HYDROLOGICAL PROCESSES
Hydrol. Process. **25**, 1784–1799 (2011)
Published online 3 December 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.7936

Models as multiple working hypotheses: hydrological simulation of tropical alpine wetlands

Wouter Buytaert[1]* and Keith Beven[2]

[1] *Civil and Environmental Engineering, Imperial College London, Skempton Building, London, SW72AZ UK*
[2] *Environmental Science, Lancaster University, Bailrigg, Lancaster, LA1 4YQ UK*

Abstract:

Tropical alpine grasslands, locally known as páramos, are the water towers of the northern Andes. They are an essential water source for drinking water, irrigation schemes and hydropower plants. But despite their high socio-economic relevance, their hydrological processes are very poorly understood. Since environmental change, ranging from small scale land-use changes to global climate change, is expected to have a strong impact on the hydrological behaviour, a better understanding and hydrological prediction are urgently needed. In this paper, we apply a set of nine hydrological models of different complexity to a small, well monitored upland catchment in the Ecuadorian Andes. The models represent different hypotheses on the hydrological functioning of the páramo ecosystem at catchment scale. Interpretation of the results of the model prediction and uncertainty analysis of the model parameters reveals important insights in the evapotranspiration, surface runoff generation and base flow in the páramo. However, problems with boundary conditions, particularly spatial variability of precipitation, pose serious constraints on the differentiation between model representations. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS modelling; hypothesis testing; wetlands

Received 25 March 2010; Accepted 11 October 2010

INTRODUCTION

Models as working hypotheses

Hydrological models are frequently used to generate time series of future discharges, or to predict the potential impacts of changes in the catchment on the discharge regime. The application of hydrological models in prediction mode is well studied. Many methods exist to find good parameter set(s) to be used for prediction (e.g. (Gupta *et al*., 1998; Beven, 2001b; Vrugt *et al*., 2003; McMillan and Clark, 2009)) and methods to predict the uncertainty in the final predictions have received much attention ((Liu and Gupta, 2007; Götzinger and Bárdossy, 2008; Stedinger *et al*., 2008; Beven, 2009), and many others).

Recently, a series of opinion papers have emerged in the hydrological literature that debate the use of hydrological models as a means of improving hydrological knowledge (e.g. (Beven, 2006b; Andréassian *et al*., 2007; Beven, 2008; Savenije, 2008; Sivapalan, 2009)). The integrative nature of computer models can indeed be a very useful tool to test hypotheses about a hydrological system. Many processes that are thought to be important at the catchment scale are nonlinear and so variable in time and space that studying them in isolation at small scales yields only limited insight about their hydrological behaviour and importance at larger catchment scales (see Beven (2006c), for example). Reconstructing these processes in a hydrological model is a way to explore the validity of process representations, interactions and scaling behaviour. Comparing different conceptual representations of hydrological processes in a model can then be used to test which of these hypotheses are compatible with observations at the catchment scale. A small number of past studies have adopted this approach (Piñol *et al*., 1997; Clark *et al*., 2008; Fenicia *et al*., 2008).

In this paper, such an approach is used to unravel the hydrological processes occurring in tropical alpine wetlands. These wetlands, locally known as páramos, provide many environmental services, of which water production is one of the most important (IUCN, 2002; Buytaert *et al*., 2006a). Due to the high and sustained baseflow and the good water quality of rivers descending from the páramos, these wetlands are the main water source for agricultural and urban use, as well as hydropower prediction. The region is experiencing rapid changes in land use. These changes, together with global climate change, are expected to have a strong impact on the water availability and quality in the páramo (Liniger *et al*., 1998; Jansky *et al*., 2002; Buytaert *et al*., 2006a; Buytaert and Beven, 2009). Therefore, an improved understanding of the hydrological processes of the Andean páramos is urgently needed. So far, most hydrological research has focused on small-scale processes such as hydrophysical soil properties (Farley *et al*., 2004; Buytaert *et al*., 2005, 2006c; Harden, 2006), erosion (Poulenard *et al*., 2001; Podwojewski *et al*., 2002; Harden, 2006) and climate variability (Timbe, 2004; Buytaert *et al*., 2006b; Célleri *et al*., 2007). Attempts to model the hydrology of the

* Correspondence to: Wouter Buytaert, Civil and Environmental Engineering, Imperial College London, Skempton Building, London, SW72AZ UK. E-mail: w.buytaert@imperial.ac.uk

páramos are thus far very scarce and have had only limited success.

