Surface Water Flooding Component for NHP HIM: Phase 2 Report

D2.5 Linking G2G to JFlow Inundation Model: SWF Hazard Footprint

Final Project Report

August 2016

Flood Forecasting Centre
Exeter, Devon, EX1 3PB
JBA Project Manager
Dr Neil Hunter
JBA Consulting
1 Belle Vue Square
Broughton Road
Skipton
BD23 1FJ

Revision History

<table>
<thead>
<tr>
<th>Revision Ref / Date Issued</th>
<th>Amendments</th>
<th>Issued to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version 1 (01/07/2016)</td>
<td>-</td>
<td>Steven Cole, Robert Moore</td>
</tr>
<tr>
<td>Version 2 (23/08/2016)</td>
<td>Minor updates based on feedback</td>
<td>Steven Cole, Robert Moore</td>
</tr>
</tbody>
</table>

Contract

This report describes work commissioned by Robert Moore, on behalf of Centre for Ecology & Hydrology, by a letter dated 20 May 2015. Centre for Ecology & Hydrology’s representative for the contract was Robert Moore. Sarah Warren, Neil Hunter and Beatriz Revilla-Romero of JBA Consulting and Steven J. Cole, Robert J. Moore and Steven C. Wells of Centre for Ecology & Hydrology carried out this work.

Prepared by .................................................Sarah Warren BSc MSc
Analyst

Beatriz Revilla-Romero BSc MSc PhD
Analyst

Neil Hunter BSc PhD MCiWEM C.WEM
Head of Flood Modelling

Reviewed by ..................................................Andrew Gubbin BSc MSc PGDip CEng MICE
FCiWEM C.WEM MBCS
Director

Purpose

This document has been prepared as a Final Report for Centre for Ecology & Hydrology. JBA Consulting accepts no responsibility or liability for any use that is made of this document other than by the Client for the purposes for which it was originally commissioned and prepared.

JBA Consulting has no liability regarding the use of this report except to Centre for Ecology & Hydrology.
Acknowledgements

We are grateful to Dr Luke Smith and Dr Qiuhua Liang of Newcastle University for providing geo-referenced social media observations and model results for the Newcastle 2012 flood event.

Copyright

© Jeremy Benn Associates Limited 2016

Carbon Footprint

A printed copy of the main text in this document will result in a carbon footprint of 399g if 100% post-consumer recycled paper is used and 313g if primary-source paper is used. These figures assume the report is printed in black and white on A4 paper and in duplex.

JBA is aiming to reduce its per capita carbon emissions.
Executive summary

WP2.5 aimed to increase understanding of the Surface Water Flooding (SWF) hazard footprint assumptions within the Surface Water Flooding Hazard Impact Model (SWF HIM). The SWF HIM is intended to provide better information to inform the surface water flooding element of the Flood Guidance Statement and combines surface runoff estimates from the Flood Forecasting Centre’s distributed Grid-to-Grid (G2G) forecasting model with impact libraries derived off-line using existing receptor information and surface water flood mapping. In the current implementation, receptor and flood map data are provided by the Environment Agency’s National Receptor Dataset (NRD) and Updated Flood Map for Surface Water (uFMfSW) respectively.

Here, instead of selecting the most appropriate footprint from pre-computed flood mapping, runoff forecasts were used to provide boundary conditions to high resolution 2D hydraulic models. Grid-to-Grid (G2G) surface runoff grids driven by observed rainfall and forecast rainfall were supplied by the Centre for Ecology & Hydrology (CEH) at 15-minute, 1km x 1km resolution for two case study flood events, Newcastle upon Tyne (28 June 2012) and Canvey Island (20 July 2014). The hydraulic models produced flood outlines associated with these events for both the observed (simulation outputs) and forecast rainfall (ensemble outputs).

This work therefore sought to compare the hazard footprints produced using the current SWF HIM “look up” approach with those produced from pseudo real-time hydraulic modelling. The aim being to confirm (or not) the suitability of the current approach for deriving SWF hazard footprints without recourse to event-specific flood modelling.

The simulation flood outlines were compared with observed flood data, including observed depths, impacted areas and photographs, to validate the linked G2G-JFlow model outputs. The model was found to perform well for Newcastle but less so for Canvey Island due the effect of a faulty rain gauge on rainfall inputs.

The simulation flood outlines were compared with observed flood data, including observed depths, impacted areas and photographs, and with nine scenarios (1-hour, 3-hour and 6-hour durations for each of the 30, 100 and 1,000-year return periods) from the uFMfSW to determine which scenario matched the simulation outline most closely in each 1km x 1km grid square in the case study areas. The patterns of the best fit scenarios were logical given the distribution of surface runoff; the 1,000-year scenarios (greater flood extents) fitted best in locations with high runoff and the 30-year scenarios (smaller flood extents) fitted best in areas with little runoff. Overall, the best fit scenarios had high “measure of fit” (MoF) values for Newcastle but fitted less well to Canvey Island. This could be a result of poorly recorded rainfall for Canvey Island which resulted in little flooding in the simulation flood outlines and fit was generally worst in areas with little flooding as there is increased sensitivity to drainage modelling assumptions in less extreme events. Both case studies produced a better overall fit (“total MoF”) when comparing the simulation outlines with the nine scenarios than with the three scenarios.

The simulation flood outlines were also compared with those of the ensemble forecasts to determine which ensemble member matched most closely and again Newcastle showed good fits so at least one ensemble was predicting the simulation well in most grid squares (the majority of grid squares had MoF values greater than 0.8). The total MoF was high (> 0.65) for two of the forecast origins for Canvey Island but low for the other two (< 0.5) but again this was influenced by the poor quality of the observed rainfall data in the simulation. Additionally, the location and extent of flooding in the ensembles vary greatly and sometimes only one or two ensembles contain flooding at
a given location, which can lead to lower MoF values as there is a limited choice of ensembles. This variability is due to the lower predictability of convective events within the Numerical Weather Prediction models used to provide the rainfall inputs to G2G.

The ensemble flood outlines were also compared with the uFMfSW scenarios for Newcastle and these were summarised as the number of ensembles that each scenario fits best, the scenario with the highest ensemble count and the maximum MoF value associated with that scenario over the ensembles for which it is the best fit. The 30-year scenarios often had the highest ensemble count but this was a likely impact of only half the ensembles containing any significant flooding. It would therefore also be useful to use the comparison with the maximum MoF to identify a worst case scenario.

The final part of the analysis compared our “best fit” scenarios from the simulation comparison with the uFMfSW scenarios for the Newcastle case study with the equivalent identified by CEH using the approach currently implemented within the SWF HIM. CEH compared maximum surface runoffs for 1, 3 and 6-hour durations over their respective whole forecast time-windows with the corresponding 1, 3 or 6-hour uFMfSW effective rainfall grids for the 30, 100 and 1,000-year return periods, noting the highest threshold crossed. Error matrices were created to compare the grids and an overall accuracy and Kappa statistic were calculated. It was found that this G2G-only method generally showed lower return periods than the linked G2G-JFlow method in many grid squares, probably due to the fact that the G2G-only method required the threshold to be crossed whereas the linked G2G-JFlow method found the uFMfSW scenario with the closest match. Some sensitivity analysis of uFMfSW thresholds was also carried out by repeating these comparisons with relaxed thresholds for the G2G-only method and lowering the thresholds did increase the similarity of the grids produced by the two methods.
## Contents

1. **Introduction**
   - 1.1 Context and drivers  8
   - 1.2 Scope of work  8

2. **Linking Grid-to-Grid hydrological and JFlow hydraulic models**
   - 2.1 Overview of models  10
   - 2.1.1 Grid-to-Grid hydrological model  10
   - 2.1.2 JFlow hydraulic model  10
   - 2.2 Linking Grid-to-Grid hydrological and JFlow hydraulic models  11
   - 2.3 Case study events  11
     - 2.3.1 Newcastle, 28 June 2012  11
     - 2.3.2 Canvey Island, 20 July 2014  12
     - 2.3.3 Boundary condition data from Grid-to-Grid  13
     - 2.3.4 Validation data  15
     - 2.3.5 Updated Flood Map for Surface Water data  16

3. **Comparison of G2G-JFlow modelling with flood observations**
   - 3.1 Approach  18
   - 3.2 Analysis and findings  18
     - 3.2.1 Newcastle case study  18
     - 3.2.2 Canvey Island case study  22

4. **Comparison of G2G-JFlow modelling using observed and forecast rainfall and uFMfSW mapping**
   - 4.1 Approach  25
     - 4.1.1 Additional analysis for comparison of G2G-JFlow modelling using observed rainfall with uFMfSW mapping and with G2G-JFlow modelling using forecast rainfall  28
     - 4.1.2 Additional analysis for comparison of G2G-JFlow modelling using forecast rainfall with uFMfSW mapping  28
   - 4.2 Comparison of G2G-JFlow modelling using observed rainfall with uFMfSW scenarios  29
     - 4.2.1 Analysis and findings  29
   - 4.3 Comparison of G2G-JFlow modelling using forecast rainfall with uFMfSW scenarios  34
     - 4.3.1 Analysis and findings  34
   - 4.4 G2G-JFlow modelling using observed and forecast rainfall  38
     - 4.4.1 Analysis and findings  38

5. **Testing SWF HIM assumption of using the accumulated G2G runoff to identify the most appropriate uFMfSW scenario**
   - 5.1 Data  47
   - 5.2 Approach  49
   - 5.3 Analysis and findings  50
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4</td>
<td>Sensitivity analysis</td>
<td>53</td>
</tr>
<tr>
<td>6</td>
<td>Conclusions and recommendations for further work</td>
<td>63</td>
</tr>
<tr>
<td>6.1</td>
<td>Conclusions</td>
<td>63</td>
</tr>
<tr>
<td>6.2</td>
<td>Recommendations for further work</td>
<td>65</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>Appendix A – Newcastle surface runoff grids provided by CEH</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Appendix B – Canvey Island surface runoff grids provided by CEH</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Appendix C – G2G SWF hazard footprints</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Appendix D – G2G SWF hazard footprints for sensitivity analysis</td>
<td>71</td>
<td></td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Context and drivers

To improve the forecasting of surface water flooding and associated impacts across England and Wales, a prototype system based on coupled hydro-meteorological hazard modelling and pre-computed impact libraries has been developed and tested by the Natural Hazards Partnership (NHP). The Surface Water Flooding Hazard Impact Model (SWF HIM) is intended to provide better information to inform the surface water flooding element of the Flood Guidance Statement (FGS) produced by the joint Environment Agency- Met Office Flood Forecasting Centre (FFC).

The SWF HIM combines 1km x 1km surface runoff estimates from the FFC’s distributed Grid-to-Grid (G2G) forecasting model with impact libraries derived off-line using existing receptor information and surface water flood mapping. In the current implementation, receptor and flood map data are provided by the Environment Agency’s National Receptor Dataset (NRD) and Updated Flood Map for Surface Water (uFMfSW) respectively. Full details of the SWF HIM approach are provided in Gunawan and Aldridge (2016).

The aim of this project is to develop the proof-of-concept SWF HIM into a real-time, end-to-end system that can then be trialled in operational use by FFC hydro-meteorologists. Work Package 2 of this project focuses on improving the SWF hazard footprint used within the HIM and WP2.5 is specifically concerned with alternative methods for downscaling the G2G runoff forecasts. Here, instead of selecting the most appropriate footprint from pre-computed flood mapping, runoff forecasts are used to provide boundary conditions to high resolution 2D hydraulic models. Surface runoff is then routed across the digital terrain model (DTM) surface within the flood model, identifying flooding pathways and areas where ponding will occur. WP2.5 therefore seeks to compare the hazard footprints produced using the current SWF HIM “look up” approach with those produced from pseudo real-time hydraulic modelling. The aim being to confirm (or not) the suitability of the current approach for deriving SWF hazard footprints without recourse to event-specific flood modelling.

1.2 Scope of work

This report contains details of the analysis carried out in this project, where Grid-to-Grid surface runoff grids driven by observed and forecast rainfall were run through JFlow to produce flood outlines based on both the observed (simulation outputs) and forecast rainfall (ensemble outputs) for two case study events, Newcastle upon Tyne (28 June 2012) and Canvey Island (20 July 2014). Section 2 provides an overview of the G2G and JFlow models and how they were linked together. Background on the two case studies and the data used throughout the analysis is provided in Section 2.3.

The simulation flood outlines were compared with observed flood data to identify how well the linked G2G-JFlow model performs and the approach and findings from this are discussed in Section 3.

Section 4.1 outlines the approach used to compare flood outlines from the G2G-JFlow modelling using observed rainfall, from the G2G-JFlow modelling using forecast rainfall and from the Updated Flood Map for Surface Water.

The simulation flood outlines were compared with three scenarios (“worst case” scenarios for the 30, 100 and 1,000-year return periods) and nine scenarios (1-hour, 3-
hour and 6-hour for each of the 30, 100 and 1,000-year return periods) from the uFMfSW to determine which scenario matched the simulation outline most closely in each 1km grid square in the case study area. The analysis and findings from this analysis are provided in Section 4.2.

The simulation flood outlines were also compared with those of the ensembles to determine which ensemble matched most closely. A “measure of fit” (MoF) statistic, described in Section 4.1, was used to quantify the fit. The results from this analysis are presented in Section 4.4.

Additionally, the ensemble flood outlines were compared with the uFMfSW scenarios for Newcastle and these were summarised as the number of ensembles that each scenario fits best, the scenario with the highest ensemble count and the maximum MoF value associated with that scenario over the ensembles for which it is the best fit. The results from this analysis are provided in Section 4.3.

The final part of the analysis compared our “best fit” scenarios from the simulation comparison with the uFMfSW scenarios for the Newcastle case study with the equivalent identified by CEH using the approach currently implemented within the SWF HIM. The G2G-only method required the uFMfSW threshold to be crossed for the scenario to be selected so some sensitivity analysis of uFMfSW thresholds was also carried out by repeating these comparisons with relaxed thresholds for the G2G-only method. Error matrices were created to compare the grids and an overall accuracy and Kappa statistic were calculated. These results are provided in Section 5.
Linking Grid-to-Grid hydrological and JFlow hydraulic models

2.1 Overview of models

2.1.1 Grid-to-Grid hydrological model

The SWF HIM is underpinned by the Grid-to-Grid distributed grid-based rainfall-runoff and routing model developed by CEH (Moore et al., 2006, 2007; Bell et al., 2009). It is already in real-time operational use by FFC for forecasting fluvial flooding across England & Wales (Price et al., 2012) and supports the fluvial elements of the Flood Guidance Statement. It runs at a 15-minute time-step and is currently configured at a 1km x 1km grid resolution. In addition to providing time- and space-varying river flow and soil moisture estimates, G2G also provides surface runoff as an average water depth over a grid square in mm and it is this that is used to estimate the SWF hazard footprint. There is no explicit routing within the calculation of surface runoff. The surface runoff does contribute to the surface flow routing scheme which routes water from grid-to-grid to yield fluvial flow estimates across the model domain.

