Evaluating G2G for use in Rapid Response Catchments: Final Report

Report – SC110003/R2

The Environment Agency is the leading public body protecting and improving the environment in England and Wales.

It's our job to make sure that air, land and water are looked after by everyone in today's society, so that tomorrow's generations inherit a cleaner, healthier world.

Our work includes tackling flooding and pollution incidents, reducing industry's impacts on the environment, cleaning up rivers, coastal waters and contaminated land, and improving wildlife habitats.

This report is the result of research commissioned by the Environment Agency's Evidence Directorate and jointly funded by the joint Environment Agency/Defra Flood and Coastal Erosion Risk Management Research and Development Programme and the Scottish Environment Protection Agency.

Published by:
Environment Agency, Horizon House, Deanery Road, Bristol, BS1 9AH
www.environment-agency.gov.uk

This report is printed on Cyclus Print, a 100% recycled stock, which is 100% post consumer waste and is totally chlorine free. Water used is treated and in most cases returned to source in better condition than removed.

Email: fcerm.evidence@environment-agency.gov.uk.

Further copies of this report are available from our publications catalogue: http://publications.environment-agency.gov.uk or our National Customer Contact Centre: T: 08708 506506
E: enquiries@environment-agency.gov.uk.

Author(s):
Cole, S.J., Moore, R.J., Robson, A.J., Mattingley, P.S., Black, K.B., Kay, A.L.

Dissemination Status:
Publicly available

Keywords:
Flood Forecasting, Ensembles, G2G, Rainfall-Runoff, Distributed, RRC, Rapid Response Catchments

Research Contractor:
Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB. Project Manager: Robert Moore, 01491 692262

Environment Agency's Project Manager:
Daniel Hine, Evidence Directorate, Sapphire East, 550 Streetsbrook Road, Solihull, B91 1QT

Project Number:
SC110003

Evidence at the Environment Agency

Evidence underpins the work of the Environment Agency. It provides an up-to-date understanding of the world about us, helps us to develop tools and techniques to monitor and manage our environment as efficiently and effectively as possible. It also helps us to understand how the environment is changing and to identify what the future pressures may be.

The work of the Environment Agency's Evidence Directorate is a key ingredient in the partnership between research, guidance and operations that enables the Environment Agency to protect and restore our environment.

This report was produced by the Scientific and Evidence Services team within Evidence. The team focuses on four main areas of activity:

- **Setting the agenda**, by providing the evidence for decisions;
- **Maintaining scientific credibility**, by ensuring that our programmes and projects are fit for purpose and executed according to international standards;
- **Carrying out research**, either by contracting it out to research organisations and consultancies or by doing it ourselves;
- **Delivering information, advice, tools and techniques**, by making appropriate products available.

Miranda Kavanagh

Director of Evidence

Executive summary

Flood impacts can be severe for rapid response catchments (RRCs). Providing targeted flood warnings is challenging using existing methodologies and on account of the typical absence of river flow gauging. The Pitt Review of the Summer 2007 floods recognised the need for new alert procedures for RRCs able to exploit the new distributed flood forecasting capability being progressed from research into operations. Work on the G2G (Grid-to-Grid) distributed hydrological model was accelerated into operational practice to support 5-day countrywide flood outlooks, a major recommendation of the Pitt Review. The present study aims to explore the potential of G2G to support more frequent and detailed alerts relevant to flood warning in RRCs. Integral to this study is the use of emerging rainfall forecast products, in deterministic and ensemble form, which allow the lead-time of G2G flow forecasts to be extended and given an uncertainty context.

This Report sets down the overall scope of the project, provides an introduction to G2G by way of background and then reports on the outcomes of the R&D study. This includes extensive preparatory work on collating historical datasets to support G2G model assessment, both relating to hydrometry and new rainfall forecast products. A framework is developed for assessing G2G in both simulation-mode and forecast-mode (as a function of lead-time) targeted at the RRC requirement. Relevant to the requirement is the RRC Register of points and areas of interest compiled by the Environment Agency, and the characteristics of RRCs (occurring in isolation or in combination): small catchment area, urban/sub-urban land-cover and steep slopes. The assessment framework is first applied assuming perfect knowledge of rainfall observations for past and future times, so as not to confound the analysis with errors from rainfall forecasts. Variability of performance measures across groups of sites is summarised through box and whisker plots, groups being differentiated on size of catchment area and nature of G2G run (simulation, and with the addition of state updating and flow insertion in turn). Skill scores judge how well the model performs in detecting a flood event exceeding a flow threshold, taken as the median annual flood (as an indicator of bankfull flow exceedance for natural channels) and fractional multipliers of it. The skill scores include POD (Probability of Detection) and FAR (False Alarm Ratio). Performance maps of R^2 Efficiency, indicating the variability in the observations accounted for by the model, are used to portray the spatial variability of G2G accuracy across the country.

G2G performance in small catchments, relevant to the RRC requirement, is best over South West, North East and North West regions; also median performance appears robust from one year to the next. Larger catchments benefit most in forecast-mode from flow insertion, whilst smaller headwater catchments gain particularly from ARMA (AutoRegressive Moving Average) error-prediction. An assessment is made of using deterministic rainfall forecasts from NWP UKV - the Numerical Weather Prediction UK Variable Resolution form of the Met Office Unified Model - in a full emulation of G2G in real-time, and using foreknowledge of rainfall observations as a reference baseline. Forecast quality can deteriorate strongly beyond 12 hours, especially for smaller catchments, whilst for some locations good performance is maintained even for long lead-times. Diagnostic analysis reveals that the UKV rainfall forecasts have patterns of overestimation in some lowland areas (e.g. over London) and leeward of high elevation areas (e.g. north and south Pennines). Overall performance is better in Scotland although there is evidence of UKV overestimating rainfall near the coast at Edinburgh and Elgin in the north.

The assessment framework is extended to include rainfall forecast ensembles and probabilistic flood forecasting, using a combination of case-study and longer-term analyses. Blended Ensemble rainfall forecasts are assessed in two forms: forecasts out to 24 hours updated 4 times a day, and nowcasts out to 7 hours updated every 15 minutes. The 24 hour forecasts generally perform well as input to G2G in the case studies, the G2G flow forecasts typically signalling a flood peak 12 to 18 hours in advance and ahead of any observed response for small catchments. New regional summary map displays of the probability of flow threshold exceedances over a forecast horizon, and for increasing levels of severity, are developed to highlight evolving hotspots of flood risk over time.

The first ever continuous assessment of G2G probability flow forecasts is reported using national maps of probabilistic skill scores - Relative Operating Characteristic (ROC) Skill Score and Brier Skill Score (BSS) - to spatially assess their performance. It is noted that the short periods available for assessment - a 7½ month period over England & Wales and 4 ½ months over Scotland - limit the analyses to low return period flow thresholds. Half the median (2-year) flood is used although a regional pooled analysis allows some assessment up to 5-year. The G2G probability forecast assessed is the probability of the chosen flow threshold being exceeded at any time over the forecast horizon (taken to be 24 hours). Comparison of these scores when applied to deterministic and probabilistic forecasts from G2G provides strong evidence of the value of G2G ensemble forecasts as an indicator of flood risk over Britain. Noticeably poorer performance indicated by the BSS across Scotland is in part attributed to the short, summer-dominated assessment period.

Operational tools available to FFC and the SFFS for using G2G flow ensembles are reviewed and options for improvement identified drawing on the experience and findings of the study. This leads to identifying some work of an operational nature for consideration in Phase 3 of the project. The report closes with a summary of project achievements grouped thematically, a set of recommendations both of a general nature and specific to FFC and SFFS needs, and finally some proposals for consideration under Phase 3 of the G2G for Rapid Response Catchments project.

Some key benefits arising from the project are summarised below.

- Evidence has been produced that shows G2G has good skill in providing strategic forecasts for RRCs. The evidence is stratified by catchment type (area, urbanisation, headwater), form of forecast (simulation or forecast mode) and nature of rainfall input (raingauge, deterministic forecast, ensemble forecast).

- Strong evidence has been presented on the advantage of using an ensemble rainfall forecast as input to G2G to obtain a probabilistic flood forecast for an RRC, relative to an approach where only a single deterministic rainfall and flood forecast is obtained. This indicates better guidance can be given on forecast flood risk for RRCs, improving the level of service provision for such catchments which are currently not well served.

- An improved G2G model configuration, exploiting gauged flows from 912 sites and including new locally calibrated parameters, has been delivered and made operational for the FFC with England & Wales coverage. The benefit is improved operational flood forecast accuracy. For Scotland, an enhanced configuration will be delivered to SFFS in Spring 2014.

- Detailed recommendations on how the visual presentation of G2G ensemble results could be improved are set down in this report. When further developed and implemented, these will prove of benefit to the preparation of Flood Guidance Statements issued by FFC and the SFFS across Britain.

Acknowledgements

The support and guidance of the following current and past members of the Project Board is acknowledged: Daniel Hine, Chrissy Mitchell, Stefan Laeger, Emily Athwal, George Merrick, Tim Harrison (all EA), Graeme Boyce (FFC) and Michael Cranston (SEPA).

Carol Langley (EA), Richard Maxey (SEPA) and Marc van Dijk (Deltares) are all thanked for their work and perseverance in providing the supporting hydrometric datasets for England, Wales and Scotland. Clive Pierce, Stephen Moseley and other colleagues at the Met Office are thanked for their help in providing various rainfall forecast products used in the project.

Contents

1 Introduction

1.1 Background

This document reports on the Environment Agency FCRM R&D Research Project SC110003, entitled “Evaluating and improving G2G flood forecasting model for rapid response catchments”. The project acknowledges the importance of providing flood forecasts for rapid response catchments (RRCs) where flood impacts can be severe and provision of targeted flood warnings is currently problematic using existing methodologies. For brevity, the project will be referred to here as “G2G for RRCs”.

The project recognises that the G2G (Grid-to-Grid) Model, developed especially with flood forecasting for ungauged catchments in mind (SC030227/SR1), is well suited to provide forecasts for rapid response catchments which are typically ungauged. Presently, the G2G operates within the Flood Forecasting Centre (England & Wales) and the Scottish Flood Forecasting Service to provide a “national flood outlook” over the next 5 days, with forecasts updated every 6 hours. It is not routinely run to address the rapid response catchment flood forecasting problem, which requires much more frequent running of the model aligned to frequent telemetry polling of raingauge and river flow data.

Forecasting for rapid response catchments is also more demanding of rainfall forecasts, required as input to the G2G Model to obtain extended lead-time forecasts and, in turn, a timely flood warning capability. In this context, the project aims to assess rainfall forecast products available now and in the future with regard to their utility for use with the G2G Model. Rainfall forecasts of high resolution and in deterministic and ensemble form are required to be assessed, the latter in relation to probabilistic flood forecasting: of particular importance for rapid response catchments during convective storms, with the flooding of Lynmouth and Boscastle being classic examples.

The research study aims to provide evidence to help define the level of service for flood warning in rapid response catchments, made possible through using the G2G Model in conjunction with rainfall predictions for flood forecasting. This evidence will be provided through a performance assessment of G2G flood forecasts for rapid response catchments. The assessment will also be used to guide improvements in the G2G model formulation, targeted at rapid response catchments.

It is noted that the study aligns to recommendations 4, 6 and 34 of the Pitt Review on the Summer 2007 floods.

1.2 Scope of work

1.2.1 G2G model assessment for RRCs

The study aims to first assess the performance of the G2G Model using flows from river gauging stations. Four forms of assessment are to be carried out as follows.

(i) Simulation-mode assessment

The initial assessment is to be carried out in *simulation-mode* using observed rainfall as input, considering both the use of raingauge and weather radar data. This allows attention to be focussed on the ability of the G2G Model to transform rainfall to river flow for rapid response catchments, and to identify any shortcomings in model formulation and configuration for such situations. An infrastructure for model assessment has previously been developed at CEH and was available as a starting point to support this form of assessment.

(ii) Forecast-mode assessment: foreknowledge of observed rainfall

A second form of assessment, still using observed rainfall, is to evaluate the model in *forecast-mode* where flows up to the time the forecast is constructed are used for data assimilation via direct flow insertion, state correction and ARMA error prediction. This emulates the real-time performance of the model and how forecast accuracy falls off with increasing lead-time, whilst avoiding the confounding effects of rainfall forecasts. Work on extending the G2G model assessment infrastructure, through inclusion of forecast lead-time information in the Performance Summary pages in current use, is to be undertaken.

(iii) Forecast-mode assessment: use of deterministic rainfall forecasts

A third form of assessment, using different sources of *rainfall forecasts in deterministic form*, is to be carried out in *forecast-mode* to emulate the operational forecasting situation, or what might be possible when new forecast products become operationally available. The different rainfall forecast products to be considered is discussed in Section 3.4.

(iv) Forecast-mode assessment: use of ensemble rainfall forecasts

A fourth form of assessment, using different sources of *rainfall forecasts in ensemble form*, is to be carried out in *forecast-mode* to emulate the current operational forecasting situation, or what might be possible when new forecast products become operationally available. The different rainfall forecast products to be considered are discussed in Section 3.4.

This builds on ways of assessing ensemble forecast products developed under previous work undertaken by CEH and others for the Environment Agency. The case study datasets will serve as indicative demonstrators whilst the rainfall forecast product datasets over longer periods (months) will be used to perform more meaningful assessments. It also provides an opportunity to assess the potential benefits of using ensemble rainfall forecasts versus deterministic rainfall forecasts.

1.2.2 G2G model improvements

The first form of assessment above aims to identify any shortcomings in model formulation for rapid responding catchments. Simulating river flows by using the G2G model to transform rainfall (and potential evaporation) estimated from observations focuses on how well the runoff production and water routing processes are represented in the model, helping target any model developments that may be required. The study will aim to use any identified shortcomings in model performance to prioritise work on model improvements.

1.2.3 Focus on rapid response catchments

The focus on rapid response catchments is to be achieved in two ways. First, through identifying all gauged catchments that conform to a "rapid response catchment" definition and using these in a national scale assessment (using long periods of observation and forecast product data). An identification procedure for RRCs using a catchment area criteria is an immediate choice to use, and could be extended if considered appropriate.

Second, certain case study catchments are to be identified for more focussed attention, sometimes in relation to specific extreme floods of convective, orographic and frontal origin. The case studies identified initially for consideration aim, in part, to capitalise on past and ongoing work in relation to:

(i) Boscastle (16 August 2004) in the South West,

(ii) Morpeth (Wansbeck, September 2008) and the Calder in the North East,

(iii) the Summer 2007 floods in the Midlands,

(iv) the Cumbria floods (21 November 2009) and Kent and Darwen catchments in the North West, and

(v) the Irvine catchment in Scotland.

During the course of the project, further case studies have been indentified and the project scope extended to include the flood-rich year of 2012, and with a specific focus on the Comrie floods in Scotland (see Section 3.2.2).

1.2.4 Assessment for ungauged RRCs

Whilst the assessment aims rightly to use gauged catchments and their measured river flows as reference, the capability to forecast for the normal ungauged RRC is to be addressed. This is achieved by assuming the gauged catchment to be ungauged and not available for data assimilation (via direct flow insertion, state updating and ARMA error prediction), and performing an assessment in this mode.

1.2.5 Choice of rainfall forecast products to use in assessment

A preliminary review of possible rainfall forecast products to use was carried out as part of pre-project planning, involving discussion with Clive Pierce (Met Office, Exeter) and Nigel Roberts (Met Office, Reading). This focussed on access to archives of new products not yet operationally used by the EA, and products scheduled for release over the next year or two.

A set of deterministic and ensemble forecast products for consideration was prepared, and decisions made on what to extract from Met Office archives, what re-runs were needed to conform with the current version, and what data needed to be supported by a live-feed to CEH Wallingford. The latter was made possible using the communication link that supports the emulation of the EA feed to Hyrad used for Support & Maintenance purposes.

Rainfall forecast products of particular importance are the deterministic NWP forecast at 1.5km over most of the domain (the UKV product) and the Blended Ensemble forecast (the STEPS-2 product). It was decided to provide a live feed of the latter to CEH. Ensemble forecasts for the Morpeth flood (5/6 September 2008) were also re-run. Details are summarised in Section 3.4.

1.2.6 Raingauge and river flow data to support study

CEH have previously developed a national modelling database of raingauge and river flow data to support the calibration and assessment of the G2G Model across England & Wales for the FFC. This spanned 2½ years encompassing the water years 2007/08 and 2008/09. The present project has updated this archive to include data for more recent years, initially to September 2011 in Phase 1.

Automating the transfer of data out of WISKI using scripts (and sharing experience gained recently from SEPA) was initially thought possible, allowing this task to be done much more efficiently than previously. However, it became apparent that the EA had not yet invested in an upgrade to WISKI so that efficient automated transfer of data would not be possible, as it had been in Scotland. This led to many months of painstaking work shared between EA and CEH staff, and supported through Deltares carrying out the level to flow transformation using FEWS (NFFS). It is important that the EA recognise the impact on productivity of this work for CEH (and its own staff) as the CEH cost implications have been met fully by CEH. The Project Board was advised through the Phase 1 Report that this should be borne in mind in future investment priorities with a view to upgrading WISKI sooner than planned. This is now being progressed by the EA and will undoubtedly benefit future projects of this kind.

Under Phase 2, hydrometric data for England & Wales were provided by Deltares from the telemetry NFFS feeds for the years 2011 and 2012. This proved to be a more straightforward data supply and acquisition procedure but with the limitations that (i) data only from sites configured in NFFS were available and (ii) the usual data issues associated with telemetry feeds had to be accommodated. The latter were ameliorated by CEH performing manual and automated quality-control of the river flow and raingauge data respectively.

A similar but longer modelling database with coverage across Scotland had been developed for SEPA shortly before the commencement of Phase 1. This database was updated to September 2012 under Phase 2 of the project and managed efficiently using scripts where possible.

1.3 Project schedule

The work is being carried out in three phases. The Phase 1 report (Environment Agency, 2012) was delivered in November 2012. This Final Report is a development of it, and aims to report on both Phase 1 and 2 that encompass the main R&D activity of the project.

Phase 1

Phase 1 aimed to provide a preliminary assessment of the G2G Model for rapid response catchments and identify work required to improve model performance for such areas. The focus was on running the G2G Model with perfect foreknowledge of future rainfall so as to understand the capability of the hydrological model in isolation from the weather model used for rainfall forecasting. Assessment was restricted to England & Wales coverage in Phase 1 with Scotland included in Phase 2.

Some preliminary investigation of the use of rainfall forecasts, using deterministic high-resolution UKV NWP was carried out to begin to gain an appreciation of the additional warning lead-times that can be obtained. The Phase 1 report made recommendations for G2G Model development for implementation and assessment in Phase 2.

Following review of the Phase 1 Report by the Project Board, the programme of work under Phase 2 was agreed. The scope of work was expanded to include analysis of the notable floods that occurred during Summer 2012. It was agreed to update the G2G hydrometric archive to at least 30 September 2012 for both FFC and SFFS. The extended dataset is of specific benefit to the G2G for RRCs project as it provides the hydrometric data that aligns to the Blended Ensemble rainfall forecasts captured and archived via the live feed to CEH Wallingford. Updating the data holdings within the time-frame of the project was identified in the original Project Plan as an option for consideration.

For reasons of efficiency, a decision was made to update the England & Wales dataset through use of the NFFS telemetry archive, managed by Deltares on behalf of the EA. Additionally, SEPA requested that the Comrie floods of November 2012 be included in Phase 2: they provided a local dataset to support this as it lay outside the period of the national dataset.

The Environment Agency and FFC were requested to provide details of specific periods and areas affected by the Summer 2012 floods and considered of particular interest.

Phase 2

The model improvements resulting from the Phase 1 work were carried through to Phase 2. More detailed analyses were carried out using deterministic and ensemble rainfall forecasts as input to the G2G Model, and procedures implemented for their assessment, including extensions to the Performance Summary template.

The outputs of Phase 2 include this Final Report on the R&D phase of the work. It documents the model improvements and performance assessments as well as giving operational recommendations for flood forecasting in RRCs.

This Final Report includes an evaluation of the G2G model for flood forecasting in rapid response catchments, when used in combination with high-resolution rainfall forecasts in deterministic and ensemble form. It also includes operational recommendations for use by the agencies and how the improved G2G model will be rolled out under Phase 3. Here, it is noted that the Phase 1 improvements to the G2G have already been "rolled-out" to the FFC and are in use operationally on a day-to-day basis. Implications to the level of service that will be provided by the G2G's utilisation in forecasting for rapid response catchments will be set out.

Phase 3

The Project Plan envisaged that an improved configuration and/or version of G2G will result from this study, for delivery to the operational agencies concerned. It is noted here that this was in part achieved under Phase 1. Also that this Phase 3 activity will form part of the S&M activity relating to the CEH Forecast Model Support service to the agencies commissioning the project (EA and SEPA) on behalf of the Flood Forecasting Centre (FFC) and Scottish Flood Forecasting Service (SFFS).

Phase 3 under the plan envisaged delivering, for operational use in the FFC and SFFS, any revised G2G executable, configurations and Performance Summary pages, under S&M arrangements in place with the agencies concerned. Any further recommended operational developments beyond this would be identified and a plan of work proposed. This Final Report, marking the end of Phase 2 of the R&D activity, includes specific proposals as to how Phase 3 should now proceed (Section 7.4) and what operational deliverables have already been achieved (Section 7.1.4).

The project started in January 2012 and was originally planned to run for one year. Delays in data take-on, and an expansion of the scope of work led to a revision of Phase 2 completion to November 2013 for the Final Report, in draft ready for review.

Raising awareness of Project Outcomes is planned to be achieved through a workshop and published paper, or as agreed in discussion with the Project Board.

1.4 Outline of Report

This Introduction has provided the background to the study, outlined the original planned scope of work and subsequent agreed expansions, and set out the project schedule and its revision along with deliverables.

Section 2 provides an outline of the G2G model, supporting datasets and model parameters, and the G2G products (gridded outputs) it can be configured to produce.

Section 3 provides details of the data used in the study covering hydrometric datasets, gridded observed and forecast rainfall products, and information on RRCs. The relevance of the data management activity to ongoing configuration work for FFC is highlighted, bringing important operational improvements with delivery of a new G2G configuration for England & Wales as an outcome of the Phase 1 work.

The methodology for assessment, aligned to the RRC requirement, is outlined in Section 4. This is applied to the current G2G model in Section 5, initially with a focus on simulation-mode performance and then in forecast-mode using both observed and forecast rainfalls as input. Assessments in ensemble-mode are then reported on. Section 6 reviews the operational tools available to FFC and the SFFS for using G2G flow ensembles. It identifies opportunities for improvement in the light of the experience of the project.

Section 7 concludes the report with a summary of achievements, a set of recommendations both of a general nature and specific to FFC and SFFS needs. It closes by setting down proposals with a focus on operational implementation for consideration under Phase 3 of the G2G for RRCs project.

2 The G2G model

2.1 G2G model formulation

The Grid-to-Grid Model, or G2G, is a physical-conceptual distributed hydrological model developed by the Centre for Ecology & Hydrology (CEH) at Wallingford (Bell *et al.* 2009; Moore *et al.*, 2006, 2007). A schematic of the model is presented in Figure 2.1.

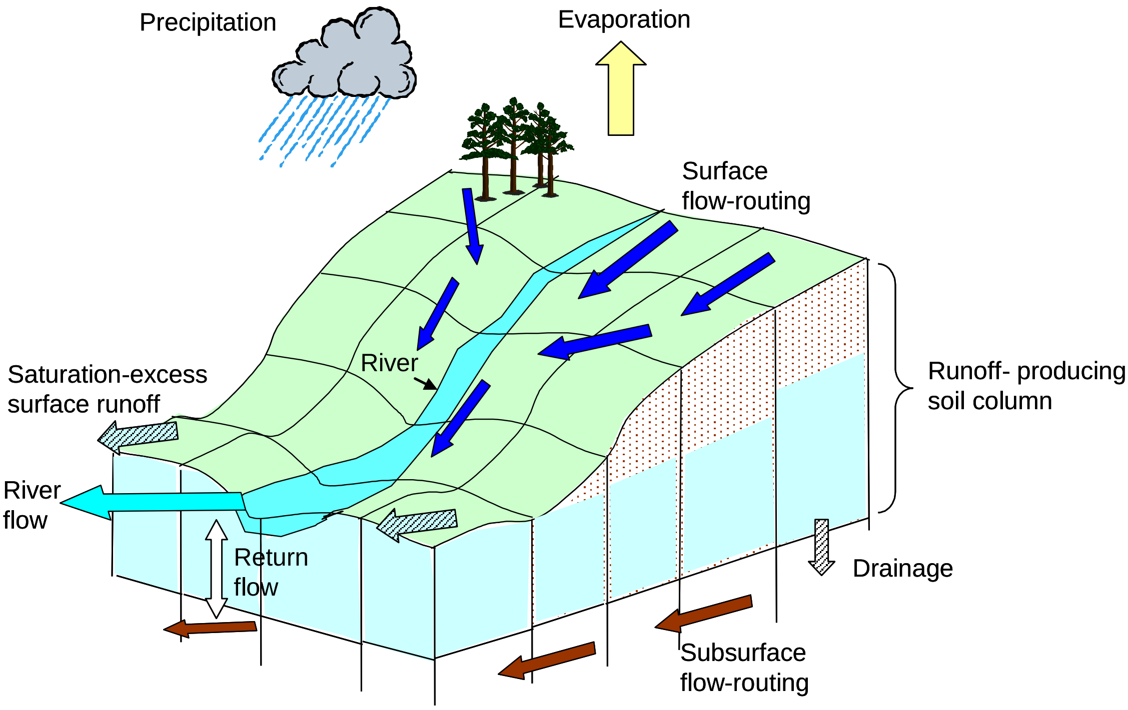


Figure 2.1 Schematic of the Grid-to-Grid Model.

G2G is formulated to represent spatial variability in catchment response and to make full use of spatially-distributed rainfall data derived from networks of radars and raingauges. The model employed operationally is configured to run on a 1 km grid and for a time-step of 15 minutes. Spatial datasets (e.g. terrain, soil/geology, land-cover) are used to support its configuration and parameterisation.

G2G is in operational use as a countrywide flood forecasting system in both the Flood Forecasting Centre across England & Wales (Price *et al.*, 2012) and the Scottish Flood Forecasting Service across Scotland (Cranston *et al.*, 2012). Five day outlook forecasts from G2G are used in preparing the Flood Guidance Statements issued by these operational bodies. At present, it is not configured or run with forecasting for rapid response catchments specifically in mind but its potential to meet this requirement is recognised.

Through adopting an area-wide formulation, in contrast to a catchment-based one, the G2G model is well suited to support forecasting at any set of locations within a defined area. As a consequence, the G2G is able to be calibrated to groups of gauged locations over the model domain and forecasts extracted for any ungauged location within the same area. It will also support modelling of nested and parallel catchments. Consequently, the G2G Model provides a flexible and natural approach to a range of gauged and ungauged flood forecasting problems, including use for flood warning in rapid response catchments.

The model employs a simple *runoff production scheme* to derive surface and sub-surface runoffs from gridded rainfall and potential evaporation inputs. The water holding capacity of each model grid-square controls runoff production and is specified using soil property and terrain slope data. Variation of this capacity from point to point within a grid-square is represented in a probability-distributed way. The processes of lateral and vertical drainage through the soil are represented and specified through soil property data. Land-cover data are used to modify runoff response in areas of urban and suburban development. A groundwater storage receives water via soil percolation (recharge) and releases groundwater flow (sub-surface runoff), as a function of water stored, for subsequent subsurface flow routing.

The G2G *water routing* component offers a choice of nonlinear and linear kinematic wave formulations. Surface and sub-surface runoffs are routed via parallel fast and slow response pathways linked by a return flow component representing stream-soil-aquifer interactions. Water is routed from grid-cell to grid-cell, with the terrain-following flow paths being configured using a digital terrain model. The nonlinear storage routing formulation, allowing conveyance to be related to channel properties (slope, width, length and roughness) through use of Manning's equation, is invoked for river reaches; it is also used for groundwater routing of sub-surface runoffs. The simple linear formulation, equivalent in conceptualisation to a network cascade of linear reservoirs, is currently used for hillslope routing of surface runoff. Floodplain storage is not explicitly represented at present. However, its effects on river flow can in part be invoked through the conveyance formulation, most readily through automated local calibration for gauged river reaches (see below).

In contrast to more complex physics-based distributed hydrological models, the physical-conceptual form of the G2G Model employs simple depth-integrated formulations of runoff production and flow routing. This means that it is computationally efficient, and therefore fast to run for nationwide real-time flood forecasting on a 1 km model grid. Because the model formulation allows model properties to be linked directly to spatial datasets on terrain, soil/geology and land-cover, only a few parameters require to be calibrated across the model domain. It is possible to *locally calibrate parameters* affecting channel flow routing (flow conductance) and return flow fraction for gauged river reaches.

Direct flow insertion allows observed flows, available up to the time the G2G model is run ("time-now"), to be used instead of modelled flows to improve forecast performance at locations downstream of river gauging stations. A simple *empirical state-correction scheme* for the G2G model is provided for forecast updating in real-time. This scheme adjusts the model states, using observed river flows up to the time of the forecast creation ('time-now'), as a way of improving the accuracy of the flood forecasts. Model simulation errors up to time now can be used to forecast future errors using an ARMA error predictor and in turn produce an internally updated flow forecast for each gauged river location.

To accommodate the effects of *artificial influences*, such as river abstractions/discharges and reservoirs/lakes, simple functionality is provided to set a constant flow value (negative or positive), to apply an annual profile of monthly flows and to represent a damped response using a conceptual storage with a rate-constant parameter. Direct insertion of flow up to time-now can be used for a gauging station that measures the outflow from a reservoir or lake; future flows can be set to the last observed flow or the functionality above invoked.

The *G2G Snow Hydrology component* uses screen-level air temperature and precipitation from a weather model as additional inputs. It is used to simulate snowpack formation, melt, storage and drainage release to the land surface within each G2G model grid-square. Key features include (i) handling of rain-on-snow, (ii) elevation-dependent rain/snow discrimination and rate of melting, and (iii) pack "ripening" controlling the rate of release of water. This component operates only over Scotland at present (a weather model estimate of snowmelt is used over England & Wales).

The G2G model has evolved as a *toolkit of modelling components* representing the runoff production and flow routing processes, and different options are available to represent a range of hydrological behaviours. A modular design allows new formulations to be added with relative ease.

2.2 G2G datasets

G2G employs gridded time-series datasets as model input, spatial datasets that support its configuration to a given model domain, and model parameters that control how the model responds. The G2G products (the G2G output grids) that can be output are also configurable. These datasets are outlined in this sub-section, to provide further background to the G2G model.

2.2.1 Gridded time-series datasets

The G2G hydrological model requires use of gridded time-series data e.g. grids of rainfall as 15 min totals, and potential evaporation hourly totals and (for snowmelt) air temperature at screen height as values every hour (alternatives are possible). To manage the large amounts of data involved, the software has been set up to be able to read from and write to a specialist database known as the Spatial Image DataBase (SIDB). The model software reads these gridded time-series and model states information from the SIDB, and optionally writes states and output data to the SIDB.

Gridded observation-based rainfall estimates are obtained either by adjusting the radar rainfalls using telemetry-raingauge data or by interpolation of the raingauge network data. This requires

- the radar rainfall data to be loaded onto the SIDB,
- pre-processing of the telemetry-raingauge data onto a local database,
- estimation of raingauge-adjusted radar rainfall or raingauge-only rainfall using HyradK,
- storage of the new spatial rainfall estimates onto the SIDB using HyradK.

It is also possible to use the radar rainfall data without adjustment within G2G (step 1 of the above list). The choice of which rainfall estimate to feed into G2G will depend on the availability and quality of the radar rainfall and telemetry-raingauge network sources for observation data. Also what forecast (e.g. weather model) products are available for future times.

2.2.2 Spatial Configuration Datasets

Grids of spatial configuration data are used by the G2G hydrological model. These include flow direction, slope and accumulated area, main river length and slope, land/river designation, standard annual average rainfall (SAAR), accumulated SAAR, land-cover (urban and suburban percentage), HOST dominant soil class, and geology permeability. Currently the model expects the grids to be provided on a 1km cell size.

The 1 km flow, slope and accumulated area DTM grids have been derived from the IHDTM (Integrated Hydrological Digital Terrain Model) based on UK Ordnance Survey 1:50000 data. The river information has additionally used a digital 1:50000 river network.

The 29 HOST soil classes are associated, through a look-up table, with the following soil properties: water content (field capacity and residual values), depth and porosity, and saturated hydraulic conductivity.

2.2.3 G2G model parameters

G2G is a parameter efficient model achieved by using spatial datasets to control how model properties vary across model cells. The table below summaries the main “regional” parameters applied across the model domain. The number is less than indicated as this table encompasses different options for some components: for example, the wave speeds and non-linear routing parameters relate to two different routing formulations; also the geology dependent parameters are optional.

Table 2.1 G2G model parameters.

Parameter		Units	Value
Runoff generation			
λ	Drainage conductivity multiplier	-	$7.0\text{x}10^{-5}$
γ	Lateral drainage scaling factor	none	40
Wave Speeds			
Surface land, c_l		ms^{-1}	0.1
Surface river, c_r		ms^{-1}	0.0015
Urban factor, f_u		none	2.0
Surface river bankfull, c_{bank}		ms^{-1}	0.3
k_g	horizontal hydraulic conductivity	h^{-1} mm^2	$2.0\text{x}10^{-7}$
Non Linear Routing			
Inverse of river roughness (conductance)		none	0.9
Nonlinear exponent m		none	2.0
Return Flow Fraction			
Soil store, r_l (or r_i)		none	0.00001
Subsurface routing store, r_r (or r_b)		none	0.00001
Return flow (r_l) by 29 HOST types		none	0.00001
Parameters by geology type (4 types)			
k_g	horizontal hydraulic conductivity	h^{-1} mm^2	$1.0\text{x}10^{-7}$
r_r	Subsurface return flow fraction	none	0.00001
λ_g	Drainage conductivity multiplier	none	0.00001

Additional parameters are involved to control the slow model **state-correction** in real-time, as indicated in the table below.

Table 2.2 G2G model state-correction parameters and example values.

State-correction parameter		Units	Value
Non-linear routing			
g_d	Max % change in soil store per day	%	1.0
L_T	Threshold soil depth	cm	40
f_T	Change tapering factor	none	30

A further set of parameters are associated with the **G2G Snow Hydrology** scheme, as indicated in the table below.

Table 2.3 G2G Snow Hydrology parameters and example values.

Parameter	Meaning	Typical value	Units
c	Precipitation representativeness factor	1	dimensionless
T_s	Temperature threshold below which precipitation is snow	1	°C
T_m	Critical temperature above which melt occurs	0	°C
f	Melt factor	4	mm/day/°C
k_1	Storage time constant: lower outlet	0.15	day^{-1}
k_2	Storage time constant: upper outlet	0.85	day^{-1}
S_c^*	Maximum liquid water content, as a proportion of total	0.04	dimensionless
T_c	Critical temperature below which no drainage occurs	0	°C

Options to accommodate **artificial influences** (abstractions and lake-reservoir damping) are associated with further configuration and parameter settings.

Flow updating by **direct flow insertion** and **ARMA error-prediction** involve further configuration and parameter settings.

2.2.4 G2G Output Grids

G2G produces optional time-series of grids as outputs for the following quantities:

River flow (m^3s^{-1})
Subsurface flow (m^3s^{-1})
Surface runoff (mm/15min)

Subsurface runoff (mm/15min)
Surface (routing) store water content (mm)
Subsurface (routing) store water content (mm)
Soil moisture deficit (mm)
Soil percent saturation (%)
Snowpack dry store water depth (mm)
Snowpack wet store water depth (mm)

The frequency of output is configurable.

2.3 Closing remarks

This Section has provided background to the G2G hydrological model with regard to its physical-conceptual formulation, the gridded time-series of rainfall and potential evaporation used as input, and the spatial datasets on terrain, soil/geology and land cover that support its configuration across modelled domains (England & Wales for FFC and Scotland for SFFS). An overview of the model parameters serve to provide some further insight into the model formulation, including those used for state correction and for the G2G Snow Hydrology component employed in G2G for Scotland.

This outline has aimed to provide a context for the G2G model assessments that follow, which are targeted at investigating the potential value of G2G for forecasting floods in rapid response catchments.

3 Data used in the study

3.1 Background

Hydrometric records of good quality and historical length are essential to support development, calibration, assessment and operational configuration of the G2G model. Also, understanding the form of river flow data available within the operational flood forecasting systems (NFFS and FEWS Scotland) is required to ensure appropriate configuration of G2G.

As a general rule, the forecast performance of G2G increases as the number of (reliable) flow sites are increased and configured for flow insertion, state updating and ARMA error-prediction: so there is a real incentive to configure more sites where practical. However, the same source of flow data and model configuration must be used for both real-time flood forecasting and offline model calibration, assessment and derivation of the Q(T) (flow Q of return period T years) grids used to assess G2G model output.

For England & Wales, the first nationally representative datasets were provided under the EA/Defra T46 “Hydrological modelling using convective scale rainfall modelling” (Environment Agency, 2010): data from 76 river gauging stations and 735 raingauges were extracted mostly from the NFFS archives for the years 2007 and 2008. The NFFS real-time telemetry archives were primarily used as data from them were readily available from Deltares. However, these archives suffer from telemetry outages and data not being subject to Environment Agency quality control processes.

Following the formation of the Flood Forecasting Centre in April 2009, a follow-on project implemented the G2G operationally within NFFS-FFC. As part of this the first truly national datasets were taken on for England & Wales. All available river flow and tipping-bucket raingauge data were extracted from WISKI for the period January 2007 to September 2009 which encompassed two water years. WISKI was chosen as it should be a more complete dataset that has benefited from some quality control. The data take-on took longer as the Environment Agency do not currently have functionality for batch exporting from WISKI. This project culminated in the current version (v1.3.001) of G2G being released to FFC in June 2011 and configured to use 708 river gauging stations.

Since 2011 there has been an ongoing activity between CEH and FFC to review which sites are in the FFC configuration and what they should be used for (flow insertion, state updating, ARMA error-prediction, reservoir configuration, etc.). This work has now been completed, but not yet implemented operationally. In scoping the work for the G2G for RRCs project, it also came to light that flow data for 269 of the 708 sites were not being used by G2G because either

(i) the WISKI rating (level to flow) used to supply river flow data to CEH had not been configured into NFFS-FFC, or

(ii) telemetered river flows (e.g. ultrasonic) were deemed not reliable enough for real-time use, whilst the logged WISKI flows were suitable for offline work.

This prompted the FFC to get additional ratings configured where available and review whether observed telemetered river flows could be used.

It also came to light that there were almost 400 other sites in the NFFS-FFC configuration that had an NFFS rating configured but no river flow data were available from WISKI. In consequence, these had never been considered for use by G2G. As river flow data from these additional sites would be of benefit to the “G2G for RRCs” project and to the G2G operational configuration, these data were requested. This activity resulted in an improved operational configuration for G2G that was an early operational benefit resulting from Phase 1 of the project. More details are given in Section 4.1

For SEPA and the Scottish Flood Forecasting Service, a “G2G for Scotland” project was carried out over 2010-2012 which included development and operational delivery of the G2G Snowmelt Hydrology module (a development of the CEH PACK model already used by the EA and SEPA in their regional and scheme-based model networks). River flow and raingauge data were extracted from WISKI for the period 1 January 2007 until at least October 2010. This was simpler than the EA data take-on, partly because there are fewer sites and regions but also because SEPA have WISKI Service Provider that allows consistent batch export of data.

3.2 Hydrometric data take-on

3.2.1 Phase 1

In Phase 1 of “G2G for RRCs”, the new hydrometric data take-on focussed on England & Wales up to the end of the 2010/11 water year. This is because these datasets were most in need of updating, ending in September 2009, and that there was the opportunity to consider the additional ~400 sites that did not have flow data available in WISKI but do have a rating in NFFS. It was anticipated that inclusion of these additional assessments sites would significantly benefit the “G2G for RRCs” project and its findings.

Also it was believed that the Phase 1 findings for England & Wales will provide some insight into the general G2G performance in RRCs that are applicable to Scotland. Phase 2 provided an opportunity to assess any Scotland specific issues: see Section 3.2.2.

Under Phase 1,the following data were requested from the EA WISKI archives:

(i) **River Level data (15 minute) for the period 1 January 2007 to 30 September 2011 extraction** were requested in EA xml format for sites that have ratings within NFFS. Occasionally the NFFS ratings differ from the WISKI ratings so this accurately reflects what is used operationally. These were converted to river flow by Deltares using the operational NFFS configuration (note the latest operational ratings are used for the entire period).

(ii) **River Flow data (15 minute) for the period 1 January 2009 to 30 September 2011** were requested for all sites in WISKI (766 were delivered). This was to update the existing G2G data holdings, including an allowance for the period January to September 2009 to have been updated by subsequent EA quality control.

(iii) **Tipping-Bucket Raingauge data for the period 1 January 2009 to 30 September 2011** were requested for all sites. Time-of-tip data were preferred but for some gauges only 15 minute totals were stored. Data for 1038 raingauges were delivered and their distribution given in Figure 3.1.

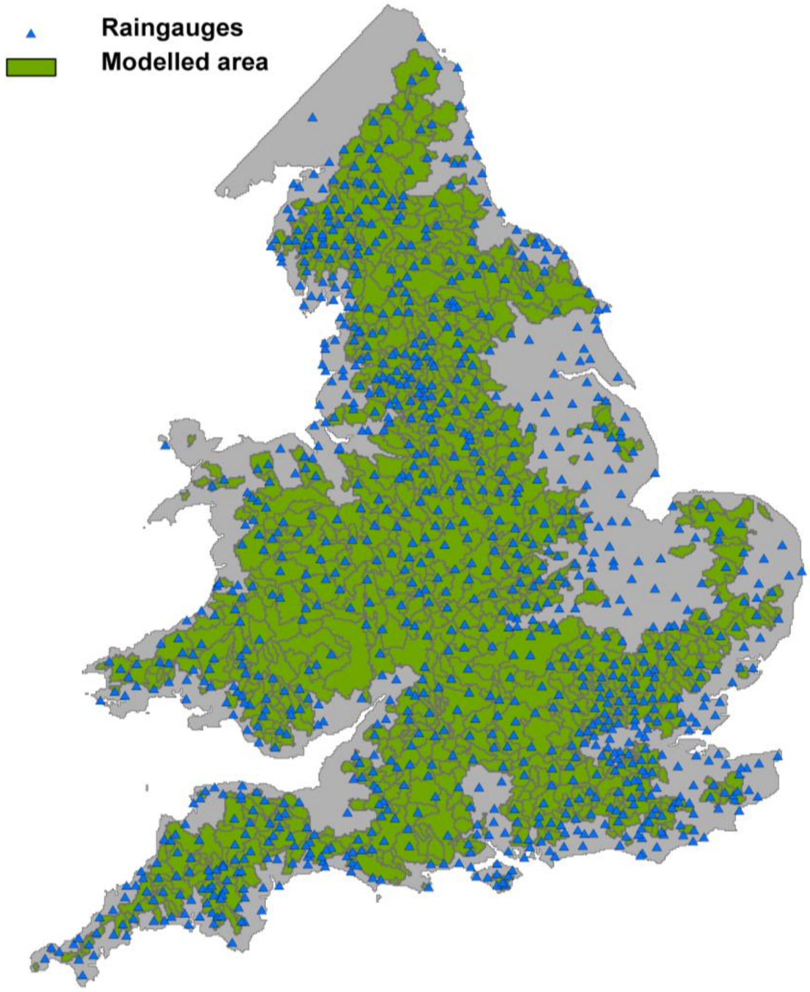


Figure 3.1 Map showing distribution of the 1038 raingauges over England & Wales and the G2G gauged modelled area.

To allow an opportunity for the Environment Agency data quality-control to take place for the most recent period of record, data were delivered in two stages: (i) up to September 2010 and (ii) October 2010 to September 2011.

Reconciling the two new data deliveries with each other and the previous WISKI data was not straightforward as many sites had changed WISKI ID and/or name. Also there were many issues such as files supplied in the wrong format, incomplete or corrupt files, wrong data periods extracted, multiple time-series delivered for some sites and some sites being delivered in one delivery and not another. It is hoped that the future WISKI 7 upgrade will allow repeatable batch exports to take place so as to minimise these difficulties: these have taken an inordinate amount of additional CEH and EA resource to resolve.

For the NFFS derived flows there were several extra checks and troubleshooting required by CEH, Deltares and the Environment Agency. The initial plan was that the WISKI extracted xml flows would replicate the telemetry feeds received by NFFS. Unfortunately there were several sites where m AOD (metres Above Ordnance Datum) were used in one case and m ALD (metres Above Local Datum) used in the other, but these were easily fixed.

Before converting the levels to flows, the NFFS-FFC rating configuration was updated by Deltares to include WISKI ratings for sites which didn't have them in NFFS (see

Section 3.1) and also included any regional rating configuration changes as at the start of 2012. For some sites the new NFFS derived flows overlapped the previously used WISKI flows and where they didn't agree clarification about the differences were sought from the regions. This and other checks identified several rating curves that were configured wrongly in NFFS for both the FFC and the Regions. The relevant EA contacts have been informed so that the necessary changes to NFFS can be made. Thus attention to detail in this data take-on activity is leading to improvements to the operational NFFS regional systems as well as NFFS-FFC.

This thorough analysis also revealed that some regions have special rules for generating observed flow series used for modelling but these hadn't been pulled through to NFFS-FFC: for example, the gauging station on the Thames at Sutton Courtney switches between ultrasonic measurement and a rating curve depending on certain rules. Following consultation with the regions, all special flow calculations have now been included in NFFS-FFC. Also, the G2G NFFS configuration files have now been improved to include extra information on what type of flow is sent to G2G (e.g. rated flows, directly observed flows or special rule) and information on the type of rating used (e.g. power law, Crump weir). Going forward this should help understand the flow calculations used in NFFS-FFC and make future updates easier.

It is worthwhile noting that the NFFS-FFC configuration uses the latest operational rating curves (as at the start of 2012) for the whole study period and does not allow for ratings to change in time. This is only a concern if a different operational rating is applicable to earlier parts of the period (e.g. pre-flood scheme or gauging station renovations). It is believed that this will only affect a small number of sites so should not adversely affect the overall assessment.

In addition, catchment areas were requested for the ~400 new river gauging stations whose flows had not previously been used in G2G. This allowed appropriate G2G pixels to be located for outputting G2G flows for assessment against the river flow records.

More details about the river flow data checking and G2G configuration are given in Section 4.1.

3.2.2 Phase 2

On completion of Phase 1 it was agreed that the G2G hydrometric archives would be updated to at least 30 September 2012 for both FFC and SFFS. This ensures both coverage of events in Summer 2012 events and overlap with the Blended ensemble datafeed starting in May 2012.

England & Wales

Taking account of the data take-on issues in Phase 1, it was decided to use the NFFS telemetry archive, managed by Deltares on behalf of the EA, as being straightforward, relatively quick and in a standard format. A drawback is that this archive will lack the continuity of data (due to real-time telemetry issues) and the quality-control provided by the WISKI archive, although CEH apply some manual and automatic quality-control to ameliorate this. Also, it will not include any of the additional assessment river gauging station (or raingauge) data that are not within NFFS.

The G2G NFFS-FFC configuration developed and delivered under Phase 1 (rolled-out for FFC operational use in March 2013 and discussed further in Section 4.1) was used to generate the river flow and raingauge data. Data across England & Wales for the calendar years 2011 and 2012 were provided from 912 river gauging stations and 975 raingauges. A period of 3 months near the end of 2012 was missing in the original river flow data delivery but this was quickly resupplied.

All raingauge data were supplied as 15 minute totals from the Deltares NFFS-FFC archive. Unfortunately there is an issue with the Midlands raingauges as Deltares receives both hourly and 15 minute rainfall totals for Midlands in support of NFFS-Midlands and NFFS-FFC and these get mixed up on import into the Deltares NFFS-FFC instance. Operationally there is no issue as only one type is sent to an instance of NFFS. CEH were sent the raw NFFS telemetry feeds and produced 15 minute rainfall totals for Midlands region. Despite this, the Phase 2 data take-on via NFFS-FFC proved to be much more efficient than in Phase 1 via WISKI.

Scotland

Prior to the G2G for RRCs project commencing, the CEH G2G hydrometric archive for Scotland consisted of river flow and raingauge data extracted from the SEPA WISKI archive for the period 1 January 2007 until at least October 2010. Within Phase 2 of the project, national raingauge (264) and river flow (252) datasets were extracted from WISKI for the period 1 January 2010 to at least 30 September 2012. There were a small number of data issues that were quickly resolved. Figure 3.2 shows the distribution of the raingauges across Scotland relative to the gauged catchments.

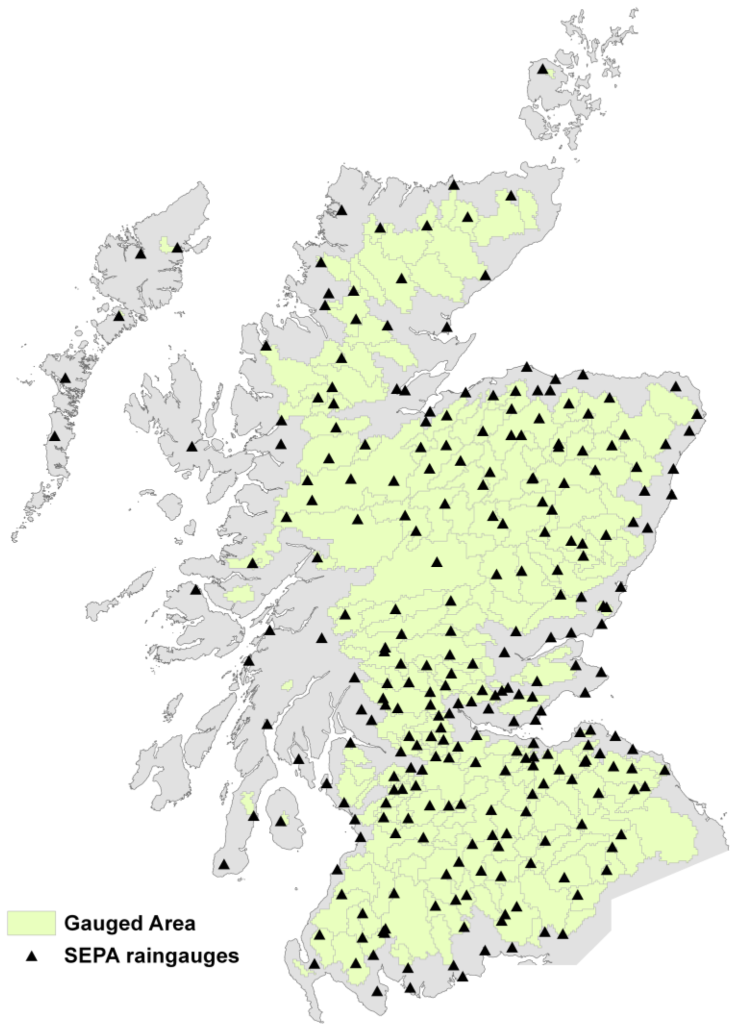


Figure 3.2 Map showing distribution of the 264 raingauges over Scotland and the G2G gauged modelled area.

In addition to the national datasets, a small dataset from eight raingauges and two river flow stations were provided to support analysis of the Comrie November 2012 case study.

