The 2013/14 Thames basin floods: do improved meteorological forecasts lead to more skilful hydrological forecasts at seasonal timescales?

Article

Accepted Version

Neumann, J., Arnal, L., Magnusson, L. and Cloke, H. (2018) The 2013/14 Thames basin floods: do improved meteorological forecasts lead to more skilful hydrological forecasts at seasonal timescales? Journal of Hydrometeorology, 19. pp. 1059-1075. ISSN 1525-7541 doi: https://doi.org/10.1175/JHM-D-17-0182.1 Available at http://centaur.reading.ac.uk/76984/

To link to this article DOI: http://dx.doi.org/10.1175/JHM-D-17-0182.1

Publisher: American Meteorological Society

The 2013/14 Thames basin floods: Do improved meteorological forecasts lead to more skilful hydrological forecasts at seasonal timescales?

Jessica Neumann[1*], Louise Arnal[1, 2], Linus Magnusson[2] and Hannah Cloke[1, 3.]

1. Department of Geography and Environmental Science, University of Reading, Reading, UK.
2. European Centre for Medium Range Weather Forecasts (ECMWF), Reading, UK.
3. Department of Meteorology, University of Reading, Reading, UK

Corresponding author email: j.l.neumann@reading.ac.uk

Abstract

The Thames basin experienced 12 major Atlantic depressions in winter 2013/14 leading to extensive and prolonged fluvial and groundwater flooding. This exceptional weather coincided with highly anomalous meteorological conditions across the globe. Atmospheric relaxation experiments, whereby conditions within specified regions are relaxed towards a reanalysis, have been used to investigate teleconnection patterns. However, no studies have examined whether improvements to seasonal meteorological forecasts translate into more skilful seasonal hydrological forecasts. This study applied relaxation experiments to reforecast the 2013/14 floods for three Thames basin catchments with different hydrogeological characteristics. The tropics played an important role in the development of extreme conditions over the Thames basin. Greatest hydrological forecasting skill was associated with the tropical Atlantic and less with the tropical Pacific, although both captured seasonal meteorological flow anomalies. Relaxation applied over the north-eastern Atlantic produced confident ensemble forecasts, but hydrological extremes were under-predicted; this was unexpected with relaxation applied so close to the UK. Streamflow was most skilfully forecast for the catchment representing a large drainage area with high peak flow. Permeable lithology and antecedent conditions were important for skilfully forecasting groundwater levels. Atmospheric relaxation experiments can improve our understanding of extratropical anomalies and the potential predictability of extreme events such as the Thames 2013/14 floods. Seasonal hydrological forecasts differed to what was expected from the meteorology alone, thus knowledge is gained by considering both components. In the densely populated Thames basin, considering local hydrogeological context can provide an effective early alert of potential high-impact events, allowing for better preparedness.

1. Introduction

The prediction of water availability over seasonal timescales is beneficial for many aspects of the water sector including flood forecasting, water supply, hydropower generation and navigation. For contingency planners, skilful seasonal hydrological forecasts (SHF) of river and groundwater levels have the potential to provide an indication of possible flood events weeks or months in advance, allowing for more optimal and consistent decisions to be made (Arnal et al. 2017). The operational use of SHF however, remains a challenge due to uncertainties posed by the initial hydrologic conditions (e.g. soil moisture, groundwater levels) and seasonal climate forcings (mainly forecasts of precipitation and temperature) which lead to a decrease in skill with increasing lead times (Wood and Lettenmaier, 2008; Svensson, 2016).

Across the UK and Europe, seasonal streamflow and groundwater forecast methods are currently being developed for application e.g., the UK Met Office Global Seasonal forecast system (GloSea5), The Hydrological Outlook UK and the Copernicus European Flood Awareness System (MacLachan et al. 2014; Mackay et al. 2015; Svensson, 2016), supported by Copernicus projects including SWICCA and EDgE (Copernicus, 2017a, b). Recent UK developments in SHF stem from a prolonged period of drought beginning in 2010, which changed rapidly to widespread flooding during the winter of 2013-14. Driven by the consecutive formation of 12 major Atlantic depressions, the period between December 2013 and February 2014 (DJF 2014) was the wettest in the UK since records began in 1910 (Huntingford et al. 2014; Kendon and McCarthy, 2015; Muchan et al. 2015) and the stormiest for at least 143 years when measured by cyclone frequency and intensity (Matthews et al. 2014). Individual storm events did not yield exceptional rainfall, but accumulated levels over the period led to extensive flooding nationwide, the costs of which were estimated at £1.3 billion (Environment Agency, 2015a). The Thames River basin (south-east UK) received more than half a year's typical rainfall during DJF 2014 (Lewis et al. 2015) which led to concurrent fluvial, pluvial (surface-water/flash) and groundwater flooding; so called compound or coincident flood events (Thorne, 2014).

To date, seasonal hydrological forecasting studies in the Thames and other lowland catchments have primarily identified initial hydrologic conditions as a dominant source of predictability (see Svensson et al. 2015; Svensson, 2016). This is because flow regimes are dominated by slowly-released groundwater and forecast skill can be derived largely from the hydrogeological memory of antecedent conditions. Research conducted elsewhere however, has found that driving hydrological models with more skilful meteorological inputs can capture observed flood events (Yossef et al. 2013) and improve hydrological prediction skill for streamflow (Shukla and Lettenmaier, 2011; Svensson et al. 2015) and groundwater levels (Almanaseer et al. 2014). The contribution of meteorological forcing to SHF skill has also been found to outweigh that provided from initial hydrologic conditions during times of transition from dry to wet climate conditions (Wood et al. 2016). In the UK, there is demand to develop and improve the characterisation and skill of meteorological inputs (Lewis et al. 2015) to improve hydrological forecasting skill at longer lead times and during winter months (e.g., see Li et al. 2009; Shukla and Lettenmaier, 2011; Thober et al. 2014). This is recognised as being particularly important for predicting groundwater levels during extreme events as currently, inadequacies in seasonal rainfall forecasts have resulted in low groundwater forecast skill in all but the most quickly-responding UK catchments (Mackay et al. 2015). Considering better predictions of meteorological conditions, alongside studies focusing on the role of initial hydrologic conditions, will help ascertain which skill improvements have the greatest potential to benefit hydrological forecasts across the Thames basin.

There has been much discussion regarding the meteorological factors that led to the DJF 2014 floods in southern England. Huntingford et al. (2014) proposed various driving mechanisms for the precipitation anomalies, with the North Atlantic Oscillation (NAO) providing the strongest relationship. A positive NAO, characterised by an atmospheric pressure difference between the Azores and Iceland, is associated with increased delivery of rain-bearing cyclonic weather systems into northern Europe during the winter months (Wilby et al. 2004; Svensson et al. 2015). The importance of NAO however, was disputed by van Oldenburg et al. (2015) who stated that a pressure pattern bearing a low to the west of Scotland (as opposed to Iceland) accounted for substantially more of the variance in precipitation during this event, and that combinations of major large-scale modes of variability are likely to have caused the stormy conditions (see also Knight, et al. 2017).

The exceptional conditions of DJF 2014 have thus also been attributed to a hemispheric pattern of severe weather. Relative to the December 1981 to February 2010 ERA-Interim climatology, sea surface temperatures in the tropical Pacific were warmer than usual which disturbed wind patterns over the north east Pacific and deflected the Atlantic jet stream northwards. This brought cold air to North America while eastern Europe was anomalously warm (Palmer, 2014; Watson et al. 2016); this temperature gradient strengthened the jet stream and provided conditions for the continued formation of depressions that affected the UK (Slingo et al. 2014). These anomalous conditions however, were not skilfully forecast by the European Centre for Medium Range Weather Forecasts (ECMWF) operational System 4 (S4) seasonal meteorological forecasting system. As a result, the ECMWF conducted a set of hindcast Atmospheric Relaxation Experiments (ARE) to better understand the role of tropical sea surface temperatures in forcing extratropical circulation response. The ARE relaxed the atmosphere towards the ERA-Interim reanalysis state within specified domains highlighted by negative Rossby Wave Source anomalies (see Rodwell et al. 2015, Magnusson, 2017), forcing S4 to more accurately represent the cyclonic weather conditions prevailing in winter 2013-14. The results provided convincing evidence that the temperature and precipitation anomalies in Europe and North America were embedded within a hemispheric regime that was partly forced by tropical and underlying sea-surface temperatures via ‘Rossby Wave Source’ forcing (associated with its convection and divergent outflow) and that increased precipitation may have acted to re-enforce the upstream wave.

This paper will use the ECMWF’s atmospheric relaxation experiments (ARE) from Rodwell et al. (2015) to relate seasonal hydrological forecasting skill to the forecasting skill of meteorological input and its traceability from different atmospheric domains. Seasonal hydrological reforecasts for DJF 2014 were conducted using the European Flood Awareness System (EFAS) with seasonal meteorological input generated from unforced S4 and three ARE. Specifically, we seek to identify 1) which seasonal meteorological reforecasts perform best, 2) whether increased skill in seasonal meteorological input translates through to more accurate streamflow and groundwater reforecasts for the 2013-14 compound flood event in the Thames River basin, and 3) how hydrological response differs for catchments with different geological and land-use characteristics. We discuss the potential for improvements to seasonal meteorological and hydrological forecasts and the practical value of more skilful seasonal flood forecasts for stakeholders to assist with decision-making in the Thames River basin.

2. Methods

a. Study catchments

The Thames River basin (containing 18 tributary catchments) covers approximately 16,200 km^2 in south-east UK. The western side is predominantly rural, comprising agriculture and woodland with rolling hills and wide, flat floodplains. Towards the centre and east, the basin becomes increasingly urbanised encompassing the towns of Reading, Slough and Greater London. The source of the River Thames is located in the west (elevation up to 350 m asl) and flows 230 km to Teddington Lock which is the official upper tidal limit (elevation 4 m asl) (Fig. 1). The basin encompasses a diverse range of lithologies which greatly influence the flow regime of the Thames and its tributaries; from seasonally spring-fed streams, chalk aquifers with high baseflow, and clay-based rivers that are characterised by a flashy response to storm events and high levels of surface runoff (Bloomfield et al. 2009). Anthropogenic channel modifications, abstraction from major aquifers and discharge points into the river also influence the flow regime; abstraction specifically represents a 5% – 12% reduction in typical annual peak flow (Thames Water, 2010). Recent estimates identified more than 200,000 properties at risk of flooding from a 1:100 year event across the basin (Environment Agency, 2009).

