What is the most useful approach for forecasting hydrological extremes during El Niño?

Environmental Research Communications

OPEN ACCESS

RECEIVED
21 January 2019

REVISED
28 February 2019

ACCEPTED FOR PUBLICATION
20 March 2019

PUBLISHED
4 April 2019

LETTER

What is the most useful approach for forecasting hydrological extremes during El Niño?

Rebecca E Emerton[1,2], **Elisabeth M Stephens**[1] and **Hannah L Cloke**[1,3,4]

1 Department of Geography and Environmental Science, University of Reading, United Kingdom
2 European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
3 Department of Meteorology, University of Reading, United Kingdom
4 Department of Earth Sciences, Uppsala University, Sweden

E-mail: rebecca.emerton@reading.ac.uk

Keywords: hydrological extremes, El Niño, seasonal forecasting, climate variability

Supplementary material for this article is available online

Abstract

In the past, efforts to prepare for the impacts of El Niño-driven flood and drought hazards have often relied on seasonal precipitation forecasts as a proxy for hydrological extremes, due to a lack of hydrologically relevant information. However, precipitation forecasts are not the best indicator of hydrological extremes. Now, two different global scale hydro-meteorological approaches for predicting river flow extremes are available to support flood and drought preparedness. These approaches are statistical forecasts based on large-scale climate variability and teleconnections, and resource-intensive dynamical forecasts using coupled ocean-atmosphere general circulation models. Both have the potential to provide early warning information, and both are used to prepare for El Niño impacts, but which approach provides the most useful forecasts? This study uses river flow observations to assess and compare the ability of two recently-developed forecasts to predict high and low river flow during El Niño: statistical historical probabilities of ENSO-driven hydrological extremes, and the dynamical seasonal river flow outlook of the Global Flood Awareness System (GloFAS-seasonal). Our findings highlight regions of the globe where each forecast is (or is not) skilful compared to a forecast of climatology, and the advantages and disadvantages of each forecasting approach. We conclude that in regions where extreme river flow is predominantly driven by El Niño, or in regions where GloFAS-seasonal currently lacks skill, the historical probabilities generally provide a more useful forecast. In areas where other teleconnections also impact river flow, with the effect of strengthening, mitigating or even reversing the influence of El Niño, GloFAS-seasonal forecasts are typically more useful.

1. Introduction

Global overviews of upcoming flood and drought events provide valuable information for organisations working at the global scale, across a range of water-related sectors from agriculture to humanitarian aid. Producing such forecasts at the global scale has only become possible in recent years due to the integration of meteorological and hydrological modelling capabilities, improvements in data, satellite observations, and increased computer power [1–4]. While several forecasting centres now produce operational forecasts of floods in the medium-range, up to ~2 weeks ahead [5], earlier indications, many weeks or even months in advance, could be beneficial for water resources and disaster risk management.

Broadly speaking, there are two key ways to extend the predictability of river flow and provide earlier indications of flood hazard: statistical forecasts, typically based on large-scale climate variability and teleconnections, and dynamical forecasts using coupled ocean-atmosphere general circulation models (GCMs).

Operational seasonal forecasts, using both statistical and dynamical approaches, are widely available for meteorological variables, but the hydrology is often not represented, particularly for large or global scales. This means that forecasts of precipitation are often used as a proxy for flooding. However, research has shown that the link between precipitation and flood magnitude is nonlinear [6], and as such, precipitation may not be the best indicator of potential flood hazard [7]. Recently, there has been an effort to provide the equivalent early awareness information for hydrological variables, as exists for meteorological variables.

Global scale statistical forecasts often rely on ENSO (El Niño Southern Oscillation) teleconnections. ENSO is the largest signal of interannual climate variability [8]; a phenomenon in which sea surface temperatures (SSTs) in the central and eastern equatorial Pacific fluctuate between warm (El Niño) and cool (La Niña) conditions. ENSO is known to influence various aspects of weather and climate, including river flow [9] and flooding [10–12], worldwide. Historical probabilities, such as those provided by the International Research Institute for Climate and Society [13] for precipitation and temperature, are an example of a statistical forecast that is often used for El Niño preparedness activities.

In response to a lack of hydrologically-relevant information on ENSO impacts, Emerton *et al* [14] estimated historical probabilities of high and low river flow during El Niño and La Niña. These historical probabilities provide statistical forecasts of extreme river flow, based on the links between past ENSO events and river flow across the globe.

The recent move towards the development of coupled atmosphere-ocean-land models means that it is also now becoming possible to produce seasonal dynamical hydro-meteorological forecasts. The first operational global seasonal river flow forecasting system was implemented in 2017, as part of the Global Flood Awareness System (GloFAS; [1]). GloFAS-Seasonal [15] provides openly-available dynamical forecasts of high and low river flow out to 4 months ahead by forcing a hydrological river routing model with seasonal forecast output from a GCM.

Both forecast approaches have the potential to provide early warning information through provision of hydrologically-relevant global scale forecasts, and both are used to prepare for El Niño impacts, but more research is required to explore whether statistical forecasts are able to provide stronger indications of changes in hydrological extremes than seasonal dynamical forecasts.

This study uses river flow observations to compare the potential usefulness of these two global scale forecasts of river flow during El Niño events. Both forecasts are compared to a forecast of climatology and then against each other, using an event-based verification approach.

2. Forecasting approaches

2.1. Dynamical approach: GloFAS-Seasonal

GloFAS-Seasonal provides global scale seasonal hydro-meteorological forecasts using a GCM. Implemented in 2017, it is run by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the European Commission Joint Research Centre (JRC), as part of the Copernicus Emergency Management Services. It uses surface and subsurface runoff forecasts from ECMWF's latest seasonal meteorological forecasting system, SEAS5 [16, 17], to drive a river routing model, Lisflood [18], producing forecasts of river flow out to 4 months ahead. The GloFAS website (www.globalfloods.eu, see figure 1(a) for example) provides openly available seasonal outlooks of the likelihood of exceeding / falling below the climatological thresholds of high (80th percentile) and low (20th percentile) weekly-averaged river flow.

For this study, we make use of the GloFAS-Seasonal *reforecasts*, which were produced using the SEAS5 reforecasts [15, 19] initialised with the ERA5-R river flow reanalysis [15]. ERA5 [20] is currently still in production, and as such, 34 years of data were available with which to produce the reforecasts: 1981–1983, and 1986–2016.

2.2. Statistical approach: historical probabilities

Historical Probabilities (hereafter referred to as HistProbs) provide information about typical El Niño impacts based on historical evidence [21, 22]. The probability of an impact is predicted based on the frequency of occurrence during past El Niños.