3.3 Gridded observed rainfall products

Based on previous experience, automated quality-control of the national sets of raingauge data is required to mitigate the effect of erroneous rainfall data on G2G modelling. Although the Environment Agency and SEPA have generally performed some form of quality-control on data stored in WISKI and quality flags exist for some raingauge records, it is not necessarily done consistently across all sites. In addition, the real-time NFFS-FFC archive data for 2011 and 2012 over England and Wales had no quality-control applied before delivery and exhibits the usual issues associated with real-time data (e.g. erroneous values, missing periods recorded as zero rainfall).

During the FFC and SFFS projects that implemented the G2G operationally, an automated quality-control algorithm was developed by CEH (Howard *et al.*, 2012). To enable real-time application and computational efficiency, all gauges are processed in parallel for each 15- minute interval by the algorithm. A set of simple tests are performed on each gauge separately, before more involved comparisons to neighbours are made. Robust statistics (median and median absolute deviation) are used as the basis for detection of outliers. When a suspicious gauge value is detected, the value is flagged as missing and the details are logged. In the case of a historical study such as here, past records are set to missing for the duration of the period over which the flag is set.

During the FFC and SFFS projects the approach described by Howard *et al.* (2012) was applied and HyradK (Moore *et al.*, 2004; Cole & Moore, 2008) used to generate a 1km grid of rainfall estimates using raingauges alone and in combination with radar data. The analysis of the G2G model outputs using the different rainfall estimates led to the raingauge-only product being recommended for use in model calibration and as the first choice to use operationally. This is consistent with previous findings and how rainfall is used operationally in G2G (Price *et al.*, 2012; Cranston *et al.*, 2012).

For the new raingauge data (2009 onwards for FFC and 2010 onwards for SFFS), an updated form of the quality-control algorithm has been applied. This includes improvements to the cross-correlation calculations and partial blockage detection along with further minor developments. HyradK has been used to generate the 1km raingauge-only rainfall estimate as before. It should be noted that any automated method, however carefully designed and calibrated, will occasionally remove valid rainfalls and let erroneous rainfalls remain but the overall benefit to the modelling process should be positive.

3.4 Rainfall forecast products

The Met Office have provided archived and live feeds of rainfall forecast data in support of the project. This section first considers the rainfall forecast domains required for the project and, separately, for the live feed to the Environment Agency for operational use. This is followed by more details about the various rainfall products received for use in the project.

3.4.1 Rainfall forecast domains

CEH provided guidance on the appropriate rainfall domains to use for extracting archived data and also the live feeds to both CEH Wallingford and Environment Agency. This considered the following factors:

(i) minimising the data volume of ensemble rainfall products,

(ii) data needed to cover England, Wales and Scotland where possible,

(iii) the live data feeds to Environment Agency and CEH could be different.

Table 3.1 below lists the rainfall domains for various Met Office products received by the Environment Agency and SEPA. The domain is defined by the National Grid References of the lower left corner of the lower left pixel and the top right corner of the top right pixel. The domains are also plotted in Figure 3.3.

Environment Agency live feed

It is clear that the current domain for STEPS/NWP rainfall forecast data is much larger than that needed for EA hydrological modelling applications. It was recommended that the UK radar composite domain be used for the new Blended Ensemble forecast rainfall data. This is about half the area and so the data volume should be about 50% smaller too.

If Environment Agency data volumes need to be reduced further, two obvious options exist:

(i) Reduce the domain further by bringing in the south-west corner of the domain to either:

   a. The Isles of Scilly (this would remove the Channel Islands from the domain)

   b. The mainland coast of Cornwall (this would remove the Isles of Scilly and the Channel Islands from the domain)

(ii) Only providing rainfall values over the land area of England & Wales (or a buffer around it) and setting the remaining pixels to missing. This should aid compression and make the files smaller whilst not affecting hydrological modelling applications. However, it will prevent visualisation of “what rainfall is coming”, although using a buffer around England & Wales would moderate this.

Consultation with the Environment Agency user community is recommended before implementing any of these changes.

Table 3.1 Details of existing operational domains used at the start of the project.

Product	**Domain (NGR)**	**Area (km^2)**	**Resolution**	**Colour on map**
UK radar composite (EA)	(20000,-75000) to (660000,700000)	496,000	1km 496,000 pixels	White
STEPS (EA)	(-167000,-145000) to (857000,879000)	1,048,576	2km 262,144 pixels	Black
NWP (EA)	(-167500,-144500) to (856500,879500)	1,048,576	4km 65,536 pixels	Not shown (almost the same as STEPS)
UK radar composite (SEPA)	(45000,450000) to (465000,1060000)	256,200	1km 256,200 pixels	Blue

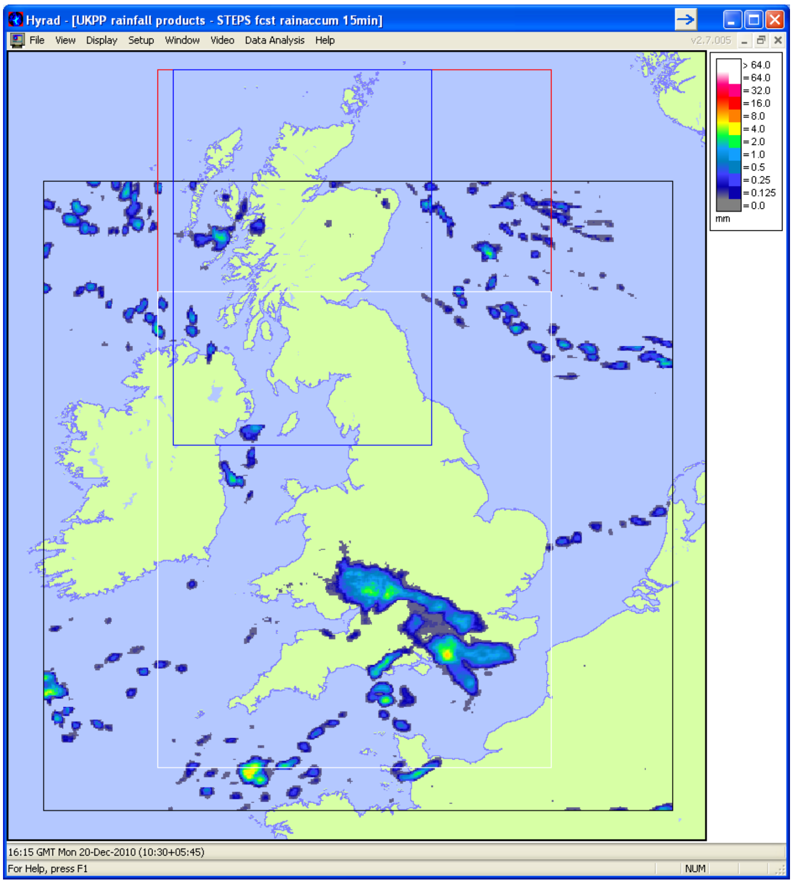


Figure 3.3 Map of existing (black, white and blue) and recommended rainfall forecast domains (red).

SEPA live feeds

If SEPA decide to implement a live datafeed of the blended ensemble rainfall forecasts it is recommended to review the domain of the supplied data with similar considerations to the Environment Agency feed above. The current Scotland radar data domain is fairly close to the hydrological drainage area relevant to Scotland (minus the Shetlands) and thus a good candidate; however, there is scope to bring the southern boundary up a little to reduce data volumes by a small amount. Again, consultation with the SEPA user community would be required before making any changes.

CEH Wallingford live feed and archived data

For CEH Wallingford, the rainfall forecast domain selected encompasses the UK radar domains used by the Environment Agency (white line in Figure 3.3) and SEPA (blue line). The dimensions are given in Table 3.2 and the extension to the Environment Agency and SEPA domains is given in red in Figure 3.3.

Table 3.2 Recommended rainfall domains for the live feeds and archived data

Product	Domain (NGR)	Area	Resolution	Colour on map
Blended Ensembles (EA live feed)	(19000,-75000) to (661000,701000)	498,192 km^2	2km 249,096 pixels	White
Blended Ensembles (CEH live feed and archived data)	(19000,-75000) to (661000,1061000)	729,312 km^2	2km 364,656 pixels	Red

3.4.2 Deterministic UKV (~1.5km) NWP data

The Environment Agency/Defra “Storm-scale numerical modelling” project (FD2207) examined the ability of a 1 km configuration of the MetUM to predict convective rainfall (Roberts, 2006). It concluded that a significant increase in forecast accuracy was achievable and that the benefits for fluvial and pluvial flood forecasting could be enhanced further by post-processing the output to take account of forecast uncertainty. Research and development by the Met Office into high-resolution NWP and the domain and lead-time over which different NWP resolutions operate has continued.

During 2012 the Environment Agency Regions used the 0-36hr UK4 rainfall forecasts and the FFC also used the extended 0-120hr UK4 rainfall forecasts. From January 2013 the 0-36hr UK4 forecasts are replaced by UKV (Tang *et al.*, 2013) and the extended UK4 rainfall forecasts made available to the Regions.

The Met Office have been running the deterministic UKV (~1.5km) NWP forecasts since 2010 non-operationally for evaluation. Data were extracted by the Met Office from their MASS archive for the period 4 January 2010 to 31 December 2011. These data were provided “as is” and known to suffer from some gaps or incomplete runs as this was a non-operational product over this period. Also, several changes to the forecast length occurred during this period. Figure 3.4 highlights the temporal distribution of the missing forecast data, assuming a 36 hour forecast was expected. Table 3.3 provides comments on the data available.

Forecasts were run with time-origins of 03:00, 09:00, 15:00 and 21:00 each day and mapped on to a 2km resolution grid.

Table 3.3 Comments on the UKV (~1.5km) NWP data

Date	Comment
4 Jan 2010 – 4 Mar 2010	Expected forecast length was 6 hours
5 Mar 2010 – 12 Jul 2010	Expected forecast length was 24 hours
12 Jul 2010 – 31 Dec 2011	Expected forecast length was 36 hours
21 Dec 2010 – 31 Dec 2011	The forecasts weren’t always complete so around 8-16% of images typically are missing due to this. These could be at any point during a forecast.
28 Aug 2011 – 24 Sep 2011	Long period where no forecasts were available

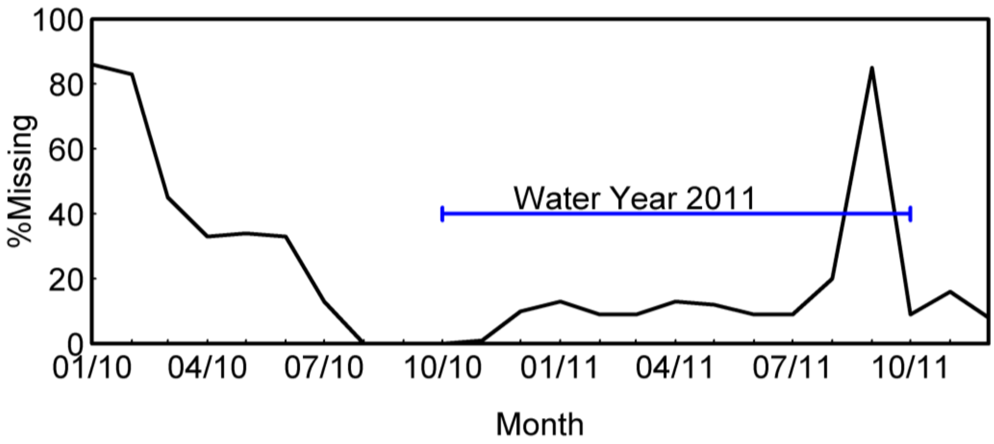


Figure 3.4 Amount of data missing, assuming a 36 hour forecast was expected.

3.4.3 Blended ensembles

The Environment Agency has a requirement for precipitation forecasts with a horizontal resolution of 2 km or finer, an accumulation time-step of 15 min or finer and a range of at least several days. There is also a desire to understand and estimate the uncertainty in the rainfall forecasts and subsequent flood forecasts. At present, no single Numerical Weather Prediction model forecast can satisfy this requirement.

Concurrent to the developments in NWP reported in Section 3.4.2, the Short Term Ensemble Prediction System (STEPS) was developed to provide ensembles of precipitation nowcasts. STEPS blends extrapolated radar and satellite observations with the most recent forecast from the 4 km configuration of the MetUM (UK4) and a time series of synthetically generated precipitation fields (noise) with space-time statistics inferred from weather radar. The noise component serves to account for uncertainties in the evolution of the extrapolation and UK4 forecast components and also to downscale the UK4 forecast.

The "*Blending convective scale NWP with ensemble nowcasts*" project was established to integrate these developments and produce a seamless, high-resolution (2 km) ensemble forecast of 15 min rainfall accumulations which are suitable for use in fluvial and pluvial flood forecasting and warning.

Outputs from the work include a *quick-win* blending algorithm designed as an upgrade for STEPS, and known as STEPS-2, was implemented in February 2012.

Benefits of the STEPS-2 algorithm are:

- near optimal combinations of an extrapolation nowcast and multiple NWP model forecasts of precipitation;
- quantification of uncertainties in radar rain-rate estimates, nowcasts and NWP forecasts, with this information conveyed to users in the form of ensembles;
- downscaling coarse resolution NWP forecasts (e.g. 18 km MOGREPS-R forecasts), correcting for the under-prediction of both orographic precipitation and the heavy and extreme precipitation associated with convective showers;
- generation of large ensembles of precipitation nowcasts and forecasts needed to propagate the associated uncertainty through hydrological models such as the G2G model.

Historical Period 1: June 2011

This Summer 2011 period was used within the STEPS-2 trial and evaluation phase. Although this is not associated with any particular major flood, it is the first month-long period of STEPS-2 ensemble data that have been made available for testing with hydrological models.

The February 2012 version of the STEPS-2 algorithm was used to reproduce the following 15 minute rainfall accumulation forecasts on a 2km grid.

- *Control + 12 member ensemble nowcast to T+7 h, Updated every 15 min.* Uses a radar extrapolation nowcast + UK4 deterministic forecast

This equates to 1,048,320 individual rainfall images over the month of June. Unfortunately due to technical issues some 306,167 images (29%) were missing.

A majority were due to forecast data either not being available in the archive or only having a small proportion available. This mainly affected 1-4, 6, 27-30 June 2011. The remainder were generally due to the end of forecasts not being completed.

Historical Period 2: 5/6 September 2008

This case study period was selected as it contained the Morpeth extreme flood and had relevant ensemble data available. There was only one UK4 forecast available. It had a forecast origin of 06:00 GMT 5 September and provided forecast data over the 30hr period 12:00 GMT 5 September to 18:00 GMT 6 September. Usually the UK4 data would update every 6 hours and the latest run would be used with STEPS-2.

The following ensemble data (15 min rainfall accumulations) were provided

- *Control +* ***23*** *member ensemble nowcast to T+7 h, Updated every 15 min.*
  Nowcast origins from 12:00 GMT 5 September 2008 to 11:00 GMT 6 September 2008.
  Uses a radar extrapolation nowcast + UK4 deterministic forecast. STEPS-2 algorithm (as at February 2012).
  Very few small periods of missing data within some members.

- ***Single 24*** *member UKV ensemble forecast to T+30 h starting 12:00 GMT 5 September 2008*
  Forerunner to current MOGREPS-UK ensemble.
  Existing run stored from previous analysis.

Live blended ensemble feeds

CEH started to receive the live feed of blended ensemble data on 10 May 2012.

Initially both the Environment Agency and CEH received:

- *12 member ensemble nowcast to T+7 h, Updated every 15 min.*
  Uses a radar extrapolation nowcast + UK4 deterministic forecast

- *12 member ensemble forecast to T+24 h, Updated every 6 h.*
  Uses a radar extrapolation nowcast + UK4 deterministic forecast

At first, the STEPS-2 algorithm used UK4 but this changed to UKV late in November 2012.

Initially the T+24h forecasts were run 4 times a day but at unequally spaced forecast origins: 00:15, 07:15, 12:15 and 19:15. From 17 September 2012, and to coincide with operational roll-out to the FFC, the origins became equally spaced: 00:15, 06:15, 12:15 and 18:15.

From 4 December 2012 the number of ensembles increased to 24 members.

The FFC received trials of this product during Summer 2012 and it went operational on 14 September 2012. Currently (August 2013), the NFFS-FFC system hosted at Capgemini is running 24 member blended G2G ensembles 4 times a day out to T+24 h and every hour to T+7 h.

3.4.4 Future developments

The exact timetable and definition of future ensemble forecast products is constantly changing due to ongoing developments and availability of computing resources related to supercomputer upgrades. Potential future and ongoing developments are:

- Potential increases in STEPS-2 ensemble size.
- MOGREPS-UK, a 2.2km NWP ensemble product, has been running for evaluation since May 2012. Performance against UKV is being assessed.

MOGREPS-UK is anticipated to pass operational acceptance testing in 2013. This may replace the UKV deterministic forecast in the blended ensemble.

- More frequent UKV runs. The update frequency may change from 6 to 3 hours.
- There is also the Nowcasting Demonstration Project which is using 4D-Var data assimilation of Doppler winds over the southern UK. This currently runs every hour out to 7 hours and soon to be 12 hours. Although initially deterministic, the frequent run-times allow for time-lagged ensembles to be produced.

3.4.5 Issues with rainfall forecast data take-on

Extraction of historical forecast data has proven to be a lengthy process and was delayed due to Met Office resourcing issues and failure of some retrievals from the Met Office MASS archive. At the time of the retrieval there was some overhead in setting it up and then, once started, the actual extraction could take several weeks due to the design of the archive. Transfer of data was made more difficult for both Met Office and CEH due to the need to use DVDs as the transfer media for Met Office IT security reasons. This highlighted the need to explore more efficient procedures (e.g. USB mass storage) as a priority to gain resource savings: this simple data transfer task involved around 30 DVDs which have proved to be error-prone adding work to all parties. More recent data transfers from the Met Office have trialled an ftp route which appears easier to use and will address some of these issues.

3.5 Rapid Response Catchment information

The Environment Agency have provided the 2010/11 Rapid Response Catchment register information. The national datasets contained 64 boundaries relating to very high and high priority RRCs (although these weren't necessarily hydrological boundaries), related WISKI surface water recording stations and a point dataset of at-risk locations. Furthermore, South West Region provided additional information including hydrological boundaries.

Figure 3.5 shows the distribution of the England & Wales RRC datasets relative to the G2G model domain that is gauged and highlights that some regions do not have any RRCs within the national dataset. Investigations matching the RRC boundary and/or point locations with the river gauging station locations configured in G2G were undertaken. It was found that this matching process would only indentify a small subset of the G2G sites and that these would not necessarily be sites with a small catchment area, due to the relative positions of the river gauging stations and the RRC locations. The Environment Agency have plans to update the national RRC information to include more details from the Regions and as such should only be used in an indicative way.

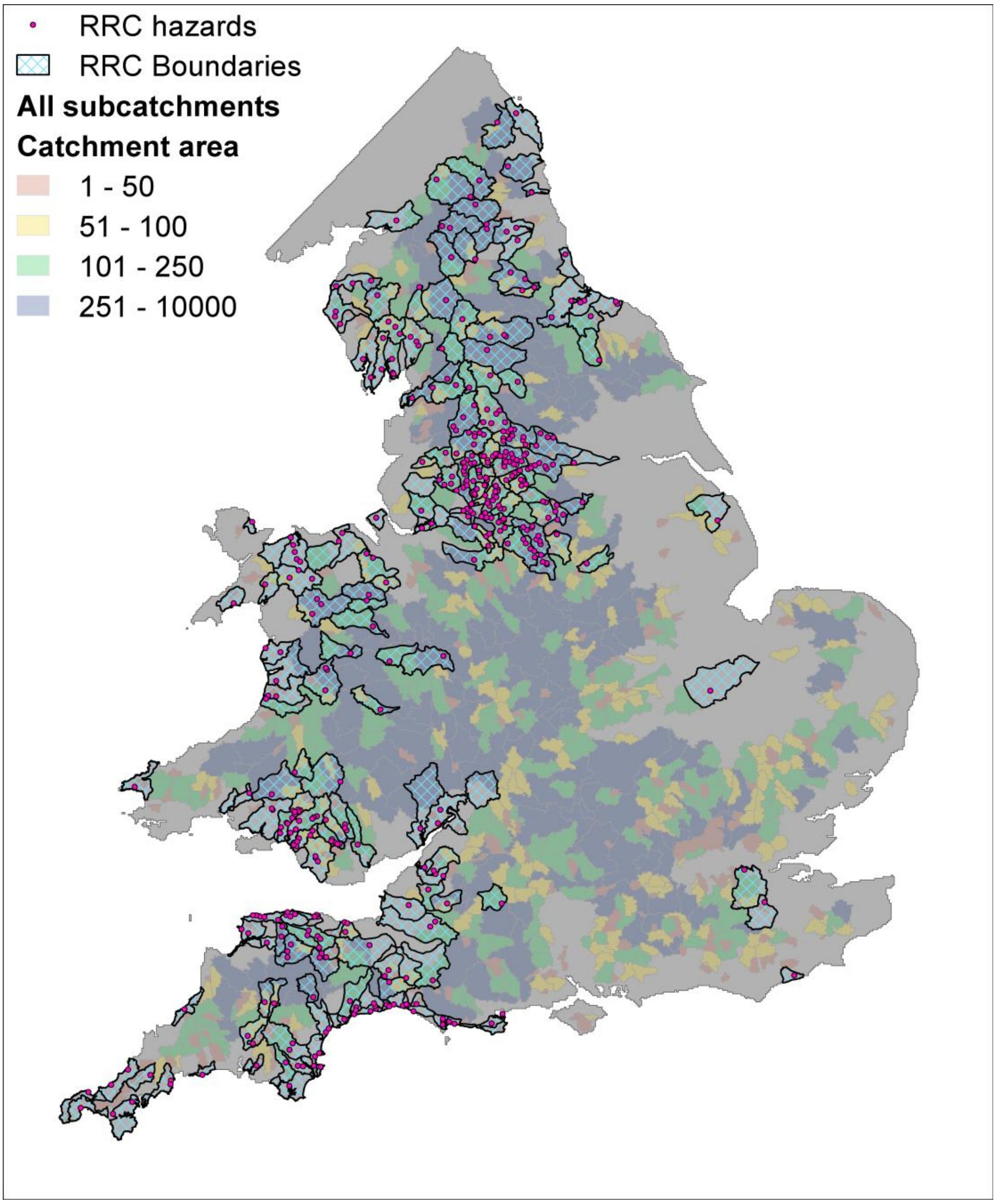


Figure 3.5 Location of national England & Wales RRC register points of interest and boundaries (note these may not be hydrological). The backdrop shows the G2G model domain that is gauged with catchments/sub-catchments coloured by total catchment area.

4 Assessment methodology

4.1 Setting up the “G2G for RRCs” configuration

England & Wales

Before analysis of G2G performance for RRCs could begin, the G2G configuration had to be reviewed and updated to make use of river flows from the newly available river gauging stations. As the England & Wales hydrometric dataset had been increased by two water years, all sites were reassessed including those previously discarded.

The first stage was to locate the appropriate G2G pixel for any new sites and then run the G2G model with all sites configured to have state-updating and flow insertion switched on. Model runs over the full 5-year dataset were assessed and decisions made on which sites to keep and which to discard. Also simple edits were performed to some time-series to remove spurious or suspect data so that model calibration, assessment and derivation of flow return period (Q(T)) grids would not be adversely affected.

This process also identified several new reservoir locations that are now included in the configuration (although the precise settings for these could be refined further in the future). Decisions were also made on which sites to use for assessment, state-updating and flow insertion.

The final “G2G for RRCs” configuration includes 1032 sites, of which 672 were from the original configuration (OP in Figure 4.1) of 708 (some sites have now closed or are now unsuitable for use), 105 were used previously for off-line model calibration (CAL), 53 were previously rejected (REJ) but now identified as useful and 202 are new sites (NEW). Figure 4.1 shows the distribution of the configured sites.

In total, 914 sites of the 1032 are currently available operationally (912 were finally included in the G2G release (v1.5.000), rolled-out to FFC in March 2013). These are the non-transparent catchments in Figure 4.1. This is a 28% increase on the 708 sites currently configured and doubles the amount of sites for which flows are currently being sent to G2G. This will have a major beneficial impact on the operational performance of G2G.

A total of 932 sites are suitable for model assessment. Whilst the “G2G for RRCs” project focuses on RRCs, it is still important to put the RRC performance in the context of countrywide model performance, particularly when trialling model improvements.

Scotland

Assessments over Scotland reported here are based on the current SEPA operational configuration (v1.4.100) minus 6 sites that have since been requested for removal by SEPA giving a total of 227 sites. Whilst all 227 sites have been used to generate the model results, only 181 are suitable for use in the model assessment presented here.

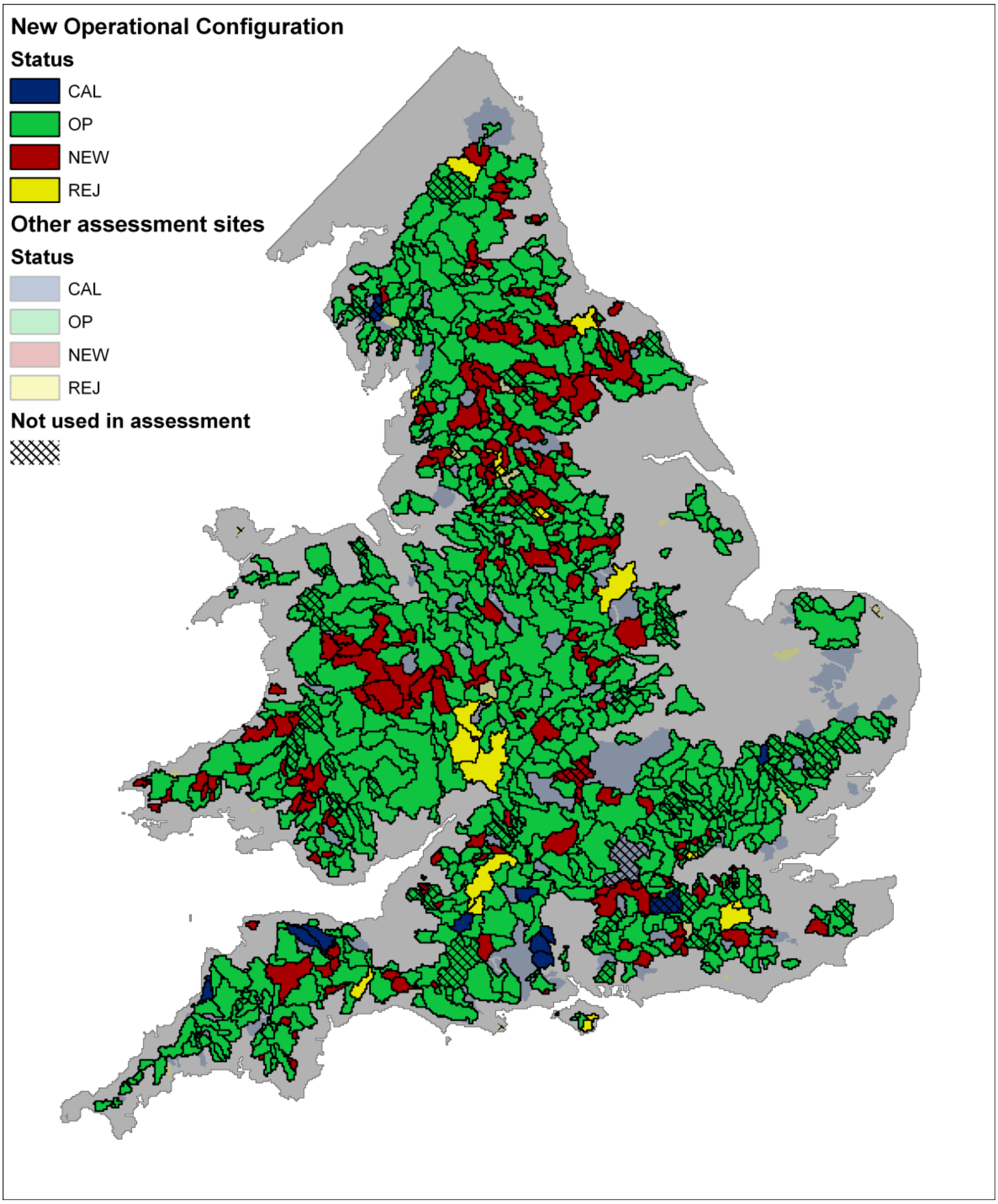


Figure 4.1 Distribution of configured sites in G2G with catchments/sub-catchments coloured according to their status: CAL, OP, NEW or REJ (see text for definitions)

4.2 G2G model assessment for RRCs

The study has assessed the performance of the G2G Model for RRCs using flows from river gauging stations but treating sites as “ungauged”. As discussed in the introduction (Section 1.2.1), four forms of assessment are involved and these are outlined next.

4.2.1 Simulation-mode assessment

The assessment will initially be reported in *simulation-mode* using HyradK raingauge-only rainfall as input: this provides a useful benchmark against which to assess the forecast performance of the model. It focuses attention on the ability of the G2G Model to transform rainfall to river flow for rapid response catchments, and identify any shortcomings in model formulation and configuration for such situations. In this mode, G2G simulations at a given site use flow-insertion from upstream sites, and state-correction from all sites, throughout the period.

The G2G model employed in this assessment is the same as that used operationally by FFC (v1.3.001) **except** the new configuration of 1032 sites has been used to perform the automated local calibration of flow routing, return flow fraction and ARMA error-prediction parameters.

Note that not all of the 1032 sites are supported in NFFS-FFC for real-time use with G2G. Nonetheless, their river flows can be employed for offline calibration of G2G local parameters. The March 2013 release (v1.5.000) of the G2G operational model allows use of these locally calibrated parameters for sites not supported in NFFS-FFC.

This assessment makes use of the four water years 2007/08, 2008/09, 2009/10 and 2010/11: a water year runs from 1 October to 30 September.

4.2.2 Forecast-mode assessment: foreknowledge of observed rainfall

A second form of assessment, still using observed rainfall, evaluates the G2G model in *forecast-mode* where flows up to the time the forecast is constructed are used for data assimilation via direct flow insertion, state correction and ARMA error prediction. This emulates the real-time performance of the model and how forecast accuracy falls off with increasing lead-time, whilst avoiding the confounding effects of rainfall forecasts. It provides a useful performance benchmark against which to assess G2G flow forecasts produced using the UKV NWP forecast rainfalls as input, for times after the time-origin of each G2G forecast.

This assessment uses the four water years spanning the period 2007-2011. For comparison with the NWP data, forecasts were made 4 times a day out to 24 hours.

4.2.3 Forecast-mode assessment: use of deterministic rainfall forecasts

A third form of assessment, carried out in *forecast-mode* using different sources of *rainfall forecasts in deterministic form*, aims to emulate the operational forecasting situation, or what might be possible when new forecast products become operationally available. The different rainfall forecast products considered are discussed in Section 3.4.

The initial focus was on assessing the UKV forecasts over a long period. As there are significant gaps in the early parts of the archived data, the analysis focused on the 2010/11 water year.

In the assessments reported later, the G2G forecasts have been set-up to run 3 hours after the forecast origin of the NWP data to emulate the delay in availability of the forecast data in an operational setting. G2G forecasts were made 4 times a day out to 24 hours using the T+3 to T+27 hour rainfall forecast from the NWP model. Note that the non-operational UKV archive is less complete than an operational feed would be. Therefore, to aid comparison with the G2G forecasts using foreknowledge of observed rainfall, missing NWP data were infilled by the observed rainfall. This means that any difference between the two G2G forecasts should only be due to differences in the rainfall estimates (NWP vs. Observed) rather than being confounded by missing forecast data.

4.2.4 Forecast-mode assessment: use of ensemble rainfall forecasts

A fourth form of assessment, carried out in *forecast-mode* using different sources of *rainfall forecasts in ensemble form*, aims to emulate the current operational forecasting situation, or what might be possible when new forecast products become operationally available. The different rainfall forecast products considered are discussed in Section 3.4.

This builds on ways of assessing ensemble forecast products developed under previous work undertaken by CEH and others for the Environment Agency. The case study datasets selected serve as indicative demonstrators whilst the use of rainfall forecast product datasets over longer periods (months) allow more meaningful assessments to be made. This form of analysis also provides an opportunity to assess the potential benefits of using ensemble rainfall forecasts versus deterministic rainfall forecasts.

4.2.5 Sites used for assessment

England & Wales

As indicated in Section 3.5, the England & Wales RRC register datasets are not entirely suitable for identifying gauged RRCs and would not fully exploit the large hydrometric and national rainfall datasets available for G2G modelling. Therefore all 932 sites suitable for model assessment have been used and the results grouped and analysed according to different criteria relevant to RRCs. Having results for all sites also allows the RRC performance to be put in context relative to the countrywide model performance.

For all the assessment sites various catchment characteristics and details have been determined to aid analysis. Particular characteristics relevant to RRCs are the catchment area, urban and sub-urban coverage, terrain slope and whether a “headwater” site (i.e. the site has no sites upstream of it in the configuration). Figure 3.5 colours the gauged catchment/sub-catchment domains according to area to give an impression of the spatial spread of catchment size whilst the left column of Figure 4.2 shows the distribution of the sites against catchment area for all sites and just headwater sites.

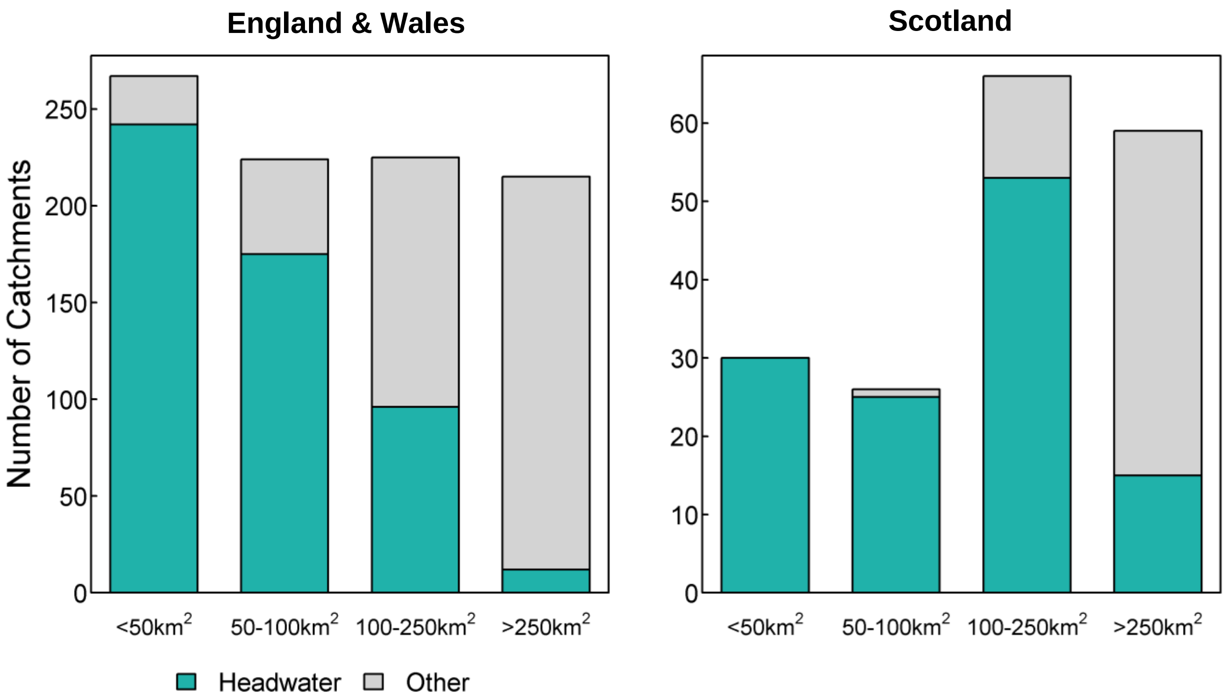


Figure 4.2 Number of assessment catchments grouped by catchment area for England & Wales (left column) and Scotland (right column). Also shown are the number of "headwater" locations (green) which have no gauged site upstream of them.

Scotland

The catchment area distribution for the 227 sites in Scotland is given in the right column of Figure 4.2. In comparison to the England and Wales (left column) distribution there are relatively fewer catchments below 100 km^2 and almost all those do not have an upstream station so are "headwater" sites. In general, Scotland also has fewer nested catchments, even for the larger catchments areas. These differences between Scotland and England & Wales reflect the different densities of the gauging station networks across the regions. Figure 4.3 provides a spatial map of the catchment areas and highlights the spatial variation in the gauge density. In particular, there are more stations (and more of the smaller gauged catchments) located near urbanised areas such across the central belt of Scotland. In contrast, many of the highland areas have few gauged catchments.

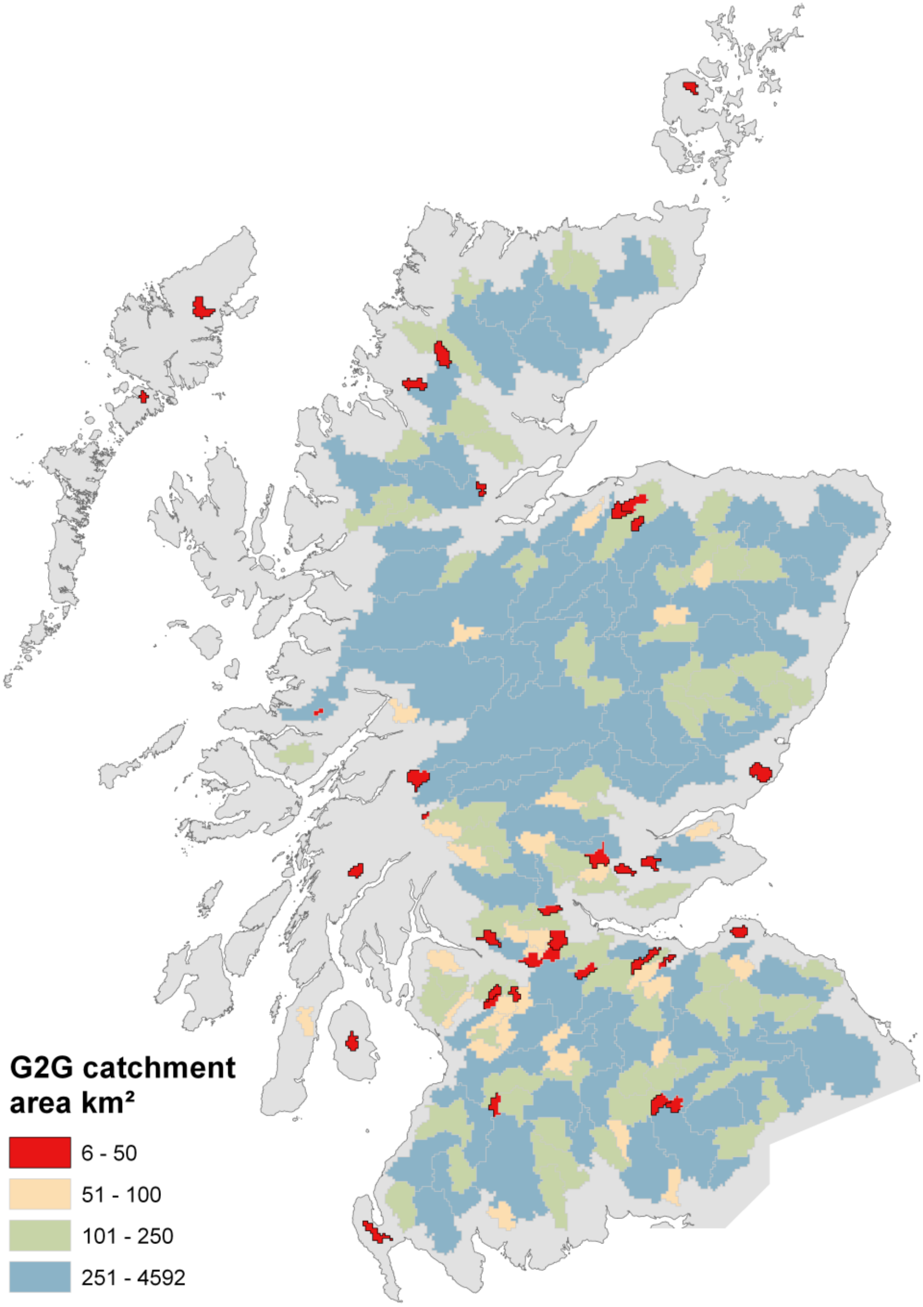


Figure 4.3 Number of assessment catchments over Scotland grouped by catchment area.

4.3 Performance measures used for assessment

The set of measures of model performance used in the assessment that follows are defined in this section. The following notation is used in defining the performance measures:

Q_t	*Observed flow at time t*
q_t	*Modelled flow at time t*
n	*Number of observations*
$e_t = Q_t - q_t$	*Model error at time t*
$\overline{Q} = \sum_{t=1}^{n} Q_t$	*Mean of the observed flow over n observations*

4.3.1 Simulation mode performance

The primary measure used to assess simulation mode performance is the R^2 Efficiency performance measure. This is defined over n observations as

$$R^2 = 1 - \frac{\sum_{t=1}^{n} e_t^{\ 2}}{\sum_{t=1}^{n} (Q_t - \overline{Q})^2}.$$

The R^2 Efficiency gives the proportion of variance in the observations that is accounted for by the model. It is a dimensionless measure and allows meaningful comparisons across different events, catchments and models.

This statistic has a value of 1 for a perfect simulation and takes negative values if the simulations are worse than that provided by the mean observed flow $\overline{Q}$.

It is useful to group and assess R^2 Efficiency measures across several sites. The main method used is to study the distribution of the outputs and look at the Maximum, Minimum, Median and the Upper and Lower Quantiles. For these purposes, the Minimum is set to 0. Standard box and whisker plots can be used to show these quantities.

4.3.2 Forecast-mode performance: deterministic rainfall forecasts

Initially, the G2G forecasts were made every 6 hours to coincide with the NWP forecasts. To gain an understanding of how performance varies with lead-time, and also provide a means to compare with the simulation-mode forecast, the R^2 Efficiency has also been calculated for each site using various portions of all forecasts, namely: 0-6 hours, 6-12 hours, 12-24 hours and the entire forecast.

Secondly, the Probability of Detection (POD) and False Alarm Ratio (FAR) skill scores have been calculated using the 24-hour forecast period. For the purposes of the categorical skill scores, a flood event is derived using the following method.

- For a given 24 hour lead-time forecast, a flood event is defined as occurring if the flow hydrograph is above the threshold at any point during the 24 hour forecast period. This is calculated separately for the observed and modelled flow time-series.

Although there are some limitations when setting a fixed window for assessing skill scores this still provides some assessment of the forecast skill.

The skill scores use the contingency table given in Table 4.1 as their basis.

Table 4.1 Contingency table for calculating skill scores

Event Forecast	Event Observed		Total
	Yes	No	
Yes	*a* hit	*b* false alarm	*a*+*b*
No	*c* miss	*d* correct rejection	*c*+*d*
Total	*a*+*c*	*b*+*d*	*n*=*a*+*b*+*c*+*d*

a = number of times threshold is crossed in both observed and modelled series within a time-interval Δt of each other. **Hit**.

b = number of times threshold is crossed in model series, but not in observed series within Δt of the model series crossing the threshold. **False alarm**.

c = number of times threshold is crossed in observed series, but not in model series within Δt of the observed series crossing the threshold. **Miss**.

d = number of times threshold is not crossed in both observed and modelled hydrographs within Δt of each other. **Correct rejection**.

*Probability of detection (*POD*) skill score*

When an observed event does occur, POD is the proportion of such events that are forecast to occur. POD can also be referred to as the 'Hit Rate' and emphasises the number of events correctly forecast. It ranges from 0 (poor) to 1 (perfect).

$$\mathrm{POD} = \frac{b}{a+c}.$$

*False Alarm Ratio (*FAR*) skill score*

When an event is forecast to occur, FAR is the proportion of such events when an observed event did not occur. This emphasises events incorrectly forecast and ranges from 0 (good) to 1 (poor).

$$\mathrm{FAR} = \frac{b}{a+b}.$$

False Alarm Rate (F) skill score

The False Alarm Rate *F* (not to be confused with FAR) is also referred to as the Probability of False Detection (POFD). This is the proportion of non-events that are forecast as false alarms. This emphasises non-events incorrectly forecast as events (i.e. false alarms) and ranges from 0 (good) to 1 (poor).

$$F = \frac{b}{b+d}.$$

This skill score is contrasted with POD in the Relative Operating Characteristics (ROC) diagram, and its associated ROC Score, in the next section. Otherwise it will not be used.

4.3.3 Forecast-mode performance: ensemble rainfall forecasts

A distinction is made here between assessments of performance over a long set of forecasts, for which skill scores are appropriate, and assessments that are case-study based where more informal displays can prove of greater utility. The two cases are treated separately below.

Continuous assessment

The ROC Score

The POD and *F* (False Alarm Rate) skill scores discussed above for assessing deterministic forecasts can be developed to also assess probabilistic flood forecasts derived from rainfall ensembles. For a given threshold, such as Q(2)/2, the POD and *F* value for different probabilities of exceedance ranging from 0 to 1 can be calculated and the paired values plotted on y- and x-axes respectively. This is called the Relative Operating Characteristics (ROC) diagram (Figure 4.4). A perfect forecast (all events detected with no false alarms) corresponds to POD=1 and *F*=0. Thus values in the top left hand corner of the ROC diagram are associated with highest skill, with a curve showing a significant bend towards this corner indicating forecast skill. When POD=*F*=1 (top right corner) corresponds to events always being warned for and when POD=*F*=0 (bottom left corner) no warning of an event is ever given. Above the 1:1 line are outcomes where POD>*F*. Note that the POD and *F* resulting from the deterministic forecast can be plotted on the same diagram as a single value for the chosen threshold.

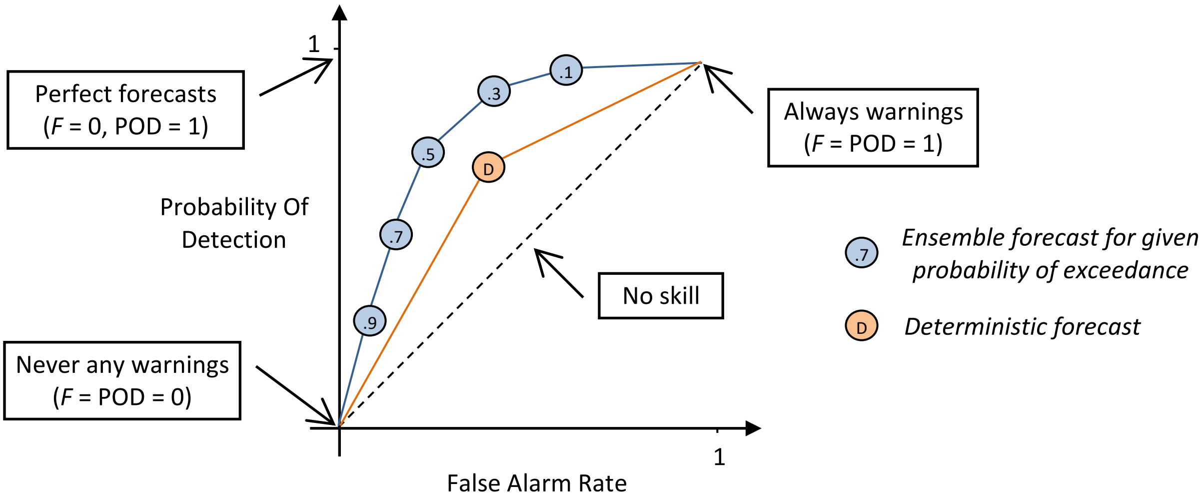


Figure 4.4 Schematic of the Relative Operating Characteristics (ROC) diagram.

The area under the ROC curve, AUC, is ideally 1, with a random forecast lying along the 1:1 line with an area of 0.5, and at worst 0 when POD is always 0. The ROC Score is defined as the AUC normalised with reference to a reference forecast AUC_{ref}, taken here to be a random forecast with no skill and AUC equal to 0.5, so that

$$\text{ROC Score} = \frac{\text{AUC} - \text{AUC}_{\text{ref}}}{1 - \text{AUC}_{\text{ref}}} = \frac{\text{AUC} - 0.5}{1 - 0.5} = 2 \times \text{AUC} - 1.$$

A ROC Score of 1 indicates a perfect forecast and if above 0 the forecast has a skill better than a random forecast.

The Brier Score

The *Brier Score*, like POD and FAR, is a categorical form of score based on crossing of a chosen threshold, but appropriate for use when the forecast being assessed is in the form of a probability. It gives the *mean square probability error* over *n* forecasts.

It is convenient to introduce some notation at this stage to allow the Brier Score to be precisely defined. Let y_i denote the observed value(s) of a quantity of interest (here, river flow) for forecast i and x denotes a threshold value of interest for this same quantity (the Q(2)/2 flow threshold, for example) and used to define the categories of event-occurrence or non-occurrence. We define Y_i as an indicator variable of the observed event, taking a value 1 if the event does occur and 0 if not. The forecast probability to be assessed, $\hat{Y}_i$, is the forecast probability of the event occurring, taking values in the range 0 to 1.

The Brier Score, giving the *mean square probability error* over *n* forecasts, can be defined through the above notation as:

$$\text{BS} = n^{-1} \sum (Y_i - \hat{Y}_i)^2 \quad .$$

It is used to express the typical size of error in probability terms on a scale of 0 to 1, with 0 being “best” (a probability forecast associated with no error).

If BS_{ref} denotes the Brier Score for a reference forecast (for example, one based on climatological relative frequencies), then the *Brier Skill Score* is given by

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{ref}}} \quad .$$

This provides a relative measure of the skill of a probability forecast, giving the proportion improvement in BS of the forecast relative to the reference forecast (e.g. climatology). At best, BSS takes a value of 1 and values less than 0 indicate the forecast performs worse than the reference over the period of assessment.

Study application of BSS and ROC Score

The i’th probability forecast derives from an ensemble of 15 min G2G forecast flow values. The probability forecast being assessed is the probability of the threshold flow (e.g. Q(2)/2) being exceeded at any time over a forecast horizon (here of duration 24 hours). This forecast probability is calculated as the proportion of flow ensemble members exceeding the flow threshold at any time over the forecast horizon. The indicator of an “event” is the exceeding of the flow threshold by flow observations (at 15 minute intervals) at any time over the forecast horizon.

Forecasts available for assessment (aligned to use of the Blended Ensemble rainfalls with G2G) are made every 6 hours (or at alternating steps of 7 and 5 hours up to mid-September 2012). In the applications considered later, the climatological reference $\mathrm{BS}_{\mathrm{ref}}$ is calculated using a probability forecast $\hat{Y}_i$ equal to the mean of the *n* indicator variables Y_i. Thus it is dependent on the period used for assessment.

Case study assessment

For the specific case-study assessments that are focused on particular times and regions, a range of quantitative and qualitative methods have been used with some concerning regional assessments and others concentrated on individual locations. Examples of some of the methods are given here with explanation of diagrams where necessary. These are used heavily in Section 5.5.