The hydrology of tropical mountain wetlands (páramos)

Neotropical alpine regions, or páramos, cover around 36 000 km^2 of the tropical Andes, between the continuous forest border (around 3500 m) and the permanent snow line (around 5000 m altitude) (e.g. (Luteyn *et al*., 1992; Hofstede *et al*., 2003; Sklenář *et al*., 2008)). They form a discontinuous belt stretching from the páramo de Mérida in Venezuela to the Huancabamba depression in north Perú. To the south, the páramo is bordered by the Jalca, which is a transition biome between the páramo and the drier puna that dominates highlands in south Peru and Bolivia (Weigend, 2004; Sánchez-Vega and Dillon, 2006). In the north, they are limited by the lower altitude of the Andes in central America. Small isolated patches occur in Costa Rica: 26 km^2 in the Cordillera Central and 126 km^2 on the Cordillera Talamanca (Kappelle *et al*., 2005).

Similar to many other mountain regions in the world (Messerli *et al*., 2004; Viviroli *et al*., 2007), páramos tend to have a higher precipitation regime than surrounding lowlands due to orographic precipitation. Ecuadorian páramos receive on average 16% more precipitation than stations in the inter-Andean valley between 2500 and 3000 m. Due to the mountainous terrain and the influence of various climate systems (Vuille *et al*., 2000), spatial variability of precipitation in the Andes is exceptionally high (Buytaert *et al*., 2006b; Célleri, 2007). Observed precipitation records ranging approximately from 500 mm in the dry páramos around Chimborazo, to more than 3000 mm in the wettest páramos of the Eastern Cordillera that are influenced by air masses from the Amazon basin.

The vegetation of the páramo consists of tussock grasses and low shrubs (Sklenář and Balslev, 2005). Local patches of shrub vegetation exist, dominated by indigenous Polylepis species. Many of the plant species are adapted to the specific physio-chemical and climatic conditions, such as the low atmospheric pressure, intense ultra-violet radiation and the drying effects of wind (Luteyn, 1999). Transpiration is limited by the low temperatures, high frequency of fog, cloud cover and high relative humidity. This results in low overall evapotranspiration rates and a high runoff ratio. Literature values of evapotranspiration range from 0·8 to about 1·5 mm day^{-1} (Hofstede *et al*., 1995; Buytaert, 2004) and runoff ratios of natural páramo vary around 50–70% (Buytaert *et al*., 2007).

Due to the lack of a dense vegetation layer, the high water attenuation capacty of the páramos is commonly attributed to their soils. Due to the cold and wet environment, and the low oxygen pressure, organic carbon accumulates in the soils. The combination of a high organic carbon content and volcanic parent material gives rise to very porous soils with a high infiltration and storage capacity, which avoids the flashy hydrological response characteristic of temperate wetlands (Bragg, 2002; Holden *et al*., 2006). Another important mechanism behind the high water regulation capacity is the abundance of hydrologically disconnected areas because of the irregular topography, which gives rise to a large number of lakes and swamps. Finally, temporal variability of rainfall is low in many páramos.

These specific hydrological processes are vulnerable to perturbation. Being headwater catchments, the páramos rely on meteorological water. Except perhaps for local fracture systems or páramos with a large upslope area, the buffering role of deep groundwater contributions is limited. Therefore, spatial and temporal changes in the precipitation pattern may have a strong impact on hydrological processes, as well as soil formation and ecosystem dynamics. As a result, hydrological modelling and prediction are important aspects of sustainable water resources management.