The HistProbs of high and low river flow during ENSO events from Emerton *et al* [14] have been reproduced in this study for weekly-averaged river flow, in order to directly compare them with GloFAS-Seasonal. Following the method of Emerton *et al* [14], we used the ERA-20CM-R 10-member, 110-year (1901–2010) river flow climatology to calculate the upper and lower 20th percentile of river flow for each grid point. We then calculate, for each week of an El Niño, the percentage of historical El Niños during which the high or low flow threshold was exceeded. The use of ERA-20CM-R allows for more El Niños to be included in the calculation of the HistProbs, with 30 El Niños identified over the 110-year period. An El Niño is identified when the SST anomaly

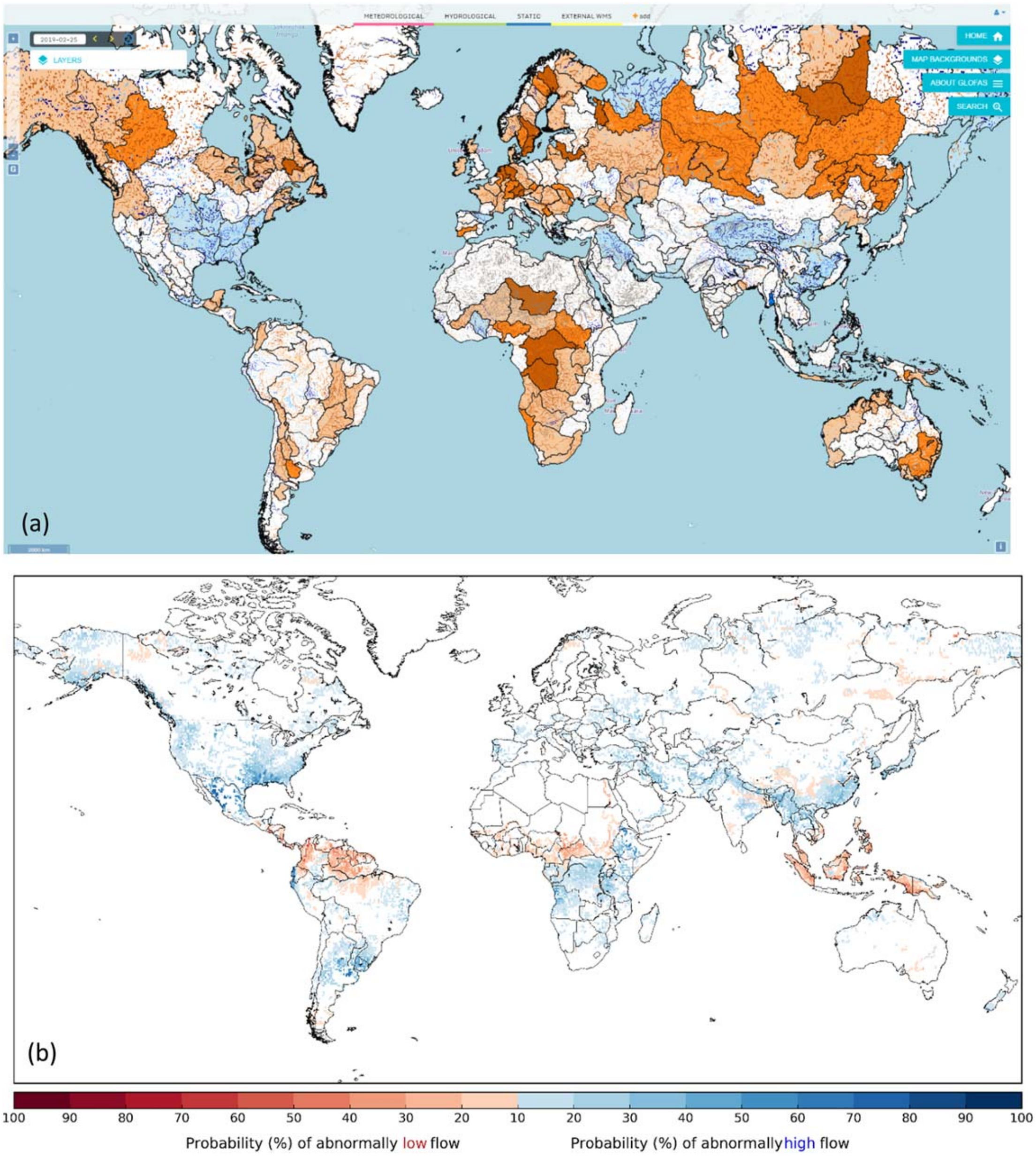

Figure 1. (a) Example of the GloFAS-Seasonal forecast website, displaying the probability of exceeding both the high (blue) and low (orange) river flow thresholds. (b) Example of the HistProbs forecast for one week during an El Niño. The map displays the probability of exceeding both the high (blue) and low (red) river flow thresholds. While both examples display forecasts for February during an El Niño event, (a) indicates the maximum probability over the 4-month lead time for a forecast started in February, and (b) indicates the probability for one week in February only.

in the central equatorial Pacific Ocean (Niño3.4 region; 5° S–5° N, 170°–120°W) exceeds +0.5 °C for at least five consecutive (overlapping) three-month periods.

The HistProbs (figure 1(b)) were estimated for each grid point, through calculation of the percentage of the 30 historical El Niños in which the river flow exceeded the high flow threshold, or fell below the low flow threshold, during the same week. This was repeated for each of the 10 ensemble members of ERA-20CM-R. The ensemble mean probability was then interpolated from the 0.5° (~50 km) resolution of ERA-20CM-R, to the 0.1° (~10 km) resolution of GloFAS-Seasonal; it is this higher-resolution ensemble mean that is used throughout this study.

3. Evaluation data and methods

This study evaluates the predictability of hydrological extremes during El Niño in both GloFAS-Seasonal and the HistProbs by assessing the ability of each system to predict high and low river flow, with the correct timing, during an El Niño. The ability of a forecast to predict events of the correct category is referred to as the 'potential usefulness' and is of particular importance for decision-making purposes [23].

The potential usefulness is calculated using the relative operating characteristic (ROC) curve, based on ratios of the probability of detection (POD) and the false alarm rate (FAR) [24]. These ratios are calculated by assessing whether a forecast correctly predicted an observed event, or whether it missed the event or provided a false alarm, and allow for estimation of the probability that an event will be predicted. The POD (equation (1)) and FAR (equation (2)) are calculated as follows:

$$POD = \frac{\text{hits}}{\text{hits} + \text{misses}} \tag{1}$$

$$FAR = \frac{\text{false alarms}}{\text{false alarms} + \text{correct negatives}} \tag{2}$$

where a *hit* is defined when the forecast correctly predicted flow exceeding [falling below] the 80th [20th] percentile during the same week that the observed river flow exceeded [fell below] the 80th [20th] percentile of the observations at that location. It follows that a *miss* is defined when an event was observed but the forecast did not exceed the threshold, a *false alarm* when the forecast exceeded the threshold but no event was observed, and a *correct negative* when no event was observed and the forecast did not exceed the threshold.

