6 National exemplar

The EU Water Framework Directive has enacted a legislative imperative to ensure all European waters bodies reach 'good ecological status' by 2015. Nationally this requires improvements to modelling hydrological quantity and quality in all water bodies and water resources in a more integrated manner and at the same time assess the uncertainties in these predictions. Assessment of uncertainty will be especially important when conducting scenario testing for quantifying likely environmental change (Cloke et al. 2012). Examples of current models that can be run at national scales include G2G (e.g. Bell et al., 2007a and 2007b), which is being used as the basis for national flood forecasting (e.g. Price et al., 2012) and mass balance mixing models of flow and water quality based on monitoring evidence such as SimCAT (Crabtree et al., 2009). However what these models do not allow for is understanding and testing where competing conceptual structures of processes best-predict stream discharge and quality and if this can be related to catchment characteristics.

The EVO project produced national exemplars for both hydrology and biogeochemistry to explore the ways in which cloud-computing could support the development of an integrated modelling framework to deliver ensemble predictions of hydrological and biogeochemical behaviour in catchments for the whole of the UK, and estimate the uncertainties associated with these predictions.

6.1 Hydrological National Exemplar

The national hydrological modelling conducted for the EVOp project is the first of its kind to test competing models of rainfall-runoff structures nationally, and in a comprehensive uncertainty analysis procedure to quantify predictive uncertainties. With regard to the modelling methodology, the model structural exploratory scheme (FUSE) developed by Clark (e.g. Clark et al., 2008) was used. FUSE is to date the most comprehensive tool available to researchers in the hydrological sciences for the exploration and assessment of both model structural and parameter uncertainties in the manner attempted here.

With regard to the geographical scope of the work, FUSE has been applied at catchment scale with almost complete national coverage; that is to say, nearly all of the regularly monitored catchments brought within the NRFA stream gauging network have been included. This extends the modelling assessment far beyond what has been attempted in previous studies for the U.K (e.g. Bell et al., 2007a and b; Arnell, 2011) and reflect the first national scale benchmarking of predictive capability. Taking these points into account, the research is therefore able to demonstrate results not only for individual catchments but also the patterns of model structural performance and parameter uncertainty that emerge across the whole landmass of mainland Great Britain, with the exception of those near-coastal catchments and adjacent areas for which no stream data were made available.

Key questions of interest are:

- What happens when we join up data nationally?
- What happens when we join up modelling nationally?
- What can this deliver in terms of national capability?

The modelling has therefore been aimed at achieving a national picture of the ability to predict streamflow, with the broadest feasible range - so far as this is possible within the National River Flow Archive (NRFA) gauging network of catchment sizes, locations and characteristics. The catchments exhibit differences in flow control and management, such that some are almost wholly natural, whereas others may be highly modified for regulation (for example, by flow controls during floods, or by abstractions for irrigation). In the UK, many catchments will lie somewhere between these two extremes; they are neither wholly natural nor wholly managed. The list drawn from NRFA records comprises 1,454 stream gauge stations in the UK (Figure 6.1), of which 1,403 are in Great Britain and another 51 in Northern Ireland. Catchment sizes range from ~1 km^2 to 104 km^2, and some catchments are located on offshore islands (the Hebrides, Orkneys and Shetlands).

By including such a wide range of geologies, there is also clear inclusion of the influence of different groundwater and base flow conditions. Thus, many catchments in areas overlying the chalk and other potentially aquiferous rocks will exhibit high base flow