The physical-conceptual formulation of the hydrological processes in the model is underpinned by spatial datasets and therefore G2G runoff production is shaped by the storm pattern, landscape properties - land-cover (e.g. urban/sub-urban), terrain, soil and geology - along with dynamically and spatially changing antecedent soil moisture as calculated through continuous water accounting within G2G.

Initial case study trials (Cole et al., 2013) suggest that using dynamic gridded surface runoff estimates from G2G can provide a potentially significant step forward in assessing the SWF hazard footprint, compared to methods primarily based on rainfall depth (e.g. the Surface Water Flooding Decision Support Tool, see Halcrow (2011)). This potential is further explored as part of Work Package 2 of the current project.

2.1.2 JFlow hydraulic model

G2G runoff forecasts are routed overland using the JFlow 2D hydraulic model (Lamb et al. (2009), Crossley et al. (2010a, 2010b), www.jflow.co.uk). JFlow solves the Shallow Water Equations using a finite volume formulation that combines the Riemann based solver of Roe with an upwind treatment of the source terms. The model is both conservative and shock capturing, and maintains water at rest over irregular topography. JFlow is implemented on a regular grid using the supplied DTM and does not require any secondary grid generation process. This simplifies the model set up and allows for direct interpretation of the model results relative to the DTM.

JFlow has been designed with the emphasis on easy set up and model specification. Models are configured using databases, and this provides a highly ordered means to store significant quantities of data. The modelling engine is controlled through a web interface, ensuring that projects and models can be set up, run and monitored from any location in which an internet connection is available.

JFlow has been benchmarked using the test cases proposed by the Environment Agency of England in the Science Report SC080035/SR2, Benchmarking of 2D
Hydraulic Modelling Packages, and the results have been submitted to the Environment Agency. The software was also used to undertake the hydraulic modelling that underpins the Updated Flood Map for Surface Water dataset (Environment Agency, 2013).

2.2 Linking Grid-to-Grid hydrological and JFlow hydraulic models

G2G runoff estimates for a particular event are provided as a time-series of ASCII raster grids at 15-minute intervals and 1km x 1km horizontal resolution. These time- and space-varying data are applied within the 2D flood model as a vertical boundary condition analogous to the "direct rainfall" approach used to produce the uFMfSW (Defra, 2010). However, unlike the uFMfSW modelling, infiltration losses are accounted for within G2G and therefore there is no further adjustment of these boundary conditions within JFlow.

Data from the uFMfSW is used to parameterise floodplain topography and hydraulic roughness across the two case-study areas (see below).

Outputs from the flood models require GIS post-processing to improve their usability for identifying areas and receptors at risk of surface water flooding. The same approach and parameter specification used for the uFMfSW is applied here (described in Section 4 of Environment Agency (2013)).

2.3 Case study events

Two case studies have been used throughout this analysis; both are recent summer events driven by convective storms. They occurred on 28 June 2012 and 20 July 2014 in Newcastle upon Tyne and Canvey Island respectively. The rainfall events and the resulting flooding are described below.

2.3.1 Newcastle, 28 June 2012

On 28 June 2012, intense convective thunderstorms delivered, in parts of the city of Newcastle upon Tyne, up to 50mm of rainfall in 2 hours – this was equivalent to the 100-year return period storm (Newcastle City Council, 2013a). This rain fell onto ground that was already saturated by high rainfall accumulations throughout June 2012 (Environment Agency, 2012). Alongside runoff from extensive impermeable urban land cover, this resulted in large volumes of surface runoff that flooded over 500 properties. Much of this flooding occurred within one hour (Newcastle City Council, 2013b). The scale of the event, combined with the rapid onset of flooding, left people with little time to react and caused significant disruption to the city.

The nature of surface water flooding, and the rapid onset of this event, means that it is rare to find conventional records of flood extent and depth (e.g. gauge data, aerial photographs, wrack marks). However, this event – which began at around 15:00GMT in a busy, urban area – was well documented by members of the public through large numbers of photographs and other eyewitness reports of flooding. Much of this information was collated and georeferenced by Newcastle University, and we are grateful to the university for providing the data.

A map of the 20km x 20km study area is shown in Figure 2.1. All the subsequent maps for Newcastle cover this domain except the flood outlines for the ensembles (Figure 4.9), which show a subset of the area. The 1km grid shown on the map is the resolution of the CEH G2G surface runoff data and is the resolution used throughout
the analysis for all comparisons. The Mean High Water Springs (MHWS) tidal boundary is also highlighted on the map.

![Map of the Newcastle study area with the 1km grid overlaid (black) and the Mean High Water Springs tidal boundary highlighted in blue.](image)

**Figure 2.1** Map of the Newcastle study area with the 1km grid overlaid (black) and the Mean High Water Springs tidal boundary highlighted in blue.

### 2.3.2 Canvey Island, 20 July 2014

Canvey Island has a history of coastal and surface water flooding. One of the largest surface water events recorded occurred on 20 July 2014, when intense rainfall – estimated approximately as a 300-year return period storm – resulted in substantial flooding to properties and infrastructure (Essex County Council, 2014).

As with the Newcastle surface water event described above, conventional observations of flood extent are sparse. The main records available to assess this event are from the media database described in Section 2.3.4.

A map of the 17km x 15km study area is shown in Figure 2.2. All the subsequent maps for Canvey Island cover this domain. The 1km grid shown on the map is the resolution of the CEH G2G surface runoff data and is the resolution used throughout the analysis for all comparisons. The Mean High Water Springs (MHWS) tidal boundary is also highlighted on the map.
Figure 2.2 Map of the Canvey Island study area with the 1km grid overlaid (black) and the Mean High Water Springs tidal boundary highlighted in blue.

2.3.3 Boundary condition data from Grid-to-Grid

Surface runoff grids from G2G were provided by CEH for both case studies. Grids were provided as ASCII files at 15-minute time steps at a 1km resolution. Simulation-mode outputs using gridded rain gauge data as inputs were provided along with ensemble outputs driven by probabilistic rainfall forecasts.

There is large variability in flooding across forecast ensemble members due to the nature of the convective storms that generated flooding across Newcastle and Canvey Island. This type of storm is harder to predict with high certainty, and therefore a greater spread of ensemble members is produced. Recent winter storms of 2015/16 should exhibit less spread across ensemble members as frontal rainfall has higher predictability. This because the spatial distribution of rainfall within frontal weather systems is generally determined by larger-scale processes, i.e. rainfall occurs where air is ascending up a frontal boundary. The spatial and temporal variability of smaller-scale convective storms is more strongly influenced by local processes and the interaction of these local processes with those operating at larger scales. Hence, where and when a convective storm may occur can vary greatly between ensemble members, where small differences in the state of the atmosphere result in differences in local-scale processes.
Each G2G dataset is based on the rainfall forecast product that was available at the time of the event. It should be noted that improvements in rainfall forecasting since the time of the case study events, particularly in the spatial and temporal resolution of the data available, will improve the quality of forecasts derived from G2G in the future.

**Newcastle**

The simulation data used as inputs to JFlow for Newcastle start from 28/06/2012 00:00 and finish at 29/06/2014 23:45. The total and maximum simulation runoff (driven by observed rainfall) for each 1km grid square is shown in Figure 2.3 and the majority of the runoff is in the north east corner of the study area.

![Figure 2.3 Total and maximum runoff for Newcastle driven by observed rainfall (simulation data).](image)

There were 12 ensemble members provided for Newcastle for one forecast origin (07:15 GMT 28 June 2012) and the time steps used as inputs to JFlow started at 28/06/2012 07:30 (15-minute lead time) and finished at 29/06/2012 07:15. G2G was driven by the Short Term Ensemble Prediction System (STEPS-2) with UK4 rainfall forecasts. Further details of this data are given in Appendix A.

**Canvey Island**

For Canvey Island, the simulation data used as inputs to JFlow started at 19/07/2014 13:00 and ended at 21/07/2014 15:00. After much of the analysis had been completed, it transpired that the rain gauge totals used for the simulation runs may have been poorly recorded due to the gauge not functioning properly, despite being well-sited. This has resulted in very little runoff in Canvey Island and therefore the simulation results are less relevant here. CEH has subsequently re-run the G2G simulations using rainfall radar although the runoff outputs from these were not available to JBA Consulting within the time frame of this analysis.

The impact of this is shown in Figure 2.4, which shows the total and maximum runoff produced from the observed rainfall. The highest runoff is away from Canvey Island.
itself and towards the north and middle of the study area. There are seven grid squares which do not contain data and these should be ignored when interpreting the data as the results will not be valid for these squares.

Figure 2.4 Total and maximum runoff for Canvey Island driven by observed rainfall (simulation data).

There were 24 ensemble members provided for ten forecast origins for Canvey Island. Four of these forecast origins covered the event very well and were considered to be most useful, hence 96 ensembles were used in the analysis. The forecast origins were as follows, with the first and last time steps used as inputs to JFlow also given (there is a 15-minute lead time for the first time step and the forecast lead time goes up to 32 hours):

- 19/07/2014 13:00 (from 19/07/2014 13:15 to 20/07/2014 21:00)
- 19/07/2014 19:00 (from 19/07/2014 19:15 to 21/07/2014 03:00)
- 20/07/2014 01:00 (from 20/07/2014 01:00 to 21/07/2014 09:00)
- 20/07/2014 07:00 (from 20/07/2014 07:15 to 21/07/2014 15:00)

Gridded rain gauge data were used to initialise the G2G model up to the forecast origin and the G2G surface runoff forecasts were based on the Short Term Ensemble Prediction System (STEPS-2) with MOGREPS-UK Blended Ensembles (provided by the Met Office). Further details of this data can be found in Appendix B.

2.3.4 Validation data

Validation data were used to establish linked G2G-JFlow model performance. Both case studies have at least several sources of spatially-distributed observations of flood extents and impacts. These data range from point based, local measurements to coarse resolution datasets, with wide spatial coverage.

A media-based impacts database was provided by the Health and Safety Laboratory (HSL), King’s College London (KCL) and the Flood Forecasting Centre (FFC) and this was available for both Newcastle and Canvey Island. This was derived by searching
media reports of flood impacts and, where possible, geo-references this information at points (34 exist for Newcastle), polylines (i.e. roads) and polygons (broad-scale impact information). In addition, post-event measurements of observed depth were available for Newcastle and the Section 19 Flood Incident Report (Government Office for Science and Essex County Council) for Canvey Island.

Further information about the validation data is given below:

**Observed depths:** Nine measurements of observed depths were available for the Newcastle event only. These were based on photographs of flooding at known locations. Depths were recorded after the event, by measuring the level of water against structures that were visible in the photographs (e.g. bins, phone boxes and other street furniture). The data were provided by Newcastle University (Bertsch, 2013).

**Photographs:** 117 photographs of the Newcastle event were provided by Newcastle University. These were filtered by known location and time, as follows:

- Most (81 photos) were attributed with the name of the location. Some (36) were of unknown locations.
- 66 photographs were georeferenced. Of those, 31 have a registered time.

These 31 georeferenced photographs with time tags provide wide spatial coverage of the city, although they miss some potential key areas shown in other georeferenced photographs without a registered time. The times of the photographs range from 16:17 until 21:18 of 28 June 2012.

**Health and Safety Laboratory (HSL)/King's College London (KCL)/Flood Forecasting Centre (FFC) database:** Contains points (34 for Newcastle, 5 for Canvey Island), roads and polygons that register a flood impact. This was based on mining media data for impact verification methodology.

**Section 19 Flood Investigation Report:** provided by the Government Office for Science and Essex County Council. Reports are only available for Canvey Island. These can be downloaded from:

**Social media information:** E.g. Twitter (#CanveyIsland flooding)

### 2.3.5 Updated Flood Map for Surface Water data

The Updated Flood Map for Surface Water was produced and published in 2013 to help the Environment Agency and Lead Local Flood Authorities meet their obligations for Flood Hazard and Risk Mapping under the Flood Risk Regulations 2009. It was also intended to provide a definitive and high quality dataset to support flood risk management and spatial and emergency planning activities at the local and national scales. The methodology and outputs available are fully described in Environment Agency (2013).

The uFmISW comprises 2m resolution flood mapping across England and Wales for the 9 combinations of 3 storm durations (1, 3 and 6 hours) and 3 rainfall probabilities (1 in 30, 1 in 100 and 1 in 1,000 years). However, the published product ([http://watermaps.environment-agency.gov.uk/wiyby/wiyby.aspx?topic=ufmfw](http://watermaps.environment-agency.gov.uk/wiyby/wiyby.aspx?topic=ufmfw)) shows only the worst case flood extents, depths, velocities and hazard rating associated with each of the three probabilities. This worst case is determined for each flood variable by
producing a grid of maximums across the different storm durations within a GIS to identify the worst case result in every 2m grid cell. These two groups of map products are referred to as the 9 and 3 scenarios respectively hereafter.

The complete uFMfSW is now freely available under HM Government’s Open Data Initiative, data.gov.uk.
3 Comparison of G2G-JFlow modelling with flood observations

3.1 Approach

Quantitative and qualitative assessments were undertaken to compare the linked G2G-JFlow model output to the validation datasets. The JFlow simulation flood outlines and depths were plotted with observed depths, with geo-referenced photographs and known impacted locations to enable validation of the modelled flood depths and extents.

3.2 Analysis and findings

3.2.1 Newcastle case study

Predictions of flood extent and depth from the G2G-JFlow model generally match well with the diverse observed datasets, with limited bias towards over- or under-prediction. The outputs from JFlow were compared with observed flood depths, photographs, impacted roads from news reports and severity at specific locations. These validation datasets proved to be very useful in identifying the accuracy of the simulation flooding in this case study. In addition, the results from the JFlow model, driven by forecast rainfall, show that areas of highest agreement between ensemble members tend to be low-lying areas and topographic depressions where flood water will pond/accumulate. These are the areas that are most susceptible to surface water flooding. However, there are wide areas of poorer agreement where flooding is shown in only one to four ensemble members (see Figure 4.9) and this has an impact on the results in Section 4.3. This results from high variability in the flood extents between ensemble members – short duration, high intensity convective storms are difficult to capture far in advance by forecast rainfall products. In general, most of the ensemble members tend to under-predict the flood extent in comparison to the JFlow simulation driven by observed rainfall.