The ROC curve is constructed from the FAR (horizontal axis) and POD (vertical axis) at different probability thresholds (in this case, in 10% bins), therefore providing information on the likelihood that an event will be predicted at a given probability threshold. The geometrical area under the ROC curve (AROC; $0 \leqslant \text{AROC} \leqslant 1$) provides a summary statistic for the performance of a probabilistic forecast, where a forecast that correctly predicts every observed event (with no recorded false alarms or missed events) would have an AROC of 1. An $\text{AROC} < 0.5$ indicates that the skill of the forecasts is less than a forecast of climatology, which has an AROC of 0.5.

The AROC is used to infer the potential usefulness of the forecast; a forecast that is more skilful than a forecast of climatology is said to be potentially useful, whereas a forecast that is less skilful than a forecast of climatology is not useful. This approach has previously been used in the evaluation of seasonal river flow forecasts [15, 23]. Often, seasonal forecasts are provided in terms of the likelihood that a given variable will be above or below normal (based on terciles) in the coming months. The evaluation technique used in this study presents a significant challenge for both forecasting systems, requiring that they predict more extreme weekly-averaged river flow, in the same week as that in which it was observed, several weeks to months ahead.

3.1. Observed data

The two forecasts are evaluated over the same 34-year period (1981–2015), using river flow observations obtained from the Global Runoff Data Centre (GRDC; [25]), alongside observations that have been made available to GloFAS [15]. To ensure a large enough sample size for the forecast evaluation, alongside the best possible spatial coverage, the following criteria are applied to the data:

- The weekly-averaged river flow record at each station must contain data for at least 50% (17 years) of the evaluation period, in order to calculate the observed high and low flow thresholds (80th and 20th percentiles) for each station, and for each week of the year.
- The weekly-averaged river flow record at each station must contain at least 6 El Niños over which to evaluate the forecasts.
- The upstream area of the corresponding grid point in the model river network must be at least 1500 km^2.

Data from human-influenced rivers have not been removed, as we are interested in identifying the ability of both forecasting approaches to predict observed events, rather than their ability to represent natural flow. Of the 2355 stations in the database, ~1250 contain enough data to meet the above criteria and are used in this study.

3.2. Calculating potential usefulness of GloFAS-Seasonal

To evaluate the potential usefulness of GloFAS-Seasonal we calculate the AROC for each season during an El Niño using the observations as a benchmark. The AROC for a season is calculated by grouping together forecasts for every week during the season for all 11 El Niño events between 1981 and 2015.

The AROC is also calculated for lead times of 1–4 months ahead, by selecting the GloFAS-Seasonal weekly-averaged river flow forecast that would have been available 1, 2, 3 and 4 months ahead of each week of the El Niño event. For example, for the fourth week in January the forecast available one month ahead would be the fourth week of the forecast produced at the start of January, the forecast available two months ahead would be the 8th week of the forecast produced in December, and three months ahead the 12th week of the forecast produced in November. Following the same method, for the second week in December, the forecast available one month ahead for that week, would be the 6th week of the forecast produced in November. This is necessary because while GloFAS-Seasonal predicts weekly-averaged river flow, the forecasts are updated just once per month.

3.3. Calculating potential usefulness of the historical probabilities

To evaluate the potential usefulness of the HistProbs we calculate the AROC for each season during an El Niño event using the observations as a benchmark.

The HistProbs are a 'static' forecast, that is, the forecasts do not change with lead time and there is just one probability for high or low river flow during each week of an El Niño. As such, the AROC is calculated by comparing the river flow in each week of the 11 El Niño events in the observations, with the HistProb of high or low river flow for the corresponding week of the year. The AROC for a season is calculated by grouping together forecasts for every week during the season, for all 11 El Niño events between 1981 and 2015.

4. Results

The results presented in this section compare the 'potential usefulness' of both GloFAS-Seasonal and the HistProbs during an El Niño. The following criteria are used to define the 'most useful' forecast, based on the null hypothesis that the potential usefulness of the two forecasts is not significantly different:

- If GloFAS-Seasonal has an AROC > 0.5 and the HistProbs < 0.5, or both exceed 0.5 but GloFAS-Seasonal has an AROC > 0.1 larger than the HistProbs, *GloFAS-Seasonal* is most useful
- If the HistProbs have an AROC > 0.5, and GloFAS-Seasonal < 0.5, or both exceed 0.5 but the HistProbs have an AROC > 0.1 larger than GloFAS-Seasonal, the *HistProbs* are most useful
- If both forecasts have an AROC > 0.5, and within 0.1 of each other, both are useful and *similar*
- If both forecasts have an AROC < 0.5, *neither* are useful

The statistical significance of the difference in AROC between the two forecasts was investigated using a bootstrap procedure. For each season and each observation location, all available forecasts for both GloFAS-Seasonal (132 forecasts per season across the 11 El Niño events, at each lead time of 1–4 months ahead) and the HistProbs (143 forecasts per season, providing an independent probability for each week of the season, but the same probability for a given week across all 11 El Niño events), were resampled with replacement, and the resulting AROC was calculated. This process was repeated 1000 times.

Figure 2 displays box plots of the global bootstrapped AROC differences (GloFAS-Seasonal - HistProbs) at lead times of 1 and 3 months ahead for high and low river flow in MAM during an El Niño. These results indicate that, aggregated globally, there is evidence that GloFAS-Seasonal provides an improved AROC for forecasts of both high and low river flow, however, this is not statistically significant. For high [low] flow 3 months ahead, the median AROC difference is 0.32 [0.18], across all stations where at least one of the two forecasts is potentially useful (AROC > 0.5). Further assessment of the bootstrapped AROC differences for each individual station indicates that at ~95.5% of the locations where the median AROC difference of the 1000-bootstrapped sample exceeds ± 0.1, the choice of the most useful forecast is statistically significant to the 95% confidence level (at $\sim$ 4.5% of stations, this is not the case, and using a threshold of ± 0.1 does not provide a statistically significant result). At locations where the median AROC difference is <0.1, choosing a 'most useful' forecast would not provide a statistically significant result, and therefore it is reasonable to class the forecasts as 'similar' (or 'not useful' depending on the AROC values).