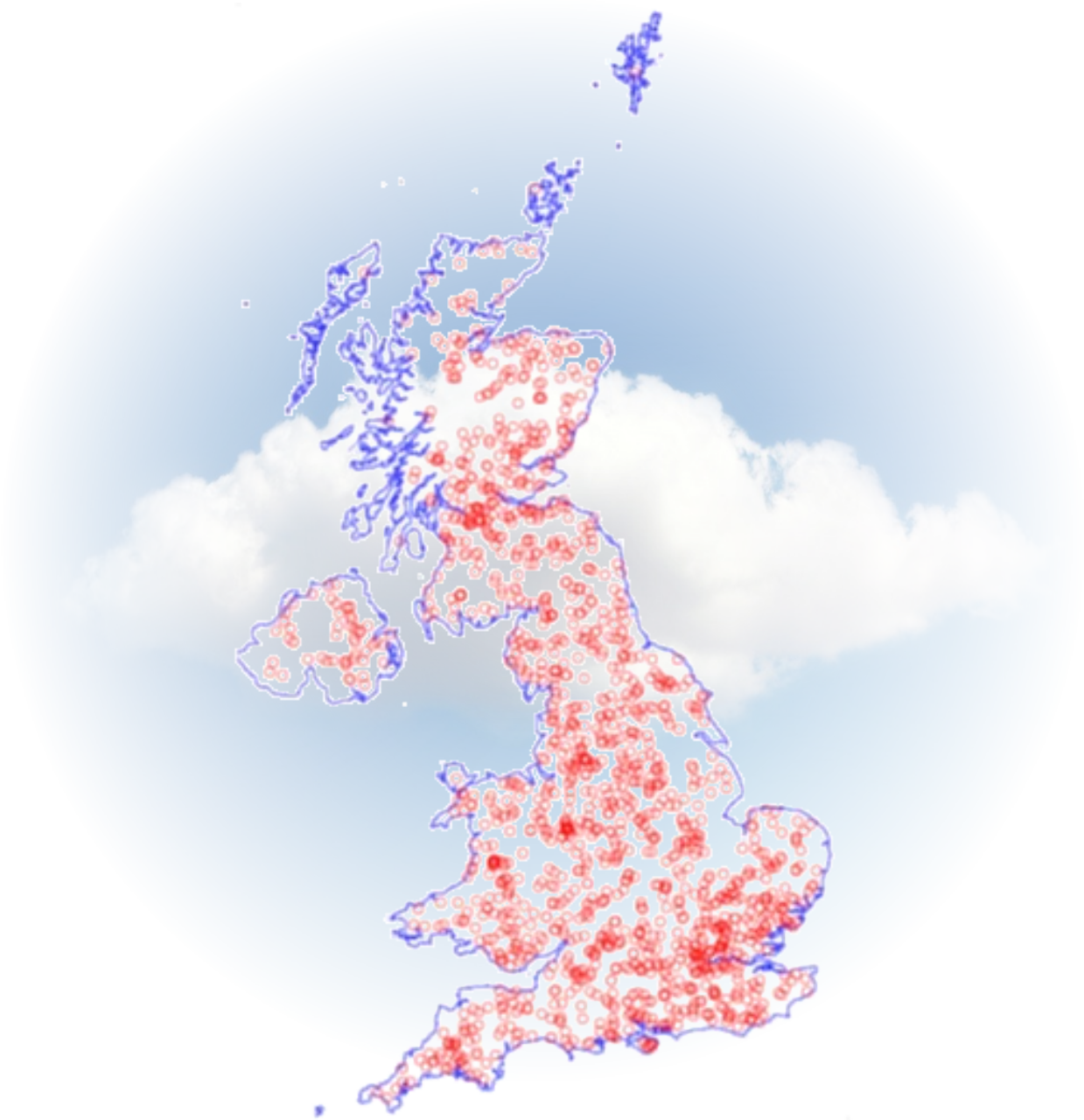


Figure 6.1: location of the 1454 stream gauging stations in the UK for which NRFA stream flow data were provided. The geographical boundary of Northern Ireland is extended south beyond the political border so as to include all of the contributing areas of the cross-border catchments.

indices (BFI) which define the potential contributions of groundwater; similarly, some catchments in these locations may also exhibit major losses to groundwater, such that runoff is much lower than would otherwise be expected from consideration alone of mean annual rainfall and evaporation from surface topographic features.

In addition to achieving the national coverage outlined above, the modelling seeks to achieve the finest spatial and temporal resolutions - with respect to inputs and outputs - commensurate with the time and resources available. A daily time step has thus been applied in all the calibration work, such that rainfall and evaporation inputs, the stream flow data and the model hydrographic outputs, are all expressed in daily intervals. With respect to the spatial resolution of the input data, rainfall and evaporation are sourced from daily values at 1 km^2 resolution and then aggregated for each catchment according to the area cut.

Two further aims in the modelling are related to dealing with calibration and prediction uncertainty. In particular, the research seeks to demonstrate the influence and effect of both by specifically including multiple model structures in the calibrations. In this way, not only are the best (and worst) performing model calibrations derived for each catchment according to particular metrics or combinations thereof and through the application of dense sampling in the parameter space, but these calibrations are achieved using a selection of different hydrological model structures. In this way, the best performing structures can also be identified for each catchment and metric, or metric combination. The research therefore highlights how model structural differences affect

model calibration; it also shows how model structural performance varies across space and by catchment.

6.1.1 Chosen modelling methodology: conception

FUSE has been developed as a tool to aid identification of the most appropriate model structure to use for a particular problem (Clark et al., 2008; Kavetski and Clark, 2010). The scheme as presently coded applies to a set of lumped catchment models.

An important research question in hydrology is to be able to understand what level of complexity is needed in a conceptual model of rainfall runoff processes to provide useful predictions, given imperfect data and knowledge of catchment behaviour. There is no doubt this will vary depending on the questions being asked of a particular modelling application, and so different types and complexities of models may be needed for different purposes. The work focused on identifying the broader patterns of model prediction capability, and hence benchmarking national capability, whilst recognising that model structures are not perfect and where the catchments themselves may be highly modified and lack critical details that might inform why model deficiencies do or do not occur. Hence, the method embraces the potential for model structural error; and this in turn is likely to vary from place to place (Beven, 2000, 2002, 2007). In this respect, a poor calibration may be caused by major incongruities between the physical properties of the system - the catchment being modelled - and the model being applied to simulate it. There also may be situations where a calibration appears to be good even though the underlying structure is a poor representation of the reality. This may provide "the right answer for the wrong reasons". In addition, there may be many ways in which an acceptable answer may be obtained as defined by good predictive capability, so that the model demonstrates a significant degree of model equifinality (Beven, 2006; Beven and Freer, 2001). The aim of the method, within a comprehensive uncertainty analysis procedure is to aid diagnosis of these possible modelling outcomes, and thereby contribute to understanding of the most appropriate structure or structures to use for a particular application. Moreover, a preliminary study using FUSE to evaluate model structures in an uncertainty analysis framework reveals that individual model performance is different in different regions and no single structure is likely to be 'best' across all catchments (Coxon, 2011). The need to assess for multiple model structures is thus clearly warranted from both theoretical and evidential standpoints.

6.1.2 Chosen modelling methodology

The FUSE scheme uses four primary source model structures, all of which have been applied widely and are well respected, these include TOPMODEL, VIC, PRMS and SAC (Clark et al., 2008) (see Figure 6.2). These are all broadly similar in that they incorporate state variables for soil water - in one or more stores - with fluxes which allow the movement of water through the system according to particular process laws. The equations of state and the parameters which are used in them govern the flux rates at any time-step in each model. Solution of the fluxes and updating of state variable quantities is carried out using an implicit Euler scheme, which is considered and demonstrated to be more conceptually and mathematically correct than other schemes typically employed in hydrological modelling (Kavetski and Clark, 2010).

An important feature of FUSE is that the sophistication of the code allows new model structures, here called 'variants', to be formed out of any component of any of the four source models. Thus one variant may comprise components of PRMS and TOPMODEL only, whereas another may use components of TOPMODEL, VIC and SAC, and so on. The mixing of the models' features in this way is not entirely without limit and there are conceptually feasible variants which are not practical to include due to their complexity and the awkward structure of the code needed to run them. Nevertheless, FUSE permits over 1,000 variants to be tested and explored, each with a parameter and equation set controlling fluxes and states in the manner described. One of the most important considerations here is to be able to simulate more than 1,100 catchments for a range of model types and within an uncertainty analysis procedure that allows understanding of the limits of the predictive capability, and to express the uncertainties in the predictions.

