Journal of Hydrology 548 (2017) 801–817

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Demonstrating the value of community-based ('citizen science') observations for catchment modelling and characterisation

Eleanor Starkey [a,*], Geoff Parkin [a], Stephen Birkinshaw [a], Andy Large [b], Paul Quinn [a], Ceri Gibson [c,1]

[a] School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK
[b] School of Geography, Politics and Sociology, Newcastle University, Newcastle upon Tyne NE1 7RU, UK
[c] Tyne Rivers Trust, Unit 8, Shawwell Business Centre, Corbridge, Northumberland NE45 5PE, UK

ARTICLE INFO

Article history:
Received 13 August 2016
Received in revised form 30 November 2016
Accepted 10 March 2017
Available online 18 March 2017
This manuscript was handled by G. Syme, Editor-in-Chief, with the assistance of M Smith, Associate Editor

Keywords:
Community-based
Citizen science
Monitoring
Flash flood
Hydrological modelling
SHETRAN

ABSTRACT

Despite there being well-established meteorological and hydrometric monitoring networks in the UK, many smaller catchments remain ungauged. This leaves a challenge for characterisation, modelling, forecasting and management activities. Here we demonstrate the value of community-based ('citizen science') observations for modelling and understanding catchment response as a contribution to catchment science. The scheme implemented within the 42 km^2 Haltwhistle Burn catchment, a tributary of the River Tyne in northeast England, has harvested and used quantitative and qualitative observations from the public in a novel way to effectively capture spatial and temporal river response. Community-based rainfall, river level and flood observations have been successfully collected and quality-checked, and used to build and run a physically-based, spatially-distributed catchment model, SHETRAN. Model performance using different combinations of observations is tested against traditionally-derived hydrographs. Our results show how the local network of community-based observations alongside traditional sources of hydro-information supports characterisation of catchment response more accurately than using traditional observations alone over both spatial and temporal scales. We demonstrate that these community-derived datasets are most valuable during local flash flood events, particularly towards peak discharge. This information is often missed or poorly represented by ground-based gauges, or significantly underestimated by rainfall radar, as this study clearly demonstrates. While community-based observations are less valuable during prolonged and widespread floods, or over longer hydrological periods of interest, they can still ground-truth existing traditional sources of catchment data to increase confidence during characterisation and management activities. Involvement of the public in data collection activities also encourages wider community engagement, and provides important information for catchment management.

1. Introduction

Under future climate change scenarios, wetter winters and more intense summer storms are expected to exacerbate already complex catchment management issues throughout the UK and western Europe (Chan et al., 2015; Forzieri et al., 2016; Kendon et al., 2014). Empirical data is therefore required to characterise catchment behaviour over time, model floods, improve forecasts and subsequently enhance community resilience as part of the wider catchment management process. The importance of meaningful data is further emphasised when considering the performance of new flood management interventions such as 'natural flood management' (Nicholson et al., 2012; SEPA, 2015). The potential benefits of engaging, collaborating and actively involving local communities within affected catchments is also rapidly being

Abbreviations: AE, Actual Evaporation; AWS, Automatic Weather Station; BADC, British Atmospheric Data Centre; CB, Caw Burn; HB, Haltwhistle Burn; Ks, Saturated Hydraulic Conductivity; NSE, Nash-Sutcliffe Efficiency; PGB, Pont Gallon Burn; P, Precipitation; PBIAS, Percentage Bias; PBSD, Physically-based spatially-distributed; PE, Potential Evapotranspiration; Q, Discharge; Q*obs*, Observed Discharge; Q*sim*, Simulated Discharge; RLGB, River Level Gauge Board; RMSE, Root Mean Square Error; R^2, Coefficient of Determination; SD, Soil Depth; SOF, Strickler Overland Flow; TRT, Tyne Rivers Trust.

* Corresponding author.
E-mail address: eleanor.starkey@ncl.ac.uk (E. Starkey).
[1] Present address: The Freshwater Biological Association, The Ferry Landing, Far Sawrey, Ambleside, Cumbria LA22 OLP, UK.

http://dx.doi.org/10.1016/j.jhydrol.2017.03.019
0022-1694/© 2017 The Authors. Published by Elsevier B.V.

recognised as a vital component of an integrated catchment management toolkit (Bracken et al., 2014; Large et al., 2017).

Despite the UK having some of the world's most reliable and dense hydrometric and meteorological monitoring networks, data remains scarce for many rural catchments (Buytaert et al., 2016; Illingworth et al., 2014; UK Met Office, 2010). A variety of methods are used for observing and/or estimating spatial rainfall patterns (Bárdossy and Pegram, 2013; Durkee, 2010; Lanza et al., 2001; Shaw et al., 2011) but data availability and accuracy issues still persist on a local level. There are a number of issues; catchments are spatially and temporally complex, and flash floods, while of particular interest and importance to both hydrologists and communities, are hard to characterise given that they are rare, spatially localised, short lived and often occur in locations without formal monitoring (Archer and Fowler, 2015; Archer et al., 2016; Perks et al., 2016).

The absence of whole-catchment data can complicate the catchment modelling process (Seibert and McDonnell, 2015), especially when attempting to replicate or predict extreme events in unique locations. While workers like Zhu et al. (2013, 2014) describe how rainfall radar observations are becoming more readily available, providing improved spatial and temporal coverage in hydrological models, errors relating to timing and magnitude can propagate through the modelling process (Harrison et al., 2000). Good quality and detailed ground-based observations are therefore required to create robust models (Beven, 2009; Beven and Westerberg, 2011; Vidon, 2015). Through incorporation of such observations, the improved predictive power of the model will then play a significant role in influencing choices made by stakeholders in the catchment characterisation and management process.

The co-production of 'indigenous' knowledge and the activity of community-based monitoring (and related activities described in the literature using a range of terminology including citizen science, volunteered geographical information (VGI), crowd-sourcing, citizen observatory and participatory monitoring) is rapidly expanding (Follett and Strezov, 2015; Pocock et al., 2014; Wentworth, 2014). The term used depends on the degree of 'volunteer' involvement and the specific techniques adopted, but in general they all refer to the participation of the public (i.e. non-professionals) in the generation of new knowledge about the natural environment (Buytaert et al., 2014; Pocock et al., 2014; Starkey and Parkin, 2015). Regardless of which term is used, encouraging general engagement, participation and empowerment on a local level means that the public have the potential to offer timely and low-cost solutions to the data collection phase in catchment science. Social benefits to the community are also valuable, supporting policies and management frameworks which increasingly request an integrated and bottom-up approach to catchment management. A relevant example includes the emerging 'Catchment Based Approach' (CaBA, 2016) which has surfaced from the EU Water Framework Directive and is managed in the UK by Defra, the Department of Environment, Food and Rural Affairs.

The growth in more readily available and low-cost technologies, such as smartphones, social media and the internet itself, is allowing community-based initiatives to grow rapidly. Areas include biodiversity (Sutherland et al., 2015), weather and climate (Burakowski et al., 2013; Muller et al., 2015) and disaster management (Aulov and Halem, 2012). Across North America the public are collecting regular rain, hail and snow observations and sharing them with the national CoCoRaHS network (http://www.cocorahs.org/), and a similar scheme is also active primarily across Europe, North America and Australia through the UK Met Office 'Weather Observations Website' (http://wow.metoffice.gov.uk/).

It is only recently that this type of data collection activity has started to flourish in hydrology and hydrogeology, for example, in Ethiopia (Walker et al., 2016). Only a few examples exist in the UK which specifically collect river and flood observations with some form of public involvement, for instance the Wesenseit (http://wesenseit.eu/) and Oxford Flood Network (http://flood.network/). Even fewer studies have explored the potential value of this data to support real hydrological applications, including catchment modelling, primarily due to data quality concerns or general lack of recognition (Buytaert et al., 2014, 2016; Muller et al., 2015). Only a small number of studies have made use of crowd-sourced data to validate their models, but they frequently discarded multiple observations as location, date and time stamps were absent (Fohringer et al., 2015; Kutija et al., 2014; Mazzoleni et al., 2015; Smith et al., 2015). In addition, these studies either involved 'reactive' data collection methodologies following large floods or used synthetic data to imitate citizen science, thus did not actually involve or even engage with the public. Full engagement is essential if ongoing community-based monitoring schemes are to be relied upon by professionals and regularly harnessed as an additional source of catchment information. Nevertheless, scientists and engineers are still generally reluctant to integrate this type of data into their work, which Barthel et al. (2016) attributes to professionals not being experienced enough to actually carry out the full range of participatory activities required. This includes engagement, facilitation, training and dissemination activities which are all prerequisites of successful community-based monitoring schemes.

This paper presents results from a catchment study which demonstrates the value of community-based observations for understanding and modelling spatial and temporal catchment response, including the ability to capture the shape, timing and magnitude of flood peaks for a sequence of flash flood events. Data quality issues are a particular concern with 'citizen science' studies and we take this into account by applying appropriate data quality checks before allowing further use of the data in the modelling process. The modelling results presented also infer additional information about the quality of the observations used. Walker et al. (2016) concluded that data quality from community-based observations can be of high quality if they are properly managed. Our study takes this approach a step further as it is one of the first assessments which embeds real community-based observations into a detailed catchment modelling study. To achieve this, work has been carried out on the Haltwhistle Burn catchment, a tributary of the River Tyne in northeast England, where a physically-based, spatially-distributed hydrological catchment model, SHETRAN (Ewen et al., 2000), has been used. The findings will be of interest to catchment managers, hydrologists, as well as community and environmental groups who have a common interest in holistic catchment management and who wish to expand their management toolkits.

2. Study area & focus community

Known for being located in the 'Centre of Britain', the 42 km^2 steep and low stream order Haltwhistle Burn catchment responds rapidly to heavy rainfall. This predominantly rural catchment suffers from multiple pressures (Fig. 1) and in recent years it has experienced a number of floods, including 2007, 2012, 2014 and winter 2015/2016. Flood risk is exacerbated as the main impact zone (the town of Haltwhistle) is located at a 'pinch-point' close to the outlet, and just downstream of an incised gorge section. The elongated shape of the catchment and resulting river network have also been influenced by the igneous Whin Sill outcrop which intersects this area.