4.1. Probability of high flow

Figure 3(a) indicates that for forecasts of high river flow 3 months ahead, for MAM during an El Niño, the most useful forecast varies by region, and there are many locations where neither forecast is more skilful than a forecast of climatology (grey dots).

Across much of North America, the HistProbs provide a more useful forecast of high river flow than GloFAS-Seasonal, except along the east coast, where GloFAS-Seasonal forecasts are more skilful. In the regions of South America that are more likely to see high flow during an El Niño, GloFAS-Seasonal is more useful at several locations, particularly in northern Peru, while the HistProbs are more useful in southern Brazil. In Europe, the HistProbs are more useful in the west, and GloFAS-Seasonal is more useful in the east.

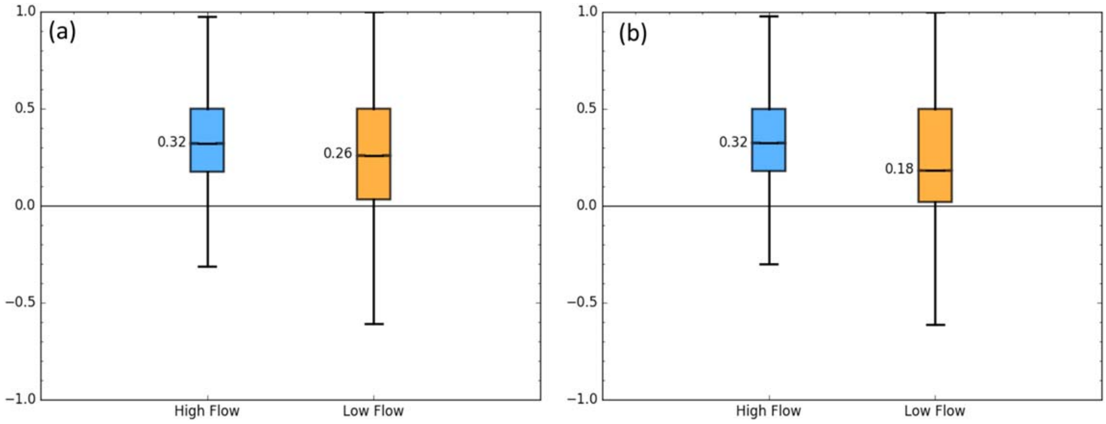

Figure 2. Box plots of the AROC differences (GloFAS - HistProbs) at lead times of (a) 1, and (b) 3 months ahead for both high (blue) and low (orange) river flow in MAM globally (for stations where at least one of the forecasts has an AROC >0.5), calculated from a bootstrap procedure that was repeated 1000 times using resampling of the 132 [144] GloFAS-Seasonal [HistProbs] forecasts, with replacement. The bottom and top of the boxes correspond to the 25th and 75th percentiles, respectively. The notch represents the 95% confidence interval around the median from a 1000-bootstrapped sample.

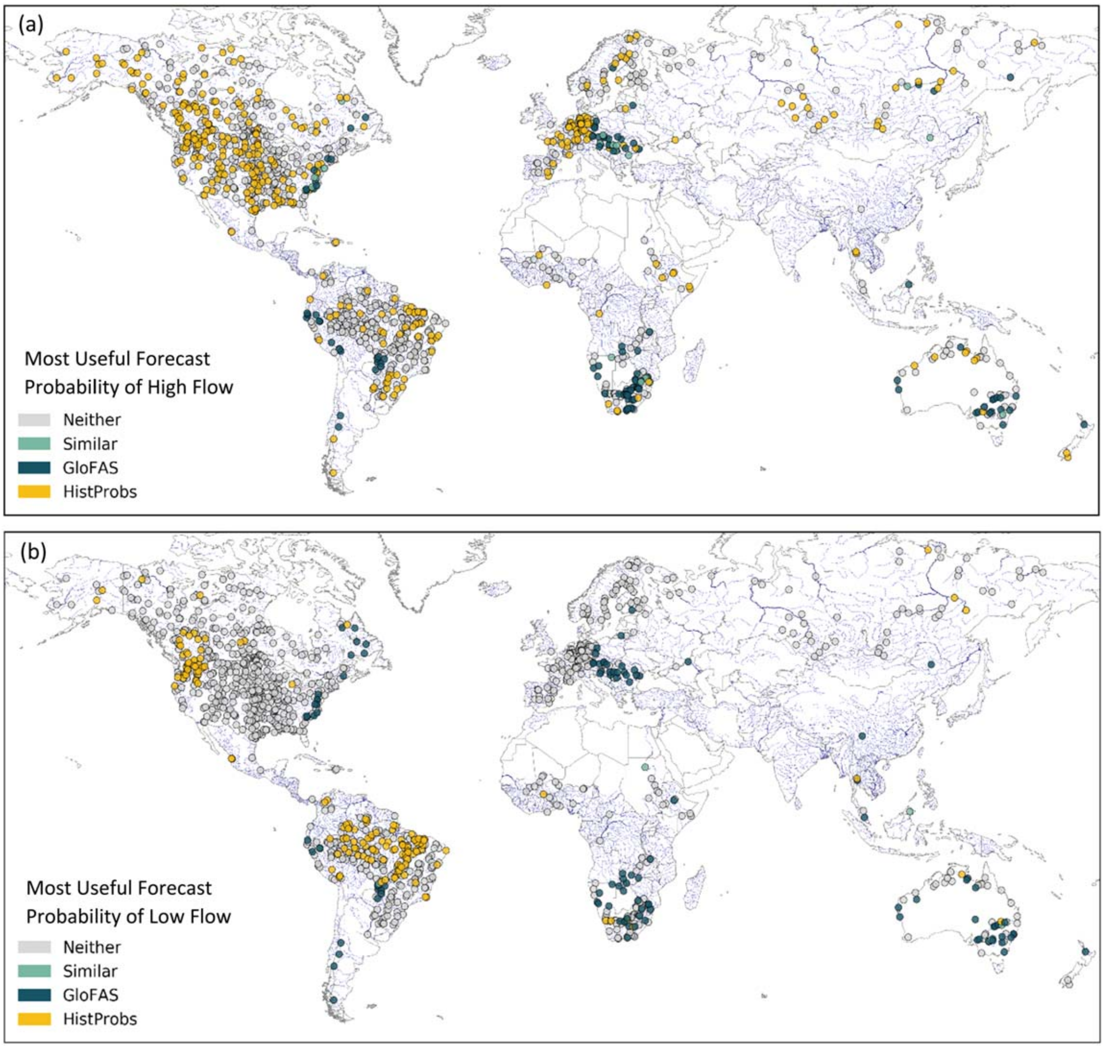

Figure 3. Maps indicating the most potentially useful forecast 3 months ahead for (a) high river flow (>80th percentile of climatology) and (b) low river flow (<20th percentile of climatology) in MAM, at each observation location.

