Journal of Hydrology 547 (2017) 664–677

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Integrated surface-subsurface model to investigate the role of groundwater in headwater catchment runoff generation: A minimalist approach to parameterisation

Pertti Ala-aho *, Chris Soulsby, Hailong Wang, Doerthe Tetzlaff

Northern Rivers Institute, School of Geosciences, University of Aberdeen, AB24 3UF, UK

ARTICLE INFO

Article history:
Received 30 August 2016
Received in revised form 19 December 2016
Accepted 12 February 2017
Available online 20 February 2017
This manuscript was handled by Prof. P. Kitanidis, Editor-in-Chief, with the assistance of Roseanna M. Neupauer, Associate Editor

Keywords:
Integrated modelling
Streamflow generation
Groundwater
Catchment modelling
Model calibration

ABSTRACT

Understanding the role of groundwater for runoff generation in headwater catchments is a challenge in hydrology, particularly so in data-scarce areas. Fully-integrated surface-subsurface modelling has shown potential in increasing process understanding for runoff generation, but high data requirements and difficulties in model calibration are typically assumed to preclude their use in catchment-scale studies. We used a fully integrated surface-subsurface hydrological simulator to enhance groundwater-related process understanding in a headwater catchment with a rich background in empirical data. To set up the model we used minimal data that could be reasonably expected to exist for any experimental catchment. A novel aspect of our approach was in using simplified model parameterisation and including parameters from all model domains (surface, subsurface, evapotranspiration) in automated model calibration. Calibration aimed not only to improve model fit, but also to test the information content of the observations (streamflow, remotely sensed evapotranspiration, median groundwater level) used in calibration objective functions. We identified sensitive parameters in all model domains (subsurface, surface, evapotranspiration), demonstrating that model calibration should be inclusive of parameters from these different model domains. Incorporating groundwater data in calibration objectives improved the model fit for groundwater levels, but simulations did not reproduce well the remotely sensed evapotranspiration time series even after calibration. Spatially explicit model output improved our understanding of how groundwater functions in maintaining streamflow generation primarily via saturation excess overland flow. Steady groundwater inputs created saturated conditions in the valley bottom riparian peatlands, leading to overland flow even during dry periods. Groundwater on the hillslopes was more dynamic in its response to rainfall, acting to expand the saturated area extent and thereby promoting saturation excess overland flow during rainstorms. Our work shows the potential of using integrated surface-subsurface modelling alongside with rigorous model calibration to better understand and visualise the role of groundwater in runoff generation even with limited datasets.

1. Introduction

Understanding streamflow generation is a fundamental challenge in hydrology. The biggest source of uncertainty resides where most of the flow takes place and vegetation taps water for transpiration: the shallow subsurface (McDonnell, 2013; Sklash and Farvolden, 1979). The challenge has been tackled by a plethora of mathematical hydrological models with very different process conceptualisations ranging from parsimonious conceptual models (Bergstrom, 1976; Beven and Kirkby, 1979) to complex and highly parameterised physically-based simulators (Aquanty, 2016; Kollet and Maxwell, 2006) – all working towards constraining the fundamentals of streamflow generation.

Physically-based hydrological models integrating flow processes in the surface and subsurface provide a promising tool to test concepts of runoff generation and have been successfully used to reveal processes responsible for streamflow generation across scales (Frei et al., 2010; Liggett et al., 2015; Park et al., 2011; Weill et al., 2013). One of their main advantages is the ability to make use of various field-based observations in both constructing the models and assessing their performance. Their physically-

* Corresponding author.
E-mail address: pertti.ala-aho@abdn.ac.uk (P. Ala-aho).

http://dx.doi.org/10.1016/j.jhydrol.2017.02.023
0022-1694/

based equations and spatially distributed parameterisation creates a consistent framework not only to mimick observed catchment behaviour, but also to formulate and test hypotheses for potential hydrological behaviour beyond existing measurements (Bolger et al., 2011; Hwang et al., 2015; Jones et al., 2006; Maxwell and Condon, 2016).

While acknowledging their apparent strengths, integrated surface-subsurface models are often criticised for the extensive data needed in parameterisation and numerical complexity leading to long model runtimes and numerical instabilities, which combine to complicate the process of catchment scale model calibration (Beven, 2002; Doherty and Christensen, 2011). Consequently, calibration of such models often resorts to 'manual' trial and error approaches (Ala-aho et al., 2015; Jones et al., 2008; Li et al., 2008), or no calibration at all (Bolger et al., 2011). Recent work has applied parameter optimisation routines, commonly the PEST software suite (Doherty, 2010) to calibrate integrated surface-subsurface models. These studies have calibrated parameters pertaining to subsurface and surface domains of catchment-scale models (Maneta and Wallender, 2013; Verbist et al., 2012; Wildemeersch et al., 2014; Yang et al., 2015) or subsurface and ET domains in more simple model configurations (Brunner et al., 2012; Schilling et al., 2014). To the best of our knowledge, studies incorporating rigorous model calibration that includes parameters from all model domains (surface, subsurface, ET), and thereby explicitly acknowledge the integrated nature of the simulated processes, are extremely rare. Overall, while catchment scale integrated surface-subsurface model applications are increasing, remote data-scarce headwater catchments are under-represented in the literature.

We propose that one can set up and calibrate a fully-integrated physically based model for a gauged basin with almost equally low data requirements as for any conceptual hydrological model. This is achieved by model calibration using the PEST framework (Doherty, 2010) where we use observations that could be expected in most gauged catchment and which are pertinent to all hydrologically relevant model domains (surface, subsurface and ET). Calibration is facilitated by novel simplifications for subsurface hydraulic and evapotranspiration parameterisation. To test which commonly available hydrological observations data are beneficial to improve simulation results, we perform a sequential calibration with incrementally more data-rich multi-component objective functions (cf. Birkel et al., 2014). We hypothesise that when introducing evapotranspiration and groundwater data in the calibration process, the model performance with respect to these variables would be improved and model parameters pertinent to these domains would be more identifiable. The study aims to achieve the classical proof of hydrological model success: a good hydrograph fit, but simultaneously produce physically meaningful output and an improved understanding of subsurface flow and runoff generation processes with minimum data requirements.

Our specific objectives are:

- To produce a spatially explicit and physically sound conceptualisation of how the subsurface operates in sustaining and generating streamflow in a glaciated headwater catchment.
- To do so by establishing a simple, novel, parameterisation inclusive of surface, subsurface, and ET domains of a fully-integrated model to facilitate a rigorous model calibration.
- Reveal and compare parameter sensitivities across all model domains (surface, subsurface, ET) for an integrated surface-subsurface model, an exercise not done before.
- Determine what data are useful to include in model calibration by examining changes in model responses and parameter sensitivities for different calibration objective functions.

2. Materials and methods

2.1. Study site

The study catchment, the Bruntland Burn, is a montane catchment (3.2 km^2) in the Scottish Highlands. The annual average precipitation approximates 1000 mm, with low seasonality, partitioning to around 600 mm annual runoff and 400 mm evapotranspiration. The catchment has an annually reoccurring but usually minor snow influence; typically accounting for<5% of annual precipitation. Recent work at the site has developed an understanding of the water storage distributions, flow paths and mixing processes by successfully combining tracer studies, hydrometric measurements and conceptual modelling approaches of various complexities (Birkel et al., 2011a,b; Blumstock et al., 2015; Huijgevoort et al., 2016; Soulsby et al., 2015). The above mentioned references contain detailed descriptions of the catchment characteristics, with a brief summary provided below.

Topographic relief in the catchment spans from 250 m (above sea level) at the outlet to 530 m at highest point on the south-east hilltop. Bedrock geology is low permeability unweathered granite and metamorphic rocks. Unconsolidated material overlying the bedrock covers ~ 70% of the catchment and consists of glacial drift deposits, primarily undifferentiated till with a silty and sandy matrix. The glacial legacy of the area has left the catchment with a flat and wide valley bottom (slope < 3^0) where geophysics estimate the drift deposits to reach 30–35 m depth, thinning out towards the steeper hillslopes (slopes > 8^0). In the valley bottom riparian areas, mineral sediments are covered by organic peat soils along with shallower peaty gleys. The soils on the hillslopes consist of podzols. Vegetation on the hillslopes is dominated by heather (*Calluna* sp. and *Erica* sp.) moorland with minor scots pine (*Pinus sylvestris*) coverage. Riparian peatland areas are dominated by *Sphagnum* spp. mosses and some grasses (*Molina caerulea*).

2.2. Integrated surface-subsurface model setup

2.2.1. Conceptual model

Based on previous work, this study simplified the landscape into two units: organic peat soils covering the riparian areas and mineral glacial drift soil/sediments mantling the hillslopes and underlying the peat soil in the valley bottom. Within the numerical model this partitioning was used to parameterise (1) the overland flow domain differently in the riparian peatlands and hillslopes and (2) subsurface flow domain differently to mineral glacial drift sediments and organic riparian peat soils. Bedrock underlying the glacial drift was assumed impermeable and comprised a hydrologically inactive no-flow boundary.

2.2.2. Simulation code HydroGeoSphere

We chose HydroGeoSphere (HGS) as the physically based simulation code because of its capabilities in representing the main hydrological processes of interest in our study – surface and subsurface flow and evapotranspiration – in a physically based manner (Aquanty, 2016). HGS uses a control-volume finite element approach to solve the interconnected flow equations in the surface and subsurface simultaneously for each time step. Flow in the subsurface is solved by a variably saturated Richard's equation and surface flow with a diffusion-wave approximation of the Saint-Venant shallow water flow equations. The model also integrates an evapotranspiration module to simulate water uptake from model nodes closest to the ground surface. The model allows water to be exchanged freely between the different flow domains allowing hydraulic gradient throughout the model to govern the flow processes. To solve the equations above, the model domain needs

to be discretised to finite elements. Elements can make up an irregular and complex domain allowing great flexibility in spatial refinement of the model. Full technical details of the model are given in (Aquanty, 2016). The code has been successfully used in simulating combined flow of water, heat and solute across catchment and hillslope scales (Ala-aho et al., 2015; Brookfield et al., 2009; Frei and Fleckenstein, 2014; Jones et al., 2008; Liggett et al., 2015).

2.2.3. Model mesh

The calculation mesh for the model was composed of unstructured finite elements. A 2-D mesh consisting of 3-node triangular elements (Fig. 1) was first created for the ground surface. The mesh refinement was focused to areas where GW-SW interaction was expected to be most active. Mesh resolution was increased in the riparian areas. To identify riparian areas we used SAGA wetness index, calculated in Lessels et al. (2015), which is well suited to predict wetness of flat valley floors that the more commonly used wetness indices, such as topographic wetness index (TWI) (Boehner et al., 2002). Streams were introduced as line features in the mesh building process. Nodes along the stream bed were assigned to overlap with highest value of TWI derived from LIDAR based DEM (Lessels et al., 2015). This ensures that the nodes along the stream bed are located at or near topographically lowest areas giving the mesh a continuously sloping streambed which reduces computational burden in the model. The mesh was further refined along the stream channel producing on average 9.6 m segments for elements connected to the stream. Finally, the mesh was adjusted so that a calculation node was assigned to each GW observation well location (Fig. 1) and was refined in the vicinity of the wells. The final 2-D model mesh consisted of 12844 elements.