Rivers Trusts exist across the UK and aim to enhance their local river basin with the help of volunteers and communities through their charitable objectives. Tyne Rivers Trust (TRT) led an

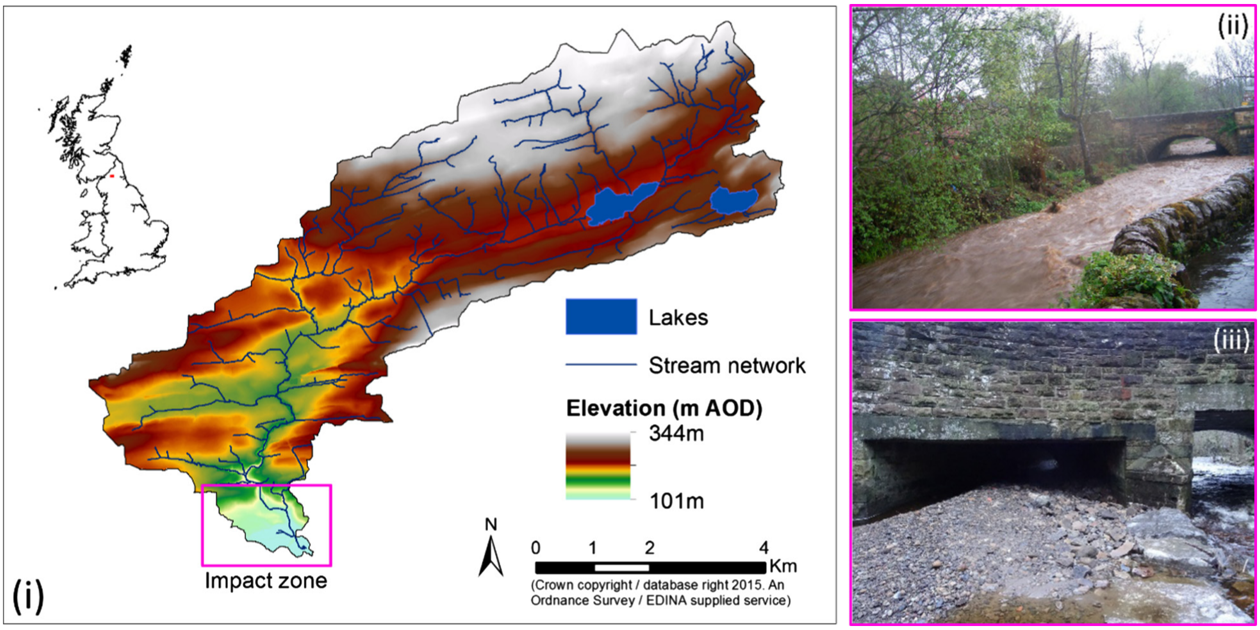


Fig. 1. (i) Location and elevation map of the Haltwhistle Burn catchment, (ii) Haltwhistle Burn at high flow and (iii) Sediment deposited under a culvert in the town following high intensity rainfall. Photographs have been provided by members of the community.

ambitious multi-partnership restoration project from 2012 to 2015 with the aim of improving the health of the Haltwhistle Burn and its tributaries, using community engagement from the onset (Tyne Rivers Trust, 2015). Although the project focused around headwater runoff and pollution, flooding was also included as an objective given that these issues are closely aligned. While TRT required evidence to characterise the catchment and assist with designing and implementing a suite of catchment management measures, no monitoring stations operated within the catchment before the project started.

The Haltwhistle Burn catchment and the already engaged 'Haltwhistle Burn River Watch Group' offered a good case study site and focus community to trial a community-based monitoring and modelling approach. Although findings are location- and community-specific, this case study site has numerous characteristics and issues which are common to many rural UK catchments. We therefore designed, implemented and facilitated a low-cost community-based monitoring programme within the catchment to support TRT's existing restoration project (Large et al., 2017 in press), to further understand flash flooding and to allow appropriate alleviation measures to be designed and implemented.

3. Methodology

3.1. Overview

The value of quantitative and qualitative observations collected by the local community have been demonstrated here by using the data alongside a traditional monitoring network to build and run a physically-based, spatially-distributed (PBSD) catchment model, SHETRAN. The community-based data includes rainfall, river level and flood observations, all of which have been used to extract timing and magnitude information for the April 2014 high intensity rainfall event which occurred in the catchment. The modelling framework involved calibrating, validating and accepting a 'baseline' model which consists of rainfall data integrated from the best available gauge combination (in this case, both community-based and traditional ground-based gauges). While keeping all other model settings and datasets the same, a 'leave-one-out' methodology allowed the effect of different combinations of these rainfall observations to be tested. All modelled outputs were statistically and visually compared with traditionally-derived hydrographs, as well as to each other. These community-based observations were also compared with the same SHETRAN model using UK Met Office rainfall radar observations over the same period.

3.2. Community-based monitoring

Participatory projects involving members of the public contain a number of stages, from initial engagement activities through to feedback and ongoing facilitation. Fig. 2 summarises the stages involved in initiating the community-based monitoring network in Haltwhistle. Key guidance documents such as those produced by Pocock et al. (2014), Science Communication Unit (2013) and Tweddle et al. (2012) were consulted for best practice during this process.

Using TRT as a 'gatekeeper', an initial workshop was held by the research team, inviting the already established River Watch Group, as well as key partners in the wider community (land owners and residents). Other engagement techniques were adopted, including social media (@HaltwhistleBurn), local newspaper articles, the project website (http://research.ncl.ac.uk/haltwhistleburn/) and leafleting. Many authors, including Tweddle et al. (2012) have argued that ongoing feedback is essential. The project website therefore acted as an ongoing community-hub and toolkit, where information and observations could be hosted.

Following these initial (but vital) engagement activities, a variety of simple low-cost citizen science style monitoring and data submission tools were sourced or developed for use. Maximising participation levels and ensuring relevant and meaningful parameters were recorded was at the forefront of the design process. Unlike many projects which strap micro-sensors to volunteers or their belongings (e.g. Castell et al., 2015; Hut et al., 2014), activities were designed here to encourage long-term monitoring beyond the lifetime of the project and for citizen scientists to physically observe and learn about their weather and water environment themselves, rather than simply distributing automatic sensors. In order to maximise the usefulness of observations and improve

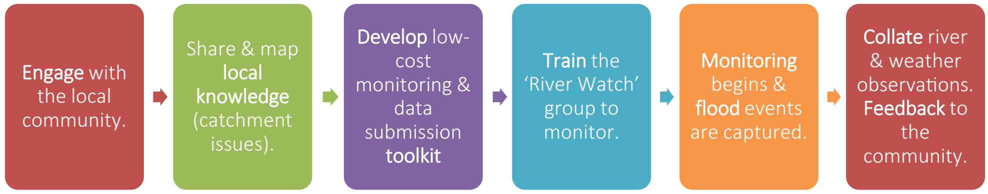


Fig. 2. Key stages involved during the community-based monitoring process to capture observations ready for the modelling activities.

their quality, a 'pro-active' monitoring approach was adopted. This involved training participants in advance so that they were confident to participate and collect good quality observations relevant to the management process. It also meant that they knew what to look out for both during and immediately after flash floods. Laminated training cards were created to ensure this awareness, and also to standardise monitoring methods (see examples in the Supplementary Material).

Although a wide range of monitoring activities were trialled, efforts ultimately focussed on rainfall, river levels and flood-related evidence (Table 1). These were the most popular and frequently observed parameters across the full monitoring period of October 2013 to February 2016. Depending on user preference, web forms, Excel spreadsheets and email, paper and face-to-face meetings, Twitter and an Android '*River and Weather*' app developed in-house were all used by volunteers to submit observations.

Once observations had been submitted and shared, datasets were anonymised and databases created. In many cases, the observations were either photographs or videos (river levels and flood information) which were named and ordered by date and time. A large quantity of flood observations obtained from multiple members of the community during the events of interest were analysed; they were generally found to be self-consistent, confirming their validity as evidence of the intense rainfall and high flow impacts experienced on the ground. Quantitative observations were manually extracted from river level photographs by the lead author in order to minimise error. Quality control checks were also manually carried out on the rainfall datasets to ensure valid observations were available for use. This involved comparing daily totals against each other, checking for gaps and outliers in the datasets, only authenticating extreme rainfall values when photographs/videos of impacts aligned, and comparing observations against average annual rainfall totals.

After establishing a network of manual rain gauges for ongoing 24-hour community observations, data from both 'Townfoot' (data quality accepted, representing the town and lower catchment) and 'Cawburn' (poor quality data sourced from the mid-catchment region) were then used within this modelling study. These two gauges offer a good comparison between datasets to emphasise the importance of good quality citizen science observations. They also contain data for the full modelling period of interest (January 2014 to May 2015). The spatial and temporal availability of community-based observations used in the SHETRAN modelling study are presented in Fig. 3, along with statistics which were used to rule out the Cawburn gauge during the quality control process. The Cawburn gauge was rejected for valid use because rainfall totals were considerably underestimated, particularly with respect to extreme events; it was, however, used in this modelling study to demonstrate the effect of a poor quality community dataset on model performance. The Cawburn observer originally highlighted that their gauge may be invalid due to lack of regular maintenance.

Flood observations provided by the community highlight three interesting high flow (flash flood) events. This paper explores all three events, focussing mainly on Event 1 (further outputs for Events 2 and 3 are in the Supplementary Material):

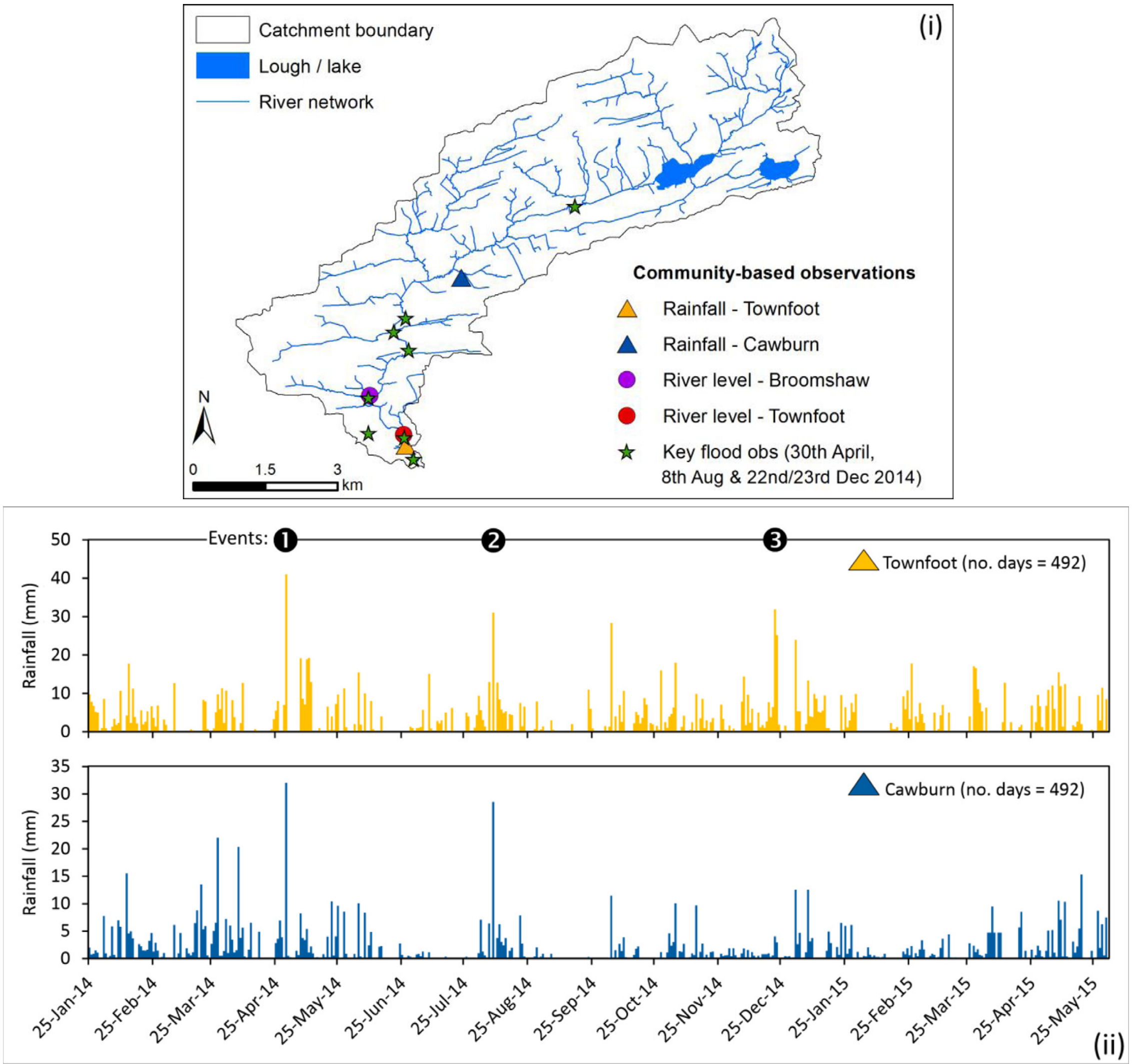
1. 30th April 2014: an intense convective storm (described as a 'cloud burst') which was localised over the town of Haltwhistle;
2. 8th August 2014: a convective summer storm falling on dry ground and mainly in the upper catchment;
3. 22nd/23rd December 2014: an intense and prolonged period of winter rainfall over a saturated catchment, causing widespread flooding, and morphological response comprising mass transportation and deposition of sediment.

