Generated using version 3.0 of the official AMS LaTeX template

A weather risk attribution forecast system for Africa and the world

Dáithí A. Stone * and Chris Lennard and Mark Tadross

CSAG, Environmental and Geographical Science, University of Cape Town, South Africa

Myles R. Allen

Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom

Peter A. Stott

Met Office Hadley Centre, Exeter, United Kingdom

Pardeep Pall

Institute for Atmospheric and Climate Science, ETH Zürich, Zurich, Switzerland

* *Corresponding author address:* Dáithí Stone, CSAG, Shell Environmental and Geographical Science Building, South Lane, Upper Campus, University of Cape Town, Private Bag X3, Rondebosch, Western Cape, 7701, South Africa.

E-mail: stoned@csag.uct.ac.za

ABSTRACT

As Russia baked in a record heatwave and floods covered much of Pakistan, or more recently as severe snowstorms disrupted western Europe and the eastern United States, the attribution question was prominent in public discourse: "Are our emissions to blame for this event?" Unfortunately, no objective system exists at the moment for dealing with this question, so responses amounted to contradictory samples of expert opinion which placed different weight on various indirect sources of evidence and different interpretations of the question. The most severe impacts of climate change over the next few decades are likely to be associated with changing risks of certain forms of extreme weather, even while other extreme weather events unrelated to climate change will continue to occur. An objective method of distinguishing between these will be essential for an evidence-based approach to the imminent distribution of climate change adaptation assistance.

Here we describe the world's first proactive system for examining how anthropogenic emissions have contributed to weather risk in our current climate. By comparing real seasonal forecasts against parallel counterfactual seasonal forecasts of "the climate that might have been" had human activities never emitted greenhouse gases, this "attribution forecast" provides real-time responses to the question: "Has this event been made more frequent by our emissions?" Output from two and a quarter years of effective operation using two separate model versions suggest that the most sensitive aspect is in the estimation of the ocean warming attributable to greenhouse gas emissions.

During development and operation of this system it has become evident that while this attribution service may be produced in a seasonal forecasting framework, it differs in fundamental ways from a seasonal forecast product and bears a much closer relation to a climate

change product. Two implementations of one method for a proactive attribution service are employed here, but other implementations and methods are possible. Given the current prominence of the attribution question, this should be a priority field for both climate research and climate service programmes. In the meantime, it is hoped that this attribution forecast service will provide valuable insight into the robustness of the attribution product, the specifics of the information required, how that information is interpreted, and how that interpretation is applied.

1. Rationale

Climate information provided to inform decisions for adaptation to climate change usually concerns the future, either the reasonably distant future a century from now or the nearer future a few decades hence. In Africa, however, such information is often of little use for a variety of reasons. Foremost amongst these is the relatively short planning horizon for most African institutions: financial constraints and historical instability on the continent both discourage long term planning. Beyond that, the climate issue in many parts of the continent is at least as much about adapting to current climate by reducing vulnerability and increasing resilience as it is about adapting to future climate change (Boko et al. 2007). Thus, there are two main questions to be targeted by producers of climate information: what is the current climate, with a particular focus on aspects that affect society and biological systems; and who will pay for reducing vulnerability to climate and climate change? The first question cannot necessarily be addressed using historical measurements, both because such measurements are often sparse over Africa and because they provide biased estimates under a changing climate (Räisänen and Ruokolainen 2008; Laepple et al. 2008), changing demography, and changing ecological landscape. As for the last question, African nations have insisted that financial transfers to fund adaptation to climate change should be additional to traditional aid which is often used to respond to weather-related disasters or reduce vulnerability to extreme weather. Meeting this additionality criterion requires an evidence-based method of distinguishing entirely natural weather events and events that have been affected by local anthropogenic activities, such as land use change, from events that have been affected by global anthropogenic emissions.

A fundamental pre-requisite for addressing both questions is a clear definition of climate. The World Meteorological Organization has traditionally used an interpretation of climate based on an observational paradigm, thus defining climate in terms of means and variances over a specific period of time, usually 30 years. This is clearly inadequate at a time when climate may be changing significantly over 30 years or less (Räisänen and Ruokolainen 2008; Laepple et al. 2008). One could adopt a more rigorous definition based on the population statistics of all possible states of the climate system given conditions external to the climate system (such as anthropogenic greenhouse gas emissions and solar luminosity). In the context of anthropogenic climate change, however, it seems ridiculous to be concerned about including the possibility of a glacial maximum, say, given that such a current cryospheric state was almost certainly not accessible to the climate system given its state at the beginning of anthropogenic interference a century or so ago. Thus a more practical definition could be along the lines of the distribution of weather states that would be obtained by a hypothetical large ensemble of realisations of the Earth System initialised with small perturbations to the initial weather state a century or so ago, but subsequently evolving with boundary conditions identical to those observed over the intervening years. This would consider millennial time scale fluctuations in the cryosphere as climate, but it would categorise both synoptic time scale fluctuations in precipitation and interannual coupled atmosphere-ocean variability associated with El Niño as weather, not climate, variations, for example.

Here we present a new type of climate product designed to supplement existing resources for addressing the first question and specifically to help answer the second question, providing evidence for or against altered weather risks and thus encouraging appropriate activities to reduce current vulnerabilities. It is the world's first system to examine in real time how anthropogenic emissions have contributed to our weather. As such, it is best considered a demonstration product highlighting both cases and issues for more in-depth investigation, rather than a definitive analysis. While currently focused on Africa, analysis is in fact also performed for other areas around the planet. The results from these areas around the globe, rather than specifically from Africa, are highlighted in this paper in order to address the broader interest of most readers.

This weather risk "attribution forecast" addresses the question of how the weather might have differed from what is currently being experienced had we never emitted greenhouse gases. The approach generally follows the method laid out in Pall et al. (2011) and also shares some aspects with the method used by Perlwitz et al. (2009), in this case comparing real seasonal forecasts against what the seasonal forecasts might have been had we never emitted greenhouse gases, as in Fig. 1. The key difference is that it is a proactive analysis, producing analyses before (and regardless of whether) an event occurs. It is therefore both timely and more aggressive in dealing with selection bias. Results are issued monthly at http://www.csag.uct.ac.za/∼daithi/forecast.

In using a seasonal forecasting framework we clearly run into trouble with our definition of climate above. As will be discussed below, these forecasts are generated with imposed sea surface states, thus precluding any century-long “spin-up” of the ocean state. This is an trade-off in experimental design between applicability of the experimental setup and applicability of the climate model. Performing the simulations with century-long simulations of a coupled ocean-atmosphere model would force the use of a low resolution model with questionable capability of simulating the synoptic type of events in which we are interested. We believe it is more desireable to do as good a job as we can of simulating atmospheric weather given surface boundary conditions than to worry about the origin of those particular surface boundary conditions. For many if not most events of interest such a short spin-up may be sufficient to explore the bulk of the weather attractor. As coupled atmosphere-ocean models improve and computing resources increase, however, more strict experimental setups will become plausible, but in the meantime results from the atmospheric seasonal forecasting setup used here must be interpreted with this qualification in mind.

The attribution forecast project presented here has three primary goals relating to the nature of the product:

i. to develop a framework for producing information about how anthropogenic emissions are affecting our current weather;

ii. to elucidate how this sort of product is interpreted by potential users;

iii. to interact with users of the product and further develop its application.

It also has two secondary goals, which relate directly to the information produced:

iv. to determine the sensitivity of estimates to different forecast models and forecast configurations;

v. to provide indicative information required for adaptation, highlighting cases for more detailed study.

This paper addresses Goal 1 by describing the setup of the system. It also presents some first observations regarding Goals 4 and 3 from the construction and first year and a half of operation of the system; more in-depth studies addressing these topics and Goal 2 will be reported later once the product has been in use for some time. Goal 5 is an operational output.