Stacking layers of triangular 6-node elements below the surface gave the model its vertical dimension. Vertical layering was refined near the ground surface to more accurately simulate the GW-SW exchange flux and water uptake by evapotranspiration. Starting from the ground surface, the first 0.2 m consisted of 4×0.05 m layers, a profile from 0.2 to 1.0 m consisted of 8×0.1 m layers, from 1 to 2 m 2×0.5 m layers. For the remainder of the vertical extent, i.e. from 2 m below ground surface down to bedrock, a sediment profile was discretised to five equally spaced layers.

2.2.4. Model boundary and initial conditions

Bedrock was treated as a hydrologically inactive unit comprising the model bottom no-flow boundary. Elevation for this boundary was interpolated primarily using data reported in Soulsby et al. (2016) (Fig. 1) who applied geophysical (electrical resistivity) techniques to estimate drift depth in four transects across the valley bottom. They report maximum drift depths of 20–40 m in the valley bottom, thinning out to 2–5 m near the hillslopes. One additional transect was surveyed in November 2015 near the catchment outlet in order to estimate the constant head boundary (Fig. 1) to supplement the data in Soulsby et al. (2016). Drift depth was estimated for discrete points with 25 m spacing along the transects. Finally, we estimated drift depth for two supplementary transects to steer the interpolation procedure in the valley bottom

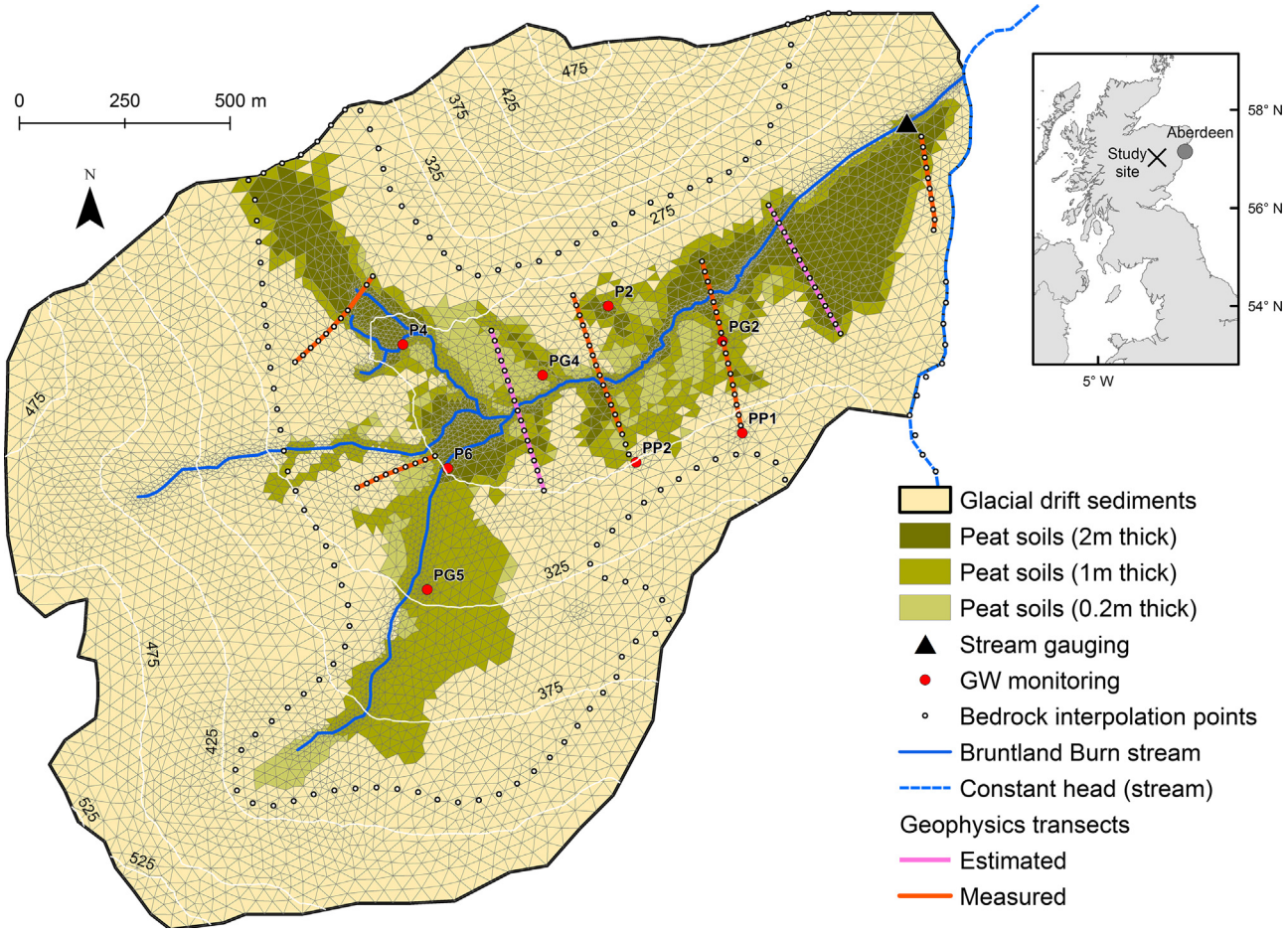

Fig. 1. Bruntland Burn study site where colour fill indicates the different landscape units; light brown for hillslopes, different shades of green for riparian peatlands. White circles show points used in interpolating bedrock surface.

areas with data gaps. Geological maps of the area were used to outline the outer margin of drift sediment depth (see 'bedrock interpolation points' in Fig. 1). Drift depth at the margin was assigned a value of 2.1 m and discrete points with 50 m spacing were set along the drift margin to be used in bedrock surface interpolation. 'Topo to raster' utility in Arcgis software utility (ESRI, 2011) was used to interpolate the bedrock surface. The method produced a continuous surface which was deemed more realistic compared to other geostatistical interpolation methods such a Kriging, which typically need a higher point density. Hillslopes not classified as glacial drift were assigned a bedrock depth of 2.1 m below ground surface.

In addition to the model lower boundary, a no-flow boundary was assigned for the porous media domain around the perimeter of the catchment except for a constant head boundary defined along regional stream specified as the stream water elevation (Fig. 1). The overland flow domain was given a 'critical depth' boundary condition for nodes on the model top boundary. The critical depth boundary allows water to leave the model domain as surface flow with respect to calculated water depth at the domain boundary, and is typically used in catchment scale simulations. The top of the domain was constrained with a combination of rain and potential evapotranspiration boundary conditions operating on a daily time step. Though the input data was of daily resolution, we used an 'adaptive timesteps' –scheme that assists model convergence by reducing or increasing the solution time step based on the transient behaviour of the system.

Model input data for daily precipitation and potential evapotranspiration are presented in Fig. 3. Precipitation data were the same as in Huijgevoort et al. (2016) based on altitudinally-adjusted inverse distance weighting from gauges surrounding the catchment. The simulation period for the year 2013 was unusual because of a 10-year return period drought in the summer of 2013 (Blumstock et al., 2015). The drought resulted in an extensive drying of the catchment in June and July followed by a brief rewetting, then a further dry spell until autumn, which is reflected in the discharge time-series (Fig. 3). This made the period particularly interesting for our study by creating unusually high evaporative conditions likely making the ET signature we set to simulate more prominent.

The time series for potential (ETp) and actual (ETa) evapotranspiration were obtained from MODIS remote sensing products providing a sum of eight day evapotranspiration (Mu et al., 2011). The study catchment encompassed 9 pixels in the dataset, which were averaged for each time step. Total depth of MODIS ETa and ETp for the year 2013 were 477 and 738 mm, respectively.

Because the simulations were operating on a daily time step, the 8-day sum of ETp needed breaking down to individual days. This was achieved by scaling the data according to a daily Penman-Monteith (P-M) based ETp estimate for the catchment given in Huijgevoort et al. (2016). The scaling resulted in daily ETp time series which respected the total ETp volume of the MODIS 8-day product. The 8-day ETa product was used in the model calibration as a part of the calibration objective function. The disaggregated MODIS ETp product was favoured instead of the local daily P-M ETp because we wanted to have model inputs for ETp consistent with the calibration dataset of ETa.

To attain initial conditions for transient model runs, the model was first run into steady state using the long term annual runoff value of 600 mm a^{-1}. After reaching steady state, the model was forced with climate data for the spin-up period between 13 Jan 2013-01 May 2013 (Fig. 3) with five iterations. Wet winter conditions during the period led to a near saturated catchment state where the storage changes in the subsurface from start to end of the spin-up were minor and the model state on 1 May was used as initial conditions for model calibration runs (see Section 2.3).

2.2.5. Model parameterisation

Organic peat soils are typically characterised by low hydraulic conductivity, and in the study area, riparian zone peat layers can reach up to several meters depth (Soulsby et al., 2016). The peat extent was obtained from Hydrology Of Soil Type (HOST) classification maps (Boorman et al., 1995), and a 0.2 deep layer of peat was assigned throughout the areas classified as peat. Areas of deeper (2 m) peat were estimated to occur within the 'raw peat' –category in the HOST classification maps. Intermediate 1-m peat depth was estimated form aerial images to occur in areas between heather covered hummocks and deep peat zones. Peat hydraulic properties presented in Table 1 were estimated from literature values. In peat parameterisation we did not follow the classical acrotelm/catotelm model with more permeable near-surface peat layers, in order to reduce the number of parameters in calibration and expecting the surface domain parameterisation to implicitly account for such near-surface flows.

The glacial drift is characterized by a medium-fine textured silty-sand matrix (Blumstock et al., 2015) and the clay content of the drifts is low due to the rock type, cool temperatures and short post-glacial history (Soulsby et al., 2016). We used the ROSETTA lookup table, i.e. the ROSETTA model H1 (Schaap et al., 2001), that reports average hydraulic parameters for different soil textural classes to get calibration boundaries for the soil hydraulic conductivity. We included textural classes (n = 8) with less than 40% clay

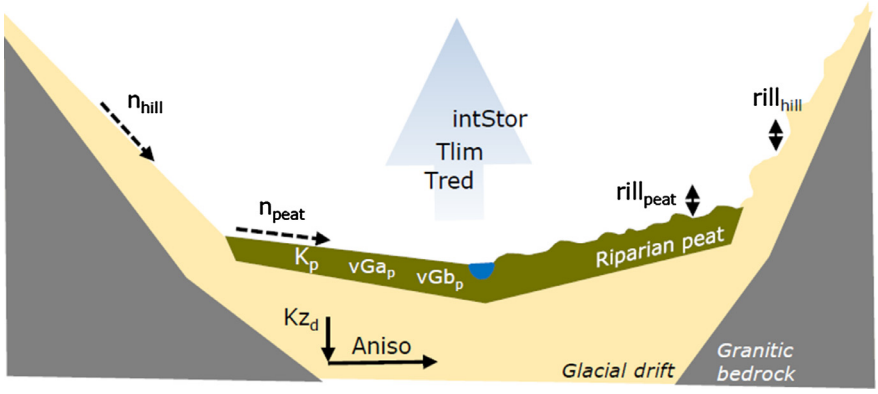

Fig. 2. Model parameters under calibration shown in a conceptual cross-section of the model domain. Subsurface: Kz_d = drift vertical hydr. cond., Aniso = Anisotropy ratio of drift vertical and horizontal hydraulic cond., K_p = Peat hydraulic cond., vGa_p = van Genuchten parameter α for peat, vGb_p = van Genuchten parameter β for peat. Surface: n_{hill} & n_{peat} = Mannings roughness coefficient for hillslopes and peatlands respectively, $rill_{hill}$ & $rill_{peat}$ = rill storage height for hillslopes and peatlands respectively. Evapotranspiration: intStor = interception storage, Tlim = soil water content where ET becomes water limited, Tred = maximum ratio ET/PET.

