Interactive comment on “A national-scale seasonal hydrological forecast system: development and evaluation over Britain” by Victoria A. Bell et al.

Response to comments from the Editor and reviewers #1 and #2

Authors’ responses in red.

Response to Anonymous Referee #1
Received and published: 22 May 2017

The manuscript illustrates the development and the evaluation of a national-scale seasonal hydrological forecast system over Britain. A high resolution hydrological model is used to estimate the initial conditions of a simple water balance model used to forecast regional seasonal flows over Britain. As meteorological forcings seasonal re-forecasts from GloSea5 and historical rainfall are used. As additional reference for verification flow persistence is used.

General:
The paper is well written, the methodology and results are nicely presented and compared. The real value of this study is the combination of a high resolution hydrological model to estimate the initial conditions and a simple water balance model which fits better to the meteorological seasonal forecast input data than a high resolution model. The regional WBM performance using observed data and the forecasting skill show promising results. The paper should be foreseen for publication in HESS after minor revisions.

Thank you for the positive comments.

Comments:
The objective of the seasonal forecasting system is not really clear. From the introduction I assume the focus on flood forecasting, from the forecast variables, regional mean flow 1-month / 3-month ahead, I assume the focus on drought forecasting. Please specify! If the focus is flood forecasting, the relationship between regional mean flow and relevant flood properties (peak, volume,...) should be shown to demonstrate the potential flood predictability of the system.

The Abstract, Introduction and Summary will be amended to make this clearer. The forecasts issued provide seasonal mean river flows over the next 1 and 3 months (not flood peaks). The objective is to provide a wide audience, ranging from water managers to members of the public, with insight into whether river flows might be expected to be above, below, or in the normal range over the next few months based on the best information available.

p 7 l 15: What is UK mean SAAR? Please add explanation / reference.

SAAR is Standard-period Average Annual Rainfall (mm) (e.g. http://nrfa.ceh.ac.uk/rainfall-statistics). However, we will amend the manuscript to say “average annual rainfall”, instead of “SAAR”. We created annual mean rainfall for a 49-year period that is longer than a standard period of 30 years, so we should not have used the acronym “SAAR” in the first place.

p 12 l 6: do you use ensemble mean of GloSea5 / historical rainfall 1-month / 3-month ahead as input for WBM or do you use the ensemble mean of the WBM seasonal flow forecasts?

We intended to say that the ensemble mean GloSea5 rainfall forecast was used as input to the WBM. The manuscript will be amended to say "the skill of ensemble mean GloSea5 rainfall forecasts has been evaluated for all seasons…".

What is the difference in skill between the two ensemble mean forecasts (rainfall ensemble mean as input vs. flow ensemble mean)?

We have not evaluated the skill in the ensemble mean flow forecast, just the skill in the flow forecast using an ensemble mean rainfall forecast. I would expect these two variants to be very similar, if not identical, as the equations are all linear (i.e. construction of the mean and the WBM are linear). It is also quicker to make one flow forecast from an ensemble mean rainfall forecast, rather than average an ensemble of flow forecasts.

Response to Anonymous Referee #2
Received and published: 2 June 2017

This manuscript explores the potential for 1 month and 3 month hydrologic forecasts on a national scale, for different seasons. As a precursor, the manuscript describes how high resolution spatial information from a hydrologic model can be used to estimate initial conditions in a simplified manner.

General: Overall, the paper is well written and the structure provides for a coherent progression through the various sections. The main takeaway points, relative to the stated intention of development an evaluation of the system, are clear in that the seasonal differences in forecast skill are prominently noted and discussed. However, as the manuscript as identifies the desire to link to decision making, I would suggest clarifying that link, maybe providing a section, as slight mentions of this link are unconvincing. I recommend the manuscript for publication in HESS after addressing minor revisions.

Thank you.
With the aim of clarifying links to decision-making, extra sentences will be added to the manuscript (in the conclusions) to say:
"*The HOUK has been in operation for 4 years (publically available from autumn 2013) and thus is a relatively new product. At present, automated web statistics indicate approximately 300 readers or users of the HOUK website per month (Prudhomme et al., under review). Exactly how water managers use the HOUK in practice has not yet been assessed, but ongoing evaluations of the skill in the different methods used in the construction of the Outlook will undoubtedly help provide the evidence required to support use of the product in decision-making."*

Specific comments:
Page 2 Line 5-7 What is meant by perceived lack of skill? Is there skill or not? I would clarify as the lack of skill could discourage development of forecasts or it can also be the case that there is skill, yet amongst key people making decisions of research priorities, it could be perceived to not have skill. The second condition is more complicated and likely will need to be resolved with social science.

This sentence will be re-worded to emphasise that published literature as recently as 2011 indicated that seasonal forecasts of rainfall in extratropical areas (such as Britain) had little skill at lead times of more than 1-month. Thus there was little incentive to use the forecasts to support decision-making.

Line 26 A general note overall, and referencing the potential value for practitioners in line 29 of the abstract, it is important not to state skill acceptability in an overarching sense. In the literature differences in both perception of skill and more importantly acceptable levels of skill in order to justify using that forecast can vary across sectors and will likely be different for users compared to forecast developers. For more on this see: Hartmann, H. C., Pagano, T. C., Sorooshian, S., & Bales, R. (2002). Confidence builders: Evaluating seasonal climate forecasts from user perspectives. Bulletin of the American Meteorological Society, 83(5), 683-698.

This is a good point. The sentence in the abstract will be clarified to say "*Given the high spatial heterogeneity in typical patterns of UK rainfall and evaporation, future development of skilful spatially distributed seasonal forecasts could lead to substantial improvements in seasonal flow forecast capability, potentially benefitting practitioners interested in predicting hydrological extremes, not only in the UK, but also across Europe*."

Page 3 Line 16 – Correct in stating that downscaling would be necessary, but is there any evidence that downscaling would be a 'worthwhile' activity for improving national estimates of water flows?
The sentence will be re-worded to make it clear that we aim to provide flow estimates nation-wide, rather than at a national scale:
"… seasonal rainfall forecasts do not provide detailed weather information at this resolution and would typically require spatio-temporal downscaling to achieve good estimates of river flow for catchments or regions nation-wide".

Line 17 – What is meant by realistic water flows? I suggest adding context or changing the word as, in the current form, this is at unnecessarily high risk for misinterpretation.

We have changed the word "realistic" to "good". Please see re-worded sentence above.

Page 4 Line 19 – I suggest clarifying the period in which the long term average is calculated over. I think it can be interpreted as monthly or seasonally, which could impact the result.

The sentence will be clarified to say "long term monthly average".

Page 7 Line 14 Does 'relative to the rainfall forecast climatological mean' imply the 3 month anomaly will be distributed over each month based on a month's relative mean contribution to the total seasonal precipitation?

Yes. The sentence can indeed be simplified to say *"Disaggregation of the 3-month ahead forecast into monthly rainfall amounts is achieved through distributing the 3-month rainfall forecast anomaly between the 3 individual months according to their relative contribution to the UK mean seasonal rainfall (1962 to 2010)."*

Page 10 Line 1 The Forth region has median model performance for 1 month or 3 month leads? Or both combined?

This is for the 1-month lead time and the text has been clarified.

Line 14 Seems that a comma is in error or a capital T is in error

Ah, thank you. The offending comma will be changed to a full stop.

Line 16 2a shows persistence on a 1 month lead more skillful than GloSea5+HIC. Can you explain why? Or I suggest noting that for overall assessments on a 1 month lead, persistence forecast should be explored in more depth.

The sentences will be clarified as follows: "*At the 1-month lead time the WBM with G2G HIC driven by an historical rainfall (climatology) ensemble performs best, and the forecasts based on persistence or Glosea5+HIC perform less well, but show some skill. For the longer 3-month lead time, the WBM with G2G HIC driven by either historical or GloSea5 rainfall perform similarly; persistence forecasts (Pers) or use of an average HIC are not recommended at this lead time.*"

The explanation for the skill in persistence forecasts is addressed in the regional skill breakdown in the next but one paragraph in the manuscript.

Page 12 Line 15 Agreed, yet this point (how inclusion of avHIC with GloSea5 lead to 0 skill during some seasons) presents another that may be worth addressing – In figure 3, there exists an interesting pattern of skill at the sub-national level. Any thoughts on why?

This probably refers to p14 Line 15. Yes, the apparent skill in use of an average HIC with GloSea5 in Figures 3 and 4 is interesting. Figure 4 indicates that these forecasts are only skilful in Autumn and Winter, and particularly at the 3-month ahead lead time. The regions where the skill is greatest (Northern and Western regions, excluding Northern Scotland) are areas with less subsurface storage than the SouthEast regions and Northern Scotland, and thus where persistence forecasts are less skilful. We would speculate that in these areas the skill arises from the improved skill in GloSea5 at the 3-month lead time, coupled with less dependence on a good HIC in these areas.

The sentence in the manuscript will be changed to say:
"…Autumn/Winter flow forecasts using ensemble mean GloSea5 rainfall and an average HIC perform surprisingly well across Britain, confirming that there is a significant element of skill associated with GloSea5 forecasts in Autumn/Winter at the 3-month lead time, often resulting in skilful flow forecasts in regions where this skill is less dependent on a good HIC."

Page 15 Figure 5 Is it possible to include a map that labels the regions (abbreviations should suffice)? I think the paper as a whole would benefit from this, even if one of the maps used in a previous figure could do this.

A map labelling the regions is already included in Figure 1b. It is rather small, and we will make it larger in the revised manuscript. The Figure 5 caption will be changed to reference this map.

Line 9 It would be useful to note which regions are included in the 'Northern' and 'Western' cluster.

The text will be amended to list the regions (Northumbria, Yorkshire, South-West, Welsh, North-West, Solway, Clyde, Forth, Tweed)

Page 17 Line 7 The presentation of this information to any user does not inherently alert them to anything. If they perceive the information (both the most likely possibility and the full range) to be trustworthy and if they are able to justify making (or changing a from normal protocol) a decision based on that information, then I can justify the use of alert in this context. I would suggest, in an over simplistic manner, what the presentation of the full range of scenarios does – presents the full range and the mean. This could be useful and interesting (and then, maybe advantageous) for a user but only if it is perceived to be relevant by them. Also this sentence uses the phrase 'presentation of the full range of scenarios' twice, so I would suggest re wording even if the above suggestion on content is not followed

This is useful feedback. The text will be modified to say *"Continued presentation of this full range of flow scenarios for the coming 1- and 3-months may be advantageous in that it informs water managers, not only of the most likely possibility, but also to the range of possibilities."*

Page 17 Line 14-16
Although the section notes recommendations will be included, I do not find any except for a weak statement regarding how an increase in spatial resolution could lead to improvement. This is not a new finding and I would consider reflecting on the content of the paper to develop recommendations that are more relevant.