2. Method

a. The seasonal forecasting system

The Climate Systems Analysis Group (CSAG) at the University of Cape Town has been issuing seasonal weather forecasts since 2002 (http://www.gfcsa.net/csag.html). Until June 2010, these forecasts were generated with a version of the U.K. Met Office's HadAM3 model (Pope et al. 2000). This version of HadAM3, run at N48 resolution (3.75° longitude by 2.5°) with 19 vertical levels, includes the mixed phase precipitation scheme as this has been found to improve the model precipitation over southern Africa (Wilson 2000). Starting in June 2010, the publicly posted forecasts have been generated using the HadAM3P version of the U.K. Met Office model which is used as the driving global model in the PRECIS regional modelling project (Jones et al. 2004). HadAM3P shares the same underlying dynamical core as HadAM3, but uses more recently developed physical parametrisation schemes and runs at N96 resolution (1.875° longitude by 1.25°) with 19 vertical levels.

The generation of the forecasts discussed in this paper is outlined in Fig. 2. For a forecast generated and issued in March, a 10-member initial condition ensemble of simulations is started from the beginning of January and run through the end of June, providing a three-month forecast coverage (i.e. April-June in this example). Monthly mean sea surface temperatures (SSTs) for mid-January and mid-February are taken from the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation version 2 (OI.v2) observational product (Reynolds et al. 2002). SSTs for later months are estimated by adding the observed February SST anomaly from the 1960–2000 climatological mean to the climatological means of these later months. Daily SSTs are estimated through linear interpolation between the middles of the months; thus, in the March forecast January is the only month to have entirely observationally-based SSTs. A seasonally varying climatology of sea ice coverage is used, with obvious issues for forecasts in high latitudes. The initial conditions are taken from the end of the first month in the previous month's forecast simulation (e.g. for the March forecast they come from the end of December in the February forecast); thus concatenation of the first months of subsequent forecast simulations results in continuous hindcast series, which are also contiguous with the ten climatological hindcast simulations started in 1960. Strictly speaking, these hindcasts are "AMIP-style" simulations because the initial conditions are not derived from actual observed states, but the distinction disappears in the structural framework of this forecasting system.

In this paper, we discuss forecasts generated from December 2008 through to November 2010 using both HadAM3 and HadAM3P. Most of these forecasts were generated after the fact rather than in real-time. The HadAM3 forecasts described here are not the same as those issued publicly prior to June 2010, and differ in some specifics and in use of computing platform from those earlier publicly issued forecasts. These new HadAM3 forecasts continue to be generated for research purposes and for the attribution forecast product described in this paper but are not currently included as part of the publicly issued seasonal forecast.

b. Generation of counterfactual forecasts

For each of these two forecast models, a second 10-member ensemble of forecasts is generated to estimate what the forecast might have been had human activities never emitted greenhouse gases. For these "non-GHG" forecasts the prescribed greenhouse gas concentrations in the model are reduced to pre-industrial values and the prescribed SSTs are altered according to an estimate of the warming attributable to our historical greenhouse gas emissions described below (Fig. 2). These non-GHG forecasts are identical to the real forecasts in all other respects, including in sea ice coverage. The non-GHG simulations were initially started from the October 2008 initial conditions of the real forecasts' hindcasts (i.e. generated in December 2008), and have been run separately from the real forecasts since then.

The warming attributable to greenhouse gas emissions is estimated using an optimal total least squares regression analysis (Allen and Stott 2003; Stott et al. 2003) on data from the HadSST2 dataset of gridded observational measurements (Rayner et al. 2006) and output from simulations of the HadCM3 coupled ocean-atmosphere climate model (Stott et al. 2006). This analysis estimates objective scaling factors by which the amplitudes of the modelled response patterns to various external forcings should be adjusted to match the observations. Details are given in the Appendix. The analysis here indicates that the response of the HadCM3 coupled ocean-atmosphere model to anthropogenic greenhouse gas forcing should be adjusted by a factor of 1.3. In order for these forecasts to be generated in real-time, we are unable to follow Pall et al. (2011) in generating large ensembles sampling the uncertainty in the attributable pattern and scaling factor; thus we use only this single pattern-scaling estimate.

A further scaling factor is estimated to account for the fact that we are outside the temporal range of the HadCM3 simulations. This factor is estimated through a linear fit to the global mean temperature change over the 1949-1998 period in the HadCM3 greenhouse gas simulations, extended out to the forecast period. For the March 2009 forecast, this factor has a value of 1.2 over the final decade in the HadCM3 simulations. The time-mean difference between that final 1989-1998 period in the HadCM3 greenhouse gas simulations and the respective periods in the related unforced simulations (which share the same initial conditions) are then multiplied by these two factors to give the attributable SST warming estimate. This attributable SST warming estimate is subtracted from the SSTs used in the seasonal forecast as indicated schematically in Fig. 2.

c. Output of the forecasts

Currently, we only publicly release the one-month forecasts with lead one month, e.g. April and April-June for the forecast issued in March. Once they have been simulated, hindcasts replace the one-month lead forecasts in the publicly issued product (i.e. in June for the April hindcast and in August for the April-June hindcast). Only temperature and precipitation are considered at the moment.

The primary outputs are regional averages calculated for the 17 regions shown in Fig. 3a. These are all existing political and/or economic regions at least 4 million km^2 in extent (except for the G3, from which Venezuela withdrew in 2006). HadAM3P output is also calculated for 10 groupings of nations of the Southern African Development Community (SADC) which are larger than 390 000 km^2, but we will focus on the global regions in this paper. Regions dominated by small islands, most notably the Pacific and Caribbean areas, are not included because temperatures (and to some degree precipitation) over these regions would be largely constrained by the imposed SSTs rather than dynamically generated by the atmospheric model.

d. Estimation of forecast quantities

For each real and non-GHG forecast pair we calculate:

i. the means of both the real and non-GHG forecasts;

ii. the probability of an unusually high (hot or wet) event in both the real and non-GHG forecasts;

iii. the probability of an unusually low (cold or dry) event in both the real non-GHG forecasts;

iv. the fraction of the risk of an unusually high event attributable to the greenhouse gas emissions;

v. the fraction of the risk of an unusually low event attributable to the greenhouse gas emissions.

The approach to estimating these values is indicated schematically in Fig. 1. Calculations for the regions are performed after spatial averaging. All precipitation calculations are performed on the logarithm (after spatial averaging in the case of the regions); this facilitates statistical tests later, because the assumption of a t-distribution is then more justified.

The mean forecast is simply the average of the values in the ten simulations of the ensemble. The threshold for an unusually low/high event (month/season) is defined as the $10^{\text{th}}/90^{\text{th}}$ percentile over the reference period covering 1960 to the year before the forecast (i.e. 1960-2008 for months/seasons in 2009). There are ten hindcast simulations covering this period for each model, yielding samples of 490 years for months/seasons in 2009. Probabilities P_{real} and $P_{\text{non-GHG}}$ of exceedance of this threshold are estimated by adopting a t-distribution characterised by the mean and standard deviation of the ten forecast ensemble members. The fraction of attributable risk (FAR) is a measure of how much the likelihood of such an unusual event differs because of the historical greenhouse gas emissions (Allen 2003; Stone and Allen 2005) and is given by

$$\text{FAR} = 1 - \frac{P_{\text{non-GHG}}}{P_{\text{real}}}. \tag{1}$$

Positive values indicate that the emissions have increased the probability, while FAR $= 0$ implies that the probability is unaffected by the emissions, FAR $= 1$ implies that the event could only occur with the emissions, and FAR $= -\infty$ implies that the event could only occur without the emissions.