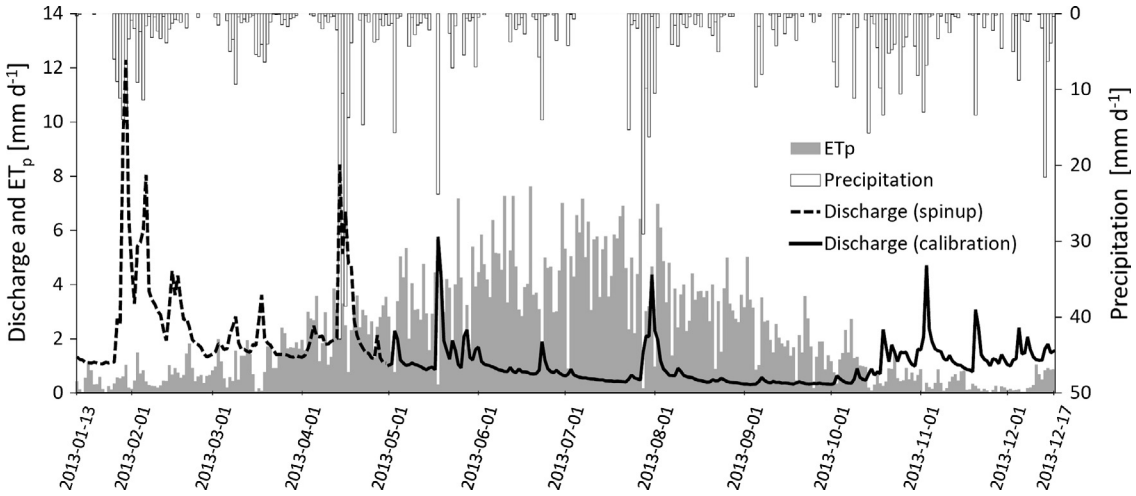

Fig. 3. Daily potential evapotranspiration (ETp) and precipitation timeseries used as driving data for model spinup and calibration periods. Solid line is the daily discharge used in model calibration and dashed line shows stream flow during model spin-up period in winter and spring.

Table 1
Model parameters that deviate from HGS model default values (Aquanty, 2016).

Parameter	Initial value	Range	Initial value	Range
Porous media domain	Drift		Peat	
Vertical hydraulic conductivity [m s-1]	2E-6	9.47E−7–1.22E−5 [2]*	3.6 E-7	9.47E−7–9.47 E−9*
Anisotropy [-]	5	1–50 [3,4]*	–	
Porosity [–]	0.25	Fixed	0.9 [5]	Fixed
Specific storage [1 m^{-1}]	2 E-4 [1]	Fixed	1E-1	Fixed
vG residual saturation [-]	0.165 [2]	Fixed	0.1 [5]	Fixed
vG α [m^{-1}]	1.47	0.95–2.71 [2] (tied)	13.17 [5]	5 – 25 [5]
vG β [–]	1.51	1.42–1.72 (tied) [2]	1.25 [5]	1.2 – 1.4 [5]
Minimum relative permeability [–]	1E-2	Fixed	1E-3	fixed
Overland flow domain	Hillslope		Riparian	
Manning's n [$m^{-1/3}$ s]	6.7	1–7 [6, 9]*	6.3	1–7 [6, 9]*
Rill Storage height [m]	0.08	0.01–0.5 [6, 7] *	0.4	0.01–0.5 [6, 7] *
Coupling length [m]	0.01 [8]	Fixed	0.01 [8]	fixed
ET domain	Initial value		Range	
Canopy storage parameter c_{int} (=intStor) [m]	1E-4		5E−5–5E−5*	
Transpiration fitting parameter c1	0		Fixed	
Transpiration fitting parameter c2 (=Tred)	0.5		0.1 –1	
Transpiration fitting parameter c3	1		Fixed	
Wilting point θ_{wp}	0.17		Fixed	
Field capacity θ_{fc} (=Tlim)	0.3		0.2–0.5	
Oxic limit θ_o	1		Fixed	
Anoxic limit θ_{an}	1		Fixed	
Leaf Area Index (LAI)	1		Fixed	
Evaporation limiting sat (θ_{e1})	100		Fixed	
Evaporation limiting sat (θ_{e2})	101		Fixed	
Root depth L_m [m]	0.5 [9]		Fixed	
Evaporation depth B_{soil} [m]	0.2		Fixed	

[1] (Shaver, 1998); [2] (Schaap et al., 2001) [3] (Sudicky et al., 2010); [4] (Krabbenhoft et al., 1990); [5] (Päivänen, 1973); [6] (Liggett et al., 2015) [7] (Frei and Fleckenstein, 2014); [8] (Ebel et al., 2009), (Dunn and Mackay, 1995).
*means that the parameter is log-transformed in the calibration process.

separate (excluding also clean sand), of which clay loam had the lowest and loamy sand the highest hydraulic conductivity, 9.47E−7 and 1.22E−5 m s^{-1}, respectively. These ranges are consistent with those measured in shallow pump tests in the shallow drift (Malcolm et al., 2004) and were used for possible hydraulic conductivity values for glacial drift in the model calibration (Table 1).

To reduce the number of parameters in the calibration we used the ROSETTA data also to estimate the parameters α [1/m] and β [–] for the van Genuchten model describing the soil pressure – saturation relationship in the drift sediment. We built a linear regression with the ROSETTA data for the 8 textural classes using log-transformed hydraulic conductivity as the explanatory variable. Using the model, high hydraulic conductivities were associated with high values of α and β. Linear models resulted in modest R^2 values of 0.32 for α and 0.49 for β. Despite the relatively poor explanatory power, the linear model achieved the purpose of having a pressure-saturation relationship prone to retain less water (under a given negative pressure) in sediments with high hydraulic conductivity than in those with low hydraulic conductivity, as is typically observed.

In the overland flow domain, a distinction in the parameterisation was made between riparian peatland, hillslopes and the

stream channel. Areas with peat soils and glacial drift sediments (Fig. 1) were assigned riparian and hillslope parameters, respectively (Table 1). Overland flow elements sharing a segment with a predefined stream channel were given stream channel properties. Manning's roughness and rill storage height are the parameters primarily responsible for water storage and transmission in the overland flow domain. These parameters were specified separately for hillslope and peatlands and were included in the model calibration with parameter values and ranges reported in table 1. Stream channel elements were assigned a Manning's n of 0.035 $m^{-1/3}$ s (Partington et al., 2013).

Simulating evapotranspiration in HGS includes processes for interception, soil evaporation, transpiration and open water evaporation, which in combination try to meet the atmospheric water demand (ETp) calculated externally (see Section 2.2.4) and provided for the model. Interception storage and open water evaporation occur at ETp rate, given there is enough water in the interception storage or that the hydraulic head is above ground level in computational node in question, respectively. Soil evaporation and transpiration were calculated according to the model suggested by Kristensen and Jensen (1975) where empirical functions are used to reduce ETa to a factor of 0–1 from ETp depending on the soil water conditions and vegetation characteristics.

Importantly, we sought to reduce the number of ET-related parameters under calibration to a minimum while allowing processes such as soil water deficit to constrain ETp down to ETa. Firstly, the ET domain in the model was parameterised spatially uniformly without distinction into riparian and hillslope areas or vegetation zones, thus assuming transpiration and interception by the localised tree canopy cover minor at the catchment scale. Secondly, parameters were reduced using unconventional parameterisation of the empirical functions in Kristensen and Jensen (1975). Parameters for the ET domain are given in Table 1, with the following motivation behind choosing the unconventional values for some of the parameters:

- The soil evaporation routine was deactivated to simplify the process by which water is removed from the porous media domain by evapotranspiration (high values for θ_{e1} and θ_{e2}).
- Leaf area index was set to unity, allowing only one parameter (transpiration fitting parameter C2, from here on Tred) to control the total fraction of ETa from ETp and parameter c_{int} (from here on intStor) to control the total interception storage.
- Oxic and anoxic limits for transpiration were set to unity, in order to allow evapotranspiration from porous media close to saturation that would we typically done by the soil evaporation routine (disabled here for process simplification).

The above formulations and assumptions reduces the empirical functions and number of parameters regulating the amount of ETa / ETp from Aquanty (2016):

$$ETa = f_1(C_1C_2, LAI) \cdot f_2(\theta_{wp}, \theta_{fc}, \theta_o, \theta_{an}, C_2) \cdot RDF(L_m) \cdot (Ep - E_{can}(LAI, c_{int})) + \alpha^*(\theta_{e1}, \theta_{e2}) \cdot (Ep - E_{can}(LAI, c_{int}) - T_p) \cdot EDF(B_{soil}) \quad (1)$$

down to:

$$ETa = T_{red} \cdot f_2(T_{lim}) \cdot RDF[Ep - E_{can}(intStor)] \quad (2)$$

where E_{can} is interception evaporation. For details about functions f_1, f_2, RDF, E_{can}, $\alpha*$ and EDF see (Aquanty, 2016), explanations and values for parameters pertaining to above functions are given in table 1.

2.3. Model calibration using PEST and high performance computing

2.3.1. Constructing three different objective functions

An innovative aspect of our model calibration was to utilise parameters from all three model domains in the calibration - five parameters pertaining to subsurface flow, four to surface flow, and three to the evapotranspiration domain (Fig. 2). Model-independent parameter estimation software (PEST) was used to optimise the fit between field observations and simulations in the weighted least squares sense (Doherty, 2010). A second novel aspect of our model calibration was to do three subsequent calibrations to test how the model performance changed and optimal parameter values and parameter sensitivities varied, when data pertinent to different flow domains was added to the calibration objective function. All of the three calibrations included streamflow data, but calibrations (2) and (3) introduced observations of ET and groundwater levels.

(1) Q calibration: only daily streamflow data was used in the objective function. Stream levels were gauged near the stream outlet at 15 min intervals, translated to discharge using a rating curve, and averaged to daily values (Figs. 1 and 3). Simulated streamflow was extracted at the overland flow domain model nodes at the gauging location and averaged to daily values.
(2) ET calibration: in addition streamflow, MODIS 8-day sum of actual evapotranspiration was included in model calibration (see Section 2.2.4). Simulated total ETa flux leaving the catchment was aggregated to 8-day values for the same time periods as the MODIS data.
(3) GW calibration: in addition to streamflow and ET data, the median of measured daily groundwater levels (Blumstock et al. (2016)) below the ground surface at eight groundwater wells was added to the objective function (Fig. 1). Simulated values were extracted from model nodes located at the observation wells. Using GW medians adhered the requirement of low resolution data that could be expected to exist in moderately monitored catchments.