This is a good point. An additional short paragraph will be added to provide recommendations:
"*Based on the skill analysis presented here, users of the Hydrological Outlook UK would be advised to have greatest confidence in Autumn and Winter flow forecasts that use GloSea5 rainfall, particularly at the 3-month lead time. For Spring/Summer flow forecasts, use of an ensemble forecast based on historical rainfall is surprisingly good and would be recommended for use across Scotland, and flow forecasts based on persistence were found to be the most skilful in South-East regions (Thames, Anglia, Wessex and Southern)."*

It may be interesting to explore the role of ENSO. Referencing 4.3.5 of van Oldenborgh et al 2005, there is potential for some skill from ENSO for parts of the UK, including Scotland. Noting the skill in Scotland (figure 3), it could be a useful exercise to disaggregate by ENSO phase both in the target month and in the month of forecast issuance.
Jan van Oldenborgh, G., Balmaseda, M. A., Ferranti, L., Stockdale, T. N., & Anderson, D. L. (2005). Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years? Journal of climate, 18(16), 3240-3249. Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-154, 2017.

This is an interesting link to make, thank you for the reference. This approach may well be worth exploring in another piece of work, but is beyond the scope of the current manuscript.

Editor Decision: Publish subject to minor revisions (further review by Editor) (07 Jul 2017) by Andy Wood
Comments to the Author:
In general, the authors' proposed revisions appear adequate to address the concerns and comments raised by the reviewers, and I encourage the authors to adjust the manuscript along the lines proposed and resubmit for my further review.

Authors response: Thank you, these revisions are now in hand.

In addition, I wish to add the following comments for the authors to consider in their revision.

Major comment:
I'm struck by the relative arbitrariness of the approach in which HICs are drawn from a high-res land surface model to initialize seasonal ensemble forecasts that are propagated forward with a conceptual scheme at the opposite end of the spectrum in terms of complexity (ie the WBM). Are there examples of this type of approach being used elsewhere that could be referenced?

Authors response: Other studies referenced in the manuscript (p2, Introduction) highlight the importance of the hydrological initial condition (HIC) to seasonal hydrological forecast skill, and our aim with the WBM formulation was to combine the high resolution, most recent, HIC available from the G2G with a simple monthly rainfall forecast from GloSea5. It is not only HICs that are extracted from the high-resolution G2G hydrological model, other spatial grids derived from G2G output (such as Smax, Smin, S_bar, q_mean) are used by the WBM. We are unaware of other studies that have used a scheme of the type presented in this manuscript, and would welcome hearing of any that we have missed.

The given rationale for the strategy is that there's no appreciable value to downscaling climate forecasts to the scale of the G2G model, yet there have been numerous examples of seasonal forecasting in which coarse climate forecasts are in fact downscaled to resolutions on the order of 10K or finer, and used to drive flow forecasts.

Authors response: To develop a seasonal hydrological forecasting approach for semi-operational use, a method was required which would make the best use of the available coarse resolution GloSea5 rainfall forecasts alongside routinely available hydrological information. Early modelling trials (unpublished) explored using a 30-year ensemble of historical rainfall observations to produce a spatially-variable, daily ensemble of 1140-member G2G-derived rainfall forecasts downscaled from national-scale to 1km resolution. Although the approach showed some promise, detecting and interpreting any forecast signal in the resulting forecast “noise” was not straightforward, and the possibility of developing a simpler, more elegant approach that could be explained clearly to a wide audience was explored in preference (this manuscript). In place of dynamic downscaling to 1km we have downscaled in space only, using climatological mean UK monthly rainfall patterns on a 1km grid to downscale GloSea5 (a similar approach was used by Bell et al., (2009) to downscale 25km RCM rainfall data to a 1km resolution grid). We have clarified the description of the spatial downscaling method applied in Section 2.3: “*This approach to spatial-downscaling using historical mean rainfall observations is similar to one used by Bell et al. (2009) to downscale 25km-resolution regional climate model data to a 1km resolution.”*

Is another practical rationale that the computational cost of running the G2G model ensembles with downscaled forecasts is not justified by any added skill from to computing a full-resolution seasonal flow forecast after downscaling?

Authors response: The added computational cost of combining multiple downscaling scenarios (in time and space) with a 38-member Glosea5 forecast ensemble is an important consideration, but was not the main one.

In any case, I think this somewhat bespoke/uncommon approach to seasonal regional forecast construction is an interesting choice that warrants a bit more motivation.

Authors response: We will amend the manuscript, particularly the motivation for the development of the WBM method (p3, lines 15-27) as follows: *“While rainfall downscaling is relatively straightforward for a particular location or catchment, using national-scale monthly rainfall forecasts to produce pixel-scale daily rainfall would require an ensemble downscaling approach based on either a weather generator or historical analogues, generating large multiples of ensemble flow forecasts. This approach has been explored in other studies (e.g. Manzanas et al. 2017, Charles et al. 2012), which showed that dynamic or statistical downscaling of seasonal forecasts can reduce local biases in variables such as temperature or rainfall, but do not necessarily improve the overall forecast skill.”.*

And is there is any way to benchmark it against any alternatives that would provide practical insights? For instance, for regional prediction of flow anomalies, is there added value from taking HICs from G2G versus from a long-term run of a simple conceptual monthly model -- could that be tested? And is there any way to assess the marginal value of this approach, versus other alternatives (such as the far more common use of one model for both HICs *and* forecast, such as at SMHI or in the US NWS), against actual obs flows -- perhaps for a handful of case study basins where such flow estimates are available? I realize that this may all lie beyond the scope of this paper. In any case a broader discussion of the rationale for this particular prediction scheme, and acknowledgement of other alternatives, could add value to the paper.

Authors response: The question about benchmarking the G2G/WBM method against alternative sources of HIC is a relevant one. We have started looking into this already, but this is ongoing research and out of scope for the current manuscript.

I find the paper's attribution of forecast skill using the component-on/component-off approach very useful and effective, as a technique that is quite common in seasonal climate prediction (ie, use a climatological Indian Ocean or continental SM, versus a dynamic one), but less common in hydrologic prediction studies. It provides a slightly different angle amidst the recent interest in the attribution of seasonal hydrologic prediction skill.

Minor comments:
p8, l8 -- 'Clim' doesn't match the 'Hist' term later used in Fig 2 and elsewhere.

Authors response: Now corrected, we meant to say “Hist”. Thank you for spotting that.

p6, l10 – the text states: “The coarse spatial resolution of the input rainfall forecasts has discouraged the development of river flow forecasts at a 1km resolution, and production of regional scale forecasts (in preference to national scale) is viewed as a pragmatic compromise. “ Is that the sole reason or is the computational cost of running a 1 km resolution model for seasonal ensembles also a limiting factor?

Authors response: Computational cost is a factor discouraging but not preventing use of the 1km resolution G2G with hundreds of ensemble members (though the linux-cluster would allow for multi-ensemble model runs if necessary). The seasonal flow forecasts from the monthly WBM are available at a 1km resolution (Section 2,2 final paragraph) before regional averages are computed, but the 1km WBM forecasts have not been assessed and are not used in practice. The methods developed here are part of a semi-operational “Hydrological Outlook” providing UK-wide seasonal forecasts of river flows to operational agencies and the general public. Using a national-scale mean rainfall product to provide a semi-operational flow forecast at a 1km resolution, when the rainfall information is simply not available at such a high resolution, seemed misleading – the provision of a regional forecast seemed more appropriate. The manuscript text will be clarified as follows: “*The coarse spatial resolution of the input rainfall forecasts has discouraged the development of river flow forecasts at a 1km resolution to ensure that users of the Hydrological Outlook do not infer that rainfall forecasts are available or skilful at this resolution. Production of regional scale forecasts (in preference to national scale) is viewed as a pragmatic compromise*”.

p6, l5 – when/why does it arise that the Qm_bar/Sm_bar >= 1? More explanation or insight would be helpful – especially as these are climatological averages. The monthly runoff ratio should almost always be between 0 and 1 unless a snowpack-related delay in runoff timing is involved.

Authors response: This is a good question, and as a result of this we have amended the preceding explanation and notation to explain the model more clearly, and to make it clear that qm_bar is G2G-derived river flow per unit catchment area, and not runoff. Across Britain qm_bar <= Sm_bar in most places, but for a few highly spatially variable catchments, and in wetter months, there will be some situations where qm_bar > Sm_bar. A note has been made in the text to say this could occur in areas of high spatial variability.

A national-scale seasonal hydrological forecast system: development and evaluation over Britain

Victoria A. Bell[1], Helen N. Davies[1], Alison L. Kay[1], Anca Brookshaw[2,3] and Adam A. Scaife[3,4]

[1] Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK
[2] ECMWF, Shinfield Park, Reading, RG2 9AX, UK
[3] Met Office Hadley Centre, FitzRoy Road, Exeter, Devon, EX1 3PB, UK
[4] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom, EX4 4QF

Correspondence to: V.A.Bell (vib@ceh.ac.uk)

Abstract. Skilful winter seasonal predictions for the North Atlantic circulation and Northern Europe have now been demonstrated and the potential for seasonal hydrological forecasting in the UK is now being explored. One of the techniques being used combines seasonal rainfall forecasts provided by operational weather forecast systems with hydrological modelling tools to provide estimates of seasonal mean river flows up to a few months ahead.

