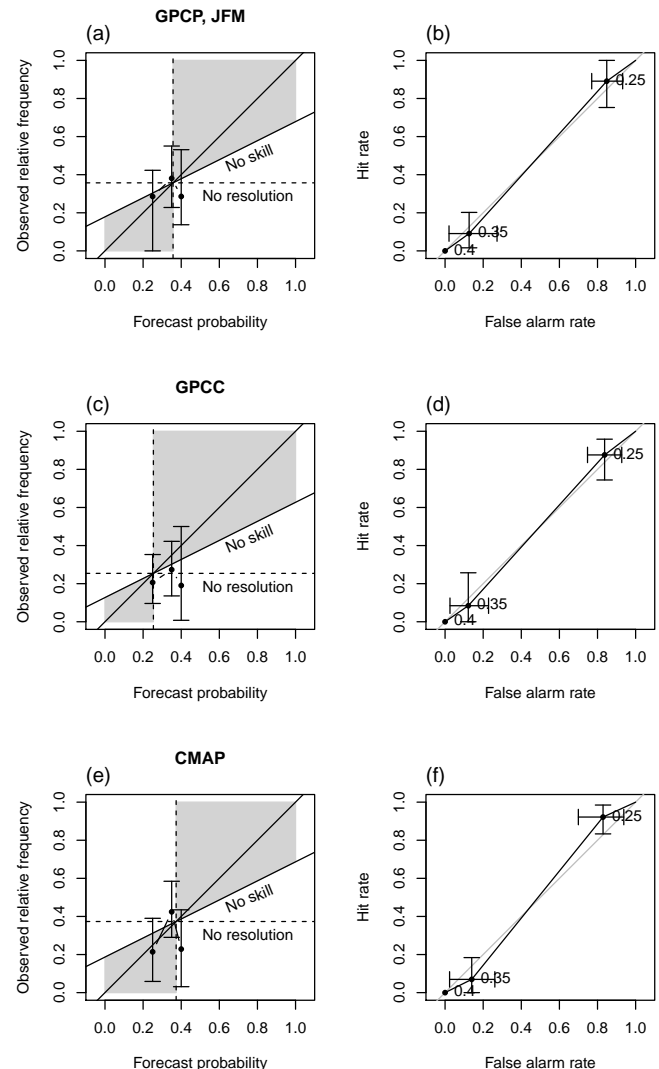
later season is exactly zero, as only forecasts of 25 % were issued for Malawi. So, the forecast skill seems to be very limited.

5 Discussion

Different data sources give slightly different results, but all results indicate that the forecasts have limited skill in the selected area. This is consistent with the results of Mason and Chidzambwa (2008) where they remark that predictability in the vicinity of Malawi is weak because of a transition between distinct ENSO teleconnection signals. It might be possible to increase the skill by recalibrating the forecasts, but that is beyond the scope of this paper.

Table 1. For the early and late season, BSS, its decomposition terms, and its CI for different data sets. The CIs are calculated using the block bootstrap ($n = 15\,000$).

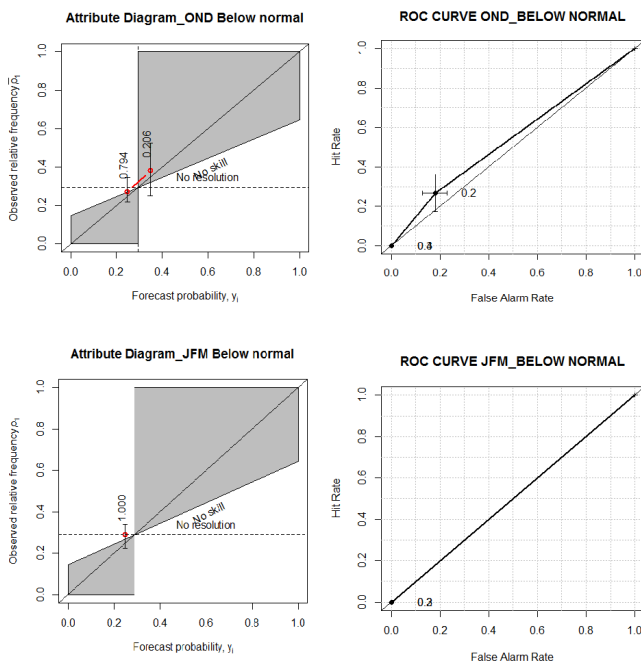| Data name	$\dfrac{\text{Resolution}}{\text{Uncertainty}}$	$\dfrac{\text{Reliability}}{\text{Uncertainty}}$	BSS	BSS CI$_{0.025}$	BSS CI$_{0.975}$
OND					
GPCP	0.098	0.192	−0.094	−0.625	0.054
GPCC	0.074	0.184	−0.110	−0.448	0.026
CMAP	0.011	0.092	−0.082	−0.350	−0.014
JFM					
GPCP	0.007	0.011	−0.003	−0.117	0.012
GPCC	0.006	0.048	−0.042	−0.331	0.006
CMAP	0.033	0.033	−0.000	−0.106	0.021

Figure 5. The attributes diagrams and the ROC curves for early and late season forecasts, based on in-situ observations for verification. The confidence intervals are calculated using the standard bootstrap ($n = 1000$).