For the 17 standard regions, uncertainties in each of these quantities are estimated through a Monte Carlo procedure of sample size 100. Random ensembles of ten members are selected from the fitted t-distributions of the forecasts and the calculations repeated on these simulated samples. An example summary of results for the odds of an unusual November 2010 is shown in Fig. 3b-e. This summary displays what we can say with confidence at the two-sided 10% significance level based on these Monte Carlo analyses. It appears to be a robust feature for all months over the past two years that a larger number of regions are estimated to be experiencing a significantly lower chance of a cold month than the number for a hot month, as seen here for November 2010. It is also common in Northern Hemisphere winter months for no confident statement to be assigned to the northern extratropical regions; this appears to be due to the higher variability in these regions at this time of year increasing the width of the distribution of probabilities, because the most likely estimates are comparably high to other regions. The maps for the precipitation events are much more patchy, often with only around two regions per map being assigned confident statements as might be expected due to Type I error.

3. Sensitivity of output to implementation

Here we examine some of the output from the first two and a quarter years of effective operation of this forecast system. This provides us the opportunity to start comparing the impact of the choice of atmospheric model and the use of forecast versus hindcast output. First we will examine one region in more detail.

a. Regional focus: ECOWAS

Figs. 4 and 5 show the single-month attribution forecasts and hindcasts produced so far over the Economic Community of West African States (ECOWAS). The most obvious feature is a clear contribution from anthropogenic greenhouse gases to temperature but little if any contribution to precipitation, as might be expected based on climate change projections. The chance of a cold month is substantially reduced by the emissions, while the chance of a hot month is more modestly increased.

Visually, the choice of atmospheric forecast model appears to make little difference to the results. During the two and a quarter years plotted in Fig. 4, there has been a general warming trend followed by a recent cooling in both the HadAM3 and HadAM3P temperature estimates. Notably, variations across lead times for individual months are generally smaller than this trend. These features are also visible in the forecasts of event probabilities, with trends consistent across all three plots. The difference between the expected temperatures forecast under the real and non-GHG scenarios is fairly steady over this period, except for the recent cooling in the real scenario which is reflected only by a leveling in the non-GHG scenario. Variations in the spread of the forecasts mean that the forecast FARs do not necessarily vary as one might at first expect by looking at the most likely realisations. Precipitation totals show a more complicated variation on time scales of several months over the period, which are less clearly reflected in the forecasts of wet event probably than of dry event probability.

In these figures we can see some suggestion that the various forecast quantities might be robust against choice of forecast lead time, atmospheric model, and forecast setup. We will now examine how this robustness holds across multiple regions.

b. *Forecast lead time, including hindcast*

With both forecast and hindcast (i.e. extrapolated versus observed SST anomalies) data available, we have the opportunity to evaluate the sensitivity of results to SSTs. Because SSTs are imposed and non-interactive in this time-slice approach, it is unknown how much attribution results for the risk of a particular event are conditioned on the exact SST realisation that occurred. The hindcast-forecast difference should provide some indication of this sensitivity, although of course the importance of coupled atmosphere-ocean mixed layer interactions are not factored in this analysis.

Figs. 6 and 7 show how the various forecast quantities at different lead times correlate spatially with the final hindcasts over the 17 global regions in Fig. 3a. The correlation of these 17 regional values are calculated at each lead time against the ultimate hindcast values for each model and month. This is not a perfect measure for comparison, because it ignores variations in continental means and also the absolute amplitudes of the differences between the regions, but it serves as a simple first-order measure of the robustness of the forecasts as a function of lead time. A first observation is that there appears to be little systematic difference across real/non-GHG scenario or atmospheric model. The spatial patterns of all temperature and precipitation quantities are clearly controlled by the imposed SSTs, because the “forecasts” at lead time of -1 month, i.e. with a hybrid of observed and extrapolated SSTs, are almost perfectly correlated with the proper hindcast generated one month later. Reduced correlation at positive lead times reflects that being able to forecast SSTs is a useful component of any seasonal forecast system. These figures differ little if we restrict the analysis to the five African regions or to the ten nation groupings within SADC (not shown), although HadAM3 shows less consistency in forecasting the probability of wet events over Africa.

Interestingly, the temperature FAR patterns are more stable across lead time than the other temperature quantities, but the reverse is true for precipitation. The implication of this is that the FAR estimates for hot and cold events may not be strongly conditioned on the SST pattern that was realised in any given month, while the FAR estimates for wet and dry events may be strongly conditioned. Notably, though, there seems to be a lower limit on the correlation of the FAR patterns for both temperature and precipitation. Thus, it appears that there may be some minimum FAR signal that is not strongly conditioned on the realised SSTs but is related to the overall attributable warming pattern, and that this signal is higher for temperature than precipitation, as might be expected.

c. Choice of atmospheric model

With the use of two atmospheric forecast models, we can take a first look at the dependence of the output on the model selection. HadAM3 and HadAM3P are two versions of the same core model, but still differences and similarities should give us a first impression of how the attribution of the risk of weather events depends on model architecture (Jones et al. 2008).

We can calculate the spatial correlation of the HadAM3 forecasts with the HadAM3P forecasts, similarly to Figs. 6 and 7 (not shown). As with the comparison against the hindcast, the correlations across models are higher for the temperature FAR estimates than for the other temperature quantities, but lower for precipitation. Perhaps surprisingly, the models agree on the most likely and event probability patterns for precipitation at least as much as they do for the temperature patterns, but the FAR the agreement is considerably higher for temperature. Once again, the plots are highly similar if the analysis is restricted to the five African regions or to the ten nation groupings within SADC.

4. Constraints and concerns

Running a product operationally places strong constraints on how thoroughly all sources of uncertainty can be examined. The most obvious constraint is on the ensemble sizes, which do not allow the direct sampling of frequency possible in the targeted study of Pall et al. (2011). A larger but still modest sample size would provide enough data to use more plausible statistical estimates of the frequency than described here, in particular applying extreme value theory. Additionally, the operational product described here does not sample uncertainty in either the pattern or amplitude of the SST warming attributable to greenhouse gas emissions. Uncertainty in the atmospheric model is examined, but only in the limited case of using two versions of the same underlying model. During development of this system, some separate attribution forecasts were generated using the setup of HadAM3 used prior to June 2010 for the seasonal forecasts. This ran on a different computing platform, included a smaller number of simulations, used a different climatological reference period, split the non-GHG forecasts from the real forecasts at a different date, and used seasonally-varying scaling factors on the attributable SST warming estimates. There were no clear differences (not shown) with the HadAM3 and HadAM3P attribution forecasts described in this paper, but this informal comparison was a weak test of the influence of all of these factors. One point to consider is that we might expect the attributable risk to be seasonally varying. Due to biases in climatology and ability to simulate relevant seasonal teleconnections, the relative precision and accuracy of members of a group of climate attribution models may easily vary according to season as well as location and field, just like the case of weather forecast models and, it seems, climate change models. Model and methodology intercomparisons, as discussed and proposed at a recent workshop on the topic (Attribution Workshop, 17-18 August 2010, Broomfield, Colorado, U.S.A., http://www.esrl.noaa.gov/psd/csi/meetings/attrworkshop_2010/Attribution_Workshop_ExecSummary.pdf) should help to clarify the relevance of these concerns if they are conducted in parallel with existing seasonal forecast and/or climate change monitoring and attribution programmes.