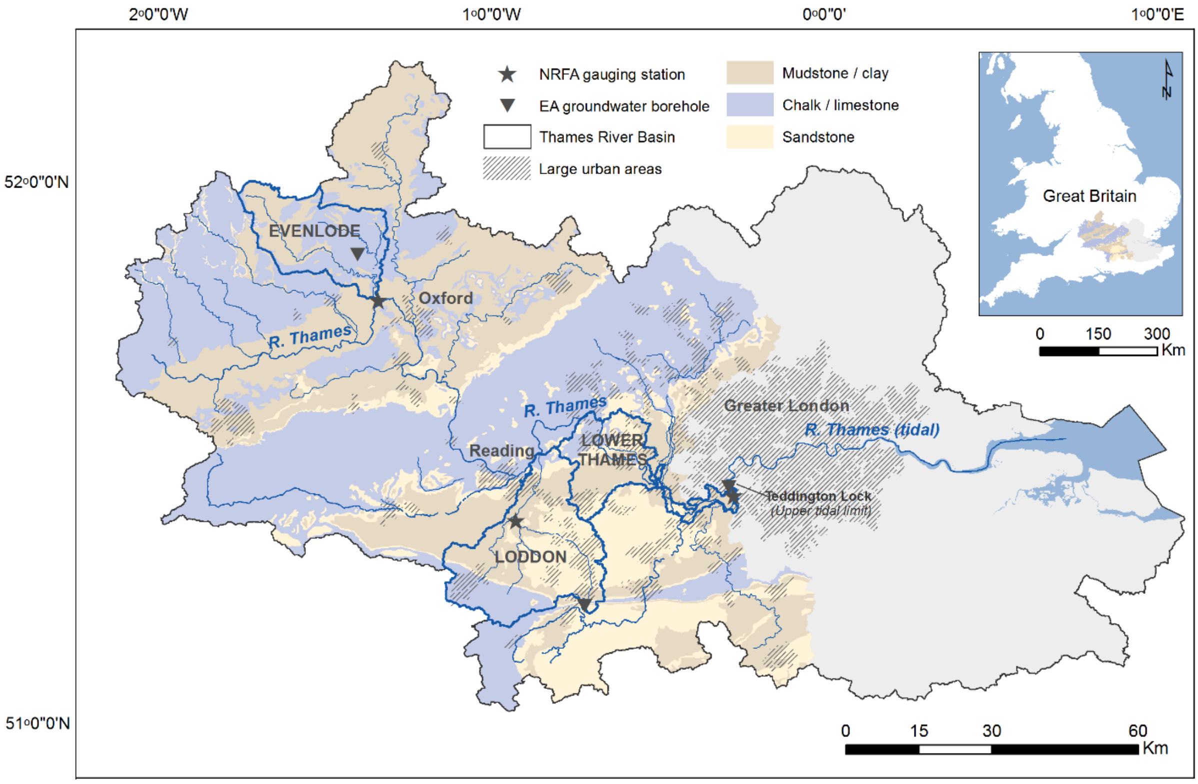


Fig. 1: Thames River Basin (UK) – geographical context and location of the West Thames National River Flow Archive (NRFA, 2017) river gauging stations and Environment Agency (EA, 2017) groundwater boreholes for the Evenlode, Loddon and Lower Thames catchments. West Thames General lithology from British Geological Survey

(after Bloomfield et al. 2011); catchments and rivers from the Water Framework Directive (EA, 2015); basemap and urban areas from Ordnance Survey (OS, 2017).

For the purposes of forecasting fluvial and groundwater floods, this study focuses on 3 catchments with contrasting geological and physical characteristics upstream of Teddington Lock that experienced compound flood events during DJF 2014. The Evenlode is a relatively small (429 sq. km) rural agricultural headwater catchment dominated by a limestone aquifer. The Loddon (682 sq. km) comprises a rural-urban gradient and variable geology. The Lower Thames (324 sq. km) is the furthest point downstream before Teddington Lock and is small and heavily urbanised, with a densely populated floodplain largely overlaying impervious London clay deposits (Fig.1, Table 1).

Table 1: Summary of catchment characteristics and flow regimes. Dominant land-use obtained from CEH LandCover Map 2007 (NERC, 2011) and general lithology (British Geological Survey after Bloomfield et al. 2011). Time-to-peak calculated according to the Revitalised Flood Hydrograph (ReFH) model (Kjeldsen, 2007). Gauged flow regimes under average (Q50) and high extreme (Q10) conditions; values from NRFA (2017).

Catchment	**Dominant land-use** (% of catchment)	**General lithology** (% of catchment)	**Time-to-peak** (*Tp*) / hrs	**Gauged flow regimes** (m^3/s)
Evenlode	Agriculture = 85% Urban = 2% Semi-natural = 13%	Mudstone / clay = 41% Chalk / limestone = 58% Sandstone = 1%	*Tp* = 14.68 hrs	Q50 = 2.52 m^3/s Q10 = 8.93 m^3/s
Loddon	Agriculture = 50% Urban = 20% Semi-natural = 30%	Mudstone / clay = 36% Chalk / limestone = 22% Sandstone = 42%	*Tp* = 9.81 hrs	Q50 = 2.31 m^3/s Q10 = 5.92 m^3/s
Lower Thames	Agriculture = 43% Urban = 35% Semi-natural = 22%	Mudstone / clay = 53% Chalk / limestone = 16% Sandstone = 31%	*Tp* = 34.31 hrs	Q50 = 40.5 m^3/s Q10 = 161.6 m^3/s

b. ECMWF Atmospheric Relaxation Experiments

The rationale behind the ARE was to investigate teleconnection patterns from specified forcing regions. The concept nudges the forecast towards the 'true state' in a pre-defined area during the forecast integration, allowing the downstream impacts from the region to be investigated. The nudging involves adding an extra term to the prognostic equations of the model. Further details about the relaxation technique can be found in section 2.2 in Magnusson (2017). The source regions in this study were selected based on their strong and persistent seasonal-mean forcing on the Rossby wave guide (Rodwell et al. 2015). As these forcing patterns in the source regions potentially had a long predictability, it was expected that the ARE should show impact on the predictability in other parts of the world.

This paper used three ARE model runs (*AR_NPAC*, *AR_WATL* and *AR_EATL*) each representing a different source of atmospheric relaxation (see Fig. 2 (d-f) and Fig. A1 (a-c) in Appendix A). The source regions were chosen where a strong average forcing on the northern mid-latitude flow during DJF 2014 was identified, using Rossby Wave Source as the diagnostic (see Rodwell et al. 2015 for details). The *AR_NPAC* region (centered at 150ºW, ºN) can be physically explained as the region where forcing from the north-east tropical Pacific acted on the mid-latitude flow (Fig. 2d). *AR_WATL* (western Atlantic) was the source region for Atlantic cyclones (75ºW, 35ºN) (Fig. 2e). *AR_EATL* was over the north-east Atlantic (15ºW, 55ºN) and associated with the heavy precipitation experienced during DJF 2014, although not directly linked to any under-lying SST anomaly (Fig. 2f). In each of the regions, the atmosphere was relaxed towards the ERA Interim reanalysis state to determine the impact of each region.

All seasonal meteorological ensemble forecasts (S4 – 51 ensemble members; *AR_NPAC*, *AR_WATL* and *AR_EATL* – 28 members) were produced by the ECWMF Integrated Forecasting System (IFS) coupled atmosphere-ocean-land model. The atmospheric model was run at T255 horizontal resolution (~ 80 km) with 60 vertical levels (91 for System 4), and the NEMO ocean model with 1º horizontal resolution in middle latitudes and higher resolution near the equator. Due to model updates, ARE used a more recent atmospheric model version (*CY40R1*) than that which is used operationally in S4 (*CY36R4*). ECMWF produced a 28-member ensemble of unforced control runs (*NO_AR* – Fig. 2c and Fig. A1d); neither cycle was able to predict the observed planetary wave anomaly in DJF 2014.

c. Seasonal hydrological modelling

Hydrological reforecasts were produced using the European Flood Awareness System (EFAS) seasonal hydrological forecasting suite. EFAS aims to increase preparedness for floods in large European river basins based on operational probabilistic flood forecasts (Bartholmes et al. 2009; Thielen et al. 2009; Smith et al. 2016). The hydrological model used in EFAS is LISFLOOD; a hybrid between a conceptual and a physical rainfall-runoff model combined with a river routing module and run on a 5 km x 5 km grid (Van der Knijff et al. 2010; Alfieri et al. 2014). LISFLOOD is calibrated using non-naturalised data. A new seasonal outlook for EFAS was recently developed by the ECMWF which uses seasonal meteorological ensemble forecasts from System 4 as input to LISFLOOD to extend the EFAS flood forecast horizon up to 7 months (Arnal et al. 2018). A reference daily simulation, termed the EFAS water balance (EFAS-WB), which starts from the initial conditions of the previous day and is forced with the most recent observed meteorological fields (interpolated point measurements of precipitation and temperature) is also run. EFAS-WB is used as initial conditions from which the seasonal forecasts are started and provides a best estimate of the hydrological states at a given time for a given grid point i.e., represents the theoretical upper limit of the model performance.

This study used the seasonal meteorological forecasts from ECMWF's S4 and the three ARE model runs as input to LISFLOOD. All hydrological forecasts were initiated on 1 November 2013 and ran for 4 months with a daily time-step to provide ensemble reforecasts for streamflow (routed river flow measured in m^3s^{-1}) and groundwater level (storage in upper groundwater zone measured in mm). Catchment-averaged daily cumulative precipitation reforecasts (mm per day) were also produced.

Raw daily observation data for streamflow (m^3s^{-1}) and groundwater level (m AOD) were obtained from National River Flow Archive (NRFA) gauging stations and Environment

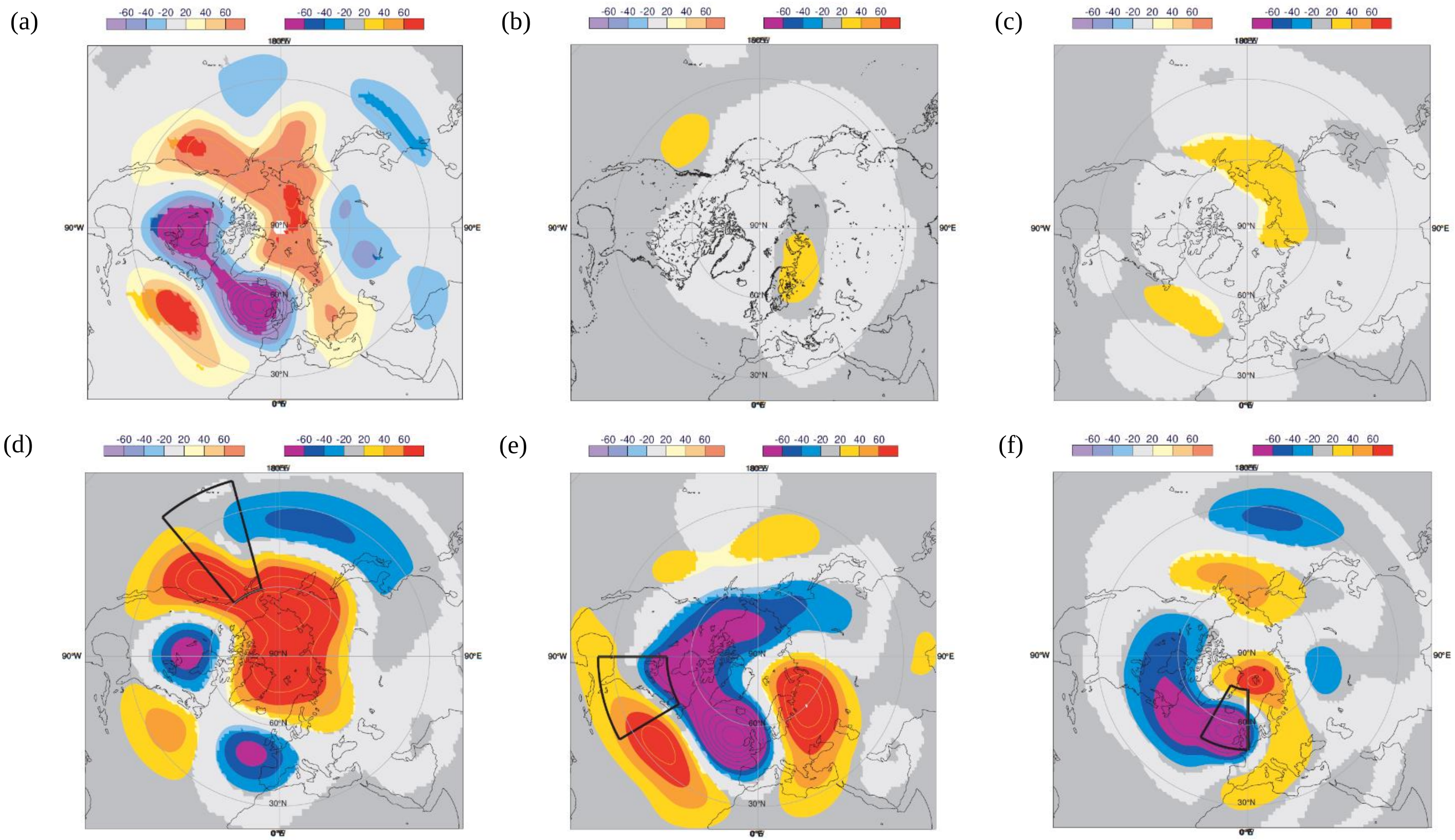


Fig. 2: DJF 2014 anomaly fields of the 500hPa geopotential height (z500) (a) ERA-Interim Analysis, (b) System 4 (*CY36R4*), (c) *NO_AR* (control) equivalent to the operational System 4 forecasts, but with the most recent model cycle (*CY40R1*). (d) – (f) show the equivalent field from ARE where the atmosphere was relaxed towards ERA-Interim reanalyses with Rossby-Wave Source centres identified by black boxes. (d) *AR_NPAC*: centered 150ºW, 35ºN, (e) *AR_WATL*: 75ºW, 35ºN, (f) *AR_EATL*: 15ºW, 55ºN. Model climatology based on 3 ensemble members, initiated from 1 November for the 30 years 1981-2010. Statistical significance at the 5% level is estimated from the 28-member distribution and indicated with saturated colours.