Firstly, it has proved to be very useful to summarise an ensemble forecast at a regional level and over a forecast duration (e.g. 24 hours). Figure 4.5 provides a summary description of the spatial plot that has been used. For a given Q(T) flow threshold, the colour circles signify how many ensemble members exceed the threshold at some point during the forecast duration. In order to compare with what was observed, a solid black dot is used to signify if the flow-inserted modelled time-series crossed the threshold during the forecast. For gauged sites where flow-insertion is turned on (which is a majority of sites) the “flow-inserted series” is simply the observed flow. For gauged sites where flow isn’t inserted, the “flow-inserted series” is simply the simulation-mode G2G flow at that location (see Section 4.2.1). This modelling approach is what is used to currently derive the G2G-adjusted Q(T) thresholds which aim to accommodate any biases in the G2G flows.

In addition to the regional summary map of performance, it is also useful to look at specific gauged locations. Figure 4.6 provides an annotated sequence of spaghetti ensemble plots for a gauged location. Both the regional summary map (Figure 4.5) and specific location plots (Figure 4.6) are used in the case-study assessments presented in Section 5.5.

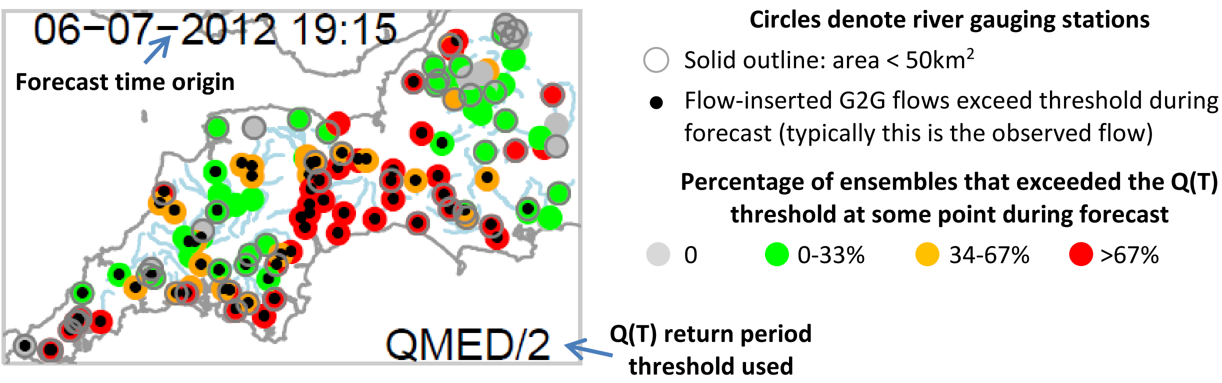


Figure 4.5 Regional summary of an ensemble G2G forecast for a given Q(T) threshold.

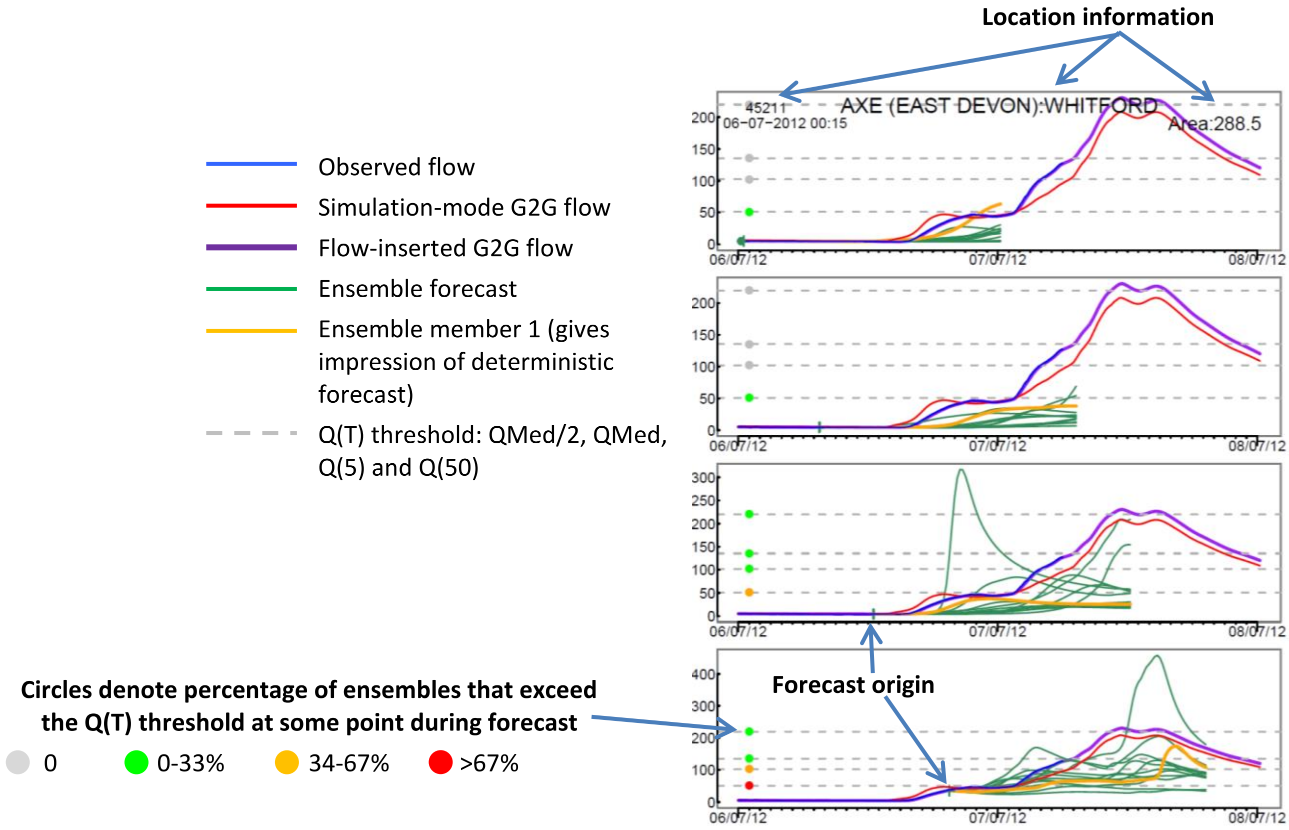


Figure 4.6 Sequence of ensemble G2G forecasts at a specific location.

4.4 Performance Summary pages

As part of previous operational G2G deliveries to both FFC and SFFS, CEH have provided "Performance Summary" pages for each gauged forecast location. The Performance Summary aims, on one page, to provide a concise indication of the performance of the G2G Model. Its purpose is to help guide the operational use of G2G forecasts for flood warning. A Performance Summary - in PDF form for subsequent rapid access and viewing via a point-and-click tool-tip in NFFS and FEWS Scotland - is provided for each river gauging station configured into the G2G Model.

These first generation Performance Summary pages only assessed simulated (simulation-mode) flows from G2G. As part of this project, there has been a significant upgrade to the template design of the Summary to include information on the performance of the model at a given site in forecast-mode.

Under Phase 1, the focus was on including a forecast-mode assessment assuming foreknowledge of forecast rainfall (as described in Section 4.2.2). The details are described fully in "A Guide to the Performance Summary" (Centre for Ecology & Hydrology, 2013) and a sample output is provided in Figure 4.7.

Major advances include:

- Performance measures over an extended set of 4 water years (for FFC) for modelled flows in simulation-mode. Measures are calculated for the whole period and each water year.

- A second page with hydrographs for each year showing (i) pure simulation (no state-updating or flow insertion), (ii) simulation with state-updating and (iii) simulation with upstream flow insertion and state-updating.
- Display of 24 hr and 5 day forecasts for the 5 largest events (top right, first page)
- For the assessment, 24 hr forecasts are made every six hours and 5 day forecasts are made every 24 hours over the whole assessment period (4 water years for FFC).
- Performance measures for forecasts at different lead-times.
    - Measures are calculated over the full 24 hr and 5 day forecasts using only those shown for the 5 largest events.
    - For the 24 hr forecasts, measures are also calculated using every forecast over the assessment period for: (i) the full 24 hr forecast, (ii) the first 6 hours of each forecast, (iii) the 6-12 hr period of each forecast and (iv) the last 12 hours of each forecast.
    - For the 24 hr forecasts, measures are also calculated using every forecast over the assessment period for: (i) the full 5 day forecast, (ii) days 2 and 3 of each forecast, and (iii) days 4 and 5 of each forecast.
- POD and FAR skill scores for forecasts over the whole period for different lead-times relative to the thresholds: Q(2), 75% of Q(2) and 50% of Q(2). The quantity Q(2) denotes the flow of magnitude Q that has a return period of 2 years: it is the median of the annual maximum flood, sometimes referred to as QMED. Quite low thresholds are used because at longer return periods very few events occur; the POD and FAR values for longer return periods have little meaning when calculated for a single site over a short record.
- Additional information about the G2G configuration for this location including local calibrated parameters, ARMA parameters and details of the gauging locations up and downstream.

The new Performance Summary has been generated for each site and has helped greatly in understanding the G2G modelled flows and provides a useful insight into the potential forecast mode performance of G2G. The new form of Performance Summary was used in the operational FFC G2G delivery following Phase 1 (rolled-out in March 2013). It will also form the basis for the Performance Summary that is delivered with future upgrades of the operational G2G Model to both FFC and SFFS.

Further development to include information on G2G performance using ensemble rainfall forecasts will be considered in the future but the utility will be limited by the time-period that ensemble data are available for.

Location ID
NFFS ID (WISKI ID, NRFA ID)

Catchment Outline and location
- G2G outlet (FI)
- G2G Outlet (no FI)
- G2G Outlet (with lake/res)
- WISKI outlet
- Raingauge

G2G boundary
NRFA catchment
River

Geological Permeability Maps:
Solid (left)
Superficial (right)

Permeability:
High
Moderate
Low
Mixed
None

1 = Artificial Influences
2 = Marked B if an NFRA Benchmark catchment
3 = Marked C if used in CEH calibration (155 sites)
4 = Gauge Data Quality for low, medium and high flows:
Green = Good, Yellow = Fair, Red/Amber= Poor
5 = Urban factor. Weighted average of urban (0.7) and suburban (0.3) fractional coverage (area of circle also).
6 = Area in km². Recorded area (G2G area in brackets), plus area of circle. If a % is shown this is the % of the area that could potentially be flow inserted.
7 = Percentage of area that is lakes
8 = Raingauge density used by CEH. (1=average) dark blue > 1, light blue < 1.
9 = Main Host Class (and percentage)

Event based hydrographs

5 biggest events in water years 2008-11 in date order. 1-day and 5-day forecasts are given for each event. Colours are alternated light and dark to help distinguish consecutive forecasts.

Events are at most 1 week before peak, 2 weeks after and are identified using an automated POT method for all flows over median flow

1-Day Forecast (every 6 hours)
5-Day Forecast (every 24 hours)

Model Setup
Grid references: G2G is 1km cell location.
Details on whether site is used in state-updating and/or flow-insertion (affects downstream).
List of adjacent down- and up-stream locations.
(Values in brackets = % of total area, grey = upstream site is not flow inserted
Local pars: gives G2G internal settings).

Summary model performance statistics
Shown for a range of periods and model outputs. NB Spring 2008 was a 4 month period used a lot during calibration.

Graphs of model performance for Spring 2008
Top graph is for flow insertion.
Lower graph shows state-updated and simulation run.

Observed
Flow inserted
State-updated
Simulation

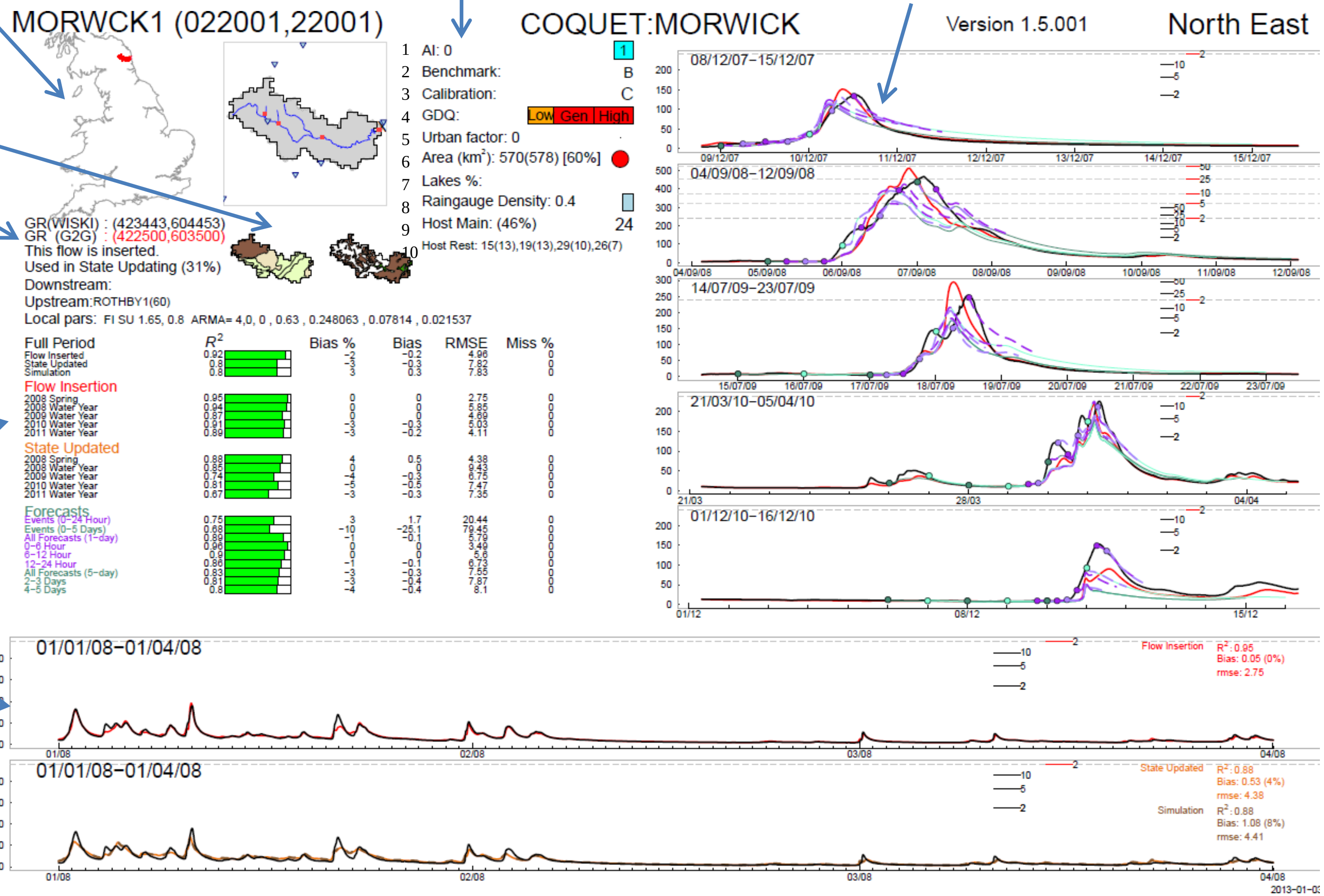


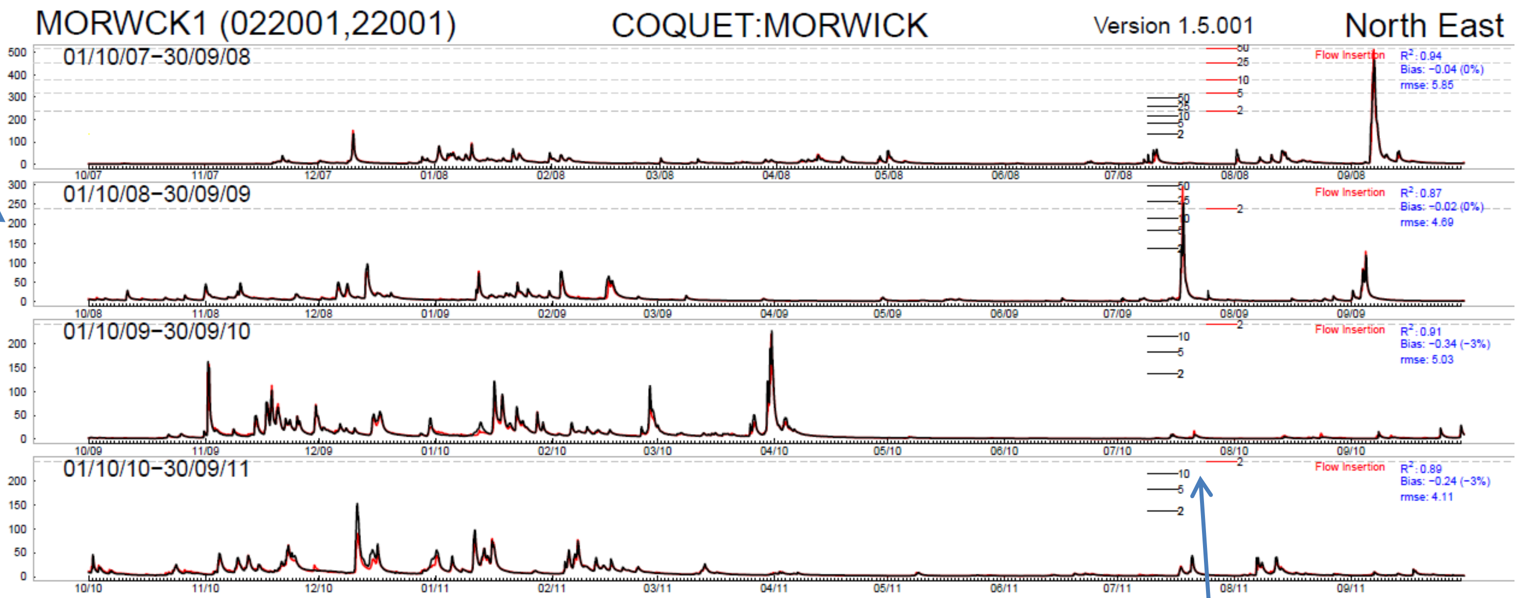
Figure 4.7 (continues)

Graphs of all water years

2008-2011 with flow insertion.

Observed (black line)

Flow inserted (red line)



Probability of Detection (POD) and False Alarm Ratio (FAR)

Shown for G2G-based thresholds of Q(2), Q(2)×3/4, Q(2)×1/2. Forecast windows of 0-24 hours (1-day forecast) and 2-3 and 4-5 day (5-day forecast) are shown.

Not shown if Q(T) values are missing

Green POD is good,

Red FAR is bad, white is good

"No." = Number of observed events

"+" = Number of observed events correctly forecast

"-"= Number of events forecast that did not occur

0−24 Hour Forecast

% Q(2)	Threshold	No.	+	−	POD	FAR
100	239.8	14	10	0	0.71	0
75	179.8	22	19	0	0.86	0
50	119.9	51	37	2	0.73	0.05

2−3 Day Forecast

% Q(2)	Threshold	No.	+	−	POD	FAR
100	239.8	14	10	0	0.71	0
75	179.8	22	19	0	0.86	0
50	119.9	51	37	2	0.73	0.05

4−5 Day Forecast

% Q(2)	Threshold	No.	+	−	POD	FAR
100	239.8	14	10	0	0.71	0
75	179.8	22	19	0	0.86	0
50	119.9	51	37	2	0.73	0.05

Q(T) values (T = 2, 5, 10, 25, 50 years depending on scale)
QMED (2 year) is always shown if it exists
Q(T) values are shown in **black for FEH** and **red for G2G (these are the ones used in FFC displays)**

Figure 4.7 Example of the new Performance Summary that includes a forecast-mode assessment (FFC example shown).

5 G2G model assessment

5.1 Introduction

The G2G model assessment for RRCs has been performed over Phases 1 and 2 of the project. Within Phase 1 the assessment used sites in England & Wales only and considered (i) simulation-mode assessment using observed rainfall (Section 5.2), (ii) forecast-mode assessment: foreknowledge of observed rainfall (Section 5.3), and (iii) forecast-mode assessment: use of UKV deterministic rainfall forecasts (Section 5.4).

Within Phase 2 the existing Phase 1 assessment over England & Wales has been extended to cover Scotland and a short summary for Scotland is given in each section. Furthermore the main focus within Phase 2 has been to assess the forecast-mode performance of G2G for RRCs using blended ensembles. This assessment is provided in detail in Section 5.5, in terms of both case-studies and longer-term assessments.

The time periods used for each type of assessment is summarised in Table 5.1 .

Table 5.1 Summary of time periods used for each type of assessment.

Assessment type	**England & Wales**	**Scotland**
Simulation-mode: observed rainfall (Section 5.2)	4 water years October 2007 – September 2011	4 water years October 2008 – September 2012
Forecast-mode: foreknowledge of observed rainfall (Section 5.3)	4 water years October 2007 – September 2011	4 water years October 2008 – September 2012
Forecast-mode: UKV deterministic rainfall forecasts (Section 5.4)	1 Water year October 2010 – September 2011	1 Water year October 2010 – September 2011
Forecast-mode: blended ensemble rainfall forecasts (Section 5.5)	7.5 months 10 May to 31 December 2012	4.5 months 10 May to 30 September 2012

It should be noted that operationally the Flood Forecasting Centre use a snowmelt scheme with G2G over England and Wales based on NWP snowmelt, snowfall and rainfall forecasts. Unfortunately the necessary supporting NWP datasets are not available for the assessment period so no snowmelt modelling is applied over England & Wales for this study. The Scottish Flood Forecasting Service use an integrated G2G snowmelt hydrology scheme that uses NWP screen level air temperature data to drive a simple temperature-excess melt formulation. This ensures water balance is maintained within the G2G hydrological modelling process. The necessary temperature data are only available in the G2G dataset archive for the 2009/10 winter so as it is not available for the whole assessment period it has not been applied in this study. However, snowmelt is only an important factor for a small fraction of rapid response catchment flood events so its omission from most of the modelling assessment period is not a major concern. Also none of the specific case study floods selected have a snowmelt element.

5.2 Simulation-mode assessment using observed rainfall

This section assesses the G2G simulation-mode runs using raingauge-only rainfall as input. In this mode, G2G simulations at a given site use flow-insertion from upstream sites, and state-correction from all sites, throughout the period, as discussed in Section 4.2.

England & Wales

The England & Wales analysis used the four water years within 2007 to 2011. The pooled performance measures for the simulation-mode G2G modelled flows, grouped by catchment area, are given in Figure 5.1. Some clear and expected findings come out of this analysis. Firstly, the pure simulation model performance (lilac bars) improves with increasing area. Secondly, although the state-updating is conservative and provides a “slow” adjustment to the modelled flow, it provides a significant improvement across all sites (green bars). Finally, using flow-insertion in addition to state-updating (blue bars) only benefits sites that have upstream sites feeding into them. Figure 4.2 provides the breakdown of which sites do or do not have upstream sites. As expected, since most of the sub 50km^2 catchments are headwater sites, this grouping only has a minor improvement due to flow-insertion whereas the improvement is significant for the larger catchments which can have gauged flows from many upstream sites feeding in.

The simulation-mode performance across assessment sites is mapped in Figure 5.2, for both pure simulation and with flow insertion (FI) and state updating (SU) switched on. In these “Performance Maps”, the gauged catchment or sub-catchment area is coloured according to the R^2 Efficiency for the gauged outlet. Also the maps emphasise certain catchments by making the others transparent: in this case it is emphasising the small (up to 50km^2) catchments. The hachured areas indicate gauged catchments not suitable for assessment (e.g. a reservoir outlet or significant artificial influences apply).

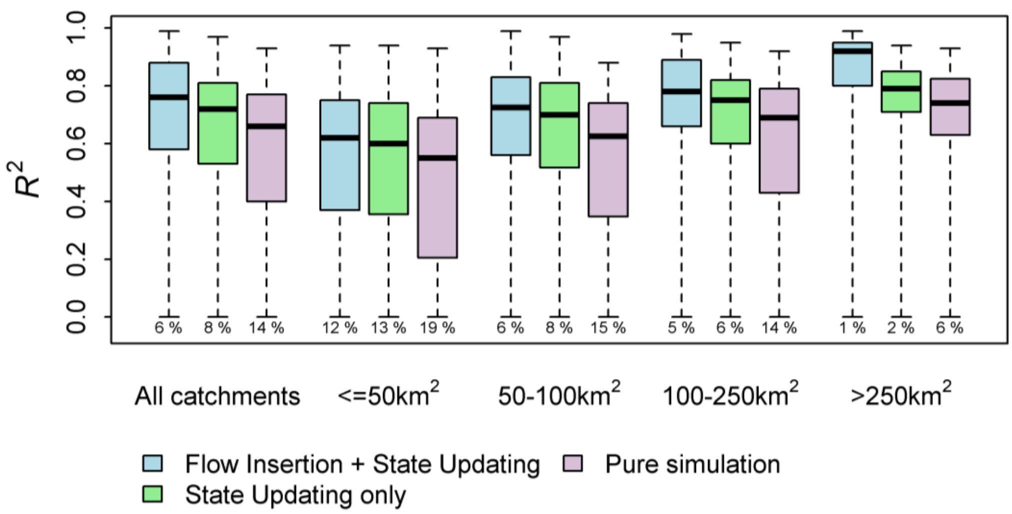
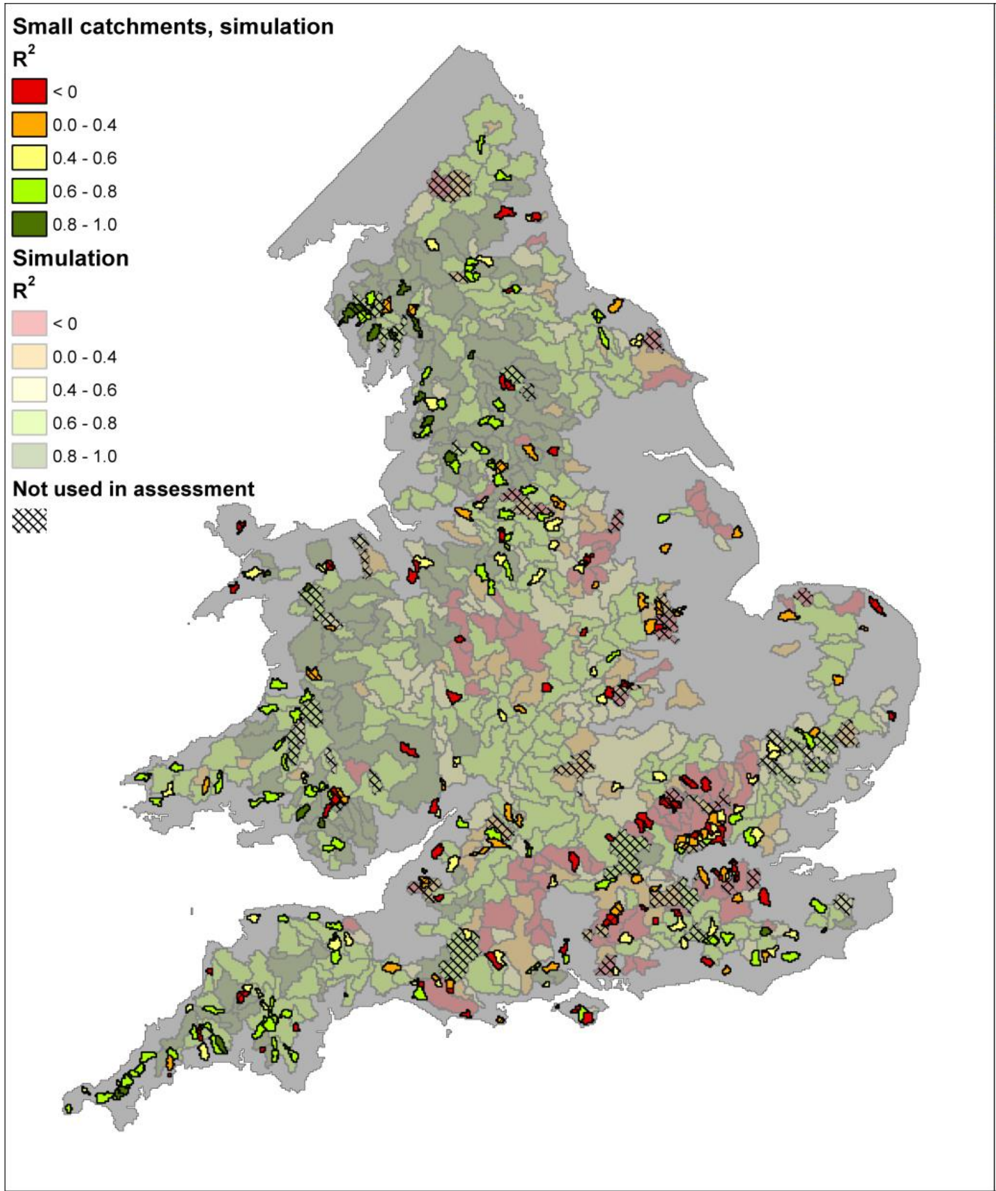


Figure 5.1 England & Wales. Box and whisker plots showing the median, inter-quartile range and the max/min values of R^2 Efficiency for simulation-mode G2G modelled flows for 2007-2011, grouped by all catchments and catchment area. The percentages indicate the fraction of sites with a negative R^2 Efficiency.



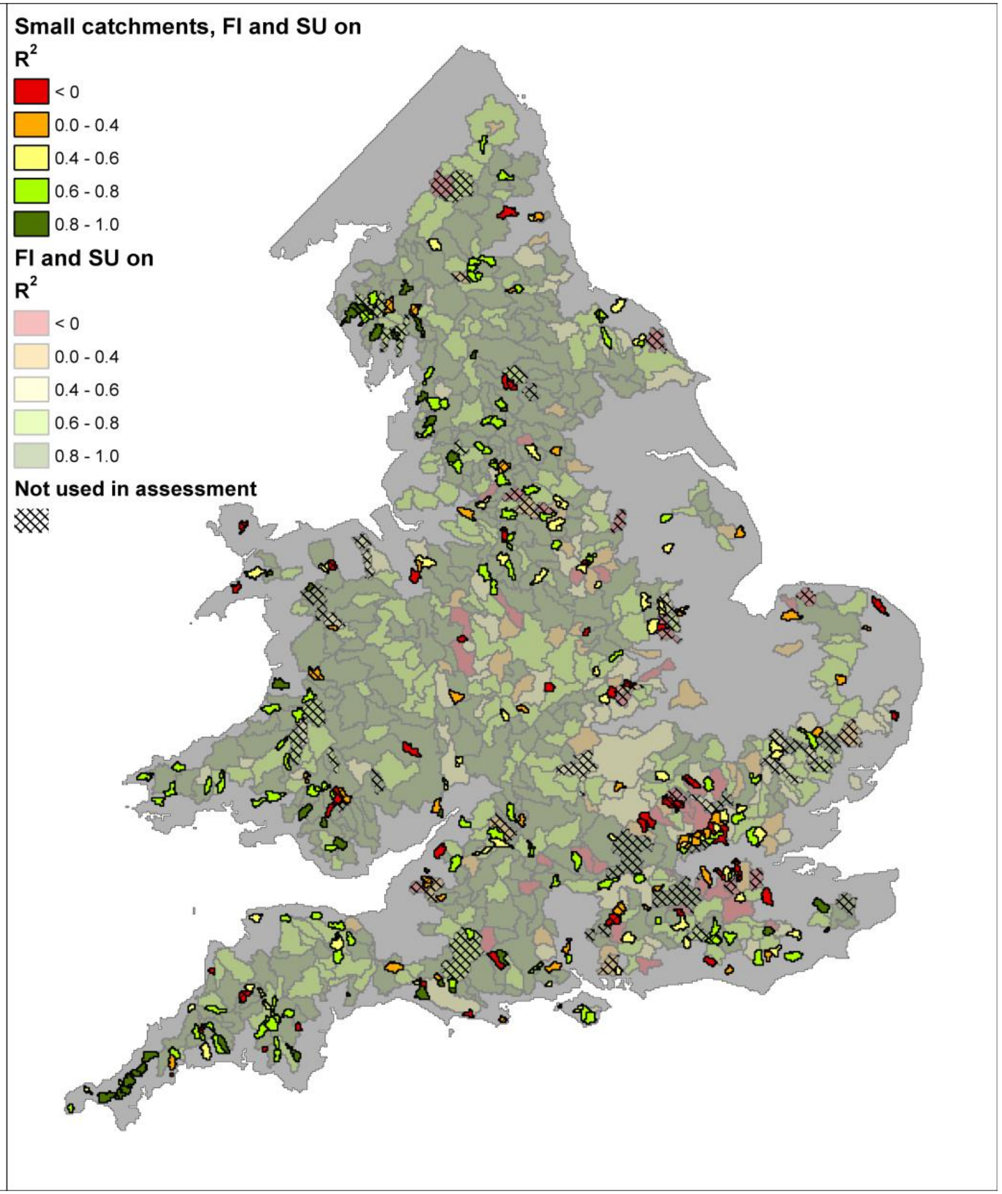


Figure 5.2 England & Wales. R^2 Efficiency performance of G2G in simulation-mode for 2007-2011. Small catchments (<50km^2) are highlighted. Left map: pure simulation performance. Right map: includes flow insertion (FI) and state updating (SU).

The Performance Maps shown in Figure 5.2 provide a useful insight into how the performance of the model and benefit of state-updating and flow-insertion varies with location. The pure simulation-mode performance is indicative of how the model performs for ungauged locations (although the routing parameters will have been calibrated for each site). Focussing on the small catchments, the maps generally show good performance over South West, Wales, North East and North West. Performance over Midlands, Southern (including Thames) and Anglian is more mixed due to a combination of reasons such as urbanisation, artificial influences, groundwater-dominated catchments and low relief. Figure 5.3 illustrates this by showing pooled performance by region for headwater catchments less than 50 km². It is also useful to assess the performance for individual water years to understand the consistency of performance over time. Box and whisker plots for the individual water years are presented in Figure 5.4 and compared with those over all years.

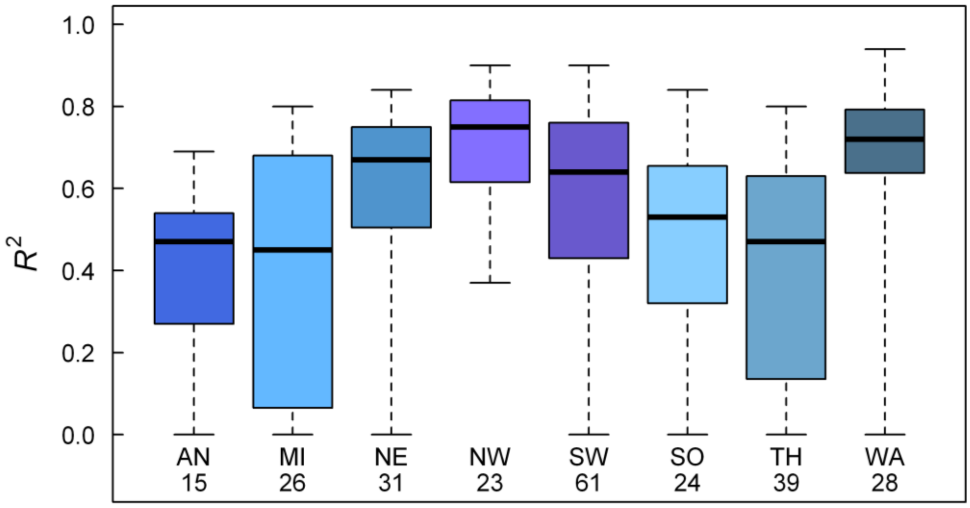


Figure 5.3 England & Wales. Box and whisker plots showing the median, inter-quartile range and the max/min values of R^2 Efficiency for simulation-mode G2G modelled flows (with flow-insertion and state-updating) for 2007-2011. Results are only for headwater catchments less than 50 km² and are grouped by region. The number of sites per region is also given.

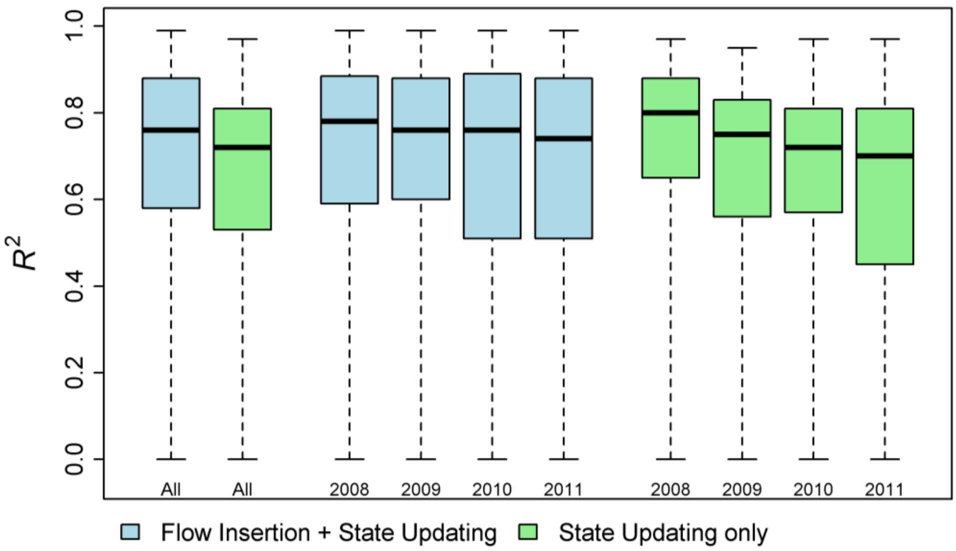


Figure 5.4 England & Wales. Box and whisker plots showing the median, inter-quartile range and the max/min values of R^2 Efficiency for simulation-mode G2G modelled flows. These are for the four individual water years and for all years together (All), and pooled over all assessment sites.

Encouragingly the median values are relatively robust, particularly when flow-insertion and state-updating are applied. The lower quartile does vary significantly between years. Further investigation has suggested this is due to occasional periods of high rainfall that are not extreme enough to have failed the automated raingauge quality-control checks but may still be erroneous. An example is given from the Performance Summary for the Leam at Kite Hardwick showing very high modelled flows in August 2010 which affected several gauged locations. Further investigation into these periods is needed.

The results have been analysed according to the urban coverage within a catchment. Scatter plots of R^2 Efficiency of the G2G flow simulations versus the urban coverage are given in Figure 5.6 grouped by catchment area. The smoothed line of local

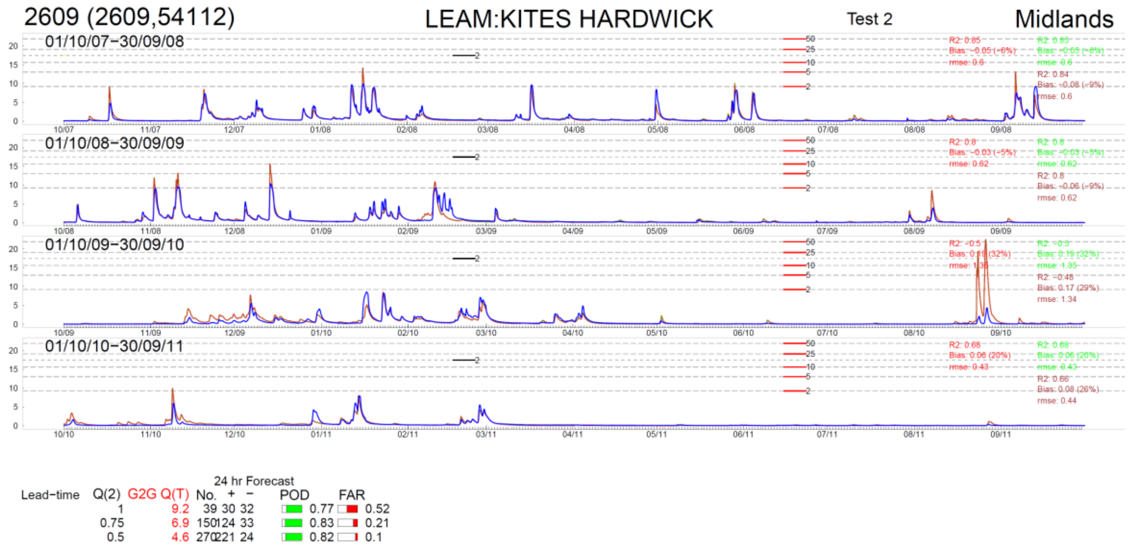


Figure 5.5 Performance Summary showing the simulation-mode G2G modelled flows for Leam at Kites Hardwick.

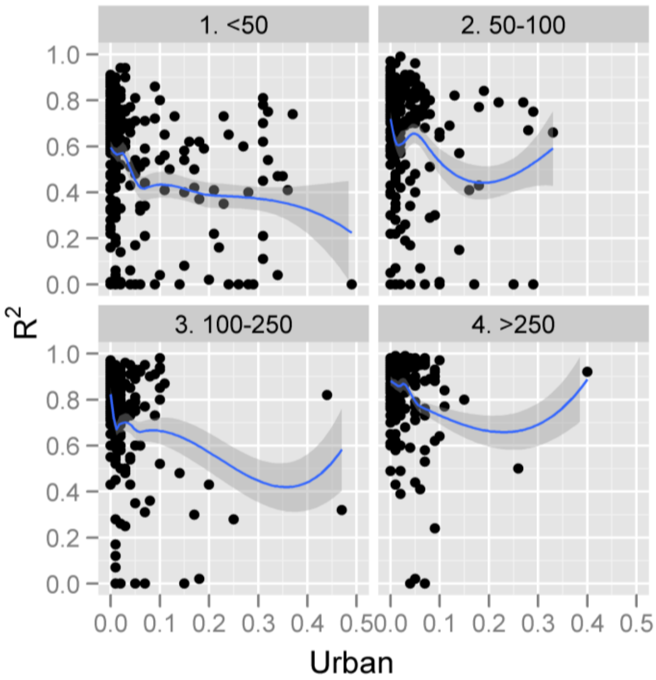


Figure 5.6 England & Wales. Scatter plots of R^2 Efficiency for simulation-mode G2G modelled flows (with flow-insertion and state-updating) versus catchment urban coverage grouped by catchment area (km²).

regression is plotted with confidence intervals. As is often the case, it is very hard to obtain a clear correlation between model performance and any one catchment property. However there appears to be a weak signal for catchments under 100km^2, especially at the lower end of urbanisation, suggesting improved modelling in urban areas could benefit model performance.

Scotland

The Scotland analysis used the four water years within 2008 to 2012. The pooled performance measures for the simulation-mode G2G modelled flows, grouped by catchment area, are shown in Figure 5.7. The trend for the “all catchments” group is similar to that for England & Wales (Figure 5.1) with model performance improving as state-updating and then flow-insertion are included. It is also noticeable that the overall model performance is generally better for Scotland for all catchment areas below 250 km^2 and is particularly evident in the smaller spread at the lower end of performance (see lower quartile and whisker). This is partly to be expected since the hydrological behaviour of the rivers in Scotland is more similar to the North East, North West, Wales and South West regions of England and Wales where G2G also performs well (Figure 5.3).

In contrast to England & Wales, the simulation-mode performance (lilac bars) starts to tail-off after 250 km^2 especially for the lower quartile sites. This is probably due to the fact that many of the larger catchments drain from the highlands where the raingauge estimates are less certain due to the relatively sparse density. Also there can be effects from lochs and hydro-power sites that are not incorporated into the model and may impact these sites. As expected, the benefit of flow-insertion is greatest for the large catchment areas greater than 250 km^2. In fact there is virtually no difference for sites less than 100 km^2 because only one site has an upstream gauge that could benefit its performance (see Figure 4.2). Figure 5.8 gives a regional breakdown of the simulation model performance (with flow-insertion and state-updating applied) for the small catchments (less than 50 km^2) and confirms that all regions perform well. Overall the good model performance for sites less than 50 km^2 is very encouraging for using G2G in Rapid Response Catchments over Scotland.

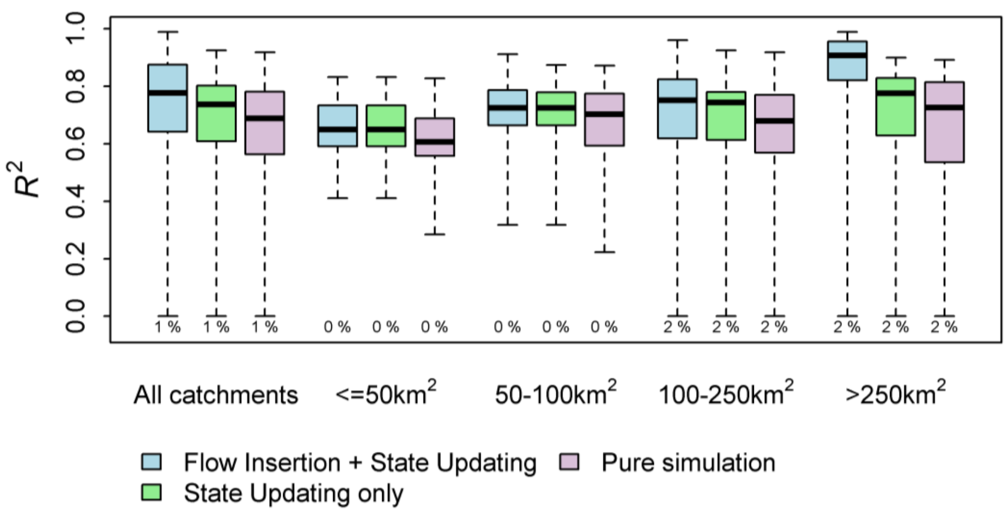


Figure 5.7 Scotland. Box and whisker plots showing the median, inter-quartile range and the max/min values of R^2 Efficiency for simulation-mode G2G modelled flows over Scotland. These are over the four water years spanning 2008-2012 and grouped by all catchments and catchment area. The percentages indicate the fraction of sites with a negative R^2 Efficiency.

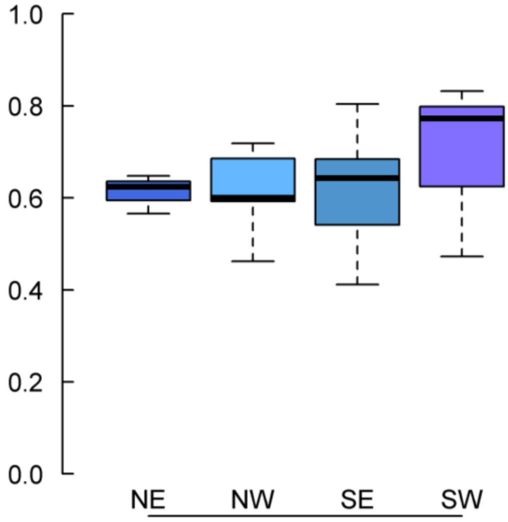


Figure 5.8 Scotland. Box and whisker plots showing the median, inter-quartile range and the max/min values of R^2 Efficiency for simulation-mode G2G modelled flows (with flow-insertion and state-updating) for 2008-2012. Results are only for headwater catchments less than 50 km^2 and are grouped by region.

5.3 Forecast-mode assessment: foreknowledge of observed rainfall

Assessments of performance in this section still use observed rainfall but evaluate the G2G model in *forecast-mode* where flows up to the time the forecast is constructed are used for data assimilation via direct flow-insertion, state-correction and ARMA error-prediction.

England & Wales

The full four water year period (2007-2011) has been used for this assessment. The first analysis aims to understand how the forecast-mode performance relates to the simulation-mode at different lead-times and is presented in Figure 5.9. *A priori* it is expected that the forecast accuracy falls off with increasing lead-time, eventually

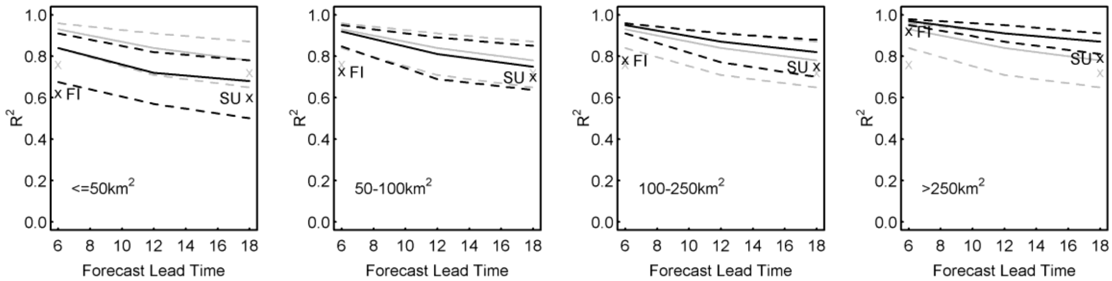


Figure 5.9 England & Wales. Median (solid lines) and inter-quartile range (dashed lines) of R^2 Efficiency for forecast-mode G2G modelled flows at different lead-times (black lines) and grouped over different catchment sizes. Those grouped over all catchments are in grey. Simulation-mode G2G modelled flows with flow-insertion and state-updating (FI) and state-updating only (SU) are indicated by crosses (median score). All four water years (2007-2011) are used.

returning to the simulation-mode level. Interestingly, the plots clearly show there is a lasting benefit of the data assimilation, with forecast-mode efficiencies being better than those in simulation-mode even out to the 12-24 hr forecast window (plotted at 18 hrs), and even with state-updating on. As expected, the forecast-mode performance increases with catchment size as the larger catchments benefit most and longest from upstream flow insertion, and consequently performance drops off with lead-time sharpest for the smaller catchments.

To understand model performance further, it is also useful to consider the difference between the “headwater” catchments which do not have upstream sites and the “downstream” catchments that do have upstream sites and can benefit from upstream flow-insertion. This is presented in Figure 5.10 for both simulation-mode and forecast-mode assessments. The results confirm that “headwater” sites perform significantly better in forecast-mode through use of ARMA error-prediction. For the “downstream” sites, there is little difference in the median performance of simulation-mode or forecast-mode performance for sites larger than 50km^2. Interestingly, for the “downstream” catchments smaller than 50km^2 there is considerable benefit to the median performance when running in forecast-mode.

This “headwater” catchment performance is explored further in Figure 5.11 which gives an indication of the benefits of forecast-mode performance for each region and also the difference between smaller catchments and larger catchments. It also gives an insight into the spread of performance and confirms the regions with poorest simulation-mode performance benefit most from the ARMA error-predicted forecasts. To gain a further appreciation of the geographical spread of forecast-mode performance, Figure 5.12 maps R^2 Efficiency by colour-coding the sub-catchment areas to which the performance assessment relates.

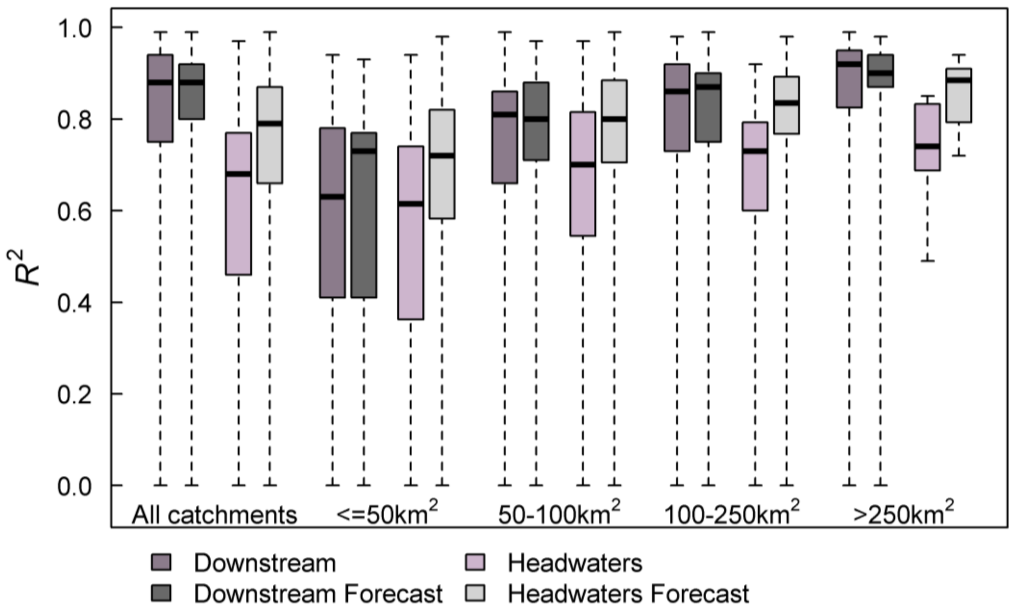


Figure 5.10 England & Wales. Box and whisker plots showing the median, inter-quartile range and the max/min values of R^2 Efficiency for G2G modelled flows produced in (i) simulation-mode (with flow-insertion and state-updating), and (ii) forecast-mode (with observed rainfall) using the full 24 hr forecast over the four water years. Assessments are grouped by catchment sizes and split between headwater (no upstream site) and downstream (has an upstream site) sites.