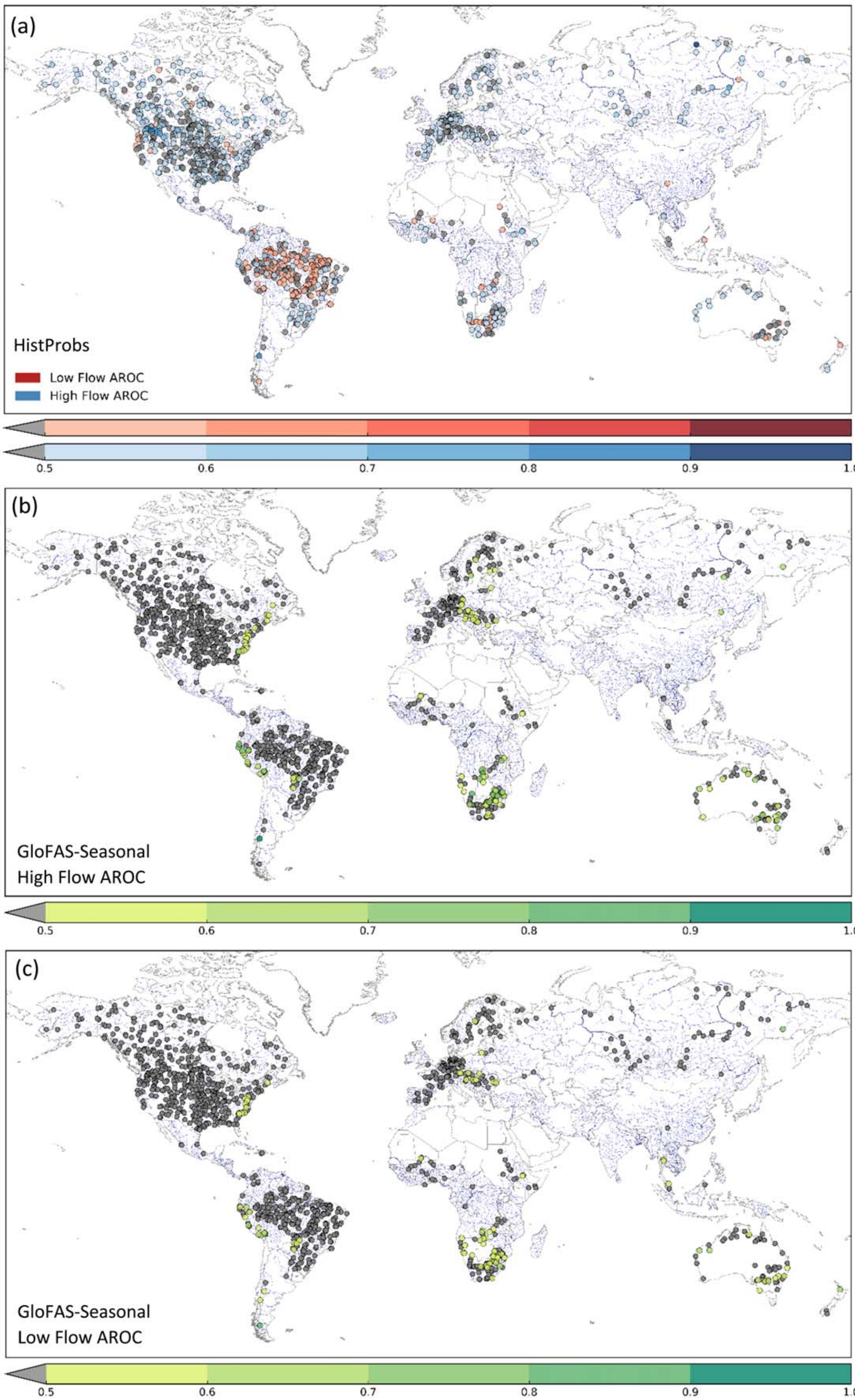

Figure 4. Maps indicating (a) the AROC of the HistProbs for both high river flow (>80th percentile of climatology, blue) and low river flow (<20th percentile of climatology, red) in MAM, (b) the AROC of GloFAS-Seasonal 3 months ahead for high river flow in MAM, and (c) the AROC of GloFAS-Seasonal 3 months ahead for low river flow in MAM. On all 3 maps, the darker the colour, the higher the skill (and potential usefulness) of the forecast. Grey dots indicate that the forecast is not useful at that location; i.e. the forecast has an AROC $\leqslant$ 0.5.

Figure 4 shows the AROC values for each forecast at locations where they are more skilful than climatology. Generally, the AROC for the HistProbs lies in the 0.5–0.6 range, meaning they are only marginally more skilful than climatology, except in some small regions, such as north-west USA where the AROC reaches 0.7–0.8. There are also regions where GloFAS-Seasonal forecasts are only marginally more skilful than climatology, such as the east coast of North America, but the majority of locations show an AROC of 0.6–0.8.

Results for all seasons and lead times are provided in the supplementary material. In general, the results tend to be consistent with lead time, although as may be expected, the skill of GloFAS-Seasonal is reduced at longer lead times in some locations. The skill of both forecasts varies more significantly with season than with lead time. Figure S1 (available online at stacks.iop.org/ERC/1/031002/mmedia) shows that areas where neither is useful are more widespread in JJA, when El Niño typically begins to develop, and both become more widely skilful through SON and DJF as El Niño intensifies. The timing of El Niño onset varies from one event to the next, which results in more uncertainty in the HistProbs for JJA than for other seasons. For GloFAS-Seasonal, forecasts made ahead of JJA are likely to be more uncertain due to uncertainty in forecasting the timing and magnitude of El Niño. Forecasts of El Niño produced before and during spring tend to be much less successful (the infamous 'spring predictability barrier'), although the cause of this remains controversial [26–29].

4.2. Probability of low flow

Figure 3(b) provides the same results for forecasts of low river flow. Locations where neither forecast is more skilful than climatology are more widespread. However, some of these regions, such as the USA, are more likely to see high river flow during an El Niño.