6.1.3 Acquisition and preparation of data

In order to conduct the calibrations, the main inputs of rainfall and evaporation need to be prepared for each catchment and calibration time period of interest. Likewise the stream flow data from the NRFA records need preparation, and in particular to be examined for gaps or errors in any record which might make it unsuitable for use. A general rule applied initially in the work conducted - the "80-10 rule" - was that in any calibration period, the stream flow record should be at least 80% complete, and the length of any single, contiguous gap in the record no longer than 10% of the total calibration period, excluding the initial warm up. The application of this rule meant that ~150 of the stream flow records were unusable.

Rainfall and evaporation input data were provided in mm per day. With respect to the rainfall data provided, this is the daily 1k m^2 resolution record prepared by the EA and the Met Office. The methodology for the preparation of which is reported in Keller et al., 2006. To use in FUSE, the daily rainfall for each square kilometre cell is first compiled into a time series for that cell, covering the period from 1st January, 1961 to 31st December, 2008, which is the entire data record period provided. The total for each day for a particular catchment is then summed by selecting the source cells (or part thereof where appropriate) relating to the catchment cut area and aggregating all of each day's total for those cells, and then repeating this for the next day, and then the next, and so on, thus forming an equivalent daily time series for the catchment as a whole.

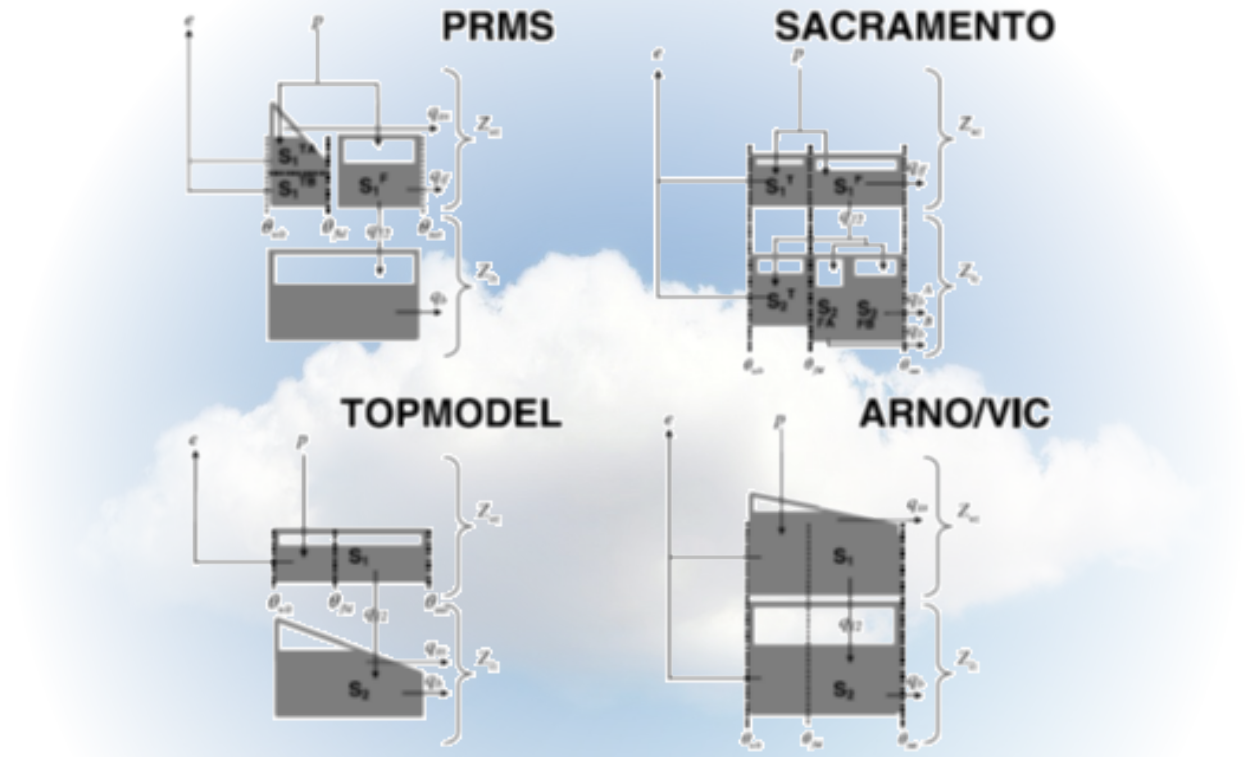


Figure 6.2 Simplified model structure diagrams, adapted from Clark et al., 2008, of each of the four source models used in FUSE.

A similar method is applied to the evaporation product, sourced from the UK MORECS records (Hough and Jones, 1997). However, the MORECS data are first provided at 40 km resolution (thus 1600 km^2 per cell), so the daily time series are first divided into single square kilometres, and these are then aggregated for each cut catchment area in the same way as for the rainfall. It should be noted that throughout the potential evaporation product ("PET") was used from the MORECS data rather than actual evaporation product. The reason being that within each model structure in the FUSE scheme a calculation is made for actual evaporation losses.