2.3.2. PEST calibration setup

We used utility program (*pwtadj1*) provided with the PEST software (Doherty, 2010) to adjust the individual observation weights in a manner that each observation group (discharge, ET, median GW levels) had approximately the same influence in the history matching process (Doherty and Welter, 2010). This ensures that observation groups with more observations (streamflow vs. GW) or observations with higher magnitude in numbers (ET vs. streamflow) do not dominate the calibration. When incorporating GW wells in calibration round 3, we avoided the problem of higher water table depth numbers in some wells dominating the fitting process, by using a cubic root transformation to bring the absolute values among different wells closer together.

Observations in the objective function (224 streamflow observations, 28 ET observations and 8 GW observations) greatly exceeded the number of parameters under calibration (12) we considered the calibration a well-posed inverse problem, though some of the data in streamflow timeseries may bear redundant data due to autocorrelation (Doherty, 2010). PEST was run in 'singular value decomposition' – mode because of its robustness in guiding the numerical solution. In the calibration process, parameters with calibration ranges spanning order(s) of magnitude were log-transformed (see table 1) to ensure a numerically more robust calibration (Doherty, 2010). Because outputs from spatially distributed fully integrated models are highly nonlinear with respect to changes in parameter values, PEST calibration is an iterative process. For each iteration, PEST needs to establish a 'Jacobian

Matrix', by calculating derivatives of model outputs with respect to parameter values to record the relationship between model parameters and model outputs of interest. This requires multiple transient runs of the model with marginally varied parameter values. Each individual model run consists of a spin-up period to stabilise the storages for the tuned parameter values, and a calibration period for which the model output is extracted for the matching process (see Fig. 3).

We extracted parameter values for each iteration to learn how extensively the calibration process explored the parameter space. Moreover, PEST gives outputs for parameter sensitivities based on the Jacobian matrix solved for each iteration. We examined the relative composite parameter sensitivities for the "best fit" model for each calibration run (Q, ET, GW) to find the most sensitive model parameters and to see if adding components to the objective function resulted in changes of parameter sensitivities. For each calibration run we scaled the composite sensitivities between 0 and 1 by dividing all sensitivities by the most sensitive parameter value to make the sensitivities comparable across calibration runs, as the absolute relative sensitive values differ between calibration runs. In PEST it is possible to extract additional information about statistics of the model calibration, such as parameter correlations or data worth. However, in the 'singular value decomposition' – regularisation mode we used this information was not recorded and we used parameter sensitivities to compare different calibration runs.To hasten the process we used PEST's Parallel PEST functionality to parallelise model calibration, namely the model runs pertaining to calculation of the Jacobian matrix, utilising high performance computing provided by the University of Aberdeen IT service. In addition, the numerical solution of the HGS was split to 4 cores using HGS's parallelisation functionalities. This allowed us to reduce the calibration time from 20 to 60 days on a desktop PC (64-bit Windows OS, 3.3 GHz processor, 8 GB RAM) down to 2–3 days.

3. Results

3.1. Parameter sensitivities and evolution in model calibration

Parameters for peat and glacial drift hydraulic conductivities were consistently most sensitive with respect to the model output (Fig. 4). However, most parameters in surface flow and ET domain were also sensitive, indicating that parameters from all domains were appropriate to include in the calibration – an aspect that is typically lacking in automated calibration of catchment scale integrated surface-subsurface models. Generally the relative sensitivity of parameters were similar across the different calibration rounds, indicating that adding observations relating to the ET or subsurface domain was not directly reflected in the sensitivity of parameters most closely associated with the domain in question.

Values that model parameters took during the calibration are shown in Fig. 5 to show the calibrated parameter values and demonstrate the extent to which the parameter space was explored in PEST. Most of the parameters, typically the most sensitive ones (e.g. Kz_d, intStor, n_{hill},) did not migrate far from the initial values. Q and ET calibrations tended to land on similar optimal parameter values, whereas some parameters in the GW calibration (K_p, Tlim, Tred, $rill_{peat}$) deviated from the other two runs.

3.2. Comparison of simulated and measured hydrological variables

The model reproduced the main characteristics of the stream hydrograph especially in the moderate flow range for all calibration setups (Fig. 6). Though streamflow dynamics were generally captured, simulated streamflow was biased to underestimating highest flows and overestimating flows in prolonged drought. Different calibration objectives resulted in only minor differences in simulated hydrographs in terms of response dynamics, the only notable difference being lower simulated streamflow in the GW calibration during catchment rewetting in late autumn. Direct exfiltration through stream bed cells accounted for less than 25% of simulated flow even during minimum flows, again with minor differences between calibrations.

Remotely sensed MODIS estimates for ETa and simulation outputs for all calibration rounds showed a poor fit (Fig. 7). Particularly the timing of highest ETa did not coincide in simulations and MODIS data, where MODIS data had highest ETa during mid-summer (July), whereas simulated ETa peaked in early summer (June) and again after mid-summer catchment rewetting (see Fig. 6) in early August. After the highest evaporative period, from September onwards the fit between simulations and MODIS ETa product was satisfactory. Contrary to our original hypothesis, including ETa to the model calibration (ET calibration) did not notably improve ETa fit. Only the GW calibration deviated from the other two resulting in considerably higher ETa values.

Groundwater levels were reasonably well simulated in the riparian areas (P and PG wells, see Fig. 1) where the water table remains close to ground surface (Fig. 8). In the riparian wells, the main discrepancy was that for wells P2 and PG4 simulated water tables were above the ground surface, whereas the measured water table remained below it. However simulations in the hillslope (PP1 and PP2) were biased towards deeper water tables in the model outputs, whereas observed water table median remained within 1 m below the ground. In contrast to the ETa calibration where adding ETa data did not improve the ETa fit, introducing GW data in model calibration improved fit of the shallow riparian wells except for well P6 and hillslope wells.

3.3. Spatial insights into runoff generation

The spatially distributed explicit model output allows evaluation of where the calibration scheme had its successes and failures. Fig. 9 presents snapshots of model output at the end of a dry period and after catchment rewetting. Supplementary Material (Vid. 1–4) extends the presentation of same data with animations showing the simulated catchment response over a 30-day period where the catchment is rewetted after a dry spell. The output demonstrates how the surface flow domain is highly dynamic with expansion and contraction of the saturated area and overland flow in response to varying precipitation inputs (Fig. 9 left column, Vid. 1). During the dry periods the majority of the riparian peatlands remain saturated (pink colour), and flow in the surface flow domain (different shades of blue colour) takes place primarily in the stream channel and some areas immediately adjacent to the stream. In response to rainfall events, the saturated areas expand and in areas of saturation the distribution and velocity of overland flow across the surface is notably increased.

During the drier periods groundwater sustains streamflow by direct channel exfiltration (Fig. 6) but equally importantly via persistent groundwater flux to riparian areas in the valley bottom (Fig. 9, Vid. 2). This groundwater inflow is responsible for the saturating the riparian areas and facilitating saturation excess overland flow, which is apparent from the close resemblance of spatial patterns of GW exfiltration and saturated areas in surface domain (Fig. 9). Despite its importance in dry period runoff generation, simulations suggest that groundwater in the deep valley bottom drift deposits has low flow velocities and remains relatively immobile through rain events (Fig. 10, Vid. 3).

Water limitation and energy inputs have a strong control on simulated ETa, even with simplified ET domain parameterisation (Fig. 9, Vid. 4). At the end of dry period with high energy inputs

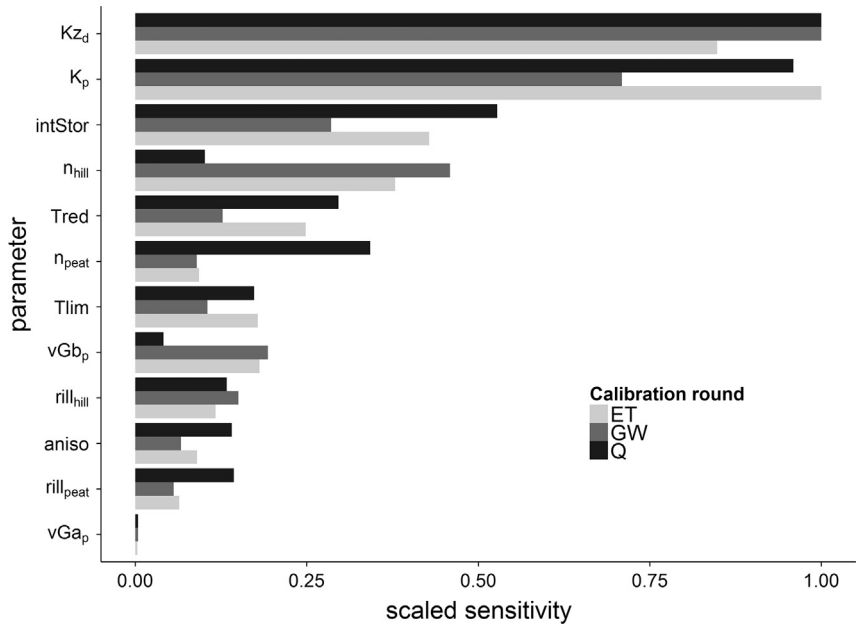

Fig. 4. PEST output for relative composite parameter sensitivities in different calibration rounds. Values are scaled from 0 to 1 by dividing the all parameter sensitivities with the highest parameter sensitivity value (Kz_d or K_p) in each calibration run.