3.3. Traditional hydrometric monitoring network

Prior to the project, there were no traditional ground-based hydrometric monitoring networks in operation within the catchment boundary. A traditional hydrometric monitoring network was therefore set up alongside the community-based scheme to fill the data gaps, capture local response and offer scientifically robust

Table 1
Examples of community-based monitoring techniques used in Haltwhistle which are relevant to this modelling study.

Parameter	24-hour rainfall totals	River (water) levels (sporadic/daily)	Flood-related information
Method	Plastic manual rain gauge in back gardens, placed at ground level. Graduated scale in millimetres for quantitative observations taken at the same time, usually every day in the same location.	Manual river level gauge boards at key (fixed) locations. 'River Watch Photo Posts' erected to provide instructions and consistency. Photographs or direct quantitative measurements taken.	• Anecdotes/eye-witness descriptions; • Photographs; • Videos; • Extra river levels. All provided with date, time and locational information.
Example			



(iii) Quality control checks (*note that the 1961-1990 average annual rainfall is 1000-1500mm, Kay et al., 2013)

Gauge	No. gaps in data	Σrainfall (mm) 492 days	Σrainfall (mm) annual*		Extremes valid?	Dataset accepted?
			01/03/14 – 01/03/15	01/05/14 – 01/05/15		
▲ Townfoot	0	1530	1107	1087	✓	✓
▲ Cawburn	0	857	587	472	x	x

Fig. 3. Spatial (i) and temporal (ii) availability of community-based observations used to model, along with a summary of the quality control checks used to accept or reject individual rain gauges (iii). The Townfoot rain gauge has also been compared with traditional gauges (see Supplementary Material). Note that Cawburn rainfall totals are significantly lower than expected, hence it was rejected. (See above-mentioned references for further information.)

hydrological data. Rainfall and discharge datasets were necessary to calibrate and validate SHETRAN, but also to demonstrate the value of community-based input data (as rainfall influences runoff).

An aerodynamic tipping bucket rain gauge and six pressure transducers for water level recording were installed between January and May 2014. Flow gauging was required to convert water level into discharge (Q) using stage-velocity-area derived rating curves (see Supplementary Material for detail). Data from a nearby UK Met Office daily rain gauge at Blenkinsopp Hall (west of the catchment boundary) was also sourced from the British Atmospheric Data Centre (BADC). The spatial and temporal availability of traditional data used in SHETRAN are shown in Fig. 4. A few gaps exist in the time series because of equipment failure, including battery failure, network issues, data storage capacities and damage caused by cattle.

Met Office 1 km NIMROD rainfall radar data was also sourced from the BADC and represents an alternate source of traditional data. It was only feasible to study the three flood events listed above due to the large the amount of processing required to extract and prepare the data, as well as run SHETRAN.

3.4. Hydrological modelling using SHETRAN

SHETRAN (Système Hydrologique Européen TRANsport) is a PBSD hydrological model which is capable of simulating spatially-distributed hydrological processes at a catchment scale (Newcastle University, 2016). Catchments are represented by a three-dimensional discretised grid and a simplified river network (known in this model as 'channel links'), thus the model can represent both surface and subsurface processes. SHETRAN is well-established and researched in the literature, with modellers utilis-

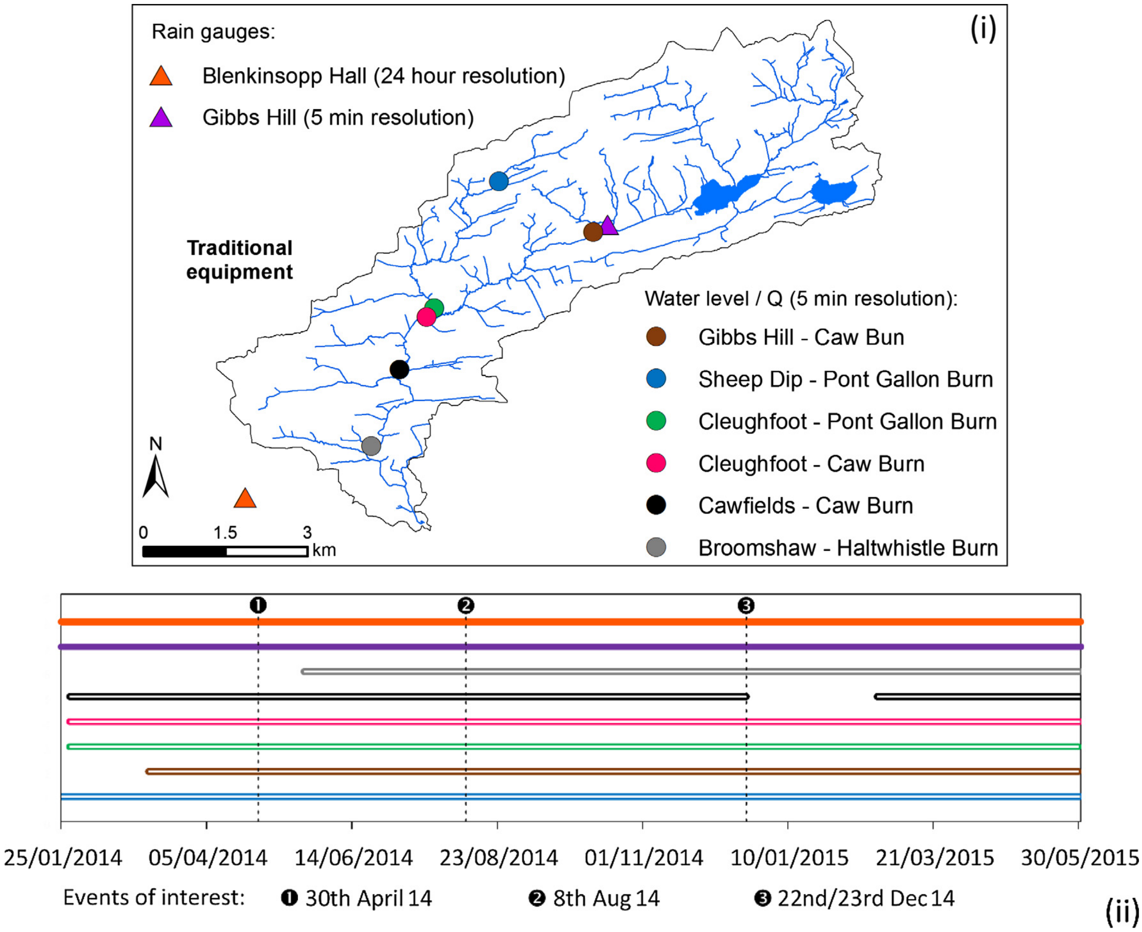


Fig. 4. Spatial (i) and temporal (ii) availability of traditional datasets used in this study. Colours correspond to each individual gauge on the map.

ing it to obtain discharge information for a variety of applications (Birkinshaw et al., 2011, 2014; Mourato et al., 2015; Parkin et al., 2007). However, SHETRAN has not yet been used to demonstrate the value of community-based observations. Being a PBSD model, it provides an opportunity to use observed data from various sources and locations, and integrate them into the hydrological cycle.

The most recent version of SHETRAN was sourced from Newcastle University (2016). Table 2 summarises the input data sourced and prepared for the Haltwhistle Burn catchment, along with other relevant model settings required. SHETRAN was set up to run between 26/01/2014 00:00 and 01/06/2015 00:00 GMT, a period of 491 days which makes use of the best available data when both community-based and traditional datasets overlap.

Based on the input layers, SHETRAN represents the Haltwhistle Burn catchment using the river network and catchment grid presented in Fig. 5. Output locations (Q*sim*) corresponding to each gauging station (Q*obs*) are also highlighted. The model described in this section is referred to as 'Model A'. Where changes have been made to input rainfall, a new model name is used.

SHETRAN has been manually calibrated using an iterative approach by systematically changing the values of input parameters. The parameters are those which are reported to be hydrologically sensitive in the literature and in SHETRAN (Birkinshaw et al., 2011, 2014; Đukić and Radić, 2016; Mourato et al., 2015), including the Strickler overland flow (SOF) roughness coefficient, soil depth (SD), saturated hydraulic conductivity (Ks) and the ratio of actual to potential evapotranspiration (AE/PE). These parameters can be adjusted within the soil and land cover layers and therefore allow the model to account for local variability in surface and subsurface properties. The aim of the calibration phase is to alter the model parameters in order to minimise the error between Q*obs* (the benchmark) and Q*sim*, whilst still being physically acceptable (Beven, 2009). The validation phase involved running the model for an independent set of data to check that the model settings still produced an acceptable simulation. A split sample test was used to divide the calibration and validation periods (see Table 3); both periods contain an adequate range of hydrological conditions.

Alongside graphical and visual inspection, it is good practice to use a combination of statistical performance indicators to assess model performance (e.g. Hall, 2001; Krause et al., 2005; Moriasi et al., 2007). The following tests, which are frequently used to assess hydrographs, were used. The acceptable performance values listed for each were chosen based on limits reported in the literature as providing reliable modelled outputs (Moriasi et al., 2007; Mourato et al., 2015):

- Coefficient of determination (R^2), with 0.7 being used as the minimum acceptable value;
- Root mean square error (RMSE), to provide an indication of performance in the same units as Q;
- Percentage bias (PBIAS), with ±25% being reported as the maximum acceptable error;
- Nash-Sutcliffe Efficiency (NSE) coefficient, with anything above +0.5 reported to provide at least a 'good' fit.

In order to demonstrate the value of community-based observations, a 'leave-one-out' methodology was adopted. The leave-one-

Table 2
Input data sourced and prepared ready for the Haltwhistle Burn SHETRAN model. Additional information is given in the text.