MATERIALS AND METHODS

The experimental catchment

The study catchment Huagrahuma is located northwest of the city of Cuenca in the south Ecuadorian Andes (Figure 1). It is part of the Rio Paute basin, which hosts the largest hydropower plant of the country. During the dry season, baseflows in the hydropower reservoir are sustained by the páramo ecosystem, which covers about 40% of the basin area. Additionally, the páramo provides around 90% of the water for Cuenca (around 500 000 inhabitants). As such, the hydrological processes of the surrounding páramos are crucial for local socio-economic development.

The Huagrahuma catchment has an area of 2·58 km^2 and extends between 3690 and 4100 m above sea level. It is a typical glacier-shaped valley with strong gradients in slopes, exposure and elevation. The soils are Hydric Andosols (FAO/ISRIC/ISSS, 1998; Buytaert *et al*., 2006c) and consist of a mixture of volcanic ashes, organic carbon and tertiary bedrock material (Figure 2). Due to the high altitude, low temperature and wet soil conditions, soil organic carbon accumulates and typical values in the topsoil vary around 30% (Buytaert *et al*., 2005, 2006c). This high carbon content results in very open and porous soils, with infiltration capacities between 15 and 150 mm h^{-1}, and water retention capacities up to 90 vol% in saturated conditions.

On average, the soil layer is about 80 cm thick, with bedrock outcroppings at convex locations and hilltops (Buytaert *et al*., 2006c). It covers a bedrock consisting of Cretaceous and early Tertiary lavas and andesitic volcanoclastic deposits, shaped and compacted by glacier activity during the last ice age (Coltorti and Ollier, 2000; Hungerbühler *et al*., 2002). The hydraulic conductivity of the bedrock is low, particularly compared to the hydraulic conductivity of the soils.

The irregular and steep topography gives rise to many local depressions in the landscape where surface and

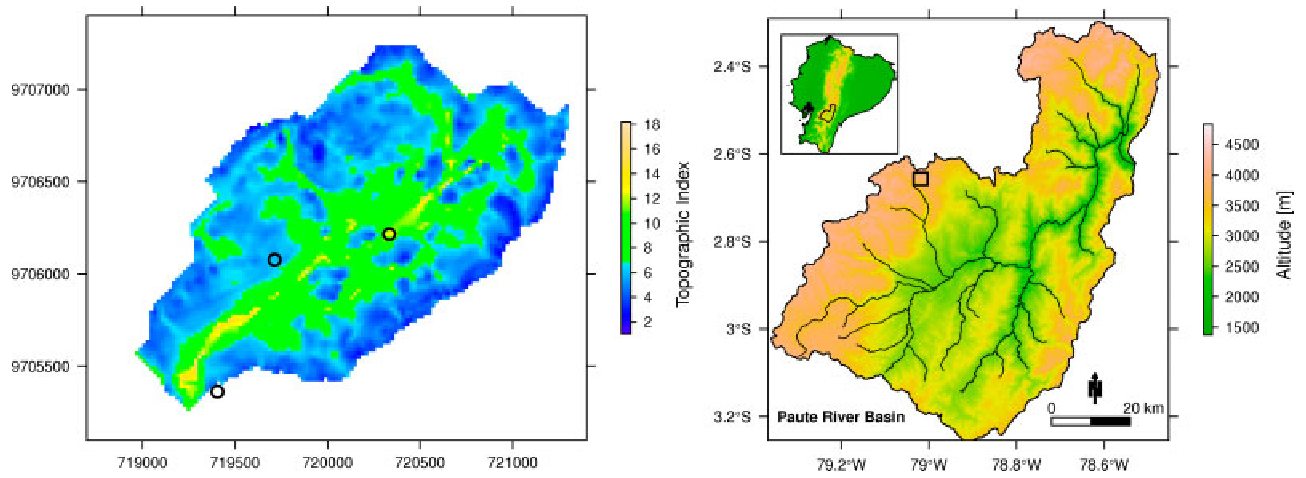

Figure 1. Location and topographic index map of the experimental catchment Huagrahuma, north-west of the city of Cuenca, south Ecuador. ○ = location of the rain gauges. Coordinates of the topographic index map are in UTM zone 17S

subsurface runoff accumulates and swamps or little lakes and ponds develop. These areas can range between a few m^2 to several hectares in the larger valleys. Soils in these areas can be several meters deep and generally remain at or very close to saturation throughout the year (Buytaert *et al*., 2006c). In the study area they are estimated to occupy around 5% of the landscape, although this depends strongly on the local topography.