Modelled depth comparison with observations

Figure 3.1 shows comparisons of modelled depths with point observations. The maps show depth of the modelled flooding, using observed rainfall on a 2m x 2m grid. At nine locations, the measured point depth for the 28/06/2012 flood event is shown. In general, the results show good agreement between the modelled and the observed depths at these locations. All these points are located in areas where water accumulates, but include a range of depths – they are not just focused where the water is very deep (>0.9m). Points E, B and I are particularly encouraging and show that the G2G-JFlow model is not grossly over- or under-predicting depths at these locations.
Figure 3.1 JFlow depth comparison with point observations (provided by Newcastle University) for Newcastle.

Spatial comparison with geo-referenced photographs

Figure 3.2 shows comparison with geo-referenced photographs. In total, the dataset provided contains 66 georeferenced photographs. Of them, 31 also have date/time information. These 31 georeferenced and time-tagged data points have a wide spatial distribution, despite missing some areas shown as flooded in other georeferenced photographs without a registered time. The images in the figure show flood depths, modelled using observed rainfall, on a 2m x 2m grid, as with the previous comparison. At seven sample locations, the pictures taken during the 28/06/2012 flood event are shown. Modelled flood extent and depth is generally consistent with the information provided from the photographs at these locations.
Spatial comparison with the HSL/KCL/FFC flood impacts database

This dataset contains 34 impacted locations for the Newcastle 28/06/2012 flood event. Furthermore, some main roads were identified as impacted, from searching databases of published media reports. In total, there were 171 non-georeferenced reported incidents in the North East (Minimal: 15; Minor: 70, Significant: 84; Severe: 2) due to disruption and damage to traffic and communities. The severity of impacts is based on scoring used by the FFC within the Flood Guidance Statement matrix. Due to the localised nature of the analysis carried out here, it was not possible to evaluate whether all the impacted areas recorded on this dataset were impacted on the ground during this flood event. Results from this comparison are given in Figure 3.3.

The model simulated flooded areas along the main roads where flooding was reported. However, the model shows that many other smaller roads were also flooded. One limitation of using the media database for validation, is that some receptors were not sufficiently noteworthy to be reported by the local or national media. However, some of these smaller scale flood impacts were shown in previous validation analysis (e.g. based on photographs in the previous section) and these were also accurately simulated by the G2G-JFlow model.
Figure 3.3 Spatial comparison with HSL/KCL/FFC dataset for Newcastle.
3.2.2 Canvey Island case study

Predictions of flood extent and depth from the G2G-JFlow model match generally well with the diverse observed datasets, with limited bias towards over- or under-prediction. For this case study, a smaller number of datasets was available for validation. The outputs from the linked G2G-JFlow model were compared against photographs, impacted roads, and news reports. The simulated flood outline in Canvey Island was underestimated due to low accuracy of the observed rainfall data used to force the model. Elsewhere within the study area, the model simulated large flooded areas. However, the validation datasets did not provide full coverage for many of those areas, as the impacts of that flood event were minor in comparison to those focused on Canvey Island itself. The results from the G2G-JFlow model driven by forecast rainfall show low agreement between the ensemble members. This can be seen in Figure 4.13, Figure 4.14, Figure 4.15 and Figure 4.16, where there are locations where only one to four ensemble members show flooding, and this has an impact on the results in Section 4.4. This is the case even in low-lying areas and topographic depressions. This is partly due to the potential under-prediction by most of the G2G ensemble members across the study area, including over Canvey Island. The results suggest that this event was difficult to correctly forecast – the short duration, high intensity convective nature of the storm would have contributed to the challenges involved in accurately predicting rainfall.

Spatial comparison with HSL/KCL/FFC flood impacts database

Figure 3.4 shows results from comparing the simulated flood outline with the impacts database. The HSL/KCL/FFC dataset contains 5 impacted points for the Canvey Island 20/07/2014 event. However, all of them are outside Canvey Island itself, as can be seen on the top left figure.

Some main roads were identified as impacted from searching databases of published media reports. However, due to the localised analysis carried out here, it was not possible to evaluate whether all the impacted areas, identified in the database, were impacted during this flood event. Unfortunately, as explained in Section 2.3.3, the modelled output failed to fully capture the extent and intensity of the flood event due to the faulty rain gauge.
Figure 3.4 Spatial comparison with HSL/KCL/FFC dataset for Canvey Island.
Spatial comparison with Essex County Council (ECC) Section 19 Flood Investigation Report

ECC compiled a dataset of roads, partially or fully affected by flooding, exclusively for the Canvey Island area. This dataset is based on reports received from professional partners, residents and local councillors who were present during the event or spoke to eyewitnesses. Figure 3.5 shows the comparison and again the modelled output failed to fully capture the extent and intensity of the flood event due to the faulty rain gauge.

Figure 3.5  Spatial comparison with ECC dataset.
4 Comparison of G2G-JFlow modelling using observed and forecast rainfall and uFMfSW mapping

4.1 Approach

Outputs from the G2G-JFlow modelling using observed rainfall (simulation outputs) and using forecast rainfall (ensemble outputs) were compared with the 3 and 9 uFMfSW scenarios to determine the closest match within each 1km grid square. Similarly, a comparison was made between the simulation and ensemble outputs to identify which ensemble was closest to the simulation in each grid square.

For each of these comparisons, a map was created such as in Figure 4.1, where 0 = correct dry, 1 = correct wet, 2 = over-prediction and 3 = under-prediction. Each 4m² pixel contains a value of zero, one, two or three.

Figure 4.1 An example of a map produced when comparing two grids. The values indicate whether the grids agree that there is flooding in the pixel (1), whether one of the grids contains flooding where the other does not (2 or 3, depending which grid contains flooding) or whether neither grid contains flooding (0).
The counts of zeros, ones, twos and threes in each 1km grid square are then used to calculate a “Measure of Fit” (MoF) statistic. This is defined as:

\[
\text{MoF} = \frac{\text{Number of correct wet pixels}}{\text{Number of correct wet pixels} + \text{Number of over-predicted pixels} + \text{Number of under-predicted pixels}}
\]

Note that the count of “correct dry” pixels is not included as this can be skewed by the extent of unfloodable areas within the model domain.

Although the comparison shown in Figure 4.1 appears poor across the whole 20km x 20km area, it has some 1km grid squares that contain a lot of values of 1 and therefore will have a very high MoF value, such as at Point A. Point B, however, is in a grid square containing a lot of over-prediction so the fit will be much poorer.

The scenario (or ensemble) with the highest MoF value is determined as having the best fit. The MoF values range from zero to one with zero implying that there are no correct wet pixels (i.e. the grids do not contain any coincident flooding). A value of −1 is used for cases where both grids are completely dry (i.e. there is no over- or under-prediction).

The MoF and number of “correct wet” pixels for each comparison (e.g. simulation JFlow output with each of the 3 or 9 uFMfSW scenarios) are output to a shapefile, along with the best fit scenario and its corresponding MoF value, for each 1km grid square.

Initial outputs showed some very high MoF values, including 1, for some of the comparisons but many of these were in grid squares with very few “correct wet” pixels and a lot of “correct dry” pixels, i.e. squares containing very little flooding. Therefore, a count threshold was implemented based on the number of “correct wet” pixels. This has been set at 250 pixels, which is 1/1000th of the number of pixels in a 1km grid square; higher thresholds can exclude a lot of grid squares. Best fits are not found for grid squares with “correct wet” counts below the threshold.

The approach was implemented in Python and the logic used to determine the best fit and associated MoF value for each grid square over the comparisons is shown in the flow chart in Figure 4.2. For example, if comparisons with every scenario (or ensemble) produced an MoF value of −1, this implies that none of the grids in any of the comparisons had any flooding for that square so “all correct dry” is recorded. Alternatively, if comparisons with every scenario produced MoF values of 0 and −1, it implies that some of the comparisons do have over- or under-prediction occurring. In this case, the best fit would be one of the “correct dry” comparisons as they do not contain over- or under-prediction and therefore “correct dry” is written out alongside a value of −1. No scenario is selected in these two cases as the study is interested in where flooding occurs and selecting an arbitrary scenario would be meaningless.

The outputs from the analysis are the “best fit” scenario (or ensemble) and its associated MoF for each 1km grid square.
Figure 4.2 Flow chart of the logic used to determine the best fit scenario or ensemble and its associated MoF.
4.1.1 Additional analysis for comparison of G2G-JFlow modelling using observed rainfall with uFMfSW mapping and with G2G-JFlow modelling using forecast rainfall

For these comparisons, a “Total MoF” was calculated for the “Best Fit” grid, which takes the scenario with the best fit for each grid square and uses the counts of 0s, 1s, 2s and 3s for those comparisons to obtain a measure of fit representing a merged grid of all the best fit scenarios. For this calculation, grid squares for which the “correct wet” count was lower than the threshold were excluded. “No correct wet” only occurs in the simulation comparison with the uFMfSW scenarios and the squares which are labelled as “no correct wet” are included in the total MoF calculation with the best fit scenario selected as the one which has the least combined over- and under-prediction. “All correct dry” and “correct dry” are excluded from the calculation given that the best fit is where there is no flooding in either grid in the square.

4.1.2 Additional analysis for comparison of G2G-JFlow modelling using forecast rainfall with uFMfSW mapping

The comparison of the G2G-JFlow modelling using forecast rainfall (ensemble flood outlines) with the three and nine uFMfSW scenarios produced 12 sets of comparisons for Newcastle so additional analysis was carried out to summarise these comparisons.

For each of the 12 ensembles, for each grid square, the above analysis identified the uFMfSW scenario with the closest fit to the ensemble. This additional analysis identifies how many ensembles had a best fit to each uFMfSW scenario (treating the three and nine scenarios separately) and also the highest MoF value for each scenario. An example for one of the grid squares is shown in Table 4.1, with best fits highlighted for the three and nine scenarios. Some ensembles have been removed from the table as they all have “correct wet” counts below the threshold for comparison with every scenario. In this example, for the three scenarios, four ensembles fit the 30-year scenario best, and the highest MoF out of those four is 0.44. The maximum ensemble count is therefore 4. For the nine scenarios, three ensembles have a best fit with the 30-year 1-hour scenario and three have a best fit with the 100-year 1-hour scenario. Hence, the maximum ensemble count is 3.

Table 4.1 An example of the additional analysis for one grid square when comparing G2G-JFlow modelling using forecast rainfall with uFMfSW mapping.

<table>
<thead>
<tr>
<th>MoF_30yr</th>
<th>Ens_2</th>
<th>Ens_3</th>
<th>Ens_6</th>
<th>Ens_10</th>
<th>Ens_11</th>
<th>Ens_12</th>
<th>Highest MoF</th>
<th>Ensemble Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.203</td>
<td>0.233</td>
<td>0.439</td>
<td>0.437</td>
<td>0.407</td>
<td>0.281</td>
<td>0.439</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>0.134</td>
<td>0.153</td>
<td>0.438</td>
<td>0.441</td>
<td>0.428</td>
<td>0.185</td>
<td>0.441</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0.066</td>
<td>0.076</td>
<td>0.289</td>
<td>0.305</td>
<td>0.375</td>
<td>0.091</td>
<td>0.375</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MoF_100yr_1hr</th>
<th>Ens_2</th>
<th>Ens_3</th>
<th>Ens_6</th>
<th>Ens_10</th>
<th>Ens_11</th>
<th>Ens_12</th>
<th>Highest MoF</th>
<th>Ensemble Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.345</td>
<td>0.381</td>
<td>0.396</td>
<td>0.381</td>
<td>0.334</td>
<td>0.410</td>
<td>0.410</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>0.239</td>
<td>0.274</td>
<td>0.441</td>
<td>0.436</td>
<td>0.395</td>
<td>0.331</td>
<td>0.441</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.203</td>
<td>0.233</td>
<td>0.439</td>
<td>0.437</td>
<td>0.407</td>
<td>0.281</td>
<td>0.439</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.207</td>
<td>0.237</td>
<td>0.461</td>
<td>0.461</td>
<td>0.430</td>
<td>0.285</td>
<td>0.461</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>0.149</td>
<td>0.171</td>
<td>0.439</td>
<td>0.449</td>
<td>0.428</td>
<td>0.206</td>
<td>0.449</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.136</td>
<td>0.155</td>
<td>0.431</td>
<td>0.433</td>
<td>0.422</td>
<td>0.187</td>
<td>0.433</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.077</td>
<td>0.088</td>
<td>0.324</td>
<td>0.354</td>
<td>0.410</td>
<td>0.106</td>
<td>0.410</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.069</td>
<td>0.079</td>
<td>0.302</td>
<td>0.319</td>
<td>0.384</td>
<td>0.096</td>
<td>0.384</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.071</td>
<td>0.081</td>
<td>0.306</td>
<td>0.323</td>
<td>0.379</td>
<td>0.098</td>
<td>0.379</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
From this table, the maximum ensemble count, its associated scenario and best MoF value were extracted to a shapefile and plotted (see Section 4.3). Where the maximum ensemble count is tied between two scenarios, both are extracted but plotted as “No clear winner” and the best MoF value is taken as the highest out of the six highlighted in yellow in Table 4.1. In this case, for the nine scenarios, the highest MoF value is 0.46. This analysis was implemented using the programming language R.

This additional analysis was only completed for the Newcastle case study.

4.2 Comparison of G2G-JFlow modelling using observed rainfall with uFMfSW scenarios

4.2.1 Analysis and findings

The simulation maximum flood outlines produced by running G2G surface runoff driven by observed rainfall through JFlow were compared to nine uFMfSW outlines (1, 3 and 6-hour for each of the 30, 100 and 1,000-year return periods) and also to three “worst case” uFMfSW outlines using the approach outlined in Section 4.1. The analysis was run for both the Newcastle and Canvey Island case studies.

Newcastle case study

The simulation flood extent for Newcastle is shown in Figure 4.3. The zoomed in extent is the extent of the plotted ensemble outlines in Figure 4.9.

Figure 4.3 Simulation flood extent for Newcastle.

Figure 4.4 shows the results for comparison with the three “worst case” uFMfSW scenarios. The scenario which fits best to the simulation flood outline in each 1km grid square is shown in the left plot and the MoF value associated with that scenario is given in the right plot. An MoF of 0 indicates that there are no pixels containing flooding in the same location in the two grids being compared and “no correct wet” appears in grid squares where all the comparisons have an MoF of 0. The white squares in the MoF (right) plot correspond to the grid squares in the best fit (left) plot which are labelled “Count for all MoFs is less than 250”. This is where the count of
“correct wet” pixels is less than the threshold and here implies that the count is too low for each of the three comparisons.

![Figure 4.4 Comparison of the G2G-JFlow modelling using observed rainfall (simulation output) with the three “worst case” uFMfSW scenarios for Newcastle. The left map gives the scenario with the best fit and the right map gives the best fit MoF value on a 0 to 1 scale from low to high. The white squares in the right plot correspond to the squares where the “correct wet” count is below the threshold for all scenario comparisons so there is no best fit scenario selected and hence no MoF. The MoF of 0 in the right plot corresponds to the “No correct wet” grid square in the left plot.]