The work presented here shows how spatial information contained in a distributed hydrological model typically requiring high resolution (daily or better) rainfall data can be used to provide an initial condition for a much simpler forecast model tailored to use low-resolution monthly rainfall forecasts. Rainfall forecasts ("hindcasts") from the GloSea5 model (1996 to 2009) are used to provide the first assessment of skill in these national-scale flow forecasts. The skill in the combined modelling system is assessed for different seasons and regions of Britain, and compared to what might be achieved using other approaches such as use of an ensemble of historical rainfall in a hydrological model, or a simple flow persistence forecast. The analysis indicates that only limited forecast skill is achievable for Spring/Summer seasonal hydrological forecasts, however, Autumn/Winter flows can be reasonably well forecast using (ensemble mean) rainfall forecasts based on either GloSea5 forecasts or historical rainfall (the preferred type of forecast depends on the region). Flow forecasts using ensemble mean GloSea5 rainfall perform the most consistently well across Britain, and provide the most skilful forecasts overall at the 3-month lead time. Much of the skill (64%) in the 1-month ahead seasonal flow forecasts can be attributed to the hydrological initial condition (particularly in regions with a significant groundwater contribution to flows), whereas for the 3-month ahead lead time, GloSea5 forecasts account for ~70% of the forecast skill (mostly in areas of high rainfall to the North and West) and only 30% of the skill arises from hydrological memory (typically groundwater-dominated areas). Given the high spatial heterogeneity in typical patterns of UK rainfall and evaporation, future development of skilful *spatially distributed* seasonal forecasts could lead to substantial improvements in seasonal flow forecast capability, potentially benefitting practitioners interested in predicting hydrological extremes, not only in the UK, but also potentially across Europe.

1 Introduction

A series of low-pressure systems crossing Britain in Winter 2015/16 resulted in some of the most widespread and severe flooding witnessed in the UK, with several rivers in the north of Britain recording their highest ever flows and thousands of properties flooded (Centre for Ecology & Hydrology 2016). Repairs to damaged homes, businesses and flood defences were required, and procedures for forecasting and mitigating the floods are understandably being examined. Until relatively recently, a perceived lack of skill in seasonal weather forecasts in extratropical regions beyond a lead time of 1 month (Lavers et al. 2009, *Arribas et al*. 2011) has until recently discouraged the development of routine seasonal hydrological forecasts using climate model output in Britain. However, the potential for seasonal hydrological forecasting in the UK is now being explored. Various seasonal forecast systems now provide skilful forecasts out to a few months ahead (e.g. MacLachlan et al. 2015, Athanasiadis et al. 2014), allowing for some form of skilful dynamical hydrological forecast. As well as using climate model output, others are investigating statistical relationships between large-scale North Atlantic climate indices (such as the North Atlantic Oscillation) and seasonal rainfall or river flow anomalies (Lavers et al. 2010a, 2010b; Macgregor and Phillips 2004; Svensson and Prudhomme 2005; Wedgbrow et al. 2002; Wilby 2001, Svensson et al. 2015), and these can provide increased skill when large scale patterns dominate regional rainfall (Scaife et al. 2014).

A recent review of seasonal hydrological forecasting methods using climate model output by Yuan et al. (2015) highlighted the dependence of predictive skill on both the large scale climate drivers and the local hydrological initial condition (HIC), which for some regions can persist for several months. The relative importance of initial conditions and boundary forcing (the meteorological forecast) on the skill of seasonal hydrological prediction has been examined by a number of authors, for example a study of skill in forecasting mean seasonal river flows across Europe concluded that much of the skill could be attributed to correct hydrological initial conditions, rather than the weather forecast (Bierkens and van Beek, 2009). In a UK-based study, also using seasonal forecasts, Svensson et al. (2015) identified a geographical complementarity in regional seasonal hydrological predictability, noting that predictability in river flows in southern and eastern Britain derived primarily from hydrological memory of antecedent conditions, and from meteorological predictability (predictions of the atmospheric circulation over the North Atlantic at the seasonal timescale) in northern and western areas. They were able to generate skilful hydrological forecasts for river flows using the large scale atmospheric circulation which governs much of UK winter (December to February) rainfall, and November initial conditions.

Advances in the performance of operational seasonal forecast systems such as the Met Office GloSea5 system (MacLachlan et al. 2015) are now encouraging the development of hydrological forecasting systems that can make best use of these more skilful seasonal forecasts. In the UK, the recently developed Hydrological Outlook UK (HOUK) provides an insight into future hydrological conditions nationwide. It describes likely trajectories for river flows and groundwater levels on a monthly basis, with particular focus on the next one and three months. A number of techniques are used to project forwards from the current state, and results from these are used to produce a summary including a highlights map. Prudhomme et al (under review) summarises the range of techniques used in the production of the HOUK, which encompass schemes using historical river flow analogues, ensembles of historical sequences of observed climate, and ensembles of seasonal rainfall forecasts. The forecasts issued provide seasonal mean river flows and instantaneous groundwater levels with a forecast horizon of up to 12 months ahead, with an emphasis on the next 1 to 3 months (http://hydoutuk.net/).

The approach using seasonal rainfall forecasts provided by the Met Office model GloSea5 is explored in more detail here. These rainfall forecasts are combined with hydrological modelling tools to provide estimates of hydrological conditions up to a few months ahead. The hydrological modelling follows on from an approach to seasonal forecasting developed by Bell et al. (2013), which used a distributed hydrological model driven by observations to provide the hydrological initial condition, and a monthly 'water-balance anomaly' model to estimate sub-surface water storage over the next 1 to 3 months as perturbations from the initial state, driven by Met Office seasonal rainfall forecasts. Forecasting UK-wide monthly-mean river flow is less straightforward than forecasting subsurface water storage, as river flow is a spatial and temporal integrator of local-scale runoff production, arising from a combination of antecedent storage and the partitioning of effective rainfall between surface /subsurface runoff and storage. This task can arguably be undertaken by a fully configured grid-based hydrological model, maintaining a continuous local water balance and using daily or sub-daily spatial rainfall estimates as input. However, seasonal rainfall forecasts do not provide detailed weather information at this resolution and would typically require spatio-temporal downscaling to achieve ~~realistic~~ good ~~national~~ estimates of river flow for catchments or regions nation-wide. While rainfall downscaling is relatively straightforward for a particular location or catchment, ~~turning~~ using national-scale monthly rainfall forecasts ~~into~~ to produce pixel-scale daily rainfall would require an ensemble downscaling approach based on either a weather generator or historical analogues, ~~thus~~ generating ~~ever-greater numbers~~large multiples of ensemble flow forecasts. This approach has been explored in other studies (e.g. Manzanas et al. 2017, Charles et al. 2012), which showed that dynamic or statistical downscaling of seasonal forecasts can reduce local biases in variables such as temperature or rainfall, but do not necessarily improve the overall forecast skill.

The alternative approach explored here is to place the greatest emphasis on the hydrological initial condition provided by an up-to-date model, while simplifying the generation of hydrological forecasts through use of a temporally-coarse water balance model with less dependence on high resolution weather information. A monthly-resolution forecast model provides additional benefits by reducing the computational overhead of the use of a rainfall forecast ensemble. This scheme is used to provide regional-scale estimates of the river flows over the coming months, and work presented here examines the skill of these forecasts for Britain, for geographical regions, for particular seasons, and at 1- and 3-month lead times.

2 Models and Methods

2.1 Hydrological initial condition

Grid-to-Grid (G2G) is a spatially-distributed hydrological model, which is generally configured to a 1km^2 grid across Britain, with a 15-minute time-step, and underpinned by digital spatial datasets of topography, soil/geology and land cover. A detailed description of G2G is presented in Bell *et al*. (2009), with a brief overview of the model's subsurface (soil and groundwater) storage formulation provided in Bell et al. (2013). Input to the model consists of gridded time-series of precipitation and potential evaporation (PE) derived from observations, numerical weather prediction or regional climate models. Model output can be in the form of area-wide, gridded time-series of river flows, runoff and soil-moisture, or time-series of river flows at gauged or ungauged locations. Applications of the model include both continuous simulation of river flows in a changing climate (Bell *et al*., 2009, 2016) and real-time flood forecasting (Moore *et al.*, 2006; Cole and Moore, 2009)**.**

For production of the HOUK, G2G is run continuously over several years to produce an estimate of the most recent hydrological condition across Britain, from which an estimate is made of the current depth of subsurface water storage. The G2G requires gridded time-series of rainfall and PE. Daily precipitation data on a 5km grid, provided by the Met Office for 1958–present (Perry et al. 2009), were used at the 15-minute G2G time-step by equally spreading them throughout the day, and downscaled to 1km using a spatial weighting based on 1km Standard Average Annual Rainfall data for 1961–1990 (Bell et al. 2007). Monthly PE data on a 40km grid from MORECS (Hough and Jones 1997) were spread equally through the month and applied equally to each 1km box within each 40km square. Here, the depth of sub-surface water storage, S, consists of the sum of the unsaturated soil, V, and the groundwater, V_g, storages. The depth of water in groundwater storage arises from the balance between recharge and groundwater outflow over long periods, and while it is unlikely to correspond directly to a groundwater level observation, it can provide an indication of whether storage in the saturated zone is greater or less than the long-term monthly average.

2.2 Water-balance model for flows

Following Bell et al. (2013) the continuity equation can be used to express change in total subsurface water storage, S, as a balance between input precipitation P and outputs through actual evaporation E and net outflow ~~(surface and sub-surface runoff)~~ per unit area Q, so $\mathrm{d}S/\mathrm{d}t = P - E - Q$, where all quantities are expressed in water depth (mm) over a model grid-cell. In terms of discrete months, if S_m and S_{m+1} represent the storage at the end of months *m* and *m+1*, and if P_{m+1}, E_{m+1} and Q_{m+1} denote mean rainfall, evaporation and ~~runoff~~ net outflow per unit area over the month *m+1*,

$$S_{m+1} \cong S_m + P_{m+1} - E_{m+1} - Q_{m+1} \quad (1)$$

At a monthly time-step it is assumed here that daily/sub-daily changes in rainfall, storage, and ~~river~~ net outflows can be neglected and gross simplifying assumptions can be made as to whether excess effective rainfall is stored in the subsurface or released via runoff from saturated pixels. Storage of water in each pixel is assumed to vary between S_{min} and S_{max}, the historical minimum and maximum G2G-simulated storage of each 1km pixel's sub-surface water store respectively. By combining the current storage S_m as estimated by the G2G at the forecast time origin with monthly seasonal forecasts of P_{m+1} and E_{m+1}, corresponding forecasts of storage~~, runoff~~ and flow can be produced as follows:

1. For a discrete month, *m*, an initial estimator of the storage in each pixel at the end of the following month *(m+1)* can be given by

$$\hat{S}^*_{m+1} \approx S_m + P_f - E_f \quad (2)$$

where P_f and E_f are seasonal forecasts of mean monthly rainfall and actual evaporation, and the * indicates an initial (as opposed to final) estimator.