Item/setting required	Data source	Preparation for SHETRAN
Model resolution	100 m chosen – maximum resolution feasible (when considering model stability and simulation time).	
Mask	Outline derived using EDINA Digimap 5 m Panorama elevation data.	Aggregated to 100 m (4110 active cells in plan view available for simulation).
Minimum & mean filled DEM	50 m panorama elevation data supplied by EDINA Digimap. Elevation ranges from 101 m to 344 m AOD.	Resampled to 100 m resolution grid using minimum and mean aggregation techniques.
Precipitation (P)	Combination of data from Figs. 3 and 4 and rainfall radar used in the main modelling framework. Refer to the Thiessen polygons in Fig. 6 for spatial interpolation and distribution. Gibbs Hill, Blenkinsopp Hall and Townfoot gauges were initially used to set up the model.	
Potential evaporation (PE)	No automatic weather stations (AWS) available within catchment boundary. Met Office Spadeadam AWS used from the BADC (located 10 km north west from the catchment): • Maximum and minimum temperature; • Wind speed; • Relative humidity. Spadeadam did not contain any sunshine data. Brampton manual weather station run by a Met Office volunteer (located 21 km west from the catchment) used instead for 'total sunshine hours'. No gaps found in datasets used.	PE calculated using five weather parameters and the UN Food and Agriculture Organization recommended Penman-Monteith approach (Raes, 2012). This approach represents evaporation from a vegetated surface with an unlimited supply of water, which was considered sufficient for this study site and land cover. An open access tool described by Raes (2012) was used to calculate PE automatically. Final PE dataset was aggregated to a 24-hour resolution and used uniformly across the catchment.
Soil & geology	Peaty (upper catchment) and loamy (mid/lower catchment) soils with a moderately productive aquifer dominate. The EU soils database and British Geological Survey hydrogeology layers (1 km resolution) initially used to obtain realistic properties and set up the model.	Resampled to 100 m resolution grid. Calibration activities later refined the soil and geology datasets to allow for local variations in runoff.
Land cover	25 m Land Cover Map 2007 supplied by EDINA Digimap. Catchment is dominated by grassland (64%), evergreen forest (18%) and Shrub (11%).	Land cover codes reclassified to fit SHETRAN (arable, bare ground, grass, deciduous forest, evergreen forest, shrub and urban). Aggregated to 100 m grid. Calibration activities later refined land cover properties to allow for local variations in runoff.
Lakes	Ordnance Survey 1:10,000 Master Map shapefiles. Includes Greenlee (0.51 km^2) and Broomlee (0.30 km^2) Loughs in the upper catchment.	Converted to 100 m raster grid.
Max & min temperature	Temperatures are used directly in SHETRAN only for simulating snowpack development and snowmelt; there were no snow events during the simulation period.	
Output resolution & locations	SHETRAN was set to produce simulated discharge (Q*sim*) every 5 min for the six gauging stations which contain observed discharge (Q*obs*).	

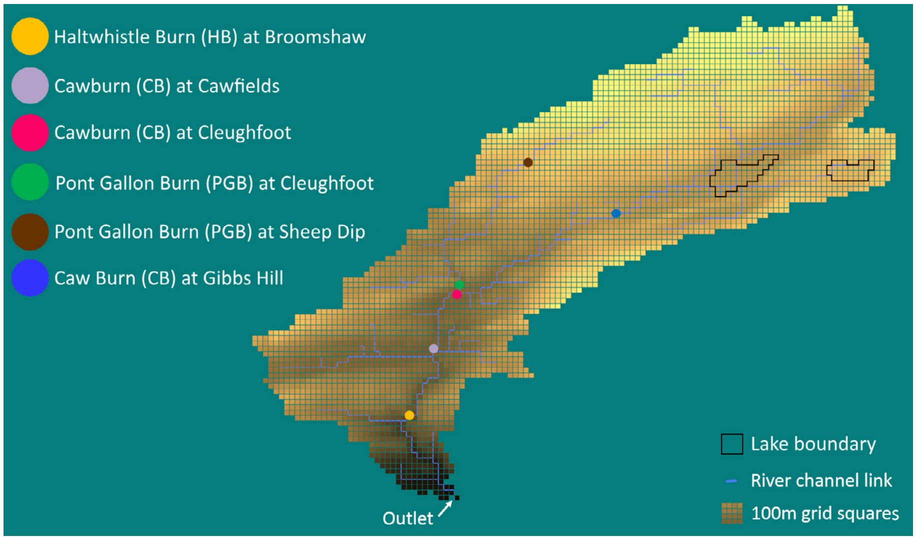


Fig. 5. SHETRAN 100 m grid and river network used to represent the Haltwhistle Burn catchment. Coloured dots represent locations where modelled discharge (Q*sim*) have been extracted. Watercourse abbreviations are referred to in later sections.

Table 3
Defining the calibration and validation periods within the full simulation period of interest.

Simulation period	Time period (from – to) (GMT)	Number of days
Calibration	28/09/2014 00:00 to 01/06/2015 00:00	246
Validation	26/01/2014 00:00 to 27/09/2014 23:55	245

out procedure involved re-running the already calibrated, therefore accepted, SHETRAN model multiple times. On each occasion different elements of information were excluded from the simulation to test how well the model performs without it. Beven (2009) and Otieno et al. (2014) advocate leaving observations out of the rainfall interpolation and modelling process as a way of demonstrating their value. Such an approach has allowed different sources (therefore combinations) of rainfall data to be used and assessed against the 'baseline' (Model A). This approach was feasible as precipitation is SHETRAN's main temporal and spatial driving variable. Making use of a 'patchwork' of heterogeneous information, combinations used were dictated by the spatial and temporal availability of input precipitation data previously described. SHETRAN was not recalibrated before each combination; other than the rainfall data, all parameters and datasets remained constant throughout. The performance of Model A was expected to degrade with diminished rainfall information, offering an opportunity to test model performance in relation to each other.

Point rainfall measurements were spatially interpolated across the catchment to create a 100 m resolution grid using conventional Thiessen polygons (Fig. 6). Although there are many other interpolation techniques available (e.g. Shaw et al., 2011), Thiessen polygons, which assign areas of the catchment to the nearest point measurement, are able to represent localised storms well if enough rain gauges are present (therefore providing a good test here). Interpolation methods, such as arithmetic mean, cannot achieve this and more advanced geostatistical techniques were not expected to yield better results. Alongside catchment-wide rainfall radar data, traditional, community-based and a combination of both data sources were used to create these spatial maps. It should be noted that the Cawburn gauge data was also incorporated into some scenarios to demonstrate potential implications when 'rejected' observations are used. Since community-based rainfall observations and the UK Met Office Blenkinsopp Hall gauge have coarser temporal resolutions, these data have been disaggregated into 5-min timesteps by imposing the same rainfall pattern from a traditional 5-min resolution rain gauge (Gibbs Hill), in model scenarios where this detail is available (Models A, B and E). Where this detail is not available (Models C, D, F and G), they have kept their original resolution to allow model performance to be evaluated whilst using these temporally coarser observations. The statistical performance indicators were then utilised to quantitatively assess the effects of each rainfall combination.

4. Results & discussion

4.1. Enhancing SHETRAN's inputs using quantitative and qualitative observations

Analysis of different sources of rainfall has highlighted the importance of spatial and temporal observations, particularly during the period of intense localised rainfall experienced on the 30th April 2014 (Event 1). Fig. 7 displays a set of 48-hour cumulative rainfall plots which represent Event 1 for each of the three gauges used to initially build Model A. It is clear that the traditional gauges observed much lower rainfall totals (17.6 mm and 17.9 mm) compared with community-based (41 mm), despite being only a few kilometres apart. If the community-based observations had not been available, the traditional gauges would have completely missed these larger totals observed over the lower catchment and the impact zone. However, Fig. 7 also confirms that while rainfall radar totals were significantly lower than those observed by the community, the radar observations did show the spatial location and extent of the storm and provided detailed temporal resolution, thus have captured steeper cumulative trends, hence implying a more intense, short-lived storm.

One obvious drawback with community-based rainfall observations is that they are usually reported on a 24-hour basis. If used in isolation at this resolution, only rainfall totals can be extracted. However, the full range of qualitative and quantitative community-based observations displayed in Fig. 8 (photographs, videos, tweets and anecdotes) illustrate how the wider community can contribute to the generation of an 'event timeline' which

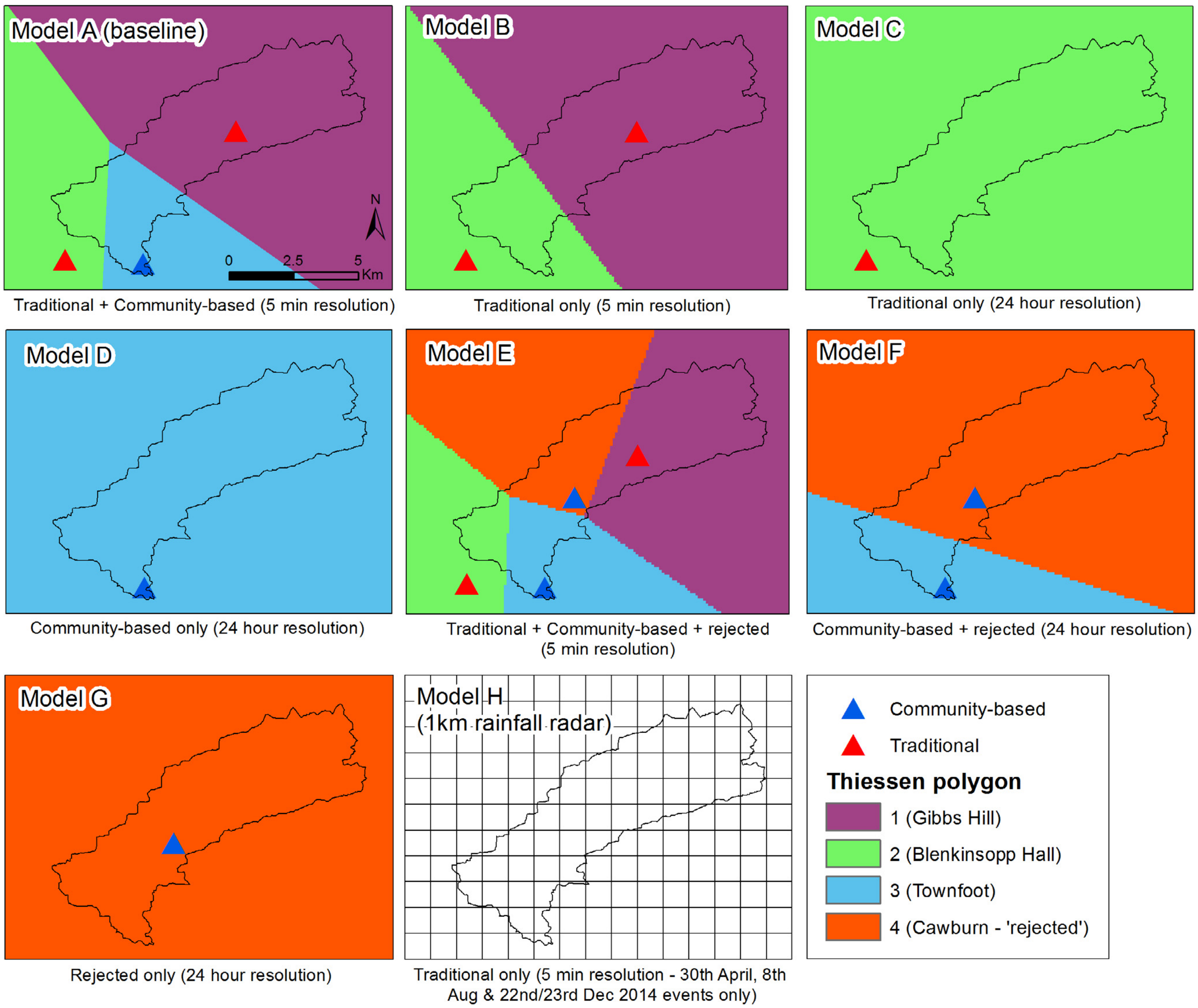


Fig. 6. Combination of rain gauges and resulting Thiessen polygons used to spatially estimate precipitation across the catchment in Models B-G (includes original Model A), as well as a 1 km resolution grid which utilises Met Office rainfall radar data (Model H). Original rainfall datasets have been directly fed into these models, rather than calculating areal rainfall, in order to capture spatial variability.