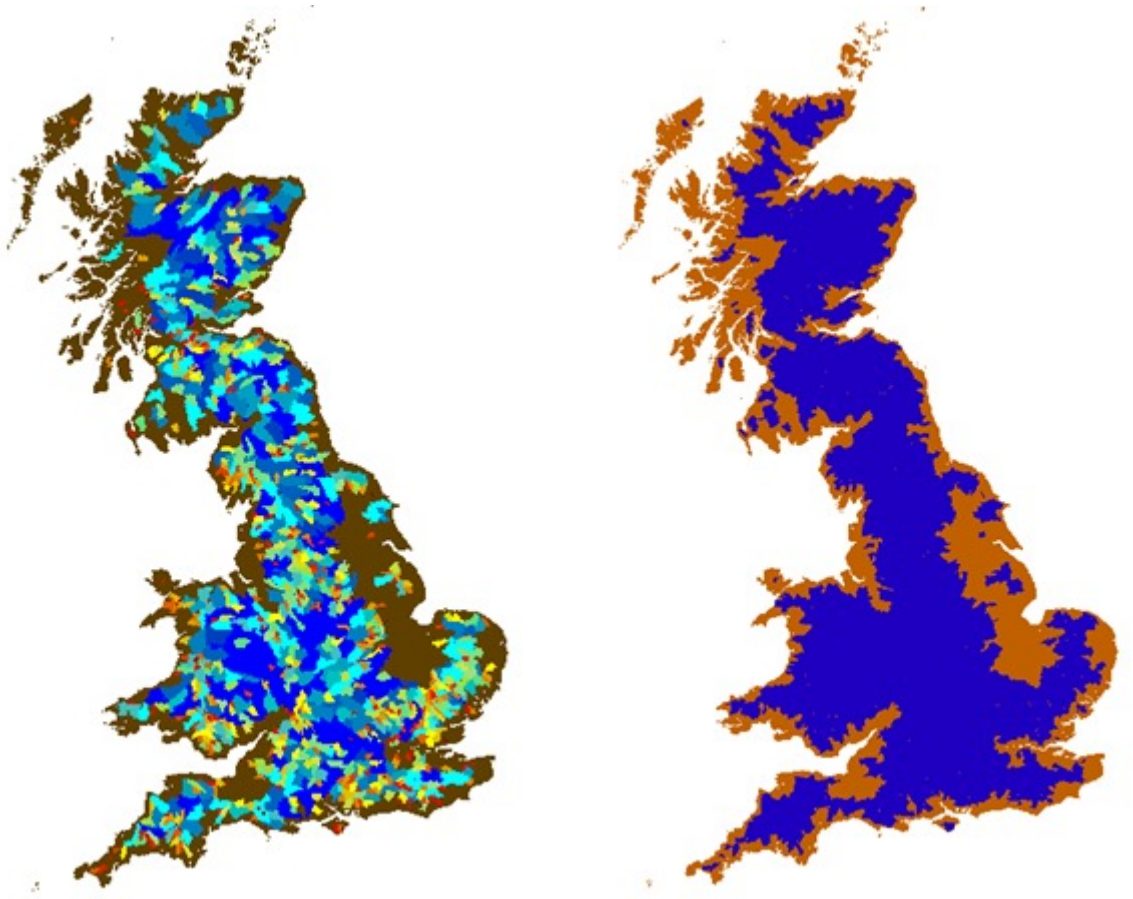
Regarding the cut catchments, it was not possible to obtain a product to use (based on the NRFA catchment outlines dataset), so the areas and outlines were calculated from a 50m resolution DEM of Great Britain, resampled from the 5m NextMap data available from NEODC data sources; a similar DEM was obtained latterly for use to cut out the catchments in Northern Ireland. Of 1,403 catchments aimed to be cut in Great Britain, over 1,250 were cut so as to produce areas within +/- 5% of those on the NRFA database, and a further 100 or so to within +/- 10%. Generally, the areas cut and the shapes obtained were judged as acceptable, and the overall extent of catchment coverage - even after removal of those not usable in the first calibrations because of gaps in the streamflow record - was considered more than adequate to satisfy the project's aim of achieving national hydrological modelling (Figure 6.3).

Regarding the streamflow data, these were all provided in units of mean flow discharge per day, expressed in cubic metres per second. These data were converted to equivalent mm per day, using the areas cut for each catchment, before being used in the FUSE calibrations.

6.1.4 Sampling, model evaluation and model performance metrics

The approach to each set of model simulations follows the same path, namely to begin with a review of the streamflow data to establish the most suitable calibration period. A further restriction made on the calibrations conducted so far is that the usable calibration period should be 10 years long, plus a one year start up period, and should end on or after 31st December, 1998. Thus, the most recent possible calibration period would cover from 1st January, 1999, to 31st December, 2008, with a warm up period from 1st January, 1998; similarly, the oldest acceptable period, based on the same selection method, would cover from 1st January 1989, running to 31st December 1998, with 1997 as the warm up year. These latter requirements were applied so as to ensure the initial set of calibrations conducted extend over comparatively recent periods which may still be considered broadly equivalent with the present in terms of climate, land use, hydrographical response and catchment flow management

Once the streamflow period had been chosen, the rainfall and evaporation data for the same period are also selected from the catchment aggregated datasets



(a) Catchment areas cut

(b) Aggregated area of catchments

Figure 6.3 (a) 1,103 catchment areas cut in mainland Great Britain, based on the 50m DEM, all to within +/-5% of the NRFA listed areas. Although all 1,403 catchments were cut, only those within the +/-5% band have been used, and the potential number is reduced after removing those with incomplete streamflow records. In (a), the smaller catchments are overlaid on the larger so as to preserve detail of the smaller areas cut. In (b), the aggregated area of the catchments cut in (a) is shown, demonstrating that the coverage is national, although there are some areas missing in the lowlands and near the coastal margins. The catchment areas in Northern Ireland are not shown.

in order to begin running the FUSE calibration. Further choices required are which model structural variants to apply in the calibration, and how many points to sample in the model parameter space.

For every catchment's results reported, the decision was made after some preliminary testing and exploration to conduct the first set of calibrations using 2,000 sample points per catchment per model. The total number of sample points per catchment is then simply a multiple of the number for each model structure. Also, the sampling scheme is a space-filling SOBOL scheme, similar to a Latin-Hypercube method, and incorporated in FUSE's general program structure; the use of SOBOL is explained in more detail by Clark et al., 2008.

With respect to the model structures of interest, a core aim within EVOp was to demonstrate the effects of evaluating multiple model structural simulations and uncertainty. This has never been demonstrated before in the UK and is only really possible with more powerful cloud computing type resources.

During the model evaluation of each catchment and each of the four model structures, the predicted discharge was compared with the observed. A set of calibration performance metrics were then calculated and were related to each sampled point in the multi-dimensional parameter space.