2. The initial estimator for forecast storage $\hat{S}^*_{m+1}$ (Eq. 2) neglects forecast ~~runoff~~ Q_{m+1} which can be significant, but is less easy to forecast directly than storage as its magnitude will depend on a number of factors including soil properties, storage, effective rainfall and topography. Typically in hydrological models, ~~runoff~~ river flow is estimated through a relationship between incoming effective rainfall and antecedent soil-moisture and sub-surface water storage, and for the seasonal forecasting application considered here, Q_{m+1} ~~runoff~~ is also estimated through a relationship of the form $Q \simeq f(S)$. For the national-scale application required for the HOUK encompassing a wide range of soils, geology and catchment characteristics, a very simple empirical relationship relating grid-cell net outflow in a month to mean monthly river flow is assumed:

$$\hat{Q}_{m+1} \simeq \hat{S}^*_{m+1} \frac{\bar{q}_{m+1}}{\bar{S}_{m+1}} . \quad (3)$$

Here, mean monthly ~~runoff~~ net outflow in month *m+1* is estimated in terms of the forecast storage $\hat{S}^*_{m+1}$ scaled by the ratio between G2G model-derived estimates of long-term mean river flow per unit catchment area, $\bar{q}_{m+1}$, and storage $\bar{S}_{m+1}$ for the month (1962 to 2010).

3. Replacing the unknown Q_{m+1} in Eq. (1) with its estimator $\hat{Q}_{m+1}$ yields an improved estimate of S_{m+1}:

$$\hat{S}_{m+1} \approx (S_m + P_f - E_f)\left(1 - \frac{\bar{q}_{m+1}}{\bar{S}_{m+1}}\right) \approx \begin{cases} \hat{S}^*_{m+1}\left(1 - \frac{\bar{q}_{m+1}}{\bar{S}_{m+1}}\right), & \text{for } \frac{\bar{q}_{m+1}}{\bar{S}_{m+1}} < 1 \\ 0, & \text{else.} \end{cases} \quad (4)$$

Eqs (1), (3) and (4) form the basis of the water balance model (WBM), which considers two situations according to whether the forecast sub-surface storage in the pixel is saturated:

- For saturated pixels, defined as $\hat{S}_{m+1} \geq S_{max}$, further excess rainfall cannot be accommodated as subsurface storage and is instead assumed to contribute directly to surface runoff and river flows. Then storage at the end of the next month $S_{m+1} = S_{max}$ and, re-arranging Eq. (1), $Q_{m+1} \approx S_m + P_f - E_f - S_{max}$.

- For unsaturated pixels, defined as $\hat{S}_{m+1} < S_{max}$, excess rainfall is assumed to contribute to both sub-surface storage and runoffnet outflow, and forecasts of these variables are estimated from Eqs (34) and (43) respectively. For a small number of locations in highly spatially variable catchments, where $\frac{\hat{q}_{m+1}}{\hat{S}_{m+1}} \geq 1$, to maintain continuity, S_{m+1}= 0, and $Q_{m+1} = S_m + P_f - E_f$.

Runoff Net outflow estimatestotals (mm) for each grid-square in each region are converted to monthly mean river flows (m^{3}s^{-1}) through lateral transfer of upstream runoff flows from each catchment to the catchment outlet for every river grid-cell, using the 1km flow directions identified for the kinematic wave routing implemented in the G2G Model (Davies and Bell 2008, Bell et al. 2009). WBM flows for every 1km river location are scaled with respect to historical mean WBM flow (1962 -2010) and these standardised flows are averaged to provide a mean value for each of 17 geographic regions (Figure 1Figure 1b). The coarse spatial resolution of the input rainfall forecasts has discouraged the development of river flow forecasts at a 1km resolution to ensure that users of the Hydrological Outlook do not infer that rainfall forecasts are available or skilful at this resolution. , and Pproduction of regional scale forecasts (in preference to national scale) is viewed as a pragmatic compromise.

2.3 Seasonal rainfall forecasts

The long range meteorological forecasts used here were produced using the Met Office Global Seasonal forecast System (MacLachlan *et al*. 2015) and consist of a multi-member ensemble of UK-average (i.e. spatially uniform) monthly total rainfall forecast for the next month, available at the start of each month. The climate model at the core of this forecast system has atmospheric resolution of 0.83 degrees longitude by 0.55 degrees latitude, 85 quasi-horizontal atmospheric levels and an upper boundary at 85km near the mesopause to represent stratospheric processes which are important for winter forecasts (Scaife et al. 2016). The ocean resolution is 0.25^{o} globally in both latitude and longitude with 75 quasi-horizontal levels. This ocean resolution is necessary to reduce key biases in the ocean and atmosphere and give a realistic winter atmospheric blocking climatology in the model (Scaife et al. 2011). A multi-member ensemble of 1- and 3-month rainfall forecasts (mm/day) was run for each season in the period 1996 to 2009 with lagged start dates centred on 1st February, 1st May, 1st August and 1st November. 12 ensemble members were available for forecasts starting in August and February and 24 for those starting in May and November. Thus ensembles of 1-month ahead forecasts are provided for December, March, June and September, and 3-month ahead forecasts for Winter (DJF), Spring (MAM), Summer (JJA) and Autumn (SON). Members from the same start date differ only by stochastic physics. Initial atmospheric and land surface data were taken from ERA interim observational reanalyses and initial conditions for the global ocean and sea ice concentration were from the FOAM data assimilation system (Blockley *et al.* 2014).

The spatially-uniform rainfall forecasts present a dilemma for hydrological modellers who typically require high spatial and temporal resolution weather information to estimate a water balance and represent the highly spatially and temporally variable nature of streamflow. An ensemble of mean UK rainfall forecasts provides no information on whether the rainfall is more likely to occur in the North or South, however, it does provide some indication of whether the rainfall totals will be higher or lower than the climatological (long-term) mean. Such spatially-uniform forecasts will be unable to provide the spatial heterogeneity observed in UK rainfall and would under/overestimate rainfall in Northern/Southern regions if used directly. Instead, the rainfall forecasts P_f are converted to spatially uniform rainfall anomalies, $a = P_f - \overline{P_f}$ (mm) relative to the GloSea5 estimate of climatological mean rainfall ($\overline{P_f}$). A spatially distributed UK monthly rainfall amount, P^*, is then calculated as $P^* = \frac{P_{ij}}{\bar{P}}(\bar{P} + a)$, where $\bar{P}$ and P_{ij} are the UK-mean and the local (1km pixel) monthly mean rainfall (1971-

2000) respectively. This approach to spatial-downscaling using historical mean rainfall observations is similar to one used by Bell et al. (2009) to downscale 25km resolution regional climate model data to a 1km resolution.

To produce the 3-month ahead flow forecasts using the GloSea5 hindcast dataset, either sequential monthly rainfall forecasts or a 3-month mean rainfall forecast were available for use as input, thus for Winter (DJF), forecasts were available for D, J and F separately and the 3-month (DJF) mean. In the analysis that follows forecast skill has been assessed using both temporal resolutions of rainfall forecast, but as the results are very similar, only results for the mean 3-month ahead forecast are presented. This is consistent with the monthly Hydrological Outlook for which forecasts at lead times of 1- and 3-months only are available. Disaggregation of the 3-month ahead forecast into monthly rainfall amounts is achieved through distributing the 3-month rainfall forecast anomaly ~~(relative to the rainfall forecast climatological mean)~~ between the 3 individual months according to ~~the ratios~~their relative ~~of~~ contribution to the~~ir~~ UK ~~mean SAAR~~ mean seasonal rainfall (1962 to 2010).

2.4 Seasonal flow forecasts

To produce seasonal flow forecasts (hindcasts), the water balance model for flows (Section 2.2) is initialised with the most recent G2G estimate of sub-surface water storage (Section 2.1). GloSea5 seasonal rainfall forecasts (1- and 3-month ahead, Section 2.3) are applied alongside climatological monthly mean actual evaporation (AE) estimated from a long Grid-to-Grid model run (1962 to 2010).

While the skill of a single set of forecasts can be compared to observations using measures such as the Pearson correlation coefficient, the performance of an ensemble of seasonal flow forecasts can more easily be assessed using the relative operating characteristic (ROC) skill score (Kharin and Zwiers 2003), used widely for probabilistic weather forecast verification. For ensembles, the ROC is a curve that indicates the relationship between hit rate and false alarm rate as different sorted ensemble members are used as decision thresholds. The ROC is commonly summarized through the integrated area under the curve, *AUC*, using $S_{ROC}=2*AUC-1$: A perfect forecast has $S_{ROC}=1$ ($AUC=1.0$), while forecasts with no skill have $S_{ROC}<=0$ ($AUC<=0.5$). The scores are calculated separately for each of three severity bands (below normal, 0 – 28%; normal, 28-72%; above normal, 72-100%), by ranking standardised river flow forecasts for the 17 geographical regions of Britain in relation to simplified percentile ranges of historical flow estimates for each month based on 49 years of WBM simulated flows (1962–2010). The relatively wide bands were selected to agree with that used by both the rest of the HOUK methods, and by the Hydrological Summary produced by the National Hydrological Monitoring Programme (Dixon et al. 2013), and serve to highlight when flows are unusually high or low.

To assess the importance of various factors involved in the seasonal flow forecasts, the performance of four alternatives is compared:

(a) WBM with GloSea5 rainfall forecasts and the most recent G2G HIC ("GloSea5+HIC");
(b) WBM with GloSea5 rainfall forecasts and a G2G historical mean HIC (1962-2010) ("GloSea5+avHIC");
(c) WBM with an historical observed (spatially averaged) rainfall ensemble (49 members, 1962-2010) and the G2G HIC ("Hist+HIC"); and
(d) Flow persistence with the G2G HIC ("Pers").

Comparing (b) to (a) gives an idea of relative contribution of the HIC to forecast skill, while comparing (c) to (a) gives an idea of the relative contribution of GloSea5 rainfall to forecast skill. Flow persistence (d) (carrying the most recent flow anomaly forward to the next 1- and 3-months) provides a much simpler form of forecast, for overall comparison. The skill of WBM with the G2G historical mean HIC and the ensemble of historical observed rainfall was also assessed but, as would be expected, the ensemble of forecasts had zero skill ($S_{ROC}=0$) and for brevity have been excluded from the analysis. Performance results for the remaining four alternatives are presented in Section 0.