Agency (EA) groundwater boreholes. Within each study catchment, one gauging station on the main river and one groundwater borehole was chosen (Fig. 1). All observation points provided a complete daily record over the 4 month reforecasting period, plus data extending back 20 years (or as far back since start of records) to identify probability exceedance thresholds for that location. Streamflow and groundwater level reforecasts were obtained for the 5 km EFAS grid tile within which the NRFA gauging station and EA groundwater borehole were located to ensure spatial consistency when comparing between forecasts and observations (Fig. 1). Areal precipitation reforecasts were calculated using the arithmetic mean for each catchment.

Continuous ranked probability scores (CRPS) (Hersbach, 2000) were used as a measure of streamflow forecast sharpness and accuracy comparing against the simulated water balance (EFAS-WB) and river gauge observation data. In order to ensure consistency when comparing against different sized ensembles, the relative percentage difference between the 28 and 51 member CRPS values for S4 were calculated; values ranged from 0% (no difference in the Evenlode) to 0.83% (in the Lower Thames) (Fig. A2, Appendix B).

Spearman's Rank Correlation Coefficient (p) was used to compare the median forecasted groundwater level against the simulated EFAS-WB and against borehole groundwater observations. Spearman's Rank is a non-parametric measure of temporal rank correlation which accounted for groundwater levels being expressed in different units.

Finally the EFAS-WB was compared against gauged daily streamflow observations and borehole groundwater observations as an evaluation of the LISFLOOD performance capability to accurately forecast the events in each catchment – this was achieved using Pearson's Correlation Coefficient (r) (to test EFAS-WB streamflow performance) and Spearman's Rank (p) (groundwater performance).

A workflow of all the forecasts, models, methods and analyses used in the paper is shown in Fig. 3.

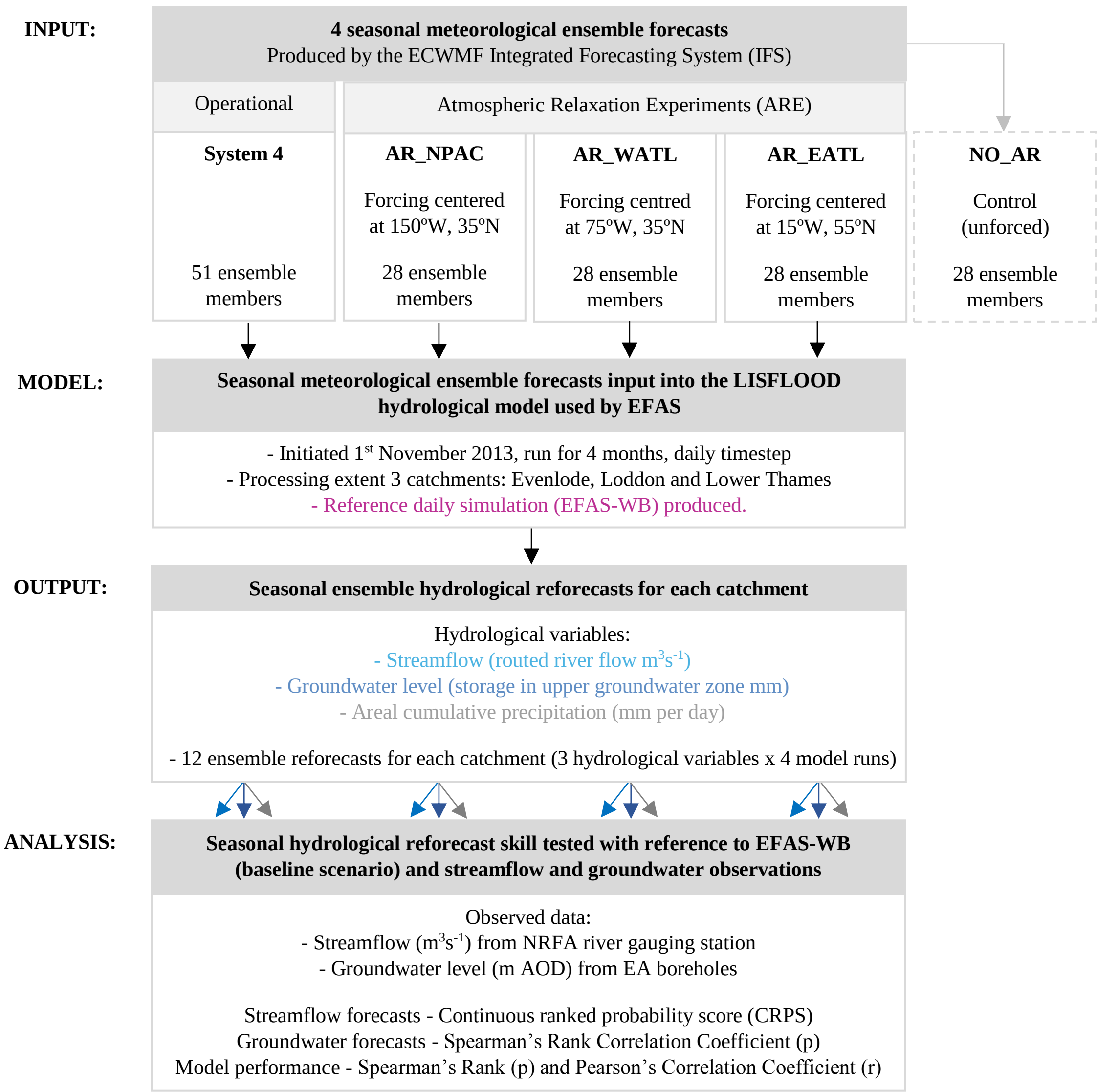


Fig. 3: Workflow detailing seasonal meteorological forecast inputs, hydrological modelling setup, seasonal hydrological reforecast outputs, observational data and analyses used in the paper.

3. Results

a. Meteorological forcing (Fig. 2)

Severe weather conditions did not originate from a single event, but from a number of events between late December 2013 and the end of February 2014 as supported by the negative seasonal average anomaly of the 500hPa geopotential height (z500) over the North-eastern Atlantic, with the UK located at the south-eastern edge (Fig. 2a). For a seasonal forecasting system, capturing this structure was key to predicting the wet anomaly over the UK, however no anomaly was present in the ensemble mean averaged over the whole season for the S4 forecast (Fig. 2b).

Figs. 2 (d-f) show the results from the three ARE. By applying the atmospheric relaxation over the north-eastern Pacific (*AR_NPAC*), the z500 anomalies over the western hemisphere were improved with a negative node over Canada and a positive node over the western Atlantic. There was also a negative anomaly present over the north-eastern Atlantic, with a similar position to the analysis but weaker in magnitude (Fig. 2d).

Relative to *AR_NPAC*, the seasonal anomaly over the eastern Atlantic was better captured both in position and magnitude, with relaxation applied over the eastern part of U.S and western Atlantic (*AR_WATL*) (Fig. 2e). In the final experiment with the relaxation applied over the eastern Atlantic (*AR_EATL*), the negative anomaly was inside the relaxation box. However, the magnitude was less than in the analysis and *AR_WATL*, and the southern extent (outside the box) was not captured (Fig. 2f). The time-series of accumulated precipitation shows *AR_EATL* under-predicted through late December but captured the rainfall better in January and February.

b. Hydrological response to meteorological forcing (Figs. 4 – 6)

1) Overview

Patterns in simulated EFAS-WB cumulative areal precipitation values (pink line) were consistent across all catchments (Figs. 4 – 6). There were a few wet days in early November, a dry period into mid-December followed by higher than average rainfall conditions, with extreme precipitation events corresponding with Atlantic depressions recorded in mid-late December, early January, late January and early February. Over the 4 months, total cumulative areal precipitation (EFAS-WB, pink line) was greatest in the Loddon catchment at 541.2 mm, and lower in the Evenlode and Lower Thames – 494.1 mm and 454.1 mm respectively (Figs. 4 – 6).

During early-mid December, observed gauged daily streamflow (black line) fell below the median (Q50) daily flow record (Table 1) in all three catchments (based on daily flow records from 1994-2014 (NRFA, 2017)). Observed streamflow in all catchments then exceeded the Q10 Exceedance Threshold (percentage of time that streamflow exceeds the 90^{th} percentile) from mid-December through to the end of the study period. Observed borehole daily groundwater levels (black line) exceeded Q50 (EA, 2017) in the Loddon towards the end of January but did not reach the Q10 level of 65.9 mm (not shown on Fig. 5). Observed groundwater levels exceeded Q10 in the Evenlode by mid-December and Lower Thames by mid-January (Figs. 4 and 6).

Comparing observations against EFAS-WB (model performance – Fig. 7), LISFLOOD was capable of predicting streamflow and groundwater levels with reasonably high accuracy in all 3 catchments; correlation coefficients ranged from a positive moderately strong 0.7 for Lower Thames groundwater to a near perfect 0.98 positive correlation for Lower Thames streamflow (Fig. 7).

2) Comparison of seasonal hydrological forecasts

Visual improvements to areal precipitation forecasts, and streamflow forecasts identified by CRPS followed the general pattern: (worst) S4 > *AR_NPAC* > *AR_EATL* > *AR_WATL* (best) (Figs. 4 – 6 and Fig. 8a). This trend was similar for groundwater correlations; all 3 ARE model runs demonstrated marked improvement compared with S4 forecasts which showed negative correlation with simulated EFAS-WB and borehole groundwater observations in each catchment (Fig. 8b).

System 4 forecasted a linear increase in rainfall from November 1st which failed to pick up the low rainfall conditions during end of November to early December, nor the extreme precipitation events in mid-December and beyond. S4 also substantially under-predicted the total amount of precipitation forecast over the 4 month period. The resulting streamflow forecasts showed minimal forecast skill across all catchments; the median did predict above-average streamflow conditions (up to 90th percentile at times) and low numbers of ensemble members forecast some extremes, however the timing and magnitude of peak events were largely incorrect, notably during the first 6 weeks (Figs. 4 – 6). S4 forecasted decreasing groundwater levels over the 4 months leading to negative correlation with borehole observations; this was most pronounced in the Evenlode where observations recorded a 6.10 m increase in groundwater levels in the aquifer (Fig. 4 and Fig. 8b).

The *AR_NPAC* precipitation forecast was similar to S4, although sharper with less spread about the median leading to minor improvements in streamflow and groundwater forecasts. The timing of peak streamflow events were more accurately represented and the magnitude was picked up by the ensemble maximum in many cases. There remained poor forecast quality during the first 6 weeks. Groundwater forecast median showed weak-moderate positive correlation with borehole observations and EFAS-WB (Fig. 8b) although there was a large ensemble spread (Figs. 4 – 6).