In the low flow regions in the USA, South America, Africa and Australia, there are locations at which the HistProbs are potentially useful (see figures 2(b) and 3(a)), but the variability from one location to the next is much higher than for forecasts of high river flow. The skill of the HistProbs increases during and after the peak of El Niño, in DJF and MAM. This is likely due to the delayed response of river flow to the El Niño-driven precipitation, which is more prominent for low flow and drought, than for high flow and flooding. This is also reflected in the HistProbs themselves (not shown), which highlight the lagged response of river flow to El Niño, and that the influence on rivers can continue beyond the return to neutral ENSO conditions.

In general, GloFAS-Seasonal is the most useful forecast for low river flow in the same regions as for high flow, while the HistProbs are more useful over the Amazon basin and north-west USA, particularly in DJF and MAM. Interestingly, figure 4 indicates that for low river flow, the AROC values for the two forecasts tend to be very similar; within ± 0.2. The GloFAS-Seasonal AROC values are similar to those for high river flow, reaching 0.6–0.8 in many locations, but where the HistProbs are potentially useful, the AROC can also reach 0.6–0.7, and 0.8 at some locations. As with the forecasts for high river flow, some variations in the results are seen with lead time, but these are less significant than the variations from one season to the next. Additional results for all seasons and lead times are provided in the supplementary material.

4.3. Discussion

The results presented in sections 4.1 and 4.2 highlight areas of the globe where potentially useful forecasts of hydrological extremes during El Niño are available, and indicate that the skill of both forecasts varies by region and season, and to some extent with lead time.

Overall, where there is a strong El Niño influence on river flow the HistProbs are able to provide a potentially useful forecast of high flow in regions where GloFAS-Seasonal lacks skill. The HistProbs presented here are estimated based only on SSTs in the Niño3.4 region in the central Pacific, and therefore are not able to reflect ENSO diversity. For example, flooding in Peru is known to be driven by El Niños which exhibit larger SST anomalies in the eastern Pacific than the central Pacific.

In fact, the impact of ENSO diversity provides some indication as to why GloFAS-Seasonal is more useful than the HistProbs in specific regions (e.g. northern Peru, east coast of North America, southern Africa, eastern Europe and Australia). All of these regions are similarly, if not more strongly, influenced by other modes of climate variability on seasonal to decadal timescales, such as the Indian Ocean Dipole (IOD), North Atlantic Oscillation (NAO) and Pacific Decadal Oscillation (PDO). A GCM, by design, should be able to better represent the impact of these other modes of variability on weather patterns, whereas the HistProbs are conditioned only on whether an El Niño was present in the historical record, and not the interaction with any other modes of climate variability.

Wang *et al* [30] show that generally, an El Niño combined with a warm phase PDO gives a similar, but stronger, pattern of influence on wet-dry anomalies. However, in some regions the wet-dry anomaly during El Niño is reversed when combined with a cold phase PDO. In regions where the impact is similar regardless of the PDO phase, the HistProbs are generally more useful than GloFAS-Seasonal, particularly for high flow. Regions where the wet-dry anomaly is reversed depending on the PDO phase, tend to correspond to those where

GloFAS-Seasonal is more useful. There are some exceptions, however, such as high latitude Canada and Siberia, where the HistProbs are more useful. These correspond to regions where GloFAS-Seasonal has been shown to generally be less skilful than climatology [15]. As the PDO is a decadal oscillation varying on much longer timescales than ENSO, it is likely to influence El Niño impacts over several events in turn. It is therefore a potential source of uncertainty in the HistProbs (see [14]), as they are conditioned only on ENSO, and a change in the PDO may represent a change in the climate state from the period over which the HistProbs are estimated. The state of the PDO, however, is accounted for within a dynamical seasonal forecasting system.

Further regions where GloFAS-Seasonal tends to provide a more useful forecast, for both high and low river flow, include southern Africa and Australia, which are known to be influenced by the IOD [31–34]. Saji and Yamagata [35] show that the IOD impacts African rain variability regardless of the ENSO phase, but ENSO only has an impact when combined with an IOD event. As mentioned previously, the skill can vary significantly by season, and recent research [36] has also shown that SEAS5, the meteorological forecast input of GloFAS-Seasonal, is more skilful at predicting short rains (OND) than long rains (MAM) in east Africa, as the short rains have much stronger teleconnections with ENSO and the IOD than the long rains. In Australia and south-east Asia, the IOD increases [decreases] the chance of rainfall during its negative [positive] phase [37]. Additionally, the NAO has been shown to influence flood occurrence in Europe, with extreme rainfall more likely in parts of eastern Europe during the positive phase of the NAO [38].

While the HistProbs are able to, in general, provide a more skilful forecast than climatology in the majority of regions influenced by El Niño, there are locations where GloFAS-Seasonal is less skilful than climatology in all seasons and at all lead times. In these locations, GloFAS-Seasonal is unable to correctly predict the magnitude, and/or the timing, of the observed events. A study by Hirpa *et al* [39] identifies regions of bias in GloFAS river flow simulations. Regions of negative bias generally correspond to those where GloFAS-Seasonal is not skilful in this study. Future work should determine whether calibration of GloFAS, such as that presented by Hirpa *et al* [39] for the medium-range GloFAS forecasts, could improve the skill of the seasonal forecasts. As GloFAS-Seasonal is further developed, it will also be important to consider a wider range of skill metrics for verification, taking into account both the skill and the value of the forecasting system [40]. The evaluation technique used in this study presents a significant challenge for both forecasting systems, requiring that they predict high or low weekly-averaged river flow, in the same week as that in which it was observed, several weeks to months ahead.

Prediction of El Niño events is also key for both types of forecast. As a dynamical model, GloFAS-Seasonal incorporates forecasts of SSTs and therefore ENSO. Decision-makers often rely on forecasts of El Niño before consulting forecasts such as the HistProbs, when an El Niño event is forecast or developing. ECMWF's seasonal forecasts of ENSO events are world-leading [19, 28], and SEAS5 represents an improvement in the skill of these forecasts over the previous version of the forecasting system, S4. However, there is a decrease in the skill of the IOD in SEAS5, with forecasts producing cold events that are too large and too frequent, alongside a slight deterioration in the skill of upper level winds [19], which are important for representing teleconnections across the globe. While dynamical models are better able to represent the complex interactions between the various modes of climate variability and their associated teleconnections by design, it is still possible that the evolution of El Niño may be uncertain or incorrectly predicted, or that even a perfect forecast of El Niño evolution may poorly simulate the teleconnections due to the nonlinearity of the teleconnections and their impacts. This can have important implications for seasonal predictability of ENSO teleconnections using GCMs [41].

A further point of consideration is that while this study makes use of $>$1200 river flow observation stations around the globe, there are large areas of the world, including some that are significantly impacted by El Niño, where there is very sparse to no data coverage. At many of the stations used, management of water resources will be evident in the river flow records, particularly during periods of low flow conditions, and this is likely to affect the evaluation results.