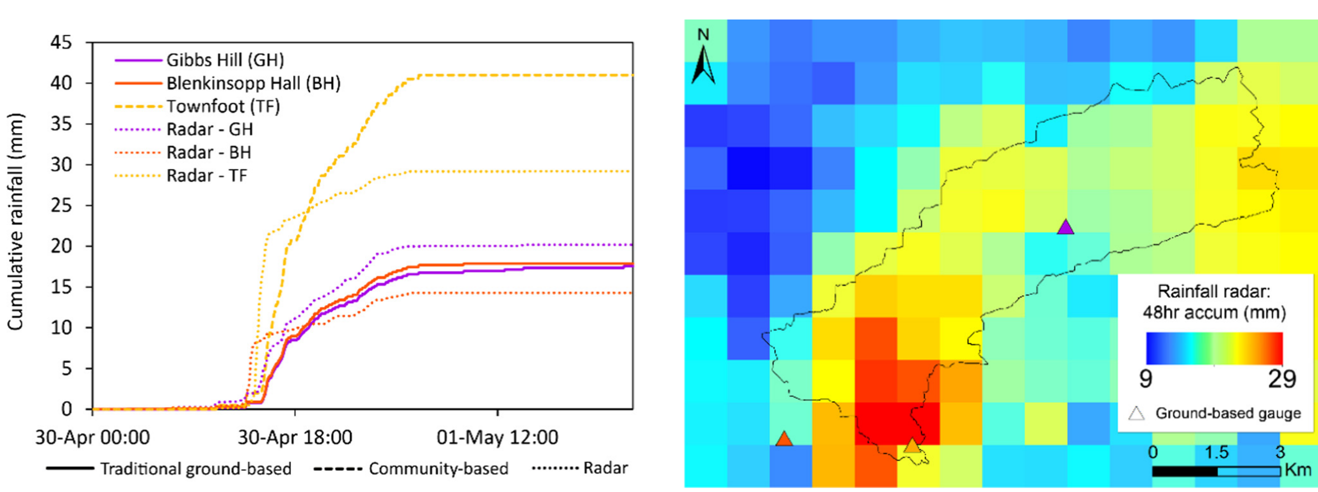


Fig. 7. Left: 48-hour cumulative rainfall plots for Event 1 (30/04/2014 00:00 to 02/05/2014 00:00 GMT) for each gauge initially used in Model A, and rainfall radar where each gauge overlaps. Right: Rainfall radar accumulations for the same period across the catchment. Ground-based gauges are overlaid onto the radar grid. Plots relating to Events 2 and 3 can be found in the Supplementary Material.

Fig. 8. A timeline of Event 1 (30th April 2014) created by harnessing a range of community-based quantitative and qualitative observations collected on the ground. Note quotes such as "*Monsoon alert. Heaviest rain I've seen in ages*!", and early warnings submitted and then crowd-sourced using Twitter.

specifically highlights when the storm started and finished. Together with the quantitative rainfall totals, this simple source of ground-based evidence allows duration, magnitude and intensity information to be inferred on a local scale. For Event 1, after observations were captured and shared by the public, it was clear that the event was extremely intense with 41 mm falling in just 30 min in the lower Haltwhistle Burn catchment. This was derived by assessing the timeline of observations presented in Fig. 8, which visually and anecdotally confirms that heavy rain was experienced locally on the ground between 15:20 and 15:50 (BST). An event as intense as this would also be required to generate the flood and debris-related impacts witnessed on the ground by the community. Rainfall totals can thus be disaggregated across the specific time period when it was physically observed (in this case, 41 mm of rain disaggregated evenly across 30 min), rather than 24 hours, to realistically replicate a high intensity storm. SHETRAN's precipitation time series were therefore updated to reflect the nature of Event 1 before Model A was calibrated.

This heterogeneous data integration process has only been possible due to the number of community-based observations being available and because the rainfall event hit the town where people live and walk past the Haltwhistle Burn. Event 2 (8th August 2014) provides an example where the storm was centred higher up in the catchment, meaning the downstream community were unable to provide information to help interpret quantitative rainfall totals. Event 3 (in December 2014) was more widespread with saturated antecedent conditions, so observations captured by the community were useful for highlighting downstream impacts. The value of the community-based rainfall observations for Event 1 have therefore been enriched as it was possible to extract important hydro-information from the patchwork of informal and heterogenic community-based observations, and utilise them within SHETRAN to characterise the high intensity storm. These sub-hourly and highly localised hydrological events, which are still poorly monitored and understood by professionals, require this level of detail in order to better characterise them and their impacts (Archer and Fowler, 2015; Archer et al., 2016; Perks et al., 2016).

4.2. Final calibration and validation results (Model A)

Initial calibration simulations for Model A reproduced the overall shape and timing of each hydrograph reasonably well. In order to improve SHETRAN's ability to reproduce Q*obs* at Gibbs Hill, the SOF values of the actual channel links of the loughs (links which overlapped the lake layer) needed to be reduced from 3.0 to 0.1. These results have subsequently highlighted the importance of the two lakes (Greenlee and Broomlee Lough – shallow water and wetland nature reserves) in the upper catchment and their ability to naturally attenuate high flows during and after rainfall. Final model settings adopted are listed in the Supplementary Material.

Final calibration and validation results are presented in Table 4. Fig. 9 also contains graphical comparisons of Q*obs* and Q*sim* (using Gibbs Hill, Sheep Dip and Broomshaw as examples) as well as Q*sim* for each gauging station. All of the statistics fall within acceptable limits, except for the Pont Gallon Burn at Sheep Dip during the validation period. This has been attributed to the Pont Gallon Burn sub-catchment not containing its own rain gauge, which would have been necessary to fully capture the localised rainfall experienced during Event 2. Despite this, the model's overall average (catchment-wide) performance is still well above the acceptance levels across the multiple indicators, so this SHETRAN model was accepted for its intended use. The multi-location and multi-response approach has highlighted the importance of sub-catchment information and catchment connectivity to the

Table 4
Final statistical results for the calibration and validation periods. Results relate to Model A using best available data, including quantitative and qualitative community-based observations (watercourse acronyms: Caw Burn, CB; Haltwhistle Burn, HB; Pont Gallon Burn, PGB).

Gauge/Output Location	R^2	RMSE (m^3/s)	PBIAS (%)	NSE
Calibration period: 28/09/2014 00:00 to 01/06/2015 00:00 (where observed data is available)				
CB at Gibbs Hill	0.92	0.26	−5.56	0.85
PGB at Sheep Dip	0.83	0.04	3.33	0.78
PGB at Cleughfoot	0.89	0.11	−13.29	0.88
CB at Cleughfoot	0.92	0.35	−9.31	0.90
CB at Cawfields	0.84	0.36	−6.71	0.86
HB at Broomshaw	0.88	0.47	0.48	0.77
Average	**0.88**	**0.27**	**−5.18**	**0.84**
Validation period: 26/01/2014 00:00 to 27/09/2014 23:55 (where observed data is available)				
CB at Gibbs Hill	0.90	0.10	10.47	0.88
PGB at Sheep Dip	0.52	0.04	−47.63	0.21
PGB at Cleughfoot	0.77	0.09	−12.20	0.76
CB at Cleughfoot	0.89	0.19	−8.34	0.86
CB at Cawfields	0.86	0.24	−4.77	0.85
HB at Broomshaw	0.87	0.14	14.86	0.72
Average	**0.80**	**0.13**	**−7.94**	**0.71**

calibration process as the Haltwhistle Burn catchment does not respond in a uniform way.

4.3. Performance of SHETRAN using different combinations of rainfall data

Models B-G have been assessed across the full modelling period to determine the change in SHETRAN's performance in relation to the calibrated and validated (therefore accepted) baseline model, A.

Table 5 (i) presents the statistical results (averaged across all six gauging stations) for each model simulated i.e. rain gauge combination tested. The most notable trends exposed are that model performance progressively deteriorates from Model A to G and, as expected, A continues to be the most acceptable model for use. These trends are strengthened by the fact that multiple statistical performance indicators express the same trends, as well as overall discharge error (as PBIAS results, which relate to mass balance, illustrate). A more pronounced case for these trends is exemplified in Table 5 (ii) which present the same set of statistics, but only for the Haltwhistle Burn at Broomshaw, where the bulk of community-based observations exist. For instance, the NSE coefficient falls by 1.30 when comparing Model G against A, whereas the difference between the same two models is only 1.09 when assessing all six gauging stations at the same time. Note that this trend is still apparent despite the Broomshaw gauge analysis excluding Event 1 (i.e. missing Q*obs*).

The following points can also be noted when assessing the full modelling period (rather than individual peaks):

- The performance of Model A is only marginally better than B, implying both should be acceptable for wider use. The use of community-based observations has not therefore degraded SHETRAN's predictive power, but similar results would have been obtained for the full modelling period if only two traditional gauges (Model B) were available. Nevertheless, this comparison emphasises that it is feasible to create an acceptable model containing community-based observations and achieve statistical results similar to those obtained in other SHETRAN studies (Birkinshaw et al., 2011, 2014);
- 'Rejected' community-based rainfall observations have significantly affected (degraded) model performance, particularly the mass balance aspect. Comparisons between Model A and E show this most clearly;
- Use of community-based observations alone significantly degrades model performance. However, the use of one good quality community-based rain gauge (Model D) produces statistical results which are similar to the outputs obtained when using one traditional rain gauge (Model C). However, this is not the case for the 'rejected' community-based data when used in isolation (Model G);
- Models containing two or three rain gauges, for which it has been possible to disaggregate time series into 5 min intervals, have produced reliable outputs. This is also true for models containing input data which had not been rejected during the quality control process. Models using only one rain gauge at a 24-hour resolution (Models C and D) would be rejected here. Nevertheless, some modelling studies regularly use these coarser resolutions.

Overall, these findings confirm that the resolution of the input data, the data quality and the total number of rain gauges used override the importance of whether or not community-based observations are used alongside traditional sources. These are obvious and important factors which modellers traditionally consider (Beven, 2009; Beven and Westerberg, 2011; Montanari and Di Baldassarre, 2013). This suggests that there is potential for integrating community-based observations with traditional sources to fill monitoring gaps, to support the modelling process and to characterise catchments on a local scale meaningful to resident communities. Findings here also complement results obtained by Mazzoleni et al. (2015) who found that synthetic intermittent observations improved model performance for streamflow. It is also important to remember that traditional observations are not free from error and can still provide incorrect information (Beven and Westerberg, 2011).