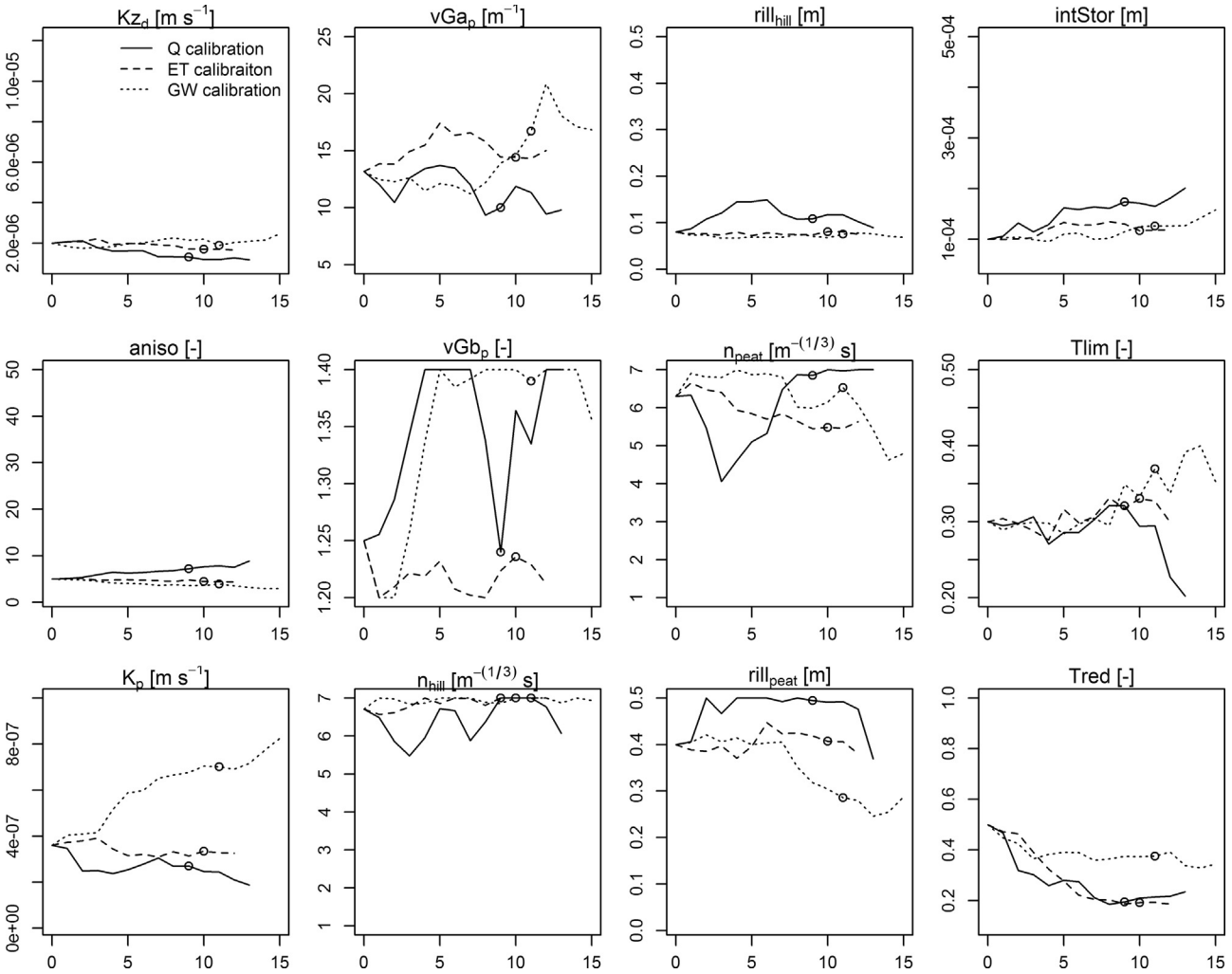

Fig. 5. Evolution of model parameters during the three different calibration rounds, x-axis showing the iteration number. The open circles denote the parameter value of the best fit simulation for each calibration. Y-axis in show the range the parameters were allowed to take during optimisation.

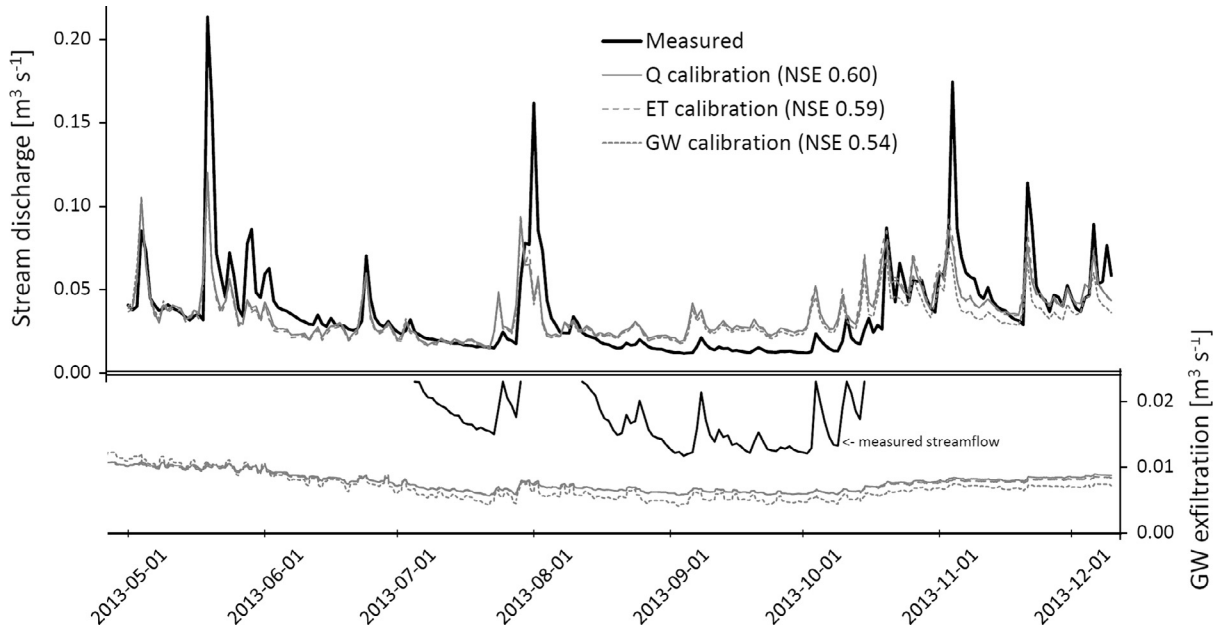

Fig. 6. Measured stream hydrograph and optimised 'best fit' simulated streamflow from Q, ET and GW model calibration rounds are shown in upper panel. Lower panel demonstrates the aggregated groundwater exfiltration directly through model elements that comprise the predefined stream channel in these different model calibrations.

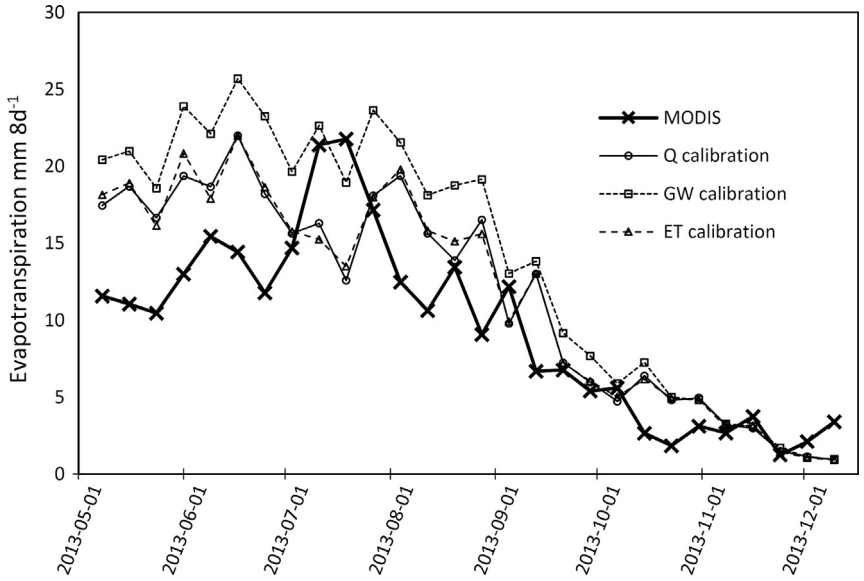

Fig. 7. Time series of 8-day sum actual evaporation (ET_a) of the MODIS product and simulation results for different calibration runs.

ET is water limited on the hillslopes (rate 1–2 mm d^{-1}) but exhibits high rates (5–4 mm d^{-1}) in the saturated riparian areas. After catchment rewetting the spatial organisation of ETa response is more varied across individual days (Vid. 4), but for energy limited conditions ET is typically more uniform across the catchment with moderate rates of 2–3 mm d^{-1}, as demonstrated in Fig. 9.

4. Discussion

4.1. Calibrated parameter values and sensitivities

To our knowledge this study is among the first to include parameters equally from all flow domains (ET, surface and subsurface) in a fully integrated modelling framework with a rigorous model calibration scheme. Sensitivity analysis revealed that hydraulic conductivities for both peat and mineral soil were the most sensitive parameters, but others relating to the subsurface domain (van Genuchten parameters α and β and anisotropy) were relatively insensitive (Fig. 4). The sensitivity of hydraulic conductivity is typically highlighted in studies exploring integrated surface-subsurface model parameter sensitivity; but in contrast to our results, work by Verbist et al. (2012) and Wildemeersch et al. (2014) suggests van Genuchten parameters to have high sensitivities. In our case, low sensitivity of the peat soil van Genuchten parameters was probably because the majority of the peat soil was always fully or nearly saturated in the model (Figs. 8 and 9), i.e. the unsaturated zone where peatland van Genuchten parameters would become important was largely absent. The mineral drift van Genuchten parameters were tied to its hydraulic conductivity as explained in Section 2.2.5, and therefore not explicitly part of

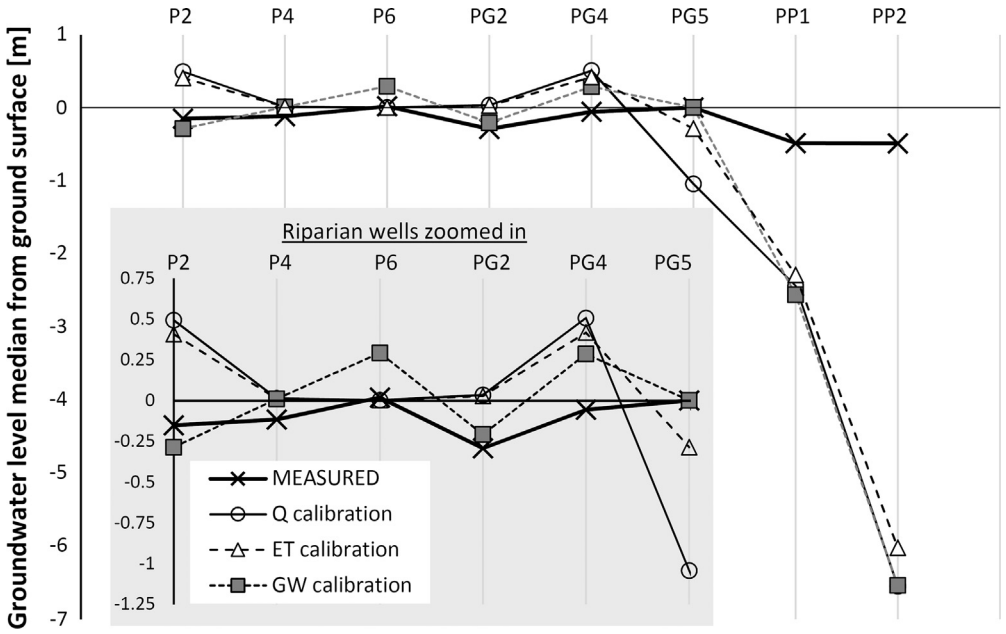

Fig. 8. Comparison of measured groundwater level medians with model output from different calibration rounds. Bottom left shows the riparian wells on a magnified y-axis to show the improvement in fit for the GW calibration round in shallow wells.