With respect to calibration there is a wide range of possible metrics that might be used to quantify the calibration adequacy, and these metrics may in turn be split up in different ways, for example by season (Figure 6.7), to provide more discrimination between one performance metric and another, or one catchment and another. The only metrics commented on within this report are the Nash Sutcliffe index ("NS"), the sum of absolute errors ("SAE"), and the Nash-Sutcliffe index of the logarithm of flows ("NSlog"). The method of calculation of the NS and NSlog indices is shown in the online Annex.

These particular metrics have been chosen because they can be used to indicate how well a model has been calibrated, not only to the overall flow record, but

also with an emphasis on certain flow magnitudes. Thus, the NS index is often found to be particularly good at calibrating for the higher flood flow peaks, whereas the NSlog score is better for evaluating model structures that perform well for low flow periods; the SAE score appears to be a useful measure for assessing general model simulations for the whole period.

With respect to the running of the FUSE scheme, this has been conducted on a high performance computing (HPC) cluster, which is, for the purposes reported here, a closed cloud computing resource. Jobs submitted permit many different catchments to be run at once. In the first main calibration exercise reported here, the run comprised some 8.8 million simulations, and took eight working days to complete on the HPC cluster.

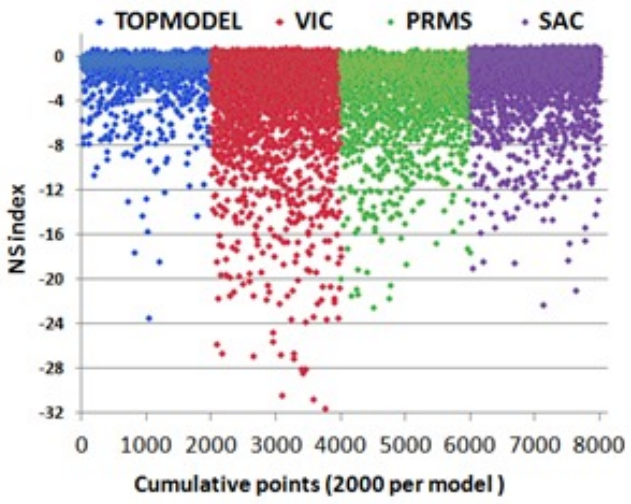
The abstracted output requires further analysis and processing before the data can be presented as usable results for inclusion in scientific literature. Dealing with the SAE metric, the top 5% of results for each model and catchment are considered usable. Although this appears on the face of it to be an arbitrary cut off, it can be considered in an equivalent way to the standard of accepting as significant a probability of 0.05 (i.e. 5%) or lower in a statistical test such as a Student's 't' test. In the same way, the best 5% of NS and NSlog scores are also treated as usable, with the proviso that all such scores must be greater than zero. A feature of NS and NSlog is that the maximum possible score is 1, denoting perfect agreement between the modelled and observed output. Similarly, a score of zero denotes a calibration result that is no better than using the mean flow (or mean of the log flow) for the whole series, which would be of no value in this work. It follows that only scores above zero are used, and these in turn must be within the top 5% of the results (Figure 6.4).

6.1.5 Initial results: examples of spatial presentation and analysis

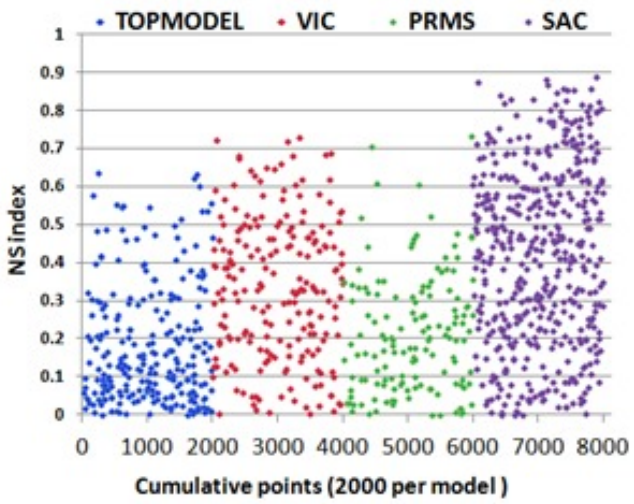
The procedure of catchment selection and calibration was conducted for all of the catchments satisfying the requirements of the most recent 10 year period selection and after application of the 80-10 rule. The abstracted results were analysed, and can then be plotted to begin to benchmark the national picture of predictive capability for all catchments analysed.

Figure 6.4 provides a useful overview of the sorts of results the national hydrological modelling has generated. The results plotted here are from 1,103 catchments, this being the number that satisfied the first calibration period, and the 80-10 rule.

Figure 6.5 shows the best NS score for each catchment and model structure from the Monte Carlo simulations, as explained above. The results clearly



(a) all NS output data across the four models. Most of these calibrations must be discarded because any NS score <0 represents a calibration worse than simply using the mean flow.



(b) the same data as in (a) but only showing the acceptable NS score (>0). The calibrations actually used for EVOp purposes are then a subset of these. Note also how performance differences between models are evident.

Figure 4, (a) and (b) Example of FUSE output, the calibration data here being the Nash-Sutcliffe index ("NS") achieved by each of the four main source models - TOPMODEL, VIC, PRMS and SAC - for the catchment of the River Thames at Kingston, stream gauge no. 39001. To aid comparison between the models, the sample points are plotted cumulatively on the x-axis, 2000 points sampled per model structure. See figure text by (a) and (b) for detail.

indicate differences between the best model simulations achieved with each model. It is striking how in closer analysis, the poor performance of some of the models can be related almost immediately to geology or land use characteristics. For example, both TOPMODEL and PRMS do poorly on areas dominated by chalk. More detailed analysis also shows that in many areas a NS score of 0.8 or higher has been calculated. Given that the inputs of rainfall and evaporation have been lumped, sometimes over quite large areas (up to 104 km^2 in the case of the Thames at Kingston), and the inputs and outputs are also aggregated to daily periods, such high calibrations scores are considered a very promising start; it will be of great interest to see to what extent these results can be improved upon with further work, for example by trialling other model variants.