Monitoring

Discharge was measured with a concrete V-shaped weir with sharp metal edges, combined with a water level sensor programmed at a 15-min interval. About halfway through the monitoring period the sensor failed and was replaced with a backup sensor on 12/05/2002. The new sensor had a lower sensitivity, resulting in measurements of slightly lower quality. The structure has a maximum capacity of 800 l s^{-1}. This value was exceeded 13 times during the monitoring period. These events were excluded from the model evaluation.

Discharge data are available from 04/09/2001 to 17/06/2005, with 31% missing data. There are two main data gaps, from 04/04/2002 to 12/05/2002 and from 07/02/2003 to 13/08/2003. Additionally, due to limited access to the study region during certain periods, the sampling interval was increased to 30 min to avoid memory overflow in the logger. This results in 91 413 discharge measurements (Table I).

Precipitation was measured with three tipping bucket rain gauges (0·2 mm resolution). The tipping bucket data were transformed to a 15-min time series to be compatible with the water level data. The precipitation data were merged by weighted averaging based on Thiessen polygons. Some gaps were filled by linear interpolation from the two other stations. The final precipitation series does not contain any gaps.

Reference evapotranspiration was calculated from meteorological data using the Penman–Monteith formula with constant canopy resistance (Allen *et al*., 1998). No synchronous meteorological measurements were available. Data from a meteorological station at the nearby Chanlud damsite (Figure 1) recorded in 2003 were used. The average intra-day curve of reference evapotranspiration was calculated and repeated for the monitoring period. Since seasonal climate variability is very low in the páramo region, this method is thought to produce an adequate approximation.

Table I. Statistical characteristics of the observed discharge during the modelled periods

Statistic	Dimension	Total period	Period 1	Period 2
Max.	m/15 min	5·70e-4	3·16e-4	4·14e-4
Q95	m/15 min	6·90e-5	8·59e-5	7·79e-5
Median	m/15 min	2·79e-5	3·13e-5	2·92e-5
Q5	m/15 min	6·61e-6	1·83e-5	1·33e-5
Min.	m/15 min	2·02e-6	1·37e-5	6·76e-6
n	—	91 413	10 000	9 372

Max, maximum observed flow; min, minimum observed flow; n, number of discharge measurements available; Q5 and Q95 are respectively the 5 and 95% quantiles.

A digital elevation model (DEM) of the catchments was generated from contour lines with an interval of 20 m, using regularized splines with tension (Mitasova and Mitas, 1993). The DEM was subsequently used to calculate the topographic index and the channel length distribution as required by the different models.

Hydrological modelling

Nine different model structures were selected to test hypothesis about the representation of the hydrology of the páramo at a catchment scale (Figure 3). A rationale for the model selection is given in the following subsections. All models were run for the entire monitoring period. Since two near-complete subsets were available, they were used for a split sample test (Klemeš, 1986). The first period (23/05/2002 to 04/09/2002) was used for model calibration and a corresponding period two years later (23/05/2004 to 04/09/2004) was used for model evaluation. Since there was no gap in the input data, the whole period was simulated at a 15 min timestep. Although the model used explicit time stepping, tests

Figure 2. Typical soil profile of the páramo region. A porous, highly organic topsoil with high permeability overlays a clayey bedrock substrate with lower permeability, giving rise to a saturated subsurface runoff zone in the lower part of the topsoil

with running the model at a higher timestep (1 min) did indeed not yield significant differences in results, and also suggests that the short stepsize constrains numerical artefacts (Clark and Kavetski, 2010). Discharge statistics of the periods are given in Table I.