The two big questions are how much accuracy is required of an operational attribution system and how this accuracy can be tested. Evaluation is clearly a major issue given the central role of a fundamentally unobservable system: the climate that would have been obtained had we not raised greenhouse gas concentrations. The fact that results for temperature appear to be dominated by the attributable change in large-scale sea-surface temperatures provides some reassurance here, because this aspect of anthropogenic climate change has been extensively studied and explored in the context of the traditional attribution question (Stott et al. 2010). The use of more coupled models to estimate patterns of attributable warming would clearly increase confidence in this aspect of the system. A more challenging problem is to assess whether the atmospheric model used is accurately simulating the statistics of weather extremes and their response to changing boundary conditions. It is important to emphasise that simple predictability is irrelevant here: there is nothing to prevent us from attributing changes in risk of weather events that are completely unpredictable in an initial-condition seasonal forecast. Seasonal forecast reliability statistics are more informative, because they provide an indication of whether the atmospheric model used is capable of simulating extreme weather with realistic frequency. In principle, reliability diagrams could be used to provide a post-hoc correction on model output statistics to compensate for biases in the atmospheric model, or, perhaps preferably, we could use them as a test of model reliability which would provide us with a means of identifying weather events for which our modelling system cannot yet provide an attribution assessment. This sort of analysis is a major endeavour over Africa because of the difficulties in interpreting available observational data and is well beyond the scope of this paper.

The level of accuracy required would seem to depend in large part on the purpose. If the attribution product is intended to directly inform legal cases or the international financing of major adaptation activities, then we would require a very rigorous system. Yet for such purposes, one would suppose that more targeted studies would be required anyway which focus on the event from a damage viewpoint rather than a climatological viewpoint (Pall et al. 2011). Thus, the most obvious use for an operational system would be in highlighting cases worthy of more rigorous targeted study. Another purpose would be for addressing questions from the public at large, usually via the media, on the cause of a particular event as the event is occurring. Tied into existing communication and education methods, this could prove a powerful tool to enhance understanding of climate change and climate research. For that, any resource must surely be better than the lack (beyond subjective expert judgement) that has existed until now.

Two major parallel challenges in the generation of this product, which operates every month covering the globe, has been to reduce the amount of output to a manageable size whilst still providing information in a readable format. While gridded maps would seem a friendly format for presentation, we have found them difficult for the presentation of uncertainty. In particular, gridded maps tend to be read at the spatial resolution shown, which is dangerous if it is the same as the model grid resolution. The restriction of output to regional averages has thus been deliberate. Further presentation measures include not publishing output for all forecast lead times or all possible season lengths. Furthermore, we have only focused on the contribution of anthropogenic greenhouse gas emissions, rather than of all anthropogenic forcing. This selection was made for consistency with international financial facilities currently being developed for funding adaptation activities (United Nations 2007), but for many purposes the total anthropogenic contribution may be more relevant. Should both the greenhouse gas and the total anthropogenic contributions both be presented in an operational product, and if so then how can we minimise the confusion that may result, especially when the contributions are opposite?

An overriding issue with any study on the attribution of weather risk is the selection effect (Chase et al. 2006), i.e. of focusing on events expected to have an increased probability attributable to historical emissions whilst ignoring the multitude of events with no such expected link or an expected decreased probability. The production of an operational system is a step toward reducing that risk, in that it produces information in a proactive and systematic way without tailoring definitions to fit a specific event that occurred. Nevertheless, with the 17 regions, 2 models, and 4 types of events considered here, one might expect 14 region-model-event combinations to have an apparently significantly increased probability at the 10% significance level purely by random chance for any given forecast. Some extra analysis needs to be applied in order to characterise whether a forecast seems to have a real signal or not. The current plan for this attribution forecast is to invoke a field significance test using a Monte Carlo analysis of simulations from the climate*prediction*.net experiment (Pall et al. 2011). There are many permutations on how such a field test can be performed, however. Should it be on all events, or separately for temperature and precipitation events? Should it be applied globally for each month, or locally over a year? These questions are so numerous and complex that they must be left for another paper.

One of the most evident issues noticed by the authors during the development of this attribution forecast system is that it differs in several major respects from a standard seasonal forecast system and thus must be treated differently too. In particular, unlike a true forecast product, the output of an attribution forecast system cannot be evaluated directly against observational records. In that sense, while it is generated in a manner akin to a seasonal forecast, this attribution forecast should be interpreted more appropriately as a climate change product. The system presented here is one realisation of a weather risk attribution system (albeit running two separate model versions). Other possible realisations exist, sampling not only different atmospheric models, attributable warming estimates, and statistical estimation strategies, but also different approaches than the time-slice method used here (e.g. Christidis et al. 2011). The sensitivity of results to each of these factors can only be revealed by sampling through the various possibilities in each, akin to what is done in the projection of climate change and, on a larger scale, in weather forecasting. Such an effort must be international, and the creators of this product encourage groups around the world to develop their own research and operational attribution programmes in order to satisfy the growing request for this resource.

Acknowledgments.

We wish to thank Nikos Christidis, Alice Favre, David Frame, Simon Wilson, and Piotr Wolski for helpful discussions. We also acknowledge the U.K. Met Office for use of the atmospheric models. D.A.S. was funded by Microsoft Research with additional support from the U.S. Department of Energy's Office of Science, Office of Biological and Environmental Research and the U.S. National Oceanic and Atmospheric Administration's Climate Program Office, via the International Detection and Attribution Group, and with travel support from: the International Meetings on Statistical Climatology; the U.K. Foreign and Commonwealth Office; and the World Climate Research Programme. We also gratefully acknowledge the Water Research Commission of South Africa and the Climate Change Adaptation in Africa programme for providing supporting funds and materials to M.T. and C.L. P.A.S. was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101).

APPENDIX

Estimation of sea surface temperatures in the non-GHG forecast

The warming attributable to greenhouse gas emissions is estimated using an optimal total least squares regression analysis (Allen and Stott 2003; Stott et al. 2003) on data from the HadSST2 dataset of gridded observational measurements (Rayner et al. 2006) and output from simulations of the HadCM3 coupled ocean-atmosphere climate model (Stott et al. 2006). If $T_{OBS}(x)$ is the observed spatio-temporal SST data, $T_{GHG}(x)$ is the corresponding data from historical simulations of HadCM3 including increases in greenhouse gas concentrations, $T_{NAT}(x)$ from simulations including changes in external natural forcings (volcanic aerosols and solar luminosity), and $T_{ALL}(x)$ from simulations including both changes in greenhouse gas and natural forcing plus changes in the concentrations of sulphate aerosols, then the regression takes the form

$$T_{OBS}(x) = \beta_1 \left(T_{ALL}(x) - \nu_{ALL}(x)\right) + \beta_2 \left(T_{GHG}(x) - \nu_{GHG}(x)\right) + \beta_3 \left(T_{NAT}(x) - \nu_{NAT}(x)\right) + \nu_{OBS}(x). \tag{A1}$$

$T_{ALL}(x)$, $T_{GHG}(x)$, and $T_{NAT}(x)$ are each the ensemble mean of a four-member initial condition ensemble of simulations, resulting in residual noise $\nu_{ALL}(x)$, $\nu_{GHG}(x)$, and $\nu_{NAT}(x)$ from the limited sampling. Similarly, $\nu_{OBS}(x)$ is a noise term accounting for the limited sampling of $T_{OBS}(x)$. The βs are the regression coefficients on the transient scenarios. Because the scenarios are not entirely independent (both NAT and GHG forcings are included in the

ALL scenario), the scaling factor on the simulated response to greenhouse gas forcing is

$$\beta_{GHG} = \beta_1 + \beta_2. \tag{A2}$$

Thus a scaling factor of $\beta_{GHG} = 1$ would imply that observational constraints agree with the HadCM3 estimate of the amplitude of the spatio-temporal response to greenhouse gas forcing.