The NS score is particularly good for calibrating for the higher flows, so it is of interest to compare results where the calibration metric better-matches the lower flows, using the NSlog score. Rather than showing this for each model individually, the results can also be viewed across the whole model ensemble, showing the best result for each catchment regardless of the model structure used (Figure 6.6).

Likewise, inter-seasonal comparisons are also made possible on the same basis, using the ensemble best result for reach catchment (Figure 6.7).

6.1.6 Single objective versus multiple objective calibration performance criteria

In the example results, only one metric at a time has been considered. However, a broader calibration parameter set for each catchment may be provided by considering multiple performance criteria i.e. using metrics in combination to calculate a more generally applicable set of calibrations across the different metrics or models. For example, it is evident from the contrast between high and low flow performance, in Figure 6.6, and the seasonal differences, in Figure 6.7, that trying to obtain an overall best fit for all seasons and flow conditions will require a degree of compromise. Similarly, when including additional metrics (for example NS, NSlog and SAE in combination) the best parameter values generating the best results for each metric individually may not be those that generate an equivalent best value for the other two metrics.

6.1.7 Other considerations and calibration issues

One aspect of the work conducted to date is that no allowance has been made for catchments departing fundamentally from the conceptual structures shown in the FUSE scheme. In particular, although the four main model structures used in FUSE are immediately applicable to a wide range of soils, geologies and climates, there are difficulties when these are applied to catchments where the flows are strongly affected by groundwater, in particular high base flow indices, losses to groundwater or from abstractions to extra-catchment areas. Abstractions and irrigation are likely to affect overall water balance. If severe this is difficult if not impossible to compensate for by parameter value adjustments alone. Similarly, a catchment may gain water from sources beyond its topographic watershed. For example for industrial discharges or domestic outflows sourced from reservoirs well outside the catchment area. In future work, it would be of great interest to see whether model calibration performance can be improved in the catchments most subject to these factors, for example by trying to take into account recorded abstraction and discharge data where these are available. This of course would be greatly simplified in a full EVOp where such data and models were more directly coupled in a more sophisticated framework than can be achieved in this demo pilot.

Another aspect to consider is the reliability of the flow gauge data itself, irrespective of whether there are groundwater or abstraction factors to account for. The modelling here demonstrates clearly that for many

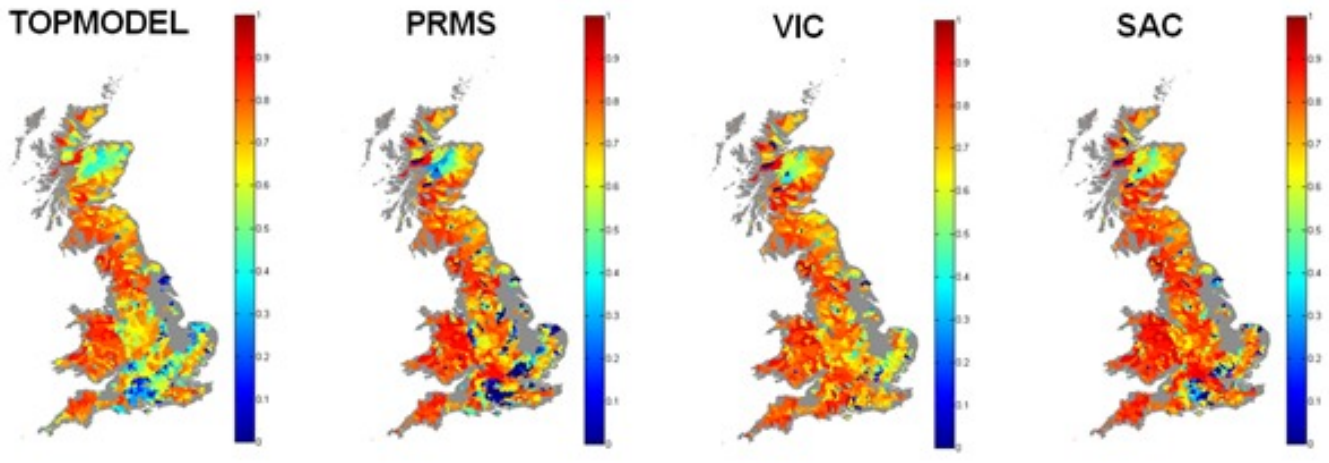


Figure 6.5 Best NS scores for each catchment and model structure; results shown for 1,103 catchments, 2,000 sample points per model structure. A score of 1 is a perfect simulation, scores below 0.6 would not be normally classified as good simulations of high flood flow behaviour.