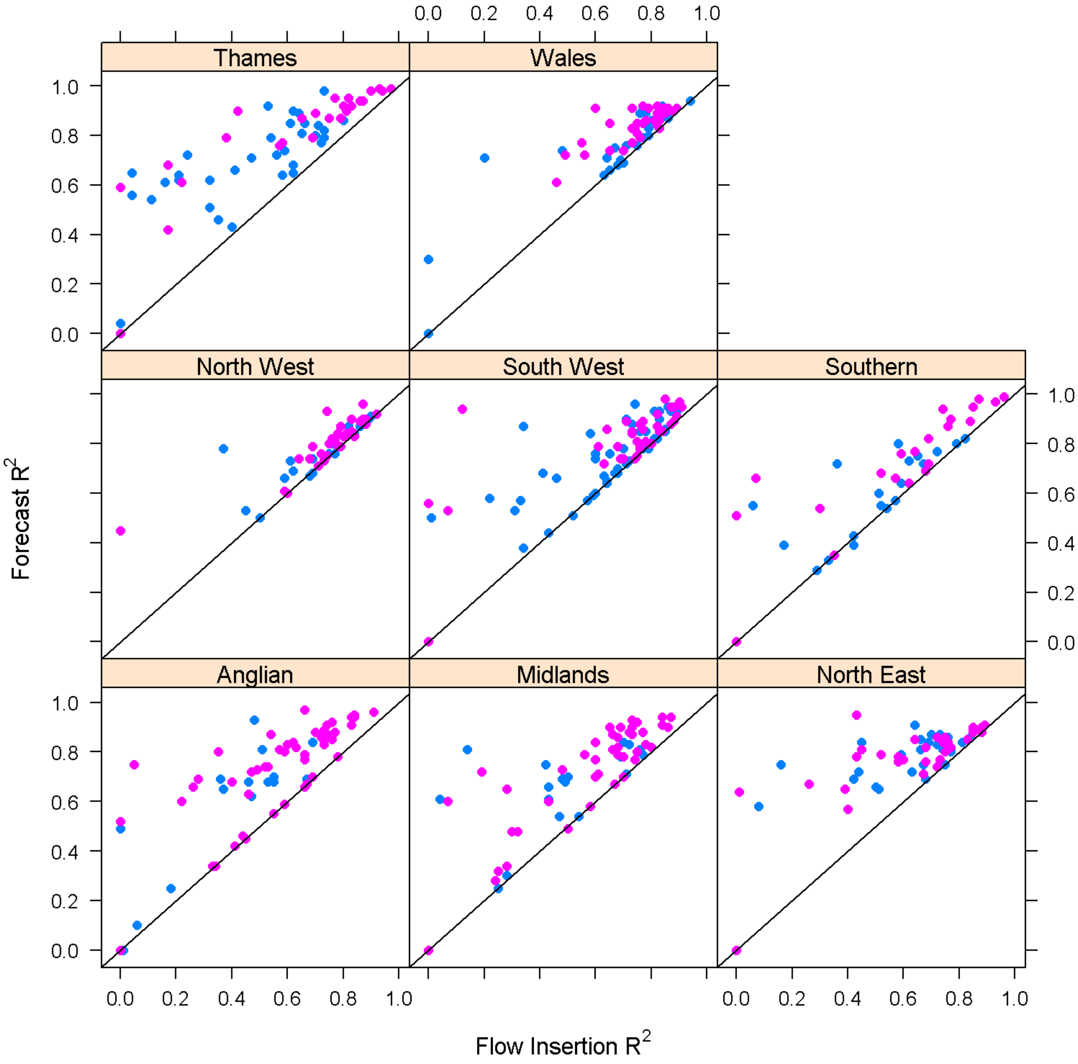


Figure 5.11 England & Wales. Scatter plots comparing the R^2 Efficiency for G2G modelled flows produced in (i) simulation-mode (with state-updating and flow insertion, x-axis), and (ii) forecast-mode (with observed rainfall, y-axis) using the full 24 hr forecast over the four water years. Only headwater catchments (no upstream site) are shown: blue dots are up to 50km^2, pink dots are over 50km^2.

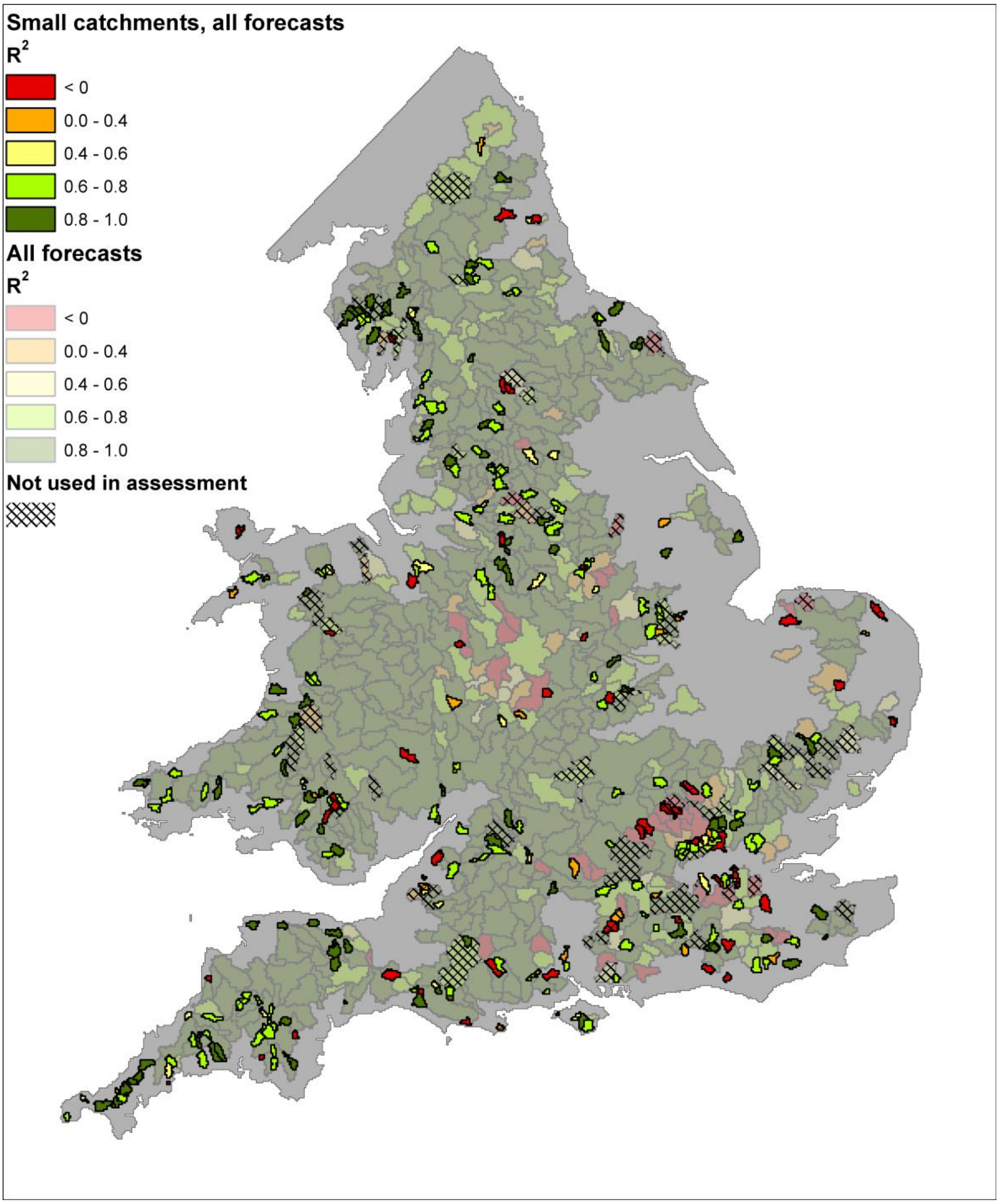


Figure 5.12 England & Wales. Forecast-mode G2G model Performance Map (with observed rainfall) using all 24 hr forecasts over the four water years (2007-2011). Catchments/sub-catchments coloured by R^2 Efficiency and small catchments (up to 50km^2) are highlighted.

Scotland

The full four water year period (2008-2012) has been used for this assessment over Scotland. Figure 5.13 shows how the forecast-mode performance relates to the simulation-mode at different lead-times. As with the England & Wales assessment (Figure 5.9), the forecast accuracy falls off with increasing lead-time and there is a lasting benefit of the data assimilation, with forecast-mode efficiencies being better than those in simulation-mode even out to the 12-24 hr forecast window (plotted at 18 hrs), and even with state-updating applied. There are also some noticeable differences to the performance seen for England & Wales. In particular, for the smaller catchments below 100 km^2, the modelled flows for Scotland have higher median R^2 Efficiencies for short lead-times and less spread in terms of efficiencies over all lead-times. This relatively better performance for small catchments in Scotland is a reflection of the good simulation-mode flows for these sites as discussed in Section 5.2. In comparison to England & Wales, the median efficiency values for larger Scottish catchments (> 100 km^2) return much closer to those in simulation-mode with state-updating applied (SU). This is a consequence of Scotland having far fewer nested gauged locations compared to England & Wales where flow-insertion can have a significant and long lead-time benefit for large catchments.

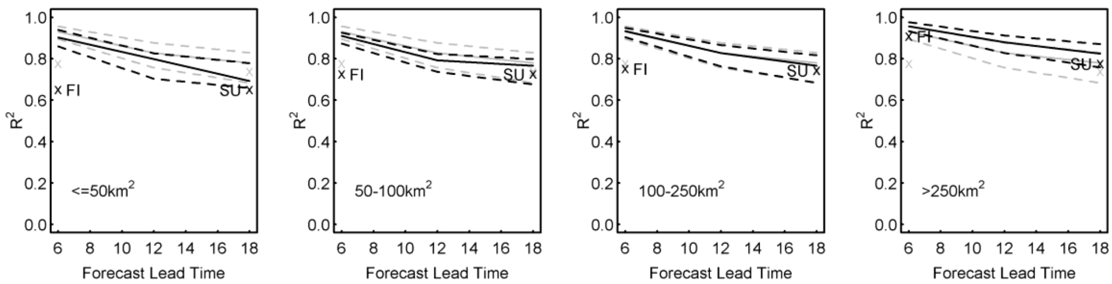


Figure 5.13 Scotland. Median (solid lines) and inter-quartile range (dashed lines) of R^2 Efficiency for forecast-mode G2G modelled flows at different lead-times (black lines) and grouped over different catchment sizes. Those grouped over all catchments are in grey. Simulation-mode G2G modelled flows with flow-insertion and state-updating (FI) and state-updating only (SU) are indicated by crosses (median score). All four water years (2008-2012) are used.

The assessment of "headwater" catchments which do not have upstream sites is explored in Figure 5.14. The results confirm that "headwater" sites in all regions perform significantly better in forecast-mode through use of ARMA error-prediction. Northern areas benefit most which is likely to be a reflection of the lower raingauge densities giving lower simulation-mode performance and more scope for benefits from assimilating flow observations for these sites. The geographical spread of forecast-mode performance is presented in Figure 5.15 which maps R^2 Efficiency by colour-coding the sub-catchment areas to which the performance assessment relates. This shows very good performance (dark green) across almost all of the southern areas and eastern coast.

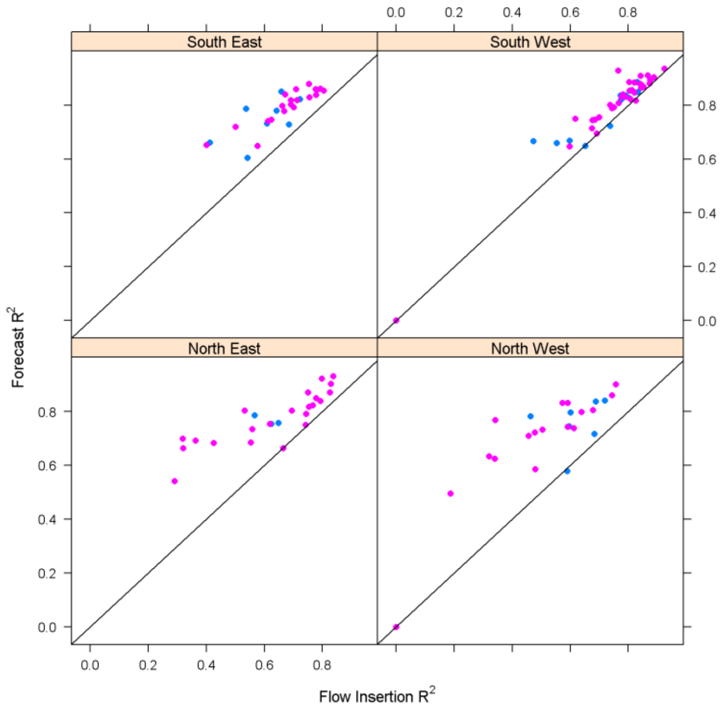


Figure 5.14 Scotland. Scatter plots comparing the R^2 Efficiency for G2G modelled flows produced in (i) simulation-mode (with state-updating, x-axis), and (ii) forecast-mode (with observed rainfall, y-axis) using the full 24 hr forecast over the four water years. Only headwater catchments (no upstream site) are shown: blue dots are up to 50km^2, pink dots are over 50km^2.

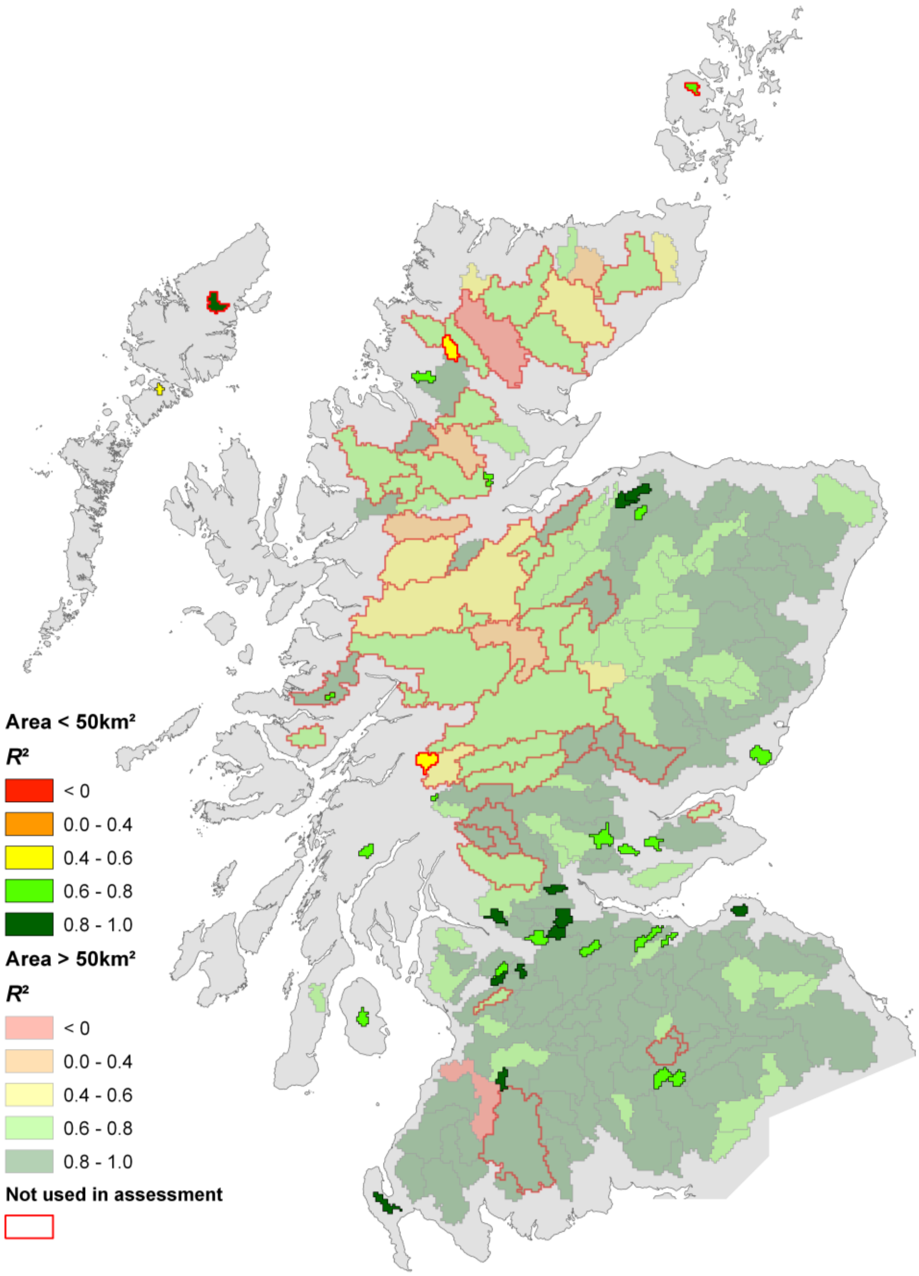


Figure 5.15 Scotland. Forecast-mode G2G model Performance Map (with observed rainfall) using all 24 hr forecasts over the four water years (2008-2012). Catchments/sub-catchments coloured by R^2 Efficiency and small catchments (up to 50km^2) are highlighted.

5.4 Forecast-mode assessment: use of UKV deterministic rainfall forecasts

This section compares the forecast-mode performance of G2G using NWP UKV forecast rainfall relative to using observed rainfall as foreknowledge of future rainfall. Due to the availability of the NWP data, the period of study is the 2010/11 water year.

England & Wales

Over England & Wales, Figure 5.16 compares the UKV-based G2G forecasts (red lines) with the raingauge-only based forecasts (black lines) with assessments grouped by catchment size as before. A strong deterioration in performance is seen beyond 12 hours lead-time for all catchment sizes and is particularly severe for the smaller catchments most relevant to RRC. This is partly to be expected as the skill of the deterministic NWP forecasts in predicting the location of rainfall (particularly when of convective origin) drops off with lead-time and will have most impact for smaller catchments. This is where the blended ensembles are expected to have most benefit and this is assessed in Section 5.5.

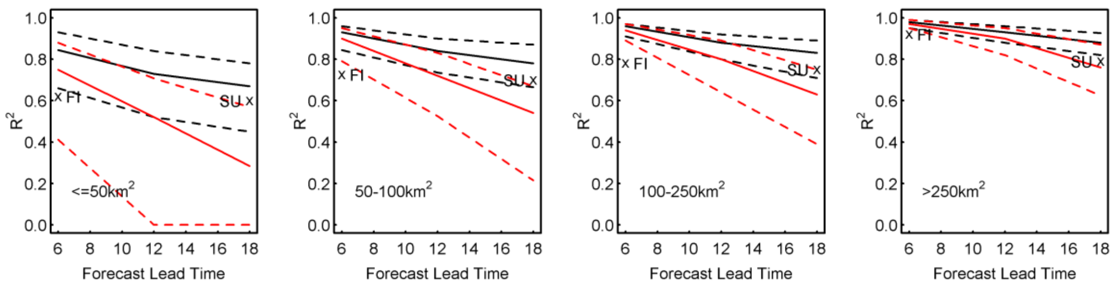


Figure 5.16 England & Wales. Median (solid lines) and inter-quartile range (dashed lines) of R^2 Efficiency for forecast-mode G2G modelled flows for different lead-times using observed rainfall (black lines) and UKV NWP (red lines) grouped over different catchment sizes.

Further insight into how the spread of model performance evolves with lead-time is presented in Figure 5.17. These scatter plots show the raingauge-based G2G forecast performance relative to the NWP UKV based G2G forecasts. These clearly show that beyond the first 6 hours, a reasonable number of NWP based G2G forecasts drop off rapidly and giving R^2 Efficiency values below zero. The plots also show that for some locations G2G forecasts are still performing very well, even at long lead-times.

Some example forecasts are given in Figure 5.18 for two locations where G2G does very well when using raingauge-only rainfall as input but less well when using NWP rainfalls. These are for two sites in North West Region. For the Sprint at Sprint Mill (draining the south east of the Lake District near Kendal), left plot, the UKV-based performance is variable with the large peak not captured well in the rainfall forecast. For the Roch at Albert Royds Bridge (right hand plot) use of UKV rainfalls appears to consistently forecast too much rain. An example rainfall forecast accumulation map for the relevant area is shown in Figure 5.19 serving to highlight the overestimation by the UKV rainfall compared to a HyradK raingauge-only estimate.

Looking at individual sites and forecasts provides some useful insights but a major interest is in the spatial variability of the forecast performance and whether there are any regional trends. This is explored in Figure 5.20 which displays a map of the differences in G2G performance between the two types of forecasts: the larger circles show where the UKV-based forecasts are performing poorly relative to the raingauge-only based forecasts. This clearly shows strong geographical biases, the largest and most notable groupings being along the northern and southern end of the Pennines and an area centred around London. The south coasts of England and Wales also shows some deterioration in NWP-based G2G forecast performance but to a lesser degree.

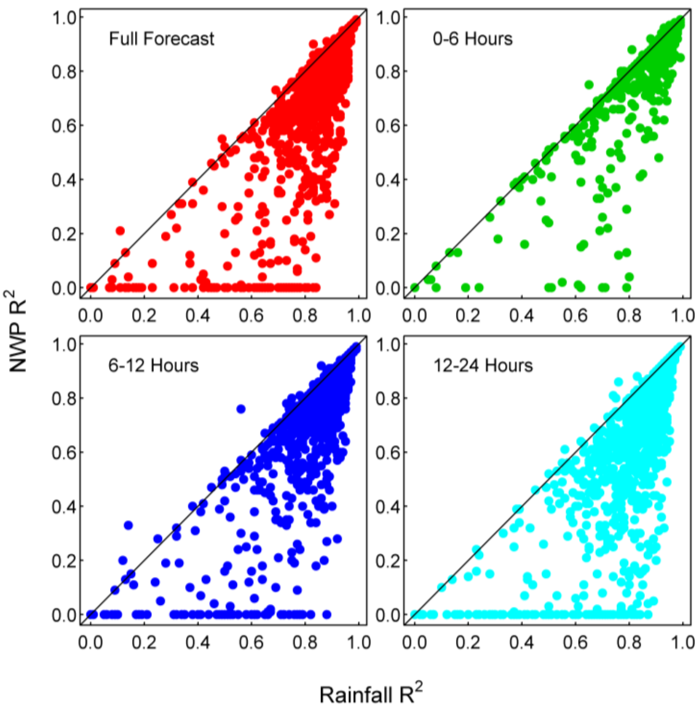


Figure 5.17 England & Wales. Scatter plots of R^2 Efficiency for forecast-mode G2G flows using raingauge-only observed rainfall and UKV NWP forecast rainfall grouped by forecast lead-time.

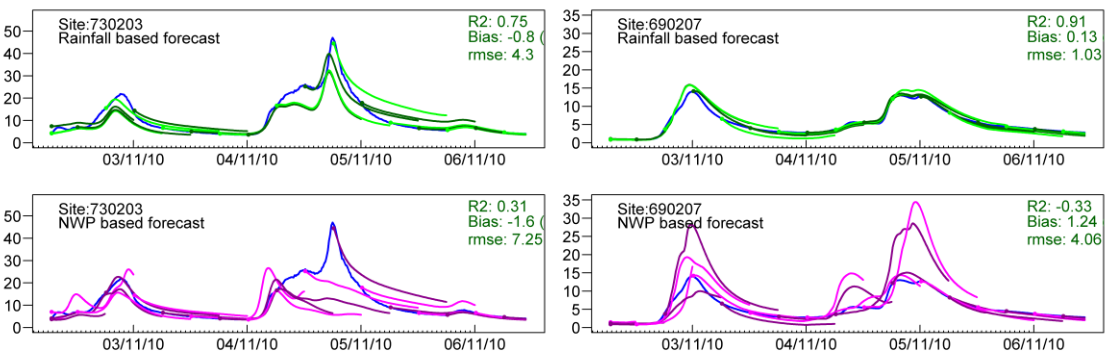


Figure 5.18 Example 24 hr lead-time G2G forecasts made every 6 hours using (i) UKV NWP forecast rainfall (bottom), and (ii) HyradK raingauge-only observed rainfall (top). Sprint Mill is shown on the left and Albert Royds Bridge on the right.

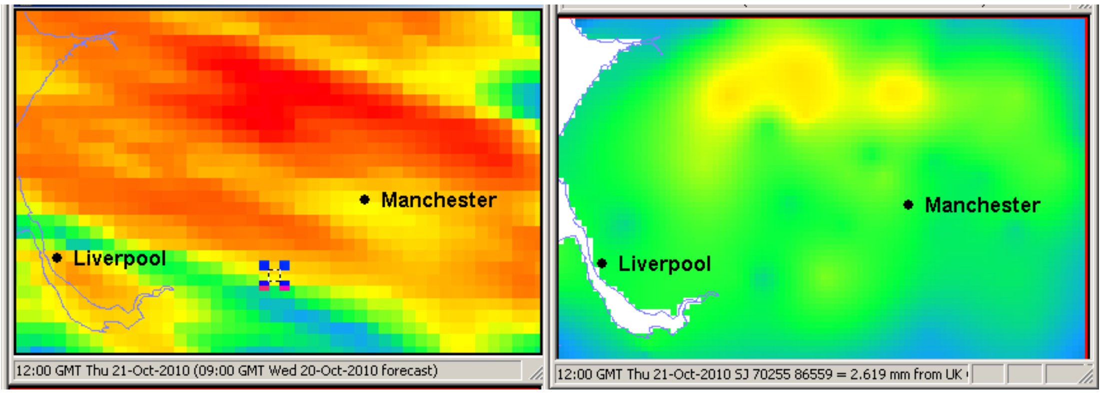


Figure 5.19 24-hr accumulated rainfalls ending 12:00 21 October 2010 using UKV NWP forecast rainfall (left) and HyradK raingauge-only observed rainfall (right).

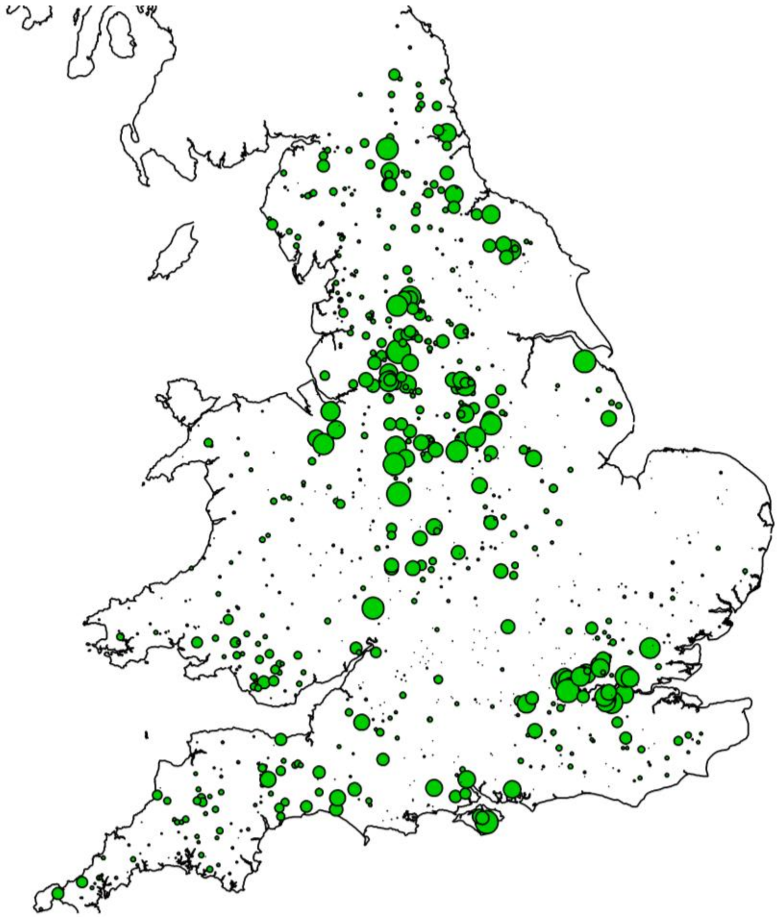


Figure 5.20 England & Wales. Difference in R^2 Efficiency between forecast-mode G2G flows using observed rainfall and using UKV NWP forecast rainfall. Large circles show where NWP G2G forecasts perform worse.

To investigate this further, NWP UKV forecast monthly rainfall accumulations are calculated (using the T+3 to T+27 part of each forecast and dividing by 4 to account for forecast overlaps) and then compared to equivalent raingauge-only accumulations. Such monthly rainfall accumulation maps for the period August 2010 to June 2011 are shown in Figures 5.21 and 5.22. Background information on elevation and Standard Annual Average Rainfall (SAAR) are presented in Figure 5.23 and provide additional context. Together, these provide some interesting insight into the possible reasons for the localised deterioration in the NWP-based G2G forecasts. In particular, there are several months where the lowland areas have too much rain forecast by the NWP: for example August to November 2010. This may have some correlation with the deterioration of G2G forecasts over London.

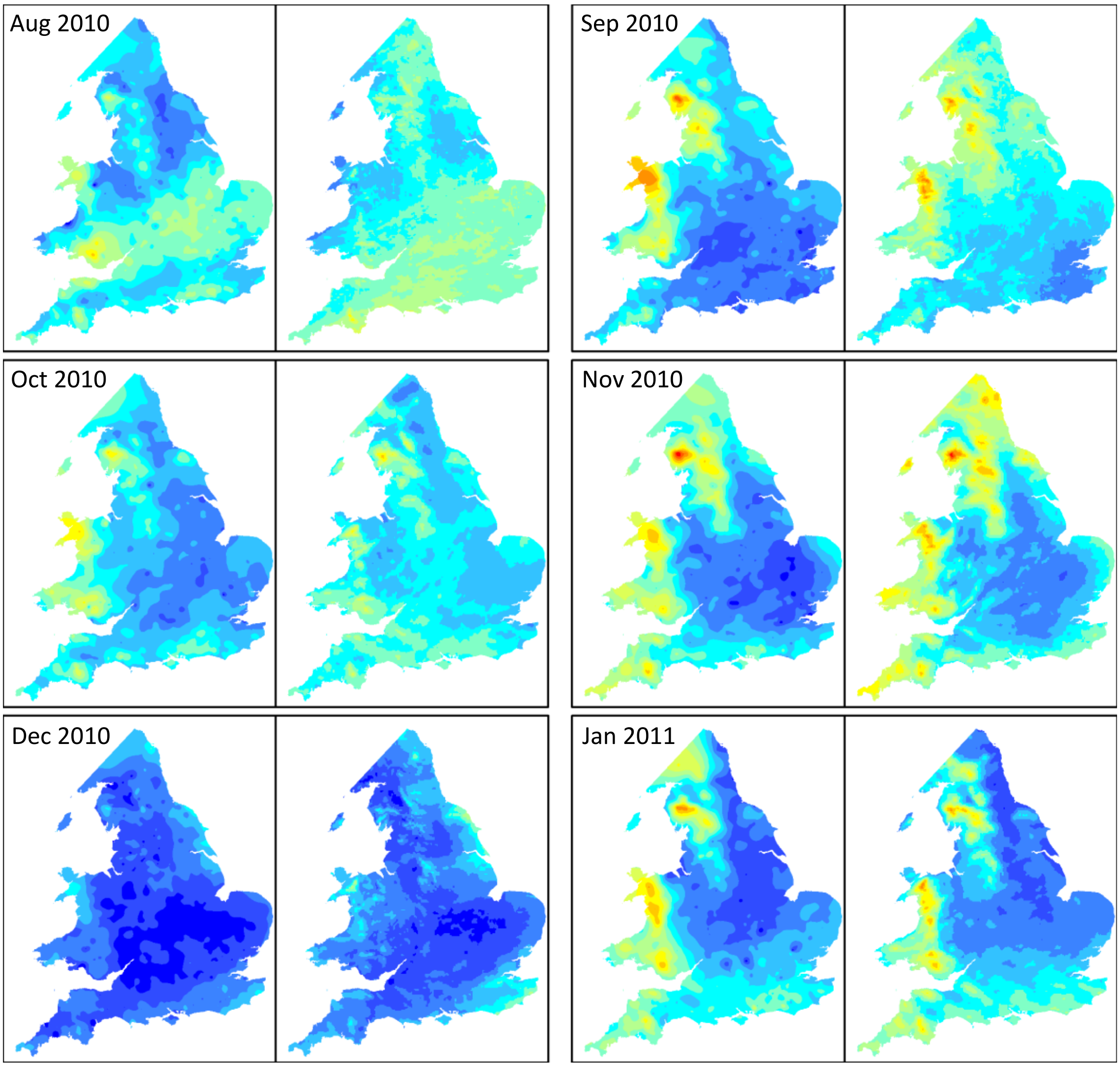


Figure 5.21 England & Wales. Accumulated rainfalls for 30 day periods covering each month from August 2010 until January 2011. Left hand plot of each pair use HyradK raingauge-only observed rainfall, right hand plot uses the UKV forecast. The legend scale is given in Figure 5.22.

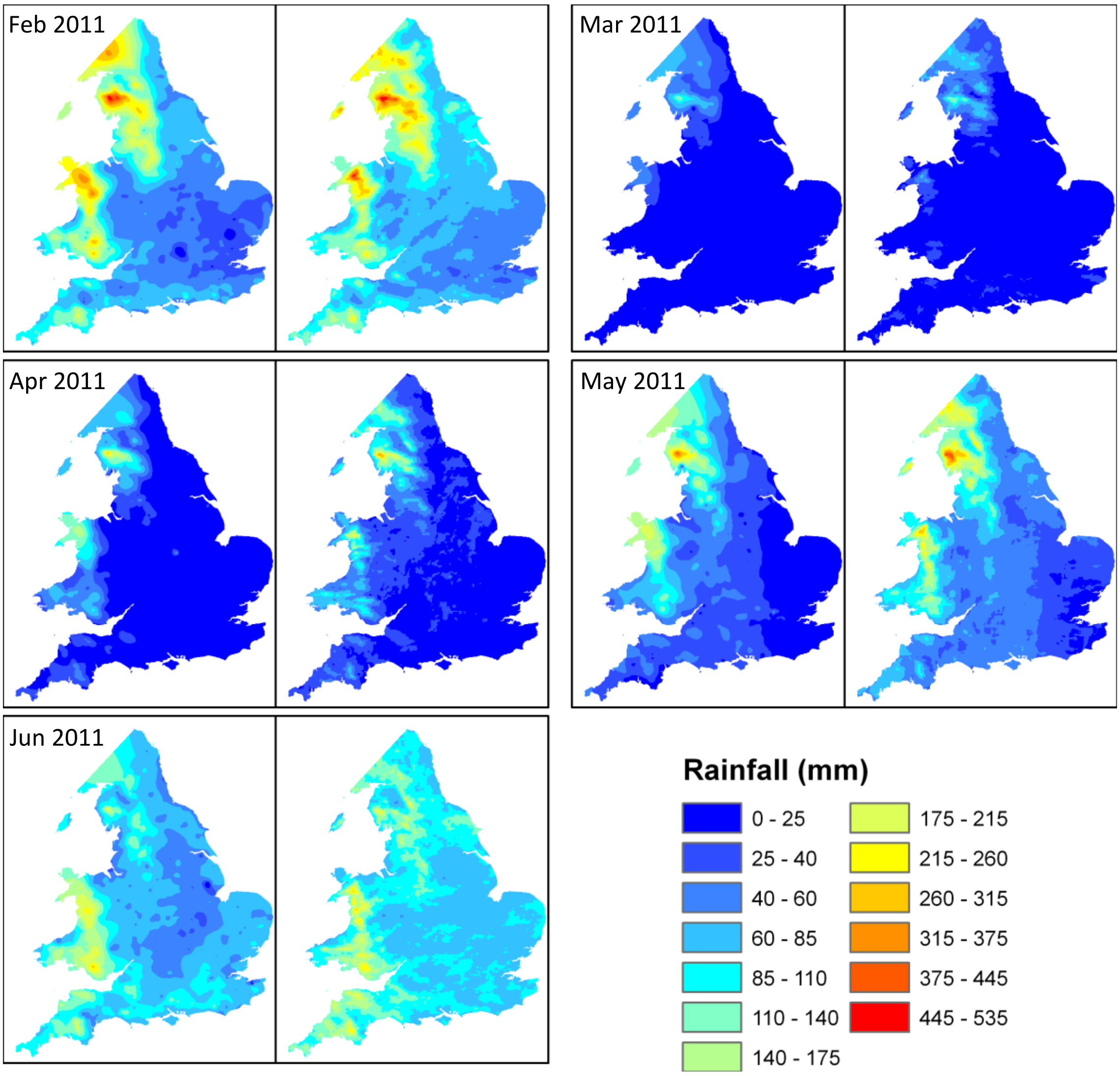


Figure 5.22 England & Wales. Accumulated rainfalls for 30 day periods covering each month from February 2011 until June 2011. Left hand plot of each pair uses HyradK raingauge-only observed rainfall, right hand plot uses the UKV forecast.

Another possible trend is related to high-elevation areas that are subject to orographic enhancement of the rain. The high-resolution UKV model represents the seeded-feeder mechanism of orographic enhancement well compared to previous, lower resolution (~12km) operational models. However, there is some evidence that there may be too much forecast precipitation in the lee of the high-elevation areas. For example, September 2010, November 2010 and March 2011 all show some evidence of the northern and southern areas of the Pennines having too much rain and corresponds with the spatial regions identified in Figure 5.20. Long-term 12 month accumulations for the raingauge-only and UKV rainfalls are given in Figure 5.24 and confirm reasonable agreement along the west coast which deteriorates further east and inland where the NWP tends to over-predict rainfall amounts.

This analysis of the UKV-based G2G forecasts has shown some very interesting regional behaviour that provides useful background for the blended ensemble analyses.

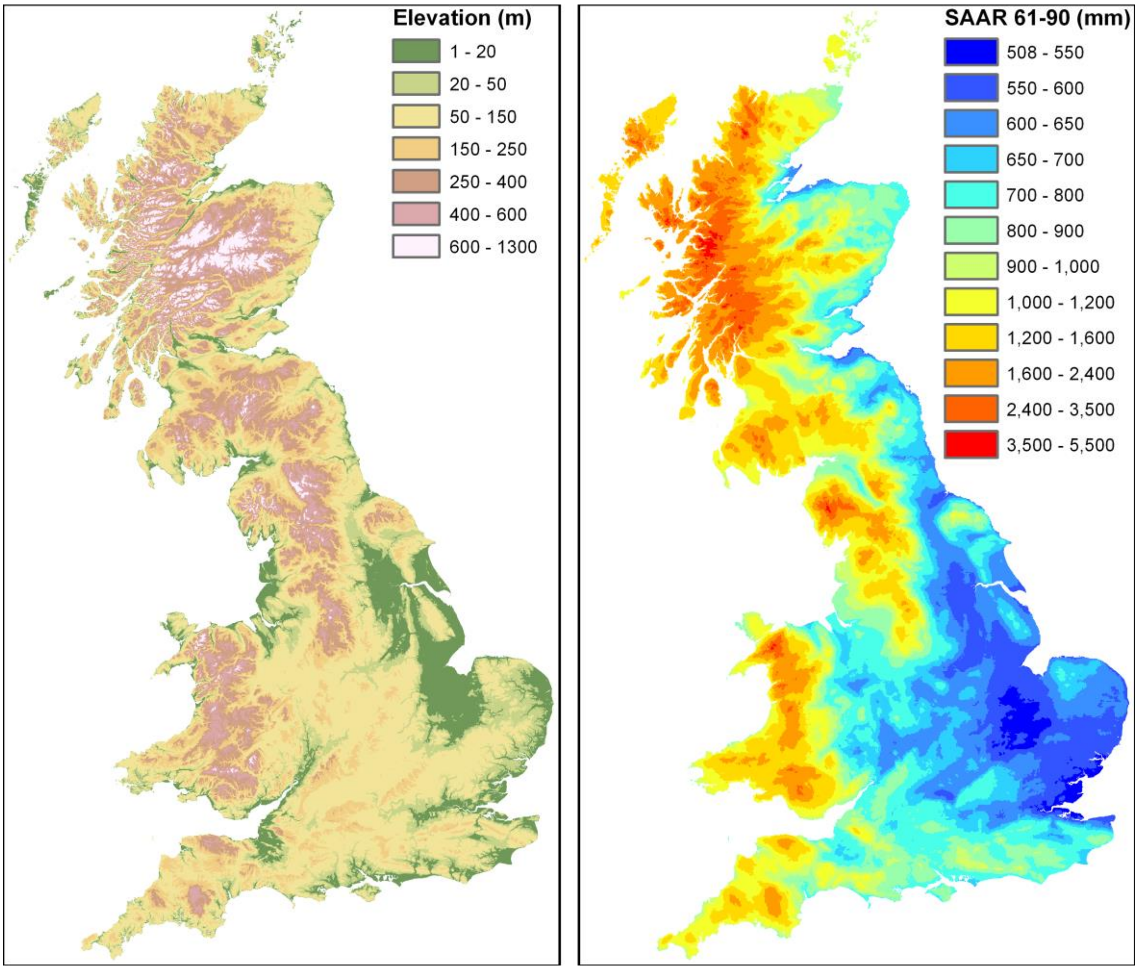


Figure 5.23 Elevation and Standard Annual Average Rainfall (SAAR) for England, Wales and Scotland. The period used for SAAR is 1961-1990.

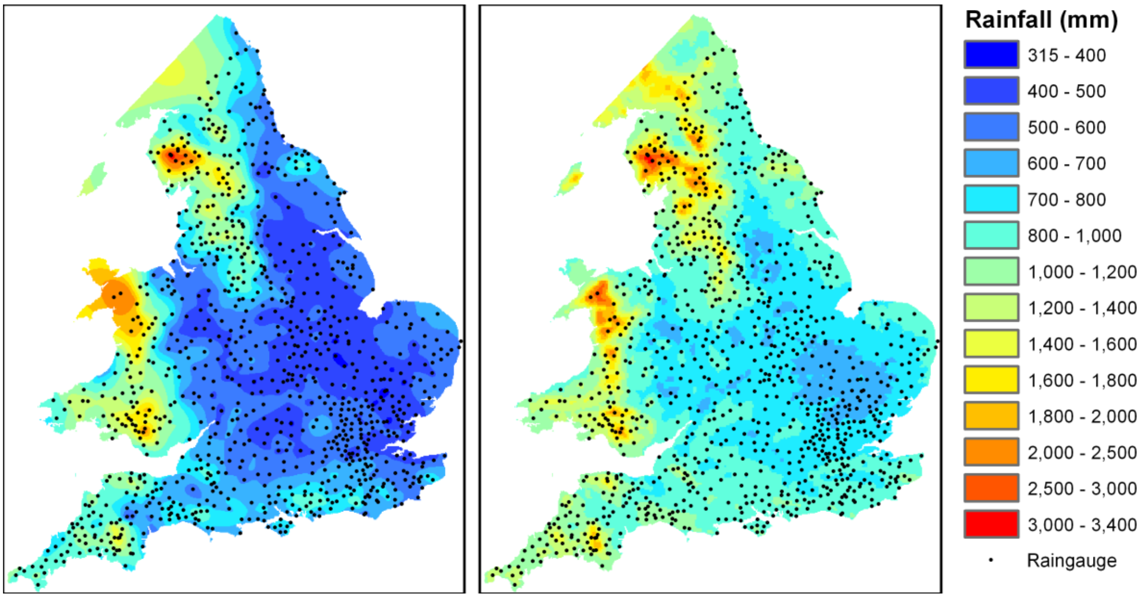


Figure 5.24 England & Wales. Accumulated rainfalls for August 2010 until July 2011. Left hand plot uses HyradK raingauge-only observed rainfall, right hand plot uses the UKV forecast, solid circles are raingauge locations.

Scotland

Over Scotland, Figure 5.25 compares the performance of G2G flow forecasts using UKV rainfall as input (red lines) with those using raingauge-only rainfall (black lines). Assessments are grouped by catchment size as before. In comparison to England & Wales (Figure 5.16), an even stronger deterioration in median R^2 Efficiency performance is evident beyond lead-times of 12 hrs for all catchment sizes (solid lines). For the small catchments (< 50 km^2), the 6 hour lead-time performance is still good with the median R^2 Efficiency above 0.8 and the spread (dashed lines) much less compared to England & Wales. This is a reflection of the good G2G simulation-mode performance for these sites. Conversely, the spread of performance for the larger catchments (>100 km^2) is greater for Scotland than for England & Wales and is probably due to Scotland having fewer nested catchments. One interesting trait is that the G2G forecast performance using UKV rainfall for Scottish catchments less than 50 km^2 is better than for the 50 to 100 km^2 catchments. The opposite is true in England & Wales and is partly due to the fact there is only one site in Scotland that has an upstream gauge to help with forecasts for these smaller catchments. An additional factor could be where the catchments are located and the performance of the UKV rainfall forecasts over these areas.

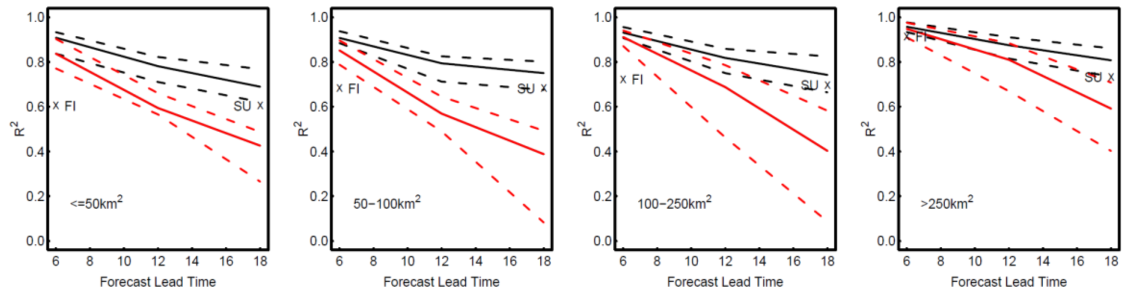


Figure 5.25 Scotland. Median (solid lines) and inter-quartile range (dashed lines) of R^2 Efficiency for forecast-mode G2G flows for different lead-times, using raingauge-only observed rainfall (black lines) and UKV forecast rainfall (red lines), grouped over different catchment sizes.

More detail on how the spread of model performance evolves with lead-time is presented in Figure 5.26. These scatter plots show the G2G flow forecast performance obtained using raingauge-only rainfall as input relative to using UKV NWP rainfall forecasts. These show two clear groups of model performance in terms of raingauge-based G2G forecast performance with a majority of sites having R^2 Efficiency values of 0.6 or more over the full range of forecast lead-times (red dots). There is then a clear break to a group of sites which have low or negative R^2 Efficiency values and generally represent sites that are not well modelled for a genuine reason, e.g. reservoir influences or an under-representative raingauge network over the catchment. The G2G forecast performance for the first 6 hours (green dots) is generally very good using the UKV rainfall and drops off only slightly in relation to use of raingauge rainfall. Beyond the first 6 hours, G2G flow forecast performance using UKV rainfall drops off more rapidly, with some sites having R^2 Efficiency values below zero. The scatter plots also show that for some locations G2G forecasts are still performing very well, even at long lead-times, with a reasonable number having R^2 Efficiency values above 0.4. In comparison, the performance for England & Wales (Figure 5.17) is much more varied.

In particular using NWP rainfall for the first 6 hours, performance drops off more quickly and is much more varied than for Scotland.

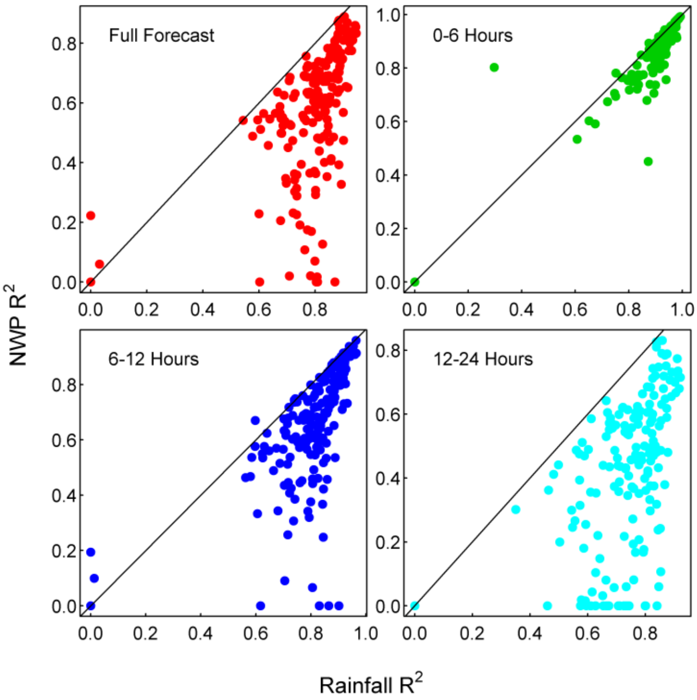


Figure 5.26 Scotland. Scatter plots of R^2 Efficiency for forecast-mode G2G flows using raingauge-only observed rainfall and UKV forecast rainfall grouped by forecast lead-time.

Assessing the spatial variability of the G2G forecast performance using UKV rainfall is explored in Figure 5.27. This maps the differences in G2G performance between the two types of forecasts: the larger circles show where the G2G flow forecasts using UKV rainfall as input are performing poorly relative to using raingauge-only rainfall. This reveals a band of gauging stations from Edinburgh northwards where the G2G forecasts using UKV rainfall appear to be performing less well, particularly towards the coast at Edinburgh and Elgin in the north.

As with the England & Wales analysis, NWP UKV forecast monthly rainfall accumulations have been calculated (using the T+3 to T+27 part of each forecast and dividing by 4 to account for forecast overlaps) and then compared to equivalent raingauge-only ones. Maps of these monthly rainfall accumulations for the period August 2010 to June 2011 are shown in Figures 5.28 and 5.29 and 12 month totals are presented in Figure 5.30. The elevation and Standard Average Annual Rainfall maps in Figure 5.23 should also be referred to. Together, these provide interesting insight into the possible reasons for the localised deterioration in the G2G forecasts using UKV NWP rainfall as input. It should be noted that the gauged catchment areas used in the assessment of G2G are predominantly in the southern and eastern areas of Scotland (see Figure 5.15). Therefore along the west coast the difference between the G2G forecasts using raingauge-only and UKV rainfall are of less relevance to this part of the analysis. But the difference does indicate that the sparseness of the raingauge network in this area can have a significant detrimental impact on rainfall estimation. It would also suggest that either including orography or radar rainfall within the rainfall estimation scheme may provide some benefits and should be considered in the future.