These results show that scenarios with bigger flood extents fit best where there is most flooding (see Figure 2.3). The fit is good for most of the grid (the majority of MoF values are greater than 0.6) and the poorest fit occurs in areas where there is less flooding and hence an increased sensitivity to drainage modelling assumptions. The total MoF for the “best fit” grid is 0.69.

Figure 4.5 shows the results for comparison with all nine uFMfSW scenarios. As above, the uFMfSW scenario which fits best to the simulation flood outline in each 1km grid square is shown in the left plot and the MoF value associated with that scenario is given in the right plot.
Figure 4.5 Comparison of the G2G-JFlow modelling using observed rainfall (simulation output) with the nine uFMsSW scenarios for Newcastle. The left map gives the scenario with the best fit and the right map gives the best fit MoF value on a 0 to 1 scale from low to high. The white squares in the left plot correspond to the squares where the “correct wet” count is below the threshold for all scenario comparisons so there is no best fit scenario selected and hence no MoF. The MoF of 0 in the right plot corresponds to the “No correct wet” grid square in the left plot.

These results show the same pattern as comparison with the three scenarios in that the scenarios with bigger extents fit best where there is most flooding. Again, the fit is good for most of the grid with most grid squares containing MoF values greater than 0.6 and over a third having MoF values greater than 0.8. The poorest fit occurs in areas where there is less flooding and hence an increased sensitivity to drainage modelling assumptions. The MoF values here are generally higher than for the three scenarios and this is reflected in the total MoF for the “best fit” grid, which is 0.78 for the nine scenarios. This is a result of having more degrees of freedom. Therefore, fitting the simulation flood extent to nine scenarios rather than the three “worst case” scenarios is beneficial, enabling better fits to be obtained.
Canvey Island case study

The simulation flood extent for Canvey Island is shown in Figure 4.6 and it can be seen that there is very little flooding modelled in Canvey Island itself.

![Map of Canvey Island with urban areas highlighted in grey and a selection of towns labelled.](image)

Figure 4.6 Simulation flood extent for Canvey Island with urban areas highlighted in grey and a selection of towns labelled.

Figure 4.7 shows the results for comparison with the three “worst case” scenarios. As with Newcastle, the uFMfSW scenario which fits best to the simulation flood outline in each 1km grid square is shown in the left plot and the MoF value associated with that scenario is given in the right plot. The interpretation of the values is the same as for Newcastle but these also contain grid squares labelled as “Correct dry”, which is where the best fit option is neither the simulation grid nor the uFMfSW scenario containing flooding in that grid square. The MoF label associated with these is −1 to distinguish it from the 0s (“No correct wet”) squares where there is over- or under-prediction in all the scenario comparisons.

There are seven grid squares in the south east corner for which G2G data does not exist (these are shown as “No data” in Figure 2.4). Any results in these grid squares are therefore not valid; these are all “Correct dry” or “No correct wet” squares (the latter if the uFMfSW scenarios contain flooding).
Figure 4.7 Comparison of the G2G-JFlow modelling using observed rainfall (simulation output) with the three “worst case” uFMfSW scenarios for Canvey Island. The left map gives the scenario with the best fit and the right map gives the best fit MoF value on a 0 to 1 scale from low to high. The white squares in the left plot correspond to the squares where the “correct wet” count is below the threshold for all scenario comparisons so there is no best fit scenario selected and hence no MoF. The MoF values of 0 in the right plot correspond to the “No correct wet” grid squares in the left plot and the MoF values of −1 correspond to the “Correct dry” grid squares.

These results show that the 30-year scenario dominates for most of the area, which is logical given that the flood outlines are not particularly large. The highest surface runoff was at the top (see Figure 2.4) and the fit appears to be better in that area. The fit is less good where the flood outlines are smaller. Given that the 30-year scenario is still the best fit even where there is higher runoff, even the 30-year scenario must be over-predicting for the rest of the study area. The total MoF for the “best fit” grid for three scenarios is 0.23. The lack of flooding due to the poorly recorded rainfall totals is likely to be affecting these results, particularly given that there is increased sensitivity to drainage modelling assumptions in less extreme events.

Figure 4.8 shows the results for comparison with all nine uFMfSW scenarios.
Figure 4.8 Comparison of the G2G-JFlow modelling using observed rainfall (simulation output) with the nine uFMfSW scenarios. The left map gives the scenario with the best fit and the right map gives the best fit MoF value on a 0 to 1 scale from low to high. The white squares in the left plot correspond to the squares where the “correct wet” count is below the threshold for all scenario comparisons so there is no best fit scenario selected and hence no MoF. The MoF values of 0 in the right plot correspond to the “No correct wet” grid squares in the left plot and the MoF values of −1 correspond to the “Correct dry” grid squares.

These results show the same pattern as comparison with the three scenarios in that the 30-year scenarios dominate with the fit being best at the top. The MoF values are generally higher than for the three scenarios and this is reflected in the total MoF for the “best fit” grid, which is 0.33 for the nine scenarios. Therefore, the Canvey Island case study also demonstrates the benefits of using nine scenarios, despite the lower quality simulation data.

4.3 Comparison of G2G-JFlow modelling using forecast rainfall with uFMfSW scenarios

4.3.1 Analysis and findings

The ensemble maximum flood outlines produced by running G2G surface runoff driven by forecast rainfall through JFlow were compared to the same uFMfSW outlines as the simulation flood outline (Section 4.2) using the approaches outlined in Section 4.1 and Section 4.1.2. The analysis was completed for the Newcastle case study only.

The ensemble flood outlines for Newcastle are shown in Figure 4.9 for a subsection of the study domain (highlighted in Figure 4.3) and those for Canvey Island are shown in Figure 4.13, Figure 4.14, Figure 4.15 and Figure 4.16. There is clearly a high amount of variation in the flood extents between the ensembles and plotting the ensembles together reveals wide areas of poor agreement between the ensembles (i.e. areas that only show flooding in 1-4 ensemble members). This is due to short duration, high
intensity convective storms being difficult to capture far in advance by forecast rainfall products. In general, most of the ensemble members tend to under-predict the flood extent when comparing with the JFlow simulation driven by observed rain gauges.

Newcastle case study

The ensemble flood extents for Newcastle are shown in Figure 4.9 and it can be seen that only four to six ensembles contain much flooding.

Figure 4.9 Ensemble flood extents for Newcastle.

Figure 4.10 shows the information described in Section 4.1.2, i.e. the maximum ensemble count, its associated scenario and best MoF value. Where the maximum ensemble count is tied between two scenarios, the square is labelled as “No clear winner” and the best MoF value is taken as the highest out of the six highlighted in yellow in Table 4.1.
Figure 4.10 For comparison with three scenarios, the left plot shows the scenario that is most often the best fit across the ensemble members, the middle plot shows the number of ensembles for which that scenario is best fit, and the right plot shows the highest MoF value associated with the scenario in the left plot. The ensemble count of zero corresponds to where the “correct wet” counts are below the threshold for comparisons between every ensemble and scenario.

The middle plot shows the maximum ensemble count associated with any scenario and so indicates the strength of agreement or consistency across ensembles. The scenario associated with that maximum count is shown in the left plot and the maximum MoF for that scenario over the ensembles for which it is best fit is shown in the right plot, which indicates the quality of the best fit.

In general, the 30-year scenario is most often the best fit across the ensembles, which is logical as most of the ensembles do not contain much flooding (see Figure 4.9) so the smallest outline would fit best most often, including locations where the simulation outputs show more flooding. However, the maximum ensemble count varies across the area, with highest agreement generally in the bottom left and in some areas with high runoff and lowest agreement in the top left and bottom right, where the flood outlines from the simulation have smaller extents. High agreement for the 30-year scenario makes sense at the bottom where there should be less flooding as the ensembles with little flooding should have a good fit and the other ensembles may have smaller flood outlines in this area so should also fit well to the 30-year scenario. The grid square with an ID of 366 is an example of very high agreement and here the highest MoF values are from comparisons with the ensembles with most flooding. This indicates that the ensembles with little flooding don’t fit as well, even to the 30-year scenario.

In squares within the storm path (i.e. areas with high runoff), where the 30-year scenario has the maximum ensemble count, results show an agreement of five to eight ensembles. This would be expected as the four to six ensembles with significant flooding are likely to fit better to the 100 or 1,000-year scenarios and the remaining seven or eight ensembles will fit best to the 30-year scenario. This is demonstrated by the grid square with an ID of 209. This also explains why the 100 and 1,000-year scenarios appear so infrequently on the left plot.

Low agreements can indicate that many of the ensembles had no “correct wet” pixels or counts below the threshold when compared with the scenarios. Grid square 22 is an example of this. This is also the reason that squares 54 and 72 (in the storm path)
have the 1,000-year scenario as the one with highest agreement; the “correct wet” count is below the threshold for most of the ensembles which contain little flooding when compared with all the scenarios and therefore there is higher agreement between the ensembles with more flooding, which fit better to the larger return period scenarios.

Consequently, these results are influenced by the threshold. A higher threshold may potentially exclude more of the ensembles with little flooding and so the 30-year scenario may not be the best over the ensembles so frequently.

Grid squares where there is no clear winner could potentially be where some ensembles are excluded due to counts below the threshold and the remaining ensembles each have best fits which are shared between the scenarios rather than the 30-year scenario dominating.

The maximum MoF values are high (greater than 0.6) in general, particularly in the storm path, including areas of high and low agreement and where there is no clear winner between the scenarios.

The same information is shown for comparison with the nine uFMfSW scenarios in Figure 4.11.

![Figure 4.11](image)

**Figure 4.11** For comparison with the nine uFMfSW scenarios, the left plot shows the scenario that is most often the best fit across the ensemble members, the middle plot shows the number of ensembles for which that scenario is best fit, and the right plot shows the highest MoF value associated with the scenario in the left plot. The ensemble count of zero corresponds to where the “correct wet” counts are below the threshold for comparisons between every ensemble and scenario.

Comparison with the nine scenarios has enabled the 100 and 1,000-year scenarios to be picked out more as being most often the best fit across ensembles in the areas with greater surface runoff. This is probably due to having a greater range of scenarios so the best fits are split across the ensembles more without one scenario dominating. This is demonstrated by the ensemble counts generally being lower than in the comparison with three scenarios.

The majority of the grid squares, however, do still show that the 30-year 1-hour scenario is most often the best fit across ensembles although there are bands of the 100 and 1,000-year scenarios in the storm path and these are often the longer
durations. As with the comparison with the three scenarios, the dominance of the 30-year 1-hour scenario is probably due to many ensembles containing little flooding.

Also similarly to the comparison with the three scenarios, the higher agreement towards the bottom of the grid is logical as there is less runoff here so more ensembles should agree with the scenario with the smallest outline (30-year 1-hour).

In the locations where the 100 and 1,000-year scenarios are most often the best fit, an agreement of four to six would be expected given that fewer than half of the ensembles contain a lot of flooding. The results agree with this expectation in general, for example, four ensembles agree that the 1,000-year 6-hour scenario is the best fit for grid square 204 and these are the four ensembles containing the most flooding. For grid square 210, five ensembles agree that the 100-year 3-hour scenario is the best fit and further investigation shows that these are five ensembles containing slightly less extensive flooding.

Overall, the pattern produced in the left plot using the nine scenarios is more logical than using the three scenarios with respect to the observed rainfall pattern although there are still locations which don’t appear to represent the observed rainfall, such as the top right corner, but these could be areas where most of the ensembles are not predicting flooding.

The right plot shows a greater number of lower maximum MoF values than the comparison with the three scenarios although the fits are reasonable in the storm path. There is a similar number in the top band for both the three and nine scenarios.

Across comparisons with both the three and nine scenarios, some spread across the ensembles would be expected given the nature of the event; the flooding was driven by a convective storm which is harder to predict so there is less consistency across the ensembles.

One consideration regarding this approach is that the MoF value shown in the right plot is only the highest out of the ensembles which contribute towards the maximum ensemble count and so there are places where a higher MoF value could be found when comparing with a different scenario that has a lower agreement (e.g. grid square 5). Therefore, it would be potentially useful to select scenarios by using the maximum MoF values as well as highest agreement.

4.4 G2G-JFlow modelling using observed and forecast rainfall

4.4.1 Analysis and findings

The simulation maximum flood outline produced by running G2G surface runoff driven by observed rainfall through JFlow was compared with ensemble maximum flood outlines produced by running G2G surface runoff driven by forecast rainfall through JFlow using the approach outlined in Section 4.1. The analysis was run for both the Newcastle and Canvey Island case studies.

Newcastle case study

Figure 4.12 shows the results for the comparison for Newcastle. The ensemble which fits best to the simulation flood outline in each 1km grid square is shown in the left plot and the MoF value associated with that scenario is given in the right plot. The interpretation of the values is the same as in Section 4.2 and the grid squares labelled
as “Correct dry” are where the best fit option is where neither the simulation grid nor the ensemble grid contain flooding in that grid square.

Figure 4.12 Comparison of the G2G-JFlow modelling using observed rainfall (simulation output) with the G2G-JFlow modelling using forecast rainfall (ensemble outputs) for Newcastle. The left map gives the ensemble with the best fit to the simulation flood outline and the right map gives the best fit MoF value on a 0 to 1 scale from low to high. The white squares in the right plot correspond to the squares where the “correct wet” count is below the threshold for comparison with every ensemble so there is no best fit ensemble selected and hence no MoF. The MoF value of −1 in the right plot corresponds to the “Correct dry” grid square in the left plot.

The six ensembles with the best fits are the ones which contain the most flooding (see Figure 4.9). Ensembles 2, 10 and 11 contain most flooding and these are in the areas where the storm produced greatest runoff.

Ensembles 3 and 12 have smaller flood outlines but still a reasonable amount and they fit fairly well in the top left and bottom right where there is less surface runoff in the simulation.

The MoF values are high with a particularly good fit in the storm path. The poorest fits are in the top left and bottom right corners where there is less flooding so different ensembles are over- and under-predicting. The total MoF is 0.79 so across the ensembles, a good fit is possible.

**Canvey Island case study**

The ensemble flood outlines for each forecast origin are shown in Figure 4.13, Figure 4.14, Figure 4.15 and Figure 4.16. It can be seen that some of the ensembles are missing from the figures and these are ensembles which had so little flooding that any minor flooding was removed when post-processing the JFlow outputs.
Figure 4.13 Ensemble flood extents for Canvey Island for the 19/07/2014 13:00 forecast origin.

Figure 4.14 Ensemble flood extents for Canvey Island for the 19/07/2014 19:00 forecast origin.
Figure 4.15 Ensemble flood extents for Canvey Island for the 20/07/2014 01:00 forecast origin.