3 Results

3.1 Assessment of water balance model for flows

The performance of the WBM to produce flow forecasts was assessed at a regional scale for the period January 1962 to December 2010 using observed gridded rainfall inputs (i.e. assuming a perfect rainfall forecast) and monthly mean AE (1962-2010) from the G2G, and initialising the WBM each month with the most recent G2G HIC. The resulting output, consisting of a temporal sequence of fixed lead-time, 1- and 3-month ahead regional flow forecasts, was compared to G2G regional mean monthly flows derived from a continuous simulation from 1962-2010 driven by observed (spatially distributed) daily 5km gridded rainfall observations and MORECS PE as input (described in Section 2.1). A comparison with measured river flows at individual sites across Britain has not been undertaken because the WBM has been developed to provide regional monthly flows, and observed mean flows are not available at a regional scale.

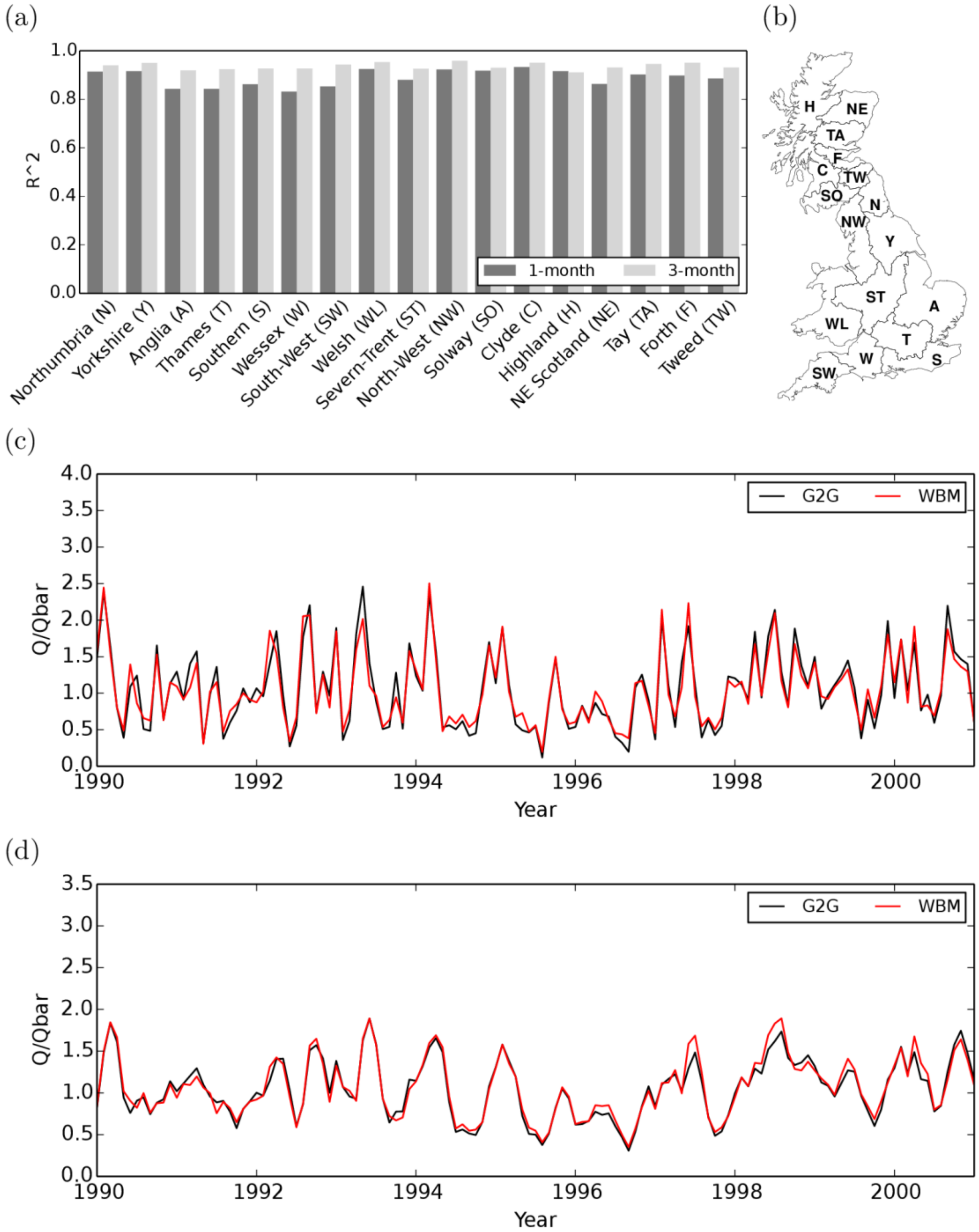

Figure 1 (a) Regional WBM performance (1-month and 3-months ahead) compared to G2G output for 17 regions, in terms of the (Pearson) r^2. The regions are shown in (b). Monthly time-series of (c) 1-month and (d) 3-month ahead forecasts for the period January 1990 to December 2000: mean G2G (black) and WBM (red) simulated flows for the Forth region.

Figure 1~~Figure 1~~a provides a summary of WBM forecast performance in terms of the (Pearson) r^2 at the regional scale when compared to G2G output. For all 17 regions, the 1-month ahead WBM simulates more than 80% of the variability in G2G flows, and for 8 regions (typically upland regions highly responsive to rainfall), it explains more than 90%. The 3-month ahead forecasts all explain more than 90% of the variability. By way of example, Figure 1~~Figure 1~~c and d show modelled regional mean river flows for the Forth region (which has the median model performance for the 1-month lead time), illustrating how closely the 1- and 3-month ahead WBM forecasts match continuous simulation G2G regional mean flows. Regional flows are estimated as the regional mean of $\boldsymbol{Q}/\overline{\boldsymbol{Q}}$ at every location (1km pixel) for which $\overline{\boldsymbol{Q}} > 0.05\ m^3 s^{-1}$. The division by $\overline{\boldsymbol{Q}}$ enables equal weighting for upstream and downstream river locations.

3.2 Assessment of seasonal flow forecasts

An assessment of model skill using the S_{ROC} skill score has been undertaken for Britain as a whole, for 17 regions, two lead-times, four forecast starting points (seasons). A skill assessment should ideally take into account all these factors, and although the average performance measure over all areas shown in Figure 2~~Figure 2~~a disguises the complexity in regional response and forecast model performance at different times of year, it does immediately highlight the utility of using the HIC with rainfall ensembles (GloSea5 or historical) over use of an average HIC or flow persistence. It is important to note that although skill scores improve with the number of ensemble members (Scaife et al., 2014), for the skill assessments here, the ensemble size varies: the historical rainfall ensemble has 49 members, while the rainfall forecast ensemble has 12 or 24 members for autumn/summer and summer/winter respectively. Thus the forecast rainfall skill scores may be an underestimate of the real-time skill (operational GloSea5 forecasts have 42 members). At the 1-month lead time the WBM with G2G HIC driven by an historical rainfall (climatology) ensemble performs best, and the forecasts based on persistence or Glosea5+HIC perform less well, but show some skill. For the longer 3-month lead time, the WBM with G2G HIC driven by either historical or GloSea5 rainfall perform similarly; persistence forecasts (Pers) or use of an average HIC are not recommended at this lead time.

When the overall performance scores shown in Figure 2~~Figure 2~~a are split between seasons, utility of GloSea5 forecasts in September/Autumn and December/Winter becomes apparent (Figure 2~~Figure 2~~b,c: GloSea5+HIC). For forecasts that use the HIC, use of an ensemble of historical rainfall provides some skill (S_ROC > 0) across all seasons, particularly at the 1-month lead time, but use of GloSea5 rainfall forecasts is more skilful in autumn (yellow bars), and also in winter (blue bars) at the 3-month lead time. There is little skill in summer flow forecasts (red bars) whatever type of rainfall forecasts is used, with the best 1-month ahead forecast performance achieved using historical rainfall, and best 3-month ahead forecasts from persistence of current flow conditions or historical rainfall. Scaife et al (2016) identify several mechanisms as to why extratropical seasonal forecast skill is most apparent in winter, and thus less apparent in summer months. Seasonal forecasts of flows in spring have only modest skill and use of historical rainfall and the HIC is recommended at both lead times.

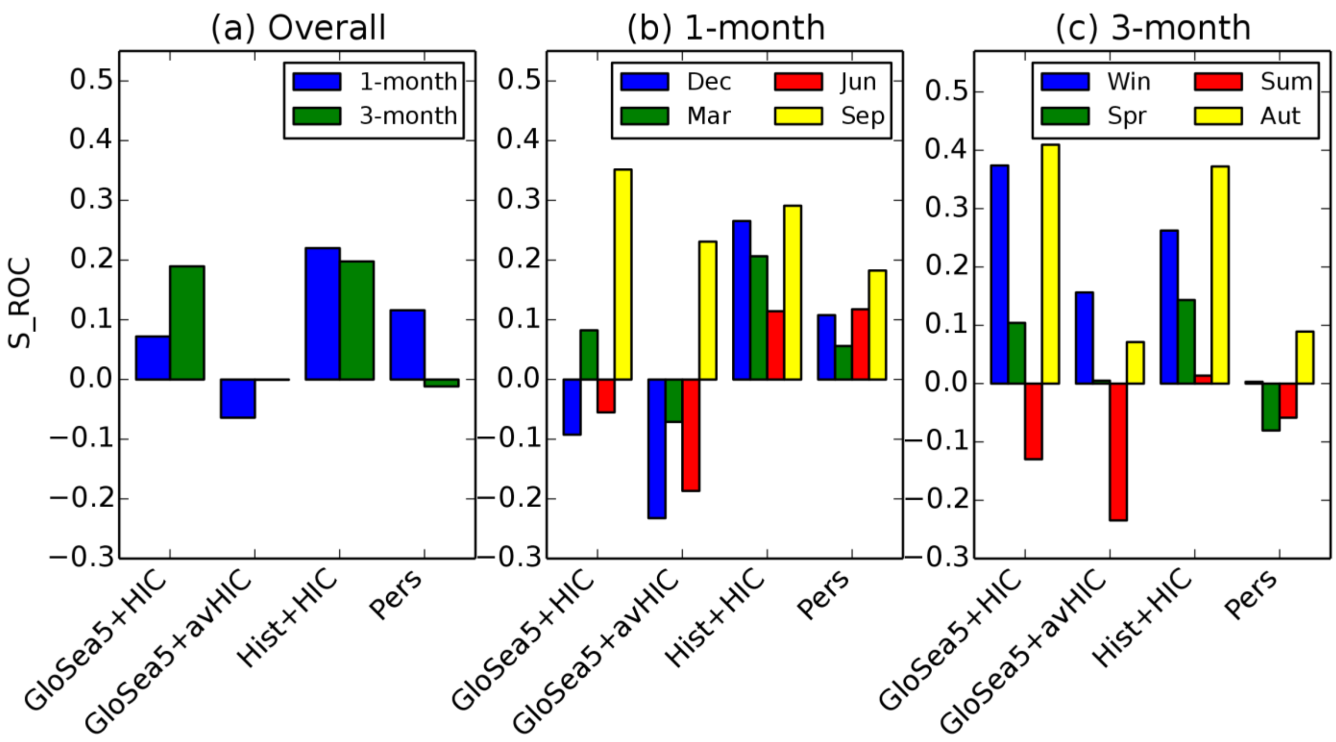

Figure 2 Bar charts showing S_{ROC} (a) averaged across three severity bands, four seasons and 17 regions, at two forecast lead times (1 and 3 months). (b,c) as (a) but presenting results for the four months/seasons separately at the two forecast lead times.