The linear store. The linear store (e.g. (Beven, 2001b)) is a commonly used structure to build simple conceptual models. Combinations of linear stores can be used to represent the hydrological system, each of which assumes a linear relation between internal storage and discharge (Figure 3). Buytaert *et al.* (2004) analysed the recession curves of Huagrahuma and a similar catchment in the Paute basin and found a good linear behaviour. They identified three main response modes, which showed different mean residence time parameters and which were attributed to overland flow, interflow and baseflow, albeit that such a process interpretation is made purely on the basis of the residence time characteristics of the stores.

Therefore, a combination of parallel stores is used as a benchmark model in this study. Since the study revealed that interflow was of lesser importance, both a combination of two and three parallel stores is used. Each store is characterized by one parameter, the mean residence time and an initialization value (Buytaert *et al*., 2004). Evapotranspiration is a fixed proportion of precipitation, the value of which is also optimied. Additionally, one or more parameters are needed to distribute precipitation over the different stores. This results in a total of six parameters for the two-store model, and nine parameters for the three-store model (Tables II and IV). However, the fast reservoir is only active during and right after precipitation events. If the modelling starts in a dry period, the initialization can therefore be set to zero.

Topmodel

As detailed in the section on 'The Experimental Catchment', topography has a strong influence on the hydrological response. Local depressions and flat areas with a large accumulated drainage area frequently generate saturated conditions. At the same time, the high porosity and infiltration capacity of the soils, in combination with the typically low rainfall intensity in the area Buytaert *et al.* (2006b) virtually eliminate infiltration excess overland

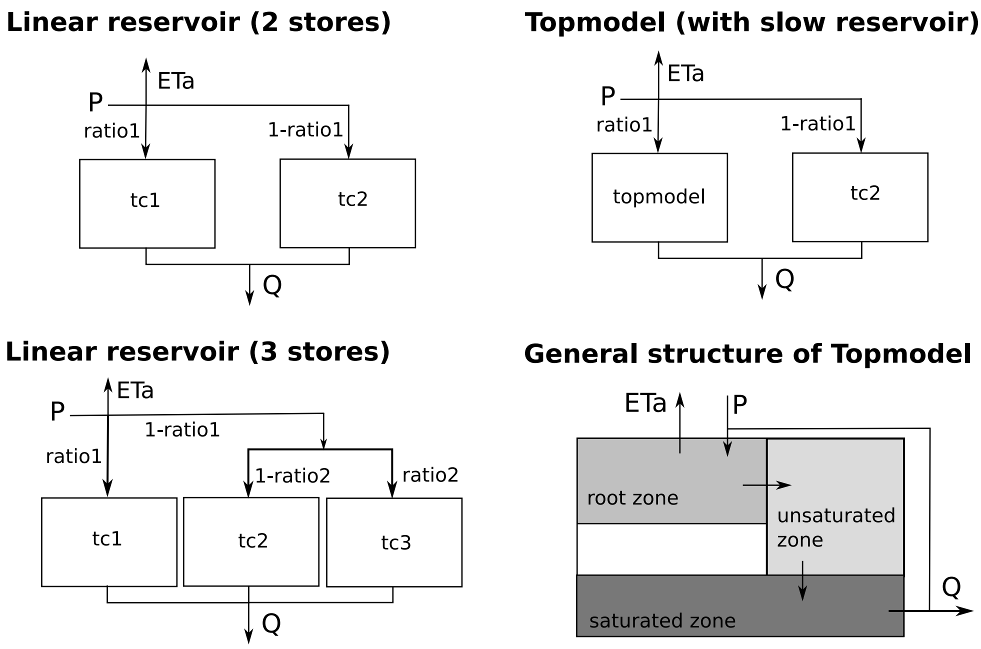

Figure 3. Conceptual representation of the models used in this study. For the meaning of the parameters see Table I. A full explanation of the implementation of the models, including equations, can be found in the study of Beven (2001b)