4.4. Importance of community-based observations during flood events

Event 1 (30th April 2014) has been isolated here for analysis to determine how SHETRAN performs during a localised flash flood

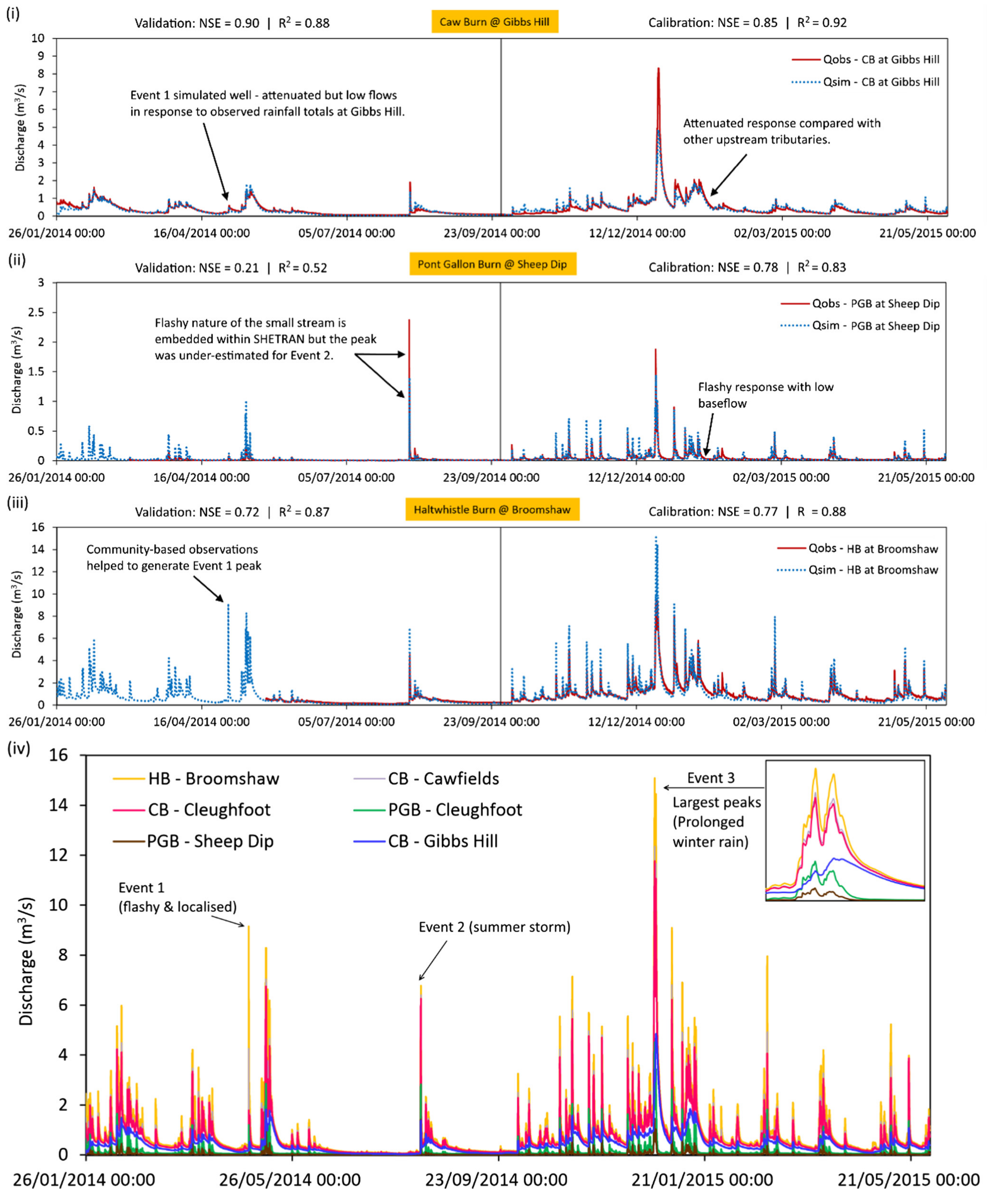


Fig. 9. Q*obs* and Q*sim* results for the Caw Burn at Gibbs Hill, the Pont Gallon Burn at Sheep Dip and the Haltwhistle Burn at Broomshaw, plotted (i-iii) for Model A over the calibration and validation periods. Q*sim* for all gauging stations are also presented together, which emphasises variation in sub-catchment response (iv).

event when a patchwork of community-based observations are most abundant, as well as rainfall radar.

Table 6 (i) contains the statistical results relating to Event 1, comprising an analysis covering four days to capture the rise and recession of a single event-based hydrograph. The dominant pattern generally involves a degradation in model performance when rain gauges are removed or rainfall radar is used. Performance diminishes when community-based observations are completely absent or when the Thiessen polygon over-exaggerates the spatial scale of the convective storm (in this case the 41 mm captured by

Table 5
Average (i) and Broomshaw only (ii) SHETRAN results for Models A-G across the full modelling period (rain gauge combinations: • traditional and community-based, * traditional only, ◆ community-based only and ◇ rejected).

Model, rain gauge combination & total number of rain gauges used (brackets)	R^2	RMSE (m^3/s)	PBIAS (%)	NSE
Full modelling period: 26/01/2014 00:00 to 01/06/2015 00:00 (where observed data is available) *(i) Average results across all six gauging stations:*				
A • (3)	0.86	0.22	−5.17	0.83
B * (2)	0.85	0.23	−4.90	0.82
C * (1)	0.61	0.33	2.98	0.61
D ◆ (1)	0.55	0.36	−5.10	0.48
E •◇ (4)	0.58	0.30	25.09	0.53
F ◆◇ (2)	0.11	0.63	59.98	−0.23
G ◇ (1)	0.05	0.64	61.08	−0.26
Full modelling period: 26/01/2014 00:00 to 01/06/2015 00:00 (where observed data is available) *(ii) Results for the Haltwhistle Burn at Broomshaw only:*				
A • (3)	0.90	0.39	2.42	0.81
B * (2)	0.89	0.40	2.71	0.80
C * (1)	0.80	0.42	12.01	0.77
D ◆ (1)	0.79	0.41	6.56	0.78
E •◇ (4)	0.93	0.32	23.48	0.86
F ◆◇ (2)	0.46	0.99	74.21	−0.26
G ◇ (1)	0.09	1.07	80.34	−0.49

Table 6
Average (i) and Cawfields only (ii) SHETRAN results for Models A-H across Event 1 (rain gauge combinations: • traditional and community-based, * traditional only, ◆ community-based only and ◇ rejected). ° Assessment excludes any Broomshaw observations.

Model, rain gauge combination & total number of rain gauges used (brackets)	R^2	RMSE (m^3/s)	PBIAS (%)	NSE
Event 1 (30th April): 29/04/2014 00:00 to 03/05/2014 00:00 (° where observed data is available) *(i) Average results across five gauging stations:*				
A • (3)	0.76	0.44	13.94	0.49
B * (2)	0.43	0.60	21.55	0.22
C * (1)	0.32	0.64	36.01	0.03
D ◆ (1)	0.53	1.55	−189.92	−134.82
E •◇ (4)	0.80	0.24	−19.10	−4.72
F ◆◇ (2)	0.65	0.82	−148.71	−50.24
G ◇ (1)	0.58	0.74	−122.50	−47.17
H Rainfall radar •	0.52	0.56	13.55	0.09
Event 1 (30th April): 29/04/2014 00:00 to 03/05/2014 00:00 *(ii) Results for the Caw Burn at Cawfields only:*				
A • (3)	0.75	1.03	28.67	0.54
B * (2)	0.09	1.50	42.53	0.02
C * (1)	0.03	1.57	52.64	−0.08
D ◆ (1)	0.92	2.46	−82.72	−1.65
E •◇ (4)	0.96	0.55	18.78	0.87
F ◆◇ (2)	0.96	1.09	−69.08	0.48
G ◇ (1)	0.95	0.99	−43.65	0.57
H Rainfall radar •	0.23	1.40	38.68	0.14

the community). Analysis confirms that the community-based observations have helped to capture river response following the storm but the spatial extent of the event is not accurately represented, even by Model A. Table 6 (ii) contains SHETRAN's response for the Caw Burn at Cawfields. This gauging station is used to represent river response upstream of the town because observed water level (therefore discharge) was not recorded at Broomshaw for this period (see data gap in Fig. 4). Compared to the catchment's average response, model performance at the Cawfields gauge is significantly enhanced when community-based observations are incorporated.

Fig. 10 presents discharge plots for each model at each gauging station, along with observed data for comparison. Manual river levels observed by the community (subsequently converted to

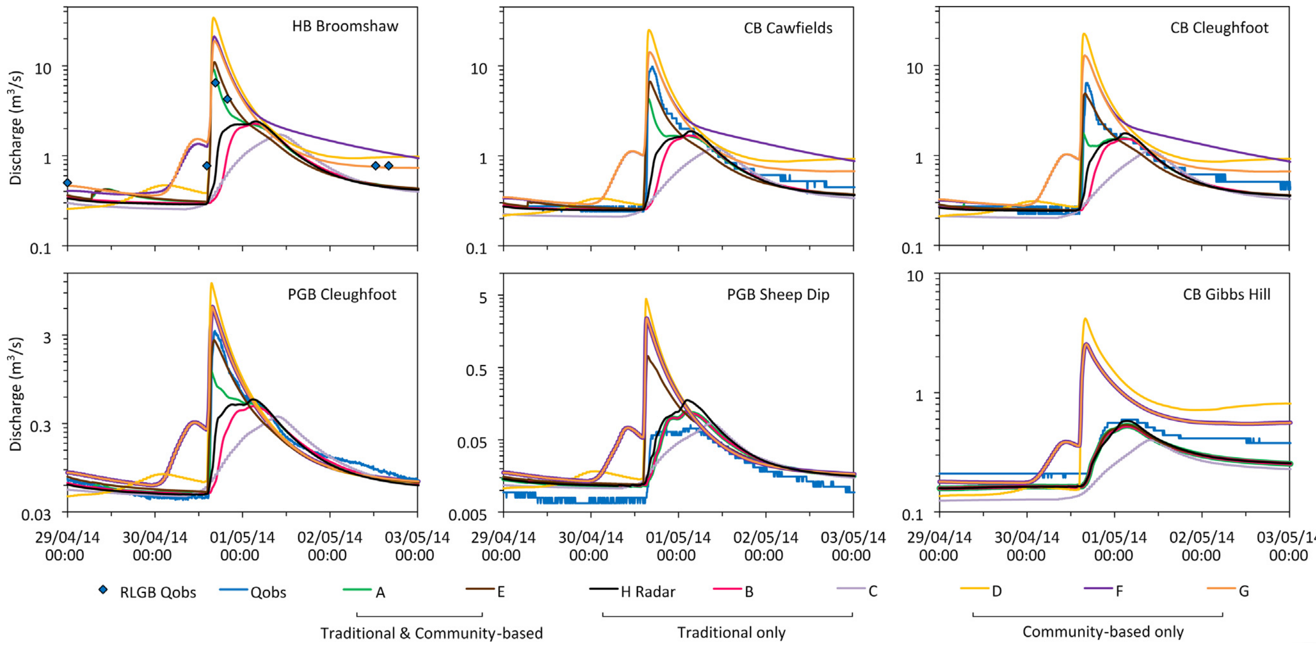


Fig. 10. Hydrograph shape: final simulated discharge obtained from SHETRAN Models A-H for all relevant gauging stations during the April 2014 event. Includes manual river level gauge board (RLGB) observations collected by the community which have been converted into discharge. Note that discharge has been plotted using a logarithmic scale.