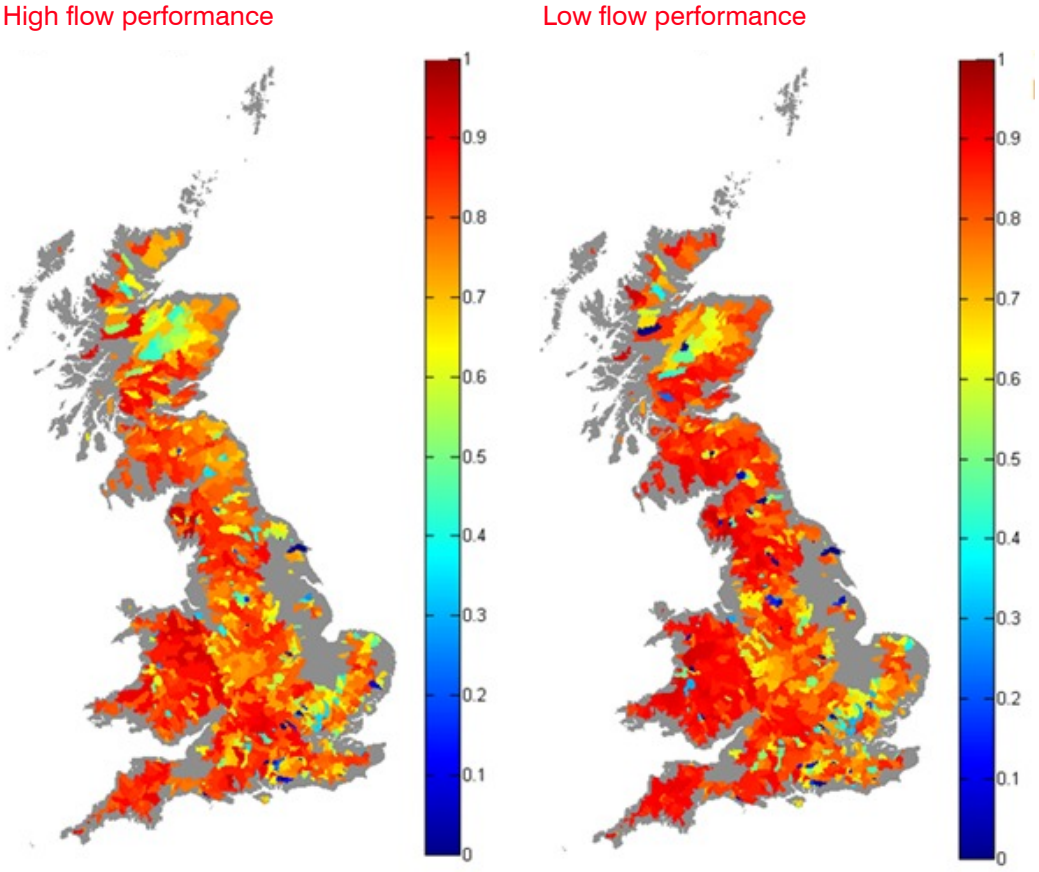


Figure 6: best NS scores contrasted with best NSlog scores, for the entire combined multi-model ensemble and each catchment. These metrics indicate extra detail in the calibration performance, NS being most indicative of a good match of the modelled output with high flows and NSlog of a good match of the model with low flows.

catchments, there is no uniquely 'best' model to use for national hydrological modelling; rather the model structure most appropriate for one catchment appears to differ from that most appropriate for another. However, the veracity of this finding may need to be tempered by assessing the value of the calibration datasets. For many of the stream gauges, the data are likely to be at their most reliable when the flows are within bank. However, once the stream or river is in flood, the discharge becomes more speculative, and during large flood events, where the water is well over bankfull, the stream discharges may be seriously in error (either over- or under-estimated). Work is in hand to undertake a review of stage-discharge relationships at various gauges in the UK, to see how consideration of the uncertainties in the stream gauging may affect overall calibration and model structural uncertainties. This work is beyond the scope of the EVOp, but presents an important and potentially valuable opportunity for further scientific research, and one that would also benefit the usefulness of the EVOp.

6.1.8 Northern Ireland

Although stream flow data were provided for 51 catchments in NI, there is no rainfall or evaporation product, equivalent to those for GB that can be used as inputs for the FUSE modelling. This point aside, the catchment cuts have been prepared for Northern Ireland and these can assist with the biogeochemistry national exemplar. Also, if suitable rainfall and evaporation products become available, calibrations using FUSE could be conducted quickly to augment the national hydrological modelling already completed for Great Britain.

6.1.9 Unique science and demonstrations of the hydrological national exemplar

- The first ever full exploration of the national hydrological modelling capability for greater than 1,100 catchments.
- The first ever national assessment of multiple model structures in a closed cloud computing resource.
- The first ever national comprehensive assessment of model uncertainty analysis to understand parameter and model structure uncertainty and predictive capability.
- The first ever national multi-criteria assessment of model simulation performance that explicitly assesses if models are fit for purposes for high flows, low flows and seasonal responses.
- Improvements to the hydrological modelling predictive capability by using a grand ensemble of model structures.
- Ability to extract model structures and parameter sets that are 'behavioural' for simulating 'tailored'