Areal precipitation forecast by the *AR_EATL* model run was sharp with a good correlation but under-prediction in respect to the simulated EFAS-WB values in all catchments. Subsequent streamflow forecasts demonstrated accuracy and sharpness but under-prediction and reduced reliability for high extremes. Groundwater forecasts were sharper than S4 and *AR_NPAC* but also under-predicted against the EFAS-WB (Figs. 4 – 6).

AR_WATL produced the best areal precipitation forecast in all catchments; the forecast median traced the simulated EFAS-WB cumulative rainfall patterns with relatively high accuracy until mid-late December when accuracy trailed off. Precipitation forecasts remained sharper than S4 and *AR_NPAC* and total rainfall was matched by the forecast maximum in the Evenlode and Lower Thames (Figs. 4 – 6). Low CRPS and strong positive correlation values indicate a marked improvement for all streamflow and groundwater forecasts (Figs. 8a, b). Extreme streamflow events were missed in late December – early January in all catchments which correlated with the decreased accuracy in the rainfall forecast. Groundwater forecasts showed regular oscillations in all 3 catchments (also apparent in *AR_EATL* and *AR_NPAC* forecasts) (Figs. 4 – 6).

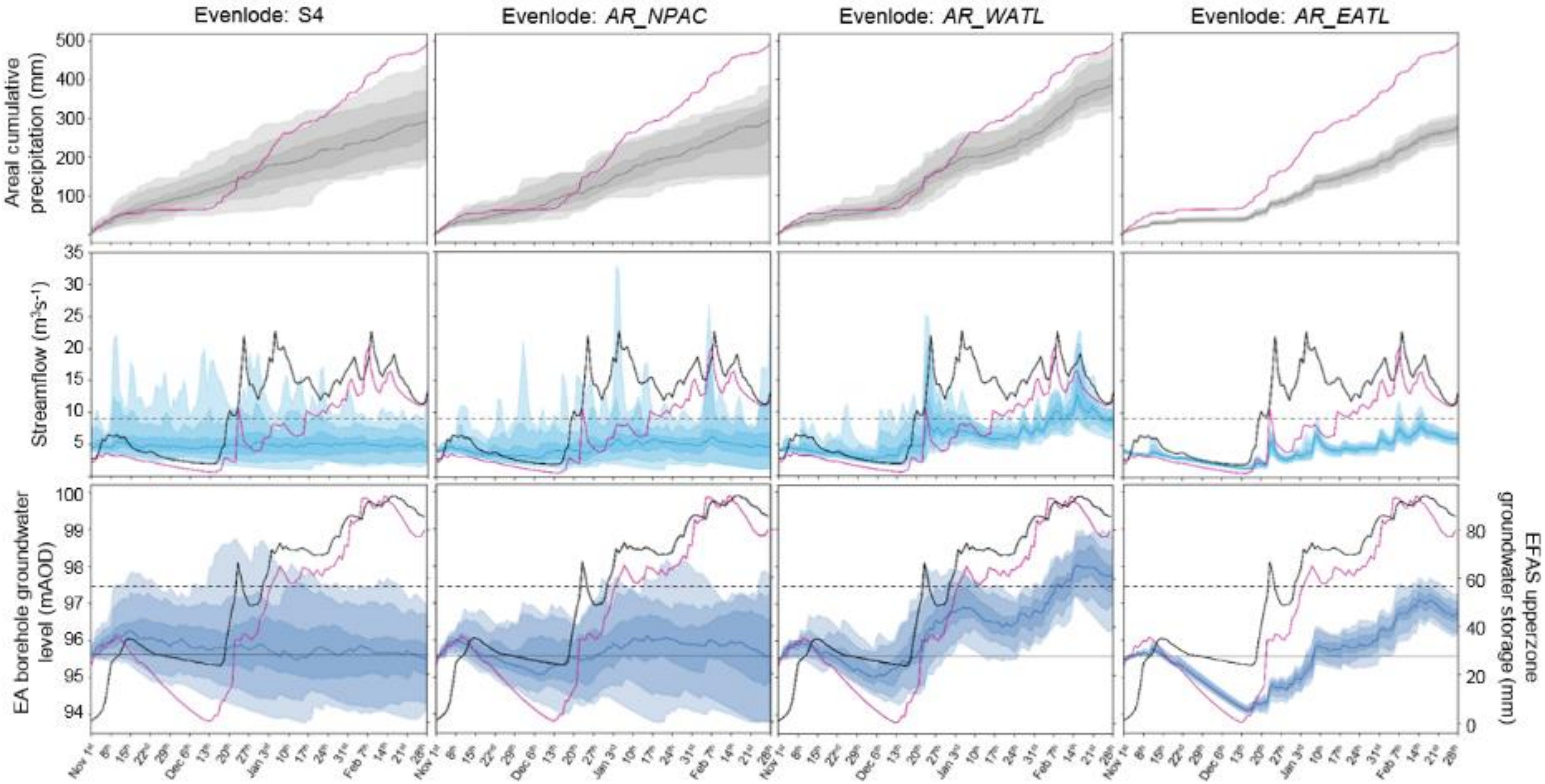


Fig. 4: Areal precipitation, streamflow and groundwater levels; a comparison between S4 and the 3 ARE model runs *AR_NPAC*, *AR_WATL* and *AR_EATL* for the Evenlode catchment. Forecast shading shows min., 5th, 25th, 75th, 95th and max. of the ensemble in all cases. Areal precipitation (mm) = catchment averaged cumulative daily forecast median values (grey). Streamflow (m^3s^{-1}) = daily forecast median (light blue) at river gauging station. Groundwater level (mm) = daily forecast median (dark blue) at groundwater borehole. Observations (black) and simulated EFAS-WB (pink) in all cases. Q10 (long dash) and Q50 (short dash) show exceedance thresholds (based on 1994-2014 observation records or longest available record).

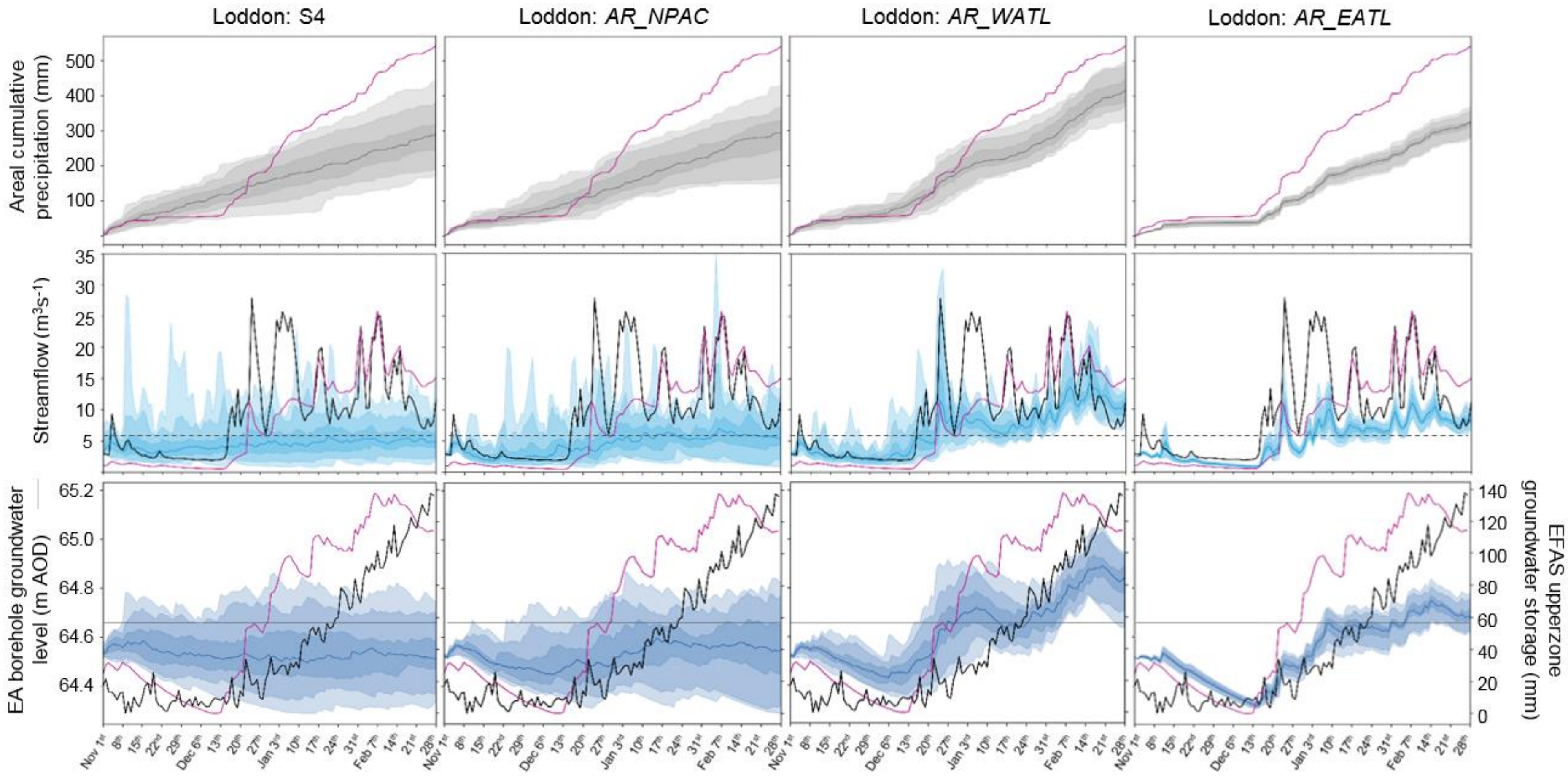


Fig. 5: Areal precipitation, streamflow and groundwater levels; a comparison between S4 and the 3 ARE model runs *AR_NPAC*, *AR_WATL* and *AR_EATL* for the Loddon catchment. Forecast shading shows min., 5th, 25th, 75th, 95th and max. of the ensemble in all cases. Areal precipitation (mm) = catchment averaged cumulative daily forecast median values (grey). Streamflow (m^3s^{-1}) = daily forecast median (light blue) at river gauging station. Groundwater level (mm) = daily forecast median (dark blue) at groundwater borehole. Observations (black) and simulated EFAS-WB (pink) in all cases. Q10 (long dash) and Q50 (short dash) show exceedance thresholds (based on 1994-2014 observation records or longest available record).