Before the regression analysis, some pre-processing is performed to ensure that the analysis focuses on larger scale features. Monthly mean HadCM3 surface air temperature data is assumed equivalent to SST data and masked according to the model's land-sea mask and the availability of HadSST2 data. All data sets are then averaged into five-year segments covering the 1899-1998 period and onto a $45° \times 45°$ longitude-latitude grid, with this new data weighted according to coverage, with a minimum of 25% internal coverage required for each grid cell. The extremely coarse spatial resolution means that only large-scale features are used in the comparison. Similar processing is performed on overlapping 100-year segments of a 550-year simulation with constant external forcing. The leading nine principal components (PCs) of this data are taken and all other data sets are projected onto these PCs before the regression analysis. In practice, other PC truncations result in almost identical estimates of the scaling amplitude, provided it is sufficiently large to represent the main features of the variability but not too large to include poorly sampled features. The regression is performed on these PCs in a procedure that optimises the signal in relation to the noise; this involves weighting in favour of lower-order PCs, thus if very low order PCs were retained these poorly sampled features would be very strongly weighted (Allen and Tett 1999).

Annual mean data are used in the estimation of the amplitude adjustment for the

HadAM3P and HadAM3 forecasts, with no seasonal variation (Pall et al. 2011). Thus we find $\beta_{GHG} = 1.3$, broadly consistent with estimates using both land and ocean temperatures and different pre-processing approaches (Stott et al. 2006). The likelihood function of this scaling factor is estimated using the covariance in 100-year segments of a separate 500-year simulation with constant external forcing; the β_{GHG} scaling factor is consistent with 1, as measured by the 90% confidence interval of 0.4–2.5, but inconsistent with zero.

REFERENCES

Allen, M., 2003: Liability for climate change. *Nature*, **421**, 891–892.

Allen, M. R. and P. A. Stott, 2003: Estimating signal amplitudes in optical fingerprinting, part I: theory. *Clim. Dyn.*, **21**, 477–491.

Allen, M. R. and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Clim. Dyn.*, **15**, 419–434.

Boko, M., et al., 2007: Africa. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, Eds., Cambridge University Press, Cambridge, U.K., 433–467.

Chase, T. N., K. Wolter, R. A. S. Pielke, and I. Rasool, 2006: Was the 2003 European summer heat wave unusual in a global context? *Geophys. Res. Lett.*, **33**, L23709, doi: 10.1029/2006GL027470.

Christidis, N., P. A. Stott, G. S. Jones, H. Shiogama, T. Nozawa, and J. Luterbacher, 2011: Human activity and anomalously warm seasons in europe. *Int. J. Climatol.*, In press.

Jones, G. S., P. A. Stott, and N. Christidis, 2008: Human contribution to rapidly increasing frequency of very warm Northern Hemisphere summers. *J. Geophys. Res.*, **113**, D02109, doi:10.1029/2007JD008914.

Jones, R. G., M. Noguer, D. C. Hassell, D. Hudson, S. S. Wilson, G. J. Jenkins, and J. F. B. Mitchell, 2004: Generating high resultion climate change scenarios using PRECIS. Tech. rep., Met Office Hadley Centre, Exeter, U. K. 40pp.

Laepple, T., S. Jewson, and K. Coughlin, 2008: Interannual temperature predictions using the CMIP3 multi-model ensemble mean. *Geophys. Res. Lett.*, **35**, L10701, doi:10.1029/2008GL033576.

Pall, P., T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. J. Hilberts, D. Lohmann, and M. R. Allen, 2011: Anthropogenic greenhouse gas contribution to flood risk in England and Wales in Autumn 2000. *Nature*, **470**, 382–385.

Perlwitz, J., M. Hoerling, J. Eischeid, T. Xu, and A. Kumar, 2009: A strong bout of natural cooling in 2008. *Geophys. Res. Lett.*, **36**, L23706, doi:10.1029/2009GL041188.

Pope, V. D., M. Gallani, P. R. Rowntree, and R. A. Stratton, 2000: The impact of new physical parametrizations in the Hadley Centre climate model – HadAM3. *Clim. Dyn.*, **16**, 123–146.

Räisänen, J. and L. Ruokolainen, 2008: Estimating present climate in a warming world: a model-based approach. *Clim. Dyn.*, **31**, 573–585.

Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. Ansell, and S. F. B. Tett, 2006: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: the HadSST2 data set. *J. Climate*, **19**, 446–469.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. C. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Clim.*, **15**, 1609–1625.

Stone, D. A. and M. R. Allen, 2005: The end-to-end attribution problem: From emissions to impacts. *Clim. Change*, **71**, 303–318.

Stott, P. A., M. R. Allen, and G. S. Jones, 2003: Estimating signal amplitudes in optimal fingerprinting. Part II: application to general circulation models. *Clim. Dyn.*, **21**, 493–500.

Stott, P. A., N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang, and F. Zwiers, 2010: Detection and attribution of climate change: a regional perspective. *Wiley Interdisciplinary Reviews: Climate Change*, **1**, 192–211.

Stott, P. A., J. F. B. Mitchell, M. R. Allen, T. L. Delworth, J. M. Gregory, G. A. Meehl, and B. D. Santer, 2006: Observational constrains on past attributable warming and predictions of future global warming. *J. Climate*, **19**, 3055–3069.

United Nations, 2007: Report of the Conference of the Parties serving as the meeting of the Parties to the Kyoto Protocol on its second session, held at Nairobi from 6 to 17 November 2006. FCCC/KP/CMP/2006/10, fCCC/KP/CMP/2006/10.

Wilson, D., 2000: The impact of a physically based microphysical scheme on the climate simulation of the meteorological office unified model. *Q. J. R. Meteorol. Soc.*, **126**, 1281–1300.

List of Figures

3 a) Map of the 17 regions for which area-averaged results are given. EEA-Europe consists of the European Econonic Area minus French territories outside of Europe. Venezuela withdrew from the G3 in 2006 but we include it here. There are a few cases of overlap between regions, denoted by the hashed shading. b-e) An example summary of the estimate of how historical greenhouse gas emissions have altered the chance of a November 2010 that is b) unusually hot, c) unusually cold, d) unusually wet, and e) unusually dry, as calculated from the HadAM3P model. This summary displays what we can say with confidence about such an event over each of the regions, rather than the estimate of the most likely contribution. 33

4 The attribution forecasts for one-month temperature over ECOWAS over the first two and a quarter years of operation. For each forecast type, the thick line marks the mean of the simulation ensemble, while the small dots denote the estimated 5-95th percentile range using a Monte Carlo procedure on the estimated t-distribution. For each month the plotted quantities progress from the three-month lead forecast (left) to the hindcast two months afterward (right), when available. Dark lines denote real forecasts from either model while light lines denote non-GHG forecasts; the dark lines are used to denote the merged quantity in the FAR (fraction attributable risk) panels. The climatological reference period is 1960-2008, 1960-2009, and 1960-2010 for all months in 2009, 2010, and 2011 respectively. The climatological cold and hot probability is defined to be 10%. Some values in the FAR panels are below the plotted range. 34

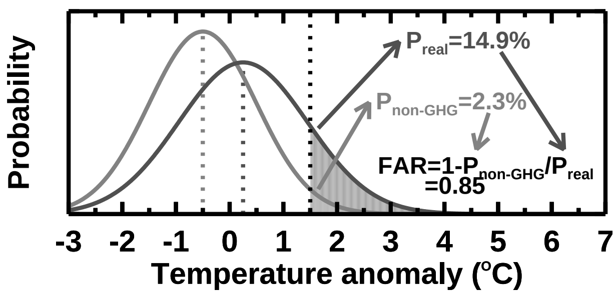

FIG. 1. Outline of the comparison made in the attribution forecast product. An estimate of the distribution of possible temperatures is made under the current real climate (red curve) and another estimate is made under a scenario in which human activities had never emitted greenhouse gases (blue curve). For both scenarios the probability of exceeding some threshold is taken, in this case a threshold for historically unusual hot events (shaded). The two probabilities are compared using the fraction attributable risk (FAR) measure described in the text. In this example, 85% of the chance of an unusually hot event is attributable to the anthropogenic greenhouse gas emissions.