Table II. Nash and Sutcliffe (NS) performance and bias for the different model structures, over the total monitoring period, the subset periods and the split sample test when parameters from Period 1 are used to predict Period 2

Model description	npar	Total period		Period 1		Period 2		Split sample	
		NS	Bias (%)	NS	Bias (%)	NS	Bias (%)	NS	Bias (%)
Parallel linear reservoirs (two stores)	5	0·64	6·0	0·72	−0·5	0·79	−2·9	0·75	10·7
Parallel linear reservoirs (three stores)	8	0·65	−4·0	0·72	0·0	0·79	−2·3	0·77	8·7
TOPMODEL	9	0·67	−10·3	0·87	−0·2	0·77	12·2	0·62	20·8
TOPMODEL (fixed evapotranspiration)	8	0·69	9·7	0·84	5·3	0·75	15·6	0·60	27·6
TOPMODEL (with slow reservoir)	11	0·69	−3·5	0·88	3·5	0·89	−5·8	0·72	21·4
TOPMODEL (with overland flow delay)	10	0·67	−10·4	0·87	−0·2	0·77	11·6	0·59	20·2
Generalized TOPMODEL ($n = 1$)	9	0·59	−11·2	0·80	−1·6	0·66	4·3	0·45	16·7
Generalized TOPMODEL ($n = 2$)	9	0·68	−13·0	0·84	−2·9	0·74	13·9	0·63	19·9
Generalized TOPMODEL ($n = 10$)	9	0·67	−12·1	0·87	−4·6	0·76	9·5	0·61	15·2
Generalized TOPMODEL ($n = 100$)	9	0·67	−10·4	0·87	−1·4	0·76	11·4	0·63	21·0

Period 1, 23/05/2002 to 04/09/2002; Period 2, 11/06/2004 to 23/09/2004; npar, total number of parameters.

Table III. Width of the 90% prediction limits for the different model structures, relative to the average observed discharge

Model description	Total period	Period 1	Period 2	Split sample	
	Width (%)	Width (%)	Width (%)	Width (%)	Accuracy (%)
Parallel linear reservoirs (two stores)	1.11	0·97	0·98	0·97	78
Parallel linear reservoirs (three stores)	NA	0·85	1·52	1·09	86
TOPMODEL	NA	0·66	1·11	0·98	78
TOPMODEL (fixed evapotranspiration)	1.27	0·75	1·27	0·96	68
TOPMODEL (with slow reservoir)	NA	0·50	0·69	0·66	68
TOPMODEL (with overland flow delay)	NA	0·72	1·1	0·99	75
Generalized TOPMODEL ($n = 1$)	1.05	0·75	1·11	1·02	84
Generalized TOPMODEL ($n = 2$)	NA	0·66	1·17	0·88	73
Generalized TOPMODEL ($n = 10$)	NA	0·67	1·04	0·96	78
Generalized TOPMODEL ($n = 100$)	NA	0·70	1·08	0·96	79

For the split sample test, when parameters from Period 1 are used to predict Period 2, both width and accuracy (% of observations within the prediction limits) are given. Period 1, 23/05/2002 to 04/09/2002; Period 2, 11/06/2004 to 23/09/2004; NA, parameter sets with a Nash–Sutcliffe efficiency above −1 are unable to generate prediction limits that bracket 90% of the observations.

flow. As a result, the dominant surface runoff generation process is saturation excess overland flow. Identifying saturated areas in the landscape is therefore essential, and field verification has shown that the topographic index (Beven and Kirkby, 1979) is able to identify such areas. Therefore, the semi-distributed conceptual hydrological model TOPMODEL (Beven and Kirkby, 1979) was chosen as the basis for a more complex representation of the catchment.

TOPMODEL is a frequently used model, based on simple physical approximations, and is well documented in the literature (for an overview see Beven *et al.* (1995); Beven (1997, 2001b)). It has been applied to a wide range of catchments (e.g., Ibbitt *et al.*, 2001; Bastola *et al.*, 2008). The model that was implemented uses an explicit timestep with a sequential treatment of the fluxes. First, infiltration excess overland flow is calculated. Subsequently, the water balances of the root zone, unsaturated zone and saturated zone are resolved. Evapotranspiration losses are calculated last.