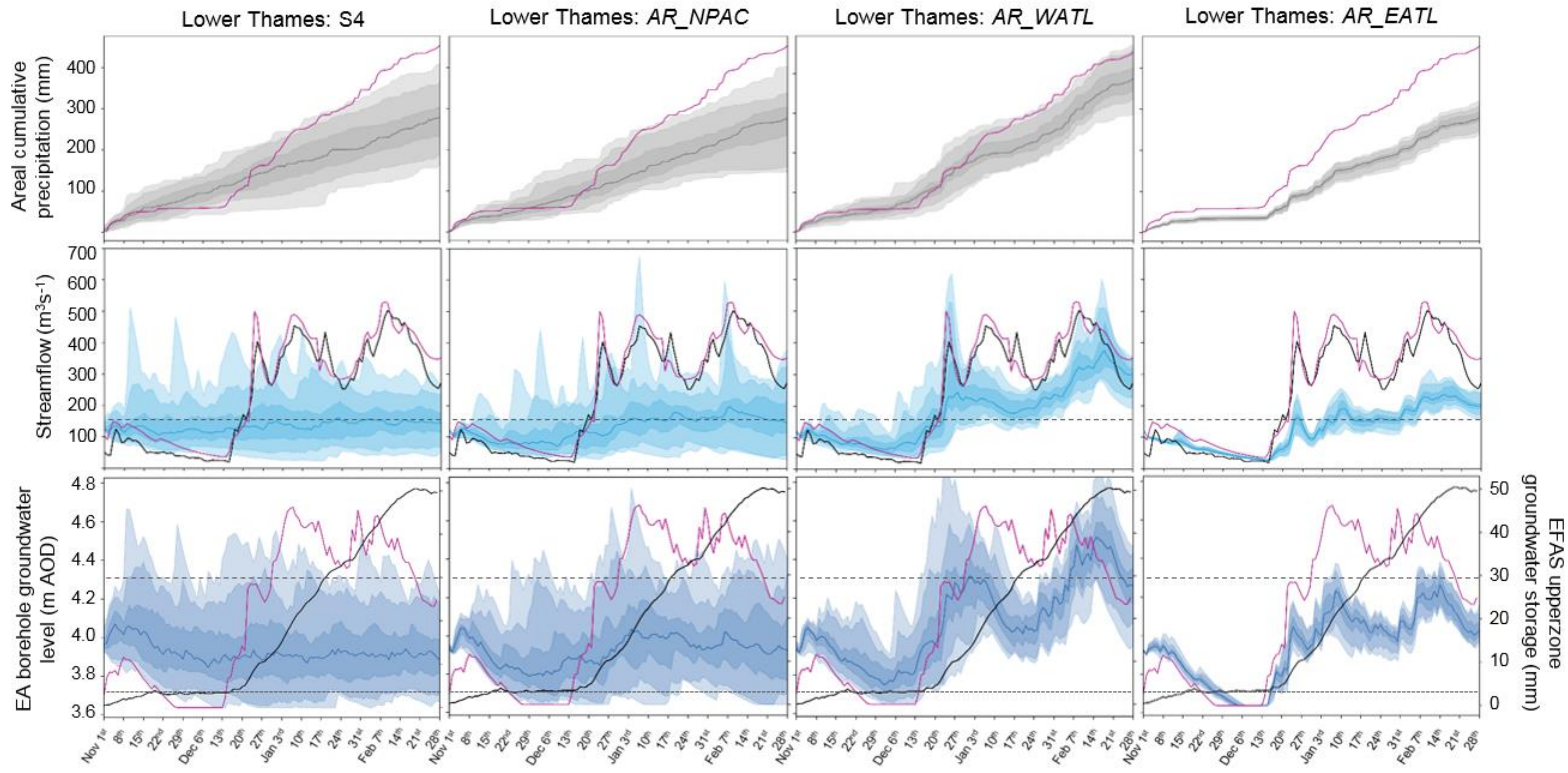


Fig. 6: Areal precipitation, streamflow and groundwater levels; a comparison between S4 and the 3 ARE model runs *AR_NPAC*, *AR_WATL* and *AR_EATL* for the Lower Thames catchment. Forecast shading shows min., 5th, 25th, 75th, 95th and max. of the ensemble in all cases. Areal precipitation (mm) = catchment averaged cumulative daily forecast median values (grey). Streamflow (m^3s^{-1}) = daily forecast median (light blue) at river gauging station. Groundwater level (mm) = daily forecast median (dark blue) at groundwater borehole. Observations (black) and simulated EFAS-WB (pink) in all cases. Q10 (long dash) and Q50 (short dash) show exceedance thresholds (based on 1994-2014 observation records or longest available record).

3) Catchment variation

Observed gauged streamflow patterns (black line), although of different orders of magnitude, were similar for the Evenlode and Lower Thames with consistently high flows from mid-December onwards with 5 – 6 clearly defined peaks (Figs. 4 and 6). LISFLOOD successfully modelled the flow dynamics in the Lower Thames (r_{sens} = 0.98, Fig. 7). The flow pattern was quite accurate in the Evenlode, but overall model performance was lower (r_{sens} = 0.81) as the simulated EFAS-WB did not capture flow pattern between mid-December and the end of January (Figs. 4, 6 and 7). The Loddon had a much flashier response with 8 clearly-defined peaks (black line), coupled with the shortest time-to peak of 9.81 hours (Table 1). Model performance was lowest of the 3 catchments (r_{sens} = 0.73) as LISFLOOD failed to detect peaks around 17th December and under-predicted the extreme events in late December to mid-January (Figs. 5 and 7).

Observed borehole groundwater levels (black line) increased in all 3 catchments; Evenlode +6.10 m, Loddon +0.90 m and Lower Thames +1.13 m (Figs. 4 – 6). The Loddon and Lower Thames recorded average or just below average (Q50) groundwater levels until mid-December when levels showed a consistent and steady rise. Groundwater levels recorded in the Evenlode were more responsive following precipitation events and mirrored streamflow dynamics (Fig. 4). LISFLOOD was best able to model groundwater levels in the Evenlode (p_{sens} = 0.92) but was oversensitive in the Lower Thames (p_{sens} = 0.70) (Fig. 7).

In respect to streamflow and groundwater level forecasting skill compared against observations in each catchment, CRPS and Spearman's Rank (p) indicated that *AR_WATL* provided the best forecast skill in all catchments (solid bars on Figs. 8a, b). $CRPS_{obs}$ for the Loddon and Lower Thames followed the before mentioned pattern S4 > *AR_NPAC* > *AR_EATL* > *AR_WATL* however, the *AR_EATL* model run performed worst in the Evenlode ($CRPS_{obs}$ = 6.32). Groundwater level forecast skill was consistent across catchments with S4 performing worst, and *AR_WATL* best (Fig. 8b).

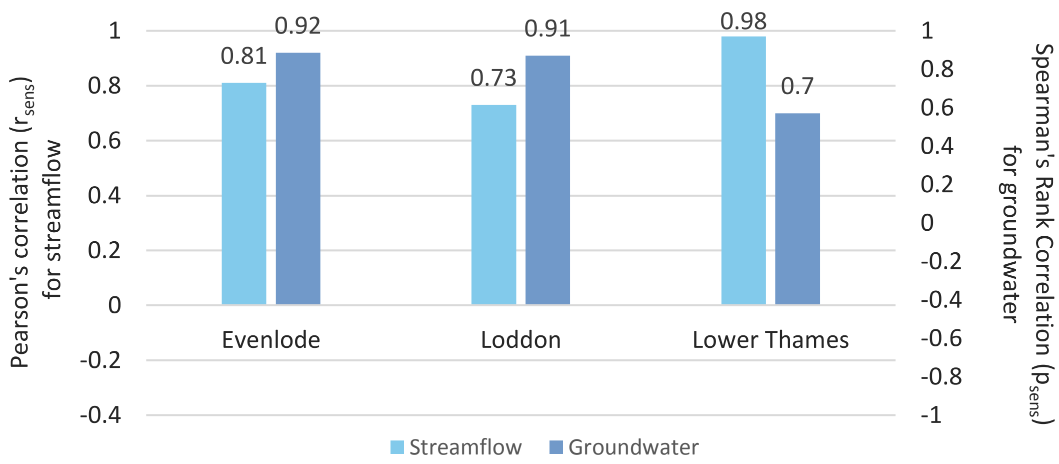


Fig. 7: Comparison between daily simulated EFAS-WB and gauged daily streamflow (m^3s^{-1}) and borehole groundwater level (m AOD) observations as a measure of the capability (sensitivity) of the LISFLOOD model performance. Streamflow performance (light blue) tested using Pearson's correlation (r_{sens}). Groundwater performance (dark blue) tested using Spearman's Rank Correlation Coefficient (p_{sens}).

(a)

(b)

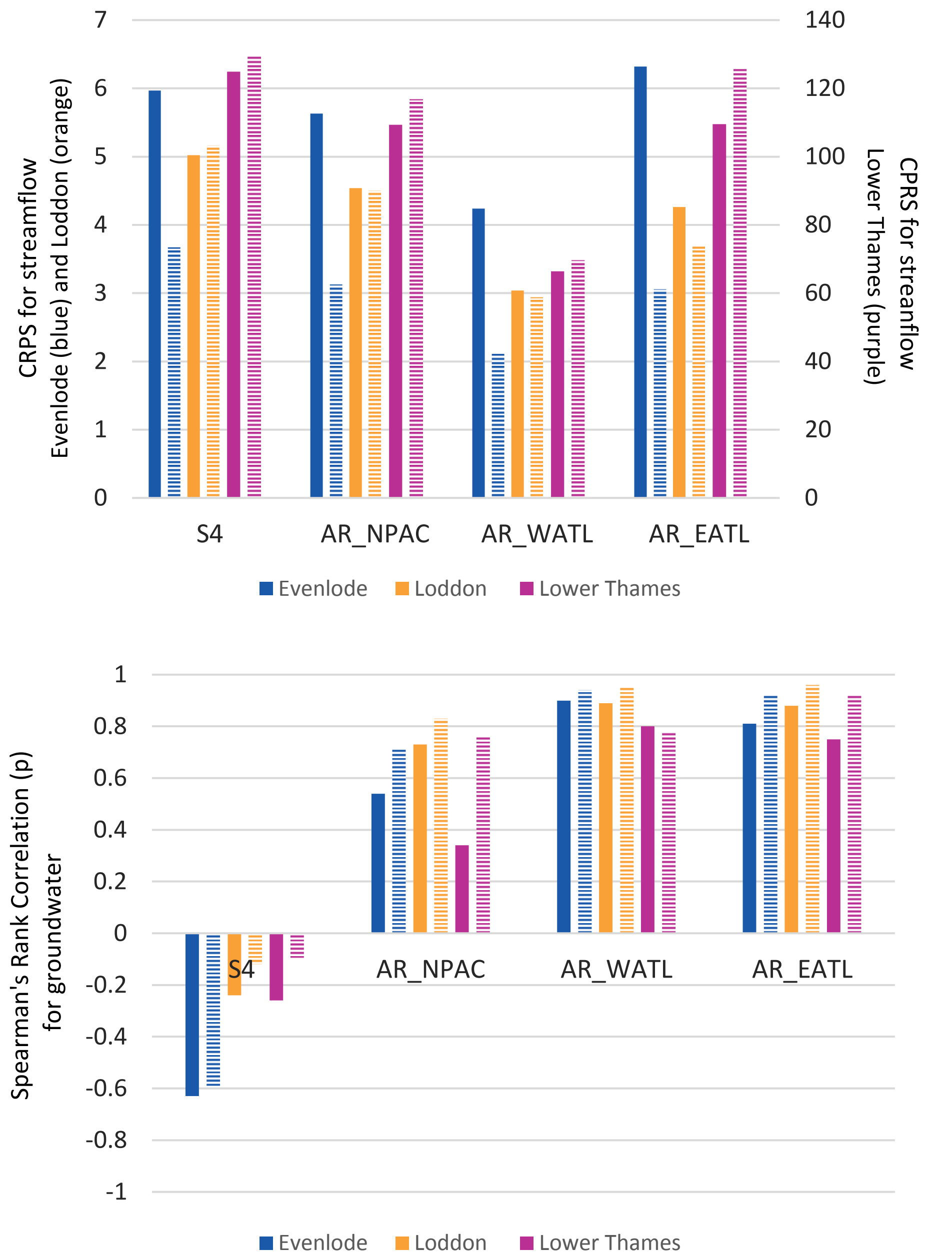


Fig. 8: Comparison of seasonal hydrological forecast skill from operational S4 and three ARE model runs; *AR_NPAC*, *AR_WATL* and *AR_EATL* for the Evenlode (blue), Loddon (orange) and Lower Thames (purple) catchments. (a) CPRS comparing daily streamflow (m^3s^{-1}) forecast against gauged daily streamflow observations (m^3s^{-1}) ($CRPS_{obs}$ shown by solid bars) and against daily simulated EFAS-WB ($CRPS_{sim}$ shown by hashed bars). (b) Spearman's Rank Correlation Coefficient (p) comparing median daily groundwater level (mm) forecast against borehole daily groundwater observations (m AOD) (p_{obs} (solid bars)) and against daily simulated EFAS-WB (p_{sim} (hashed bars)).