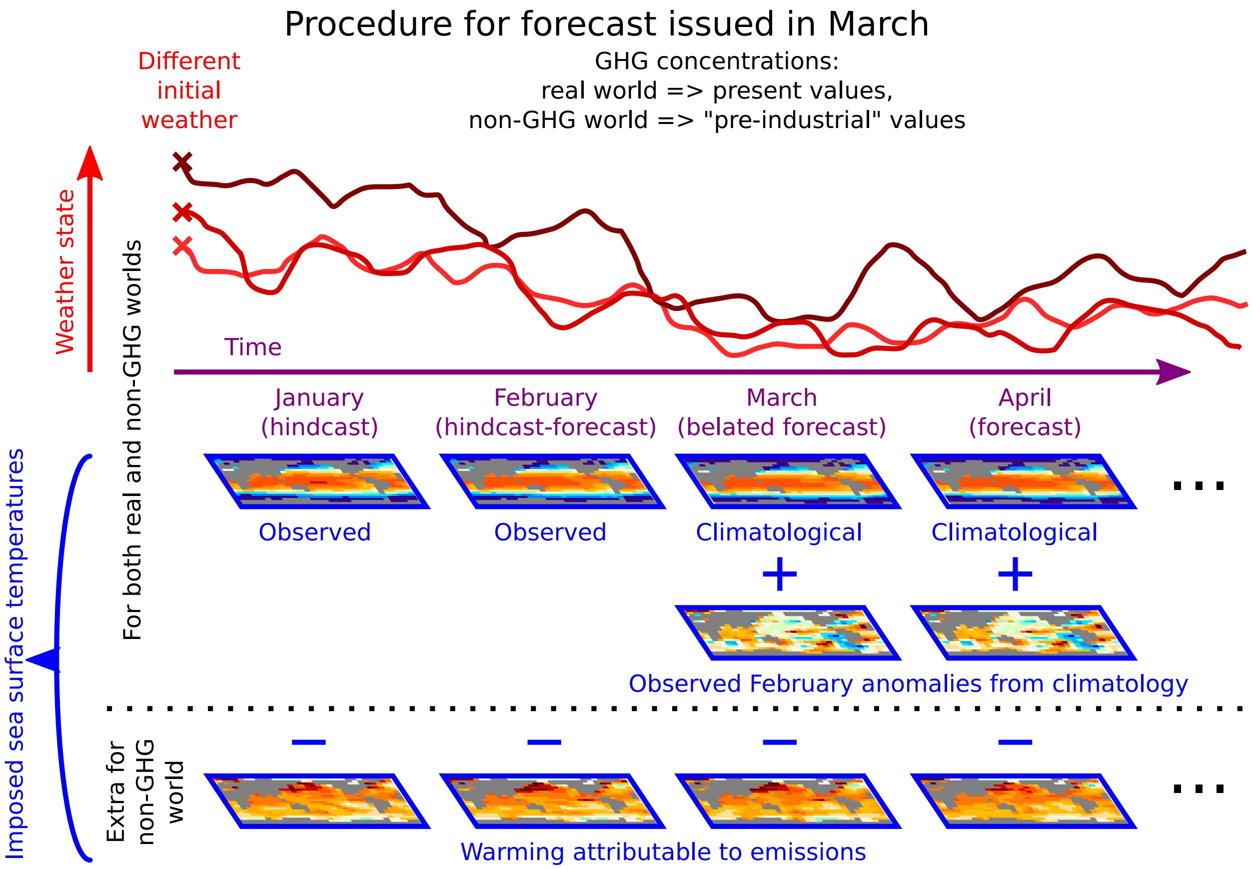

FIG. 2. Schematic of the procedure for producing the seasonal forecasts used in this attribution forecast system. Observed greenhouse gas concentrations are imposed for the real forecasts, while pre-industrial concentrations are imposed for the non-GHG forecasts. Sea surface temperatures for the non-GHG forecasts are modified by subtracting an estimate of the geographical warming attributable to historical greenhouse gas emissions.

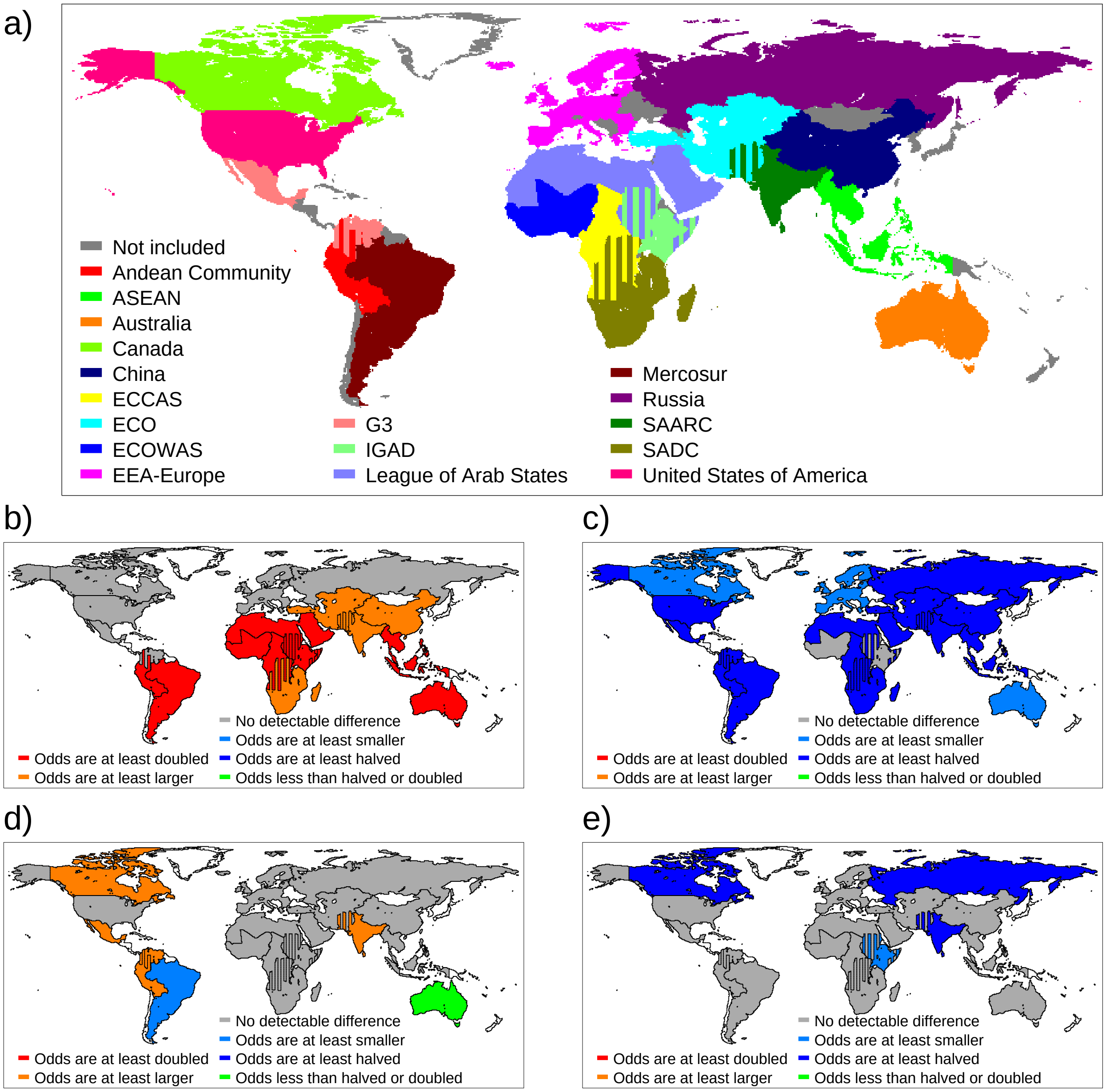

FIG. 3. a) Map of the 17 regions for which area-averaged results are given. EEA-Europe consists of the European Econonic Area minus French territories outside of Europe. Venezuela withdrew from the G3 in 2006 but we include it here. There are a few cases of overlap between regions, denoted by the hashed shading. b-e) An example summary of the estimate of how historical greenhouse gas emissions have altered the chance of a November 2010 that is b) unusually hot, c) unusually cold, d) unusually wet, and e) unusually dry, as calculated from the HadAM3P model. This summary displays what we can say with confidence about such an event over each of the regions, rather than the estimate of the most likely contribution.