Table 2. For the early and late season, BSS, and its decomposition terms for in-situ observations in Malawi.

	$\dfrac{\text{Resolution}}{\text{Uncertainty}}$	$\dfrac{\text{Reliability}}{\text{Uncertainty}}$	BSS
OND	0.009	0.003	0.006
JFM	0.000	0.008	−0.008

However, the question which arises is if verification should focus on validating the past performance or be more future-oriented and estimate how forecasts will perform in the future. Our verification is based on historical data (what was really forecasted), not on reprocessed data (what would have been forecasted, if the present system had been available then). Moreover, because of the unreproducible way RCOF forecasts are made, getting large enough data sets will take years, and even then they will not be of constant quality as the underlying forecast systems continue to evolve. Therefore the results can be too pessimistic, and for a more optimistic outlook, we should compare our results with results using hindcasts. This is a natural way of presenting verification results by the model developers (e.g., Landman, 2014). Some data is available also for the larger verification community from, for example, the Climate-system Historical Forecast Project of the World Climate Research Program (Kirtman and Pirani, 2009), but then the temporal range of data might not be as comprehensive.

Of course, seasonal forecasts can be very beneficial for agriculture. For example, a review (Hansen et al., 2011) shows how farmers can use and benefit from the information, given enabling environment and improved communication of the information and its uncertainties, and bundled with historic observations. Furthermore, as long as food insecurity remains a threat in Malawi and Zambia, the government sector, supported by major UN organisations and NGOs, benefits from the information, for instance, in vulnerability assessments and contingency plans (personal communication with the end users). However, the uncertainty of the information needs to be clearly communicated and understood by the end users to provide benefits and avoid losses from unskilful forecasts compared to efficient, science-based use of climatological information. Furthermore, the forecast, currently issued once during the rainy season, could be updated in the light of new information closer to the rainy season. And finally, the seasonal information must be complemented with accurate and efficiently communicated short-range weather

forecasts to reap the full benefits of weather and climatological information in southern Africa.

6   Conclusions

Based on our dataset, the SARCOF forecasts seem to have limited skill, which is partly explained by the climatological conditions in the area under study. However, it might not be prudent to make drastic conclusions about the usability of the current SARCOF forecasts based on this dataset alone, because ten years of seasonal forecasts is a rather small data set, as shown by large confidence intervals for all measures. This small number of forecasts, that are also constantly evolving, presents a challenge for verification in an operational setting. Especially, due to the partly subjective nature of RCOF, getting ten data points takes ten years of time. Therefore, it can take decades before a reasonable number of forecasts is available. For more automatic forecasts, hindcasts offer more forecasts in a shorter time frame, and new forecast systems can be assessed as new versions becomes available.

Edited by: P. Nurmi
Reviewed by: two anonymous referees

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), J. Hydrometeorol., 4, 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2, 2003.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Hamill, T.: Hypothesis Tests for Evaluating Numerical Precipitation Forecasts, Weather Forecast., 14, 155–167, 1999.

Hansen, J. W., Mason, S. J., Sun, L., and Tall, A.: Review of Seasonal Climate Forecasting for Agriculture in sub-Saharan Africa, Exp. Agr., 47, 205–240, doi:10.1017/S0014479710000876, 2011.

Kirtman, B. and Pirani, A.: The State of the Art of Seasonal Prediction: Outcomes and Recommendations from the First World Climate Research Program Workshop on Seasonal Prediction, B. Am. Meteorol. Soc., 90, 455–458, doi:10.1175/2008BAMS2707.1, 2009.

Landman, W. A.: How the International Research Institute for Climate and Society has contributed towards seasonal climate forecast modelling and operations in South Africa, Earth Perspectives, 1, 22, doi:10.1186/2194-6434-1-22, 2014.

Mason, S. and Chidzambwa, S.: Verification of African RCOF Forecasts, in: World Meteorological Organization RCOF Review 2008, IRI Technical Report 09-02, 2008.

Ogallo, L., Bessemoulin, P., Ceron, J.-P., Mason, S., and Connor, S. J.: Adapting to climate variability and change: the Climate Outlook Forum process, WMO Bulletin, 57, 93–101, 2008.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theor. Appl. Climatol., 115, 15–40, doi:10.1007/s00704-013-0860-x, 2013.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, 3 edn., 2011.

Xie, P. and Arkin, P. A.: Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs, B. Am. Meteorol. Soc., 78, 2539–2558, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2, 1997.