Figure 4.16 Ensemble flood extents for Canvey Island for the 20/07/2014 07:00 forecast origin.

Figure 4.17, Figure 4.18, Figure 4.19 and Figure 4.20 show the results for the comparison for Canvey Island. The plots show the same information as for Newcastle except that there are four forecast origins for Canvey Island so there is a plot for the ensembles in each forecast origin. The plots can be interpreted in the same way as the previous example from the Newcastle case study with the addition of “All correct dry” squares, which are where none of the ensembles nor the simulation contain flooding.

As well as indicating that neither grid in the best fit comparison contains flooding in the grid square, “All correct dry” and “Correct dry” squares can also indicate that an ensemble map may not exist in that location as some of the outline extents did not cover the entire study area.
Figure 4.17 Comparison of the G2G-JFlow modelling using observed rainfall (simulation output) with the G2G-JFlow modelling using forecast rainfall (ensemble outputs) for Canvey Island with a forecast origin of 19/07/2014 13:00. The left map gives the ensemble with the best fit to the simulation flood outline and the right map gives the best fit MoF value on a 0 to 1 scale from low to high. The white squares in the right plot correspond to the squares where the “correct wet” count is below the threshold for comparison with every ensemble so there is no best fit ensemble selected and hence no MoF. The MoF value of −1 in the right plot corresponds to the “Correct dry” and “All correct dry” grid squares in the left plot.

Figure 4.17 shows the best fit ensembles and associated MoF values for the 19/07/2014 13:00 forecast origin. The simulation runoff is greatest towards the top of the grid although the total runoff is not extremely high and Figure 4.17 shows that Ensembles 13 and 17 fit best here. These have quite a lot of flooding in this area but it’s not hugely extensive so it makes sense that they fit well. Ensembles 15, 16, 18, 19, 25, 27, 29, 30 and 31 also fit well where they have some flooding but not extensively so. Ensembles don’t generally fit well where they have too much flooding and equally none of the ones with very little flooding are best fits.

Overall, the best fit values are good across the grid and the poorest fits are where there is little flooding. The total MoF for a combination of the best fit ensembles is 0.69.
Figure 4.18 Comparison of the G2G-JFlow modelling using observed rainfall (simulation output) with the G2G-JFlow modelling using forecast rainfall (ensemble outputs) for Canvey Island with a forecast origin of 19/07/2014 19:00. The left map gives the ensemble with the best fit to the simulation flood outline and the right map gives the best fit MoF value on a 0 to 1 scale from low to high. The white squares in the right plot correspond to the squares where the “correct wet” count is below the threshold for comparison with every ensemble so there is no best fit ensemble selected and hence no MoF. The MoF value of $-1$ in the right plot corresponds to the “Correct dry” and “All correct dry” grid squares in the left plot.

Figure 4.18 shows the best fit ensembles and associated MoF values for the 19/07/2014 19:00 forecast origin. Ensembles 35, 41, 45, 47 and 53 are best fits where there is most flooding in the simulation; these are the ensembles with extensive flooding. Ensembles 36, 38, 41, 45, 47, 53 and 57 all tend to be the closest matches to the simulation flood outline where they have some flooding but it is not extensive, implying that they are probably over-predicting in areas where they have extensive flooding but the simulation does not. Hence they don’t fit well in those areas.

The MoF values are generally good although less so at the top where there is more flooding. The total MoF over the best fit ensembles is 0.70.
Figure 4.19  Comparison of the G2G-JFlow modelling using observed rainfall (simulation output) with the G2G-JFlow modelling using forecast rainfall (ensemble outputs) for Canvey Island with a forecast origin of 20/07/2014 01:00. The left map gives the ensemble with the best fit to the simulation flood outline and the right map gives the best fit MoF value on a 0 to 1 scale from low to high. The white squares in the right plot correspond to the squares where the “correct wet” count is below the threshold for comparison with every ensemble so there is no best fit ensemble selected and hence no MoF. The MoF value of −1 in the right plot corresponds to the “Correct dry” and “All correct dry” grid squares in the left plot.

Figure 4.19 shows the best fit ensembles and associated MoF values for the 20/07/2014 01:00 forecast origin. Here, the six ensembles with most flooding have the best fits. Ensemble 78 is the closest match to the simulation flood outline over most of the grid. This is the only ensemble to contain extensive flooding in the middle section so had to fit best but the fit is not very good as it has too much flooding and is over-predicting. The fit is better towards the right where the ensemble has less extensive flooding.

The other five ensembles tend to be the best fit in areas where only one or two of the ensembles actually contain flooding so the fit is not always very good with over-prediction likely in some squares. The total MoF for this forecast origin is 0.37, much lower than the previous forecast origins.
Figure 4.20 Comparison of the G2G-JFlow modelling using observed rainfall (simulation output) with the G2G-JFlow modelling using forecast rainfall (ensemble outputs) for Canvey Island with a forecast origin of 20/07/2014 07:00. The left map gives the ensemble with the best fit to the simulation flood outline and the right map gives the best fit MoF value on a 0 to 1 scale from low to high. The white squares in the right plot correspond to the squares where the “correct wet” count is below the threshold for comparison with every ensemble so there is no best fit ensemble selected and hence no MoF. The MoF value of −1 in the right plot corresponds to the “Correct dry” and “All correct dry” grid squares in the left plot.

Figure 4.20 shows the best fit ensembles and associated MoF values for the 20/07/2014 07:00 forecast origin. Similar to the previous forecast origin, Ensemble 83 is the only ensemble to contain significant flooding in the top right and hence it must be the best fitting ensemble even though the fit is poor, particularly at the top where there is little flooding in the ensemble and more in the simulation. Ensembles 86, 88, 90, 95, 98 and 100 all are best fits where they contain some flooding but not a significant amount and where other ensembles either under-predict or over-predict as the simulation flood outlines are not very large. The total MoF here is 0.48.

Overall, the ensembles are able to match the simulation flood outlines well although the ensembles can be so varied in where they contain flooding that sometimes only one or two ensembles contain flooding a given location (due to the lower predictability of convective events). This can lead to lower MoF values as there is a limited choice of ensembles and often an ensemble containing flooding can over-predict in these locations. The best total MoF of 0.70 is from comparison of the simulation flood outlines with those from the ensembles from the 19/07/2014 19:00 forecast origin so these ensembles are able to match the simulation outlines best at a 1km resolution.

It would be expected that a higher overall MoF may be achieved for a winter storm where the rainfall is more predictable by NWP models and the resultant flood outlines would be more consistent and so there would be a greater choice of members to evaluate best fit.
5 Testing SWF HIM assumption of using the accumulated G2G runoff to identify the most appropriate uFMfSW scenario

To test the effectiveness of the SWF HIM assumption of using the accumulated G2G runoff to identify the most appropriate uFMfSW scenario, outputs using this method were compared to the grids from Section 4.2, which use JFlow modelling to determine the most appropriate uFMfSW scenario (this will be referred to as the G2G-JFlow method). This analysis was carried out for the simulation data for Newcastle only.

The two methods are outlined in Figure 5.1 and the grids being compared are shown as the outputs.

Figure 5.1 Method currently used in the SWF HIM compared with the G2G-JFlow method. The respective outputs are being compared in this section.
5.1 Data

The G2G SWF hazard footprints are based on the same G2G configuration used in the previous analysis and further details can be found in Appendix C.

CEH provided grids showing the highest threshold crossed when comparing maximum surface runoffs with uFMfSW effective rainfall grids. For each forecast period, a sequence of surface runoff totals was calculated for 1, 3 and 6-hour durations. Within each forecast time-window the maximum surface runoffs were stored for each duration and compared to the corresponding 1, 3 or 6-hour uFMfSW effective rainfall grids for the 30, 100 and 1000-year return periods, noting the highest threshold crossed for each pixel. The following forecast periods have been used (these are the whole forecast time-windows for each accumulation duration):

- 1-hour: T+1h to T+24h
- 3-hour: T+3h to T+24h
- 6-hour: T+6h to T+24h

This method of identifying the return periods using the maximum G2G surface runoff grids and runoff exceedance thresholds for assumed critical storm durations of 1, 3 and 6 hours will be referred to as the G2G-only method. The outputs are shown in Figure 5.2.
Figure 5.2 Return periods for assumed critical storm durations of 1, 3 and 6 hours identified using the G2G-only method. The symbology corresponds to that of the G2G-JFlow method as shown in the left map of Figure 4.4.

In addition to these grids, further G2G-JFlow outputs were produced showing the same information as Figure 4.4 but using the nine scenarios separated into the three durations. This resulted in plots for each duration showing which return period fitted the simulation flood extent most closely (Figure 5.3). This will allow comparison with the grids shown in Figure 5.2.
Figure 5.3 Best fit scenarios and associated MoF values for comparison of the simulation flood extent for Newcastle with the uFMfSW scenarios for 1-hour (left), 3-hour (middle) and 6-hour (right) durations. The interpretation of the “No correct wet” and “Count for all MoFs is less than 200” grid squares is the same as previous plots.

Figure 5.3 shows that in the urban areas (most of the grid except the very top, bottom and left-hand side), the pattern in the 1-hour duration plot matches that of the “three scenarios” plot (Figure 4.4) quite well whereas in the more rural areas, the pattern of the 6-hour duration plot is similar to the “three scenarios” plot. This makes sense as the urban areas are dominated by shorter durations due to the 12mm/hr runoff rate. This means that the longer durations may remove more water in urban areas, resulting in a smaller extent.

5.2 Approach

Comparisons were carried out for six combinations of G2G-only and G2G-JFlow outputs. These are as follows:

- G2G-only 1-hour (top left grid in Figure 5.2) with G2G-JFlow three scenarios (left grid in Figure 4.4). This is useful as 1-hour is currently assumed as the Critical Storm Duration (CSD) for the maximum output maps. This is called “G2G-only 1-hour vs G2G-JFlow 3-scenarios” in the error matrices.

- G2G-only 3-hour (top right grid in Figure 5.2) with G2G-JFlow three scenarios (left grid in Figure 4.4). This is called “G2G-only 3-hour vs G2G-JFlow 3-scenarios” in the error matrices.

- G2G-only 6-hour (bottom left grid in Figure 5.2) with G2G-JFlow three scenarios (left grid in Figure 4.4). This is called “G2G-only 6-hour vs G2G-JFlow 3-scenarios” in the error matrices.

- G2G-only 1-hour (top left grid in Figure 5.2) with G2G-JFlow 1-hour (left grid in Figure 5.3). This is called “G2G-only 1-hour vs G2G-JFlow 1-hour” in the error matrices.
- G2G-only 3-hour (top right grid in Figure 5.2) with G2G-JFlow 3-hour (middle grid in Figure 5.3). This is called “G2G-only 3-hour vs G2G-JFlow 3-hour” in the error matrices.

- G2G-only 6-hour (bottom left grid in Figure 5.2) with G2G-JFlow 6-hour (right grid in Figure 5.3). This is called “G2G-only 6-hour vs G2G-JFlow 6-hour” in the error matrices.

In order to assess the similarity of the grids, error matrices were created for each comparison. The grids were given common ID values (0, 30, 100 and 1000) indicating the uFMfSW return period with the best fit to the simulation data for the G2G-JFlow outputs and the threshold crossed for the G2G-only outputs. For the G2G-JFlow grids, the value 0 has been applied to grid squares showing either no correct wet pixels or fewer than the set threshold. The error matrices show the number of grid squares fitting into each category so counts on the diagonal are the number of squares which agree in the G2G-only and G2G-JFlow outputs. An overall accuracy (sum of the diagonal values divided by the total count, which is 400 here) and Kappa statistic (with 95% confidence interval) were also calculated to give an indication of the level of agreement (Yu and Lane, 2006). The Kappa statistic is often cited as a more reliable measure of agreement (Cohen, 1960) as it accounts for agreement by chance. Judgements for the level of agreement indicated by the Kappa statistic are given in Table 5.1 (based on Landis and Koch, 1977).

### Table 5.1 Judgements for level of agreement indicated by the Kappa statistic

<table>
<thead>
<tr>
<th>Kappa statistic</th>
<th>Judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0</td>
<td>No agreement</td>
</tr>
<tr>
<td>0 – 0.2</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>0.2 – 0.4</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>0.4 – 0.6</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>0.6 – 0.8</td>
<td>Substantial agreement</td>
</tr>
<tr>
<td>0.8 – 1.0</td>
<td>Almost perfect agreement</td>
</tr>
</tbody>
</table>

This analysis was carried out in R.

### 5.3 Analysis and findings

The error matrices are provided in Table 5.2, Table 5.3, Table 5.4, Table 5.5, Table 5.6 and Table 5.7 and their associated statistics are given below the error matrices. The G2G-only thresholds are along the top of the matrix and the G2G-JFlow best fit scenarios are down the left-hand side.
Table 5.2 Error matrix for comparison of the G2G-only 1-hour output with the G2G-JFlow three scenarios output.

<table>
<thead>
<tr>
<th>G2G-only 1-hour vs G2G-JFlow 3-scenarios</th>
<th>surfacerunoff_maxtotal_1hr_T+1h_T+24h</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit</td>
<td></td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>102</td>
<td>37</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>13</td>
<td>71</td>
<td>62</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>1</td>
<td>5</td>
<td>93</td>
<td>0</td>
</tr>
<tr>
<td>Column total</td>
<td></td>
<td>130</td>
<td>114</td>
<td>156</td>
<td>0</td>
<td>113</td>
</tr>
</tbody>
</table>

Overall accuracy = 113/400 = 0.28.

Kappa statistic = 0.04 (−0.02, 0.10), which implies a slight agreement.

The Z statistic indicates the significance of Kappa (i.e. whether the agreement between the grids is significantly different from no agreement). Here the Z statistic is 1.30 with a p-value of 0.10, which results in failure to reject the null hypothesis that the extent of agreement is the same as random (Kappa=0) at the 5% level (i.e. not significant).

Table 5.3 Error matrix for comparison of the G2G-only 3-hour output with the G2G-JFlow three scenarios output.

<table>
<thead>
<tr>
<th>G2G-only 3-hour vs G2G-JFlow 3-scenarios</th>
<th>surfacerunoff_maxtotal_3hr_T+3h_T+24h</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit</td>
<td></td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>76</td>
<td>54</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>2</td>
<td>25</td>
<td>117</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>1</td>
<td>1</td>
<td>66</td>
<td>32</td>
</tr>
<tr>
<td>Column total</td>
<td></td>
<td>91</td>
<td>82</td>
<td>193</td>
<td>34</td>
<td>216</td>
</tr>
</tbody>
</table>

Overall accuracy = 216/400 = 0.54.