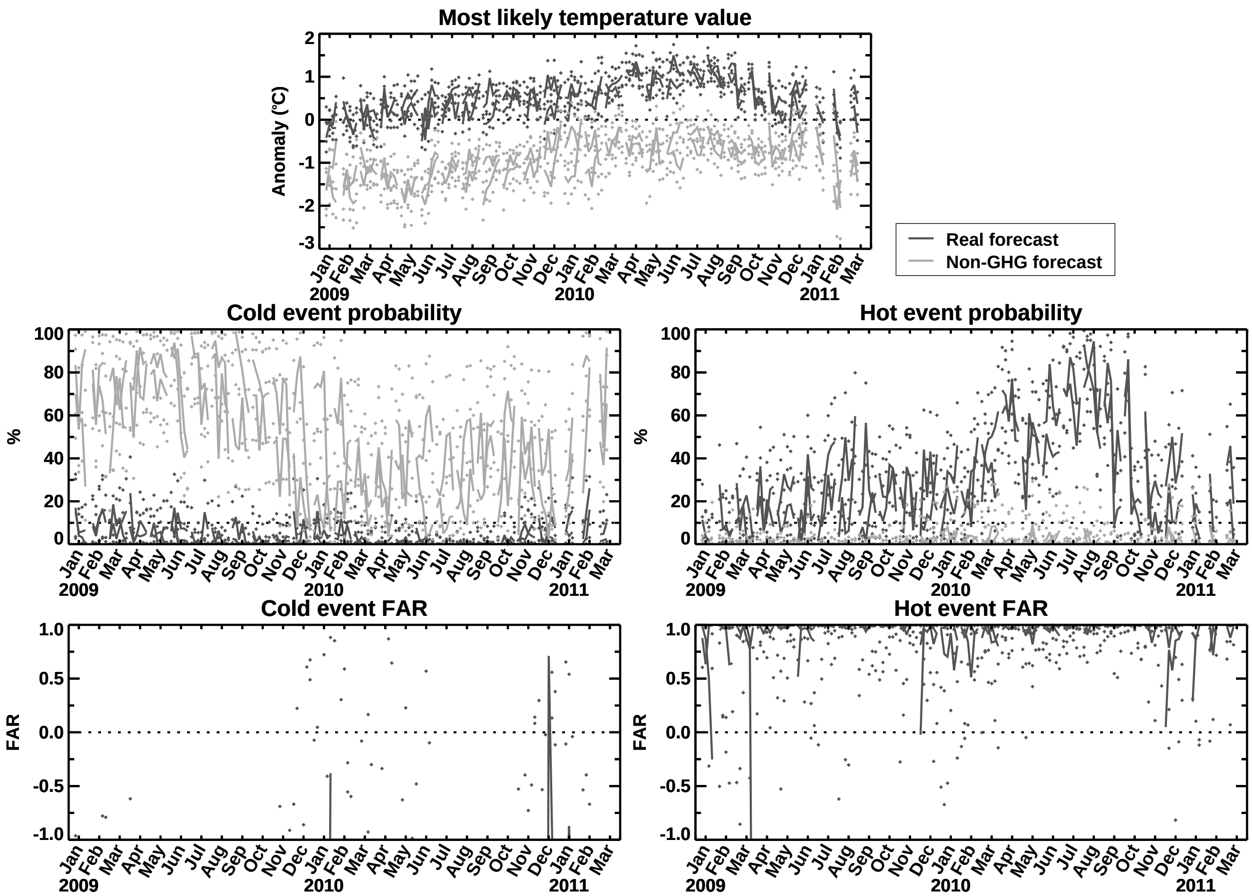

FIG. 4. The attribution forecasts for one-month temperature over ECOWAS over the first two and a quarter years of operation. For each forecast type, the thick line marks the mean of the simulation ensemble, while the small dots denote the estimated 5-95th percentile range using a Monte Carlo procedure on the estimated t-distribution. For each month the plotted quantities progress from the three-month lead forecast (left) to the hindcast two months afterward (right), when available. Dark lines denote real forecasts from either model while light lines denote non-GHG forecasts; the dark lines are used to denote the merged quantity in the FAR (fraction attributable risk) panels. The climatological reference period is 1960-2008, 1960-2009, and 1960-2010 for all months in 2009, 2010, and 2011 respectively. The climatological cold and hot probability is defined to be 10%. Some values in the FAR panels are below the plotted range.