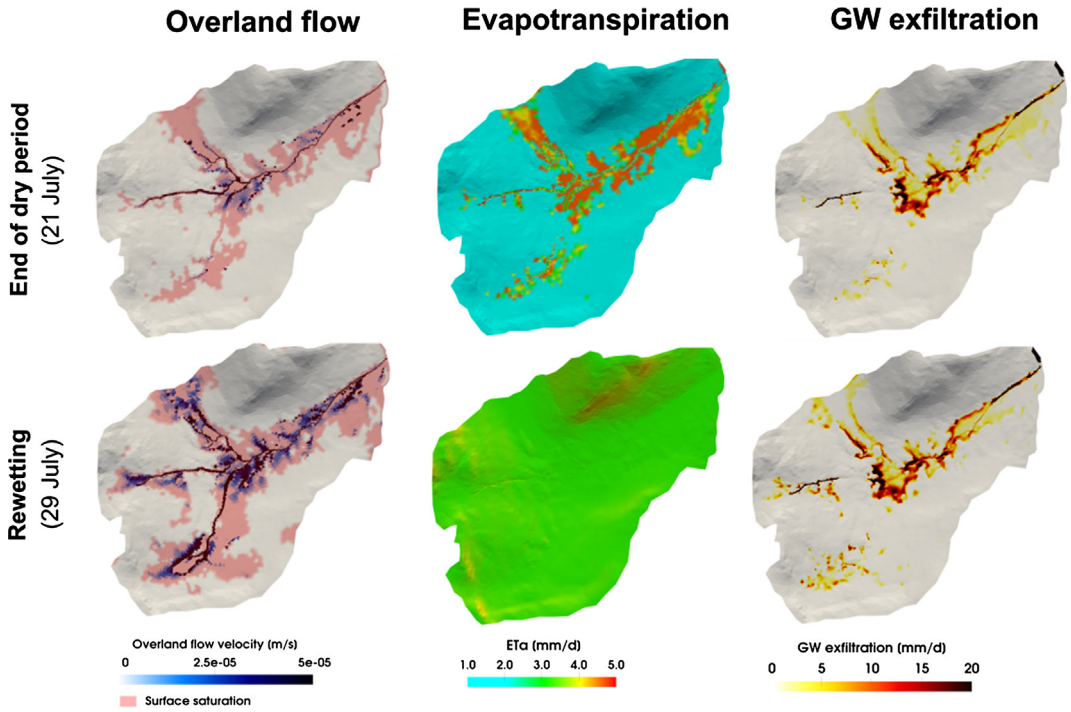

Fig. 9. Spatial snapshots for selected model output variables at one day at the end of dry period (top row) and one day after approximately 100 mm/week of rain rewetting the catchment in mid-summer (bottom row).

the model calibration. It should be noted that this linkage is likely to further increase the sensitivity mineral drift hydraulic conductivity. The calibrated value for mineral drift hydraulic conductivity was in the lower end of the range corresponding to silt and loam and silty loam soils is USDA classification (Schaap et al., 2001). The calibrated anisotropy ratio was less than 10, which is rather low compared to what is typically found in glacial deposits (Krabbenhoft et al., 1990; Sudicky et al., 2010).

For the overland flow parameters, Manning's n calibrated in the high end of literature derived range for both hillslopes (~7 s $m^{-1/3}$) and peatlands (5.5–7 s $m^{-1/3}$). High values are consistent with the variable micro-topography and extensive shrub and grass cover providing resistance to flow in the catchment. Manning's n values were also amongst the most sensitive parameters, in contrast to other work that has found Manning's n to be relatively insensitive (Maneta and Wallender, 2013; Verbist et al., 2012). The high

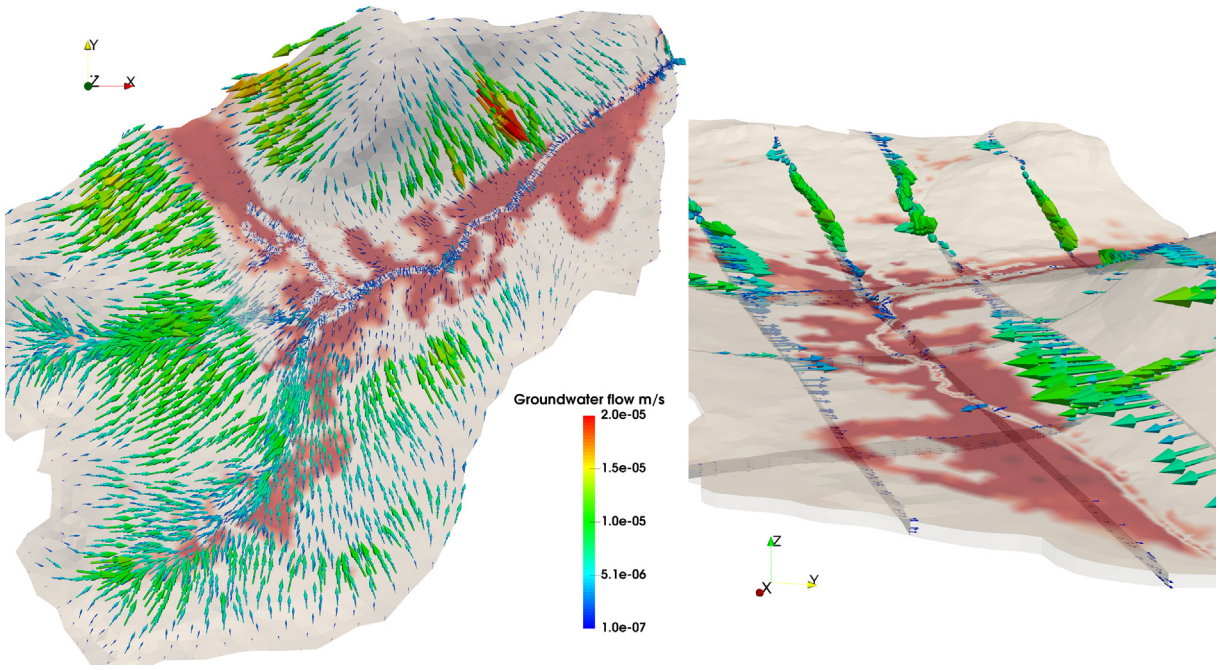

Fig. 10. Groundwater flow vectors at the end of dry period demonstrate how magnitude of flow is greater on the hillslopes with steeper gradients and low in the flat valley bottom. Panel on the right offers a landscape view from east to west and plots flow vectors along slices subsurface. Red colour in the landscape shows areas of surface saturation and both colour and size of the flow vector arrow reflect the flow rate.

sensitivity in this study likely stems from saturation excess overland flow being an active and dominant runoff generation mechanism in our simulations (Fig. 9), making the Manning's n important parameter in regulating the flow response.

Rill storage height for hillslopes calibrated around the value of 0.1 m, whereas for the peatlands the calibrated values were higher, and more variable across different calibration rounds (Fig. 5). Work by Frei et al. (2010) and Frei and Fleckenstein (2014) demonstrates how the concept of rill storage in the overland flow domain can be successfully used in mimicking the influence of micro-topography on nonstationary runoff generation. Rill storage height also regulates water ponding in the overland flow domain, affecting water sources available for evapotranspiration. Despite the importance of the rill storage height on two flow domains, in our simulations it was surprisingly among the least sensitive parameters. Moreover, calibrated values for rill storage height for riparian peatlands were particularly high (>0.3 m) in all runs (Fig. 5). We hypothesise that the surface water storage capacity created by high value of rill storage height parameter is an important factor in damping the simulated hydrograph response during high flow events (Fig. 6), leading to mismatch between simulations and observations. The accumulated surface storage may bias the simulations also by inflating the subsequent simulated base flows and exposing excessive amounts open water available for evaporation.

Parameters governing evapotranspiration functions (int, Tlim, Tred) were all moderately sensitive (Fig. 5). Calibrated values for these parameters were similar for Q and ET calibrations, whereas the GW calibration resulted in a diverging parameter set producing more ET output. In this work the ET parameters underwent an additional simplification from standard values used in HGS. Because the ET formulation following (Kristensen and Jensen, 1975) is readily largely empirical and can create inconsistent parameter behaviour for different seasons (Li et al., 2008), we argue that physical validity of the formulation was not greatly reduced. Furthermore, having only remote sensing ETa data with high level of spatial aggregation and no distinction between soil evaporation and transpiration in calibration justified merging the two evaporative processes.

The sensitivity analysis has its limitations, because the composite sensitivity is provided only for the best model run (i.e. one realisation of parameter values). Therefore the parameter sensitivity for diverging parameter combinations is not accounted for in the analysis. Null Space Monte Carlo analysis (Tonkin and Doherty, 2009) would be a robust method to further explore the parameter space around the calibrated model to increase the understanding of parameter sensitivity and model output uncertainty, but in our case the computational burden for this method would be too high. Other more advanced sensitivity analysis methods that have been successfully used in integrated hydrological modelling such as active subspace method (Jefferson et al., 2016) could be adopted to explore parameter sensitivities in future work.

4.2. Changes in hydrological response between calibration setups

Surprisingly, the three different model calibrations led mostly to similar parameter values and model output regardless of the formulation of the objective function. Analysis did not corroborate our hypothesis that remote sensing ET would bring new information to improve the model calibration. This may partially be caused by a mismatch in scales: hillslopes in the model experience water limited conditions during high evaporative demand (Fig. 9), whereas MODIS product integrates a flux over a larger spatial extent being less sensitive to water deficits in parts of the landscape. Another reason for misfit especially in midsummer might be that MODIS products are modelled data themselves which can have biases for certain land covers and overestimate the actual ET (Sharma et al., 2016). We would expect that smaller scale ETa data for sapflow or surface energy balance that is presently being collected in the catchment are equivalent in scale to the GW and streamflow measurements, and will be useful in future model iterations.

Adding GW level data to calibration objective function on the other hand had a positive influence on model fit for many of the

riparian groundwater wells, but wells on the hillslope (PP1 and PP2, see Figs 1 and 8) had notably higher measured water tables even after calibration. These wells were located at slope transition areas where hillslopes steepen above the valley bottom area, and where the model created a fringe of unusually low water tables (~5 m below ground surface, data not shown). The low water tables may have been caused by erroneous over-estimates of deep drift deposits in these transition areas, which would allow the simulated water table to remain lower than observed. However, for the catchment as a whole, geophysical surveys showed that 5–10 m depths of unsaturated drift are not uncommon on some of the steeper hillslopes (Soulsby et al., 2016).

The most notable difference between the calibrations were the high simulated ET values in GW calibration when compared with the two other calibrations (Fig. 7). The rationale here was that when PEST calibration was trying to bring down the water table at the riparian peatlands to match measured water levels in GW calibration, easiest way to do so was by promoting ET in the model. This is a classic example where model calibration is trying to compensate a structural error by adjusting parameter values. Rather than high evaporative effects, we expect that the water remains below ground surface primarily because of transmissivity feedback in the shallow peat acrotelm (see next section) which was not conceptualised in the model, but will need resolving for future model iterations.

Similar parameter values and model output across calibrations do not necessarily mean that other parameter values outside the sampled range could not lead to equally good model performance, commonly referred to as parameter equifinality (Beven, 2006). Rather, the PEST calibration procedure is conservative in varying the sensitive parameters in the gradient based search of minimum objective function (Fig. 5). Therefore the parameter space is not as extensively explored as in stochastic calibration techniques. This could be amended by adjusting the PEST calibration control options, but which could in turn lead to instabilities in the calibration process (Doherty, 2010).

4.3. Successes and failures in simulating hydrological response

Our novel simplified parameterisation and calibration of the integrated surface-subsurface simulator created a learning framework to better understand groundwater in the runoff generation process without extensive datasets for model parameterisation. Groundwater in the glacial drift valley bottom had low flow velocities and did not show an active dynamic that corresponded to the variability of runoff generation Fig. 10, Vid. 3), but rather provided a consistent water source. Direct groundwater exfiltration to stream made up only around 25% of the total flow even during simulated lowest flows (Fig. 6). An important simulated stream flow generation mechanism during the lowest flows was groundwater exfiltration to riparian wetlands, from where water is carried to the stream channel as saturation excess overland flow (Fig. 9, Vid. 1) similar to (Partington et al., 2013; Weill et al., 2013).