Svensson et al. (2015) and Svensson (2016) highlight the value of a flow persistence forecast in Southern and Eastern regions of Britain in catchments with a high subsurface aquifer storage component and for which river flows respond slowly to rainfall. Here, analysing the ensemble results for each of the 17 regions, the skill of a flow persistence forecast in Southern and Eastern areas is apparent, but using a rainfall forecast ensemble (historical or GloSea5) and the most recent HIC is more skilful. The S_{ROC} skill scores for each of the 17 regions (Figure 3), indicate that skill (averaged over all seasons) is greatly dependent on the geographical region, with the historical rainfall ensemble with a HIC providing the best forecast in most regions, although at the 3-month lead time, a GloSea5 forecast ensemble with the HIC performs well.

Although an ensemble of climate forecasts can provide some indication of the range of possible rainfall totals over the next few months, ensembles of seasonal climate predictions have been shown to possess a low ratio of predictable signal to unpredictable noise (Kumar 2009, Eade et al. 2014, Scaife et al. 2014). These authors indicate that a single ensemble mean forecast can have greater skill than that of the constituent ensemble members, and Murphy (1990) provides a quantification of the apparent improvement in skill through its reduction in initial state uncertainty. More recently, Eade et al. (2014) suggest that improvements in forecast skill could be achieved through using the mean of a large ensemble, followed by a post-processing step to adjust the ensemble mean so its variance agrees with the predictable component of the observed variance. Svensson et al. (2015) restricted their UK-wide analysis of methods for winter flow forecasts to ensemble mean forecasts (3-month ahead) from GloSea5 and rainfall climatology together with a persistence forecast, and found that skilful long-range forecasts of winter flows could be achieved through a combination of the hydrogeological memory of antecedent conditions in southern and eastern parts of the UK, and from meteorological predictability in northern and western areas. Here, it has not been possible to include the effect of the NAO index that was used by Svensson et al. (2015) alongside the GloSea5 winter forecasts, but the skill of ensemble mean GloSea5 rainfall forecasts has been evaluated for all seasons (not just winter). Results as Pearson correlation *(r)* are summarised for UK regions in Figure 4 for both Spring/Summer (labelled "SprSum") and Autumn/Winter ("AutWin"). The forecast methods used are labelled as in Figure 3, but for GloSea5 and Historical rainfall forecast ensembles, only the ensemble mean forecast is used.

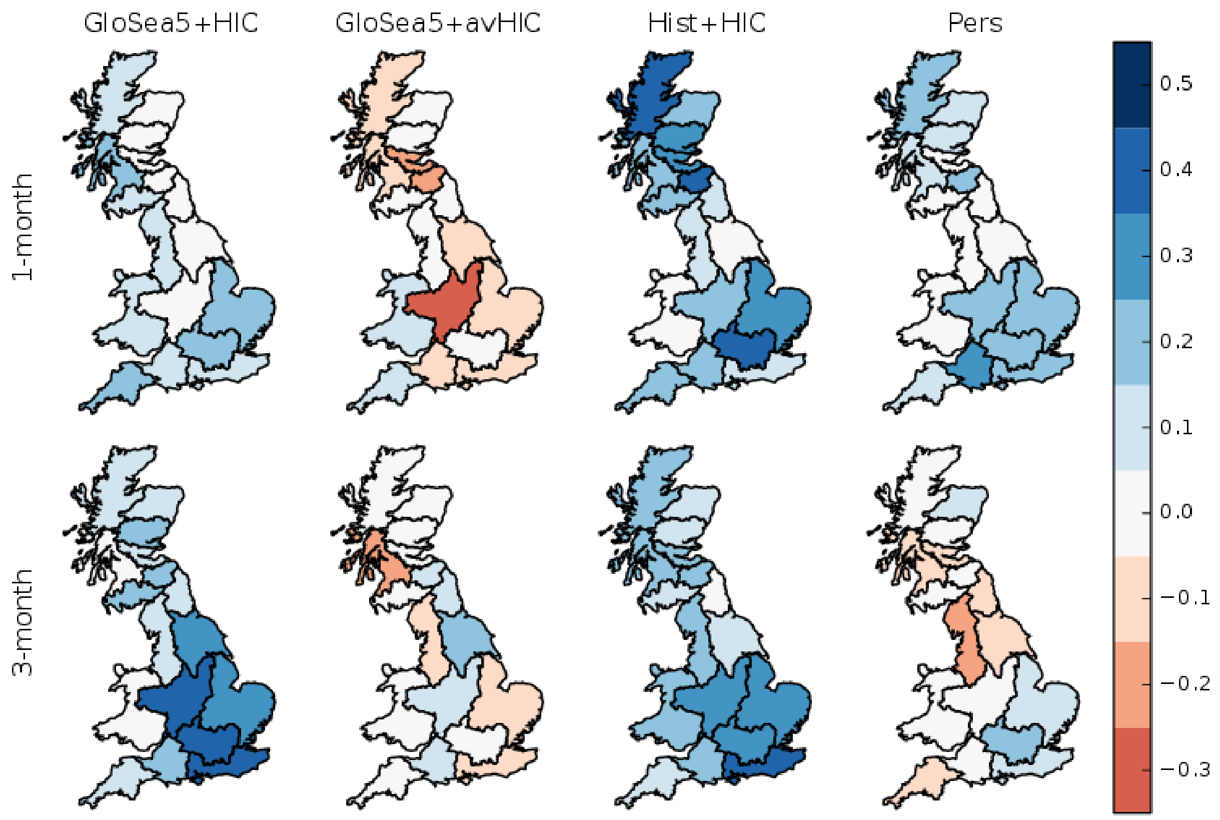

Figure 3 UK regional maps showing mean forecast skill for four seasons in terms of the S_{ROC} ensemble skill score (higher skill shown in blue)

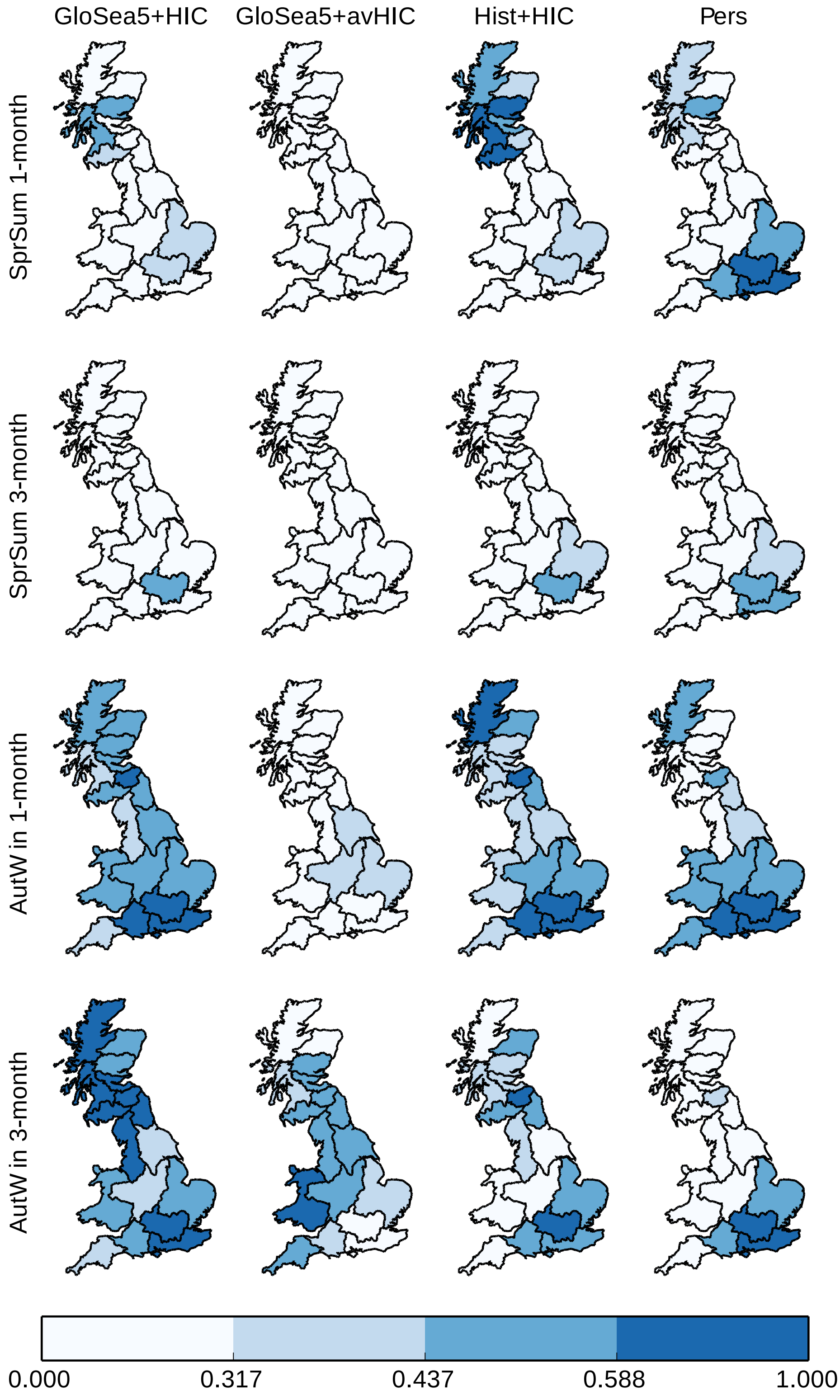

Figure 4 Maps showing 1- and 3-month ahead forecast skill (correlation *r*) for Spring/Summer (top) and Autumn/Winter (bottom). Thresholds are shown for *r*<0.317 (not significant); 0.317<*r*<0.437 (significant at 5% level); 0.437<*r*<0.588 (significant at 1% level) and *r*>0.588 (significant at 0.05% level).