4. Discussion

The winter of 2013-14 was exceptional in regard to the large number of Atlantic depressions that affected the UK – the Thames basin saw record precipitation levels which led to widespread and prolonged fluvial and groundwater flooding (Kendon and McCarthy, 2015);

the impacts of which have been well documented (e.g. Slingo et al. 2014; Thorne et al. 2014; Muchan et al. 2015). The drivers of these extreme conditions have also been debated, with papers seeking to identify the atmospheric influences via reviews of multi-scale model simulations investigating factors such as atmosphere, ocean, land-use and demographics (see Huntingford et al. 2014), correlation analyses (van Oldenburg et al. 2015) and relaxation experiments (Rodwell et al. 2015; Watson et al. 2016; Knight et al. 2017). It's largely accepted that a combination of global meteorological influences were important, however studies that link different meteorological inputs and how these translate through to hydrological forecasting skill have not been conducted. Below, we discuss how identification of skill through the meteorological (ARE) and hydrological (EFAS) seasonal forecasting chain may provide indication as to the origins of extreme events and the level of predictability that can be gained if the evolution in parts of the system are known. We also highlight the value of more skilful hydrological forecasts during extreme events for stakeholders, taking into account the variation in catchment properties that exist across the Thames basin.

a. Translating meteorological improvements into more skilful hydrological forecasts

From a meteorological perspective, *AR_EATL* was expected to give the rainfall closest to that observed during DJF 2014 due to the location of southern England at the edge of the relaxation box. Although this experiment provided the most confident hydro-meteorological ensemble forecasts, their value was limited due to under-prediction, likely because *AR_EATL* missed the southward extent of the atmospheric trough and hence, did not fully capture the details of the flow anomaly affecting southern England.

Atmospherically speaking, *AR_NPAC* captured the representation of large-scale flow over the northern Atlantic better than S4, yet this did not translate into an improved precipitation forecast, resulting in low hydrological forecasting skill over the Thames basin. As *AR_NPAC* gave a stronger anomaly in geopotential height over the eastern Atlantic, one could speculate that systematic model errors affected the Rossby wave train from the Pacific to the Atlantic leading to misplacement of the anomaly over the north-eastern Atlantic. Given the relationship between the tropical Pacific and El Niño–Southern Oscillation (ENSO) (Doblas-Reyes et al. 2013) there was hope that seasonal hydrological predictability could be improved in the future with a better modelled teleconnection from ENSO. Rather, the results point to the importance of the western Atlantic, and pose the open question about whether the forcing into this box is linked with the Pacific and/or tropical Atlantic.

The best hydrological forecasts were obtained by the *AR_WATL* experiment. Climatologically, the eastern U.S and Gulf Stream is the most active region for cyclogenesis in the Atlantic (Hoskins and Hodges, 2002) and representation of the anomaly in this region also captured the downstream anomalies over the northern Atlantic. Whether this is a result of the cold anomaly over North America giving a strong temperature contrast (baroclinicity) over the Gulf Stream, or related to the anomaly in the divergent flow from South America as discussed in Knight et al. (2017) has yet to be confirmed. Nonetheless, with future improvements to coupled models, there is scope for an improvement in the teleconnections whereby the results for this study could be revised (Magnusson et al. 2013).

b. More skilful hydrological forecasts; but missed events, oscillations and uncertainty

All three ARE led to improvements in meteorological input which translated through to more skilful streamflow and groundwater level reforecasts compared to S4, with *AR_WATL* performing best. However, there were consistent trends observed for all ARE model runs across the 3 catchments. Poor representation of the hydrological variables during the first 6 weeks coincided with the end of the drought period which preceded the extreme wet conditions. Wood et al. (2016) found that this climatological transition period produced the lowest seasonal predictability as initial hydrologic conditions provide minimal contribution; an effect that may have been heightened in the Thames basin which is largely groundwater-driven (Svensson et al. 2015).

Streamflow forecasts also missed peak streamflow events observed between the end of December and mid-January; timewise these correlated with the point at which precipitation forecasts diverged from the simulated EFAS-WB indicating potential meteorological forcing errors. This was likely due to the extreme nature of the rainfall experienced at this time which was undetected in the meteorological forecast, propagating the error into the hydrological forecast (Davolio et al. 2008) coupled with the uncertainty prevalent at longer lead times (Wood and Lettenmaier, 2008). Structural issues in LISFLOOD however, cannot be ruled out as the EFAS-WB also failed to capture these peak streamflow events in the Evenlode and Loddon catchments. Factors including the variable density of the rain gauge network, lack of horizontal flow (from pixel to pixel) of water in the topsoil and subsoil, and inability to represent fine-scale geological and morphological characteristics in smaller sub-basins for example, may have limited forecast skill and model performance in these catchments. Nonetheless, recovery of the EFAS-WB towards observations later in the seasonal streamflow forecast (mid-January onwards) suggests that these missed events may relate more strongly to meteorological forcing errors.

Groundwater level forecasts showed oscillations of increasing amplitude as precipitation forecasts improved (most obvious for *AR_WATL*) with troughs corresponding with the November dry period and missed rainfall events, and peaks shortly following periods of intense precipitation. A rapid response to rainfall has been observed for aquifer recharge rates and groundwater level time-series (Lee et al. 2006; Bloomfield and Marchant, 2013) indicating that there can be sensitivity of groundwater forecasts to meteorological forcing data. Here, we investigated LISFLOOD upper groundwater zone response where processes represent a mix of fast groundwater including preferential flow rates and sub-surface flow through soil macro-pores (Thielen et al. 2009), thus a quicker response to rainfall, following a dry period was expected (Hassan and Gregory, 2002; Lee et al. 2006). The cyclical dynamic of the forecast may also represent model processes whereby outflow from the upper zone is released once the amount of water being stored reaches a threshold (van der Knijff et al. 2010). As such, it is likely that the observed oscillations represent combined effects of LISFLOOD model set-up and sensitivity to rainfall input.

c. Catchment controls on the variation in hydrological skill improvements

There were differences in observed hydrological response and model performance between catchments, likely explained by the EFAS setup, plus local weather conditions and geographical differences acting at the catchment level. Simulated EFAS-WB streamflow values were low compared to observations in the largely groundwater-driven Evenlode and flashy-responding Loddon (discussed previously), but were well captured in the Lower Thames where peak streamflow observations exceeded ~ 500 m^3s^{-1}. This increased performance may be attributed to the fact that the Thames basin in LISFLOOD is calibrated using gauged daily flow records from the Lower Thames (at Kingston / Teddington Lock). The geographical position of the Lower Thames also represents drainage from the entire upstream catchment, essentially representing a larger basin for which LISFLOOD was designed. The greater coverage of impervious surfaces where LISFLOOD assumes no soil or groundwater storage may also have played a role (Burek et al. 2013).

By contrast, groundwater levels were most accurately modelled and forecast in the Evenlode. Despite its small size and position at the headwater, this suggests that LISFLOOD is well set-up to capture upper zone processes in rural land-use catchments dominated by chalk and limestone lithology (see also Mansour et al. 2013). Antecedent dry conditions are also likely to have played an important role, allowing percolation into the aquifer as explained by the 6.10 m increase in observed groundwater levels (Svensson et al. 2015). By contrast, the relatively small, yet linear increase in observed groundwater levels in the Loddon and Lower Thames could be attributed to the locations of boreholes within less permeable lithologies, and in the case of the Lower Thames, a heavily urbanised area (MacDonald et al. 2012). Bloomfield and Marchant (2013) recognised clear effects of fractured chalk and granular sandstone aquifer characteristics on saturated flow and storage during UK drought conditions and it's not unreasonable to expect differences to carry forward into a period of extreme rainfall. The cumulative effects of upstream groundwater abstractions not accounted for in LISFLOOD, may also explain the notable difference between simulated EFAS-WB and borehole observations in the Lower Thames – an effect less prevalent in the Loddon and Evenlode where boreholes are located towards the top of the catchments. Interestingly, the observed groundwater storage in the Loddon and Lower Thames was more consistent with that of the lower (saturated) zone in LISFLOOD (not considered in this study due to data issues), where water is either stored, or enters the channel via baseflow producing a very slow, seasonally linear response to meteorological forcings (van der Knijff et al. 2010; Mackay et al. 2015). Whether the over-sensitivity of the simulated upper zone response in these catchments (notably the Lower Thames where the EFAS-WB captured groundwater variability at an entirely different frequency) is a result of finer-scale geological and land-use heterogeneity not captured by LISFLOOD (Svensson et al. 2015) or the saturated nature of the impervious deposits which may be better represented by lower zone processes requires further work.

d. Stakeholder implications and future developments

1) Improve climate forcings to deliver more skilful hydrological forecasts

There is currently a lot of focus on improving operational flood forecasts at seasonal timescales and extreme events such as those experienced in DJF 2014 raise important questions about whether there are elements of predictability that are being missed by seasonal forecasting systems (Scaife et al. 2014; van Oldenburg et al. 2015; Watson et al. 2016; Knight et al. 2017). Driving hydrological models with inputs from atmospheric relaxation experiments provides a valid indication of what can be achieved from an operational forecasting system if the determinants of prolonged seasonal mean forcing e.g. ENSO could be captured in the future. While S4 was unable to skilfully capture the seasonal average forcing for DJF 2014, updates such as ECMWF SEAS5, which will shortly replace S4, indicate substantial improvements to SST bias in the tropical Pacific, increased model resolution and a greater ensemble size (Lucas, 2017) that may go some way to improving seasonal hydrological predictions.

2) Different hydrological model set-up to extract skill

While the uncertainty in the forecasts appear to be largest, further analysis might consider adjusting the LISFLOOD model parameters through a process of model calibration (Shi et al. 2008) and/or comparing results with those obtained from a local-scale hydrological model which better captures streamflow and groundwater dynamics in smaller basins. Use of multiple different hydrological models could also help capture a fuller representation of the uncertainty that comes from the hydrology and land surface (e.g. see EDgE, Copernicus, 2017b).

3) DJF 2014 flood signals detected weeks in advance

Based on numerical weather prediction, the fluvial flood events of DJF 2014 were well forecast at a lead-time of 2-3 days and with reasonable accuracy up to 2 weeks ahead of time (Lewis et al. 2015). Groundwater floods, which acted over a longer timescale and were triggered by exceptional aquifer recharge and saturation of permeable deposits, were not well predicted due to the complex dynamics and interactions of the groundwater system with atmosphere and land processes (Mackay et al. 2015). The Environment Agency (EA) are responsible for managing flood risk in the UK. Taking the Loddon December floods as an example, a flood alert based on the EA streamflow thresholds at the river gauging station would have been triggered from 11 m^3s^{-1}: in the case of S4 and *AR_NPAC*, the forecast median did not cross this threshold, although maximum extremes of the ensemble did. For *AR_WATL* and *AR_EATL*, a flood alert for the local area would have been observed with 6 weeks lead time (November 1st) based on the forecast median. This would have allowed mitigation strategies and low-cost preventative actions to be carried out well in advance whilst also highlighting an ‘area to watch’ as the season progressed.

The importance of SHF for advance warning should not be underestimated in densely populated areas such as the Thames basin. Increasing pressures for urban development, intensification of agriculture and clean water, demand a more spatially and temporally integrated approach to management of the water sector (Mansour et al. 2013; Lewis et al. 2015). There is also growing evidence to support an increasing likelihood of Atlantic storms that take a more southerly track akin to DJF 2014 (Slingo et al. 2014) and while the contribution of climate change cannot be definitively related to changes in UK hydrological response (Hannaford et al. 2015), even a small shift in mean climate variability could substantially shorten return periods of such events (Knight et al. 2017). Further studies that trace the meteorological input improvements right through the meteorological-hydrological forecasting chain are therefore strongly advocated.