Kappa statistic = 0.36 (0.30, 0.43), which implies a fair agreement. The Z statistic is 11.73 with a p-value of <2.2e-16, which indicates that the null hypothesis can be rejected at the 5% level (i.e. significant).

Table 5.4 Error matrix for comparison of the G2G-only 6-hour output with the G2G-JFlow three scenarios output.

<table>
<thead>
<tr>
<th>G2G-only 6-hour vs G2G-JFlow 3-scenarios</th>
<th>surfacerunoff_maxtotal_6hr_T+6h_T+24h</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit</td>
<td></td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>105</td>
<td>24</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>6</td>
<td>31</td>
<td>96</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>0</td>
<td>1</td>
<td>59</td>
<td>39</td>
</tr>
<tr>
<td>Column total</td>
<td></td>
<td>126</td>
<td>56</td>
<td>166</td>
<td>52</td>
<td>174</td>
</tr>
</tbody>
</table>

Overall accuracy = 174/400 = 0.44.

Kappa statistic = 0.25 (0.19, 0.32), which implies a fair agreement. The Z statistic is 8.87 with a p-value of <2.2e-16, which indicates significance at the 5% level.
Table 5.5 Error matrix for comparison of the G2G-only 1-hour output with the G2G-JFlow 1-hour output.

<table>
<thead>
<tr>
<th>G2G-only 1-hour vs G2G-JFlow 1-hour</th>
<th>surfacerunoff_maxtotal_1hr_T+1h_T+24h</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit for 1hr duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>70</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>44</td>
<td>99</td>
<td>41</td>
<td>0</td>
<td>184</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>2</td>
<td>8</td>
<td>114</td>
<td>0</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>Column total</td>
<td>130</td>
<td>114</td>
<td>156</td>
<td>0</td>
<td>61</td>
<td>Diagonal total</td>
</tr>
</tbody>
</table>

Overall accuracy = 61/400 = 0.15.

Kappa statistic = –0.12 (–0.17, –0.08), which implies no agreement. The Z statistic is –4.36 with a p-value of 1, which does not indicate significance at the 5% level.

Table 5.6 Error matrix for comparison of the G2G-only 3-hour output with the G2G-JFlow 3-hour output.

<table>
<thead>
<tr>
<th>G2G-only 3-hour vs G2G-JFlow 3-hour</th>
<th>surfacerunoff_maxtotal_3hr_T+3h_T+24h</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit for 3hr duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>70</td>
<td>33</td>
<td>1</td>
<td>0</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>8</td>
<td>46</td>
<td>74</td>
<td>1</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>1</td>
<td>118</td>
<td>33</td>
<td>152</td>
<td></td>
</tr>
<tr>
<td>Column total</td>
<td>91</td>
<td>82</td>
<td>193</td>
<td>34</td>
<td>153</td>
<td>Diagonal total</td>
</tr>
</tbody>
</table>

Overall accuracy = 153/400 = 0.38.

Kappa statistic = 0.18 (0.11, 0.24), which implies slight agreement. The Z statistic is 6.13 with a p-value of 4.28e-10, which indicates significance at the 5% level.

Table 5.7 Error matrix for comparison of the G2G-only 6-hour output with the G2G-JFlow 6-hour output.

<table>
<thead>
<tr>
<th>G2G-only 6-hour vs G2G-JFlow 6-hour</th>
<th>surfacerunoff_maxtotal_6hr_T+6h_T+24h</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit for 6hr duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>99</td>
<td>17</td>
<td>6</td>
<td>0</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>12</td>
<td>36</td>
<td>32</td>
<td>2</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>3</td>
<td>128</td>
<td>50</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>Column total</td>
<td>126</td>
<td>56</td>
<td>166</td>
<td>52</td>
<td>114</td>
<td>Diagonal total</td>
</tr>
</tbody>
</table>

Overall accuracy = 114/400 = 0.29.

Kappa statistic = 0.11 (0.05, 0.16), which implies slight agreement. The Z statistic is 4.34 with a p-value of 7.05e-06, which indicates significance at the 5% level.

The grids and error matrices show that generally there is an underestimate of severity by the G2G-only grids, which is probably partly due to the requirement to exceed the uFMfSW effective rainfall threshold for it to be classified as at least that return period. Although the grids in Figure 4.4 and Figure 5.2 show that the 6-hour duration G2G-only output looks closest to the G2G-JFlow three scenarios due to the increase in the 1,000-year return period squares, the grids that have highest agreement in terms of overall accuracy and Kappa statistic are the G2G-only 3-hour grid and the G2G-JFlow three.
scenarios grid. For this comparison the overall accuracy is 0.54 and the Kappa statistic is 0.36, representing a fair agreement. The statistics for the comparisons with the G2G-JFlow outputs split into the three durations show less agreement than for comparisons with the three scenarios.

5.4 Sensitivity analysis

The G2G-only method will be sensitive to the uFMfSW effective rainfall thresholds that have to be exceeded for the return period classification to change. However, the G2G-JFlow outline matching approach will switch at lower effective rainfall thresholds and so the G2G-only method will be comparatively conservative in its return period estimate.

To investigate this further, CEH produced new G2G-only return period estimates that are less conservative by lowering the effective rainfall (G2G surface runoff) thresholds. Details of how the additional G2G SWF hazard footprints have been generated are provided in Appendix D. The uFMfSW effective rainfall thresholds have been relaxed for each return period by the inclusion of a relaxation factor. The original data essentially had a factor of 100 and outputs using five new factors were provided; the values of the factors are 50, 60, 70, 80 and 90.

The three duration outputs for each factor are shown in Figure 5.4, Figure 5.5, Figure 5.6, Figure 5.7 and Figure 5.8.
Figure 5.4 Return periods for assumed critical storm durations of 1, 3 and 6 hours identified using the G2G-only method with a factor of 90. The symbology corresponds to that of the G2G-JFlow method as shown in the left map of Figure 4.4.
Figure 5.5 Return periods for assumed critical storm durations of 1, 3 and 6 hours identified using the G2G-only method with a factor of 80. The symbology corresponds to that of the G2G-JFlow method as shown in the left map of Figure 4.4.
Figure 5.6 Return periods for assumed critical storm durations of 1, 3 and 6 hours identified using the G2G-only method with a factor of 70. The symbology corresponds to that of the G2G-JFlow method as shown in the left map of Figure 4.4.
Figure 5.7 Return periods for assumed critical storm durations of 1, 3 and 6 hours identified using the G2G-only method with a factor of 60. The symbology corresponds to that of the G2G-JFlow method as shown in the left map of Figure 4.4.
Figure 5.8 Return periods for assumed critical storm durations of 1, 3 and 6 hours identified using the G2G-only method with a factor of 50. The symbology corresponds to that of the G2G-JFlow method as shown in the left map of Figure 4.4.

Visual inspection of these grids suggests that they match the G2G-JFlow outputs better, particularly when applying a factor of 50. Error matrices were created and statistics calculated for the same comparisons as those outlined in Section 5.2 using these new grids rather than the original ones with a factor of 100.

Table 5.8 provides a summary of the statistics with the comparisons with highest overall accuracy, Kappa statistic and Z statistic highlighted for each factor.
Table 5.8 Summary statistics for the six comparisons using the original data and the data adjusted by the factors. The factors are included in brackets in the comparison name in the first column (e.g. a factor of 90 is denoted by F90). The comparisons with the highest agreement are highlighted for each factor.

<table>
<thead>
<tr>
<th>Factor Comparison</th>
<th>Overall Agreement</th>
<th>Kappa statistic</th>
<th>Z statistic</th>
<th>p-value</th>
<th>Judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2G-only 1-hour vs G2G-JFlow 3-scenarios</td>
<td>0.28</td>
<td>0.04</td>
<td>1.30</td>
<td>0.10</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour vs G2G-JFlow 3-scenarios</td>
<td>0.54</td>
<td>0.36</td>
<td>11.73</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 6-hour vs G2G-JFlow 3-scenarios</td>
<td>0.44</td>
<td>0.25</td>
<td>8.87</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 1-hour vs G2G-JFlow 1-hour</td>
<td>0.15</td>
<td>−0.12</td>
<td>−4.36</td>
<td>1.00</td>
<td>No agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour vs G2G-JFlow 3-hour</td>
<td>0.38</td>
<td>0.18</td>
<td>6.13</td>
<td>0.00</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>G2G-only 6-hour vs G2G-JFlow 6-hour</td>
<td>0.29</td>
<td>0.11</td>
<td>4.34</td>
<td>0.00</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>G2G-only 1-hour (F90) vs G2G-JFlow 3-scenarios</td>
<td>0.34</td>
<td>0.09</td>
<td>3.09</td>
<td>0.00</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour (F90) vs G2G-JFlow 3-scenarios</td>
<td>0.59</td>
<td>0.43</td>
<td>14.05</td>
<td>0.00</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>G2G-only 6-hour (F90) vs G2G-JFlow 3-scenarios</td>
<td>0.48</td>
<td>0.30</td>
<td>10.17</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 1-hour (F90) vs G2G-JFlow 1-hour</td>
<td>0.20</td>
<td>−0.09</td>
<td>−2.91</td>
<td>1.00</td>
<td>No agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour (F90) vs G2G-JFlow 3-hour</td>
<td>0.45</td>
<td>0.26</td>
<td>8.89</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 6-hour (F90) vs G2G-JFlow 6-hour</td>
<td>0.36</td>
<td>0.18</td>
<td>6.78</td>
<td>0.00</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>G2G-only 1-hour (F80) vs G2G-JFlow 3-scenarios</td>
<td>0.39</td>
<td>0.14</td>
<td>4.55</td>
<td>0.00</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour (F80) vs G2G-JFlow 3-scenarios</td>
<td>0.64</td>
<td>0.50</td>
<td>16.03</td>
<td>0.00</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>G2G-only 6-hour (F80) vs G2G-JFlow 3-scenarios</td>
<td>0.51</td>
<td>0.33</td>
<td>11.06</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 1-hour (F80) vs G2G-JFlow 1-hour</td>
<td>0.23</td>
<td>−0.06</td>
<td>−2.08</td>
<td>0.98</td>
<td>No agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour (F80) vs G2G-JFlow 3-hour</td>
<td>0.53</td>
<td>0.36</td>
<td>11.99</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 6-hour (F80) vs G2G-JFlow 6-hour</td>
<td>0.44</td>
<td>0.26</td>
<td>9.42</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 1-hour (F70) vs G2G-JFlow 3-scenarios</td>
<td>0.43</td>
<td>0.19</td>
<td>5.69</td>
<td>0.00</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour (F70) vs G2G-JFlow 3-scenarios</td>
<td>0.67</td>
<td>0.54</td>
<td>17.47</td>
<td>0.00</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>G2G-only 6-hour (F70) vs G2G-JFlow 3-scenarios</td>
<td>0.54</td>
<td>0.37</td>
<td>12.21</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 1-hour (F70) vs G2G-JFlow 1-hour</td>
<td>0.28</td>
<td>−0.02</td>
<td>−0.65</td>
<td>0.74</td>
<td>No agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour (F70) vs G2G-JFlow 3-hour</td>
<td>0.62</td>
<td>0.47</td>
<td>15.22</td>
<td>0.00</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>G2G-only 6-hour (F70) vs G2G-JFlow 6-hour</td>
<td>0.51</td>
<td>0.34</td>
<td>11.52</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 1-hour (F60) vs G2G-JFlow 3-scenarios</td>
<td>0.47</td>
<td>0.23</td>
<td>6.98</td>
<td>0.00</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour (F60) vs G2G-JFlow 3-scenarios</td>
<td>0.71</td>
<td>0.59</td>
<td>18.86</td>
<td>0.00</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>G2G-only 6-hour (F60) vs G2G-JFlow 3-scenarios</td>
<td>0.57</td>
<td>0.41</td>
<td>13.49</td>
<td>0.00</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>G2G-only 1-hour (F60) vs G2G-JFlow 1-hour</td>
<td>0.31</td>
<td>0.02</td>
<td>0.47</td>
<td>0.32</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>G2G-only 3-hour (F60) vs G2G-JFlow 3-hour</td>
<td>0.72</td>
<td>0.61</td>
<td>18.86</td>
<td>0.00</td>
<td>Substantial agreement</td>
</tr>
</tbody>
</table>
The statistics reveal that the agreements improve as the threshold is relaxed with the highest agreements generally for the comparisons using the grids where a factor of 50 was applied. The error matrices for the comparisons where a factor of 50 was applied to lower the threshold for the G2G-only method are given in Table 5.9, Table 5.10, Table 5.11, Table 5.12, Table 5.13 and Table 5.14. The factors are included in brackets in the comparison name in the top left box in the matrices with a factor of 50 denoted by F50.

**Table 5.9 Error matrix for comparison of the G2G-only 1-hour output with a factor of 50 applied to lower the threshold with the G2G-JFlow three scenarios output.**

<table>
<thead>
<tr>
<th>G2G-only 1-hour (F50) vs G2G-JFlow 3-scenarios</th>
<th>surfacerunoff_maxtotal_1hr_T+1h_T+24h</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>30</td>
<td>32</td>
<td>97</td>
<td>11</td>
<td>0</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>42</td>
<td>99</td>
<td>3</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>1</td>
<td>89</td>
<td>9</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>Column total</td>
<td></td>
<td>44</td>
<td>144</td>
<td>200</td>
<td>12</td>
<td>215</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td>215/400 = 0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kappa statistic</td>
<td>0.32</td>
<td>0.25, 0.39, which implies fair agreement</td>
<td>The Z statistic is 9.32 with a p-value of &lt;2.2e-16, which indicates significance at the 5% level.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.10 Error matrix for comparison of the G2G-only 3-hour output with a factor of 50 applied to lower the threshold with the G2G-JFlow three scenarios output.**

<table>
<thead>
<tr>
<th>G2G-only 3-hour (F50) vs G2G-JFlow 3-scenarios</th>
<th>surfacerunoff_maxtotal_3hr_T+3h_T+24h</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>21</td>
<td>91</td>
<td>28</td>
<td>0</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>5</td>
<td>83</td>
<td>56</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>96</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>Column total</td>
<td></td>
<td>33</td>
<td>101</td>
<td>114</td>
<td>152</td>
<td>280</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td>280/400 = 0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kappa statistic</td>
<td>0.58</td>
<td>0.51, 0.64, which implies moderate agreement</td>
<td>The Z statistic is 18.10 with a p-value of &lt;2.2e-16, which indicates significance at the 5% level.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.11 Error matrix for comparison of the G2G-only 6-hour output with a factor of 50 applied to lower the threshold with the G2G-JFlow three scenarios output.