Apart from the use of the topographic index, other assumptions of TOPMODEL make it a logical choice for the páramo ecosystem. The absence of a dry season (Buytaert *et al.*, 2006b), and the marked drop of soil hydraulic conductivity in non-saturated conditions result in continuously wet soils (>60 vol% (Buytaert *et al.*, 2005)). Field research has shown also shown that a saturated soil layer persists above the bedrock, even on steep slopes and in dry periods (Buytaert *et al.*, 2005). This suggests that the variation in the extent of the effective upslope contributing areas is minimal, and that the entire catchment contributes to base flow most of the time.

Finally, the high porosity and low bulk density (typically below 0·6 g/cm^3 (Buytaert *et al.*, 2006c)) give rise to easily compressible soils. Bulk density tends to rise and hydraulic conductivity tends to fall with depth (Buytaert *et al.*, 2006c), giving support to the use of a nonlinear transmissivity profile. The classic TOPMODEL assumption of an exponential function of the storage deficit appears to give a good representation of the recession curves in these catchments.

Generalized TOPMODEL

Highly organic soils often exhibit swell- and shrink properties (see review in Smiles (2000)) and may develop an extensive network of cracks and pipes (Holden and

Table IV. Description of the parameters and the sampling interval for calibration

Param.	Dim.	Model	Min.	Max.	Description
tc1	h	1,2	0	30&100[1]	Linear reservoir time constant
tc2	h	1,2,5	100	2500	Linear reservoir time constant
tc3	h	2	30	100	Linear reservoir time constant
init1	$m\ h^{-1}$	1,2	0	0	Initial flow reservoir 1
init2	$m\ h^{-1}$	1,2	0	1·5e-4	Initial flow reservoir 2
init3	$m\ h^{-1}$	2	0	1·5e-4	Initial flow reservoir 3
Reff	—	1,2	0	1	Ratio of effective over total precipitation
ratio1	—	1,2,5	0	1	Distribution over stores 1 and 2
ratio2	—	2	0	1	Distribution over stores 2 and 3
Q_0	m	3–9	0	6e-5	Initial subsurface flow 0
LnTe	$m\ h^{-1}$	3–9	−2	3	Transmissivity (log transformed)
m	—	3–9	0	0·06	Shape of the transmissivity curve
Sr_0	m	3,5–9	0	0·1	Initial root zone storage deficit
Sr_{max}	m	3,5–9	0·1	0·2	Maximum root zone storage deficit
td	$h\ m^{-1}$	3–9	0	3	Unsaturated zone time delay
v_r	—	3–9	200	2500	Channel flow velocity
CD	m	3–9	0	5	Capillary drive
K_0	m^{-1}	3–9	0	1	Surface hydraulic conductivity
v_o	—	6	1	500	Overland flow velocity
K_v	—	4	0	1·5	Vegetation coefficient for evapotranspiration

[1] 100 for the 2 store linear model, and 30 for the 3 store linear model.
The models are: 1, parallel linear reservoirs (two stores); 2, parallel linear reservoirs (three stores); 3, TOPMODEL; 4, TOPMODEL (fixed evapotranspiration); 5, TOPMODEL (with slow reservoir); 6, TOPMODEL (with overland flow delay); 7, generalized TOPMODEL ($n = 1$); 8, generalized TOPMODEL ($n = 2$); 9, generalized TOPMODEL ($n = 10$); Dim, dimensions; Param, parameter.

Burt, 2002; Uchida *et al.*, 2005). Both effects have been observed in the field for the study catchment. As a result, the response at the hill slope may be very different from a simple aggregation of laboratory measurements on small soil samples (Beven, 2006a).

The complexity of preferential flow paths and hysteretic behaviour in páramo soils, and the low data availability does not permit to construct a complex model of subsurface flow. However, it stresses the need for testing alternative hypotheses to the exponential transmissivity profile assumed by the original TOPMODEL.