5. Conclusions

Atmospheric relaxation experiments can improve our understanding of extratropical anomalies and the potential predictability of extreme events such as DJF 2014. Our results highlight that there is meteorological knowledge to be gained by considering the hydrology, i.e., although large-scale seasonal flow anomalies were picked up in the meteorology, these did not always translate through to more skilful hydrological forecasts. Extreme events such as DJF 2014 are difficult to predict with confidence at seasonal timescales, however considering local hydrogeological context for streamflow and groundwater levels can provide an effective early alert of potentially high impact events, allowing for better preparedness and greater confidence in forecasts as an event approaches.

Acknowledgements

This work was supported and funded by the EU Horizon 2020 IMPREX project (http://www.imprex.eu/) (641811). Borehole groundwater level data and flood alert exceedance thresholds were made available by the Environment Agency and we thank Simon Lewis and Stuart Hyslop at the Environment Agency for useful discussions on seasonal forecasting in the Thames basin. The authors have no conflicts of interest to declare.

Appendix A

Atmospheric Relaxation Experiments (ARE) – DJF 2014 total precipitation (TPO) anomaly fields.

Appendix B

Continuous Ranked Probability Scores for System 4 run using 2 to 51 ensemble members.

References

Alfieri, L., F. Pappenberger, F. Wetterhall, T. Haiden, D. Richardson, and P. Salamon, 2014: Evaluation of ensemble streamflow predictions in Europe, *J. Hydrol.*, **517**, 913-922, http://dx.doi.org/10.1016/j.jhydrol.2014.06.035.

Almanaseer, N., A. Sankarasubramanian, and J. Bales, 2014: Improving groundwater predictions utilizing seasonal precipitation forecasts from general circulation models forced with sea surface temperature forecasts, *J. Hydrol. Eng.*, **19**, 87-98, doi:10.1061/(ASCE)HE.1943-5584.0000776.

Arnal, L., H. L. Cloke, E. Stephens, F. Wetterhall, C. Prudhomme, J. Neumann, B. Krzeminski, and F. Pappenberger, 2018: Skilful seasonal forecasts of streamflow over Europe?, *Hydrol. Earth Syst. Sci*., **22,** 2057–2072, https://doi.org/10.5194/hess-22-2057-2018.

Arnal, L., A. W. Wood, E. Stephens, H. Cloke, and F. Pappenberger, 2017: An Efficient Approach for Estimating Streamflow Forecast Skill Elasticity, *J. Hydrometeor.*, **18**(6), 1715-1729, https://doi.org/10.1175/JHM-D-16-0259.1.

Bartholmes, J. C., J. Thielen, M. –H. Ramos, and S. Gentilini, 2009: The European Flood Alert System EFAS - Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts, *Hydrol. Earth Syst. Sci.*, **13**, 141-153, https://doi.org/10.5194/hess-13-141-2009.

Bloomfield, J. P., D. J. Allen, and K. J. Griffiths, 2009: Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin, UK, *J. Hydrol.*, **373** (1-2), 164-176, http://dx.doi.org/10.1016/j.jhydrol.2009.04.025.

Bloomfield, J. P., S. H. Bricker, and A. J. Newell, 2011: Some relationships between lithology, basin form and hydrology: A case study from the Thames basin, UK, *Hydrol. Process.*, **25**, 2518–2530, doi:10.1002/hyp.8024.

Burek, P. A., J. van der Knijff, and A. De Roo, 2013: *LISFLOOD Distributed Water Balance and Flood Simulation Model – Revised User Manual 2013*, Publications office for the European Union.

Copernicus, 2017a: *SWICCA Service for Water Indicators in Climate Change Adaptation*, [www] Available at: http://swicca.climate.copernicus.eu/ (10.08.17).

Copernicus, 2017b: *EDgE*, [www] Available at: http://edge.climate.copernicus.eu/ (10.08.17).

Davolio, S., M. M. Miglietta, T. Diomede, C. Marsigli, C, A. Morgillo, and A. Moscatello, 2008: A meteo-hydrological prediction system based on a multi-model approach for precipitation forecasting, *Nat. Hazards Earth Syst. Sci*., **8**, 143-159, https://doi.org/10.5194/nhess-8-143-2008.

Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues, 2013: Seasonal climate predictability and forecasting: status and prospects, *WIREs Clim. Change*, **4**, 245–268, doi:10.1002/wcc.217.

Environment Agency, 2009: *Thames Catchment Flood Management Plan – Managing Flood Risk*, Summary Report December 2009: EA, Kings Meadow House, Reading.

Environment Agency, 2015a: *The costs and impacts of the winter 2013 to 2014 floods*. Project Summary SC140025. Defra/Environment Agency Joint R&D programme.

Environment Agency, 2015b: *WFD River Basin Districts Cycle 2,* Data contains Environment Agency information © Environment Agency and/or database right. All rights reserved. Data sourced under Open Government Licence. Available at: https://data.gov.uk/dataset/wfd-river-basin-districts-cycle-2 (10.08.16).

Environment Agency, 2017: *Groundwater Level Measurements (AfA075),* Data contains Environment Agency information © Environment Agency and/or database right. All rights reserved. Data sourced under Environment Agency Conditional Licence.

Hassan, M., and P. J. Gregory, 2002: Dynamics of water movement on Chalkland, *J. Hydrol.,* **257**, 27–41, https://doi.org/10.1016/S0022-1694(01)00530-3.

Hersbach, H., 2000: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, *Wea. Forecasting,* **15**, 559-570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hoskins, B. J., and K. I. Hodges, 2002: New Perspectives on the Northern Hemisphere Winter Storm Tracks, *J. Atmos. Sci*., **74**, 1041-1061, https://doi.org/10.1175/1520-0469(2002)059<1041:NPOTNH>2.0.CO;2.

Huntingford, C., and Coauthors, 2014: Potential influences on the United Kingdom's floods of winter 2013/14. *Nature Clim. Change,* **4**, 769-777, http://dx.doi.org/10.1038/nclimate2314.

Kendon, M., and M. McCarthy, 2015: The UK's wet and stormy winter of 2013/2014, *Weather*, **70** (2), 40-47, doi:10.1002/wea.2465.

Kjeldsen, T. R., 2007: *Flood Estimation Handbook Supplementary Report No. 1. The revitalised FSR/FEH rainfall-runoff method,* Centre for Ecology and Hydrology, Wallingford.

Knight, J. R., and Coauthors, 2017: Global meteorological influences on the record UK rainfall of winter 2013–14, *Environ. Res. Lett.,* **12**, 074001.

Lee, L. J. E., D. S. L. Lawrence, and M. Price, 2006: Analysis of water-level response to rainfall and implications for recharge pathways in the Chalk aquifer, SE England, *J. Hydrol.,* **330**, 604-620, https://doi.org/10.1016/j.jhydrol.2006.04.025.

Lewis, H., and Coauthors, 2015: From months to minutes – exploring the value of high-resolution rainfall observation and prediction during the UK winter storms of 2013/2014, *Met. Apps*., **22**, 90-104, doi:10.1002/met.1493.

Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, *J. Geophys. Res.,* **114**, D04114, doi:10.1029/2008JD010969.

Lucas, D., 2017: *Implementation of seasonal forecast SEAS5*, [www] https://software.ecmwf.int/wiki/display/FCST/Implementation+of+Seasonal+Forecast+SEAS5#ImplementationofSeasonalForecastSEAS5-References (18.09.17)

Macdonald, D., A. Dixon, A. Newell, and A. Hallaways, 2012: Groundwater flooding within an urbanised flood plain, *J. Flood Risk Manage*, **5**, 68–80, doi:10.1111/j.1753-318X.2011.01127.x.

Mackay, J. D., C. R. Jackson, A. Brookshaw, A. A. Scaife, J. Cook, and R. S. Ward, 2015: Seasonal forecasting of groundwater levels in principal aquifers of the United Kingdom, *J. Hydrol.*, **530**, 815-828, http://dx.doi.org/10.1016/j.jhydrol.2015.10.018.

MacLachlan, C., and Coauthors, 2015: Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system, *Q.J.R. Meteorol. Soc*., **141**, 1072–1084, doi:10.1002/qj.2396.

Magnusson, L., M. Alonso-Balmaseda, S. Corti, F. Molteni, and T. Stockdale, 2013: Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors, *Clim. Dyn.*, **41**, 2393 – 2409, doi.org/10.1007/s00382-012-1599-2.

Magnusson, L., 2017: Diagnostic methods for understanding the origin of forecast errors, *Q.J.R. Meteorol. Soc*., **143**, 2129–2142, doi.org/10.1002/qj.3072.

Mansour, M., J. Mackay, C. Abesser, A. Williams, L. Wang, S. Bricker, and C. Jackson, 2013: Integrated Environmental Modeling applied at the basin scale: Linking different types of models using the OpenMI standard to improve simulation of groundwater processes in the Thames Basin, UK, In: MODFLOW and More: Translating Science into Practice, IGWMC, 2-5 June 2013. Golden, Colorado, USA (unpublished) {http://nora.nerc.ac.uk/501789/}.

Matthews, T., C. Murphy, R. Wilby, and S. Harrigan, 2014: Stormiest winter on record for Ireland and UK, *Nat. Clim. Change*, **4**, 738–740, doi:10.1038/nclimate2336.

Molteni, F., and Coauthors, 2011: *The new ECMWF seasonal forecast system (*System 4*)*, ECMWF Tech. Memo., **656**, 1–49.

Muchan, K., M. Lewis, J. Hannaford, and S. Parry, 2015: The winter storms of 2013/2014 in the UK: hydrological responses and impacts, *Weather*, **70** (2), 55-61, doi:10.1002/wea.2469.

National River Flow Archive, 2017: *Search for Gauging Stations*, [www] http://nrfa.ceh.ac.uk/data/search. (10.07.17).

Natural Environment Research Council (NERC), 2011: *Landcover Map 2007 Dataset documentation*, Version 1.0, Oxfordshire: Centre for Ecology and Hydrology (CEH).

Ordnance Survey, 2017: *OS Open Data – Strategi and Miniscale*, Data contains Ordnance Survey information ©. All rights reserved. Data sourced under Open Government Licence. Available at: https://www.ordnancesurvey.co.uk/opendatadownload/products.html (16.08.16).

Palmer, T., 2014: Record-breaking winters and global climate change, *Science*, **344**, 803–804, doi:10.1126/science.1255147.

Rodwell, M. J., L. Ferranti, L. Magnusson, A. Weisheimer, F. Rabier, and D. Richardson, 2015: *Diagnosis of northern hemispheric regime behaviour during winter 2013/14*, ECMWF Tech. Memo., **769**, 1-12.

Shi, X., A. J. Wood, and D. P. Lettenmaier, 2008: How Essential is Hydrologic Model Calibration to Seasonal Streamflow Forecasting? *J. Hydrometeor.*, **9**, 1350-1363.

Shukla, S., and D. Lettenmaier, 2011: Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill, *Hydrol. Earth Syst. Sci*., **15**, 3529–3538, https://doi.org/10.5194/hess-15-3529-2011.

Slingo, J., and Coauthors, 2014: *The recent storms and floods in the UK*, Met. Office Briefing Report with Centre for Ecology and Hydrology (CEH). Available at: http://www.metoffice.gov.uk/media/pdf/n/i/Recent_Storms_Briefing_Final_07023.pdf (01.03.17).

Smith, P., and Coauthors, 2016: *On the operational implementation of the European Flood Awareness System (EFAS)*, ECMWF Tech. Memo., **778**, 1-34.

Svensson, C., and Coauthors, 2015: Long-range forecasts of UK winter hydrology, *Environ. Res. Lett.*, **10**, 064006, doi:10.1088/1748-9326/10/6/064006.