As opposed to the valley bottom, groundwater on the hillslopes with shallow soil layers, and therefore less storage, is activated during rain events. This is seen from moderately increased area of GW exfiltration (Fig. 9, Vid. 2) and increased flow velocities (Vid. 3) on the upper hillslopes. The observed groundwater response on the hillslopes in itself did not bring much water to stream directly via increased exfiltration. Rather it helped to maintain and extent saturated conditions thus promoting saturation excess overland flow of direct precipitation on the saturated areas (Soulsby et al., 2015; Tetzlaff et al., 2007).

Therefore for both the hillslopes and valley bottom the simulated runoff generation was predominantly invoked by saturation of the near-stream areas with subsequent saturation excess overland flow – the process being a more permanent state in the valley bottom and more dynamic on the hillslopes. In this respect the model results are generally consistent with the independent insights for streamflow generation at the experimental catchment built over a decade of conceptual modelling studies (Birkel et al., 2011a; Huijgevoort et al., 2016; Soulsby et al., 2007). Spatially explicit flow processes identified with field mapping (Birkel et al., 2011b) and tracer techniques (Lessels et al., 2015; Blumstock et al., 2015), such as locations of groundwater exfiltration and temporally variable streamflow source areas, show remarkable convergence with the simulation outputs, bringing confidence to model realism. Moreover, the general quantitative partitioning of stream flow sources into overland flow, direct groundwater discharge to the stream and groundwater exfiltration in lower slope areas is also consistent with results from conceptual tracer-aided modelling (Soulsby et al., 2016).

Simple, uniform structure of spatial parameter fields in the *overland flow domain* was a likely reason for some nuances in the observation data not being captured by the simulations. Spatial variability in the parameter rill storage height may be necessary to create conduits of high flow in the landscape and thereby capture the nonlinearity in the streamflow response as suggested by Frei et al. (2010) and Frei and Fleckenstein (2014). The overland flow parameters might also have and aspect of temporal variability due to seasonal vegetation growth and senescence. Furthermore, conceptual model for the peatland parameterisation may not have been adequate: the riparian areas were conceptualised as one low permeability peat layer, whereas peat soil can exhibit transmissivity feedbacks in the less decomposed and water conductive top peat acrotelm layer (Holden and Burt, 2003). This is prone to misrepresent the water flow in the surficial soil profile allowing too little horizontal flow. As an example, water ponding in the riparian area was simulated persistently (wells P2 and PG4) but measured only sporadically (Blumstock et al., 2016). Another factor that could cause excessive water stored in the valley bottom is the assumption of now-flow boundary at the drift-bedrock interface. Substantial leakage to bedrock groundwater, if present, would drain the catchment from below bring down the groundwater levels. However, this is unlikely to be a significant factor in the granitic/metamorphic rocks dominating the catchment.

The concept of using zones of uniforms parameter values in spatially distributed models is justifiably criticised in the literature for imposing strong constraints on the parameter values in model calibration (Hunt et al., 2007). Our calibration approach was designed to utilise the conceptual model of streamflow generation governed by different landscape units (hillslopes, riparian areas) developed in prior work at the site, and put our understanding to test in a fully-distributed physically based modelling framework. With this approach, simplification of the spatial structure in model parameters will inevitably bring structural error to the model and can miss for example high conductivity zones or preferential flow paths in the subsurface.

In our model application we would expect that fine-tuning the subsurface parameter heterogeneity as is often done in groundwater simulations (Dausman et al., 2010; Gonçalves et al., 2013; Jones et al., 2008) would not bring considerable improvements in the simulations for runoff generation, because of low amount of explicit data to constrain spatial calibration in the subsurface and the simulated limited role of groundwater in runoff generation. Instead, we would expect crucial improvements to streamflow simulations to come from focusing attention to overland flow parameters and their spatial structure (Maneta and Wallender, 2013). Using techniques such as pilot points (Doherty, 2003) to introduce spatial variability in overland flow rill storage heights and manning's n could create conduits for fast flow response (Frei and Fleckenstein, 2014) which were not well captured in

the simulations. We are presently collecting data to better characterise soil hydraulic properties and quantify spatially distributed evapotranspiration in the catchment. With these data available, the spatial constraints on the parameters could be relaxed having better data to facilitate a more spatially distributed model parameterisation and calibration. Obstacles brought by long runtimes in more spatially explicit calibration could be overcome by using novel calibration techniques such as pairing simple and complex models (Doherty and Christensen, 2011) to reduce model runtimes.

5. Conclusions

Fully-integrated surface-subsurface models have been shown to be useful in understanding the role of the shallow subsurface in runoff generation. However, data requirements to parameterise such models are typically considered unrealistic for catchments not instrumented specifically for hydrogeological research. Our novel way of calibrating an integrated surface-subsurface model, whilst of being inclusive of the main hydrological processes and parameters relating to them, required minimal field data to simulate the main characteristics of runoff generation in a montane headwater catchment.

We learned from model calibration that firstly, model output was sensitive to parameters from all model domains (subsurface and surface flow and ET). This implies that calibration of integrated surface-subsurface models should carefully consider inclusion of all model domains in the calibration, whereas presently focus is typically on the subsurface and sometimes surface flow parameters. Secondly, adding observational data in the calibration improved the model performance against the related output for some observations (GW levels), but not for all (ET timeseries). In both cases including new observations worsened the fit of streamflow: an optimal situation was not reached.

We reached a satisfactory fit between the observations and simulation output, but equally importantly, the simulations gave a physically grounded learning experience of the reasons for model defects and success. Comparing simulation results to observations made it possible to reflect upon the original conceptual model. Based on some mismatches in simulated groundwater levels and stream hydrograph we hypothesised the simplified structure of the peat layer (acrotelm/catotelm model not used) and uniform parameter fields in the surface flow domain as likely reasons for model-to-measurement mismatch. Visually intuitive model outputs improved understanding of how groundwater operates in generating streamflow at the catchment primarily via facilitating saturation excess overland flow. Steady groundwater inputs created saturated conditions in the valley bottom riparian peatlands, leading to overland flow even during dry periods. Groundwater on the hillslopes was more dynamic in its response to rainfall, acting to expand the saturated areas on the hillslopes and thereby promoting saturation excess overland flow during rainstorms.

Acknowledgements

This work was funded by NERC/JPI SIWA project (NE/M019896/1) and the European Research Council ERC (project GA 335910 VeWa). Numerical simulations were performed using the Maxwell High Performance Computing Cluster of the University of Aberdeen IT Service, provided by Dell Inc. and supported by Alces Software. Aquanty Inc. is acknowledged for support in providing HGS simulation software compatible with the Maxwell High Performance Computing Cluster. We would also like to thank the anonymous reviewers for their constructive comments that improved the manuscript.

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.2017.02.023.

References

Ala-aho, P., Rossi, P.M., Isokangas, E., Kløve, B., 2015. Fully integrated surface-subsurface flow modelling of groundwater-lake interaction in an esker aquifer: model verification with stable isotopes and airborne thermal imaging. J. Hydrol. 522, 391–406. http://dx.doi.org/10.1016/j.jhydrol.2014.12.054.

Aquanty, 2016. HydroGeoSphere user manual. Release 1.0. Aquanty Inc., Waterloo, Ontario, Canada.

Bergstrom, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments. SMHI RHO 7, Norrköping.

Beven, K., 2002. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrol. Processes 16, 189–206. http://dx.doi.org/10.1002/hyp.343.

Beven, K., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36. http://dx.doi.org/10.1016/j.jhydrol.2005.07.007.

Beven, K., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. J. 24, 43–69.

Birkel, C., Soulsby, C., Tetzlaff, D., 2011a. Modelling catchment-scale water storage dynamics: reconciling dynamic storage with tracer-inferred passive storage. Hydrol. Processes 25, 3924–3936. http://dx.doi.org/10.1002/hyp.8201.

Birkel, C., Tetzlaff, D., Dunn, S.M., Soulsby, C., 2011b. Using time domain and geographic source tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models. Water Resour. Res. 47, W02515. http://dx.doi.org/10.1029/2010WR009547.

Birkel, C., Soulsby, C., Tetzlaff, D., 2014. Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. Water Resour. Res. 50, 3481–3501. http://dx.doi.org/10.1002/2013WR014925.

Blumstock, M., Tetzlaff, D., Malcolm, I.A., Nuetzmann, G., Soulsby, C., 2015. Baseflow dynamics: multi-tracer surveys to assess variable groundwater contributions to montane streams under low flows. J. Hydrol. 527, 1021–1033. http://dx.doi.org/10.1016/j.jhydrol.2015.05.019.

Blumstock, M., Tetzlaff, D., Dick, J.J., Nuetzmann, G., Soulsby, C., 2016. Spatial organisation of groundwater dynamics and streamflow response from different hydropedological units in a montane catchment. Hydrol. Processes. http://dx.doi.org/10.1002/hyp.10848.

Boehner, J., Koethe, R., Conrad, O., Gross, J., Ringeler, A., Selige, T., 2002. Soil regionalisation by means of terrain analysis and process parameterisation. In: Micheli, E., Nachtergaele, F., Montanarella, L. (ed.): Soil Classification 2001. 213–222 European Soil Bureau, Research Report No. 7, EUR 20398 EN, Luxembourg.

Bolger, B.L., Park, Y., Unger, A.J., Sudicky, E.A., 2011. Simulating the pre-development hydrologic conditions in the San Joaquin Valley. California. J. Hydrol. 411, 322–330. http://dx.doi.org/10.1016/j.jhydrol.2011.10.013.

Boorman, D.B., Hollis, J.M., Lilly, A., 1995. Hydrology of soil types: a hydrologically-based of the soils of the United Kingdom. Report No.126. Institute of Hydrology, Oxfordshire, the UK.

Brookfield, A., Sudicky, E., Park, Y., Conant, B., 2009. Thermal transport modelling in a fully integrated surface/subsurface framework. Hydrol. Processes 23, 2150–2164. http://dx.doi.org/10.1002/hyp.7282.

Brunner, P., Doherty, J., Simmons, C.T., 2012. Uncertainty assessment and implications for data acquisition in support of integrated hydrologic models. Water Resour. Res. 48, W07513. http://dx.doi.org/10.1029/2011WR011342.

Dausman, A.M., Doherty, J., Langevin, C.D., Dixon, J., 2010. Hypothesis testing of buoyant plume migration using a highly parameterized variable-density groundwater model at a site in Florida, USA. Hydrogeol. J. 18, 147–160. http://dx.doi.org/10.1007/s10040-009-0511-6.

Doherty, J., 2003. Ground water model calibration using pilot points and regularization. Ground water 41, 170–177. http://dx.doi.org/10.1111/j.1745-6584.2003.tb02580.x.

Doherty, J., 2010. PEST: Model-Independent Parameter Estimation User Manual. Watermark Numerical Computing, Brisbane, Australia.