Statistical forecasts such as the HistProbs are limited in that they can only forecast the response to events which we have previously observed. With recent research suggesting that the frequency of extreme El Niño events, such as those in 1982–83, 1997–98 and 2015–16, is likely to increase with future climate change [42, 43], this limitation could become more and more relevant. The HistProbs were also estimated using the longer ERA-20CM-R dataset. This dataset provides more El Niños over which to calculate the probabilities, and has been shown to represent ENSO teleconnections, but is unable to reproduce synoptic situations as no atmospheric observations were assimilated [44]. Future work should explore whether the skill of statistical forecasts such as the HistProbs could be improved using different reanalysis products, such as ERA5.

While currently there are areas of the globe where GloFAS-Seasonal is less skilful than climatology, this is the just the first version of the first global scale operational seasonal river flow forecasting system. Future improvements to the input datasets (e.g. topography, river flow observations, lakes and reservoirs), seasonal precipitation forecasts and hydrological models could result in a dynamical forecasting system that consistently provides a more useful forecast of hydrological extremes, with the benefit that such dynamical forecasts are not constrained to periods of time when there is an El Niño. A third approach, not considered in this study, could be to combine statistical and dynamical forecasts to produce a hybrid system; recent studies suggest this approach could enhance prediction skill at seasonal timescales [45, 46]. Research shows that seasonal hydrological forecasts are able to inform local decisions and actions, and that while uncertainty is not necessarily a barrier to the use of such forecasts, a range of information, including forecast skill, different forecast types and local knowledge are important, alongside a need for higher resolutions to aid local decision-making [47].

5. Conclusions

This paper has evaluated the ability of two different seasonal forecasting approaches, statistical historical probabilities and the dynamical GloFAS-Seasonal, to predict both high and low river flow during El Niño, with the correct timing. Previous research has highlighted the importance of considering the hydrology in addition to meteorological variables, with precipitation often used by decision-makers as a proxy for river flow. These recently-developed forecasts, both of which are used for El Niño preparedness activities, aim to provide hydrologically relevant predictions of hydrological extremes.

While the results presented indicate that the skill of both forecasts varies by location, season and lead time, and it is important to remember that both approaches have uncertainties associated with them and regions where they lack skill, we are able to draw the following conclusions, to answer the question: what is the most useful approach for forecasting hydrological extremes during El Niño?

1. In regions that are strongly influenced by central Pacific El Niños, and in those where GloFAS-Seasonal forecasts currently lack skill, Historical Probabilities generally provide a more useful forecast.
2. In regions where river flow is also influenced by other teleconnections, GloFAS-Seasonal forecasts are typically more useful, as they are better able to account for the characteristics of each El Niño, including the location, timing and magnitude of the SST anomalies, and simulate the response to other modes of climate variability coinciding with El Niño. For example, the phase of the PDO, IOD, NAO, can act to strengthen, mitigate or even reverse the river flow response to El Niño at a regional scale.
3. At lead times of a season ahead, dynamical seasonal forecasts, such as the GloFAS-Seasonal river flow forecasts and seasonal precipitation forecasts, are better able to account for the interaction between various modes of climate variability. Historical Probabilities are, however, available at even earlier lead times, when an El Niño is first forecast or begins to develop.

We further emphasise that while there is often significant interest in the impacts of El Niño due to its global teleconnections, in some regions, it is important to consider that other modes of climate variability can play a key role in addition to ENSO, or may be able to provide added predictability over the use of ENSO as a predictor of hydrological extremes. As more global scale seasonal hydro-meteorological forecasting systems are developed and forecasts are improved, it will be important to revisit the question of which approach is more useful for forecasting hydrological extremes. To forecast high and low river flow on seasonal timescales, and with the correct timing, is a challenging endeavour. That either or both of these forecasts has some ability to predict these events, several weeks to months in advance, provides optimism for the future of seasonal hydro-meteorological forecasting and its use in decision-making across many water-related sectors.

Acknowledgments

The authors acknowledge financial support from the Natural Environment Research Council (NERC) as part of the SCENARIO Doctoral Training Partnership (grant NE/L002566/1), from NERC and the Department for International Development (DFID) under the Science for Humanitarian Emergencies and Resilience (SHEAR) research programme (project FATHUM (Forecasts for AnTicipatory HUManitarian action), grant number NE/P000525/1), and from the Horizon 2020 IMPREX project (grant number 641811). We would like to thank the Environmental Forecasts team at ECMWF for their support, and in particular David Lavers, Christel Prudhomme and Florian Pappenberger for their advice. We are also grateful for the provision of data by ECMWF, the Global Runoff Data Centre (GRDC; [25]), and the 13 other partners and organisations who have made river flow observations available to GloFAS [15].

ORCID iDs

Rebecca E Emerton https://orcid.org/0000-0003-0707-3993

Elisabeth M Stephens https://orcid.org/0000-0002-5439-7563
Hannah L Cloke https://orcid.org/0000-0002-1472-868X