The difference in forecast skill between the seasons is immediately apparent, with a significant level of skill achieved for Autumn/Winter forecasts in many parts of Britain by including GloSea5 seasonal dynamical rainfall predictions:

- For Spring/Summer hydrological forecasts, some skill (at the 1% significance level in the West of Scotland) is afforded through the use of a mean historical rainfall forecast in Scotland (1-month ahead only) and Southeast regions (1- and 3-months ahead), however none of the methods tested is able to provide skilful seasonal forecasts of Spring/Summer flows in Northern and Western England or Wales. Across Britain as a whole, only the use of a mean historical rainfall forecast for a 1-month ahead flow forecasts shows any significant skill (mean correlation of 0.33, significant at the 5% level).

- By comparison, Autumn/Winter flows can be reasonably well forecast across Britain using ensemble mean rainfall forecasts based on GloSea5 or historical rainfall, with mean correlations of 0.53 and 0.50 respectively for 1-month ahead forecasts and 0.59 and 0.43 for 3-months ahead. Forecasts using historical rainfall perform better at the 1-month ahead lead-time than at 3-months ahead, and again, skill is greater in Scotland and Southeast Britain than in Wales and Northern Britain.

The use of an average HIC with Spring/Summer rainfall forecasts from GloSea5 leads to forecasts with no significant skill as it removes the main component of forecast skill which in Spring/Summer is associated with hydrological persistence. However, Autumn/Winter flow forecasts using ensemble mean GloSea5 rainfall and an average HIC perform surprisingly well across Britain, confirming that there is a significant element of skill associated with GloSea5 forecasts~~, particularly~~ in Autumn/Winter at the 3-month lead time, often resulting in skilful flow forecasts in regions where this skill is less dependent on a good HIC.

By comparing forecast skill scores from different model configurations, it becomes possible to attribute overall forecast skill to the different model components such as HIC, GloSea5 ensemble and GloSea5 ensemble mean. Figure 5 provides an *indication* of the source of the forecast skill in Autumn/Winter for each region, alongside critical values for significance levels of Pearson's *r*, 5%, 1% and 0.05% (for a one-tailed test). For each region, the HIC skill is assumed to be the difference between the forecast skill for GloSea5 with HIC and with the long-term mean HIC ("avHIC"). Any GloSea5 skill beyond that associated with HIC can then be attributed to either the mean skill of the individual ensemble members, or to the ensemble mean forecast (if they are greater than the HIC skill).

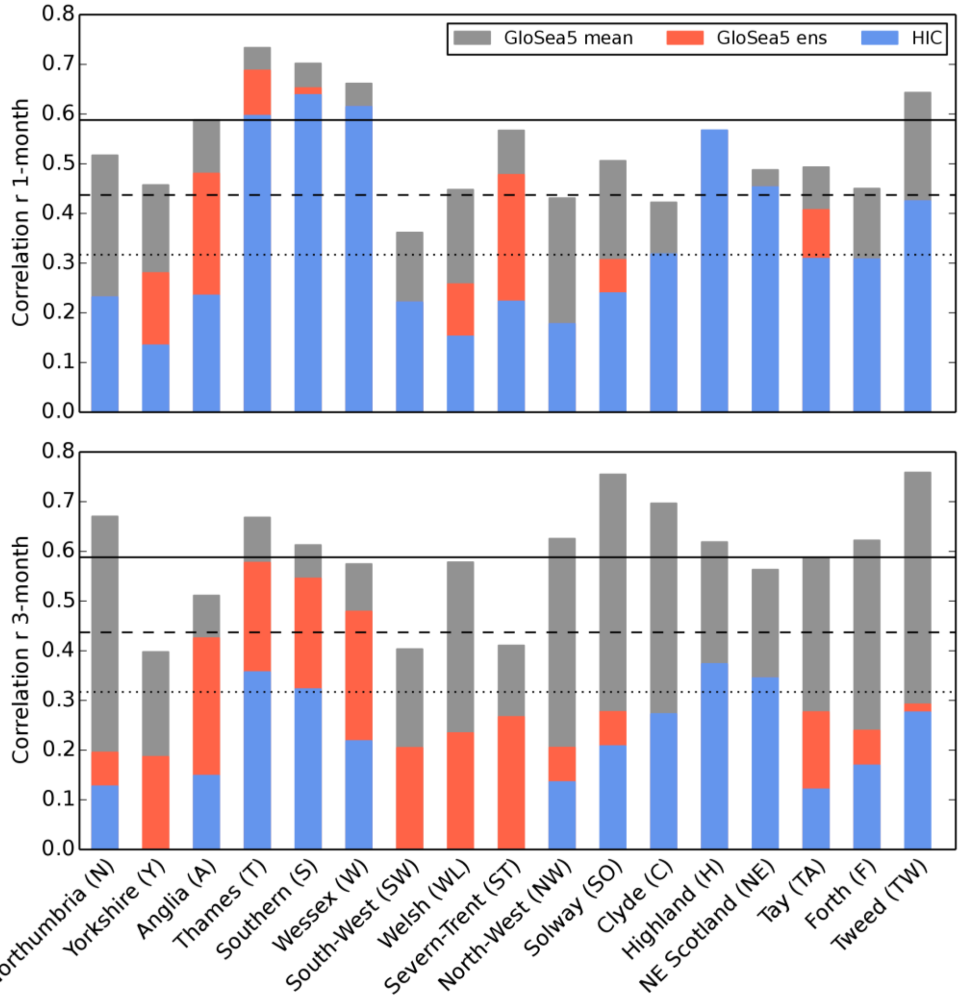

Figure 5 Sources of flow forecast skill (Pearson's *r*) in Autumn/Winter: (a) 1-month ahead; (b) 3-months ahead. Significance levels are shown with black horizontal lines (see Figure 4). A map showing regions geographically is provided in Figure 1b.

At the 1-month ahead lead time, the skill of Autumn/Winter GloSea5-derived forecasts in regions with long-term memory in the Southeast (Thames, Southern, Wessex) and Scotland (Highland, NE Scotland, Tweed) is primarily attributable to the dependence of the flows on the antecedent conditions provided by the HIC (blue bars in Figure 5). Averaged across all regions, the largest source (64%) of skill in the 1-month ahead seasonal flow forecasts comes from the hydrological initial condition. This component of skill is also key to the success of the historical rainfall and persistence forecasts in Autumn/Winter. In many regions, particularly Northern and Western areas (Northumbria, Yorkshire, South-West, Welsh, North-West, Solway, Clyde, Forth, Tweed) a further 20 to 30% increase in skill arises from the GloSea5 ensemble mean (grey bars), and for a few regions (e.g. Anglia, Severn-Trent, Clyde) modest skill (10 to 20%) is derived from the mean skill of the individual ensemble members (red bars). At the 3-month ahead lead time, the influence of the HIC on forecast skill is less apparent, and only 4 regions have levels of HIC-related skill significant at the 5% level. Averaged across all regions, the HIC contributes to only a modest 30% of the forecast skill, whereas GloSea5 forecasts account for ~70% of the forecast skill, of which 46% comes from mean ensemble skill and 23% from the mean skill of the individual ensemble members. This confirms the findings of Section 3.2 (Figure 2) that indicated that at the 3-month lead time, an ensemble of GloSea5 forecasts and a good HIC performs well. A similar analysis of the source of skill in Spring/Summer forecasts (not shown here) indicates that almost all the skill in the forecasts comes from the HIC, with little skill arising from the GloSea5 rainfall forecasts (Figure 2 indicates that an historical rainfall ensemble would be a better choice in Spring and Summer).

Summary and recommendations

The Hydrological Outlook UK (HOUK, Prudhomme et al., under review) provides an insight into future hydrological conditions nationwide across Britain. It uses a range of techniques to provide likely trajectories for seasonal mean river flows and instantaneous groundwater levels on a monthly basis, with particular focus on the next one and three months. One of the techniques uses ensembles of UK-mean, monthly resolution seasonal rainfall forecasts provided by the Met Office GloSea5 model with hydrological modelling tools. The approach combines a high resolution, spatially distributed hydrological initial condition provided by a hydrological model (Grid-to-Grid) driven by weather observations up to the forecast time origin, with a monthly resolution water balance model (WBM) to forecast regional mean river flows for the next 1 and 3 months ahead.

The forecast skill of these regional-scale estimates of the river flows has been assessed for Britain, with results broken down between geographical regions, seasons, and at 1- and 3-month lead times. Every month, the whole ensemble of rainfall forecasts is used in the operational HOUK to provide a range (median and four quartiles) of seasonal forecast flows over the next few months. However, recent literature (discussed in Section 3.2) suggests that ensembles of seasonal climate predictions can have such a low ratio of predictable signal to unpredictable noise that the ensemble mean forecast has much greater skill than the constituent ensemble members. Here, forecast skill has been assessed using both the whole rainfall ensemble and the ensemble mean. By comparing forecast skill scores from different model configurations, it has been possible to attribute overall forecast skill to the different model components such as hydrological initial condition (HIC), GloSea5 ensemble and GloSea5 ensemble mean.

The analysis indicates that only limited forecast skill is achievable for Spring/Summer hydrological forecasts (through the use of historical rainfall rather than rainfall forecasts), however, Autumn/Winter flows can be reasonably well forecast across Britain using ensemble mean rainfall forecasts based on either GloSea5 forecasts or historical rainfall (the preferred type of forecast depends on the region). Flow forecasts using ensemble mean GloSea5 rainfall perform the most consistently well across Britain, and provide the most skilful forecasts overall at the 3-month lead time. Most of the skill (64%) in the 1-month ahead seasonal flow forecasts can be attributed to the hydrological initial condition, whereas for the 3-month ahead lead time, GloSea5 forecasts account for ~70% of the forecast skill.