The relative sparseness of the raingauge network across parts of the highlands and northern Scotland makes the analysis more difficult. However, there are still some strong regional signatures coming through. In general, agreement between the UKV and raingauge accumulations is good on the west coast near the raingauge locations, except for the very northern part where raingauges appear to give lower values (possibly due to the representativity of the raingauge network). Towards the east coast, UKV appears to overestimate rainfall, particularly along the north coastline near Edinburgh and Elgin. The raingauge-only rainfall mapped in Figures 5.28 to 5.30 shows a similar trait for these coastal locations to the SAAR mapped in Figure 5.23 and gives some confidence that the raingauge locations are reasonably representative of the surrounding area in terms of rainfall experienced. These locations correspond to the catchments with largest deterioration (biggest circles in Figure 5.27) in G2G forecast performance when using UKV rainfall rather than raingauge-only rainfall.

An area of interest is the Cairngorms where the UKV NWP rainfall shows an increase, produced through the orographic enhancement scheme, in the annual UKV totals relative to the raingauge totals (Figure 5.30). Whilst some enhancement is expected in these areas (see SAAR map in Figure 5.23) the difference compared to the local raingauge network can be quite significant at times (e.g. November 2010 in Figure 5.28). The deterioration in G2G forecast performance in Figure 5.27 suggests the UKV NWP is over-predicting the rainfall here.

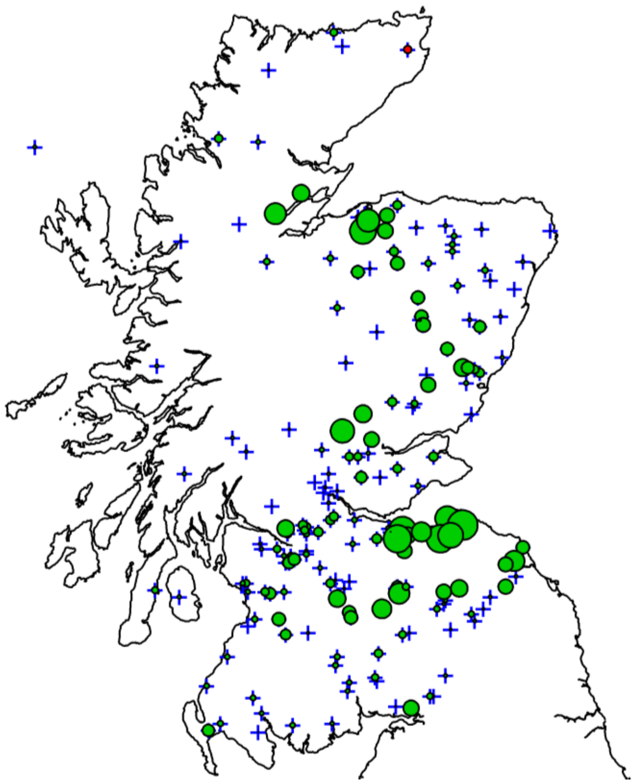


Figure 5.27 Scotland. Difference in R^2 Efficiency between forecast-mode G2G flows using raingauge-only observed rainfall and using UKV NWP forecast rainfall. Large circles show where G2G forecasts using UKV perform worse.

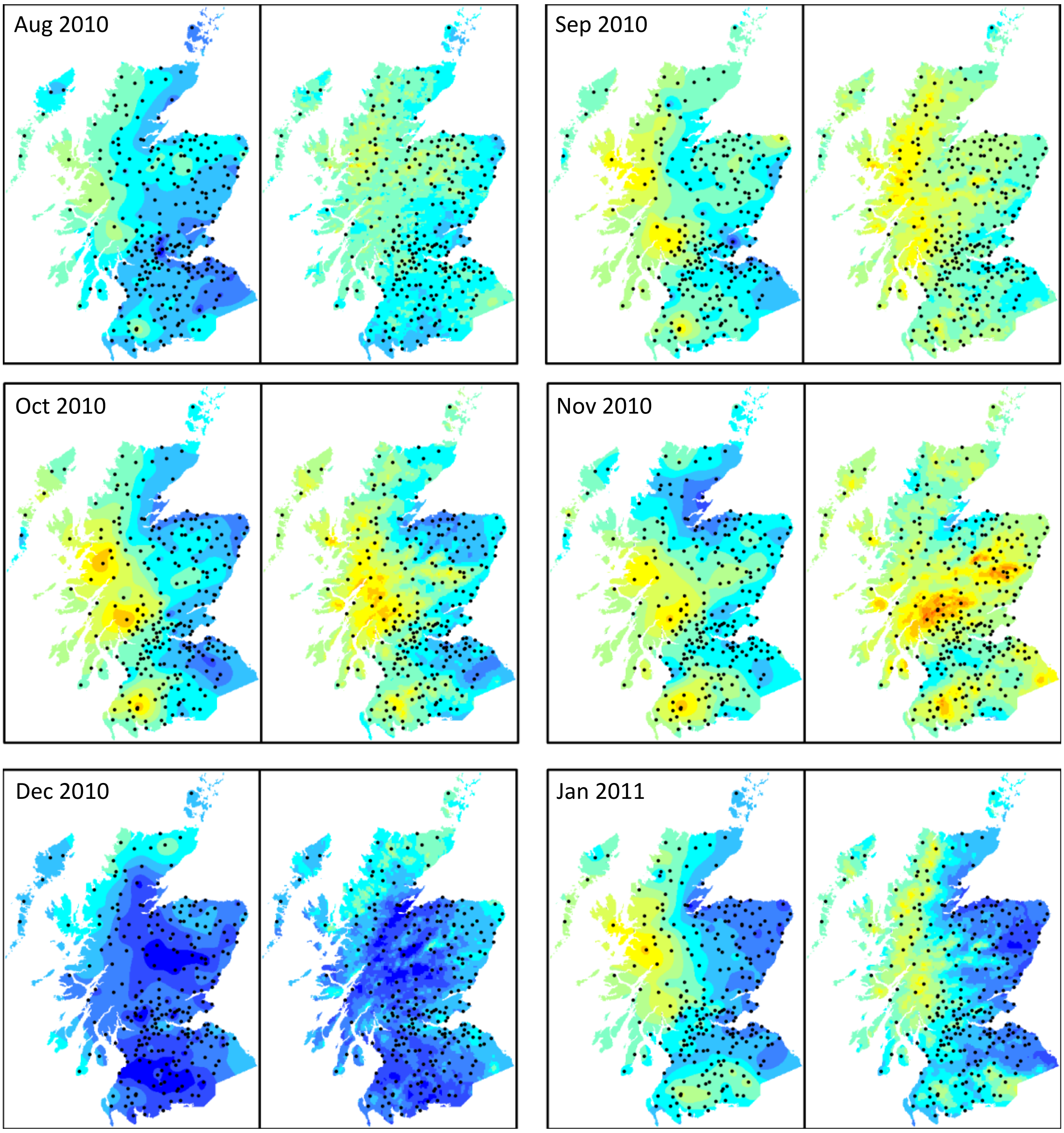


Figure 5.28 Scotland. Accumulated rainfalls for 30 day periods covering each month from August 2010 until January 2011. Left hand plot of each pair use HyradK raingauge-only observed rainfall, right hand plot uses the UKV rainfall forecast, solid circles are raingauge locations. The legend scale is given in Figure 5.29.

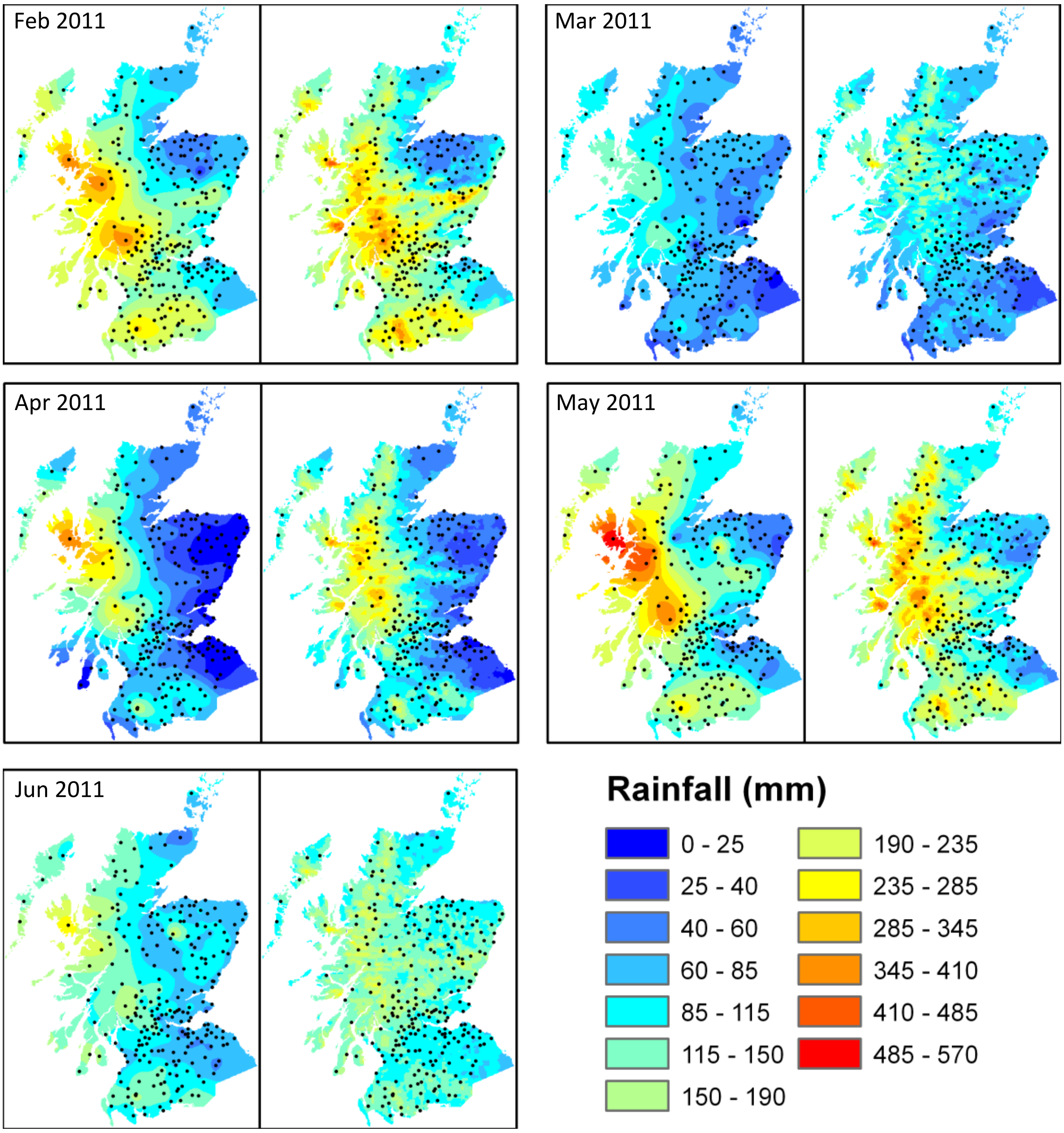


Figure 5.29 Scotland. Accumulated rainfalls for 30 day periods covering each month from February to June 2011. Left hand map uses HyradK raingauge-only observed rainfall, right hand map uses UKV forecast rainfall, solid circles are raingauge locations.

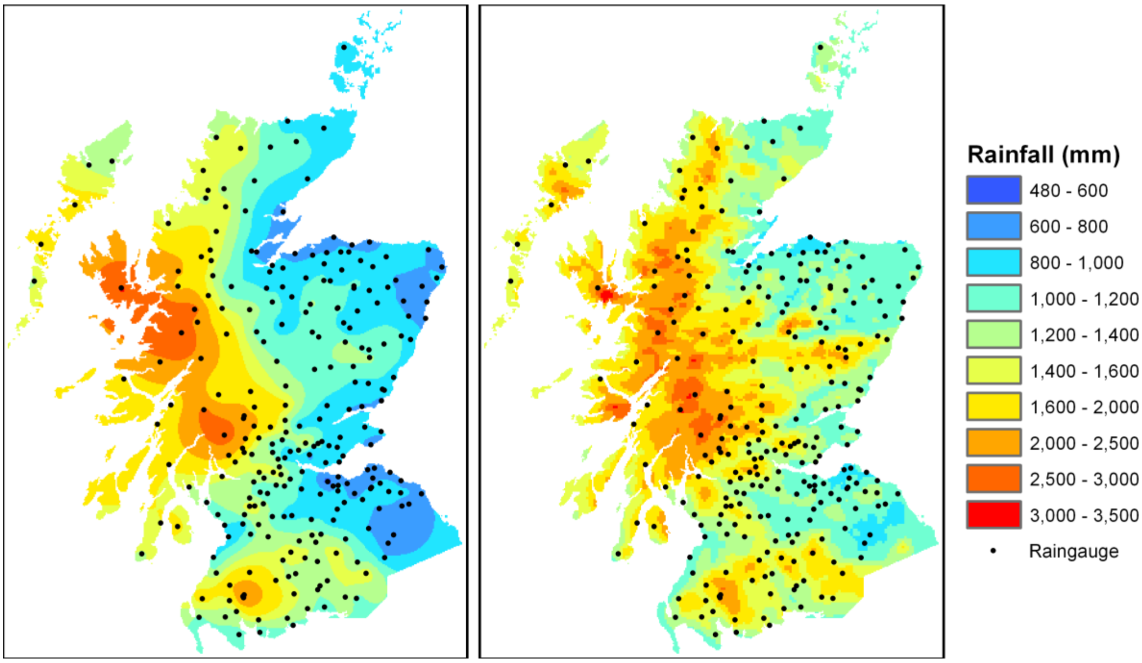


Figure 5.30 Scotland. Accumulated rainfalls for August 2010 to July 2011. Left hand map uses HyradK raingauge-only observed rainfall, right hand map uses the UKV forecast rainfall, solid circles are raingauge locations.

5.5 Forecast-mode assessment: use of Blended Ensemble rainfall forecasts

This section assesses the forecast-mode performance of G2G using Blended Ensemble rainfall forecasts as input. Firstly, case-study datasets are used and will serve as indicative demonstrators whilst using the rainfall forecast datasets over longer periods (months) will be used to perform more meaningful assessments

5.5.1 Computing infrastructure to run G2G ensembles

Running extended periods of G2G ensemble forecasts offline within an acceptable time requires significant computer processing power and storage. CEH have a new Linux Cluster consisting of 12 nodes. Each node has 2 Intel Xeon X5650 6-core (2.66GHz speed) processors so the cluster has 144 cores available. G2G has been ported from Windows code to run on Linux and recent enhancements now allow model states to be stored every day during a historical run of the model. This allows G2G to be easily restarted at any day so running case studies and splitting longer ensemble runs is much easier. Also significant developments to the environment used to run G2G models has been undertaken to maximize use of the Linux Cluster, manage running both case-study and long-term ensemble runs, and help collate and analyse the G2G outputs.

5.5.2 Case-study analyses

At the outset of the G2G for RRCs project, the Morpeth flood on 5/6 September 2008 was identified and had a 24 member UKV ensemble dataset available: this is a forerunner of the soon to be operational MOGREPS-UK ensemble. During the lifetime of the project the UK suffered a series of flood events from April through to December 2012 following an extended drought period from 2010-12 (Parry et al., 2013). An Environment Agency internal report documents 11 distinct flood periods during 2012 (Environment Agency, 2013). A full list of known notable flood events during this time are listed in Table 5.2 and a sub-set have been identified (grey shading) for detailed case-study analysis. The full list is useful background as they occur during the longer-term assessment period (Section 5.5.3). The case-study events are considered in more detail in the following sections. SEPA also provided information on known flood events over Scotland during the summer 2012 period.

Additional sources used for background information on the rainfall and flood events are: (i) the National River Flow Archive Monthly Hydrological Summary (http://www.ceh.ac.uk/data/nrfa/nhmp/monthly_hs.html), (ii) the Met Office monthly weather summary (http://www.metoffice.gov.uk/climate/uk/summaries), and (iii) the British Isles Daily Weather Diary maintained by Roger Brugge from the National Centre for Earth Observation (http://www.met.reading.ac.uk/~brugge/diary.html).

The analysis of the case study events is not in chronological sequence but is aligned to the order in which they have been assessed.

Table 5.2 Flooding events from 10 May to 31 December 2012 plus the Morpeth 2008 flood. Case-study events are shaded in grey.

Date	Summary	Locations
5/6 September 2008	Deep low pressure bringing a band of heavy rain. 150mm estimated over Wansbeck catchment.	Morpeth, north-east England plus Wales, south-west England and Midlands.
8-9 June 2012	A slow-moving area of low pressure maintained steady heavy rainfall across parts of Midlands and West Wales. 183mm in 36h at Dinas.	Wales
10-11 June 2012	Low pressure system spread up from France and became near-stationary across the south east of England. 30-50mm widespread, 70mm in West Sussex, 100mm Bognor Regis.	West Sussex
22-24 June 2012	Slow-moving low pressure system brought a prolonged period of heavy rain to northern England. 50-80mm in 24h across Lancashire, Cumbria and Pennines. 48hr totals exceeded 100mm in some places, >200mm at Honister in Cumbria. RRCs significant concern.	Northern England
27-28 June 2012	Central, eastern and northern England where severe supercell thunderstorms brought locally torrential rain, large hail and further flooding from surface water and small rivers	Midlands & Northern England, Southern Scotland
5 July 2012	Thundery down pours. 90 mm in ~ 3 hours, 20.6 in 1h.	Ayrshire, Cessnock Water (River Irvine)
6-7 July 2012	Slow moving low pressure over south-west England. >100mm recorded in 18hr in south-west. Another band of heavy rain moved northwards with 40-50mm widespread.	South-west and northern England. Edinburgh and borders.
8-9 July 2012	Heavy rain.	Calder valley, Yorkshire
14 July 2012	Intense thunder storm development initially across Wales and Shropshire (41.6mm at Pennerley) before spreading eastwards across the Midlands. Thunderstorms and flooding also over Southern and central England and East Anglia.	Wales, central, south-west and eastern England
18 July 2012	Prolonged rain over eastern and central Scotland spreading from the west. 55mm at Edinburgh.	Fife and Perthshire
5-6 August 2012	Localised heavy thundery showers causing some localised flooding. 30mm common.	Pembrokeshire, Cheshire, Devon, Tyneside and the Scottish Borders (Jedburgh)
15-16 August 2012	Heavy downpours caused localised surface flooding.	Wales, south-west England, Northampton
22-23 August 2012	Heavy showers caused localised surface flooding.	North-west Scotland, Lancashire

25 August 2012	Depression move eastwards through Wales and Midlands. Showers, some heavy and thundery. Festival at Daresbury, Cheshire closed due to flooding.	Flash floods in Debyshire (Dronfield, Barlow). Daresbury, Cheshire
28 August 2012	A frontal system from the south-west brought rainfall over most of the UK followed by some localised showers. Around 40mm in 6 hours caused flooding on the River Ruchill, Comrie.	River Ruchill, Comrie.
29 August 2012	Low pressure bringing stormy conditions with some thunder. 23.2 mm in Edinburgh, 42.6 mm St Bees Head, Cumbria.	Cumbria (River Ehrn), southern Scotland.
23-27 September 2012	A low pressure system to the west of Portugal moved north-eastwards towards the UK. The storm rapidly intensified and became slow-moving across the UK on the 24 to 26 September. Southern England and central Wales seeing 40-60mm and 70-90 mm in some places on 23rd. Northern England hit on 25th and 26th with 144.2mm at Richmond.	Flooding in Midlands, south-west (Che Magna) and northern England (Morpeth, York, Selby, River Ouse)
18-19 November 2012	Passage of a very vigorous frontal system. Heavy rain across south west and western Scotland.	Serious flooding at Comrie.
21-27 November 2012	7-day period from 20 to 26th November was one of the wettest weeks in the last 50 years for England and Wales. A succession of heavy rain events affected a swathe from Devon to north-east England.	Widespread, West Scotland, Northallerton (North Yorkshire), Malmesbury (Wilts.), Kennford (Devon)
19 -31 December 2012	A series of Atlantic depressions spread across the UK from the south west. Widespread flooding on 20th then small catchments on 22nd/23rd (Devon, Scotland) followed by larger slowly responding rivers.	Widespread flooding. Rapid response at Braunton (river Caen), Devon and Stonehaven (river Carron), Aberdeenshire

19-20 November 2012

During the evening of 18 November and through most of the following day, a very vigorous frontal system moved across Scotland from the west/south-west bringing heavy rain across south west and western Scotland. The major flooding impacts were at Comrie, situated near where Ruchill Water joins the River Earn.

This event was the first case-study undertaken for the G2G for RRCs project. It is slightly different in that data were only provided by SEPA for the 8 nearest raingauges and two river gauging stations so is very catchment specific. A joint SEPA/CEH paper was presented at the European Geosciences Union General Assembly (Cranston et al., 2013) covering this case-study and is the primary source for the results presented here.

Providing flood warnings to Comrie is difficult because of the very quick response time of the Ruchill Water catchment. It has also been flooded several times in recent years and is why SEPA have been interested in using it as a case study for the G2G for

RRCs project. The case study area is mapped in Figure 5.31 along with the hydrometric network (top figures). The most relevant river gauging station is the Ruchill at Cultybraggan (99km^2) which is situated just upstream of the confluence at Comrie. A second gauging station is the River Earn at Kinkell Bridge (591km^2) which is someway downstream of Comrie.

The simulation-mode performance of G2G using raingauge-only HyradK gridded rainfall estimates is shown in the bottom row of Figure 5.31. For the Ruchill catchment there is good confidence in the rating curve so the underestimation of the river flow appears to suggest that there is an underestimate of the rainfall. Cranston (2006) deployed additional temporary raingauges in the vicinity of the Ruchill catchment and found that there can be strong orographic enhancement for south westerly storms such as this. This means that the current configuration of operational raingauges is unlikely to capture this enhancement and is consistent with the G2G simulation results.

At Kinkell Bridge the observed flow response is more damped and lower than the G2G flow simulation. It is thought that this is due to the operation of Loch Earn upstream which is not currently included in the G2G model configuration for Scotland.

The first trial was to use the T+24 Blended Ensembles to assess if these gave any useful advanced warning. A sequence of blended ensemble G2G forecasts is given in Figure 5.32 and show a strong signal, particularly for the 12:15 and 18:15 forecasts on 18 November 2012. These forecasts would have been available around 20 and 14 hours before the peak of the flood and could have potentially triggered an early operational response. Surprisingly, the 00:15 19 November 2012 forecast just before the event appears to have some timing issues with the peak rainfall and flow occurring too early.

Secondly, the case study looked at using the T+7 Blended Ensemble nowcast of rainfall as input to G2G over the flow forecast horizon. An example snapshot of the nowcast ensemble rainfall is provided in Figure 5.33 to give an appreciation of what the blended rainfall forecasts look like. Following guidance from Clive Pierce (Met Office),

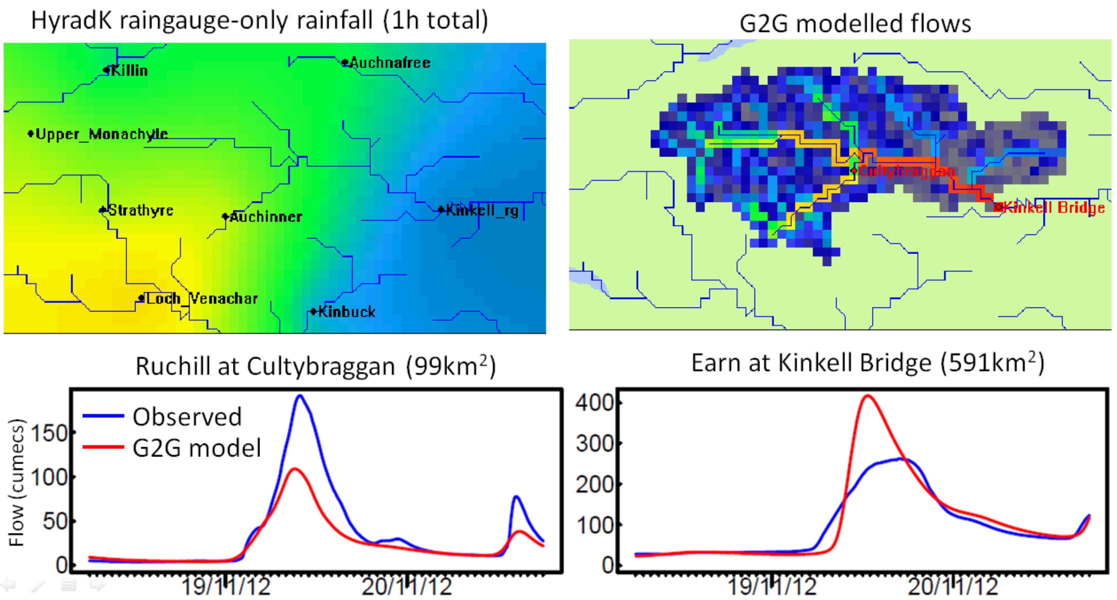


Figure 5.31 Comrie case study, 19-20 November 2012. Top left: snapshot of 1h total from HyradK raingauge-only rainfall, raingauge locations shown. Top right: corresponding 1km grid of G2G river flows, the two gauging locations are shown. Bottom row: G2G model performance in simulation-mode using HyradK raingauge-only rainfall as input

four consecutive G2G flow ensemble nowcasts were time-lagged to produce the final G2G flow ensemble nowcast for each forecast time-origin. Figure 5.34 presents a sequence of the time-lagged ensembles. The first six of these forecasts show a strong tendency for the ensemble members to be over-predicting the actual rainfall. Further discussion with Clive Pierce has conjectured that this is probably related to the orographic enhancement schemes used by the STEPS-2 algorithm.

Essentially the algorithm identifies orographic enhancement in the initial radar rainfall observations through use of Numerical Weather Prediction model information. This is then subtracted from the radar rainfall field before the advection scheme is applied and then the enhancement is added back in appropriately. The aim is to stop orographic enhancement, which is a consequence of topography and concentrated at high elevations, being advected to lowland areas where enhancement would not occur. It appears that in this case the enhancement was not correctly identified (if at all) resulting in artificially high rain-rates being advected downwind. This is also the likely reason for the early peaks in the T+24 ensemble forecast at 00:15 19 November 2012 in Figure 5.32 since advection of the radar rainfall is dominant in the early parts of the

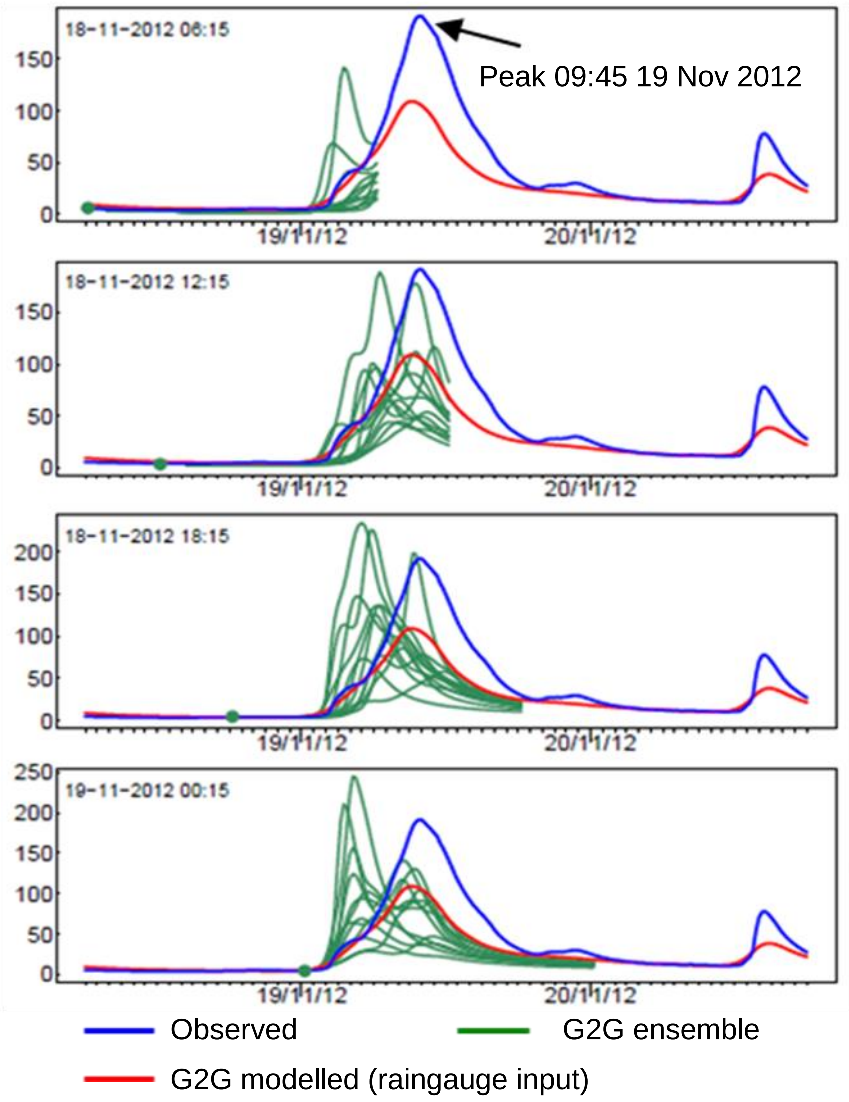


Figure 5.32 Comrie case study for 19-20 November 2012. Sequence of Blended Ensemble T+24 G2G flow forecasts for the Ruchill at Cultybraggan. The time and date shown in each plot is the forecast origin.

forecasts. It should be noted that the longer-range blended ensemble (beyond T+7) should not be affected in the same way. This is because the high-resolution UKV aims to explicitly model the orographic enhancement process and so this is partly why it may have proved a useful early indicator for this event.

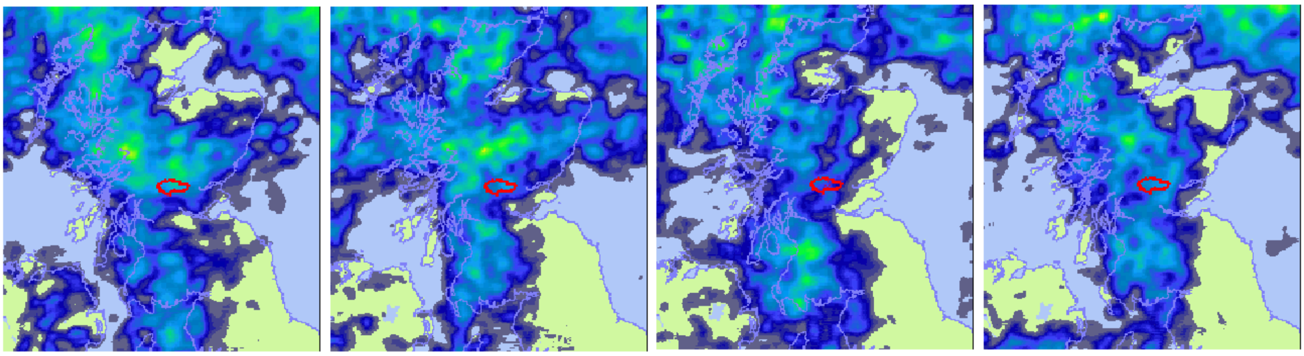


Figure 5.33 Comrie case-study for 19-20 November 2012. Example snapshot of 15 minute rainfall accumulations during a nowcast for 4 ensemble members. The River Earn at Kinkell Bridge catchment is shown in red.

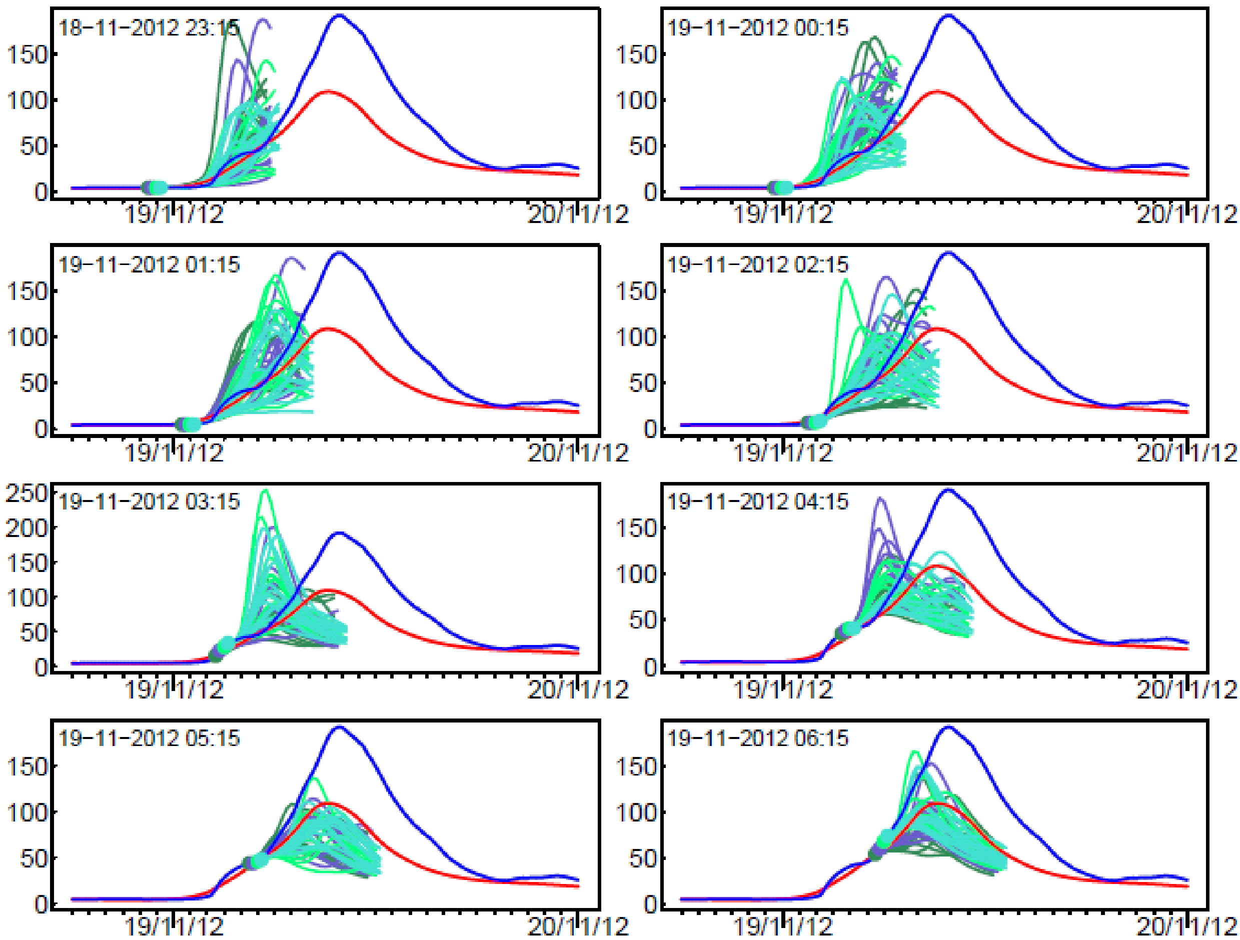


Figure 5.34 Comrie case study for 19-20 November 2012. Sequence of Blended Ensemble T+7 G2G flow forecasts for the Ruchill at Cultybraggan. The time and date shown in each plot is the forecast origin. Note four consecutive ensembles are time-lagged.

27-28 June 2012

Warm, humid air moving northwards brought rather cloudy, muggy conditions. Hill and coastal fog became a feature over western areas as sporadic rain spread to all parts with some heavy showery bursts on 27th. During the 28th hot, humid air tracked up on southerly winds from the Azores far to the south in the Atlantic. The heat and moisture in the air were enough to cause thunderstorms, but the really intense storms were formed as an Atlantic weather front moved in from the west causing air to rapidly rise and create towering cumulonimbus storm clouds laden with water. Several distinct lines of thunderstorms developed along the boundary where the air masses met.

Two particularly vigorous lines of thunderstorms formed. The first line originated in the Cardiff area of south Wales in the early morning and moved in an east-north-east direction across Worcestershire, Shropshire, the West Midlands and Leicestershire to clear Lincolnshire by late afternoon. The second line of thunderstorms reached the Lancashire coast around late morning and moved in a north-east direction to reach the Newcastle area later in the day, clearing the north east coast by late evening. Frequent lightning and large hail accompanied the storms. Rainfall totals of 20 to 30 mm in an hour were reported widely (e.g. Scampton, Lincolnshire recorded 28.4 mm in an hour) whilst 40-50 mm was reported at isolated locations. Elsewhere there were also torrential downpours across parts of Northern Ireland and western Scotland.

Flooding from surface water and small rivers was reported in central, eastern and northern England. Particularly badly affected was north-east England and Tyneside with major disruption to infrastructure including closure of the A1 and many minor roads. Flooding of Newcastle Rail Station led to train services being suspended and landslides caused the metro line to close. Houses, schools and businesses were flooded in Barrow-in-Furness, Kendal and the Penrith area. West Midlands Fire Service dealt with 282 incidents in 90 minutes whilst. Lincolnshire Fire and Rescue Service attended more than 200 incidents of flooding during the afternoon.

This case study has focussed on the North East region of England and two small catchments in the Newcastle/Tyneside region: the Ouse Burn at Woolsington (WOOLSN1, 9.0 km^2) and the Team at Team Valley (TEAMVL, 61.9 km^2). A third gauging station at Shillmoor on the Usway Burn (SHILMR1, 21.4 km^2) in the Cheviots is also considered. The study area is presented in Figure 5.35

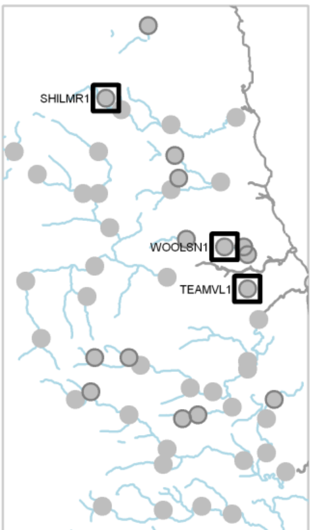


Figure 5.35 Case study area within North East Region for flood event of 27-28 June 2012. River gauging station locations are indicated by grey circles. The three highlighted by square boxes are used in subsequent point analysis whilst circles with a solid outline indicate a catchment area less than 50 km^2.

and a sequence of regional summary maps using the T+24h blended ensembles are presented in Figure 5.36 for forecast origins during the 27 and 28 of June. The observed response (black dots) show a cluster of stations in the Tyneside (Newcastle) area affected at the Q(5) threshold along with a couple of individual sites further south.

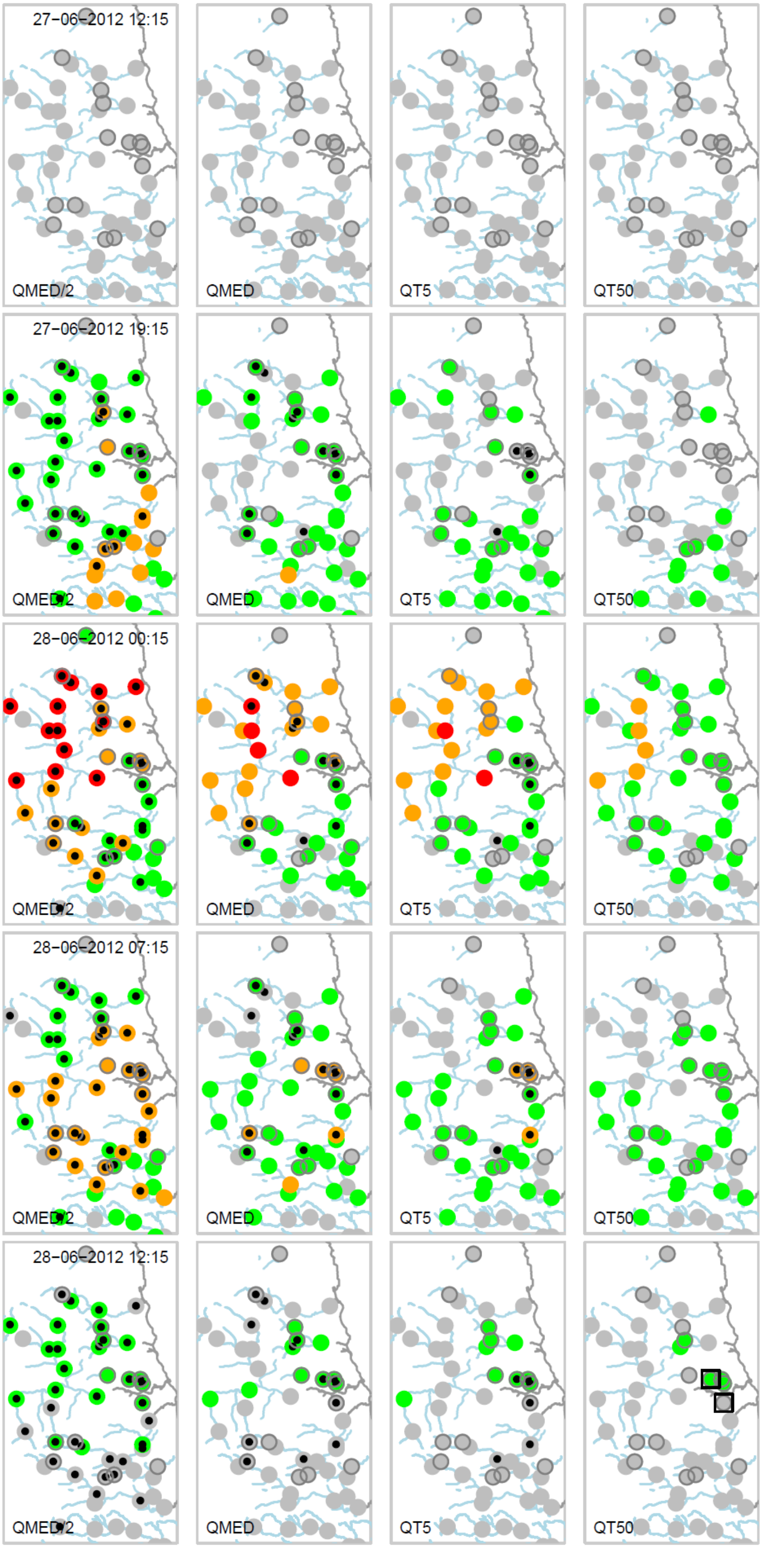


Figure 5.36 Regional summary maps of ensemble G2G flow forecasts for the North East using a sequence of T+24 blended ensemble rainfall forecasts during 27-28 June 2012. See Figure 4.5 for a full definition of the maps.

At the QMED threshold, another clustering of sites further north is apparent for headwater catchments near the Cheviots and tributaries of the River Wansbeck catchment which peaked earlier than the Tyneside catchments.

The sequence of regional summary maps show some interesting characteristics. Forecasting the formation and location of the thunderstorms that caused the floods is a significant challenge, especially when making forecasts out to 24 hours. It should be noted that as the blended ensemble scheme essentially uses one deterministic NWP forecast, the ensemble spread in location is likely to be less than would be possible from an NWP ensemble (e.g. MOGREPS-UK).

Analysing the forecasts in sequence reveals the location of highest flow response varies in location and is in partial agreement with the observed flow repsonse which peaked slightly earlier in northern areas. The first forecast to show a significant response is 19:15 on 27 June. The low Q(T) thresholds identify wide areas at low risk with the south of the region being picked out as having potentially more extreme responses. The next forecast at 00:15 28 June shows an increase in the likelihood of a flood location (more orange and reds) and moves the focus of the most likely and highest thresholds to be the north/north west of the region which has reasonable correspondence with the observed flooding in the Cheviots. The 07:15 forecast is less certain in terms of number of thresholds crossed (less red) but does well at picking out the Newcastle area as potentially most at risk (oranges in QMED and Q(5) maps). Interestingly the forecast made at 12:15, which is closest to the flood event, is also the least certain with only a low number of threshold crossings. However, Newcastle and the area north west of it are identified as having a low likelihood of exceeding a high threshold (Q(50)).

Figure 5.37 focuses on the two small catchments near Newcastle with Woolington having a catchment area of only 10 km^2. The G2G simulated flow using raingauge-only rainfall (red lines) shows flows for the Team Valley (82 km^2) location are reasonably well modelled for this event, particularly in terms of peak magnitude and timing although the fine temporal fluctuations are not captured. The Ouse Burn catchment is not so well modelled with the G2G response being too early and too sharp. The first five rows correspond to the regional summary maps in Figure 5.35 and provide additional site specific insight into the high level summary. The two forecasts (00:15 and 07:15 28 June) had a reasonable number of threshold crossings and were flagged in the regional maps. Looking at the spread in peak flow for these ensembles shows there are generally two distinct classes: (i) very high peaks for heavy storms that “hit” the catchments, and (ii) low peaks for lower intensity storms or heavy storms that only brushed the catchment. This highlights how, for these small catchments, the spatial uncertainty in the deterministic NWP forecast is captured by the blended ensemble algorithm. The 12:15 plot reinforces the conclusion of the regional summary maps that this particular forecast didn’t capture the rainfall of the event very well. The main reason is that the line of storms responsible for the flood response formed and passed over the region within the first six hours so is a challenge for the nowcasting element of the blended ensemble algorithm to capture.

This is the first case study to analyse the T+7 nowcast ensembles for a large region. A sequence of regional summary maps using the T+7h blended ensembles are presented in Figure 5.38 for forecast origins during 28 June. The regional summaries are formed every hour and use a time-lagged set of 4 ensemble nowcasts. The forecast origin of the first ensemble used is indicated, so the first row of the left hand column has a date of 06:15 28 June 2012 meaning ensembles starting 06:15, 06:30, 06:45 and 07:00 were used to form the time-lagging. In addition, point time series are given for two gauged locations. The forecasts at Shillmoor (21.4 km^2) in the Cheviots are presented in Figure 5.39 whilst the forecasts at Team Valley near Newcastle are presented in Figure 5.40. Each row of the ensemble forecast plots corresponds to the four time-lagged ensembles that are used for the regional summaries in Figure 5.38.

For the Cheviots region, the main rainfall fell between 11:30 and 15:30 with the response at Shillmoor peaking around 16:00. This part of the storm was reasonably well

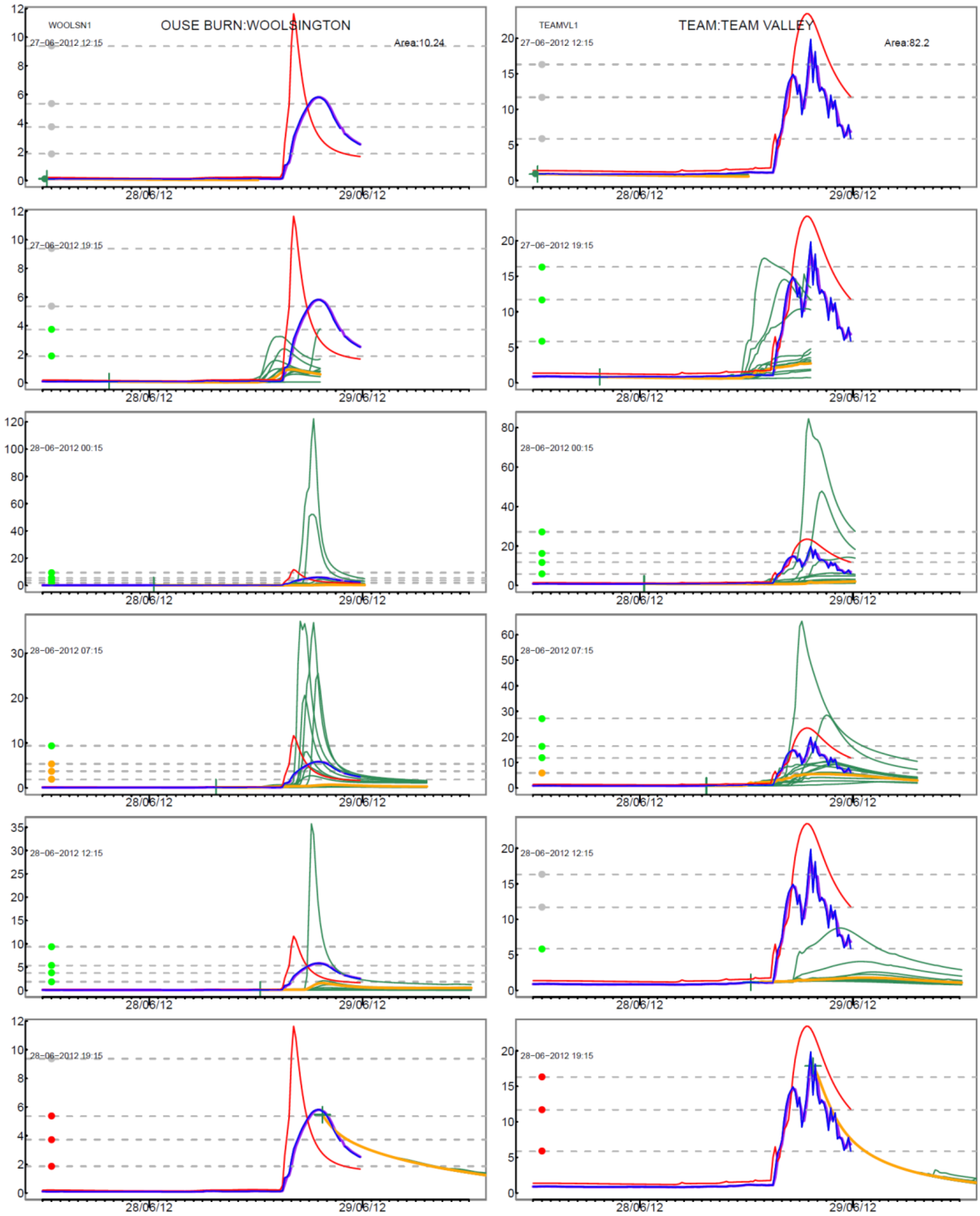


Figure 5.37 Ensemble G2G forecasts for the Ouse Burn at Woolington (WOOLSN1) and the Team at Team Valley (TEAMVL1) using a sequence of T+24 blended ensemble forecasts during 27-28 June 2012. See Figure 4.6 for a full definition of the time-series display.

organised and probably easier to forecast. The regional maps first pick up a weak signal in the Cheviot area at 11:15 which then becomes stronger in the 12:15 and 13:15 forecasts which are both before the observed river level starts to rise (see Figure 5.39). The 13:15 map in particular identifies the Cheviots and surrounding areas as the main areas of concern.

For the Newcastle area the most intense storms passed over later on between 15:00 and 18:00 with the peak gauge response at around 19:00. These thunderstorms were short-lived and very localised so more difficult to forecast. The sequence of regional maps show the first weak signal of an event for the 11:15 forecast with the Newcastle area identified as having very low likelihood of crossing the Q(50) flow threshold. For 12:15 to 13:15 the Newcastle/Tyneside area isn't identified at all, but then from 14:15 to 16:15 the signal becomes increasingly strong and localised in part due to the observed river levels at Team Valley beginning to rise by 15:00 (see Figure 5.40).