<table>
<thead>
<tr>
<th>G2G-only 6-hour (F50) vs G2G-JFlow 3-scenarios</th>
<th>surfacerunoff$_{maxtotal}$<em>6hr$</em>{T+6h}$<em>T$</em>{+24h}$</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit</td>
<td></td>
<td>0</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>39</td>
<td>84</td>
<td>17</td>
<td>0</td>
<td>140</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>2</td>
<td>24</td>
<td>63</td>
<td>57</td>
<td>146</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>76</td>
<td>99</td>
</tr>
<tr>
<td>Column total</td>
<td></td>
<td>51</td>
<td>113</td>
<td>103</td>
<td>133</td>
<td>233</td>
</tr>
</tbody>
</table>

Overall accuracy = 233/400 = 0.58.

Kappa statistic = 0.42 (0.35, 0.49), which implies moderate agreement. The Z statistic is 13.48 with a p-value of <2.2e-16, which indicates significance at the 5% level.

Table 5.12 Error matrix for comparison of the G2G-only 1-hour output with a factor of 50 applied to lower the threshold with the G2G-JFlow 1-hour output.

<table>
<thead>
<tr>
<th>G2G-only 1-hour (F50) vs G2G-JFlow 1-hour</th>
<th>surfacerunoff$_{maxtotal}$<em>1hr$</em>{T+1h}$<em>T$</em>{+24h}$</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit for 1hr duration</td>
<td></td>
<td>0</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>28</td>
<td>50</td>
<td>1</td>
<td>0</td>
<td>77</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>7</td>
<td>89</td>
<td>86</td>
<td>2</td>
<td>184</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>1</td>
<td>1</td>
<td>112</td>
<td>10</td>
<td>124</td>
</tr>
<tr>
<td>Column total</td>
<td></td>
<td>44</td>
<td>144</td>
<td>200</td>
<td>12</td>
<td>156</td>
</tr>
</tbody>
</table>

Overall accuracy = 156/400 = 0.39.

Kappa statistic = 0.11 (0.04, 0.18), which implies slight agreement. The Z statistic is 3.33 with a p-value of 4.29e-04, which indicates significance at the 5% level.

Table 5.13 Error matrix for comparison of the G2G-only 3-hour output with a factor of 50 applied to lower the threshold with the G2G-JFlow 3-hour output.

<table>
<thead>
<tr>
<th>G2G-only 3-hour (F50) vs G2G-JFlow 3-hour</th>
<th>surfacerunoff$_{maxtotal}$<em>3hr$</em>{T+3h}$<em>T$</em>{+24h}$</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit for 3hr duration</td>
<td></td>
<td>0</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>17</td>
<td>80</td>
<td>7</td>
<td>0</td>
<td>104</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>6</td>
<td>16</td>
<td>92</td>
<td>15</td>
<td>129</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>137</td>
<td>152</td>
</tr>
<tr>
<td>Column total</td>
<td></td>
<td>33</td>
<td>101</td>
<td>114</td>
<td>152</td>
<td>319</td>
</tr>
</tbody>
</table>

Overall accuracy = 319/400 = 0.80.

Kappa statistic = 0.71 (0.65, 0.77), which implies substantial agreement. The Z statistic is 21.39 with a p-value of <2.2e-16, which indicates significance at the 5% level.

This is the comparison with the highest overall accuracy and Kappa statistic.
Table 5.14  Error matrix for comparison of the G2G-only 6-hour output with a factor of 50 applied to lower the threshold with G2G-JFlow 6-hour output.

<table>
<thead>
<tr>
<th>G2G-only 6-hour (F50) vs G2G-JFlow 6-hour</th>
<th>surfacerunoff_maxtotal_6hr_T+6h_T+24h</th>
<th>0</th>
<th>30</th>
<th>100</th>
<th>1000</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 scenarios best fit for 6hr duration</td>
<td></td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>38</td>
<td>77</td>
<td>7</td>
<td>0</td>
<td>122</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>3</td>
<td>30</td>
<td>44</td>
<td>5</td>
<td>82</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>0</td>
<td>1</td>
<td>52</td>
<td>128</td>
<td>181</td>
</tr>
</tbody>
</table>

Column total: 51 113 103 133 259

Diagonal total: 10 30 44 5 82

Overall accuracy = 259/400 = 0.65.

Kappa statistic = 0.50 (0.43, 0.57), which implies moderate agreement. The Z statistic is 15.51 with a p-value of <2.2e-16, which indicates significance at the 5% level.

For factors 100, 90, 80 and 70, the highest agreement in terms of overall accuracy and Kappa statistic is between the G2G-only 3-hour output and the G2G-JFlow three scenarios output. For factors 60 and 50, however, the highest agreement is between the G2G-only 3-hour output and the G2G-JFlow 3-hour output so it is beneficial to compare to the G2G-JFlow separate durations, particularly for factor 50. The overall accuracy for this highest agreement when using the factor 50 grids is 0.80 and the Kappa statistic is 0.71, representing a substantial agreement. Therefore, the two approaches can produce similar results if the threshold used by the G2G-only method is relaxed.
6 Conclusions and recommendations for further work

6.1 Conclusions

WP2.5 aimed to increase understanding of the Surface Water Flooding (SWF) hazard footprint assumptions within the Surface Water Flooding Hazard Impact Model (SWF HIM). This work compared the hazard footprints produced using the current SWF HIM “look up” approach with those produced from pseudo real-time hydraulic modelling (Figure 5.1). The aim was to confirm (or not) the suitability of the current approach for deriving SWF hazard footprints without recourse to event-specific flood modelling.

Here, instead of selecting the most appropriate footprint from pre-computed flood mapping, runoff forecasts were used to provide boundary conditions to high resolution 2D hydraulic models. Grid-to-Grid (G2G) surface runoff grids driven by observed rainfall and forecast rainfall were supplied by CEH at 15-minute, 1km x 1km resolution for two case study flood events, Newcastle upon Tyne (28 June 2012) and Canvey Island (20 July 2014). The hydraulic models produced flood outlines associated with these events for both the observed (simulation outputs) and forecast rainfall (ensemble outputs).

The simulation flood outlines were compared with observed flood data, including observed depths, impacted areas and photographs, to validate the linked G2G-JFlow model outputs. The model was found to perform well for Newcastle but less so for Canvey Island due to the effect of a faulty rain gauge on rainfall inputs.

The simulation flood outlines were compared with three scenarios (“worst case” scenarios for the 30, 100 and 1,000-year return periods) and nine scenarios (1, 3 and 6-hour durations for each of the 30, 100 and 1,000-year return periods) from the uFMfSW to determine which scenario matched the simulation outline most closely in each 1km x 1km grid square. The patterns of the best fit scenarios were logical given the distribution of surface runoff; the 1,000-year scenarios (greater flood extents) fitted best in locations with high runoff and the 30-year scenarios (smaller flood extents) fitted best in areas with little runoff. Overall, the best fit scenarios had high measure of fit (MoF) values for Newcastle but the fit was less good across Canvey Island. This could be a result of poorly recorded rainfall for Canvey Island which resulted in little flooding in the simulation flood outlines and fit was generally poorest in areas with little flooding as there is increased sensitivity to drainage modelling assumptions in less extreme events. Both case studies produced a better overall fit (“total MoF”) when comparing the simulation outlines with the nine scenarios than with the three scenarios, a result of having more degrees of freedom. The total MoF for the “best fit” grid for Newcastle when comparing with three scenarios was 0.69 and was 0.78 when comparing with nine scenarios, which is a significant increase. The total MoF for comparison with nine scenarios for Canvey Island was only 0.33, an increase from 0.23 for comparison with three scenarios. Therefore, fitting the simulation flood extent to nine scenarios rather than the three “worst case” scenarios is beneficial, enabling better fits to be obtained.

The simulation flood outlines were also compared with those of the ensemble forecasts to determine which ensemble member matched most closely and again Newcastle showed good fits with at least one ensemble predicting the simulation well in most grid squares (the majority of grid squares had MoF values greater than 0.8). The total MoF
was high (greater than 0.65) for two of the forecast origins for Canvey Island but low for the other two (less than 0.5) but again this was influenced by the poor quality of the observed rainfall data in the simulation. Additionally, the location and extent of flooding in the ensembles vary greatly and sometimes only one or two ensembles contain flooding at a given location, which can lead to lower MoF values as there is a limited choice of representative ensembles. This variability is due to the lower predictability of convective events within the Numerical Weather Prediction models used to provide the rainfall inputs to G2G. Overall, the ensembles are able to match the simulation flood outlines well with Newcastle having a total MoF of 0.79. The highest total MoF for Canvey Island was 0.70 from comparison of the simulation flood outline with those from the ensembles from the 19/07/2014 19:00 forecast origin. It would be expected that a higher MoF could be achieved for a winter storm where the rainfall is more predictable as the ensemble flood outlines would be more consistent and so there would be a greater choice of potentially representative ensembles to enable better fit in many locations (i.e. more degrees of freedom).

The ensemble flood outlines were also compared with the uFMfSW scenarios for Newcastle and these were summarised as the number of ensembles that each scenario fits best, the scenario with the highest ensemble count and the maximum MoF value associated with that scenario over the ensembles for which it is the best fit. The 30-year scenarios often had the highest ensemble count but this was a likely impact of only half the ensembles containing significant flooding as the smallest outlines would fit best most often. For example, for the comparison with the three scenarios, in areas with high runoff, the 100 and 1,000-year scenarios were infrequently selected as the best fit to most ensembles. This is probably due to the four to six ensembles with significant flooding fitting best to the 100 or 1,000-year scenarios and the remaining seven or eight ensembles fitting best to the 30-year scenario. Grid squares where the larger return periods have been selected are generally ones which contain no “correct wet” pixels or counts below the threshold for some of the comparisons. This removes some of the ensembles with little flooding from the ensemble count, allowing higher agreement between ensembles with extensive flooding, which fit better to the larger return period scenarios. Setting a higher threshold may exclude more of the ensembles with little flooding and so the 30-year scenario may not be the best over the ensembles so frequently.

Comparison with the nine scenarios, however, allowed the 100 and 1,000-year scenarios to be picked out more as being most often the best fit across ensembles in areas with greater surface runoff. This is probably due to a greater range of scenarios resulting in the best fits being split across the ensembles rather than one scenario dominating. This is demonstrated by the ensemble counts generally being lower than in the comparison with three scenarios.

Overall, the pattern of scenarios which fit best to most ensembles appears more intuitive for comparison with the nine scenarios than with the three scenarios with respect to the observed rainfall pattern although there are still locations with high runoff where the pattern of scenarios doesn’t appear to represent the observed rainfall, but these could be areas where most of the ensembles are not predicting flooding.

The maximum MoF values are high in general for comparison with both the three and nine scenarios, particularly in the storm path, including locations with high and low agreement and no clear winner (most grid squares contain a value greater than 0.6). Comparison with the nine scenarios does show a few more maximum MoF values in the lowest band (0.0-0.2) than the comparison with the three scenarios but the fits are reasonable in the storm path and there is a similar number in the top band (0.8-1.0) for both the three and nine scenarios.

One consideration regarding this approach is that the maximum MoF value is only the highest out of the ensembles which contribute towards the maximum ensemble count.
and so there are places where a higher MoF value could be found when comparing with a different scenario that has a lower agreement. Therefore, it would be potentially useful to select scenarios by using the maximum MoF values as well as highest agreement. This would also allow a worst case scenario to be identified as the larger return periods would be selected in more locations rather than the 30-year scenarios dominating.

The final part of the analysis compared the “best fit” scenarios from comparing the simulation flood outlines with the uFMfSW scenarios (linked G2G-JFlow approach) for the Newcastle case study with the equivalent identified by CEH using the approach currently implemented within the SWF HIM (G2G-only approach). Error matrices were created to compare the grids and an overall accuracy and Kappa statistic were calculated. These results showed that generally there is an underestimate of severity by the G2G-only grids, which is probably partly due the requirement to exceed the uFMfSW effective rainfall threshold for it to be classified as at least that return period.

The grids that have highest agreement in terms of overall accuracy and Kappa statistic are the G2G-only 3-hour grid and the G2G-JFlow three scenarios grid. For this comparison the overall accuracy was 0.54 and the Kappa statistic was 0.36, representing a fair agreement. The statistics for the comparisons with the G2G-JFlow outputs split into the three durations show less agreement than for comparisons with the three scenarios.

Some sensitivity analysis of uFMfSW thresholds was also carried out by repeating these comparisons with outputs from the G2G-only approach that used relaxed thresholds. Five relaxation factors were used (90, 80, 70, 60 and 50) and these outputs were compared with the original outputs where the factor was effectively 100. Lowering the thresholds did increase the similarity of the grids produced by the two approaches with a factor of 50 producing the highest agreement; this was between the G2G-only 3-hour grid and the G2G-JFlow 3-hour grid. For the lowest two factors it was beneficial to compare the G2G-only outputs with the G2G-JFlow separate duration outputs as they produced higher agreements. The overall accuracy for the highest agreement when using the factor 50 grids was 0.80 and the Kappa statistic was 0.71, representing a substantial agreement. Therefore, the two approaches can produce similar results if the threshold used by the G2G-only method is relaxed.

6.2 Recommendations for further work

The analysis in Section 4.3 comparing the G2G-JFlow modelling using forecast rainfall with the uFMfSW scenarios focused on the scenarios which were most often the best fit across the ensemble members (i.e. the mode). The HIM, however, considers each ensemble member separately, selecting scenarios based on the highest threshold crossed, identifying the impact severity and calculating the likelihood for each severity threshold. Therefore, it would be potentially useful to select scenarios by using the maximum MoF values for each G2G-JFlow ensemble member. This would allow a worst case scenario to be identified rather than excluding the ensembles with significant flooding if only a few ensembles show flooding in a given location. This occurs in some areas with high surface runoff where the 30-year scenarios dominate but there are higher MoF values associated with other scenarios. Using the maximum MoF values would allow further consideration of risk; high impact can give higher risk in the risk matrix approach.

The analysis in Section 5 comparing the outputs from the G2G-only approach with those from the G2G-JFlow approach could also be extended to compare with the scenarios selected using the ensemble flood outlines.
References


Appendix A – Newcastle surface runoff grids provided by CEH

Information provided in an email from Steven Cole to Neil Hunter 24/06/2015.

The grids contain 15 minute G2G surface runoff totals (mm) on a 1km grid. These are national grids with the following Arc header:

NCOLS 540
NROWS 700
XLLCENTER 120500
YLLCENTER 500
CELLSIZE 1000
NODATA_VALUE −9999

**Jun28-29_Simulation_Only.zip**

This contains simulation outputs from G2G using raingauge data as input (transformed into a 1km grid using HyradK, multi-quadric surface fitting technique).

The file names are of the form:

surfacerunoffYYYYMMDDmmmm.asc

where YYYY is year, MM is month, DD is day and mmmm is minutes past midnight and are time-stamped in GMT at the end of the 15 minute time-step. For example: surfacerunoff201206290570.asc is for the 15 minutes ending 09:30 GMT, 29 June 2012.