Svensson et al (2015) highlighted that skilful seasonal predictions of UK river flows are “now a viable proposition” provided by the HOUK every calendar month at a national scale from (http://hydoutuk.net/). Currently, the whole (~42 member) ensemble of GloSea5 rainfall forecasts is used to provide a range (median and four quartiles) of seasonal forecast flows over the next few months. The enhanced level of skill that can be achieved through the use of the ensemble mean forecast alone is an important consideration, but in practice this will be very close to the ensemble median already presented alongside the minimum, maximum and mid-quartile seasonal flow forecasts. Continued presentation of this full range of flow scenarios for the coming 1- and 3-months may be advantageous in that it informs water managers, not only of the most likely possibility, but also to the ~~full range of possible scenarios~~range of possibilities. Based on the skill analysis presented here, users of the Hydrological Outlook UK would be advised to have greatest confidence in Autumn and Winter flow forecasts that use GloSea5 rainfall, particularly at the 3-month lead time. For Spring/Summer flow forecasts, use of an ensemble forecast based on historical rainfall is surprisingly good and would be recommended for use across Scotland, and flow forecasts based on persistence were found to be the most skilful in South-East regions (Thames, Anglia, Wessex and Southern). The HOUK has been in operation for 4 years (publically available from autumn 2013) and thus is a relatively new product. At present, automated web statistics indicate approximately 300 readers or users of the HOUK website per month (Prudhomme et al., under review). Exactly how water managers use the HOUK in practice has not yet been assessed, but ongoing evaluations of the skill in the different methods used in the construction of the Outlook will undoubtedly help provide the evidence required to support use of the product in decision-making.

~~It is hoped that the skill assessment presented here will lead to greater confidence in the use of HOUK seasonal flow forecasts using GloSea5, particularly in autumn and winter months,.~~

Despite the relatively low temporal and spatial resolution of the GloSea5 UK rainfall forecasts (currently: monthly time-step and national-scale), they can be used to provide skilful flow forecasts at a regional/national scale when combined with a hydrological model-simulated estimate of the hydrological initial condition. Given the high spatial heterogeneity in typical patterns of UK rainfall and evaporation, future development of higher resolution seasonal forecasts could lead to substantial improvements in seasonal flow forecast capability, benefitting practitioners interested in predicting flooding and water resources, not only in the UK, but potentially across Europe.

Acknowledgements

This work was funded by the NERC/CEH Water and Pollution Science Programme. AS was supported by the joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101), the UK Public Weather Service research programme.

References

Arribas A, Glover M, Maidens A, Peterson K, Gordon M, MacLachlan C, Graham R, Fereday D, Camp J, Scaife AA, Xavier P, McLean A, Colman A and Cusack S. 2011. The GloSea4 ensemble prediction system for seasonal forecasting. Monthly Weather Review, 139(6), 1891-1910.

Athanasiadis P, Bellucci A, Hermanson L, Scaife AA, MacLachlan C, Arribas A, Materia S, Borrelli A and Gualdi S. 2014. The representation of atmospheric blocking and the associated low frequency variability in two seasonal prediction systems. Journal of Climate, 27, 9082–9100.

Bell VA, Davies HN, Kay AL, Marsh TJ, Brookshaw A and Jenkins A. 2013. Developing a large-scale water-balance approach to seasonal forecasting: application to the 2012 drought in Britain. Hydrological Processes, 27(20), 3003–3012.

Bell VA, Kay AL, Davies HN and Jones RG. 2016. An assessment of the possible impacts of climate change on snow and peak river flows across Britain. Climatic Change, 136(3), 539–553.

Bell VA, Kay AL, Jones RG and Moore RJ. 2007. Development of a high resolution grid-based river flow model for use with regional climate model output. Hydrology and Earth System Sciences, 11(1), 532–549.

Bell VA, Kay AL, Jones RG, Moore RJ and Reynard NS. 2009. Use of soil data in a grid-based hydrological model to estimate spatial variation in changing flood risk across the UK. Journal of Hydrology, 377(3–4), 335–350.

Bierkens MFP and van Beek LPH. 2009. Seasonal predictability of European discharge: NAO and hydrological response time. Journal of Hydrometeorology, **10**, 953–968.

Blockley EW, Martin MJ, McLaren AJ, Ryan AG, Waters J, Lea DJ, Mirouze I, Peterson KA, Sellar A, and Storkey D 2014. Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts. Geosci. Model Dev., 7, 2613-2638.

Centre for Ecology & Hydrology 2016. Hydrological Summary for the United Kingdom: December 2015. Centre for Ecology & Hydrology, pp.12.

Charles A, Timbal B, Fernandez E, Hendon H. 2012. Analog downscaling of seasonal rainfall forecasts in the Murray Darling basin. Monthly Weather Review, 141(3), 1099–1117.

Cole SJ and Moore RJ. 2009. Distributed hydrological modelling using weather radar in gauged and ungauged basins. Advances in Water Resources, 32(7), 1107-1120.

Davies H and Bell V. 2008. Assessment of methods for extracting low resolution river networks from high resolution digital data. Hydrol. Sci. J, 54(1), 17-28.

Dixon H, Hannaford J and Fry MJ. 2013. The effective management of national hydrometric data: Experiences from the United Kingdom. Hydrological Sciences Journal, 58: 1383-1399.

Eade R, Smith D, Scaife AA, Wallace E, Dunstone N, Hermanson L and Robinson N. 2014. Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? Geophys Res Lett., 41(15), 5620–5628.

Hough MN and Jones RJA. 1997. The United Kingdom Meteorological Office rainfall and evaporation calculation system: MORECS version 2.0 - an overview. Hydrol. Earth Syst. Sci., 1(2), 227-239.

Kharin VV and Zwiers FW. 2003. On the ROC Score of Probability Forecasts. Journal of Climate, 16(24), 4145–4150.

Kumar A. 2009. Finite Samples and Uncertainty Estimates for Skill Measures for Seasonal Prediction. Monthly Weather Review, 137, 2622-2631.

Lavers D, Luo L and Wood EF. 2009. A multiple model assessment of seasonal climate forecast skill for applications. Geophys. Res. Lett., 36, L23711.

Lavers DA, Prudhomme C and Hannah DM. 2010a. Large-scale climate, precipitation and British river flows: identifying hydroclimatological connections and dynamics. Journal of Hydrology, 395, 242-255.

Lavers DA, Prudhomme C and Hannah DM. 2010b. Large-scale climatic influences on precipitation and discharge for a British river basin. Hydrological Processes 24 (18), 2555–2563.

McGregor GR and Phillips ID. 2004. Specification and prediction of monthly and seasonal rainfall over the south-west peninsula of England. Quarterly Journal of the Royal Meteorological Society, 130, 193-210.

MacLachlan C, Arribas A, Peterson KA, Maidens A, Fereday D, Scaife AA, Gordon M, Vellinga M, Williams A, Comer RE, Camp J and Xavier P. 2015. Description of GloSea5: the Met Office high resolution seasonal forecast system. Q. J. R. Met. Soc., 141(689), 1072–1084.

Manzanas R, et al. 2017. Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications. Climate Services, http://dx.doi.org/10.1016/j.cliser.2017.06.004

Moore RJ, Cole SJ, Bell VA and Jones DA. 2006. Issues in flood forecasting: ungauged basins, extreme floods and uncertainty. In: I Tchiguirinskaia, KNN Thein & P Hubert (eds.), Frontiers in Flood Research, 8th Kovacs Colloquium, UNESCO, Paris, June/July 2006, IAHS Publ. 305, 103-122.

Perry M, Hollis D, Elms M. 2009. The generation of daily gridded datasets of temperature and rainfall for the UK. Climate memorandum No. 24, UK Met Office.

Prudhomme C, Hannaford J, Boorman D, Knight J, Bell V, Jackson C, Svensson C, Parry S, Bachiller-Jareno N, Davies HN, Davis R, Harrigan S, Mackay J, Mackenzie A, Rudd AC, Smith K, Ward R, Jenkins A. 2017. Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales (under review).

Scaife AA, Arribas A, Blockley E, Brookshaw A, Clark R.T, Dunstone N, Eade R, Fereday, D, Folland CK, Gordon M, Hermanson L, Knight JR, Lea DJ, Maclachlan C, Maidens A, Martin M, Peterson AK, Smith D, Vellinga M, Wallace E, Waters J and Williams A. 2014. Skillful long-range prediction of European and North American winters, Geophys. Res. Lett., 41, 2514–2519.

Scaife AA, Copsey D, Gordon C, Harris C, Hinton T, Keeley S, O'Neill A, Roberts M, and Williams K. 2011. Improved Atlantic winter blocking in a climate model. Geophys. Res. Lett. 38, L23703.

Scaife AA, Karpechko AY, Baldwin MP, Brookshaw A, Butler AH, Eade R, Gordon M, MacLachlan C, Martin N, Dunstone N and Smith D. 2016. Seasonal winter forecasts and the stratosphere. Atmosph. Sci. Lett., 17: 51–56.

Svensson C and Prudhomme C. 2005. Prediction of British summer river flows using winter predictors. Theoretical and Applied Climatology, 82 (1-2), 1-15.

Svensson C, Brookshaw A, Scaife AA, Bell VA, Mackay JD, Jackson CR, Hannaford J, Davies HN, Arribas A, Stanley S. 2015. Long-range forecasts of UK winter hydrology. Environmental Research Letters, 10(6), 064006.

Svensson C. 2016. Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues. Hydrological Sciences Journal, 61, 19-35.

Wedgbrow CS, Wilby RL, Fox HR, O'Hare G. 2002. Prospects for seasonal forecasting of summer drought and low river flow anomalies in England and Wales. International Journal of Climatology, 22, 219-236.

Wilby RL. 2001. Seasonal forecasting of river flows in the British Isles using North Atlantic pressure patterns. Journal of the Chartered Institution of Water and Environmental Management, 15, 56-63.

Yuan X, Wood EF and Ma Z. 2015. A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development. WIREs Water, 2, 523–536.