These case study results using the blended nowcast T+7hr ensembles has shown that using a range of Q(T) thresholds can help narrow down the areas of main concern. Reasonably focussed regions – flood risk "hotspots" - can be identified an hour or two before any response is seen in the gauged river flows (even for small catchments). The sequence of maps also correctly identifies the northern areas being at risk earlier than the Newcastle/Tyneside areas. An interesting point to note is that any given set of maps can include (i) gauges that have peaked but are still flagged red due to the recession forecasts still exceeding thresholds, and (ii) gauges that are yet to peak or even started to register an observed response. Although the black dots on the regional summary maps indicate whether the observed flow has crossed the threshold during the forecast horizon it could be useful to show a similar dot if the observed flow has crossed this threshold in the previous X hours. This would then help to identify if a catchment has already crossed the threshold level.

Comparison with the T+24hr ensemble forecasts is very interesting. The 00:15 T+24hr ensemble forecast (Figure 5.36) gives a very similar picture to the 13:15 T+7hr forecast (Figure 5.38) with the Cheviot area being identified as a potential flood risk hotspot with very similar likelihoods. This is in part due to the flood-producing rainfall here being more well organised and easier to predict by NWP. For the Newcastle area the T+24hr gives a medium signal at 07:15 which is similar to the 14:15 T+7hr forecast. However the T+24hr forecast has a weaker signal at 12:15 whilst the T+7hr forecasts intensifies from 14:15 to give a good indication that the Tyneside area is more at risk, albeit with a very short lead-time. It should be noted that many of the G2G ensembles for the T+7hr nowcasts are still rising at the end of the forecast and there may be some benefit to extend the forecast lead-time to allow the flows to peak at the catchment outlet.

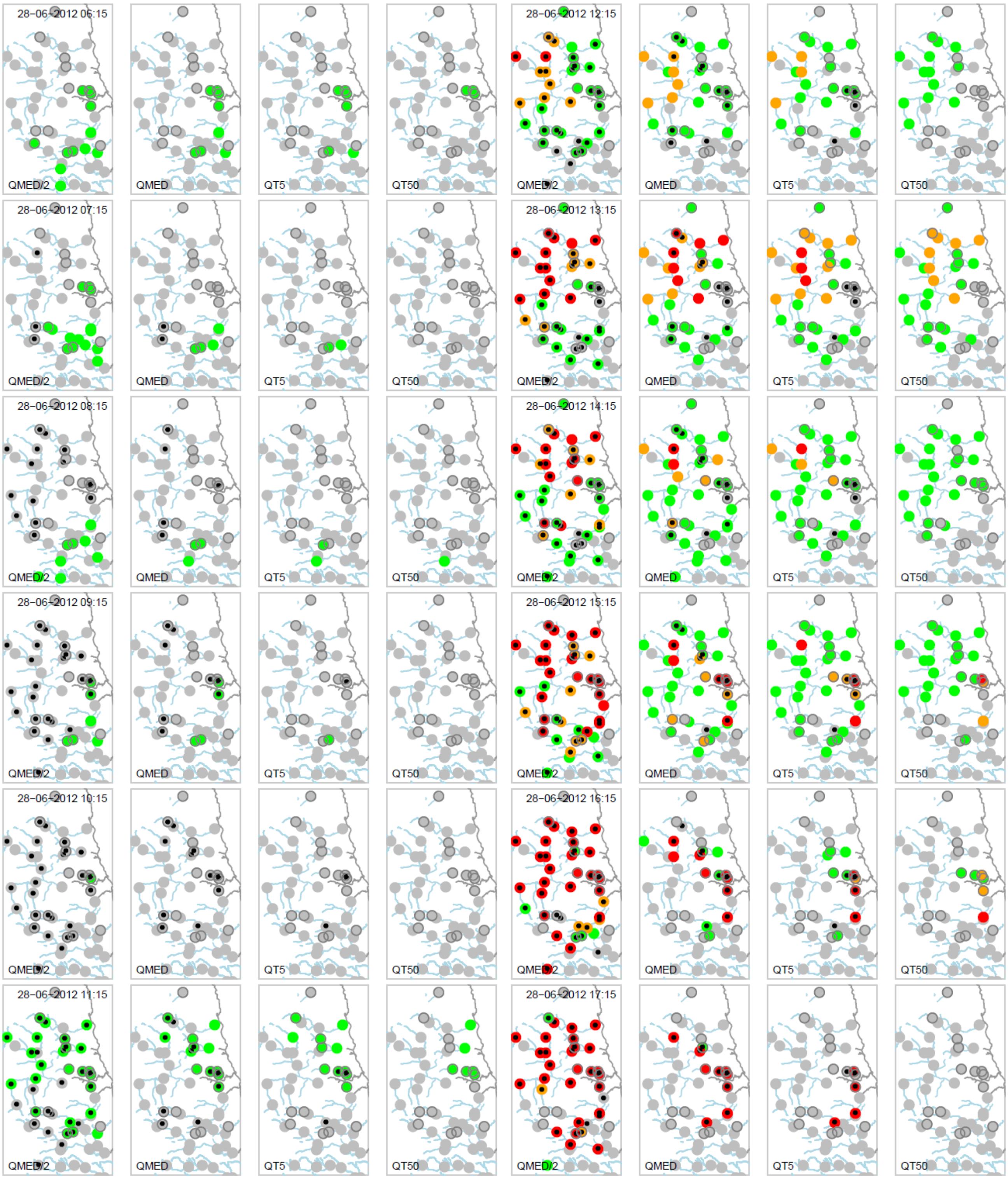


Figure 5.38 Regional summary maps of ensemble G2G flow forecasts for the North East using a sequence of T+7 blended ensemble nowcasts of rainfall during 27-28 June 2012. Note four time-lagged sets of ensembles are used for each map. See Figure 4.5 for a full definition of the maps.

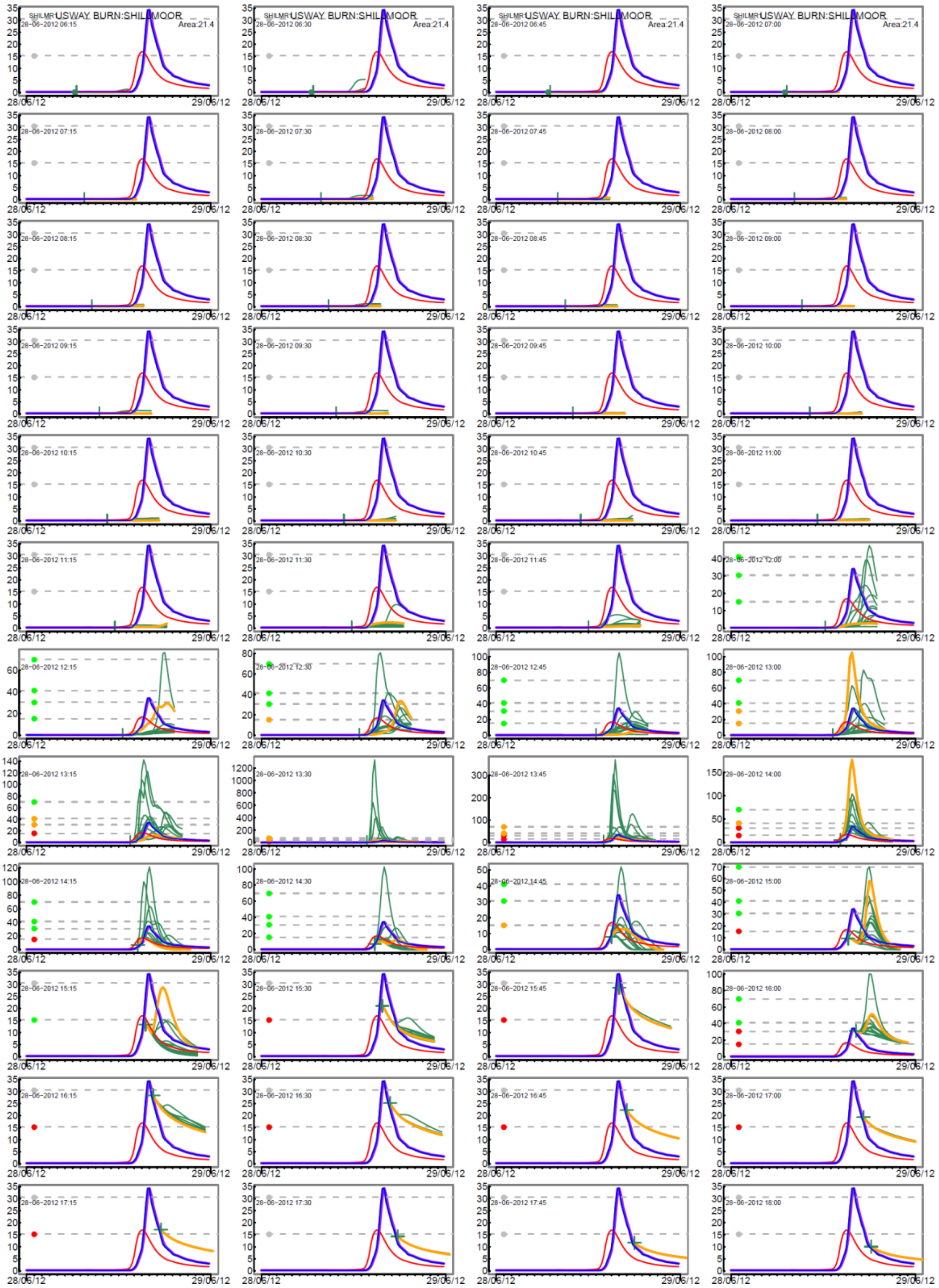


Figure 5.39 Ensemble G2G flow forecasts for the Usway Burn at Shillmoor (SHILMR1) using a sequence of T+7 blended ensemble nowcasts of rainfall during 28 June 2012. See Figure 4.6 for a full definition of the time-series display.

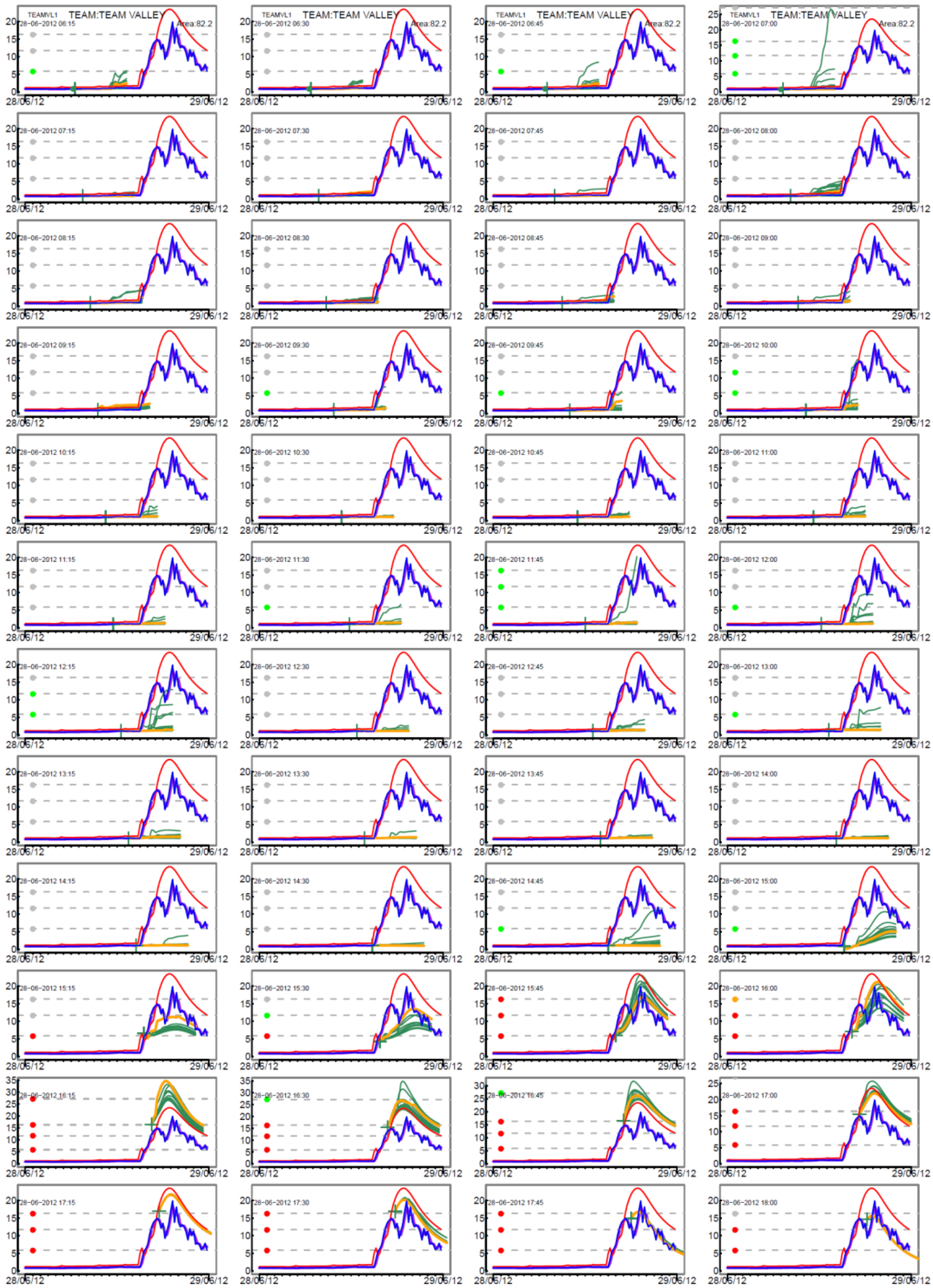


Figure 5.40 Ensemble G2G flow forecasts for the Team at Team Valley (TEAMVL1) using a sequence of T+7 blended ensemble nowcasts of rainfall during 28 June 2012. See Figure 4.6 for a full definition of the time-series display.

22-24 June 2012

On 22 June a slow-moving low pressure system brought a prolonged period of heavy rain with the heaviest accumulations in a band from Northern Ireland across the Isle of Man to Lancashire and the Pennines. Rainfall totals of 50-80mm in 24 hours were reported widely across Lancashire, Cumbria and Pennines area of Yorkshire with Honister Pass in Cumbria recording 208mm and some areas exceeding 100mm over 48 hours.

The Environment Agency reported that the main issue concerning this event was vulnerability of rapid response catchments. The severest impacts were seen in northern England where the exceptional rainfall led to record river levels and extensive flooding. The River Caldew, which drains into Carlisle before joining the Eden, reported river levels beyond those seen in the severe 2005 floods and the neighbouring Petrill also burst its banks. The River Darwen burst its banks causing floods at Darwen, Wigan and Oldham whilst the nearby River Yarrow also burst its banks affecting Croston in particular. In Rochdale, the River Roch flooded causing impacts in the surrounding areas. West Yorkshire and Calder Valley were also badly hit including locations such as Todmorden. Flash flooding in the Yorkshire Dales stranded cars near Hawes.

This case study focuses on the North West region of England and three small catchments (two less than 50km^2) that experienced serious floods: the Darwen at Ewood (39.08 km^2) and two nested sites on the River Roch at Littleborough (14.83 km^2) and Albert Royds Bridge (64 km^2). The study area is shown in Figure 5.41. A sequence of regional summary maps using the T+24h blended ensembles are presented in Figure 5.42 for forecast origins over the

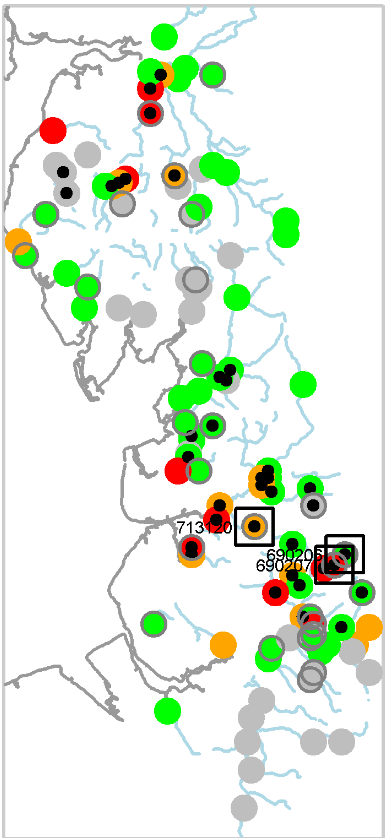


Figure 5.41 Regional summary map for case study area over North West for event of 22-24 June 2012. River gauging station locations are indicated by circles, those with a solid outline having a catchment area less than 50 km^2. Those highlighted by square boxes are used in Figure 5.43 and Figure 5.44.

period 21 to 22 June. The observed response (black dots) identifies two areas with notable flood responses. Firstly an area centred on the Darwen catchments was badly affected. Slightly later, the Caldew and Petteril that drain northwards from the Lake District to Carlisle cross the Q(5) thresholds. From the 19:15 21 June 2012 forecast onwards, the area focussed on the Darwen and Roch in the south is consistently highlighted. From 07:15 22 June 2012 the areas to the North around the Lake District start to be highlighted and correspond reasonably well to observed peaks in these regions. The sequence of maps show that assessing over a range of Q(T) grids is useful to identify potential areas with heightened flood risk. For example, studying the 12:15 22 June 2012 forecast in more detail for the Q(5) and Q(50) thresholds suggests that two areas are at particular

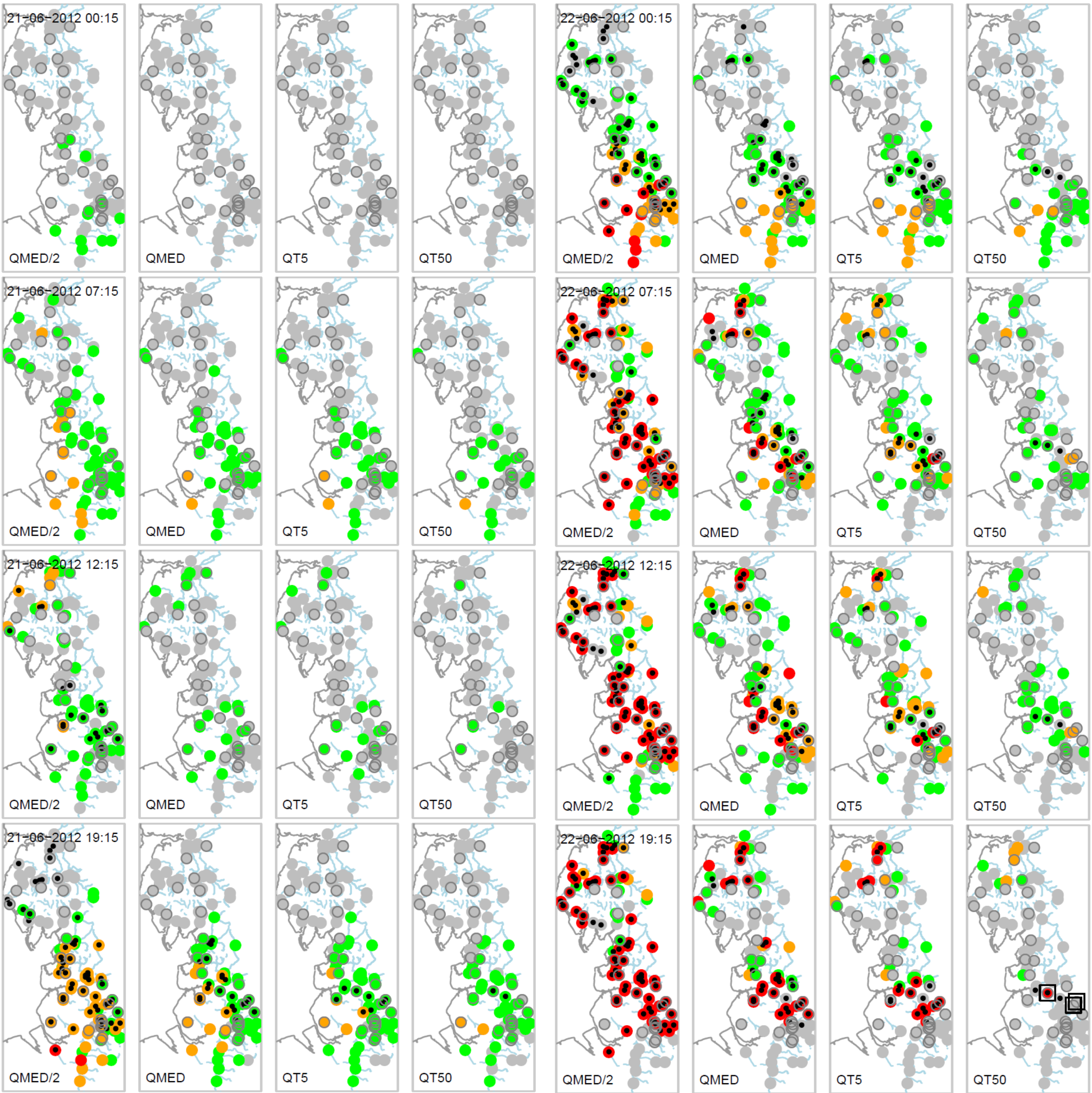


Figure 5.42 Regional summary maps over North West for 21-22 June 2012 obtained from ensemble G2G flow forecasts using T+24 blended ensemble rainfall forecasts. See Figure 4.5 for a full definition of the maps.

risk: the area centred on the Roch and Darwen and the area around the Lake District. However, the Q(50) map suggests that the Roch and Darwen area was more likely to experience a more significant flood (more sites identified and some orange) and this aligns well with the observed response (black dots).

Site specific forecasts are presented for the three small catchments in Figures 5.43 and 5.44 for the same sequence of 8 forecasts used in Figure 5.42. The simulation-mode performance (red lines) at these study sites is not particularly good for this event. The first signal of a possible event is from the forecasts at 00:15 and 07:15 21 June 2012 but the peak is forecast far too early. Nevertheless, the later forecasts appear to be useful and confirm that forecasts made on or after 19:15 21 June 2012 do a reasonable job of forecasting the event including the timing of the peak. It is worth noting that the 19:15 21 June 2012 forecast is 12 hours before observed river flows at any of the catchments start to rise and 24 hours before the study catchments peak.

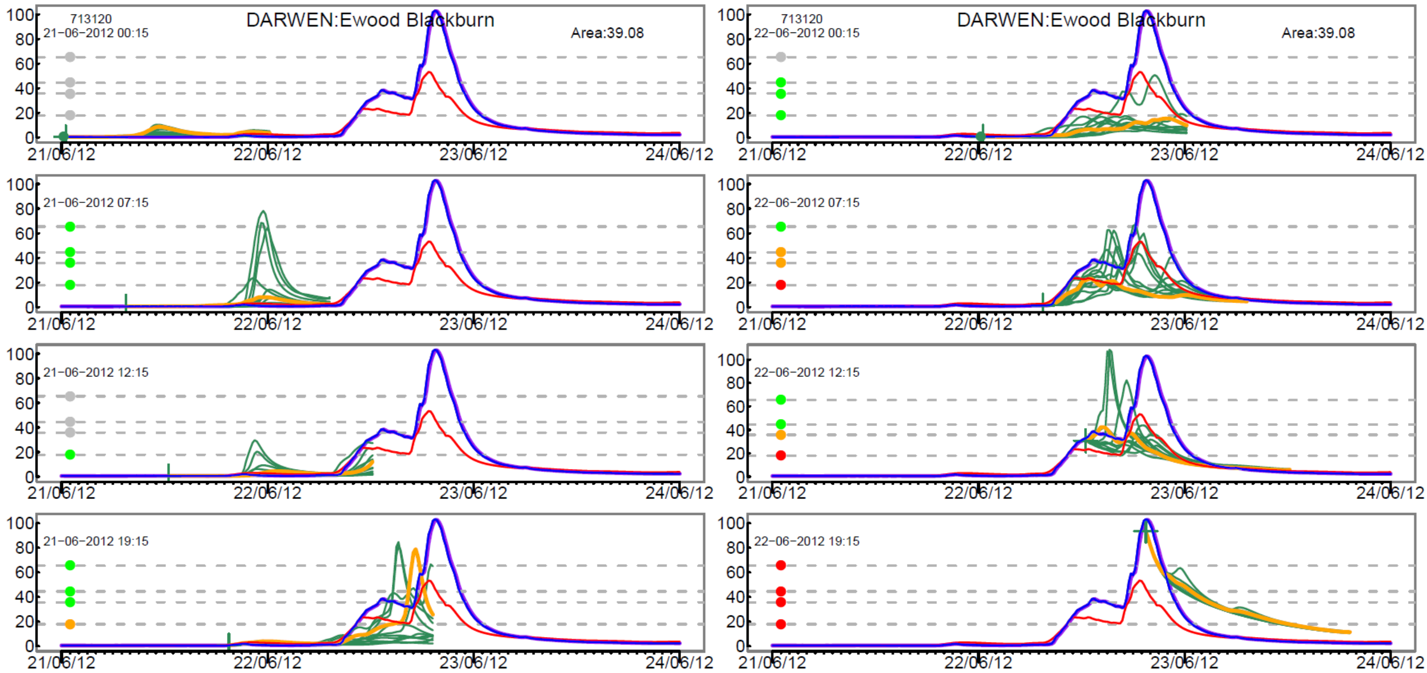


Figure 5.43 Ensemble G2G flow forecasts for the River Darwen at Ewood (713120) using a sequence of T+24 blended ensemble rainfall forecasts during 21-22 June 2012. See Figure 4.6 for a full definition of the time-series display.

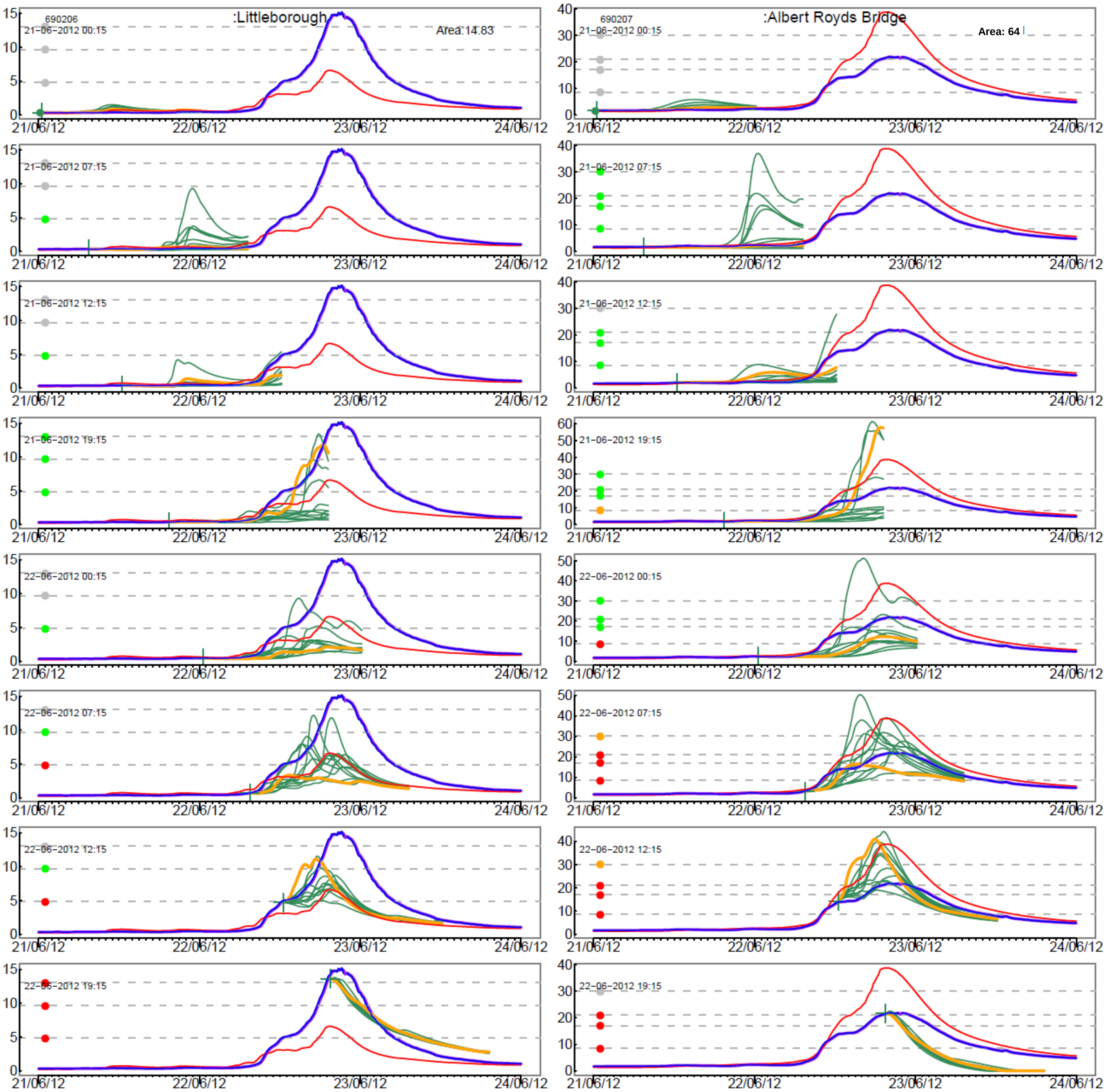


Figure 5.44 Ensemble G2G flow forecasts for Littleborough (690206) and Albert Royds Bridge (690207) using a sequence of T+24 blended ensemble rainfall forecasts during 21-22 June 2012. See Figure 4.6 for a full definition of the time-series display.

6-7 July 2012

A detailed account of the meteorology of the event is given by Almond (2013).During the night of 5 July and the following morning a band of heavy rain move northwards across East Anglia and into the north Midlands, north Wales and southern parts of northern England with 40-50mm accumulations widespread and nearly 60mm recorded in some south Pennine locations. In the afternoon and evening of 6 July, part of this band intensified and moved quickly north giving a period of heavy rain across Northumberland and reaching the Borders and Edinburgh on 7 July (15mm in 1 hour, 42mm in the day). Meanwhile the slow moving low pressure centred on the south-west of England bringing a period of persistent rain. Many raingauges recorded in excess of 100mm in the 18 hour period ending 12:00 on 7 July.

The impacts were widespread with the most sever being in the south-west of England. The additional information from Andrew Sibley in Almond (2013) details the impacts on the River Axe and its tributaries. Several gauging stations on the Axe recorded their maximum ever level: Whitford Bridge at 12:00 on 7 July (record starts 1964), Weycroft Bridge at 13:30 (record starts 1995). Also Winsham recorded its second highest level since records began in 1995. Further south, near Plymouth, the Yealm at Cornwood recorded a new maximum with records starting in 1995. Other impacts were noted in Yorkshire, north-east England, the Midlands, Anglia, Edinburgh, the Lothians and Borders.

This case study focuses on the South West region of England and four small catchments that experienced significant floods: two sites on or near the Axe (Chard Junction, 85.28 km^2; Asker at Bridport East Bridge, 46.6 km^2) and two on the south Cornish coast (Yealm at Puslinch, 54.9 km^2; Allen at Idless, 24.57 km^2). The study area is presented in Figure 5.45 and highlights the four sites. A sequence of regional summary maps using the T+24h blended ensembles are presented in Figure 5.46 for forecast origins during 6-7 July 2012. The observed river flows (black dots) show the south of Devon being most severely affected and generally lesser affects along localised areas of the south Cornwall coast. The 00:15 6 July forecasts gives the first indication of an event somewhere across Cornwall. Subsequent forecasts towards 19:15 6 July start to correctly identify specific areas of concern such as the areas

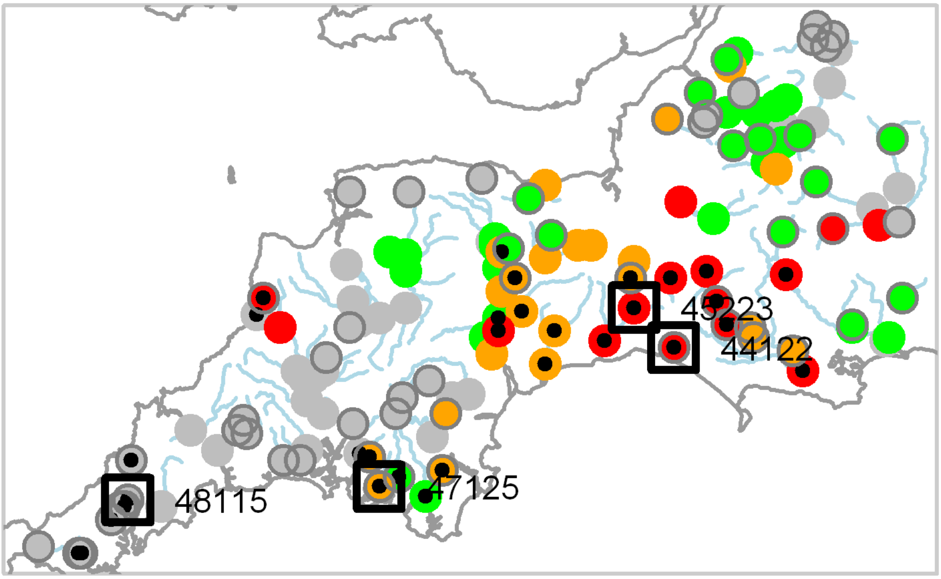


Figure 5.45 Regional summary map over the South West for event on 6-7 July 2012. River gauging station locations are indicated by circles, those with a solid outline having a catchment area less than 50 km^2. Those highlighted by square boxes are used in Figure 5.47 and Figure 5.48.

near and to the west of Plymouth, the southern tip of Cornwall and areas along the south Devon coast. This corresponds with the fact that Cornwall experienced the rainfall earlier and the catchments responded earlier. This is confirmed when comparing the gauge specific forecasts for the Allen near the south tip of Cornwall and the Yealm which drains from Dartmoor near Plymouth in Figure 5.47 with those for the two sites in or near the Axe catchment in east Devon that are presented in Figure 5.48. The 12:15 and 19:15 6 July forecasts, in Figure 5.47, show that the reasonable signal from the regional summary maps is still well in advance of any observed response at these two small catchments.

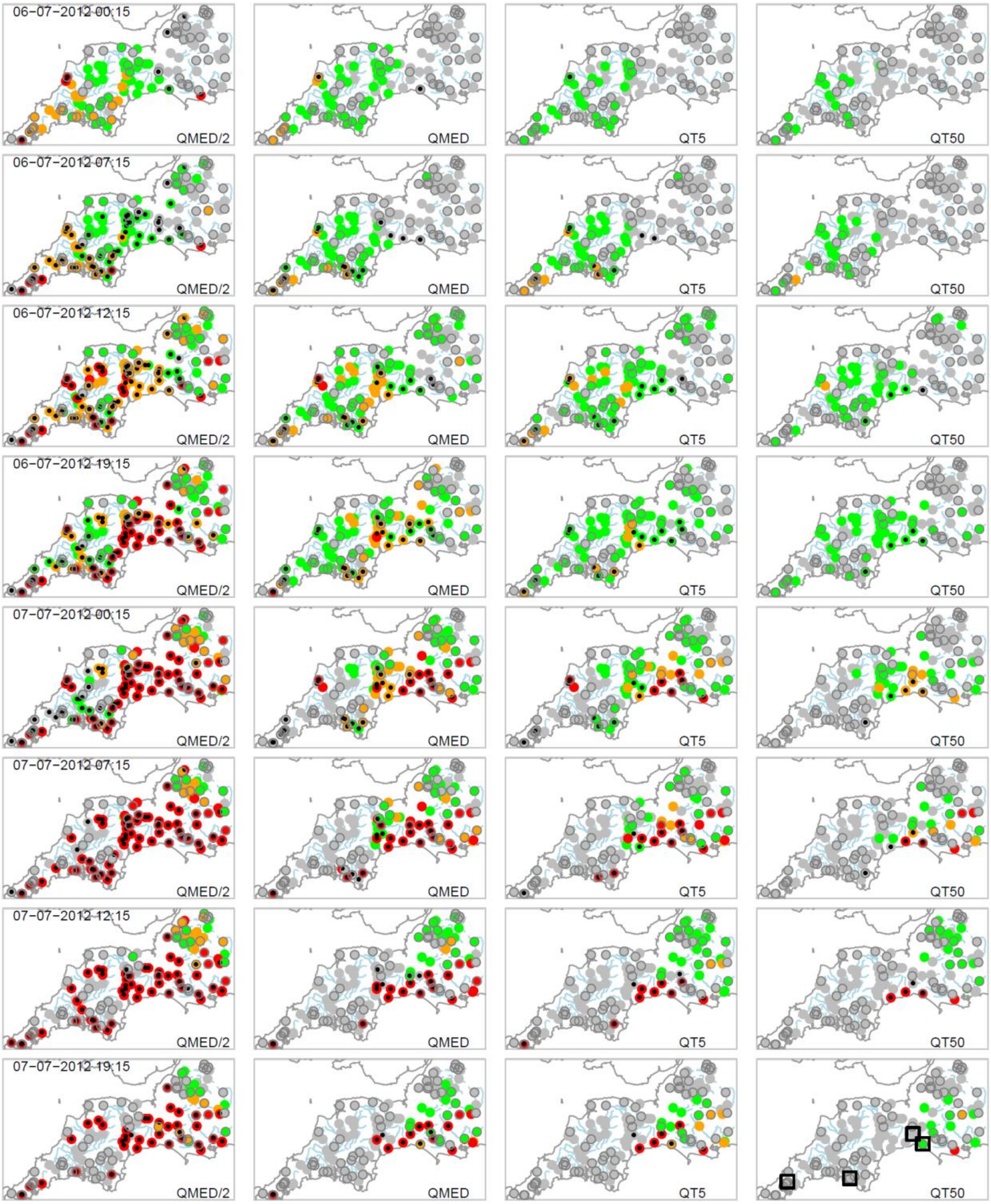


Figure 5.46 Regional summary maps over the South West during 6-7 July 2012 obtained from ensemble G2G flow forecasts using T+24 blended ensemble rainfall forecasts. See Figure 4.5 for a full definition of the maps.

For forecasts from 19:15 6 July onwards, the area of concern moves to further east and increasingly identifies the River Axe and surrounding catchments as the area most likely to experience a severe flood although it extends a little too far east when compared to observations (e.g. the 07:15 7 July forecast). The forecasts also suggest that the floods here are likely to be more severe (in terms of return period) than the slightly earlier floods in Cornwall.

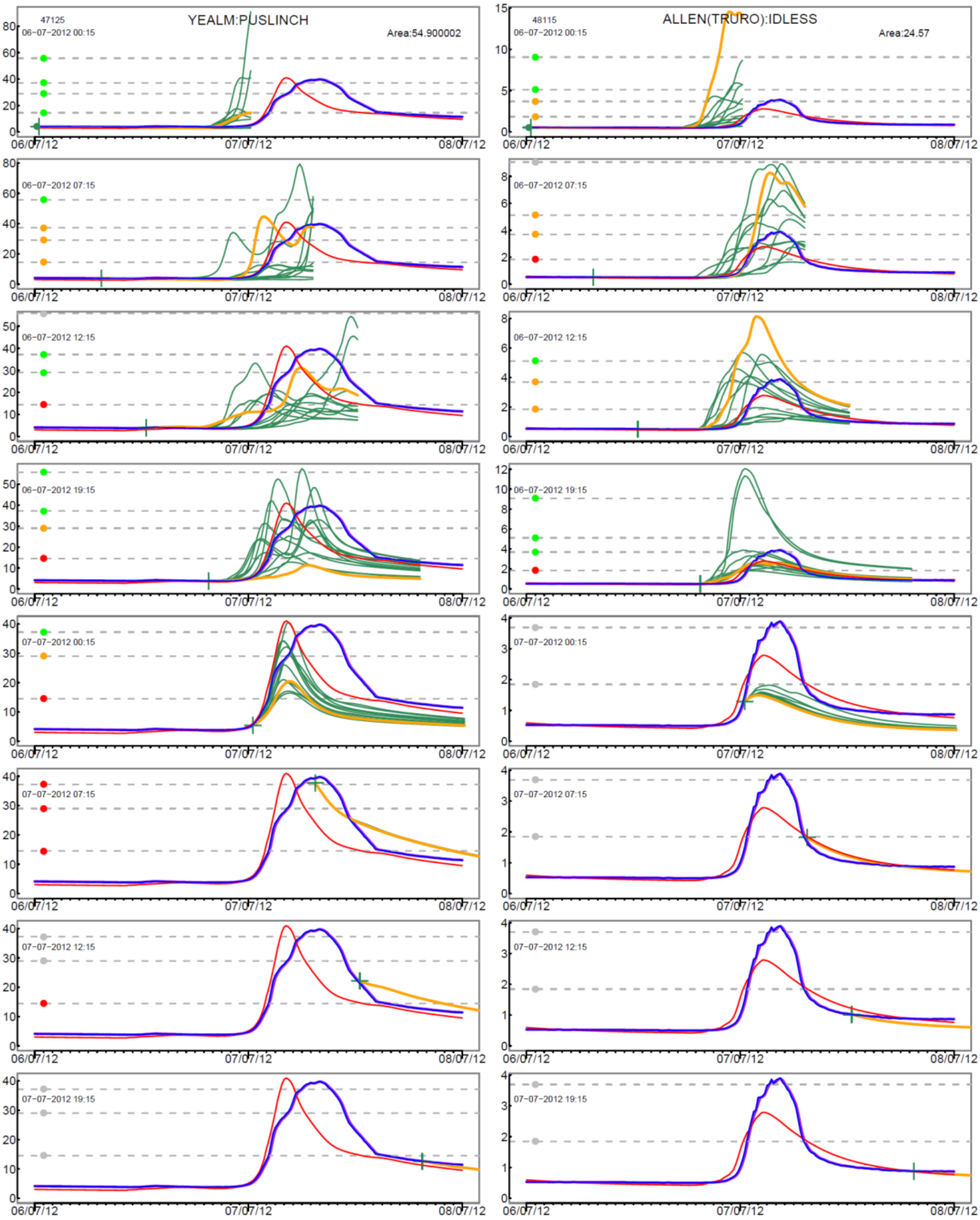


Figure 5.47 Ensemble G2G flow forecasts for the Yealm at Puslinch (47125) and Allen at Idless (4815) using T+24 blended ensemble rainfall forecasts during 6-7 July 2012. See Figure 4.6 for a full definition of the time-series displays.

In summary, the large-scale low pressure system dominating the rainfall generation for this case study is generally well forecast. At a high level, the T+24hr blended ensembles successfully identify the sequence of areas affected and also the relative severity with the River Axe area being forecast later and with a high likelihood that flooding will be more severe. Also a reasonable lead-time is achieved for the site specific spaghetti plots with the first major signal from the forecasts being associated with forecast origin times that are before the first observed response.

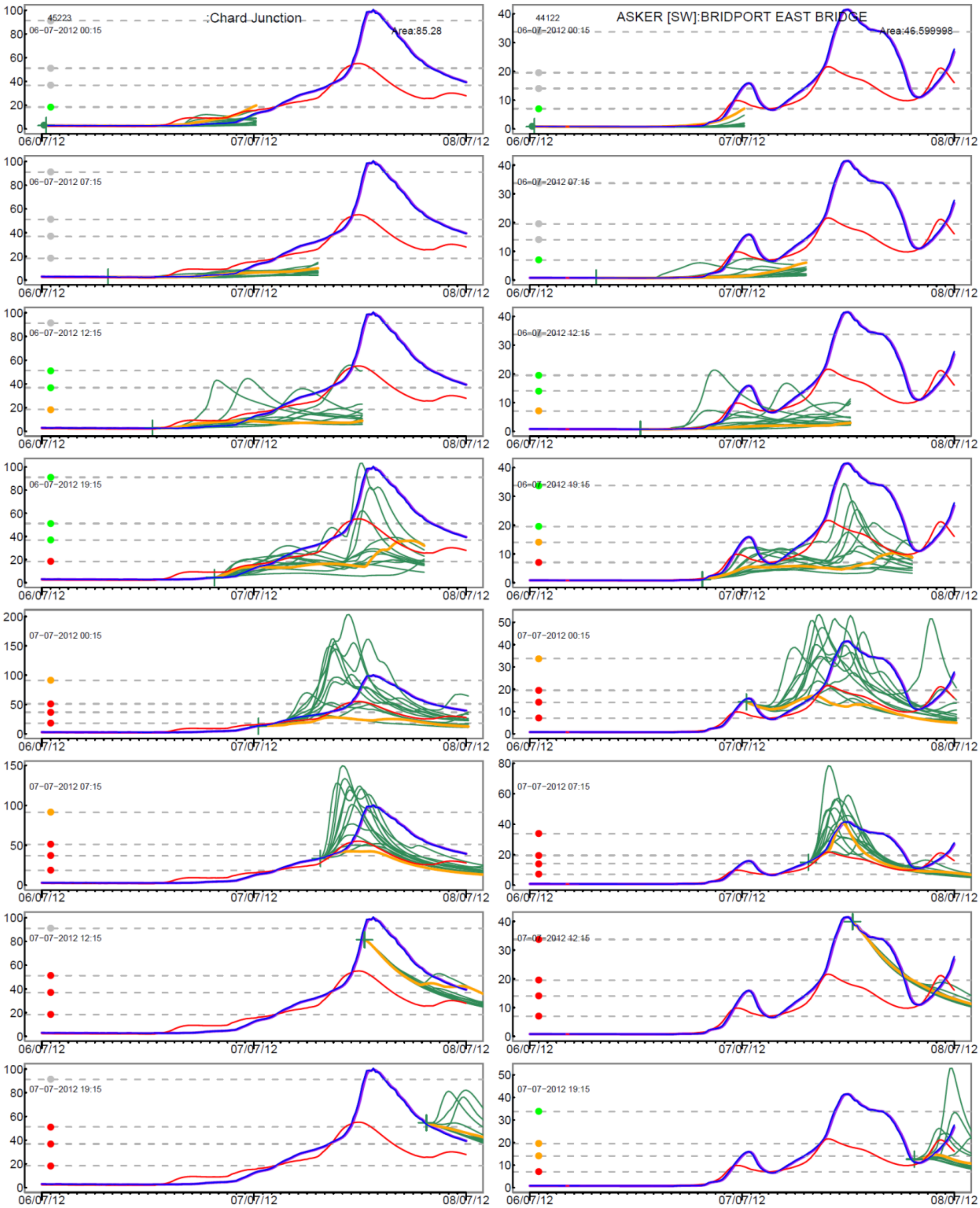


Figure 5.48 Ensemble G2G flow forecasts for Chard Junction (45223) and Asker at Bridport East Bridge (44122) for 4-7 July 2012 using T+24 blended ensemble rainfall forecasts during 6-7 July 2012. See Figure 4.6 for a full definition of the time-series displays.

5-6 September 2008

A deep low pressure system moved north-eastwards across south-west England later on the 5th, to become slow moving across the Midlands and eastern England on the 6th and 7th, before finally moving out into the North Sea on the 8th. This brought periods of heavy and thundery rain during the period to all parts and prolonged heavy rainfall. Flooding was reported in south-west England on the 5th and in north-east England on the 6th. High 48-hour rainfall totals were recorded in Northumberland on the 5th/6th, with 158.3mm at Chillingham Barns and 151.5mm at Morpeth Cockle Park (80mm in 24 hours, the highest since records began in 1898), both having return periods in excess of 200 years. Godscleugh recorded over 250mm in a 3-day period.

There were over 100 flood warnings issued by the Environment Agency over the two days with flooding reported across Wales, south-west England, the Midlands and the north-east. The most severe flooding was on the River Wansbeck through Morpeth with a new record flow at Mitford (gauging started in 1963) estimated to have a return period exceeding 100 years. Around 1000 properties were flooded in Morpeth. Nearby catchments also recorded new maxima including the Till catchment just to the north as well as the Yscir in south Wales.

In contrast to the previous case studies, this analysis uses a single 24 member UKV ensemble rainfall forecast to T+30 h starting at 12:00 5 September 2008. This is a forerunner to the MOGREPS-UK product planned to become operational towards the end of 2013. A regional summary map is displayed in Figure 5.49 for the North East This map suggests that there would have been a reasonably strong signal of flooding from the forecast even though the peak at Mitford in Morpeth was over 24 hours later. Over the North East as a whole the ensemble appears to perform reasonably well at a QMED/2 and QMED threshold. For the more extreme flow thresholds of Q(5) and Q(50) the maps correctly identify a catchment to the south of Wansbeck which observed a high flow but also incorrectly highlighted two clusters of river gauging locations further south, one being at the fluvial lower end of the Yorkshire Ouse.

The forecasts for the Wansbeck at Mitford (287.3 km^2) and three upstream sites (Middleton Bridge, 62.83 km^2; Hartburn 53.07 km^2; Nunnykirk, 33.12 km^2) are considered in more detail in Figure 5.50. Simulation-mode G2G flow forecasts using raingauge–only rainfall (red lines) show that the catchment is reasonably well modelled with the exception of Nunnykirk where peaks are generally underestimated. Observed flows at all sites cross the Q(50) thresholds except for Middleton Bridge which only cross at the Q(5) level. The ensemble flow forecasts (green lines) show a reasonable signal of a serious event occurring with green for Q(50) and red for Q(5) except Middleton which has orange for Q(5) and corresponds with the site with the least severe observed flow. Notably, almost all the ensemble peaks are significantly earlier than the observed flow peaks.

This is the only case study that uses a “true” NWP-based rainfall ensemble where each member has slightly different initial conditions that effect the subsequent evolution of the weather (the other blended ensembles in this study use the STEPS methodology to generate further ensemble members from a single deterministic NWP forecast). As such, there is a limit to what can be drawn from a single case study but there are some encouraging signs that the approach shows some promise and adds value beyond using a single deterministic forecast (orange line). However, although the Q(5) threshold identified several flood risk areas correctly, there are other areas highlighted orange or red that did not observe the same level of flood peak. This merely serves to highlight that longer periods of analysis are required to fully assess these probabilistic forecasts. This also applies to the soon-to-be operational MOGREPS-UK 2.2km rainfall ensemble which will require analysis over a reasonable length of time to make meaningful assessments of its performance for flood forecasting in rapid response catchments.

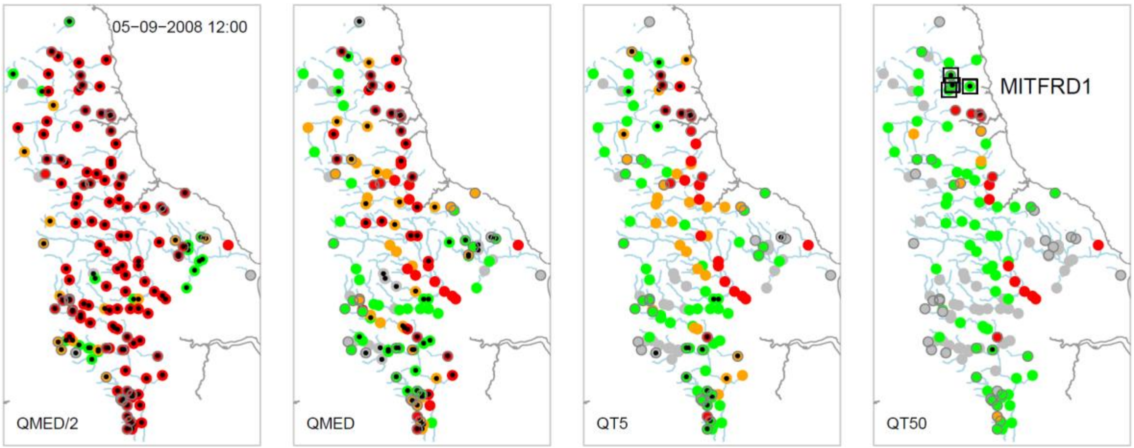


Figure 5.49 Regional summary maps of ensemble G2G flow forecasts for the North East using a 24 member UKV rainfall ensemble starting 12:00 5 September 2008. The Wansbeck at Mitford and gauges upstream of it are highlighted by square boxes and used in Figure 5.50. See Figure 4.5 for a full definition of the maps.