Data are provided for time-stamps 00:00 28 June 2012 – 23:45 29 June 2012.

**Jun28_0715_EnsForecast.zip**

This contains one folder for each ensemble member (ens1, ..., ens12) for the forecast starting 07:15 GMT 28 Jun 2012 (this should correspond to Fig 6.2 and Table 6.1 of the Phase 1 NHP SWF report).

Within each ensemble folder, the file names are of the form:

surfacerunoff201206280435_mmmm.asc

where mmmm is minutes after the forecast origin. The forecasts are 24 hours long so the last file is surfacerunoff201206280435_1440.asc
Appendix B – Canvey Island surface runoff grids provided by CEH
D2.2a SWF Hazard Footprints hindcasts using existing G2G case study outputs for JBA

1. Background

SWF Hazard Footprints from G2G are required by JBA to support the WP 2.5 case study, linking G2G surface runoffs with JFlow+ to understand the SWF Hazard Footprint assumptions within the SWF HIM.

An initial delivery of G2G outputs has previously been provided to JBA (24 June 2015) for the Newcastle 28 June 2012 case study. This consisted of ensemble outputs for one forecast origin: 07:15 GMT 28 June 2012 and a period of simulation only output that uses raingauge data as input. It should be noted that the initial delivery used the STEPS-2 with UK4 rainfall forecasts (12 ensemble members) from the Phase 1 prototype work.

This delivery provides G2G outputs for the Canvey Island 20 July 2014 case study. It uses the latest STEPS-2 with MOGREPS-UK Blended Ensembles as provided by the Met Office (a description of these forecasts is on Huddle here). For completeness, G2G outputs using STEPS-2 with MOGREPS-UK Blended Ensembles have also been delivered for the 2012 Newcastle case study, although there is no expectation of JBA using these unless time is available.

2. Data Format

The G2G SWF Hazard Footprints are based on the same G2G configuration used within the Phase 1 prototype work but using the latest STEPS-2 with MOGREPS-UK Blended Ensembles provided by the Met Office within Phase 2. Gridded raingauge data are used for input for simulation runs, and for initialising the G2G model up to the forecast origin. A G2G Surface Runoff forecast was produced for each of the 24-member Blended Ensemble rainfall forecasts out to T+32h. The G2G Surface Runoff forecasts have a model time-step of 15 minutes.

The G2G outputs have been provided in a standard ArcASCII (.asc) format, as per previous deliveries.

**Ensemble forecasts**

G2G outputs for each ensemble member are in a folder named `ensX` where X ranges from 1 to 24 for the 2012 case study and 0 to 23 for the 2014 case study. There is one `*.asc` file for each gridded image. The files use the following naming convention:

```
surfacerrunoffyyyymmddmmmm_ffff.asc
```

`yyyymmddmmmm` is the time and date of the forecast origin – note `mmmm` is minutes past midnight

`ffff` is the forecast lead-time in minutes from 0015 to 1920 (T+32h)

For example:

The file `ens2/surfacerrunoff201407200060_0135.asc` refers to a lead-time of 2h 15m (135 mins) for the forecast starting at 01:00 20 July 2014 so has a validity time of 03:15 20 July 2014.

Details of the forecasts provided are given in the following Table.
### Available forecasts

<table>
<thead>
<tr>
<th>Case study</th>
<th>No.</th>
<th>First</th>
<th>Last</th>
<th>zip file</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 June 2012</td>
<td>2</td>
<td>01:00</td>
<td>07:00</td>
<td>for_JBA_20120628_EnsForecast.zip</td>
</tr>
<tr>
<td></td>
<td></td>
<td>28 Jun 2012</td>
<td>28 Jun 2012</td>
<td>522Mb</td>
</tr>
<tr>
<td>20 July 2014</td>
<td>10</td>
<td>13:00</td>
<td>19:00</td>
<td>for_JBA_20140718_EnsForecast.zip</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18 Jul 2014</td>
<td>20 Jul 2014</td>
<td>1.9Gb</td>
</tr>
</tbody>
</table>

It should be noted that the MOGREPS-UK based STEPS-2 Blended Ensemble rainfall forecasts have different forecast origins (01:00, 07:00, 13:00, 19:00) compared to the UK4 based STEPS-2 Blended Ensembles.

For the previous 2012 Newcastle study, one forecast origin was used. For the Canvey Island 2014 case study, 10 forecast origins have been provided but it is likely that focussing effort on only one or two forecasts would be beneficial.

A hyetograph for the Easthaven Tide Barrier raingauge is given below suggesting the main rainfall fell between 14:00 and 16:00 on 20 July 2014. The forecasts 13:00 19 July through to 07:00 20 July would be the most interesting to run.

![Hyetograph Easthaven Tide Barrier](image)

### Simulation runs

The simulation-mode outputs using raingauge data as input have previously been provided for the 2012 Newcastle case study. They are provided here for the 2014 case study in the zip file for JBA_20140718_Simulation_Only.zip.

There is one *.asc file for each gridded image. The files use the following naming convention:

```
surfacerunoffyyyyymmddmmmm.bil
```

`yyyyymmddmmmm` is the validity time and date of the image.

For example:

The file `grids/surfacerunoff201407200210.asc` refers to a validity time of 03:30 20 July 2014 for the image.

These have been provided in the zip file (22Mb) for JBA_20140718_Simulation_Only.zip and cover the period 00:00 18 July 2014 to 11:45 22 July 2014.

Centre for Ecology & Hydrology, Wallingford  
Version 1.0, 4 February 2016
Appendix C – G2G SWF hazard footprints
1. Background

SWF Hazard Footprints from G2G are required by HSL to support the WP 3.2b on Sensitivity Testing. Following discussions between CEH and HSL, various outputs were agreed on that would allow investigation of using shorter summary time-steps down to one-hour and 9 versus 3 uFMfSW scenarios, as requested by FFC.

This note summarises the data supplied by CEH to HSL.

2. Data Format

The G2G SWF Hazard Footprints are based on the same G2G and forecast data configuration used within the Phase 1 prototype work. Gridded raingauge data are used for input for simulation runs, and for initialising the G2G model up to the forecast origin. A G2G Surface Runoff forecast was produced for each of the 12 member Blended Ensemble (STEPS-2 with UK4) rainfall forecasts. The G2G Surface Runoff forecasts have a model time-step of 15 minutes.

For each ensemble member, a sequence of surface runoff totals were calculated for 1, 3 and 6 hour durations. Within each forecast time-window the maximum surface runoffs were stored for each duration and compared to the corresponding 1-, 3- or 6-hr uFMfSW effective rainfall grids for 30, 100 and 1000 year return periods, noting the highest threshold crossed for each pixel.

For each of the ensemble members, and a simulation run using raingauge data as input, files using the following naming convention are provided to HSL:

```
surferunoff_maxtotal_nh_T+ah_T+bh_yyyymmdmmmm_RP.asc
```

- $n$ is the accumulation duration (1, 3 or 6 hours)
- $a$ is the start of the forecast time-window being considered
- $b$ is the end of the forecast time-window being considered
- $yyymmdmmmm$ is the time and date of the forecast origin – note $mmmm$ is minutes past midnight

For example: `surferunoff_maxtotal_3h_T+3h_T+6h_201206280435_RP.asc`

The following forecast time-windows have been used:

<table>
<thead>
<tr>
<th>Duration (n)</th>
<th>Hourly windows</th>
<th>Periodic windows</th>
<th>Whole forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-hr</td>
<td>T+1h to T+2h, ..., T+23h to T+24h</td>
<td>T+1h to T+6h, T+6h to T+12h, T+12h to T+24h</td>
<td>T+1h to T+24h</td>
</tr>
<tr>
<td>3-hr</td>
<td>T+3h to T+4h, ..., T+23h to T+24h</td>
<td>T+3h to T+6h, T+6h to T+12h, T+12h to T+24h</td>
<td>T+3h to T+24h</td>
</tr>
<tr>
<td>6-hr</td>
<td>T+6h to T+7h, ..., T+23h to T+24h</td>
<td>T+6h to T+12h, T+12h to T+24h</td>
<td>T+6h to T+24h</td>
</tr>
</tbody>
</table>
The dataset comprises of values with the following meanings:

<table>
<thead>
<tr>
<th></th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>no threshold crossed</td>
</tr>
<tr>
<td>1</td>
<td>1 in 30 year threshold crossed</td>
</tr>
<tr>
<td>2</td>
<td>1 in 100 year threshold crossed</td>
</tr>
<tr>
<td>3</td>
<td>1 in 1000 year threshold crossed</td>
</tr>
</tbody>
</table>

Each ensemble member output is in a folder named ensX where X ranges from 1 to 12. The simulation output is in folder sim.

3. Forecasts included

It was agreed that the focus of the initial testing would be during the Newcastle, 28 June 2012 case study. Outputs for one forecast origin have been provided: 07:15 GMT 28 June 2012.
Appendix D – G2G SWF hazard footprints for sensitivity analysis
Natural Hazards Partnership
Surface Water Flooding Hazard Impact Model (NHP SWF HIM): Phase 2

D2.5 Linking G2G to JFlow+ inundation model: SWF Hazard Footprint Case Study
Sensitivity analysis of uFMfSW thresholds

1. Background

JBA and CEH have been using the 28 June 2012 Newcastle flood event to test the NHP SWF HIM approach. G2G surface runoff outputs have been used as effective rainfall input into JFlow+ and maximum flood outlines produced using the same uFMfSW procedure. These outlines have then been compared to various uFMfSW outlines and “Measure of Fit” (MoF) statistics used to determine which uFMfSW outline gives the closest fit. An example is shown in Figure 1 identifying the best match with one of the three “maximum output” flood extents (for 30, 100 and 100 year return periods) from the uFMfSW.

Figure 1 Best match of maximum flood outlines produced by G2G surface runoff input to JFlow+ with three “maximum output” uFMfSW outlines. Left map gives best fit in terms of return period identified (30, 100 or 1000 years) and right map gives the measure of fit on a 0 to 1 scale from low to high.

The map of return period for the best fitting “maximum output” uFMfSW flood outline on the left can be compared to the return periods identified by looking at the maximum G2G surface runoff grids using runoff exceedance thresholds for assumed critical storm durations of 1, 3 and 6 hours (Figure 2). The latter will be referred to as the G2G-only method.
Figure 2 Return periods for assumed critical storm durations of 1, 3 and 6 hours identified using the G2G-only method. Colour classes of return period as in left map of Figure 1 (pink 30-yr, green 100-yr, blue 1000-yr).

It is recognised that the G2G-only method will be sensitive to the uFMfSW effective rainfall thresholds that have to be exceeded for the return period classification to change. However, the G2G and JFlow+ outline matching approach will switch at lower effective rainfall thresholds and so the G2G-only method will be comparatively conservative in its return period estimate.

To investigate this further, CEH have produced new G2G-only return period estimates that are less conservative by lowering the effective rainfall (G2G surface runoff) thresholds. This note describes that approach and the additional data sent to JBA.

2. Data Format

The G2G SWF Hazard Footprints are based on the same G2G configuration used within the Phase 1 prototype work and previous Hazard Footprints provided to HSL and JBA. Gridded raingauge data are used as input for simulation runs of G2G, and for initialising the G2G model up to the forecast origin. As with previous deliveries, the outputs are provided for forecast periods. The G2G Surface Runoff forecasts have a model time-step of 15 minutes.

For each forecast period, a sequence of surface runoff totals were calculated for 1, 3 and 6 hour durations. Within each forecast time-window the maximum surface runoffs were stored for each duration and compared to the corresponding 1-, 3- or 6-hr uFMfSW effective rainfall...
grids for 30, 100 and 1000 year return periods, noting the highest threshold crossed for each pixel.

The following forecast time-windows have been used:

<table>
<thead>
<tr>
<th>Duration, n</th>
<th>Whole forecast, T+ah to T+bh</th>
</tr>
</thead>
<tbody>
<tr>
<td>1h</td>
<td>T+1h to T+24h</td>
</tr>
<tr>
<td>3h</td>
<td>T+3h to T+24h</td>
</tr>
<tr>
<td>6h</td>
<td>T+6h to T+24h</td>
</tr>
</tbody>
</table>

The dataset comprises of values with the following meanings:

0  no threshold crossed
1  “1 in 30 year” threshold crossed
2  “1 in 100 year” threshold crossed
3  “1 in 1000 year” threshold crossed

Outputs for only one forecast origin are provided, 07:00 28 June 2012, as this covers the event.

3. New G2G SWF Hazard Footprints for sensitivity analysis

The above describes the data first used by JBA and has been redelivered here in folder 100.

Additional G2G SWF Hazard Footprints have been generated by relaxing the uFMfSW effective rainfall scenarios for a given duration as follows:

\[
new_{-}RP_i = RP_{i-1} + (RP_i - RP_{i-1}) \frac{\text{factor}}{100}
\]

where \( new_{-}RP \) is the new return period threshold value (of effective rainfall in mm), \( RP \) is the existing return period threshold value, and \( \text{factor} \) is the relaxation factor that lowers the threshold values. Here the suffix \( i \) ranges from 1 to 3 on the left hand side representing the 30, 100, and 1000 year return periods and \( RP_0 \) is taken to be 0mm.

Outputs have been grouped in folders according to the factor with values of 50, 60, 70, 80 and 90 produced. The filenames also now have a slightly updated convention:

surfacerunoff_maxtotal_nh_T+ah_T+bh_yyyyymmddmmmm_newRP_ff.asc

\( n \) is the accumulation duration (1, 3 or 6 hours)
\( a \) is the start of the forecast time-window being considered
\( b \) is the end of the forecast time-window being considered
\( yyyyymmddmmmm \) is the time and date of the forecast origin – note \( mmmmm \) is minutes past midnight
\( ff \) is the \( \text{factor} \)

Example: surfacerunoff_maxtotal_1h_T+1h_T+24h_201206280420_newRP_70.asc
Also included are png files displaying the three duration outputs for each factor. Not all factors need to be used in the JBA analysis but at least the 50 and 80 ones would be useful. Visual inspection suggests these match the G2G plus JFlow+ outputs better.

Centre for Ecology & Hydrology
Wallingford
Version 1.0, 14 June 2016
Offices at
Coleshill
Doncaster
Dublin
Edinburgh
Exeter
Glasgow
Haywards Heath
Isle of Man
Limerick
Newcastle upon Tyne
Newport
Peterborough
Saltaire
Skipton
Tadcaster
Thirsk
Wallingford
Warrington

Registered Office
South Barn
Broughton Hall
SKIPTON
North Yorkshire
BD23 3AE
United Kingdom

Jeremy Benn Associates Ltd
Registered in England
3246693

Visit our website
www.jbaconsulting.com