Svensson, C., 2016: Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues, *Hydrol. Sci. J.*, **61**, 19-35, doi:10.1080/02626667.2014.992788.

Thames Water, 2010: *Hydrological Context for Water Quality And Ecology Preliminary Impact Assessments*, Technical Appendix B, Thames Water Utilities Ltd 2W0H Lower Thames Operating Agreement (Cascade Consulting).

Thielen, J., J. Bartholmes, M. –H. Ramos, and A. P. J. De Roo, 2009: The European Flood Alert System – Part 1: Concept and development, *Hydrol. Earth Syst. Sci., Discuss.*, **5**, 257-287, https://doi.org/10.5194/hess-13-125-2009.

Thober, S., A. Wood, L, Samaniego, M. Clark, R. Kumar, and M. Zink, 2014: The elasticity of hydrological forecast skill with respect to initial conditions and meteorological forcing for two major flood events in Germany, Copernicus. EGU General Assembly 2014, Vienna, Austria, id.3089.

Thorne, C., 2014: Geographies of UK flooding in 2013/14, *The Geogr. J.*, **180**, 297-309, doi:10.1111/geoj.12122.

Van der Knijff, J. M., J. Younis, and A. P. J. De Roo, 2010: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, *Int. J. Geogr. Inf. Sci.* **24**, 189-212, doi:10.1080/13658810802549154.

Van Oldenburg, G. J., D. B. Stephenson, A. Sterl, R. Vautard, P. Yiou, S. S. Drijfhout, H. von Storch, and H. van den Dool, 2015: Drivers of the 2013/14 winter floods in the UK, *Nature Clim. Change*, **5,** 490–491, doi:10.1038/nclimate2612.

Watson, P. A. G., A. Weisheimer, J. R. Knight, and T. N. Palmer, 2016: The role of the tropical West Pacific in the extreme Northern Hemisphere winter of 2013/2014, *J. Geophys. Res. Atmos.*, **121**, 1698–1714, doi:10.1002/2015JD024048.

Wilby, R. L., 2001: Seasonal forecasting of river flows in the British Isles using North Atlantic pressure patterns, J. *Chart. Inst. Water Environ. Manage*., **15,** 56–63, doi:10.1111/j.1747-6593.2001.tb00305.x.

Wood, A. W., and D. P. Lettenmaier, 2008: An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophys. Res. Lett.*, **35**, L14401, doi:10.1029/2008GL034648.

Wood, A. W., T. Hopson, A. Newman, L. Brekke, J. Arnold, and M. Clark, 2016: Quantifying Streamflow Forecast Skill Elasticity to Initial Condition and Climate Prediction Skill, *J. Hydrometeor.*, **17**, 651-667, doi:10.1175/JHM-D-14-0213.1.

Yossef, N.C., H. Winsemius, A. Weerts, R. van Beek, and M. F. P. Bierkens, 2013: Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing, *Water Resour. Res*,, **49**, 4687-4699, doi:10.1002/wrcr.2035.

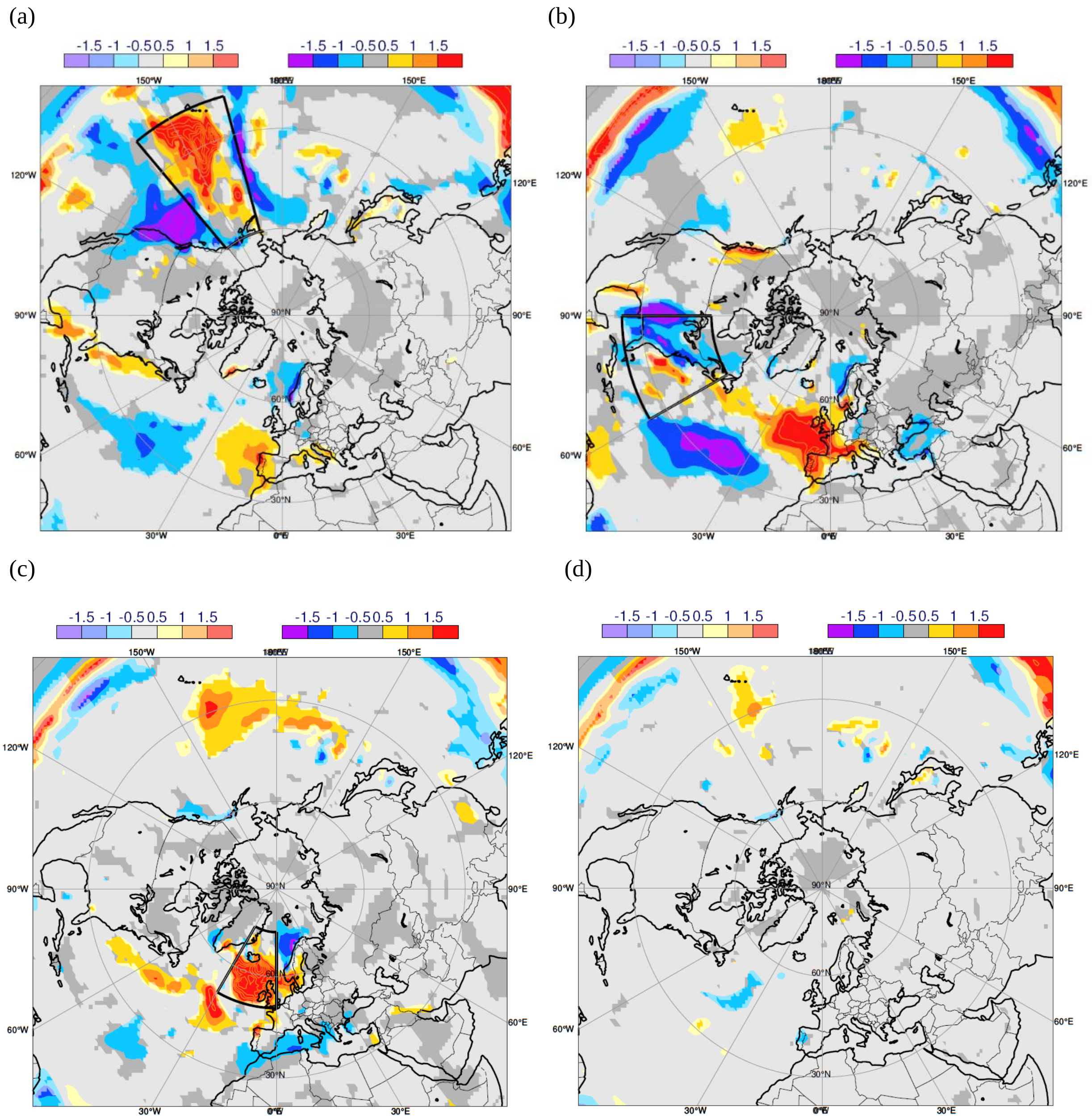
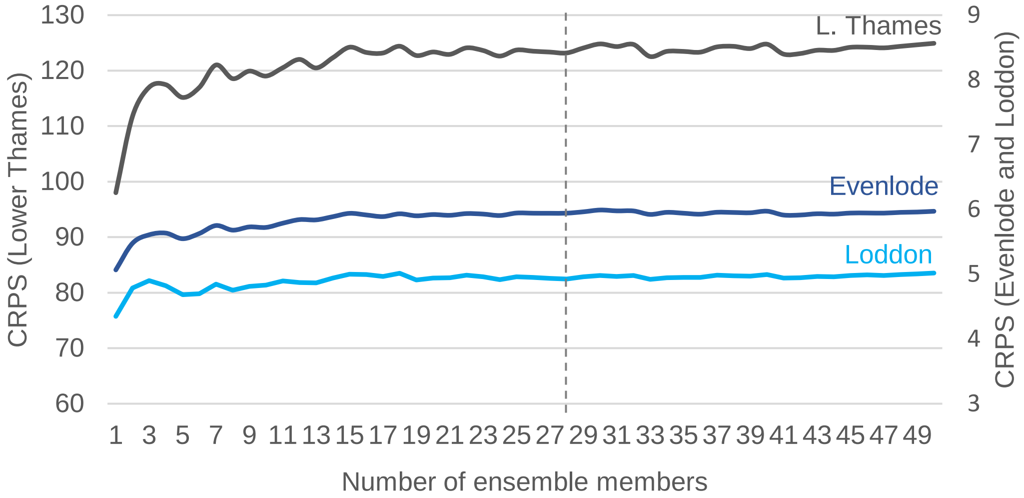


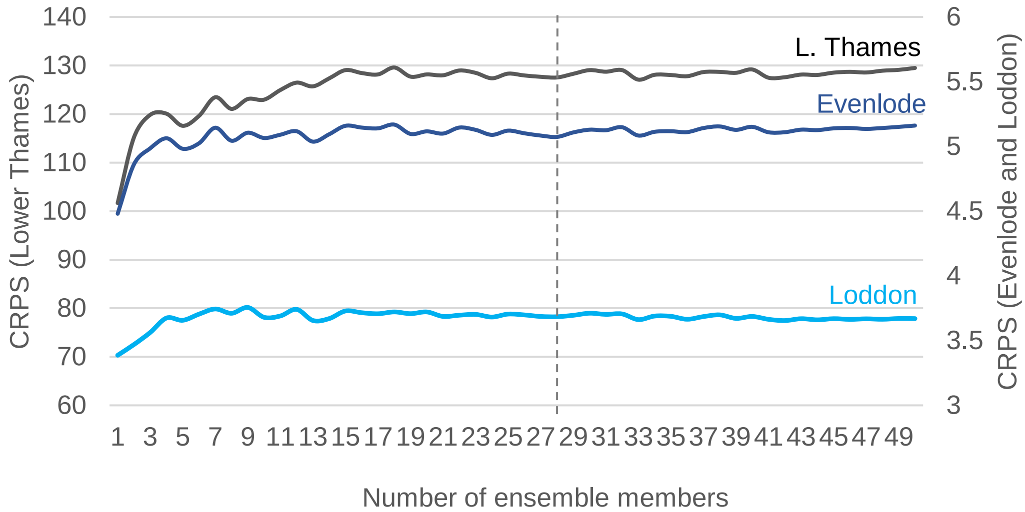
Fig. A1: DJF 2014 anomaly fields of total precipitation (TPO) relative to model climatology for 28-member hindcasts made with the coupled model (*CY40R1*). (a)-(c) Show the equivalent field from ARE where the atmosphere was relaxed towards ERA-Interim reanalyses with Rossby-Wave Source centres identified by black boxes. (a) *AR_NPAC*: centered 150ºW, 35ºN, (b) *AR_WATL*: 75ºW, 35ºN, (c) *AR_EATL*: 15ºW, 55ºN, (d) *NO_AR* (control) equivalent to the operational System 4 forecasts (*CY36R4*), but with the most recent model cycle. Model climatology based on 3 ensemble members, initiated from 1 November for the 30 years 1981-2010. Statistical significance at the 5% level is estimated from the 28-member distribution and indicated here with saturated colours.

(a)



CRPS$_{obs}$	Evenlode	Loddon	Lower Thames
28 mem average	5.066	4.985	123.187
51 mem average	5.066	4.995	124.225
Relative % difference	0.00 %	-0.19 %	-0.83 %

(b)



CRPS$_{sim}$	Evenlode	Loddon	Lower Thames
28 mem average	3.756	4.986	123.346
51 mem average	3.756	4.999	124.271
Relative % difference	0.00 %	-0.25 %	-0.74 %

Fig. A2: Continuous Ranked Probability Scores for System 4 run using 2 to 51 ensemble members. (a) Shows forecast against streamflow observations (CRPS$_{obs}$), (b) shows forecast against simulated EFAS-WB (CRPS$_{sim}$). Tables outline relative percentage difference in CRPS achieved with 28 member and 51 member ensembles.