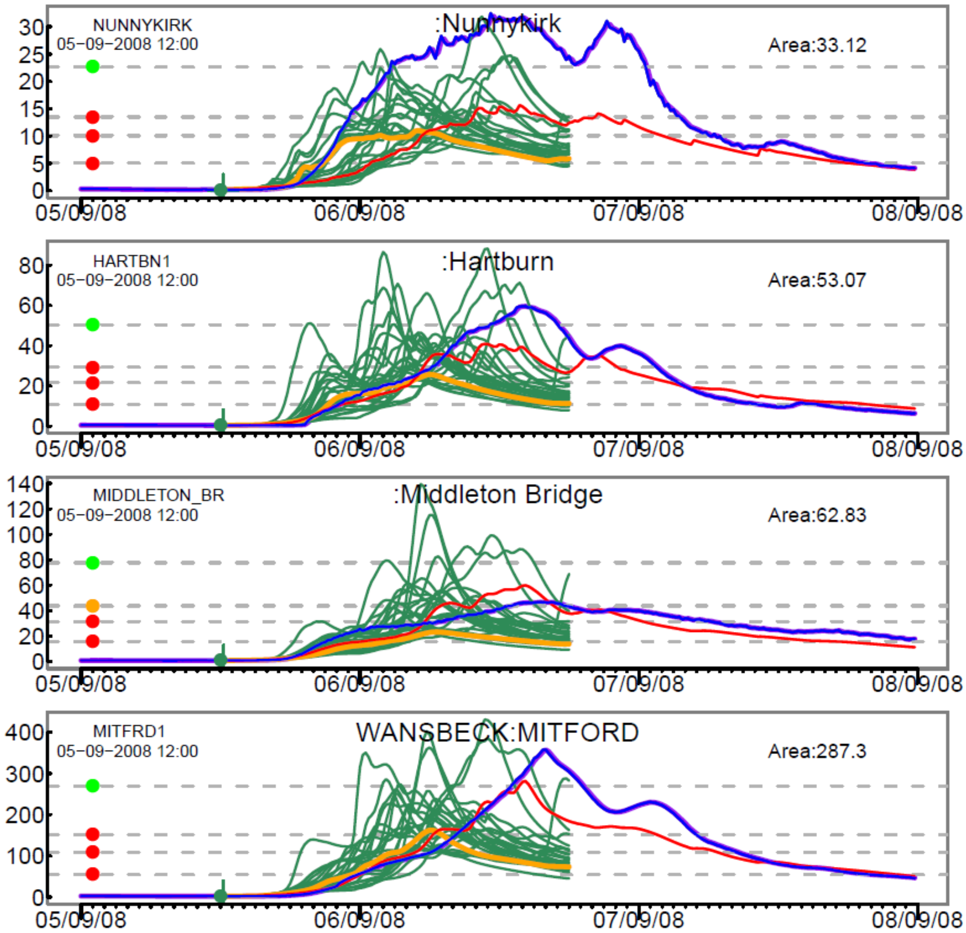


Figure 5.50 G2G ensemble flow forecasts for the Wansbeck at Mitford (MITFRD1) and gauges upstream of it using the UKV ensemble rainfall forecast at 12:00 5 September 2008. See Figure 4.6 for a full definition of the time-series display.

5.5.3 Longer-term assessment of ensembles

The live datafeed to CEH of Blended ensemble rainfall forecasts started on 10 May 2012. Based on availability of the hydrometric datasets, the longer-term assessment periods for analysis were selected: they are set down in Table 5.1 and repeated below for completeness.

- England & Wales. 10 May to 31 December 2012 (~7½ months)
- Scotland. 10 May to 30 September 2012 (~4½ months)

Fortunately for the analysis, many flood events occurred during these periods (see Table 5.2), particularly across England and Wales, although the unusual frequency of events may need to be considered when interpreting the results.

The continuous assessment measures used are discussed in detail in Section 0 and a brief summary is given here. The Relative Operating Characteristic (ROC) Score (ranges between 0 and 1) is used and assesses the ability of probabilistic forecasts to discriminate between event and non-events with 1 being a perfect score. The Brier Skill Score (1 or less) is also used and assesses the relative skill of the probabilistic forecast over that of a "climatology" or "reference" forecast, in terms of predicting whether or not an event occurred.

Longer-term assessment of G2G ensemble forecasts has focussed on using the T+24 blended ensemble rainfall forecasts that have 12 members. The definition of an event for the continuous assessment is exceedance of the QMED/2 threshold during the 24 hour forecast horizon and aligns with how the ensemble information is summarised in the case-study assessments. A relatively low threshold is used to try and ensure a reasonable number of events for each site. For the Brier Skill Score, it is important to note that the reference or climatology Brier Score is calculated using a constant probabilistic forecast that equates to the observed frequency of the event happening over the short ensemble assessment period (i.e. the observed flow exceeding the QMED/2 threshold during the forecast horizon).

The assessments have been made using (i) all ensemble members, and (ii) one ensemble member. This is done in order to make some assessment of the benefit ensemble forecasts may have over using a single deterministic forecast. Also, as an observed time-series is needed to calculate the scores, the assessment only considers gauged locations.

England & Wales

Spatial assessment maps are provided for ROC scores and Brier skill scores in Figures 5.51 and 5.52 respectively for the 7.5 month assessment period. Small catchments less than 50km^2 have been highlighted and a further breakdown by catchment area is given in the bar charts. The ROC scores in Figure 5.51 for the ensembles (top left map) show good scores in excess of 0.6 for large areas. There are a few clusters where the scores appear to be relatively lower such as Dartmoor, southern and north western areas of Wales and areas of the Lake District and Pennines. The scores obtained using only one ensemble member, akin to a deterministic forecast, are displayed in the top right map and perform markedly less well than the full ensemble. In addition to the areas of relatively poor performance highlighted by the full ensemble, the south east and areas of the southern coast are identified as having relatively poorer performance when using a single ensemble member. The bottom left map shows the difference in ROC score between using all members and only one and highlights the areas that benefit most by using 12 rather than 1 ensemble member. Analysing the breakdown by catchment area (bottom right map) shows a general trend of the ROC score improving

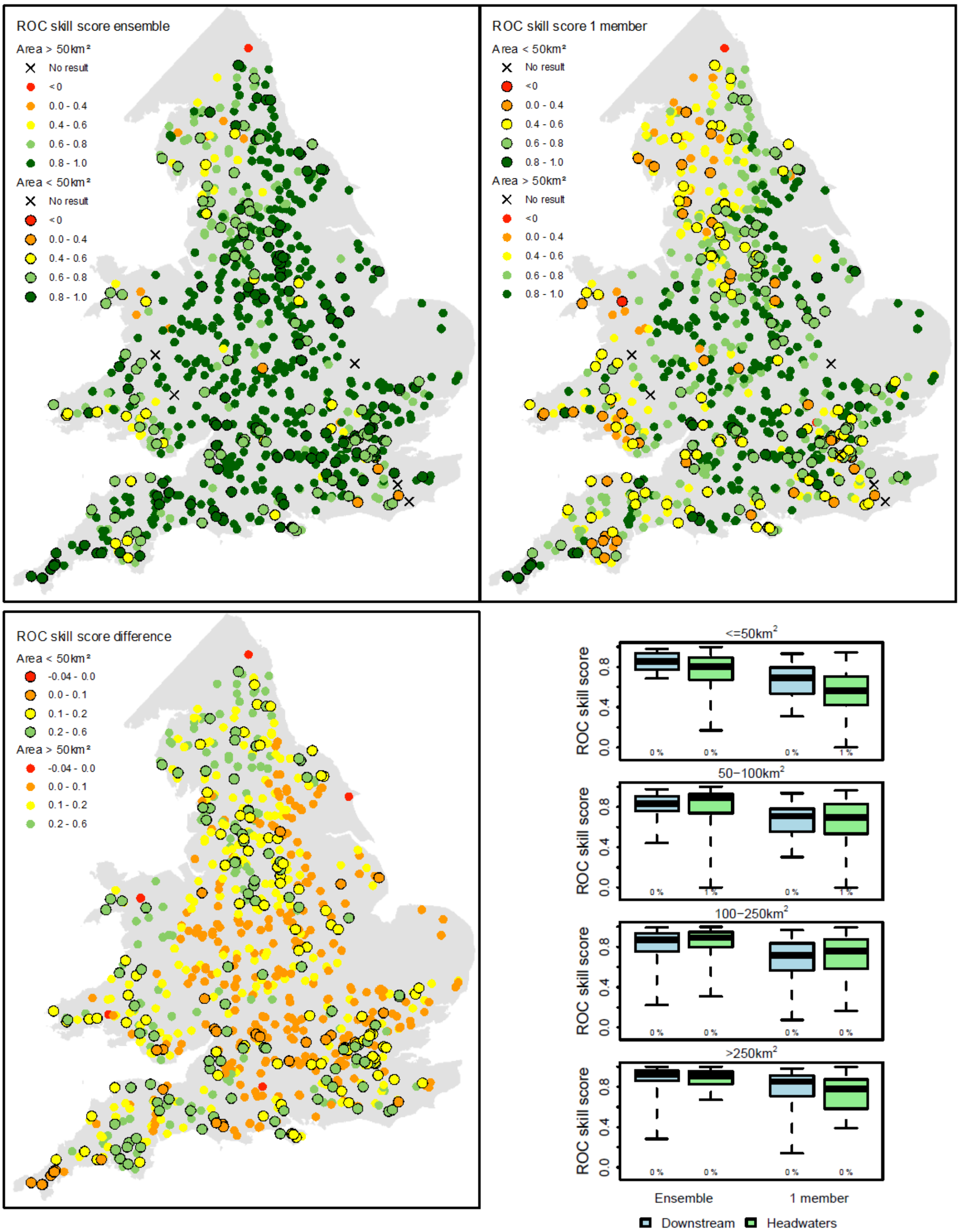


Figure 5.51 England & Wales. ROC skill score analysis of ensemble G2G flow forecasts using the QMED/2 threshold. Uses T+24 blended ensembles over the period May to December 2012. Top left: uses all ensemble members. Top right: uses just one ensemble member. Bottom Left: difference between using all 12 members and only one member. Bottom right: pooled analysis by catchment area.

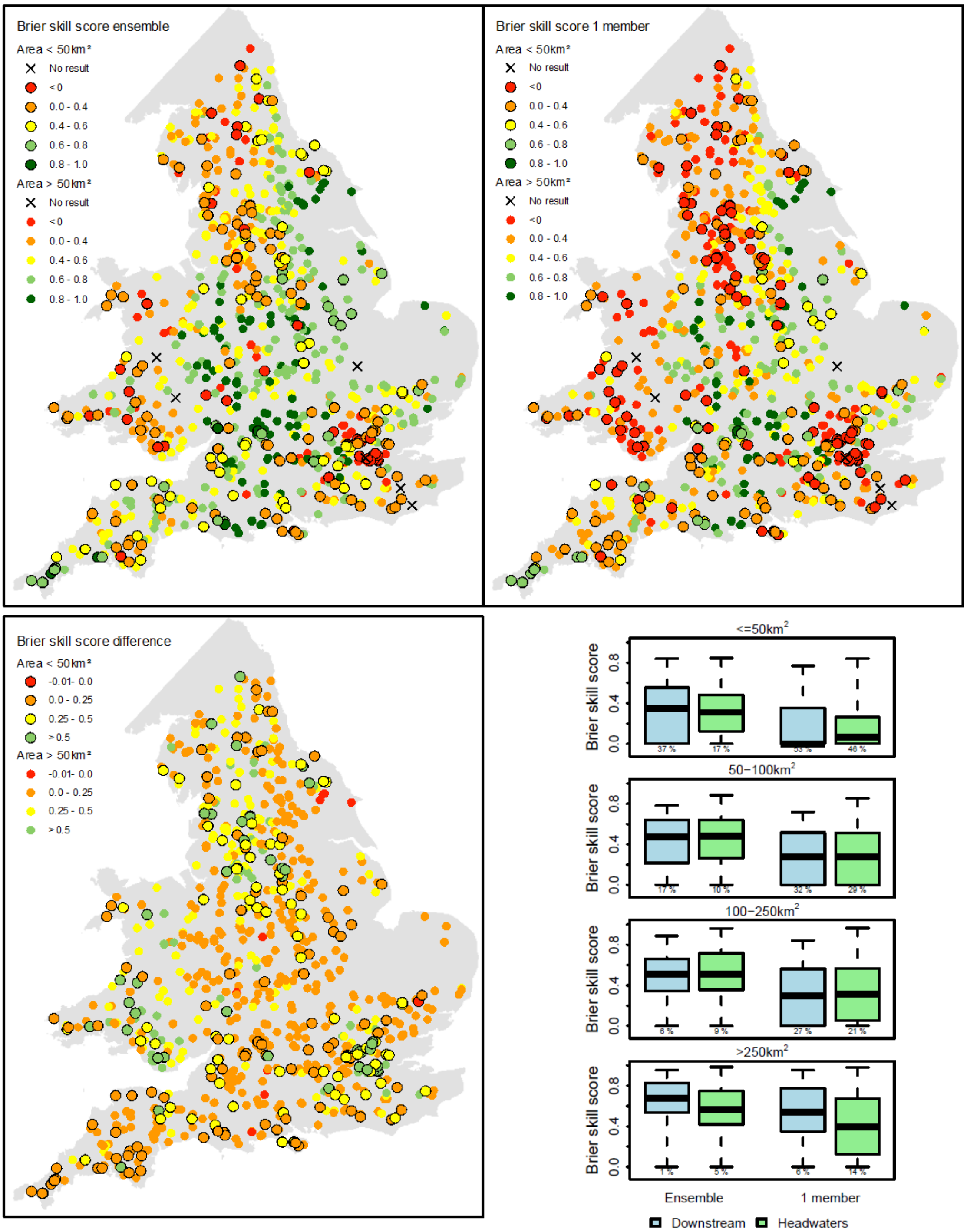


Figure 5.52 England & Wales. Brier skill score analysis of ensemble G2G flow forecasts using the QMED/2 threshold. Uses T+24 blended ensembles over the period May to December 2012. Top left: uses all ensemble members. Top right: uses just one ensemble member. Bottom Left: difference between using all 12 members and only one member. Bottom right: pooled analysis by catchment area.

with increasing catchment area and the most benefit from using all ensemble members being for the smaller catchments.

The Brier skill score analysis is presented in Figure 5.52. Broadly the spatial patterns of relatively good and poor performing regions agree well with those for the ROC score. However some additional areas are identified as having relatively poorer performance such as the London, South East, parts of the southern coast and the Southern Pennines and Peak District. Interestingly, both Brier skill score plots (top row) show significant numbers of sites where the reference Brier score performs better than the probabilistic forecast (red dots). At this stage the precise reasons for this are not entirely clear but are likely to be due to a combination of the low exceedance threshold (QMED/2) and relatively short assessment period.

An alternative approach for making a pooled analysis of the results at a regional scale is presented in Table 5.3. Here the definition of an event is 10% of sites exceeding the QMED threshold over the forecast horizon. Pooling over sites in this way has allowed an assessment at the higher thresholds of QMED and Q(5) but as the threshold increases, the number of observed "events" becomes less and make the assessment less useful. For example, at the Q(5) threshold, Wales treated as a region in the analysis has no observed events so no calculation can be performed. The results show that good ROC scores are achieved for both thresholds but this decreases with increasing thresholds. The Brier skill scores are positive for all regions at the QMED level showing an improvement over the reference forecast. For the Q(5) threshold the Brier skill score drops off with two regions (North West and Southern) performing less well than the reference forecast. Again the comments about the site Brier skill scores above are valid here.

Table 5.3 Regional pooled analysis of G2G flow ensembles using the T+24 blended ensemble rainfall forecasts.

Event	Skill score	Region							
		North West	North East	Midlands	South West	Southern	Thames	Anglian	Wales
10% of sites exceeding QMED	ROC score	0.996	0.988	0.983	0.988	0.974	0.968	0.973	0.973
	BSS	0.403	0.435	0.429	0.339	0.226	0.379	0.345	0.035
10% of sites exceeding Q(5)	ROC score	0.971	0.987	0.970	0.979	0.985	0.947	0.968	-
	BSS	-0.520	0.363	0.300	0.181	-0.211	0.177	0.173	-

Scotland

Spatial assessment maps are given for Scotland with the ROC score presented in Figure 5.53 and the Brier skill score presented in Figure 5.54. It must be remembered that the Scotland assessment is only using 4.5 months and mainly covers the summer months (May to September) so cannot be directly compared to the longer period used for the England & Wales analysis. The ROC scores when using the entire ensemble are good (0.6 or above) for a reasonable number of sites. A few clusters are identified as having relatively poorer performance, particularly some sites in the central belt and near Edinburgh, a group near Elgin and a group midway along the east coast. Interestingly these areas correspond reasonably well with the sites where the G2G flow forecasts using UKV rainfall performed relatively poorly (see Figure 5.4). Comparing the scores for the full ensemble and for the "deterministic" single ensemble member (

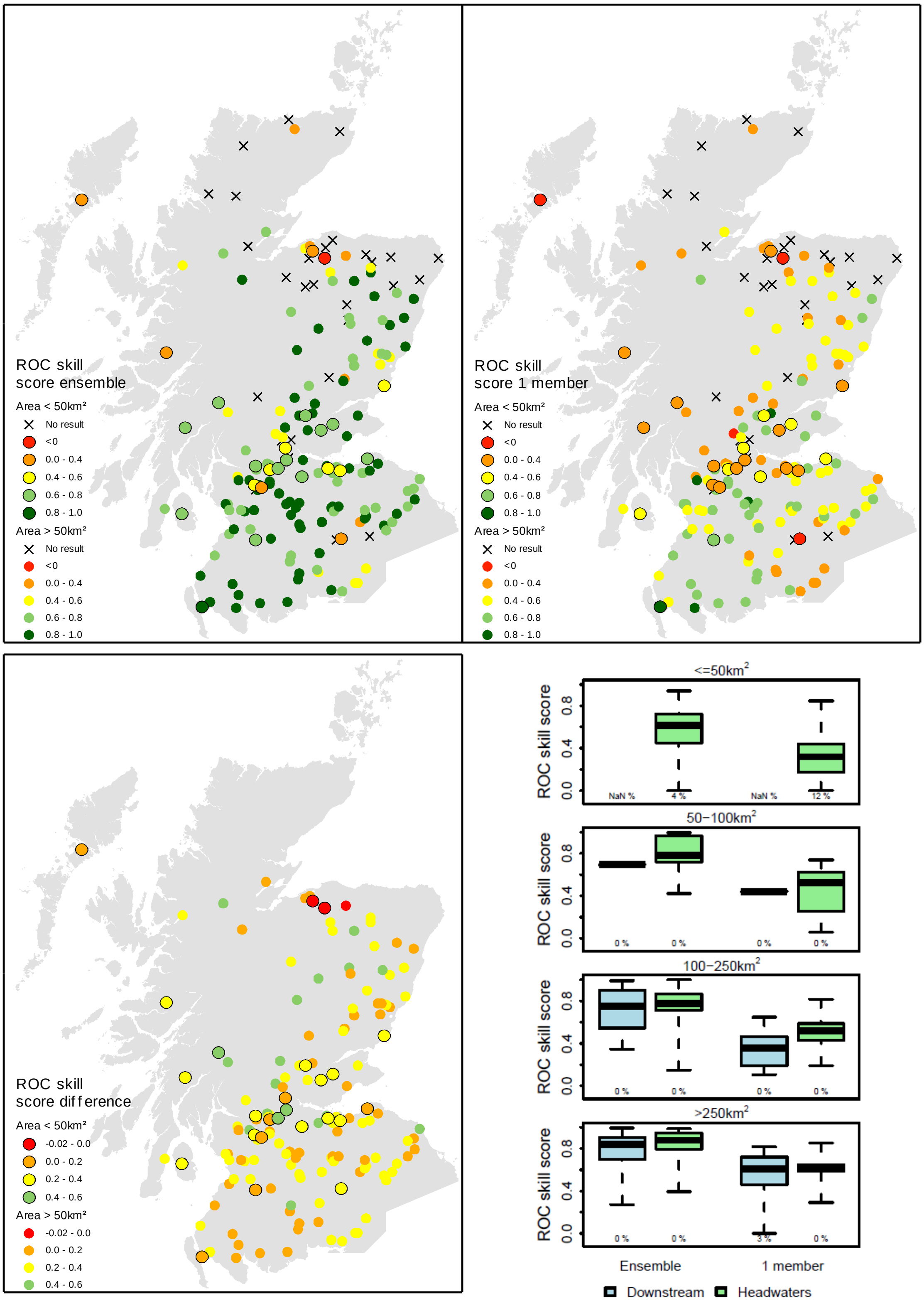


Figure 5.53 Scotland. ROC skill score analysis of ensemble G2G flow forecasts using the QMED/2 threshold. Uses T+24 blended ensembles over the period May to September 2012. Top left: uses all ensemble members. Top right: uses just one ensemble member. Bottom Left: difference between using all 12 members and only one member. Bottom right: pooled analysis by catchment area.

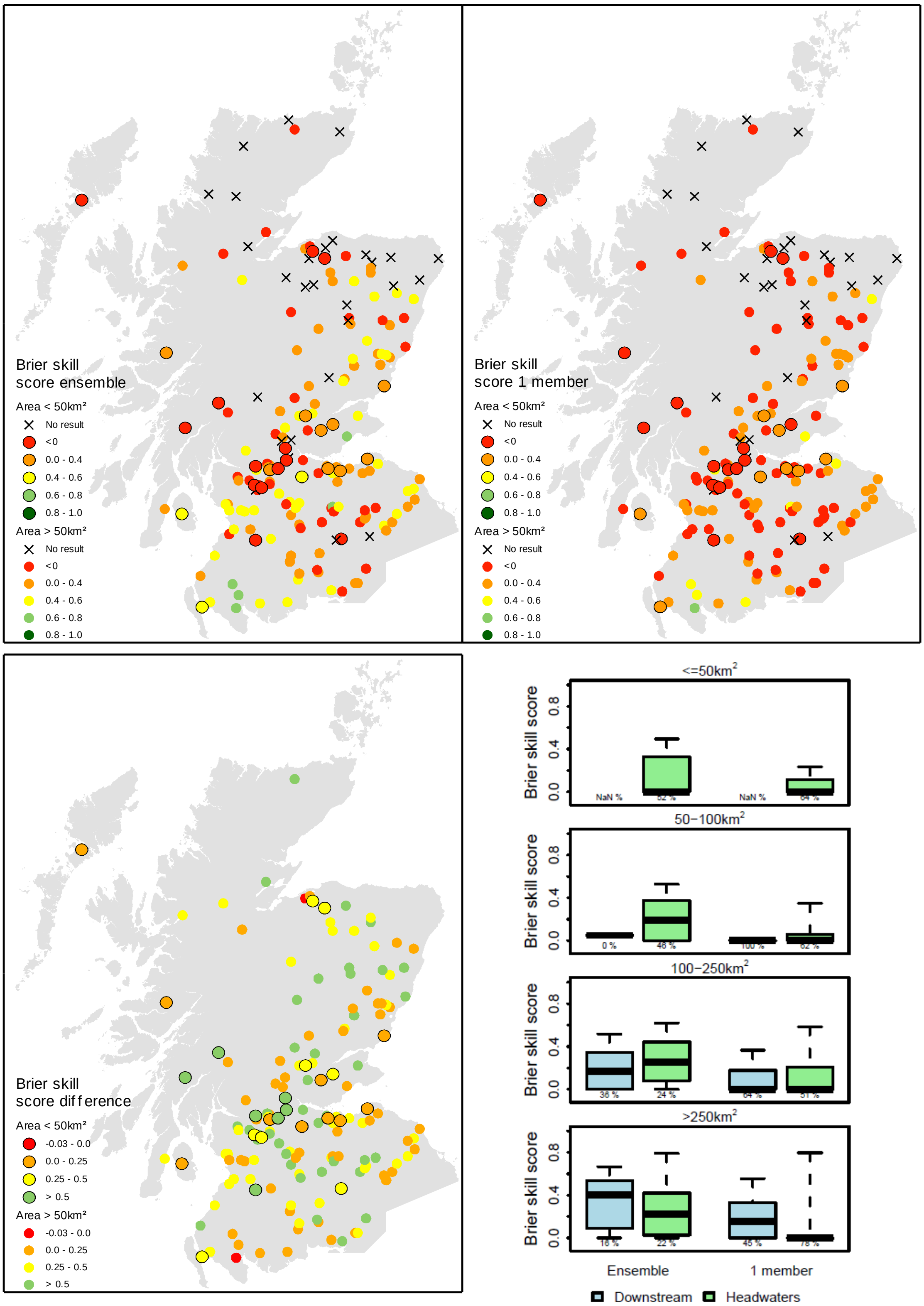


Figure 5.54 Scotland. Brier skill score analysis of ensemble G2G flow forecasts using the QMED/2 threshold. Uses T+24 blended ensembles over the period May to September 2012. Top left: uses all ensemble members. Top right: uses just one ensemble member. Bottom Left: difference between using all 12 members and only one member. Bottom right: pooled analysis by catchment area.

top row of Figure 5.53) confirms the improvement gained by using ensemble forecasts and is reiterated by the bottom left map of the differences in the ROC scores. Similarly to England & Wales, the box plots broken down by catchment area (bottom right of Figure 5.53) show a trend of the ROC score increasing with area. An even more dramatic improvement is seen when moving from use of deterministic to ensemble rainfall forecasts for the small catchments.

The Brier skill score analysis presented in Figure 5.54 is noticeably worse than the England & Wales equivalent. The top row of maps shows that a significant number of forecasts sites do better when using the reference forecast (constant probabilistic forecast equal to the observed frequency of the event happening) rather than forecasts using the full ensemble or a single ensemble member. As discussed earlier, the likely cause for this is due to the much shorter Scottish assessment period, summer months being prevalent, and possibly the low number of observed crossings over this period.

Summary of the long-term assessment of ensembles

This is the first time that there has been any longer-term assessment of the G2G river flow ensembles using the T+24 blended ensemble rainfall forecasts. The analysis is constrained by the length of observed verification data available (4.5 to 7.5 months) but still marks an improvement over individual case-study analysis. Ideally longer periods of assessment is preferred and would allow assessments of the extreme events or at higher Q(T) thresholds. Due to the short period of record, a low assessment threshold of QMED/2 has been used for site specific analysis. By defining an event to require pooling over all sites in a region has allowed some assessment at QMED and Q(5) thresholds.

The analysis presented shows that over England, Wales and Scotland, the ROC score and Brier skill score generally increase when using the full 12 member ensemble compared to the using one member as representative of a deterministic forecast. This provides strong evidence that using the full ensemble and probabilistic forecasts should allow better guidance to be given on forecast flood risk. Some regional trends have been identified, particularly areas in England and Wales, where performance appears to be relatively poorer when compared to other areas.

In terms of the quantitative score values, Scotland appears to perform less well than England & Wales but it is believed this is mainly due to the much shorter period used for assessment (May to September) and its dominance by the summer season. As commented on earlier, the Brier score and consequently also the Brier skill score will be sensitive to the short assessment periods available and is the likely cause of the noticeably poorer performance of the Brier skill score over Scotland.

Overall this first assessment of a long period of G2G ensemble flow forecasts obtained using ensemble forecast rainfall as input has been very useful and adds extra information beyond the traditional case-study assessment approach. Nevertheless, further work on analysing the flow and rainfall ensembles is likely to prove beneficial and yield deeper insights of value to better understanding and improvement. The assessment would also benefit from looking at a longer period that encompasses all seasons and allow analysis at the more extreme thresholds, especially if sites are pooled in some way.

6 Operational tools for using G2G ensembles

Operationally the Flood Forecasting Centre (FFC) use NFFS-FFC and the Scottish Flood Forecasting Service (SFFS) use FEWS-Scotland to run and view G2G forecasts in deterministic and ensemble form. Both systems are based on the Deltares Flood Early Warning System (FEWS) so have common functionality and are used to help inform the FFC and SFFS Flood Guidance Statements. It has become clear from the long-term and case-study analysis of G2G ensemble flows that the appropriate presentation of the large ensemble outputs is key in realising the operational benefits for forecasting in Rapid Response Catchments. This has been confirmed through discussions with operational hydrometeorologists at the FFC and SFFS. It is envisaged that initially the operational benefits for Rapid Response Catchments are best achieved at a national scale through constructing better post-processed G2G outputs to support the existing Flood Guidance Statement processes. Following this there is scope for more direct engagement and use within the regions (particularly of the Environment Agency and Natural Resources Wales) but this is probably best viewed as a separate and later activity.

Both FFC and SFFS currently use and view G2G forecasts in slightly different ways. A brief but not exhaustive overview of the current approaches used by each organisation is given below followed by some initial suggestions of alternative post-processing and display options.

6.1 Flood Forecasting Centre

The Flood Forecasting Centre currently run several different types of G2G flow forecast at varying frequency. A brief overview is given below.

- Deterministic 5-day G2G forecast using a combination of NWP products, run 4 times a day.
- Ensemble G2G forecast, using a 24 member blended rainfall forecast out to T+24 h, run 4 times a day.
- Ensemble G2G forecast, using a 24 member blended rainfall forecast/nowcast out to T+7 h, run every hour.

The Flood Forecasting Centre is moving to replace the UKV element of the blended ensemble forecast with the MOGREPS-UK 2.2km NWP ensemble, with a plan for this to be in place by the end of 2013.

The deterministic 5-day forecast is used to support the Flood Guidance Statement and is often the starting point. There are a number of displays used to assist this. For the spatial countrywide displays, 15 minute gridded values of flow and equivalent Q(T) threshold are available. The remaining spatial displays are all based on Q(T) thresholds (referred to as “warning level”) and include the warning level at gauged locations only (15 minute time-step) which are represented by a coloured circle. Finally there are post-processed products that show the maximum warning level on a county (days 1-3) or regional (days 4-5) basis for every 15 minutes and also over a daily time period. The maximum warning level displays are simply the maximum value across an area. These have proved problematic in the past as they are sensitive to outliers such as single “rogue” pixels where the Q(T) values may be too low and incorrectly signal a flood event. Some effort has been made to mitigate these issues by masking out problem pixels but some issues remain. Other ongoing work funded by FFC – the "G2G Q(T) Grid Developments: Scoping Study" - is looking at the Q(T) grids in more detail.

All the 5-day spatial displays are presented as 15 minute time-series. Example snapshots of some of the displays are given in Figure 6.1. For England and Wales, it is not easy to see the gridded values at a national level (bottom left) even though the time bar at the top of the display shows the maximum grid value for each time-step and highlights that something is occurring at this time-step. The regional and gauged location displays (top row) are helpful in this context as they locate the spatial areas where the maximum Q(T) levels have been reached and then the user can zoom in to that area on the gridded display to investigate further (bottom right).

For ensemble forecasts, a spatial summary is presented as numbers of ensemble members crossing given Q(T) levels: the QMED and Q(50) thresholds are currently used. These are presented as gridded values and at gauged locations. In addition, a county maximum of the number of pixel crossings is given: an example is given in the left hand side of Figure 6.2. All these displays are presented as sequences of 15 minute images with the top time bar giving the summary of maximum values across the domain for each time-step to help identify the timing of a potential flood event.

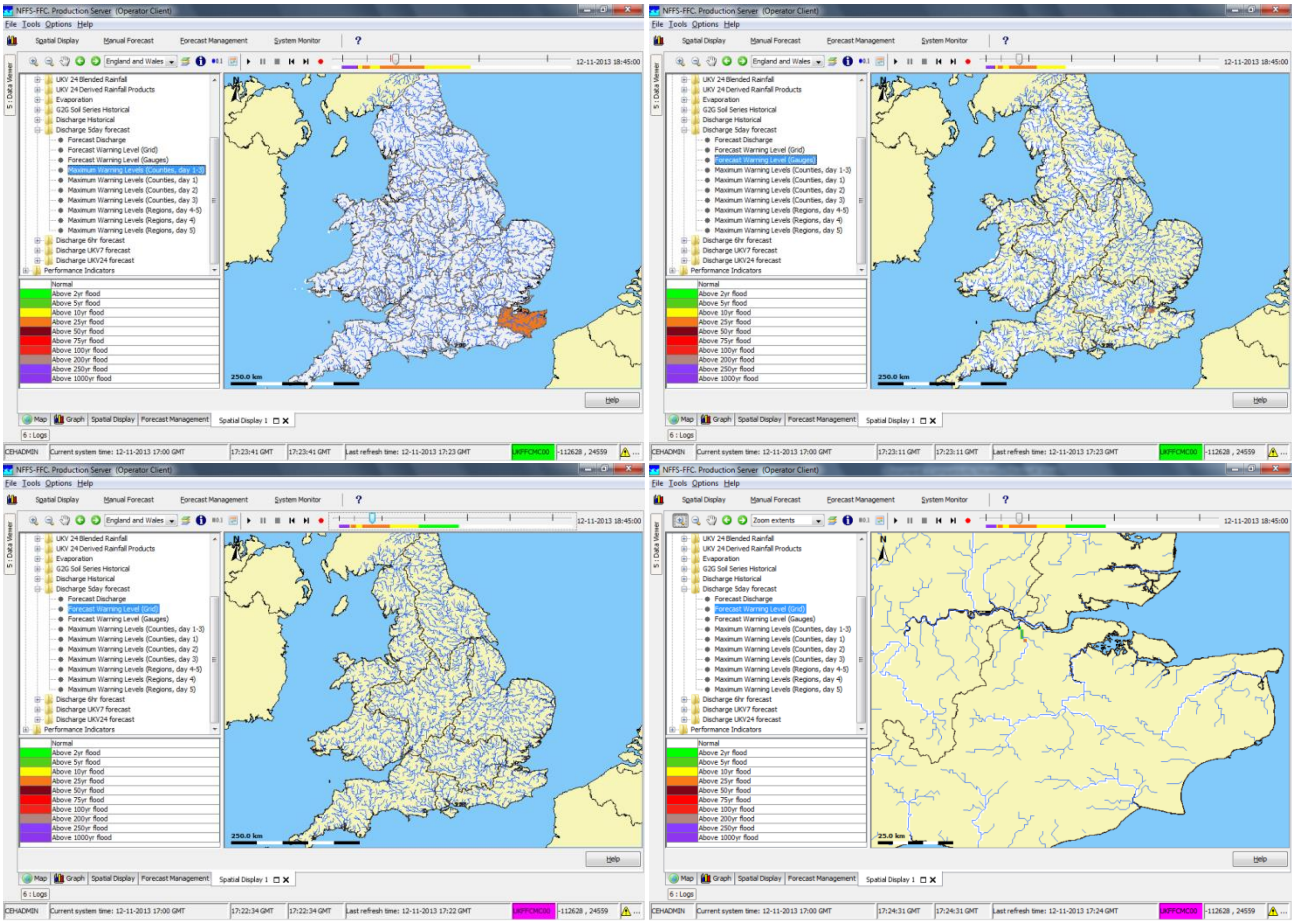


Figure 6.1 NFFS-FFC spatial displays of the deterministic 5-day G2G forecast. All displays are snapshots at a 15 minute time-step. Top left: Maximum Q(T) value over a county. Top right: Q(T) severity at gauged locations. Bottom left: grid of Q(T) severity. Bottom right: zoomed in grid of Q(T) severity.

In addition to the spatial displays, various displays at gauged locations are provided. These include hydrographs of the deterministic 5-day and 6-hour forecasts and time-lagged ensembles of the last six 5-day forecasts. For the ensemble outputs, forecast quantile plumes (rather than spaghetti plots) are provided showing the catchment rainfall and G2G river flows together: an example is given in the right hand side of Figure 6.2.

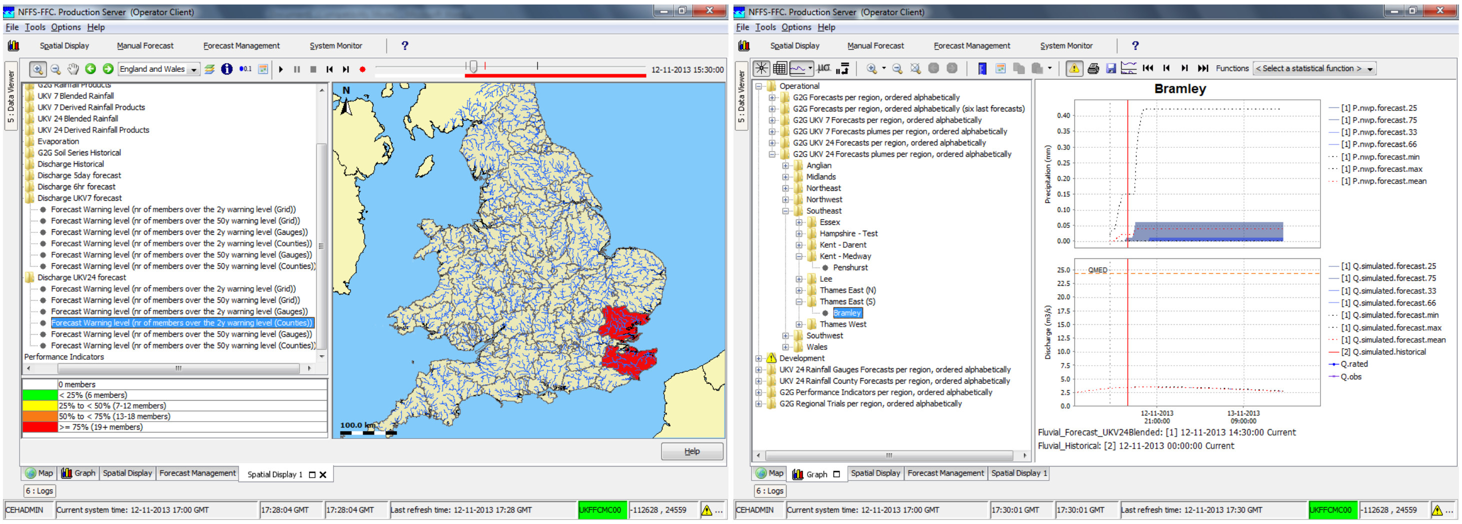


Figure 6.2 NFFS-FFC spatial displays of ensemble G2G forecast using the T+24 blended ensemble rainfall forecast. Left: Maximum number of ensembles exceeding QMED per county. Right: plume plots of rainfall and flow ensembles.

6.2 Scottish Flood Forecasting Service

The Scottish Flood Forecasting Service receives slightly different types of forecast rainfall to the Flood Forecasting Centre. The primary sources for making G2G forecasts are given below.

- Deterministic 5-day G2G forecast, using a combination of NWP products, run once a day.
- Ensemble G2G forecast, using a 24 member MOGREPS-R-R ensemble out to T+54h, run once a day routinely; since the ensemble rainfall forecast is received 4 times a day G2G can be run more frequently on demand.

It is worth noting that whilst FFC stopped using MOGREPS-R (~18km) sometime ago, SFFS found it useful and continued to use it until it was decommissioned earlier this year. Now SFFS use a replacement MOGREPS-R-R feed which is based on MOGREPS-UK 2.2km out to T+36h and MOGREPS-Global 33km from T+36 to T+54h. The product provides 3 hour precipitation totals and is mapped onto an 18km grid size to match the original MOGREPS-R feed. It is also worth noting that the product is purely NWP and the blended ensemble algorithm is **not** used to include radar–rainfall extrapolation. The initial impression from SFFS is that the new MOGREPS-R-R feed is performing better than MOGREPS-R, particularly for the first 36 hours when MOGREPS-UK is used.

Interestingly for SFFS, the interrogation of G2G outputs is made slightly simpler as it is possible to view gridded outputs sensibly across the entire mainland of Scotland at once as shown in Figure 6.3. In generating the Flood Guidance Statement, a primary source of information is the threshold exceedance display that shows the number of ensemble members exceeding a given Q(T) threshold: an example is given in Figure 6.3. This display is available for 15 minute time-steps and can be viewed on screen to identify flood risk "hotspots". In addition, gridded values of G2G forecast flow and associated Q(T) severity are available for the deterministic 5-day forecast.

The gridded spatial displays are supplemented by various hydrograph displays at gauged locations including deterministic time-lagged 5-day forecasts and ensemble plume and spaghetti plots using the MOGREPS-R-R rainfalls. The ensemble hydrograph outputs (but not gridded outputs) are also made available through web reports facilitating remote access to the forecasts. An example from 4 January 2012 is given in Figure 6.4 showing the gridded ensemble flow exceedance map highlighting particular areas and the corresponding ensemble plume hydrographs for two gauged locations in the highlighted area.

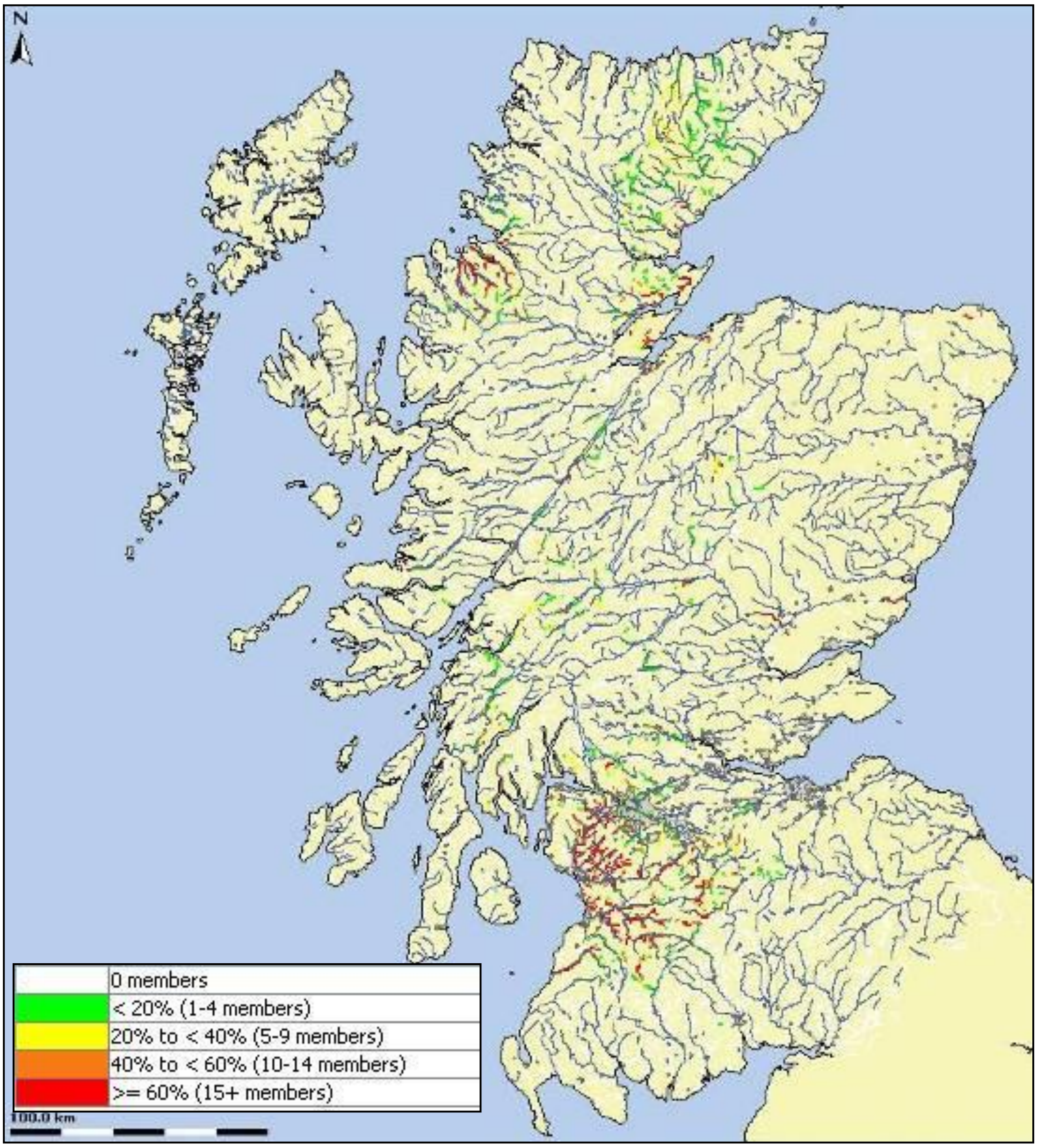


Figure 6.3 FEWS-Scotland spatial display of an ensemble G2G flow forecast using a MOGREPS ensemble rainfall forecast (source: Michael Cranston, SEPA).

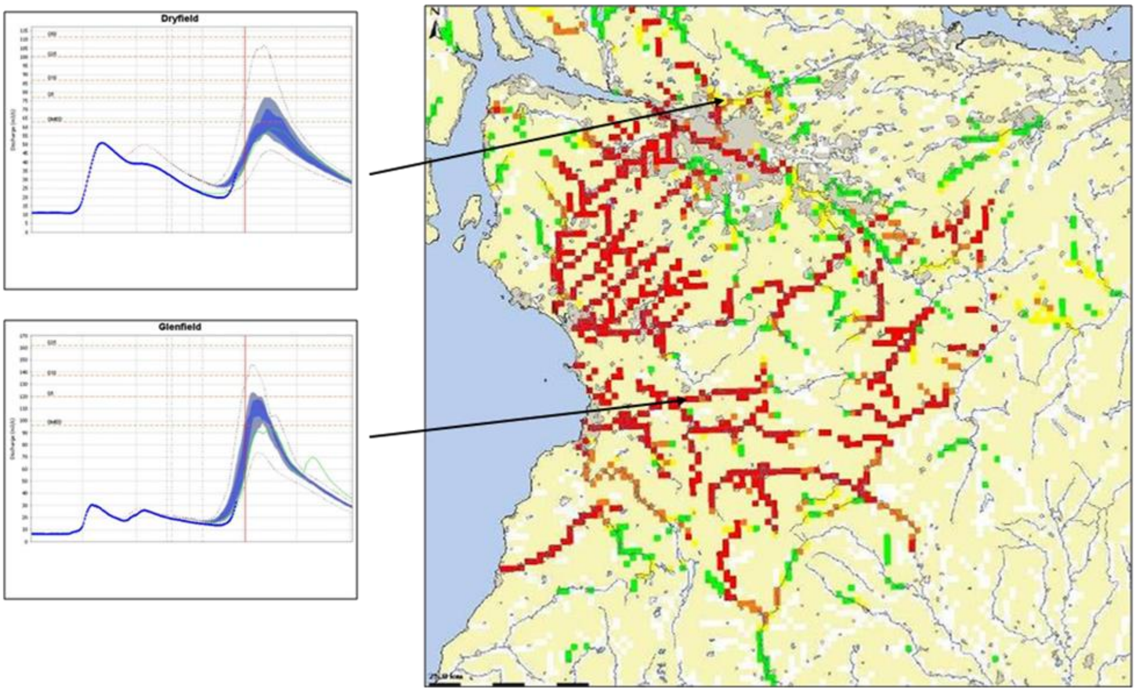


Figure 6.4 FEWS-Scotland spatial display of an ensemble G2G flow forecast using a MOGREPS ensemble rainfall forecast and plume forecast hydrographs at selected gauging stations (source: Michael Cranston, SEPA).

6.3 Potential options for improved display of G2G ensemble outputs

Through performing the case-study and long-term analysis of the G2G ensemble river flow forecasts, and having discussions with FFC and SFFS, it has become apparent that extracting useful information quickly from the ensemble outputs is essential to make the best operational use of the ensemble outputs for Rapid Response Catchments. This becomes even more important when more frequent ensemble outputs become available (e.g. every hour) and Flood Guidance Statements need to be updated quickly during a developing event.

This section provides some ideas for improving the presentation of G2G ensemble information available to end-users. It is proposed for Phase 3 that one of the main focuses should be on CEH working with FFC/SFFS and Deltares to develop these ideas further and implement them operationally. Summarising the ensemble forecasts over time, space and ensemble members in some way is an attractive approach as it condenses a lot of space-time information into concise map form. This high-level information can then act as the starting point for any further detailed interrogation of the output that may be required.

During the analysis of the case studies, it became clear that a summary map over the forecast horizon and a range of Q(T) thresholds is very useful at identifying (i) if an event is likely, and (ii) hotspot areas that are likely to be affected more severely than others. An example is given in Figure 6.5. Here, it is important to note that the averaging over time is done on an ensemble by ensemble basis: that is “how many ensemble members cross the threshold at any point during the forecast horizon”. This is different to calculating the number of exceedances every 15 minutes and then selecting the maximum number over time as this can underestimate the risk at a site

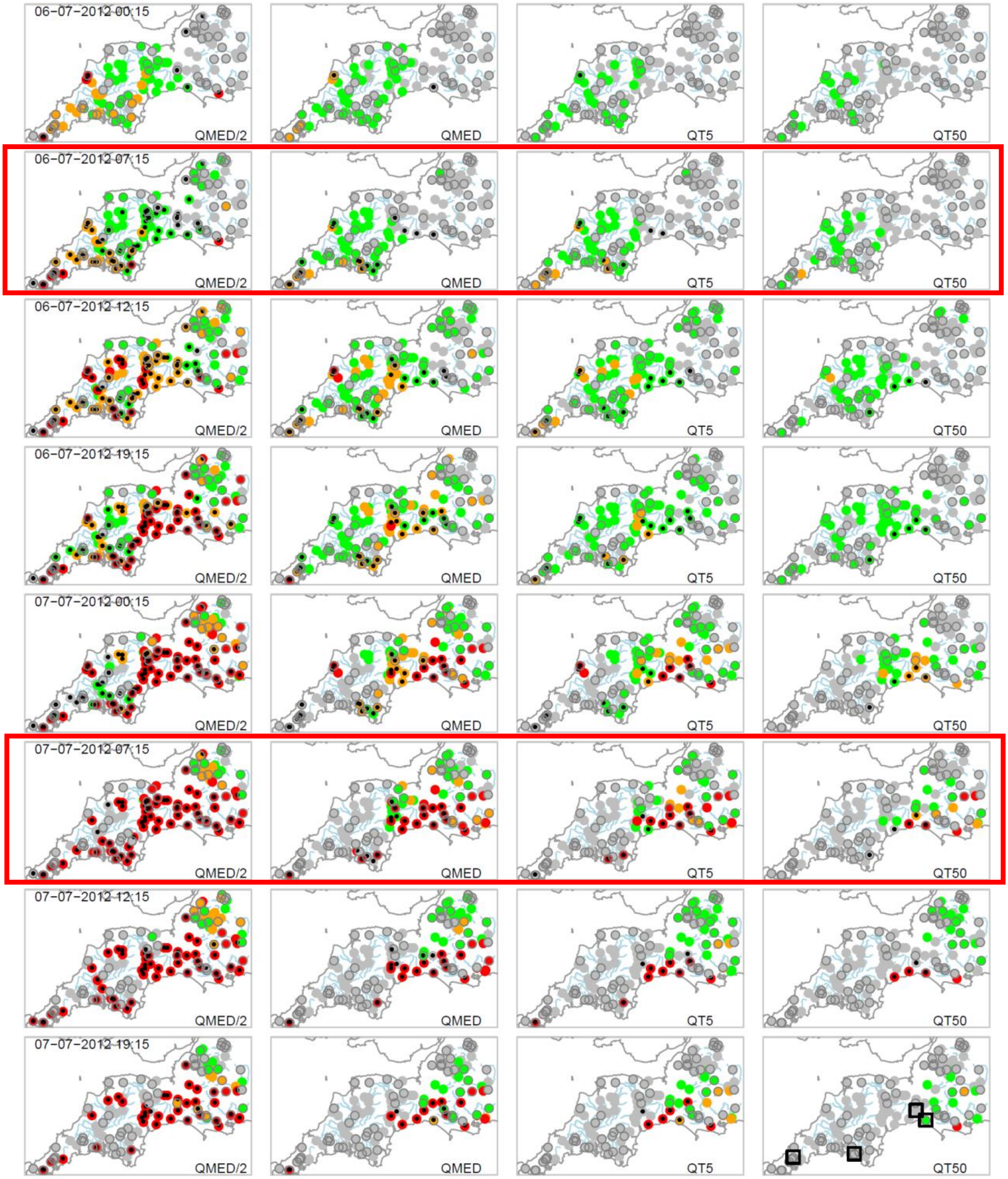


Figure 6.5 Examples of the regional summary map over the South West during 6-7 July 2012 obtained from ensemble G2G flow forecasts using T+24 blended ensemble rainfall forecasts. See Figure 4.5 for a full definition of the map.

since different members can peak at different times. For any given forecast origin (row) the regional summary over a range of thresholds gives more insight than any individual map. Secondly, viewing a sequence of maps is useful (even if they cover different time periods) as it helps to understand the strength of the forecast signal of flooding. This is similar to viewing a time-lagged ensemble. Two maps a day apart have been highlighted by red boxes to emphasise the extra information gained from the three intervening forecasts. This is especially relevant to SFFS as currently ensemble and long-range deterministic forecasts are only routinely made once a day.