References

[1] Alfieri L *et al* 2013 GloFAS—global ensemble streamflow forecasting and flood early warning *Hydrol. Earth Syst. Sci.* **17** 1161–75
[2] Alfieri L, Salamon P, Pappenberger F, Wetterhall F and Thielen J 2012 Operational early warning systems for water-related hazards in Europe *Environ. Sci. Policy.* **21** 35–49
[3] Bierkens M F P 2015 Global hydrology 2015: state, trends, and directions *Water Resour. Res.* **51** 4923–47
[4] Brown A *et al* 2012 Unified modeling and prediction of weather and climate: a 25-year journey *Bull. Am. Meteorol. Soc.* **93** 1865–77
[5] Emerton R E *et al* 2016 Continental and global scale flood forecasting systems *Wiley Interdiscip. Rev. Water.* **3** 391–418
[6] Stephens E, Day J J, Pappenberger F and Cloke H 2015 Precipitation and floodiness *Geophys. Res. Lett.* **42**
[7] Coughlan De Perez E *et al* 2017 Should seasonal rainfall forecasts be used for flood preparedness? *Earth Syst. Sci.* **21** 4517–24
[8] McPhaden M J, Zebiak S E and Glantz M H 2006 ENSO as an integrating concept in earth science *Science* **314** 1740–5
[9] Chiew F H S and McMahon T A 2002 Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability *Hydrol. Sci. J.* **47** 505–22
[10] Ward P J, Jongman B, Kummu M, Dettinger M D, Sperna Weiland F C and Winsemius H C 2014 Strong influence of El Niño Southern Oscillation on flood risk around the world *Proc. Natl. Acad. Sci. USA* **111** 15659–64
[11] Ward P J, Eisner S, Flörke M, Dettinger M D and Kummu M 2014 Annual flood sensitivities to El Niño–Southern Oscillation at the global scale *Hydrol. Earth Syst. Sci.* **18** 47–66
[12] Ward P J, Kummu M and Lall U 2016 Flood frequencies and durations and their response to El Niño Southern Oscillation: Global analysis *J. Hydrol.* **539** 358–78
[13] IRI International Research Institute for Climate and Society - Climate and Society Map Room. Available from http://iridl.ldeo.columbia.edu/maproom/ENSO/Impacts.html
[14] Emerton R, Cloke H L, Stephens E M, Zsoter E, Woolnough S J and Pappenberger F 2017 Complex picture for likelihood of ENSO-driven flood hazard *Nat. Commun.* **8** 14796
[15] Emerton R *et al* 2018 Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0 *Geosci. Model Dev.* **11** 3327–46
[16] Stockdale T, Johnson S, Ferranti L, Balmaseda M and Briceag S 2018 ECMWF's new long-range forecasting system SEAS5 *ECMWF Newsl.* **154** 15–20
[17] ECMWF 2017 SEAS5 user guide. Available from https://ecmwf.int/sites/default/files/medialibrary/2017-10/System5_guide.pdf [accessed 26/2/2019]
[18] Van Der Knijff J M, Younis J and De Roo A P J 2010 LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation *Int. J. Geogr. Inf. Sci.* **24** 189–212
[19] ECMWF 2018 SEAS5 and the future evolution of the long-range forecast system *ECMWF Sci. Advis. Comm. 47th Sess.* **47** 1-81
[20] Hersbach H and Dee D 2016 ERA5 reanalysis is in production *ECMWF Newsl.* **147** 7
[21] Bradley R S, Diaz H F, Kiladis G N and Eischeid J K 1987 ENSO signal in continental temperature and precipitation records *Nature* **327** 497–501
[22] Mason S J and Goddard L 2001 Probabilistic precipitation anomalies associated with ENSO *Bull. Am. Meteorol. Soc.* **82** 619–38
[23] Arnal L *et al* 2018 Skilful seasonal forecasts of streamflow over Europe? *Earth Syst. Sci.* **225194** 2057–72
[24] Mason S J and Graham N E 1999 Conditional probabilities, relative operating characteristics, and relative operating levels *Weather Forecast.* **14** 713–25
[25] GRDC The Global Runoff Data Centre, 56068 Koblenz, Germany.
[26] McPhaden M J 2003 Tropical Pacific Ocean heat content variations and ENSO persistence barriers *Geophys. Res. Lett.* **30** 1480
[27] Duan W and Wei C 2013 The 'spring predictability barrier' for ENSO predictions and its possible mechanism: results from a fully coupled model *Int. J. Climatol.* **33** 1280–92
[28] Barnston A G *et al* 2012 Skill of Real-Time Seasonal ENSO Model Predictions during 2002–11: Is Our Capability Increasing ? *Bull. Am. Meteorol. Soc.* **93** 631–51
[29] Wang-Chun Lai A, Herzog M, Graf H-F, Lai A W-C, Herzog M and Graf H-F 2018 ENSO Forecasts near the spring predictability barrier and possible reasons for the recently reduced predictability *J. Clim.* **31** 815–38
[30] Wang S, Huang J, He Y and Guan Y 2015 Combined effects of the pacific decadal oscillation and El Niño-southern oscillation on global land dry–wet changes *Sci. Rep.* **4** 6651
[31] Hoell A, Gaughan A E, Shukla S and Magadzire T 2017 The hydrologic effects of synchronous El Niño–southern oscillation and subtropical indian ocean dipole events over Southern Africa *J. Hydrometeorol.* **18** 2407–24
[32] Marchant R, Mumbi C, Behera S and Yamagata T 2007 The Indian Ocean dipole ? the unsung driver of climatic variability in East Africa *Afr. J. Ecol.* **45** 4–16
[33] Behera S K *et al* 2005 Paramount Impact of the Indian Ocean dipole on the East African short rains: a CGCM study *J. Clim.* **18** 4514–30
[34] Washington R and Preston A 2006 Extreme wet years over southern Africa: role of Indian Ocean sea surface temperatures *J. Geophys. Res.* **111** D15104
[35] Saji N H and Yamagata T 2003 Possible impacts of Indian Ocean dipole mode events on global climate *Clim. Res.* **25** 151–69
[36] MacLeod D 2018 Seasonal predictability of onset and cessation of the east African rains *Weather Clim. Extrem.* **21** 27–35
[37] Ashok K, Guan Z and Yamagata T 2003 Influence of the Indian Ocean dipole on the Australian winter rainfall *Geophys. Res. Lett.* **30**
[38] Nobre G G, Jongman B, Aerts J and Ward P J 2017 The role of climate variability in extreme floods in Europe *Environ. Res. Lett.* **12** 084012
[39] Hirpa F A *et al* 2018 Calibration of the global flood awareness system (GloFAS) using daily streamflow data *J. Hydrol.* **231** Accepted
[40] Cloke H L, Pappenberger F, Smith P J and Wetterhall F 2017 How do I know if I've improved my continental scale flood early warning system ? *Environ. Res. Lett.* **12** 044006
[41] Turner A, Inness P M and Slingo J M 2005 The role of the basic state in the ENSO-monsoon relationship and implications for predictability *Q. J. R. Meteorol. Soc.* **131** 781–804
[42] Cai W *et al* 2014 Increasing frequency of extreme El Niño events due to greenhouse warming *Nat. Clim. Chang.* **4** 111–6
[43] Cai W *et al* 2015 ENSO and greenhouse warming *Nat. Clim. Chang.* **5** 849–59

[44] Hersbach H, Peubey C, Simmons A, Berrisford P, Poli P and Dee D 2015 ERA-20CM: a twentieth-century atmospheric model ensemble *Q. J. R. Meteorol. Soc.* **141** 2350–75
[45] Slater L J and Villarini G 2018 Enhancing the predictability of seasonal streamflow with a statistical-dynamical approach *Geophys. Res. Lett.* **45** 6504–13
[46] Schepen A, Wang Q J and Robertson D E 2012 Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall *J. Geophys. Res. Atmos.* **117**
[47] Neumann J L, Arnal L, Emerton R, Griffith H, Hyslop S, Theofanidi S and Cloke H L 2018 Can seasonal hydrological forecasts inform local decisions and actions ? a decision-making activity *Geoscience Communication* **1** 35–57