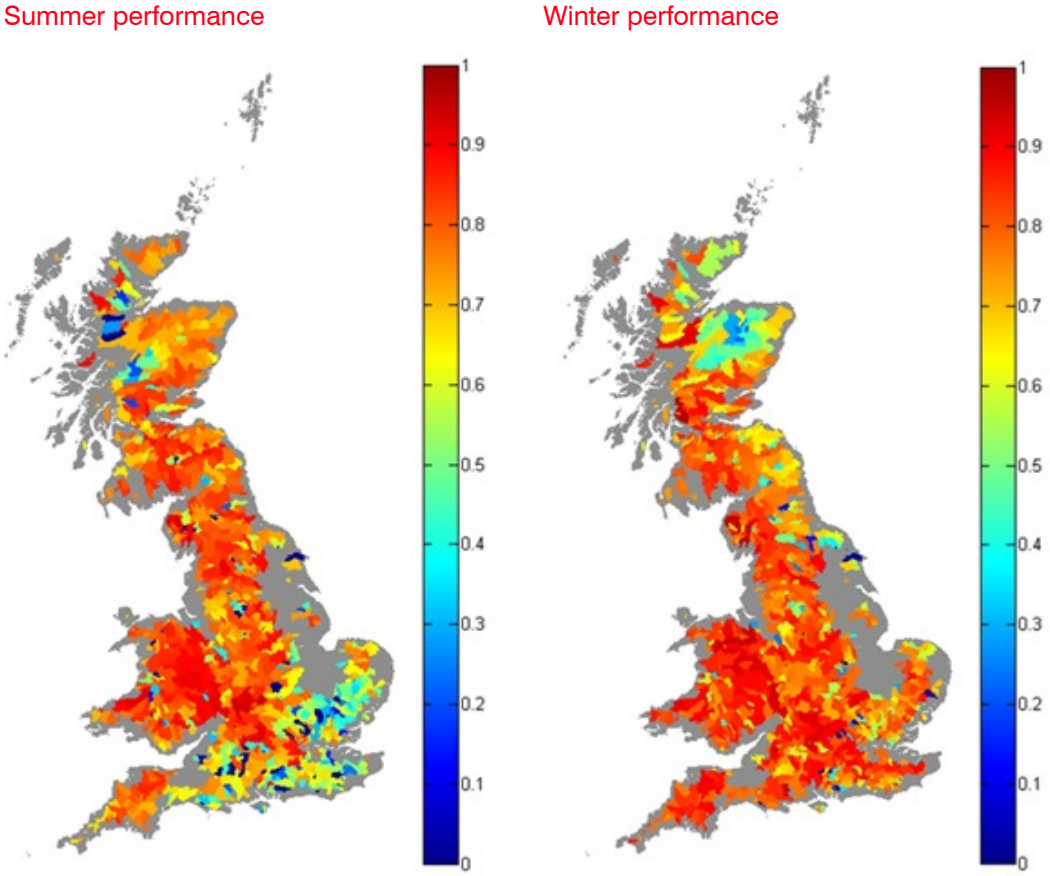


Figure 6.7 Best NS scores for the entire model ensemble for each catchment,, contrasting summer with winter performance. The poorer calibrations in the summer in the south, particularly in areas dominated by the chalk, will be noted; similarly the poorer calibrations in winter in Scotland, over the northern highlands, are striking, and are possibly due to snow melt.

performance metrics for different types of catchment flow responses.

- Identify linkages between model structures and parameters to catchment characteristics than can aid regionalisation and predictions in 'ungauged' catchments.

6.1.10 Potential 'near possible' opportunities due to this research

- More explicit coupled linkages between discharge fluxes and the biogeochemistry components to understand the monthly and seasonal dynamics of water quantify and quality fluxes.
- Improve the diagnostics (i.e. understanding model deficiencies to build better models) of model streamflow predictions by explicitly taking into account the uncertainties in the observed data products.
- Develop the first national flood inundation simulation framework that explicitly included the uncertainties in the predicted upstream boundary conditions. This will allow the inundation uncertainties in given flood return periods to be explicitly quantified
- Further increase the spatial complexity of model structures used in analysis to understand the limits of the predictive capability
- Enhance the understanding of highly modified river systems by including the national EA abstractions licence and time series information for the UK (some 22,000 licence agreements) and improve the predictive capability in these areas and lead towards an improved national water resources model of the UK to assess water security issues
- The ability to improve the assessments of environmental change scenarios on water quality and through the biogeochemistry modelling component the associated water quality.

6.1.11 Barriers

- The project has identified the current difficulty in the UK to bring together all the required observational and catchment characteristics to make this modelling possible. This was a considerable time sync and we now have a framework to make any additional simulations relatively easily to achieve
- Although we have been able to demonstrate a multi-model ensemble that includes a comprehensive assessment of uncertainty for >1,100 catchments without a fully implemented cloud computing resource we are still limited in the demonstration we have conducted.

Within a fully functioning cloud computing resource it would have been possible to -

i. Analyse a range of model spatial complexities rather than just 'lumped' conceptual structures;
ii. Run models for sub-daily input-output simulations and therefore better able to capture more convective storm responses;
iii. Run simulations for greater than 10 years to understand multi-decadal model behaviour and trends;
iv. Simulate multiple scenarios of input rainfall uncertainties and thus improve model diagnostics.

6.1.12 References

Arnell NW. 2011. Uncertainty in the relationship between climate forcing and hydrological response in UK catchments. Hydrology and Earth System Sciences, 15, 897-912.

Bell VA, Kay AL, Jones RG, and Moore RJ. 2007a. 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, and Moore RJ. 2007b. Use of a grid-based hydrological model and regional climate model outputs to assess changing flood risk. International Journal of Climatology, 27, 1657-1671.

Beven K. 2000. Uniqueness of palce and process representations in hydrological modelling. Hydrology and Earth System Sciences, 4(2), 203-213.

Beven K. 2002. Towards a coherent philosophy for modelling the environment. Proceedings of The Royal Society, series A, 458, 2465-2484.

Beven K. 2006. A manifesto for the equifinality thesis. Journal of Hydrology, 320, 18-36.

Beven K. 2007. Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. Hydrology and Earth System Sciences, 11(1), 460-467.

Beven K and Freer J. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249, 11-29.

Beven K and Westerberg I. 2011. On red herrings and real herrings: disinformation and information in hydrological inference. Hydrological Processes, 25, 1676-1680.

Clark MP, Slater AG, Rupp DE, Woods RA, Vrugt JA, Gupta HV, Wagener T, and Hay LE. 2008. Framework for Understandign Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44, article number W00B02, DOI: 10.1029/2007WR006735.

Cloke, H.L., Wetterhall, F., He, Y., Freer, J., and Pappenberger, F. 2012. Modelling climate impact on floods with ensemble climate projections. Quarterly Journal of the Royal Meteorological Society, early view 13 AUG 2012, DOI: 10.1002/qj.1998.

Coxon G. 2011. An Evaluation of Multiple Hydrological Model Hypotheses in the UK using a Framework for Understanding Structural Errors. Unpublished MSc thesis, School of Geographical Sciences, University of Bristol, University Road, Bristol, UK 62 pages.

Crabtree, B., Kelly, S., Green, H., Squibbs, G., Mitchell, G., 2009. Water Framework Directive catchment planning: a case study apportioning loads and assessing environmental benefits of programme of measures. Water Science and Technology, 59(3): 407-416.

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

Kavetski D and Clark MP. 2010. Title: Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. Water Resources Research, 46, article number W10511, DOI: 10.1029/2009WR008896

Keller V, Young AR, Morris D, and Davies H. 2006. Technical Report: Task 1.1: Estimatioin of Precipitation Inputs. Environment Agency R & D Project W6-101 - Continuous Estimation of River Flows (CERF). Main report and appendix, 36 pages.

Price D, Hudson K, Boyce G, Schellekens J, Moore RJ, Clark P, Harrison T, Connolly E, and Pilling C. 2012. Operational use of a grid-based model for flood forecasting. Proceedings of the ICE: Water Management, 165 (2), 65-77.