Although the black dots on the regional summary map indicate whether the observed flow has crossed the threshold during the forecast horizon, it could be useful to show a similar dot if the observed flow has crossed this threshold in the previous X hours. This would then help to identify if a catchment has already crossed the threshold level. Such a display can be realised in real-time in support of operational guidance, in contrast to the black dot which is useful for *post hoc* ensemble forecast verification.

Currently, the FFC summarise individual pixel outputs in space for 5-day deterministic forecasts at county (day 1-3) and regional (day 4-5) levels by using the “maximum” warning levels. This produces a sequence of maps at every 15 minute time-step. Furthermore, these are summarised over time by producing daily maximums for each day of the forecast. Ensemble G2G outputs of the number of threshold exceedances are only summarised over space at a county level and produce a sequence of maps showing the maximum exceedance per county at every 15 minutes. No summary over time is currently done for the ensemble outputs.

There is a significant limitation to the current NFFS-FFC method of summarising over space. Taking the maximum value over space is not a robust statistical measure and can be sensitive to outliers. For example, if there is an individual Q(T) value that is too low for some reason, this could falsely flag up a flood event when taking the maximum. Using an alternative statistic that is more robust, such as the 90th percentile, would potentially make the summary map more useful.

Some initial ideas building on the regional summary map are provided in Figure 6.6. All these displays summarise the flow forecast over time, and report the number of ensemble members that exceed the QMED/2 threshold at some point during the 24hr forecast horizon. Only river grid squares and those with QMED/2 greater than 1 m^3s^{-1} are considered. Firstly, it appears possible to view the gridded G2G ensemble outputs when little or no background layers are present (top left) and is a good way of identifying hotspots of flood risk, including areas that are ungauged. Reference to gauged sites could be incorporated via overlaying circles at their location coloured by their forecast probability. Alternatively, the catchment area could be colour-coded by the forecast probability at the gauged outlet: this gives some indication of the size of the catchment it relates to (bottom left). Keeping the gridded values in the background allows information across the ungauged areas to be viewed too. For example, areas in north west Wales that were potentially at risk would not have been identified if attention was restricted to gauged locations only. Finally, a very simple regional summary histogram of the number of pixels contained in each exceedance category is provided as a way of comparing across regions at a glance. This form of summary is more robust in the sense that it is less sensitive to a few rogue cells. A more sophisticated or RRC-focused summary could be designed.

Part of the proposed work in Phase 3 is to focus on developing improved operational tools and products based on the G2G ensembles, building on the initial options presented above. Following the work in Phase 1 and 2, and discussions with FFC and SFFS hydrometeorologists, the main points identified to date are summarised below.

- A high-level summary map that condenses information over space, time and ensembles is required. This could be national and/or regional. This is even more critical when using frequently updated nowcast rainfall ensembles.
- Viewing multiple maps together can be informative.
- For the longer blended ensemble forecast, an assessment of the likely time window in which thresholds would be crossed could be useful, e.g. 0-6hr, 6-12hr, 12-24hr for a T+24hr forecast, in additional to the overall risk of a flood occurring.
- Making better use of the gridded information provided by G2G, relevant for example to ungauged catchment areas, with a shift away from only focusing on gauged locations. The spatial gridded picture of flood risk is a key strength of G2G.
- Rapid Response Catchments can be highlighted by considering an appropriate subset of pixels and/or gauging stations.
- Assessing a range of Q(T) thresholds simultaneously appears useful to identify potential hotspot areas that might be most severely affected by flooding.
- Use a more robust statistic than the maximum for summarising flood risk over space and/or time.
- Care is needed when summarising ensemble outputs over time, e.g. “number of ensembles that cross a flow threshold at any point during a time window” is better than “calculate the number of flow exceedances every 15 minutes and then take the maximum of that over a time window”.
- Using the T+7 and T+24 ensembles effectively together. Can the nowcast blended ensembles improve confidence as a flood event approaches?
- On the flow threshold exceedance map, consider including a symbol showing if the observed flow has crossed the threshold in the last X hours. This provides some situational awareness and can also help interpret outputs.

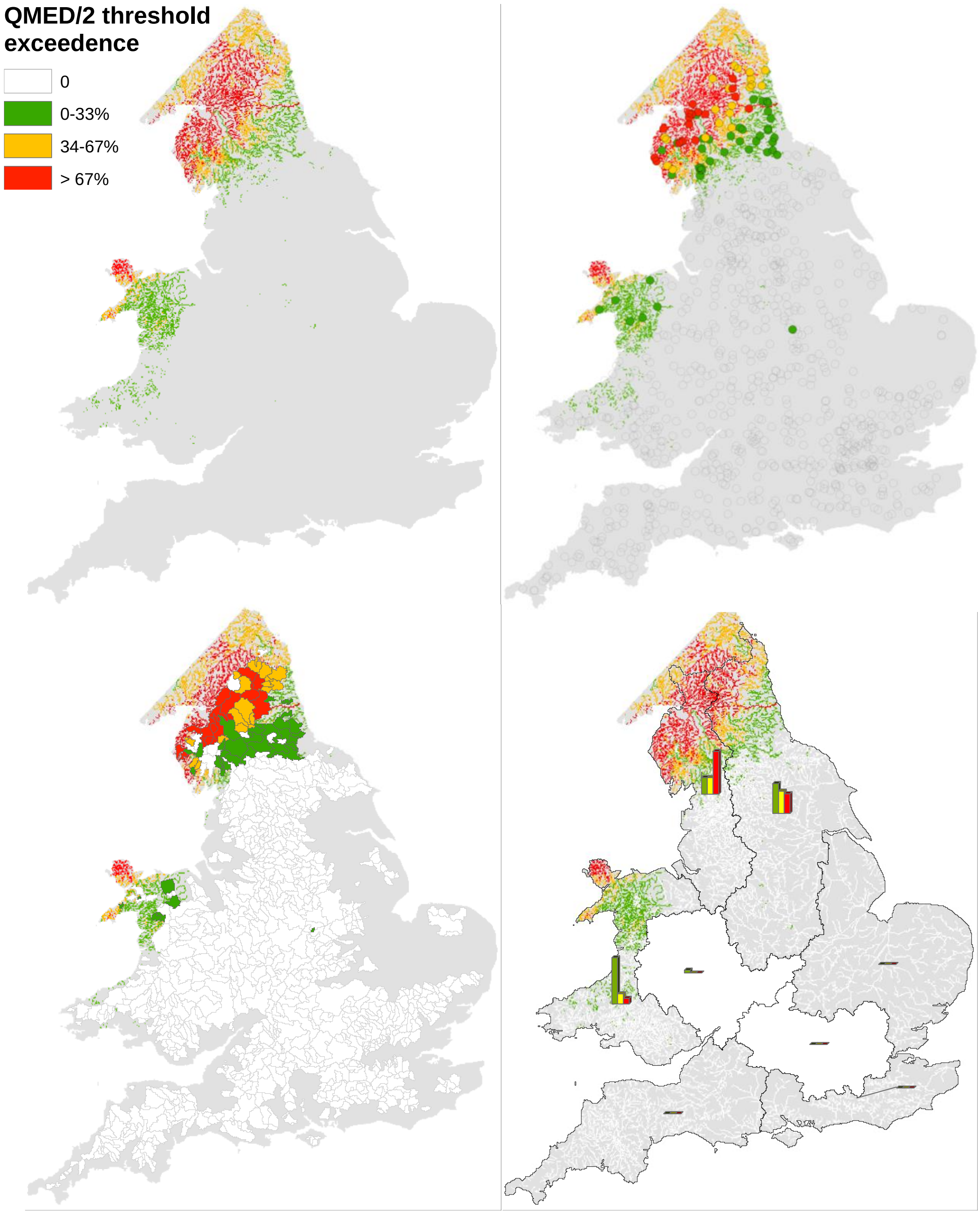


Figure 6.6 Potential display options for summarising G2G flow ensembles over a 24 hour forecast horizon using the QMED/2 threshold. Top left: grid of exceedance. Top right: grid of exceedance plus circles for exeedance at gauged locations. Bottom left: grid of exceedance plus gauged catchment areas coloured by exceedance at gauged location. Bottom left: grid plus regional histograms.

7 Summary and recommendations

7.1 Summary of Achievements

A summary of the key achievements of the “G2G for Rapid Response Catchments” project are set down below, grouped under four topic headings.

7.1.1 Data used in the study

- Take-on of Rapid Response Catchment (RRC) information, helping develop a strategy for assessing G2G for RRCs.
- Update to G2G data holdings for England and Wales to 31 December 2012 and for Scotland to 30 September 2012 in support of G2G assessment.
- Creation of data holdings for high-resolution NWP deterministic (UKV) rainfall forecasts (4 January 2010 to 31 December 2011) and Blended Ensemble rainfall forecasts, including real-time update of the latter via a live datafeed to CEH Wallingford (since 10 May 2012).
- A detailed consideration of optimal rainfall forecast domains for project use indentified inefficiencies in the operational domains used for the Environment Agency/FFC data supply. A revised domain for the new Blended Ensemble rainfall forecasts is recommended that will reduce data volumes by 50%. Further savings can be made if coverage for the Isles of Scilly and/or the Channel Islands is not required and by only keeping values over the land. User consultation is needed here.

Further details of the data used in the study are given in Section 3 of this report.

7.1.2 Assessment methodology

- Development and implementation of a strategy for assessing G2G for RRCs in three modes of forecasting:

  (i) simulation-mode (with state-updating and flow-insertion using gauging stations upstream) and foreknowledge of rainfall observations to help identify any shortcomings in G2G model formulation;

  (ii) forecast-mode using river flows up to time-now, but with foreknowledge of rainfall observations for future times, allowing an assessment not confounded by rainfall forecast errors;

  (iii) forecast-mode fully emulating operational conditions using deterministic rainfall forecasts.

  (iv) forecast-mode fully emulating operational conditions using ensemble rainfall forecasts.

- Selection of suitable sites for assessment accounting for the needs of G2G for RRCs. Grouping of sites for assessment by catchment area, and identifying which are "headwater" catchments (with no gauges upstream). Also including consideration of urban/suburban coverage and steepness of terrain slope as characteristics inducing rapid runoff response.

- Identification of performance measures to be used in assessment, including R^2 Efficiency and POD and FAR skill. Assessments are made in simulation-mode, and in forecast-mode as a function of lead-time. Box and whisker plots – displaying the minimum/maximum, lower/upper quantile and median values - are used to summarise the variation of a measure across groups of sites. Skill scores are calculated in relation to exceedance threshold crossings of Q(T), the flow Q of return period T years. The median of the annual maximum flood, Q(2), is indicative of a level a little above bankfull discharge for natural channels and used as a reference frame, and values 50 and 75% of this. (Noting that records are not sufficiently long to usefully consider longer return period flow thresholds when calculating skill scores.)

- For ensemble forecasts, long-term continuous assessment and case-study assessments have been developed.

  The continuous assessments primarily use Relative Operational Characteristic (ROC) based scores which identify how well the ensemble discriminates between events and non-events, and the Brier Skill Score which assesses how much improvement the ensemble forecasts give beyond a "reference" forecast. Again a low threshold of 50% of Q(2) was used due to the short period of record. The assessments have been made using (i) all ensemble members, and (ii) one ensemble member. This is done in order to make some assessment of the benefit an ensemble forecast may have over using a single deterministic forecast.

  Case-study assessments have focused on different methods of viewing and assessing the G2G ensemble outputs and have relevance to how best to present G2G ensemble outputs to support operational decision-making.

Further details of the G2G model assessment methodology is given in Section 4 of this report.

7.1.3 Assessment results

Simulation-mode assessment using observed rainfall (Section 5.2)

- Assessment of the G2G in simulation-mode over 4 water years using raingauge-only rainfall, broken down by catchment area and nature of G2G run (pure simulation, state-updating, flow-insertion plus state-updating); displayed as box and whisker plots and Performance Maps of R^2 Efficiency.

- *England & Wales.* Small catchments relevant to the RRC objective have the best and good performance over South West, Wales, North East and North West. Median performance is robust across different water years while some variation in the lower quartile performance can be attributed to raingauge errors. There is a weak signature of poorer model performance for small urban catchments suggesting some potential for G2G model improvement in these locations.

- *Scotland.* The performance for small catchments (less than 50 km^2) relevant to RRC is very good for Scotland in all regions assessed and the pooled performance of sites with catchment areas less than 250 km^2 is noticeably better in Scotland than England & Wales. This is partly to be expected since the hydrological behaviour of the rivers in Scotland is more similar to that in the North East, North West, Wales and South West regions of England and Wales where G2G also performs well.

Forecast-mode assessment: foreknowledge of observed rainfall (Section 5.3)

- Assessment of the G2G in forecast-mode using raingauge-only rainfall over the 4 water years. Analysis used: pooled forecast lead-time plots grouped by catchment area, box and whisker plots and scatter plots comparing simulation- and forecast-mode performance and spatial maps of forecast-mode performance.

- *England & Wales.* Demonstrates the value of data-assimilation even out to a lead-time of 12-24hr. Larger catchments benefit in performance most and for longest, on account of flow-insertion for gauges upstream. Headwater sites benefit significantly from ARMA error-prediction, when compared to simulation-mode. Regions with the poorest simulation-mode performance benefit most from ARMA error-prediction. A majority of small catchments in the west and north perform well. There are some issues in Southern, Thames and Anglian regions. Some of these are due to groundwater-dominated catchments not relevant to RRCs.

- *Scotland*. In agreement with England & Wales, all sites benefit greatly from ARMA error-prediction, particularly the Northern areas where this can ameliorate degradation in simulation-mode performance due to sparse raingauge networks. The median model performance for small catchments is noticeably better in Scotland (reflecting the simulation-mode performance) and maps show forecast performance is generally very good over the lower half and north eastern parts of Scotland.

Forecast-mode assessment: use of UKV deterministic rainfall forecasts (Section 5.4)

- Assessment of G2G in forecast-mode using NWP UKV rainfall forecasts for emulated future times (and using raingauge-rainfall up to this forecast time-origin) for the water year 2010/11; and using foreknowledge of raingauge-only rainfall as a relative performance reference frame. Monthly and annual accumulation maps of UKV and HyradK rainfall estimates are used to help diagnose biases.

- *England & Wales.* Forecasts using UKV rainfall deteriorate markedly beyond 12 hours ahead and especially for smaller catchments relevant to the RRC objective. However, for some locations G2G forecasts continue to perform very well even at long lead-times. A case study on 21 October 2010 in part of North West Region highlights overestimation of rainfall by UKV as a cause. A broader look at regional trends in G2G relative performance over England & Wales reveals geographical biases induced through use of UKV rainfall, notably for the northern and southern ends of the Pennines and in the vicinity of London. Monthly and annual rainfall accumulation maps expose (i) overestimation by UKV in lowland areas (such as over London) for several months and (ii) evidence of UKV overestimating rainfall leeward of high elevation areas (such as in the north and south Pennine areas).

- *Scotland.* For small catchments less than 50 km^2 the median performance of UKV-based forecasts drops off with lead-time but performs better than over England & Wales and with less spread in results. Performance over the first 12 hours is generally still good. Spatial analysis reveals a band of river gauging stations from Edinburgh northwards where G2G forecasts using UKV rainfall appear to be performing less well. In particular, UKV is over-estimating rainfall near the coast at Edinburgh and Elgin in the north. Along the Cairngorms the UKV orographic enhancement scheme is providing enhanced rainfall that appears realistic but the sparseness of the raingauge network in these areas prevents more detailed assessment.

Forecast-mode assessment: use of Blended Ensemble rainfall forecasts (Section 5.5)

Case-study analyses (Section 5.5.2)

- Analysis of the G2G ensemble flow forecast obtained using as input the Blended Ensemble rainfall forecast in both the longer T+24hr (updated 4 times a day) and the shorter T+7hr (updated every 15 minutes) nowcast forms. A catalogue of notable events was made and five case study events selected for detailed analysis.
- During the 2012 study period the events were dominated by frontal storms but there were still some convective events, the most significant being that for 6-7 July 2012 which was selected as a case study.
- It should be noted that the methodology for the Blended Ensembles only uses a single deterministic high-resolution NWP forecast to generate the other members. Therefore the performance of the deterministic NWP forecast has an impact on all ensemble members. This can be particularly significant in convective situations where ensemble NWP rainfall forecasts are really required to capture the uncertainty.
- Case studies showed G2G ensemble flow forecasts using the longer T+24hr blended ensemble rainfall forecasts generally performed well for all cases, often giving a reasonable signal of an event 12-18 hours in advance of the peak and also in advance of any observed response for the small catchments.
- The 27-28 June 2012 event was dominated by thunderstorms and presented most difficulties for the longer T+24hr blended ensemble rainfall forecasts, although they still provided a reasonable signal of the flood event and its location. In this case the hourly sequence of G2G ensemble exceedance maps using the T+7hr blended ensemble rainfall nowcast added extra value to the maps using the 6-hourly T+24hr blended ensemble rainfall forecasts and correctly closed in on the main flood risk hotspots. A lead-time of one to two hours before the observed flows began to rise was possible even for small catchments.
- The Comrie case-study in Scotland focussed on a rapid response catchment that currently has limited forecast capability. The G2G flow forecasts using the T+24hr blended ensemble rainfalls performed well and gave a strong indication of an event around 14-20 hours before the peak. Having 4 forecasts a day would have been necessary to capture and have confidence in the forecast signal.
- The Comrie case-study also identified an issue where the orographically enhanced rain was not correctly identified by the STEPS scheme so was inadvertently advected into lowland areas of the catchment giving too high rainfall. This caused the T+7hr forecasts and the T+24hr forecasts nearer to the event to be less useful and tended to overpredict the events in this situation.
- The final case study - concerning the Morpeth floods in September 2008 - used a true UKV ensemble and is a forerunner to the MOGREPS-UK product planned to become operational for FFC in 2013. This case study used only one forecast origin so, although it added value beyond a deterministic forecast,

more extensive assessments using the MOGREPS-UK 2.2km product are needed to draw wider conclusions.

- For all case studies, viewing regional maps of successive forecasts at a range of Q(T) thresholds was useful to highlight if an event was likely, hotspot areas that are likely to be more severely affected than others, and often captured the timing of which areas would be affected first.

Longer-term assessment of ensembles (Section 5.5.2)

- The G2G for RRCs project delivered the first continuous assessment of G2G ensemble flow forecasts. Although limited by the relatively short durations - ~7.5 months (May to December 2012) over England & Wales and ~4.5 months (May to September 2012) over Scotland – the assessment proved very useful and added extra information beyond the traditional case study assessment.
- Due to the short period of record, a low assessment threshold of QMED/2 has been used for site-specific analysis. By defining an event to require pooling over all sites in a region has allowed some assessment at QMED and Q(5) thresholds. It should be noted that the Brier Score, and consequently also the Brier Skill Score, will be sensitive to the short assessment periods available and is the likely cause of the noticeably poorer performance of the Brier Skill Score.
- Across England, Wales and Scotland, the ROC Score and Brier Skill Score generally increase when using the full 12 member ensemble compared to using one member as representative of a deterministic forecast. This provides strong evidence that using the full ensemble and probabilistic forecasts should allow better guidance to be given on forecast flood risk.
- Some regional trends have been identified, particularly areas in England and Wales, where performance appears to be relatively poorer when compared to other areas.
- In terms of the quantitative score values, Scotland appears to perform less well than England & Wales but it is believed this is mainly due to the much shorter period used for assessment, particularly for the Brier Skill Score, and its dominance by the summer season.

7.1.4 Immediate operational benefits

- Improvements to operational configurations - of both the NFFS regional systems and the NFFS-FFC system running G2G - arising from attention-to-detail hydrometric data take-on. This identified **269** of the 708 sites configured for use by G2G were not being used because either no rating had been configured into NFFS-FFC or they were deemed unreliable as being derived from telemetered flows (e.g. ultrasonic gauged sites). Ratings were subsequently configured in and flow sites were subject to review. Also data take-on identified almost **400** sites that were not being considered in the operational G2G configuration, because flows for these sites were not included in WISKI (only river levels). A total of 912 sites is now included in the operational G2G configuration giving accuracy benefits to FFC since March 2013.
- The G2G configuration for Scotland has been updated to remove 6 sites (at the request of SEPA).
- The G2G Performance Summary, used operationally to summarise the G2G model performance at gauged sites in simulation-mode, has been extended to

encompass forecast-mode assessment, and using the POD and FAR performance measures as indicated above along with revised hydrograph displays.

- For FFC, the revised G2G configuration and updated Performance Summary pages were delivered at the start of 2013 and implemented operationally in March 2013. This will give significant improvements in G2G performance relevant to RRCs. As part of this, the assessment period of the Performance Summary was extended to a longer period of 4 water years (spanning 2007 to 2011).
- For SFFS, the revised G2G configuration and updated Performance Summary pages are scheduled for delivery in the first quarter of 2014.

7.1.5 Key benefits

Some key benefits arising from the project are identified below.

- Evidence has been produced that shows G2G has good skill in providing strategic forecasts for RRCs. The evidence is stratified by catchment type (area, urbanisation, headwater), form of forecast (simulation or forecast mode) and nature of rainfall input (raingauge, deterministic forecast, ensemble forecast).
- Strong evidence has been presented on the advantage of using an ensemble rainfall forecast as input to G2G to obtain a probabilistic flood forecast for an RRC, relative to an approach where only a single deterministic rainfall and flood forecast is obtained. This indicates better guidance can be given on forecast flood risk for RRCs, improving the level of service provision for such catchments which are currently not well served.
- An improved G2G model configuration, exploiting gauged flows from 912 sites and including new locally calibrated parameters, has been delivered and made operational for the FFC with England & Wales coverage. The benefit is improved operational flood forecast accuracy. For Scotland, an enhanced configuration will be delivered to SFFS in Spring 2014.
- Detailed recommendations on how the visual presentation of G2G ensemble results could be improved are set down in this report. When further developed and implemented, these will prove of benefit to the preparation of Flood Guidance Statements issued by FFC and the SFFS across Britain.

7.2 Strategic Conclusions

The aim here is to set down a set of specific conclusions in a form that provides a more comprehensive strategic picture of how G2G performs, considering a range of aspects for forecasting in rapid response catchments. This is done by posing a set of strategic questions and using evidence from the report, and some further analysis of the assessment results, to address them.

- **What general skill does G2G have in forecasting in RRCs?**

G2G has the capability to shape a given storm pattern in space and time into a spatial flood response, taking account of the river network topology and slope, properties of the land-cover, soil and geology, and prevailing moisture conditions of the land. Given good gridded time-series of rainfall at 1 km scale, G2G can transform these in a hydrologically sensible way into gridded time-series of river flow. Evidence in this report shows the river flow simulation accuracy can be good, even for RRCs with areas below 50 km^2.

An important attribute of G2G is its area-wide formulation capable of forecasting “everywhere” across an RRC, river basin or country. This contrasts with regional flood forecasting systems normally configured as networks of models making forecasts for specific locations, often where river flows are gauged and rarely aligned to RRCs. G2G is therefore well suited to the RRC problem as such catchments are typically small and ungauged. G2G aims to complement regional forecasting systems, which typically will be expected to give higher accuracy for gauged sites whose flow records have been used for model calibration.

A further attribute of G2G stems from the gridded time-series form of river flow forecasts it produces. Once visualised as real-time animated maps, these forecasts provide a strong spatially-continuous image of the evolution of a flood over time: flood peaks moving from within RRCs to increasingly larger river networks downstream. Such information should prove invaluable to flood preparedness activities aimed at damage mitigation.

- **How does skill vary with scale and different uses?**

River flow forecasts from G2G for RRCs appear to be sufficiently good to provide a “strategic heads up” of the possibility of future flooding in a region. Notably, the probabilistic displays developed here provide the capability to identify likely “hotspots” of future flood risk. Such displays (or similar) should prove a useful aid in regional planning/preparedness activities and allow FFC/EA/SFFS/SEPA to provide a better level of service in terms of more targeted and informed outputs (e.g. Flood Guidance Statements). Site specific forecasts, particularly in convective storm situations, are unlikely to have strong skill for small RRCs when presented in deterministic form. However, the new probabilistic flood forecasts for RRCs derived from the latest forms of ensemble rainfall forecast have been shown here to be of value - both locally and regionally - though capturing, especially, the spatial uncertainty in storm location inherent in the forecast. This represents an important step forward for an inherently difficult forecasting and warning problem where uncertainty may severely compromise the value of deterministic flood forecasts for RRCs.

- **For what types of catchments and sizes does G2G have some skill?**

(i) Smaller (0.5 to 20 km^2) and larger RRCs (20 to 50 km^2)

The results reported show performance for small catchments less than 50 km^2 is good, notably for upland areas such as those prevalent in Scotland, Wales, the South West and North West. Performance is robust across the years assessed, although raingauge errors can impact on accuracy if not detected. If a larger RRC has a gauged catchment within it, data assimilation (flow insertion and error prediction) using the upstream flow can improve accuracy for the RRC downstream. The benefit can persist in forecasts for longer lead-times depending on catchment response times. However, larger catchments (rather than RRCs) benefit most from data assimilation.

Within Phase 1 of the project a minimum catchment area of 50 km^2 was agreed and used for pooled performance analysis. During the final review of Phase 2, interest was raised from the Project Board on discriminating performance for the smaller RRCs, with areas below 20 km^2. Although a full re-analysis at this scale was not possible, an exploratory analysis has been done using the simulation-mode G2G river flows and is reported here. Figure 7.1 gives the breakdown of gauged catchment areas for England & Wales and Scotland using 20, 50 and 100 km^2 area partitions, revealing only 94 assessment sites in England & Wales with an area less than 20 km^2 and as few as 7 in Scotland. Figures 7.2 and 7.3 repeat the analysis of Section 5.2 but using the 20 km^2 breakpoint. For Scotland, the G2G performance is seen to be robust across the <20 km^2 and 20-50 km^2 categories. For England & Wales the pooled performance shown in Figure 7.2 is relatively poorer for the

<20 km^2 category. To gain further insight into this, a breakdown by region is given in Figure 7.4 for catchment areas <20 km^2 and <50 km^2. These clearly show the G2G performance is robust across catchment sizes less than 50 km^2 for all regions except Anglian, Thames and Midlands where the smaller catchments perform relatively less well. These results are very encouraging and show that the underlying G2G simulation-mode performance is robust for all assessment sites under 50 km^2 for those areas most concerned with flooding from RRCs (North East, North West, Wales, South West and Scotland).

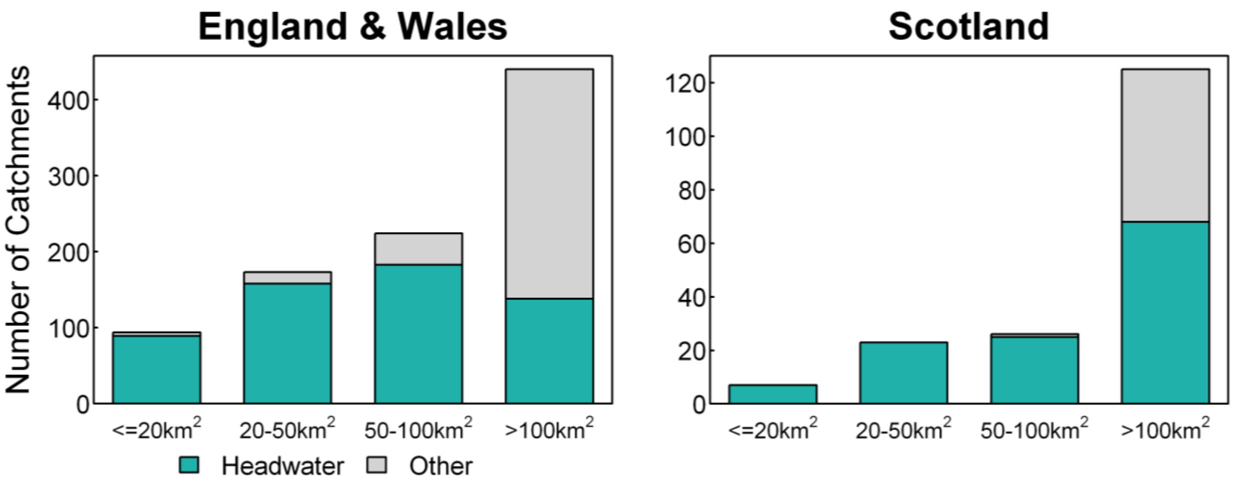


Figure 7.1 Number of assessment catchments grouped by catchment area for England & Wales (left column) and Scotland (right column). Also shown are the number of "headwater" locations (green) which have no gauged site upstream of them.

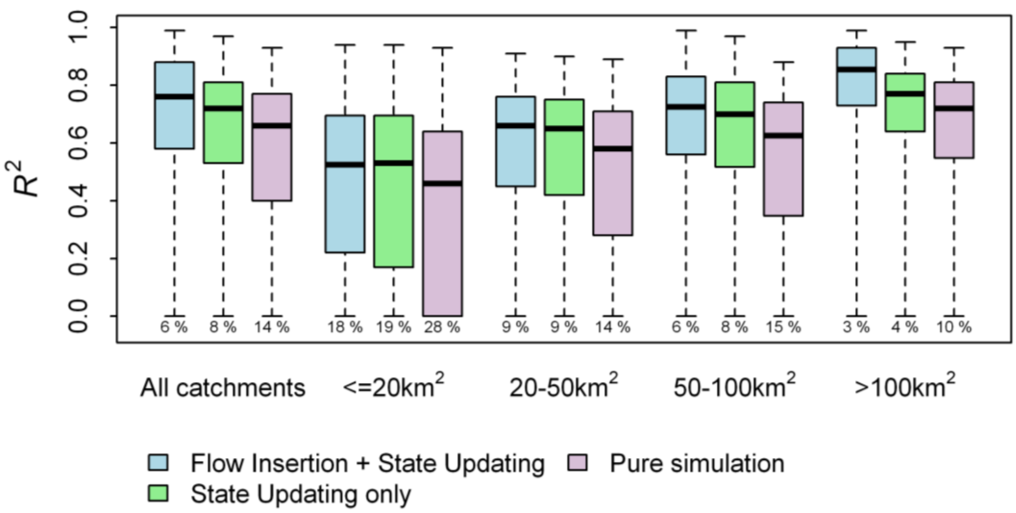


Figure 7.2 England & Wales. Box and whisker plots showing the median, inter-quartile range and the max/min values of R^2 Efficiency for simulation-mode G2G modelled flows for 2007-2011, grouped by all catchments and catchment area. The percentages indicate the fraction of sites with a negative R^2 Efficiency.

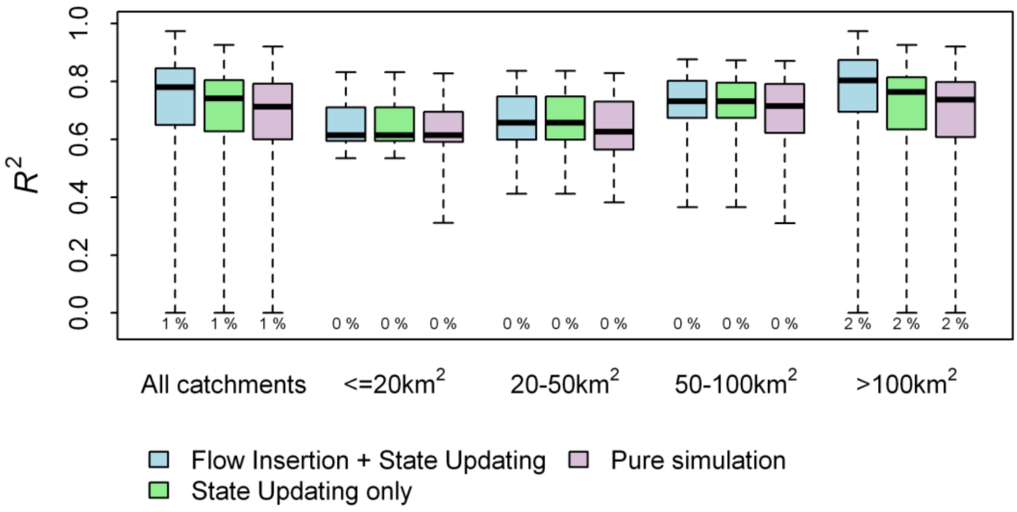


Figure 7.3 Scotland. Box and whisker plots showing the median, inter-quartile range and the max/min values of R^2 Efficiency for simulation-mode G2G modelled flows over Scotland. These are over the four water years spanning 2008-2012 and grouped by all catchments and catchment area. The percentages indicate the fraction of sites with a negative R^2 Efficiency.

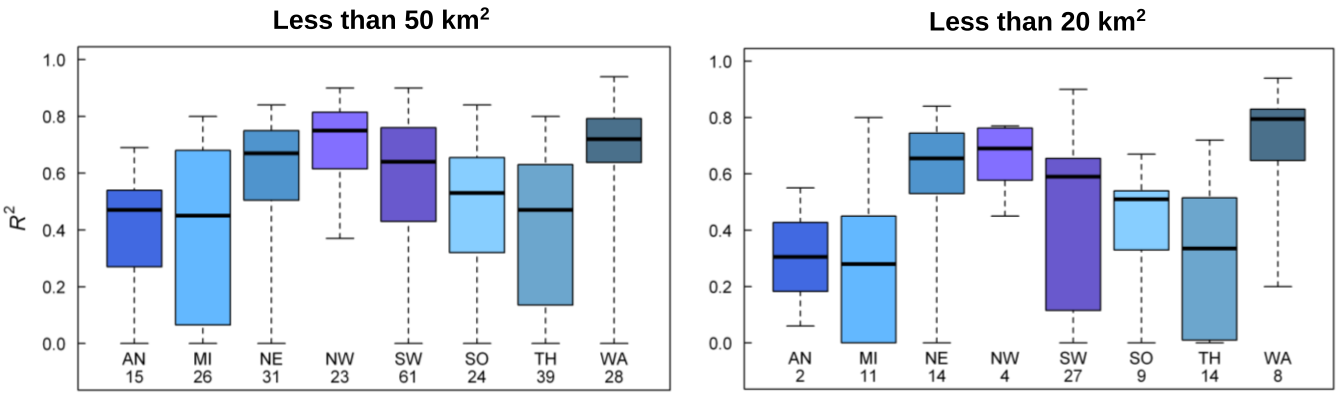


Figure 7.4 England & Wales. Box and whisker plots showing the median, inter-quartile range and the max/min values of R^2 Efficiency for simulation-mode G2G modelled flows (with flow-insertion and state-updating) for 2007-2011. Results are only for headwater catchments less than 50 km^2 (left) and less than 20 km^2 (right). Results are grouped by region and the number of sites per region is also given.

(ii) Small urban catchments.

There is a weak signature of poorer model performance for small urban catchments suggesting some potential for G2G model improvement in such locations.

- **What is the model performance with forecast rainfall?**

Uncertainty in forecast rainfall, not unreasonably, impacts greatly on the performance one can expect from G2G flow forecasts. G2G performance in the first 12 hours of a forecast from the time it is made is generally good for RRCs across Britain. However, strong deterioration in flow forecast accuracy is observed beyond 12 hours using the deterministic UKV rainfall forecasts, especially for RRCs.

The study identified geographical biases in the UKV rainfalls, overestimating leeward of hills (northern and southern ends of the Pennines) and near the coast in parts of Scotland (Edinburgh and Elgin in the north) and underestimating in lowland areas (around London). On the other hand, the UKV orographic enhancement scheme is providing realistic rainfall forecasts in upland areas although verification is problematic.

Case studies using blended ensemble rainfall forecasts illustrate a capability to signal flooding some 12 to 20 hours in advance using G2G. Longer term assessments of ensembles using skill scores provide strong evidence of the benefits of an ensemble probabilistic flood forecast over a deterministic one: allowing better guidance to be given on forecast flood risk.

7.3 Recommendations

Recommendations general

- Project results reported here confirm that G2G has utility for forecasting in Rapid Response Catchments. New and improved methods for viewing, analysing and summarising ensemble outputs are required to realise this potential operationally. The suitability of these methods needs to be assessed using the existing G2G ensemble forecast database. See Phase 3 proposal.
- Although G2G has shown utility, a specific project focussing solely on improving the simulation model performance of G2G should be considered. The scope of the ensemble and NWP analysis in this project has left little time for focussed model development. Specific areas of opportunity particularly relevant to RRCs include improvements to:

    (i) Flow routing, for example:

    - how wave speeds vary spatially in channel reaches of rapid response catchments and their association with channel properties
    - allocation of grid squares to land or river, for which different flow routing speeds are assigned

    (ii) Runoff response, for example:

    - the possibility of introducing the proportion of each HOST soil class within a 1km grid-square to better resolve heterogeneity in runoff response; this is linked to considering any benefits from configuring G2G at a finer resolution
    - improved use of land-cover dataset for rapid response urban catchments

    (iii) Gridded rainfall input, for example:

    - better gridded rainfall inputs for upland areas

    (iv) Data assimilation, for example:

    - advances in state updating and/or error prediction

  Focussed case studies exploring specific opportunities would allow ideas to be investigated locally in detail and then trialled nationwide.

- An agreed register of Rapid Response Catchments and associated river gauges (where available) should be identified. This should include information on the appropriate hydrological catchment boundaries and drainage point associated with the RRCs. This information could be used to tailor the Phase 3 products further. This information is not available in the current (2012-13) RRC Register for England & Wales.
- The ensemble G2G forecast dataset produced requires further analysis to yield deeper insights and understanding, particularly on the benefits of using frequent updates of the radar-rainfall nowcast. Analysis should be extended to give a holistic assessment of ensemble rainfall and G2G flow forecasts. In part, this will be done under the FFC-funded project on "Verification of River Flow Ensembles". Even though the operational ensemble is changing to use MOGREPS-UK, additional analysis will provide useful benchmark information.
- When the G2G hydrometric river and raingauge datasets are next updated for FFC and SFFS it is recommended to extend the G2G ensemble database to encompass ensemble rainfalls for all seasons. The longer period of hydrometric record with concurrent ensemble rainfall forecasts would allow ensemble analysis to be carried out for more extreme thresholds, especially if sites are pooled in some way.
- The new G2G forecast and analysis infrastructure should be exploited to assess G2G flow forecasts using the Met Office archive of MOGREPS-UK forecasts which starts in July 2012. Although this doesn't include the radar-rainfall nowcast element it will still provide a useful long-term assessment. In part this will be achieved under the FFC-funded project on "Verification of River Flow Ensembles".
- Data volumes of ensemble rainfall forecasts can be a limiting factor for their use in flood forecasting due to (i) the volume of data transferred, (ii) the number of files produced (e.g. one file with multiple images or one file per image), and (iii) the computer processing time for data processing, access and hydrological model runs. For ensemble rainfall forecast datafeeds, it is recommended that FFC and SFFS consider the following.
    (i) Reconsider the domain that is used for ensemble forecasts (particularly relevant for FFC/EA is whether to include Channel Islands and/or Isles of Scilly as this significantly increases the domain size).
    (ii) Consider using missing values over the sea to aid data compression, thus reducing data volumes and potentially reduce processing time.
    (iii) Ensure appropriate data compression is used to transfer data when possible (e.g. zipping files sent from Met Office).
    (iv) Ensure data files are organised optimally for the receiving system (e.g. one file per forecast or ensemble is preferred by some systems).
- If computational limits allow, consider extending the nowcast-based G2G flow forecasts by 2 hours to allow flood peaks to travel some way through the modelled river network. Several of the T+7hr G2G forecasts were still rising at T+7hr, even for the smaller catchments. There is less need to extend the longer T+24hr forecast.
- All hydrological forecasts, especially those for Rapid Response Catchments, benefit from having the most recent hydrometric data available at the forecast time-origin. It is recommended that advances in low-cost hydrometric data capture (e.g. GPRS) be exploited to achieve rapid data access in the future.

Recommendations specific to FFC

- Continue to produce G2G ensemble forecasts 4 times a day using the T+24hr Blended Ensemble rainfall forecast (and its replacement).
- Produce updated G2G flow forecasts every 15 minutes using the T+7hr blended ensemble (already included in FFC future planning). This will be achieved once the G2G is running operationally for FFC on the dual NFFS-FFC and HPC (High Performance Computer: the “Met Office supercomputer”) environment. Production of suitable G2G outputs and products will be further considered under Phase 3 (see proposal). Clear benefits to the level of service offered for flood guidance in RRCs will result from this enhanced operational capability.
- The information technology implications of running “G2G for RRCs” for the FFC requires detailed attention. This recommendation is already being largely addressed within the “G2G on HPC” Project for FFC. This is anticipating running G2G with rainfall forecast ensembles with a 15 minute update frequency, aligned to future plans for ensemble generation. Historical runs of the G2G will be carried out as now on the NFFS-FFC platform, a states file passed across to the HPC which will generate G2G ensemble flow forecasts. The NFFS-FFC platform will be used for visualising the forecast outputs and will also be capable of performing *ad hoc* deterministic forecast runs. The details of sizing data volumes and computing resource requirements have been progressed under the project. G2G is now transformed from Windows (used by NFFS-FFC) to AIX (a Unix form of operating system used by the HPC) and operational trials are planned in 2014, including running with ensembles and with frequent updates.

Recommendations specific to SFFS

- The raingauge network along the east coast of Scotland is sparse and can have a detrimental impact on rainfall estimation and G2G modelling. There are few small gauged catchments in this area. For locations with an important need for RRC forecasting then research into improved rainfall estimation algorithms is recommended (e.g. including elevation effects and/or combining radar and raingauge rainfall data).
- SFFS to consider routinely running ensemble forecasts more frequently using the current MOGREPS-R-R 18km rainfalls that are available 4 times a day. For most case-study events confidence in a strong signal of flooding was only possible 6-18 hours ahead so daily ensemble runs may not capture this. Note this product has no radar-rainfall nowcast element.
- SFFS to consider using MOGREPS-R-R at a higher resolution out to T+36hr. This is when the MOGREPS-UK is available so a resolution of 2km would be preferred. This would avoid any loss in information introduced by smoothing to the current 18km resolution used by SFFS.
- The detailed analysis over Comrie highlighted potential issues with the STEPS component of the blended ensemble that had a detrimental effect on the G2G forecasts. Further analysis of the nowcast-based ensembles forecasts is recommended (the “Verification of River Flow Ensembles” project for FFC will help progress this).
    (i) Depending on the outcome of the above analysis, SFFS should move to using the STEPS/MOGREPS-UK blended ensemble for the T+36hr element of their current MOGREPS-R-R feed.

- SFFS should consider, in the first instance, using their current STEPS deterministic nowcast feed (T+6hr) to produce G2G forecasts. These can be time-lagged. If IT infrastructure limits allow, this could be extended to blended STEPS/MOGREPS-UK T+7hr ensembles.

7.4 Proposal for Phase 3

The proposed work under Phase 3 has a specific focus to support FFC and SFFS derive more targeted information in real-time on flooding risk to Rapid Response Catchments from G2G ensemble forecasts. The initial communication route of this information to the responder community is likely to be through the Flood Guidance Statements prepared by FFC and SFFS.

- The improved operational tools and products for Rapid Response Catchments would be delivered through the NFFS-FFC/FEWS-Scotland. Based on discussions with hydrometeorologists at FFC and SFFS, the products and tools need to provide concise, robust information in a quick and clear manner. This is particularly true if using the frequently updated short-term nowcasts.
- Incorporate any emerging national datasets on Rapid Response Catchments to tailor the products further.
- An initial outline and prototype of possible new tools and products is given in Section 6.3 with a summary of points at the end of the Section. These will be further developed by CEH in conjunction with FFC, SFFS and Deltares to form the first list of viable options. This will also be informed by parallel work under the FFC-funded “Verification of River Flow Ensembles” project.
- A workshop with CEH, FFC, SFFS and Deltares is suggested to finalise the tools and products that will be implemented within NFFS-FFC and FEWS-Scotland.
- For the products and tools selected, a past analysis of performance will be made using the G2G ensemble forecast dataset. This is to provide confidence about the new products and may also help guide which ones to select. Again this will benefit from the ongoing “Verification of River Flow Ensembles” project.
- At this stage, direct use by the regions in England and Wales is anticipated in a follow on work package. This is intended to give the scope of the work in Phase 3 a clear and specific focus and would allow FFC and SFFS to trial the new tools and products before wider use. Engagement with the regions would be achieved through the Project Board and representation at the workshop.

References

Almond, C., 2013. Heavy rain and flooding in southwest England on 6–7 July 2012. Weather, 68(7),171-175.

Bell, V.A., Kay, A.L., Jones, R.G., Moore, R.J. & Reynard, N.S., 2009. Use of soil data in a grid-based hydrological model to estimate spatial variation in changing flood risk across the UK. J. Hydrol., 377, 335-350.

Bowler ,N.E., Pierce, C.E. & Seed, A.W., 2006. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. Quart. J. Roy. Met. Soc., 132(620), 2127–2155.

Centre for Ecology & Hydrology, 2013. Performance assessment of the G2G Model: a guide to the Performance Summary. Report to the Flood Forecasting Centre, Centre for Ecology & Hydrology, Wallingford, UK, 10pp. (first published 2010)

Cole, S.J. & Moore, R.J., 2008. Hydrological modelling using raingauge-and radar-based estimators of areal rainfall. J. Hydrol., 358(3–4), 159–181.

Cranston, M. D., 2006. Weather Radar and Flood Forecasting in Upland Scotland. M.Phil. Dissertation, University of Dundee, 112pp.

Cranston, M., Maxey, R., Speight, L., Tavendale, A., Cole, S., Robson, A. & Moore, R., 2013. The policy and science supporting flash flood forecasting in Scotland. Geophysical Research Abstracts. Vol. 15, EGU2013-12468.

Cranston, M., Maxey, R., Tavendale, A., Buchanan, P., Motion, A., Cole, S., Robson, A., Moore, R.J. & Minett, A., 2012. Countrywide flood forecasting in Scotland. challenges for hydrometeorological model uncertainty and prediction. In: Weather Radar and Hydrology (ed. by R.J Moore, S.J. Cole & A.J. Illingworth) (Proc. Exeter Symp., April 2011), IAHS Publ. no. 351, 538-543.

Environment Agency, 2010. Hydrological modelling using convective scale rainfall modelling – phase 3. Project: SC060087/R3, Authors: J. Schellekens, A. R. J. Minett, P. Reggiani, A. H. Weerts (Deltares); R. J. Moore, S. J. Cole, A. J. Robson & V. A. Bell (CEH Wallingford). Research Contractor: Deltares and CEH Wallingford, Environment Agency, Bristol, UK, 231pp. http://nora.nerc.ac.uk/13829/.

Environment Agency, 2012. Evaluating G2G for use in Rapid Response Catchments – Phase 1. R&D Project Report SC110003/R1, Joint Defra/EA Flood and Coastal Erosion Risk Management, (Authors: Moore, R.J., Cole, S.J., Robson, A.J., Howard, P.J., Black, K.B.), Research Contractor: CEH Wallingford, 52pp.

Environment Agency, 2013. 2012 Flooding events – England and Wales. Environment Agency internal report, 5pp.

Howard, P.J., Cole, S.J., Robson, A.J. & Moore. R.J., 2012. Raingauge quality-control algorithms and the potential benefits for radar-based hydrological modelling. In: Weather Radar and Hydrology (ed. by R. J Moore, S. J. Cole & A. J. Illingworth) (Proc. Exeter Symp., April 2011), IAHS Publ. no. 351, 219-224.

Jones, A.E., Jones, D.A. & Moore, R.J., 2003. Development of Rainfall Forecast Performance Monitoring Criteria. Phase 1: Development of Methodology and Algorithms. Report to the Environment Agency and the Met Office, CEH Wallingford, 291pp.

Moore, R.J., Bell, V.A., Cole, S.J. & Jones, D.A., 2007. Rainfall-runoff and other modelling for ungauged/low-benefit locations. Science Report – SC030227/SR1, Research Contractor: CEH Wallingford, Environment Agency, Bristol, UK, 249pp.

Moore, R.J., Cole, S.J., Bell, V. A. & Jones, D.A., 2006. Issues in flood forecasting: ungauged basins, extreme floods and uncertainty. In: Frontiers in Flood Research (ed. by I. Tchiguirinskaia, K. N. N. Thein & P. Hubert), 8th Kovacs Colloquium, UNESCO, Paris, June/July 2006, IAHS Publ. 305, 103–122.

Moore, R. J., Jones, A.E., Jones, D.A., Black, K.B. & Bell, V.A., 2004. Weather radar for flood forecasting: some UK experiences. 6th Int. Symp. on Hydrological Applications of Weather Radar, 2-4 February 2004, Melbourne, Australia, 11pp.

Parkin, G., 2010. The September 2008 Morpeth flood: information gathering for dynamic flood reconstruction. Summary report as part of the NERC FREE research programme. Newcastle University, UK, 22pp. Available online: http://www.nerc.ac.uk/research/programmes/free/resources/geoff-parkin-morpeth-report.pdf

Parry, S, Marsh, T., & Kendon, M.C. 2013. 2012: from drought to floods in England and Wales. Weather (In press), DOI: 10.1002/wea.2152.

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

Roberts, N., 2006. Storm scale numerical modelling. R&D Technical Report FD2207/TR, Research Contractor: Met Office, Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme, Defra, London, UK, 88pp.

Tang, Y., Lean, H.W. & Bornemann, J., 2013. The benefits of the Met Office variable resolution NWP model for forecasting convection. Meteorol. Appl., 20, 417-426..

Webb, S., 2013. Heavy rain and flooding in and around Aberystwyth on 8-9 June 2012. Weather, 36(6), 162-165.

We are The Environment Agency. It's our job to look after your environment and make it **a better place** – for you, and for future generations.

Your environment is the air you breathe, the water you drink and the ground you walk on. Working with business, Government and society as a whole, we are making your environment cleaner and healthier.

The Environment Agency. Out there, making your environment a better place.

Published by:

Environment Agency
Horizon House, Deanery Road,
Bristol, BS1 9AH
Tel: 0870 8506506
Email: enquiries@environment-agency.gov.uk
www.environment-agency.gov.uk