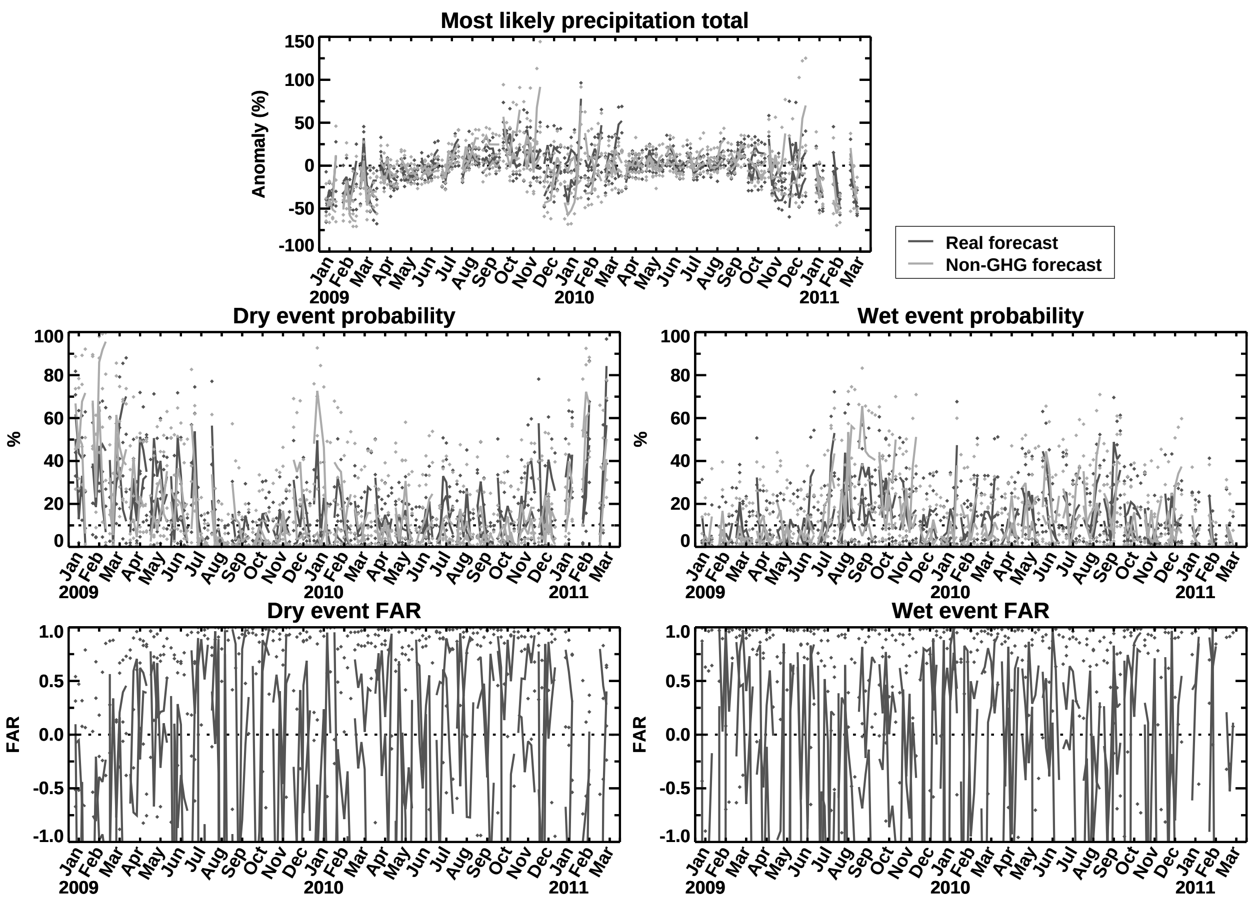

FIG. 5. As in Fig. 4 but for regional mean precipitation.

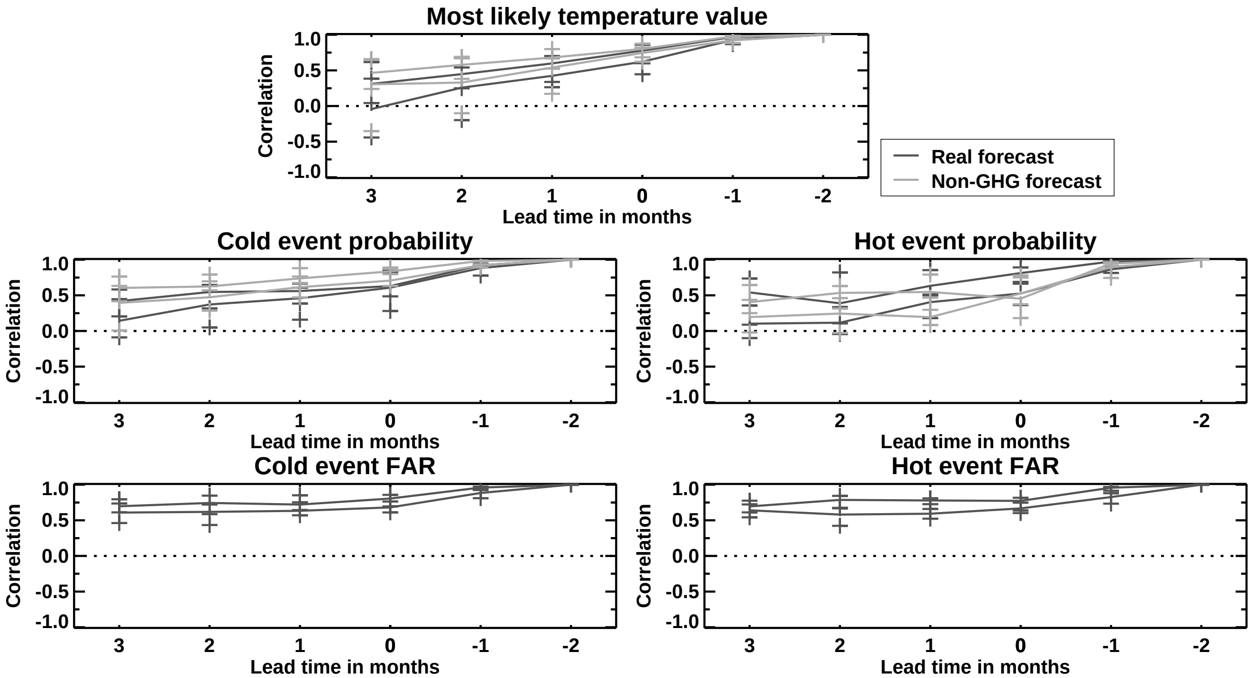

FIG. 6. The spatial correlation over the 17 regions in Fig. 3a of various forecast quantities against the final hindcast, as a function of lead time. The average values across all months are shown at each lead time for clarity; the crosses denote the approximate quartiles across the monthly samples. Only the direct estimates are shown here, because the Monte Carlo sampling used for uncertainty estimates for individual months and regions do not account for spatio-temporal covariance. All months from January 2009 to December 2010 are included, but values are not available for all lead times for the first few months. FAR values are first transformed to $\log(1 - \mathrm{FAR})$, in order to moderate the effect of negative outliers.

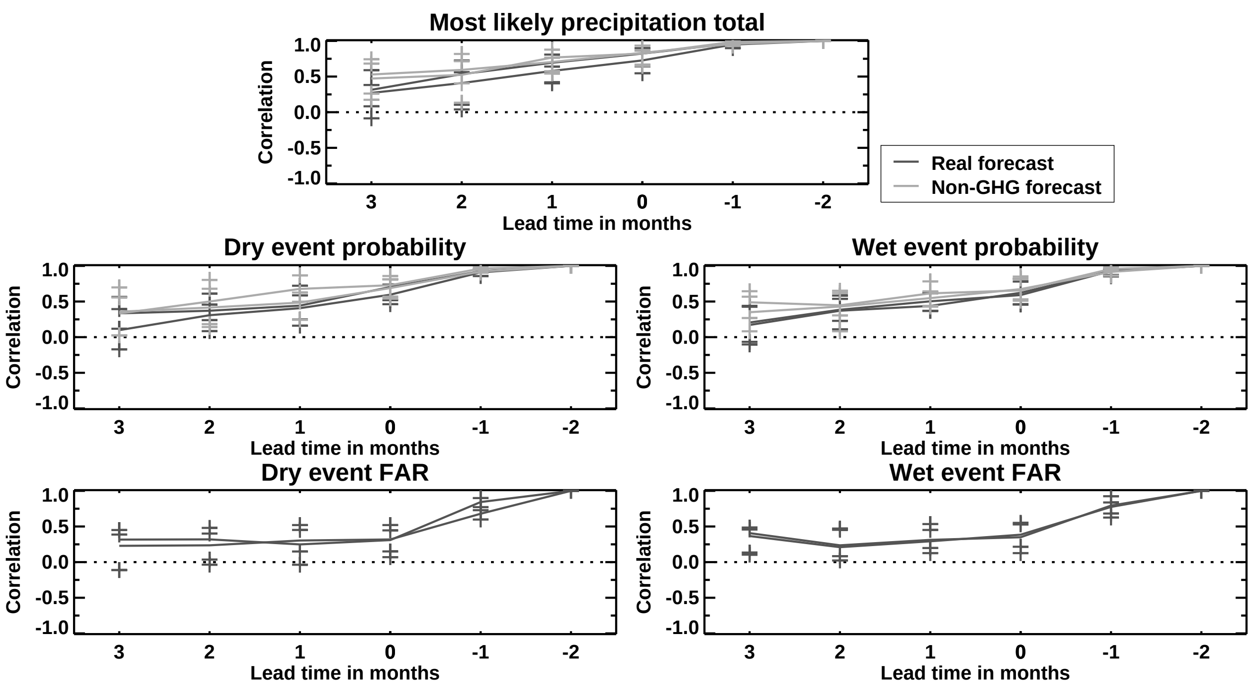

FIG. 7. As in Fig. 6 but for precipitation over the 17 regions.