Doherty, J., Christensen, S., 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resour. Res. 47, W12534. http://dx.doi.org/10.1029/2011WR010763.

Doherty, J., Welter, D., 2010. A short exploration of structural noise. Water Resour. Res. 46, W05525. http://dx.doi.org/10.1016/10.1029/2009WR008377.

Dunn, S.M., Mackay, R., 1995. Spatial variation in evapotranspiration and the influence of land use on catchment hydrology. J. Hydrol. 171, 49–73. http://dx.doi.org/10.1016/0022-1694(95)02733-6.

Ebel, B.A., Mirus, B.B., Heppner, C.S., VanderKwaak, J.E., Loague, K., 2009. First-order exchange coefficient coupling for simulating surface water–groundwater interactions: Parameter sensitivity and consistency with a physics-based approach. Hydrol. Processes 23, 1949–1959. http://dx.doi.org/10.1002/hyp.7279.

ESRI, 2011. ARcGIS Desktop: Release 10.

Frei, S., Fleckenstein, J.H., 2014. Representing effects of micro-topography on runoff generation and sub-surface flow patterns by using superficial rill/depression

storage height variations. Environ. Modell. Softw. 52, 5–18. http://dx.doi.org/10.1016/j.envsoft.2013.10.007.
Frei, S., Lischeid, G., Fleckenstein, J.H., 2010. Effects of micro-topography on surface–subsurface exchange and runoff generation in a virtual riparian wetland — A modeling study. Adv. Water Resour. 33, 1388–1401. http://dx.doi.org/10.1016/j.advwatres.2010.07.006.
Gonçalves, T.D., Fischer, T., Gräbe, A., Kolditz, O., Weiss, H., 2013. Groundwater flow model of the Pipiripau watershed, Federal District of Brazil. Environ. Earth Sci. 69, 617–631. http://dx.doi.org/10.1007/s12665-013-2400-5.
Holden, J., Burt, T., 2003. Hydrological studies on blanket peat: the significance of the acrotelm-catotelm model. J. Ecol. 91, 86–102. http://dx.doi.org/10.1046/j.1365-2745.2003.00748.x.
Huijgevoort, M.H.J., Tetzlaff, D., Sutanudjaja, E.H., Soulsby, C., 2016. Using high resolution tracer data to constrain water storage, flux and age estimates in a spatially distributed rainfall-runoff model. Hydrol. Processes. http://dx.doi.org/10.1002/hyp.10902.
Hunt, R.J., Doherty, J., Tonkin, M.J., 2007. Are models too simple? Arguments for increased parameterization. Ground Water 45, 254–262. http://dx.doi.org/10.1111/j.1745-6584.2007.00316.x.
Hwang, H., Park, Y., Frey, S.K., Berg, S.J., Sudicky, E.A., 2015. A simple iterative method for estimating evapotranspiration with integrated surface/subsurface flow models. J. Hydrol. 531, 949–959. http://dx.doi.org/10.1016/j.jhydrol.2015.10.003.
Jefferson, J.L., Gilbert, J.M., Constantine, P.G., Maxwell, R.M., 2016. Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. Comput. Geosci. 90, 78–89. http://dx.doi.org/10.1016/j.cageo.2015.11.002.
Jones, J., Sudicky, E., Brookfield, A., Park, Y., 2006. An assessment of the tracer-based approach to quantifying groundwater contributions to streamflow. Water Resour. Res. 42, W02407. http://dx.doi.org/10.1029/2005WR004130.
Jones, J., Sudicky, E., McLaren, R., 2008. Application of a fully-integrated surface-subsurface flow model at the watershed-scale: a case study. Water Resour. Res. 44, W03407. http://dx.doi.org/10.1029/2006WR005603.
Kollet, S.J., Maxwell, R.M., 2006. Integrated surface–groundwater flow modeling: a free-surface overland flow boundary condition in a parallel groundwater flow model. Adv. Water Resour. 29, 945–958. http://dx.doi.org/10.1016/j.advwatres.2005.08.006.
Krabbenhoft, D.P., Anderson, M.P., Bowser, C.J., 1990. Estimating groundwater exchange with lakes: 2. Calibration of a three-dimensional, solute transport model to a stable isotope plume. Water Resour. Res. 26, 2455–2462. http://dx.doi.org/10.1029/WR026i010p02455.
Kristensen, K.J., Jensen, S.E., 1975. Model for estimating actual evapotranspiration from potential evapotranspiration. Nord. Hydrol. 6, 170–188.
Lessels, J.S., Tetzlaff, D., Birkel, C., Dick, J., Soulsby, C., 2015. Water sources and mixing in riparian wetlands revealed by tracers and geospatial analysis. Water Resour. Res. 52, 456–470. http://dx.doi.org/10.1002/2015WR017519.
Li, Q., Unger, A.J.A., Sudicky, E.A., Kassenaar, D., Wexler, E.J., Shikaze, S., 2008. Simulating the multi-seasonal response of a large-scale watershed with a 3D physically-based hydrologic model. J. Hydrol. 357, 317–336. http://dx.doi.org/10.1016/j.jhydrol.2008.05.024.
Liggett, J.E., Partington, D., Frei, S., Werner, A.D., Simmons, C.T., Fleckenstein, J.H., 2015. An exploration of coupled surface–subsurface solute transport in a fully integrated catchment model. J. Hydrol. 529, 969–979. http://dx.doi.org/10.1016/j.jhydrol.2015.09.006.
Malcolm, I., Soulsby, C., Youngson, A., Hannah, D., McLaren, I., Thorne, A., 2004. Hydrological influences on hyporheic water quality: implications for salmon egg survival. Hydrol. Process. 18, 1543–1560. http://dx.doi.org/10.1002/hyp.1405.
Maneta, M., Wallender, W., 2013. Pilot-point based multi-objective calibration in a surface–subsurface distributed hydrological model. Hydrol. Sci. J. 58, 390–407. http://dx.doi.org/10.1080/02626667.2012.754987.
Maxwell, R.M., Condon, L.E., 2016. Connections between groundwater flow and transpiration partitioning. Science 353, 377–380. http://dx.doi.org/10.1126/science.aaf7891.
McDonnell, J.J., 2013. Are all runoff processes the same? Hydrol. Processes 27, 4103–4111. http://dx.doi.org/10.1002/hyp.10076.
Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115, 1781–1800. http://dx.doi.org/10.1016/j.rse.2011.02.019.
Päivänen, J., 1973. Hydraulic conductivity and water retention in peat soils. Acta Forestalia Fennica, vol. 129. Society of Forestry in Finland, Helsinki, Finland.
Park, Y., Sudicky, E.A., Brookfield, A.E., Jones, J.P., 2011. Hydrologic response of catchments to precipitation: quantification of mechanical carriers and origins of water. Water Resour. Res. 47, W12515. http://dx.doi.org/10.1029/2010WR010075.
Partington, D., Brunner, P., Frei, S., Simmons, C., Werner, A., Therrien, R., et al., 2013. Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment. Water Resour. Res. 49, 5501–5519. http://dx.doi.org/10.1002/wrcr.20405.
Schaap, M.G., Leij, F.J., van Genuchten, M.T., 2001. Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. Journal of Hydrology 251, 163–176. http://dx.doi.org/10.1016/S0022-1694(01)00466-8.
Schilling, O.S., Doherty, J., Kinzelbach, W., Wang, H., Yang, P.N., Brunner, P., 2014. Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water–vegetation interactions. J. Hydrol. 519 (Part B), 2258–2271. http://dx.doi.org/10.1016/j.jhydrol.2014.08.063.
Sharma, V., Kilic, A., Irmak, S., 2016. Impact of scale/resolution on evapotranspiration from Landsat and MODIS images. Water Resour. Res. 52, 1800–1819. http://dx.doi.org/10.1002/2015WR017772.
Shaver, R.B., 1998. The Determination of Glacial Till Specific Storage in North Dakota. Ground Water 36, 552–557. http://dx.doi.org/10.1111/j.1745-6584.1998.tb02828.x.
Sklash, M.G., Farvolden, R.N., 1979. The Role Of Groundwater In Storm Runoff. Dev. Water Sci. 12, 45–65. http://dx.doi.org/10.1016/S0167-5648(09)70009-7.
Soulsby, C., Tetzlaff, D., van den Bedem, N., Malcolm, I.A., Bacon, P.J., Youngson, A.F., 2007. Inferring groundwater influences on surface water in montane catchments from hydrochemical surveys of springs and streamwaters. J. Hydrol. 333, 199–213. http://dx.doi.org/10.1016/j.jhydrol.2006.08.016.
Soulsby, C., Birkel, C., Geris, J., Dick, J., Tunaley, C., Tetzlaff, D., 2015. Stream water age distributions controlled by storage dynamics and nonlinear hydrologic connectivity: Modeling with high-resolution isotope data. Water Resour. Res. 51, 7759–7776. http://dx.doi.org/10.1002/2015WR017888.
Soulsby, C., Bradford, J., Dick, J., McNamara, J.P., Geris, J., Lessels, J., et al., 2016. Estimating storage in headwater catchments: integrating insights from hydrogeophysics and tracers. Hydrol. Processes. http://dx.doi.org/10.1002/hyp.10889.
Sudicky, E., Illman, W., Goltz, I., Adams, J., McLaren, R., 2010. Heterogeneity in hydraulic conductivity and its role on the macroscale transport of a solute plume: From measurements to a practical application of stochastic flow and transport theory. Water Resour. Res. 46, W01508. http://dx.doi.org/10.1029/2008WR007558.
Tetzlaff, D., Soulsby, C., Waldron, S., Malcolm, I.A., Bacon, P.J., Dunn, S.M., et al., 2007. Conceptualization of runoff processes using a geographical information system and tracers in a nested mesoscale catchment. Hydrol. Processes 21, 1289–1307. http://dx.doi.org/10.1002/hyp.6309.
Tonkin, M., Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. Water Resour. Res. 45, W00B10. http://dx.doi.org/10.1029/2007WR006678.
Verbist, K.M.J., Pierreux, S., Cornelis, W.M., McLaren, R., Gabriels, D., 2012. Parameterizing a coupled surface-subsurface three-dimensional soil hydrological model to evaluate the efficiency of a runoffwater harvesting technique. Vadose Zone J. http://dx.doi.org/10.2136/vzj2011.0141.
Weill, S., Altissimo, M., Cassiani, G., Deiana, R., Marani, M., Putti, M., 2013. Saturated area dynamics and streamflow generation from coupled surface–subsurface simulations and field observations. Adv. Water Resour. 59, 196–208. http://dx.doi.org/10.1016/j.advwatres.2013.06.007.
Wildemeersch, S., Goderniaux, P., Orban, P., Brouyère, S., Dassargues, A., 2014. Assessing the effects of spatial discretization on large-scale flow model performance and prediction uncertainty. J. Hydrol. 510, 10–25. http://dx.doi.org/10.1016/j.jhydrol.2013.12.020.
Yang, J., Graf, T., Ptak, T., 2015. Impact of climate change on freshwater resources in a heterogeneous coastal aquifer of Bremerhaven, Germany: A three-dimensional modeling study. J. Contam. Hydrol. 177–178, 107–121. http://dx.doi.org/10.1016/j.jconhyd.2015.03.014.