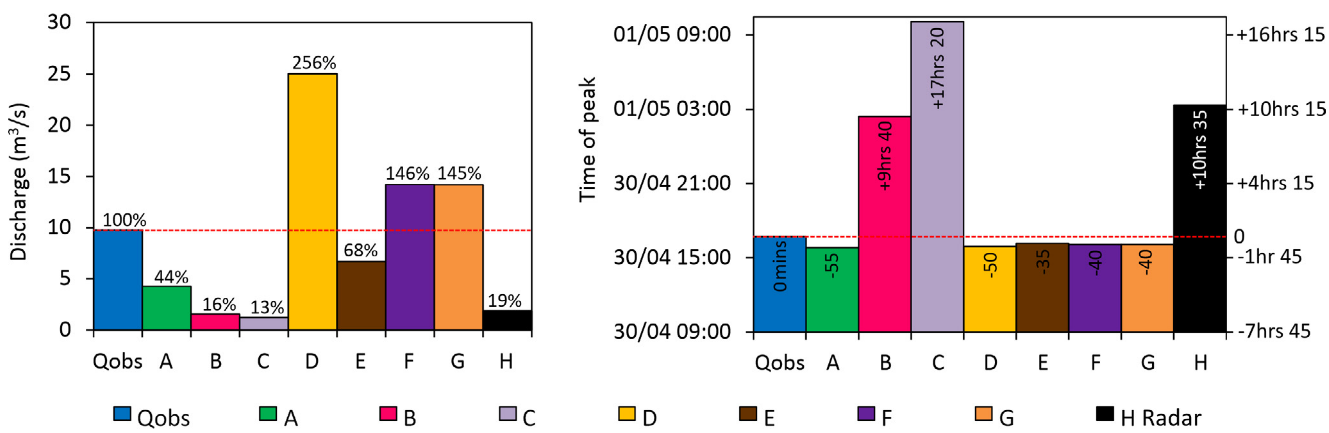


Fig. 11. A comparison between observed Q (Q*obs*) and modelled discharge (Q*sim*) for Models A-H at Cawfields Caw Burn during the 30th April 2014 event: peak discharge (left) and timing of the peak discharge (right).

discharge using the site's rating curve) have also been added to the Broomshaw comparison. Graphs help to interpret model performance relating to the shape of the hydrographs, and more specifically, the rapid rise which is only reproduced when community-based observations are integrated. Use of rainfall radar appears to improve the response of the model compared with use of only the two traditional rain gauges, but a flashy response is still absent. Although the community failed to record a river level (therefore river level gauge board (RLGB) Q*obs*, once converted to discharge) as the burn peaked at Broomshaw, the modelled hydrographs did correlate well with the six spot readings that they did manage to observe. This is true for all but the 'traditional only' models. A variety of quantitative and qualitative community-based observations have therefore been beneficially incorporated into SHETRAN and used to validate the model. However, the value of these observations are governed by a number of factors, for instance, when the peak exactly occurs (time of day, week and season) and proximity of monitoring sites to residents' homes.

Fig. 11 quantifies the impacts of each rain gauge combination on timing and magnitude of the flood peak for the Caw Burn at Cawfields. For this particular case, the following findings are highlighted when compared with observed peak discharge:

- Models B and C (traditional only combination) underestimate the flood peak by 84% and 87% respectively. Rainfall radar closely follows with 81%;
- Model D, which used a uniform grid of community-based observations, overestimates the flood peak by 156%;
- The best representation of magnitude comes from Model E, a combination of four gauges which underestimates the flood peak by 32%. This is better than Model A, and despite containing the rejected rain gauge, Model E is likely to have created a better representation of the rainfall extent;

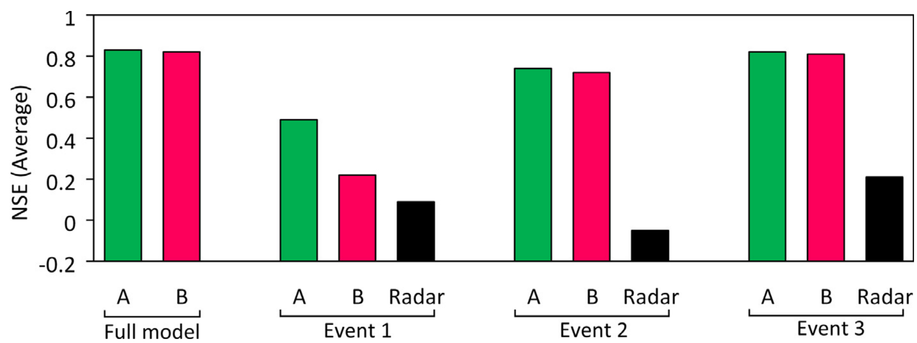


Fig. 12. NSE coefficients obtained from three key models of interest (Model A, B and Rainfall radar), each shown for the full modelling period (Jan 2014 – May 2015) and Events 1, 2 and 3. Graphs display average NSE results across all six gauging stations.

- All models containing community-based observations produce peaks which arrive within 55 min of the observed, with Model E being the closest at 35 min. Extra rain gauges above the town would have captured the extent of this intense storm more precisely, which in turn would generate a more accurate time lag;
- The timing of the traditional only combinations were considerably delayed because the hydrographs were too attenuated. The peak of the flood was over 9 h (Model B), 10 h (rainfall radar) and even as delayed as 17 h (Model C).

Event 1 has also been compared here against Event 2 (August) and 3 (December) to determine how far the value of community-based observations varies depending on the nature and length of the hydrological event (the same set of statistics and plots as those in this section are available in the Supplementary Material for these two additional events). Fig. 12 highlights the key differences between Events 1, 2, and 3, and the full 491 days modelled. The comparison uses NSE coefficients obtained, on average across the six gauging stations, from Model A and also B and H (radar) as these models alone present practical combinations which stakeholders would typically use (i.e. the best combination of traditional ground-based gauges (B) or rainfall radar (H) data which would normally be available) if the community-based observations did not exist to create Model A. Based on these plots, it is clear that the inclusion of community-based observations alongside traditional data (Model A) adds most value (higher NSE) to the localised flash flood event in April. Very little value is added during the longer modelling period and the prolonged winter storm, meaning that the traditional gauges alone were sufficient. Little value is also added to Event 2, a short-lived storm which was concentrated over the upper catchment. Nevertheless, the outcome obtained from Event 2 was significantly governed by the location of this particular storm and the fact that there were no community-based rain observations to represent it. Models containing rainfall radar observations consistently reduced model performance, thus has not been affected by the nature or length of the storm.

The patchwork of quantitative and qualitative community-based observations used here were required to help capture the intense rainfall and flash flood response during Event 1. Smith et al. (2015) and Kutija et al. (2014) also emphasise the value of community-based observations during these hydrologically important events given that they are short-lived. Accurate coverage of the rainfall extent is also required, however, as it can cause significant over- or under-estimation if incorrect. Timing and magnitude are important factors which affect public response on the ground, response by organisations responsible for flood forecasting and warning, as well as catchment managers designing intervention measures to withstand or relieve short-lived floods. Community-based observations can therefore make a difference; they have the potential to increase the spatial resolution of ground-based gauges, as well as ground-truth rainfall radar observations which are routinely adjusted using gauge-based factors (Wang et al., 2015). Our findings also compliment Seibert and McDonnell (2015), who found that a small number of 'soft' and 'fuzzy' qualitative (knowledge-based) observations are extremely useful for understanding and modelling how catchments work, particularly under high flow conditions. Seibert and McDonnell (2015) also suggest combining these informal observations with the often limited network of traditional gauges. However, such an approach relies on unpaid members of the public to be physically present, actively monitoring and collecting good quality observations, which cannot always be guaranteed.

In this case study, seven manual rain gauges were originally distributed within the Haltwhistle Burn catchment ready for community-based monitoring, but only two of these (Townfoot and Cawburn) returned data covering the full modelling period. Due to the nature of citizen science and the practicality of getting volunteers to observe parameters manually over time, it is to be expected that datasets may be missing or incomplete from some monitoring sites. If the community were to be informed that their observations are most useful during localised flash flood events, then they can prioritise their monitoring efforts and pinpoint these specific occasions. In turn, the most valuable observations are more likely to be captured for a greater number of monitoring sites, and with an increased temporal resolution. There are obvious health and safety implications for members of the general public with this regard and the engagement, training and facilitation activities required to activate community-based monitoring schemes should be prioritised.

5. Conclusions

The Haltwhistle Burn catchment and focus community have been used to demonstrate the value of real community-based observations using a PBSD catchment model (SHETRAN) under a range of scenarios. It is clear that the wider public can provide valuable inputs via citizen science style data collection activities pertinent to catchment characterisation, modelling and management. Community-based activities are less complicated, significantly cheaper and less demanding (e.g. for power and processing) than their traditional counterparts, yet results here highlight how effective and valuable they can be. Examples presented here emphasise the importance of spatial and temporal information at a sub-catchment scale. Two key conclusions can be drawn from this work:

1. Our modelling results illustrate how a patchwork of quantitative and qualitative community-based observations (which together yield information relating to rainfall totals, timing, duration, and therefore intensity) are required alongside tradi-

tional sources of hydro-information in order to fill spatial and temporal data gaps, and to characterise local catchment response more accurately than using traditional data alone. This includes the behaviour, timing and magnitude of river response during and after floods;
2. Evidence presented here confirms that community-based rainfall observations are most valuable during local flash flood events. This information would otherwise often be missed, be under-unrecorded by existing ground-based gauges, or else be significantly underestimated by rainfall radar. Community-based observations are less valuable during prolonged and widespread floods, or over longer hydrological periods of interest.

Community-based observations have the potential to add spatial detail and to ground-truth existing traditional sources of catchment data, providing accurate information to support monitoring applications nationally, including weather and flood forecasting, modelling and longer-term catchment management initiatives. If community-based monitoring efforts are to be prioritised or streamlined, then, as with any hydrological monitoring, this potential can only be realised if appropriate procedures for quality control checking are established and followed. If the public recognise which of their observations are most valuable, and they are properly trained, then they are more likely to continue monitoring and providing good quality datasets which can contribute to the catchment management toolkit in the longer term.

It is acknowledged that the results presented here are location-, community-, event- and equipment-specific. However, this case study provides an early insight into what can be achieved and the value that is added when public participation is integrated into the catchment characterisation and management process. Data outcomes will evolve and improve over time given that citizen science is flourishing in line with technological advances, but will be naturally limited by participation levels. Overall, we conclude that a citizen science approach offers local communities an exciting way to learn about their local water environment, engage with professional stakeholders, and be actively part of the catchment management process.

Acknowledgements

This work has been part-funded by Tyne Rivers Trust's Catchment Restoration Fund project (Defra funded 2012-2015) and Eleanor Starkey's NERC PhD studentship (NE/L501748/1). We would like to thank all staff from Tyne Rivers Trust for support and access to the case study site. Additional guidance from Dr Mark Powell (community engagement) and Phil James (Android App) is also appreciated. James Neasham and Michael Pollock also provided valuable fieldwork assistance. The continued enthusiasm and participation of the Haltwhistle Burn community has been essential, and for that we are extremely grateful. We would also like to thank the editors and anonymous reviewers for their valuable feedback.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhydrol.2017.03.019. These data include Google maps of the most important areas described in this article.

References

Archer, D., Fowler, H., 2015. Characterising flash flood response to intense rainfall and impacts using historical information and gauged data in Britain. J. Flood Risk Manage. http://dx.doi.org/10.1111/jfr3.12187.

Archer, D.R., Parkin, G., Fowler, H., 2016. Assessing long term flash flooding frequency using historical information. Hydrol. Res. 47 (3). http://dx.doi.org/10.2166/nh.2016.031.

Aulov, O., Halem, M., 2012. Human sensor networks for improved modeling of natural disasters. Proc. IEEE 100 (10), 2812–2823.

Bárdossy, A., Pegram, G., 2013. Interpolation of precipitation under topographic influence at different time scales. Water Resour. Res. 49, 4545–4565.

Barthel, R., Seidl, R., Nickel, D., Büttner, H., 2016. Global change impacts on the Upper Danube Catchment (Central Europe): a study of participatory modeling. Reg. Environ. Change, 1–17. http://dx.doi.org/10.1007/s10113-015-0895-x.

Beven, K., 2009. Environmental Modelling: An Uncertain Future? Routledge, London.

Beven, K., Westerberg, I., 2011. On red herrings and real herrings: disinformation and information in hydrological interference. Hydrol. Process. 25, 1676–1680.

Birkinshaw, S.J., Bathurst, J.C., Iroume, A., Palacios, H., 2011. The effect of forest cover on peak flow and sediment discharge – an integrated field and modelling study in Central-Southern Chile. Hydrol. Process. 25, 1284–1297.

Birkinshaw, S.J., Bathurst, J.C., Robinson, M., 2014. 45 years of non-stationary hydrology over a forest plantation growth cycle, Coalburn catchment, Northern England. J. Hydrol. 519, 559–573.

Bracken, L., Bulkeley, H.A., Whitman, G., 2014. Transdisciplinary research: understanding the stakeholder perspective. J. Environ. Planning Manage. 58 (7), 1291–1308.

Burakowski, E., Wake, C.P., Dibb, J.E., Stampone, M., 2013. Putting the capital 'A' in CoCoRAHS: an experimental programme to measure albedo using the Community Collaborative Rain, Hail & Snow (CoCoRaHS) Network. Hydrol. Process. 27, 3024–3034.

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T.C., Bastiaensen, J., De Bièvre, B., Bhusal, J., Clark, J., Dewulf, A., Foggin, M., Hannah, D.M., Hergarten, C., Isaeva, A., Karpouzoglou, T., Pandeya, B., Paudel, D., Sharma, K., Steenhuis, T.S., Tilahun, S., Van Hecken, G., Zhumanova, M., 2014. Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. Front. Earth Sci. 2 (26), 1–21.

Buytaert, W., Dewulf, A., De Bièvre, B., Clark, J., Hannah, D., 2016. Citizen science for water resources management: toward polycentric monitoring and governance? J. Water Resour. Planning Manage. 10. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000641.

CaBA, 2016. *Catchment Based Approach*. Available at: http://www.catchmentbasedapproach.org/ (last accessed: 22nd November 2016).

Castell, N., Kobernus, M., Liu, H., Schneider, P., Lahoz, W., Berre, A.J., Noll, J., 2015. Mobile technologies and services for environmental monitoring: the Citi-Sense-MOB approach. Urban Climate 14 (3), 370–382.

Chan, S.C., Kendon, E.J., Roberts, N.M., Fowler, H.J., Blenkinsop, S., 2015. Downturn in scaling of UK extreme rainfall with temperature for future hottest days. Nat. Geosci. 9, 1–6.

Đukić, V., Radić, Z., 2016. Sensitivity analysis of a physically based distributed model. Water Resour. Manage. 30, 1669–1684.

Durkee, J., 2010. Precipitation measurement and the advancement towards global observations. Geography Compass 4 (8), 956–978.

Ewen, J., Parkin, G., O'Connell, P.E., 2000. SHETRAN: distributed river basin flow and transport modelling system. J. Hydrol. Eng. 5, 250–258.

Fohringer, J., Dransch, D., Kreibich, H., Schröter, K., 2015. Social media as an information source for rapid flood inundation mapping. Nat. Hazards Earth System Sci. 15, 2725–2738.

Follett, R., Strezov, V., 2015. An analysis of citizen science based research: usage and publication patterns. PLoS ONE 10 (11). http://dx.doi.org/10.1371/journal.pone.0143687.

Forzieri, G., Feyen, L., Russo, S., Vousdoukas, M., Alfieri, L., Outten, S., Mogliavacca, M., Bianchi, A., Rojas, R., Cid, A., 2016. Multi-hazard assessment in Europe under climate change. Clim. Change. http://dx.doi.org/10.1007/s10584-016-1661-x.

Hall, M.J., 2001. How well does your model fit the data? J. Hydroinf. 3 (1), 49–55.

Harrison, D.L., Driscoll, S.J., Kitchen, M., 2000. Improving precipitation estimates from weather radar using quality control and correction techniques. Meteorol. Appl. 7 (2), 135–144.

Hut, R., Jong, S., Giesen, N., 2014. Using umbrellas as mobile rain gauges: prototype demonstration. Geophysical Research Abstracts, EGU2014-16418.

Illingworth, S.M., Muller, C.L., Graves, R., Chapman, L., 2014. UK Citizen Rainfall Network: a pilot study. Weather 69 (8), 203–207.

Kay, A.L., Bell, V.A., Blyth, E.M., Crooks, S.M., Davies, H.N., Reynard, N.S., 2013. A hydrological perspective on evaporation: historical trends and future projections in Britain. J. Water Clim. 4 (3), 193–208.

Kendon, E.J., Roberts, N.M., Fowler, H.J., Roberts, M.J., Chan, S.C., Senior, C.A., 2014. Heavier summer downpours with climate change revealed by weather forecast resolution model. Nat. Clim. Change 4, 570–576.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97.

Kutija, V., Bertsch, R., Glenis, V., Alderson, D., Parkin, G., Walsh, C.L., Robinson, J., Kilsby, C., 2014. Model validation using crowd-sourced data from a large pluvial flood. In: 11th International Conference on Hydroinformatics, New York City College, New York.

Lanza, L.G., Ramirez, J.A., Todini, E., 2001. Stochastic rainfall interpolation and downscaling. Hydrol. Earth Syst. Sci. 5 (2), 139–143.

Large, A., Gilvear, D., Starkey, E., 2017. Ecosystem service-based approaches for status assessment of Anthropocene riverscapes. In: Kelly, J.A. (Ed.), Rivers of the Anthropocene. University of California Press.

Mazzoleni, M., Alfonso, L., Chacon-Hurtado, J., Solomatine, D., 2015. Assimilating uncertain, dynamic and intermittent streamflow observations in hydrological models. Adv. Water Resour. 83, 323–339.
Montanari, A., Di Baldassarre, G., 2013. Data errors and hydrological modelling: the role of model structure to propagate observation uncertainty. Adv. Water Resour. 51, 498–504.
Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transformations ASABE 50 (3), 885–900.
Mourato, S., Moreira, M., Corte-Real, J., 2015. Water resources impact assessment under climate change scenarios in mediterranean watersheds. Water Resour. Manage. 29 (7), 2377–2391.
Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., Leigh, R., 2015. Crowdsourcing for climate and atmospheric sciences: current status and future potential. Int. J. Climatol. 35 (11), 3185–3203.
Newcastle University, 2016. *SHETRAN Hydrological model*. Available at: http://research.ncl.ac.uk/shetran/ (last accessed: 22nd November 2016).
Nicholson, A., Wilkinson, M.E., O'Donnell, G.M., Quinn, P.F., 2012. Runoff attenuation features: a sustainable flood mitigation strategy in the Belford catchment, UK. Area 44, 463–469.
Otieno, H., Yang, J., Lui, W., Han, D., 2014. Influence of rain gauge density on interpolation method selection. J. Hydrol. Eng. 19 (11). http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000964.
Parkin, G., Birkinshaw, S.J., Younger, P.L., Rao, Z., Kirk, S., 2007. A numerical modelling and neural network approach to estimate the impact of groundwater abstractions on river flows. J. Hydrol. 339, 15–28.
Perks, M.T., Russell, A.J., Large, A.R.G., 2016. Technical Note: Advances in flash flood monitoring using UAVs. Hydrol. Earth Syst. Sci. Discuss. http://dx.doi.org/10.5194/hess-2016-12.
Pocock, M.J.O., Chapman, D.S., Sheppard, L.J., Roy, H.E., 2014. Choosing and Using Citizen Science. a guide to when and how to use citizen science to monitor biodiversity and the environment. Centre Ecol. Hydrol.
Raes, D., 2012. The ETo Calculator: Evapotranspiration from a reference surface. Reference Manual Version 3.2, September 2012. Food and Agriculture Organization of the United Nations, Rome, Italy.
Science Communication Unit, University of the West of England, Bristol, 2013. Science for Environment Policy in depth Report: Environmental Citizen Science. Report produced for the European Commission DG Environment, December 2013. Available at: http://ec.europa.eu/environment/integration/research/newsalert/pdf/IR9_en.pdf (last accessed: 22nd November 2016).
Seibert, J., McDonnell, J.J., 2015. Gauging the ungauged basin: relative value of soft and hard data. J. Hydrol. Eng. 20 (1), A4014004.
SEPA, 2015. Natural flood management handbook. Stirling: SEPA. Available from: http://www.sepa.org.uk/ (last accessed: 22nd November 2016).
Shaw, E.M., Beven, K.J., Chappell, N.A., Lamb, R., 2011. Hydrology in Practice. Spoon Press, Oxon.
Smith, L., Liang, Q., James, P., Lin, W., 2015. Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. J. Flood Risk Manage. http://dx.doi.org/10.1111/jfr3.12154.
Starkey, E., Parkin, G., 2015. Community Involvement in UK Catchment Management. Foundation for Water Research. Available at: http://www.fwr.org/Catchment/frr0021.pdf (last accessed: 22nd November 2016).
Sutherland, W.J., Roy, D.B., Amano, T., 2015. An agenda for the future of biological recording for ecological monitoring and citizen science. Biol. J. Linn. Soc. 115, 779–784.
Tweddle, J.C., Robinson, L.D., Pocock, M.J.O., Roy, H.E., 2012. Guide to Citizen Science: Developing, Implementing and Evaluating Citizen Science to Study Biodiversity and the Environment in the UK. NERC/Centre for Ecology & Hydrology, Wallingford, p. 29.
Tyne Rivers Trust, 2015. Haltwhistle Burn – a Comprehensive Catchment Approach to Headwater Runoff and Pollution: Technical Report of the 2012–2015 Catchment Restoration Fund Project. Tyne Rivers Trust, Corbridge.
UK Met Office, 2010. Observations: National Meteorological Library and Archive Fact sheet 17 – Weather observations over land. Exeter: SEPA. Available at: http://www.metoffice.gov.uk/learning/library/publications/factsheets (last accessed: 22nd November 2016).
Vidon, P.G., 2015. Field hydrologists needed: a call for young hydrologists to (re)-focus on field studies. Hydrol. Process. 29, 5478–5480. http://dx.doi.org/10.1002/hyp.10614.
Walker, D., Forsythe, N., Parkin, G., Gowing, J., 2016. Filling the observational void: scientific value and quantitative validation of hydrometeorological data from a community-based monitoring programme. J. Hydrol. 538, 713–725. http://dx.doi.org/10.1016/j.jhydrol.2016.04.062.
Wang, L.P., Ochoa-Rodríguez, S., Onof, C., Willems, P., 2015. Singularity-sensitive gauge-based radar rainfall adjustment methods for urban hydrological applications. Hydrol. Earth Syst. Sci. 19, 4001–4021.
Wentworth, J., 2014. Environmental Citizen Science POST Note. Houses of Parliament: Parliamentary Office of Science and Technology POST Number 476 12th August 2014.
Zhu, D., Xuan, Y., Cluckie, I., 2013. Statistical analysis of error propagation from radar rainfall to hydrological models. Hydrol. Earth Syst. Sci. 17, 1445–1453.
Zhu, D., Xuan, Y., Cluckie, I., 2014. Hydrological appraisal of operational weather radar rainfall estimates in the context of different modelling structures. Hydrol. Earth Syst. Sci. 18, 257–272.