Iorgulescu and Musy (1997) present equations for a power law vertical profile of hydraulic conductivity. In this formulation, the exponential profile of the original TOPMODEL is obtained as a limit case of the new general form (though see also Lamb *et al.*, (1997) as an alternative way of generalizing the transmissivity function). In this study the generalized topmodel is run with four different exponents ranging from a linear profile ($n = 1$) to $n = 100$. Higher values asymptotically approach the exponential formulation of the original topmodel (Iorgulescu and Musy, 1997).

Custom versions of TOPMODEL

Finally, three custom versions of TOPMODEL were developed to improve the representation of specific processes of the páramo wetlands. By developing separate models, the improvements can be assessed individually.

In a first version, the evapotranspiration mechanism was simplified. Earlier research has shown that soil wetness in the studied ecosystem does not fall below 60 vol% in the upper soil horizon, which is close to the saturation point of these soils (Buytaert *et al.*, 2005). It is therefore unlikely that water stress occurs. Evapotranspiration is instead limited by plant physiological properties, which are independent of the season or the soil water content. A custom version therefore replaces the soil moisture-dependent evapotranspiration mechanism by a simple vegetation coefficient K_v, linking the reference evapotranspiration (E_0) to actual evapotranspiration (E_a):

$$E_a = K_v\, ET_0 \quad (1)$$

This also enables a more explicit representation of different vegetation properties in the model, since K_v is well described for several vegetation types, particularly crops (Allen *et al.*, 1998). This is useful to predict the impact of cultivation on the hydrological response (Buytaert and Beven, 2009).

A second shortcoming of the usual implementation of TOPMODEL is the lack of accounting for overland flow delay (although this was a feature of the original Beven and Kirkby (1979) version of the model). If TOPMODEL is applied to small catchments at a high temporal resolution, the delay caused by overland flow may become an important factor in the timing of peak discharge. High-resolution digital elevation models enable the extraction of the distribution of flow lengths from all pixels to the nearest river cell (e.g. (Quinn *et al.*, 1991)). In combination with an overland flow velocity parameter, this distribution can be converted to the time domain. Finally, overland flow is convoluted through the overland flow delay function before being routed through the channel. This enhancement requires one additional parameter, the overland flow velocity.

A last version of TOPMODEL improves the representation of hydrological connectivity in the landscape.

The irregular landscape gives rise to numerous hydrologically disconnected depressions and swamps. They trap overland flow and act as an additional storage reservoir with a very slow release. It is hypothesized that these depressions contribute significantly to base flow during extensive periods without rainfall (Mena and Medina, 2001; Hofstede *et al*., 2003).

The original TOPMODEL does not represent this process. Individual depressions are often irregular and small and therefore difficult to identify if no very high-resolution digital elevation model is available (Lane *et al*., 2004, 2009). Instead, a conceptual representation was used. Part of the precipitation is diverted to this reservoir and routed through a linear reservoir with a large residence time. This process adds two additional parameters to the model (Tables II and IV).

THE HYPOTHESES TO BE TESTED

The nine model structures outlined in Table II are used here to test a number of different hypotheses related to the complexity of the different model representations. First, the ability of TOPMODEL to model the hydrology of the páramo is evaluated. Subsequently, alternative representations and improvements are tested:

H1: The TOPMODEL structure is a good over-all representation of the hydrology of the páramo
H2: Recession curves can be represented by one or more linear reservoirs
H3: A generalized power law transmissivity profile is a better representation than an exponential profile
H4: Adding a linear reservoir to TOPMODEL to represent lakes and swamps improves predictions
H5: Antecedent moisture conditions are an important factor in predicting peak flows
H6: An overland flow delay function improves the timing of modelled peak flows
H7: Actual evapotranspiration is independent of soil water contents

Model evaluation

There are some important issues in testing hypotheses of this type in hydrological modelling. One is the issue of equifinality (Beven, 2006a): that for any one of the model structures tested there may be many different parameter sets giving acceptable fits to the discharge observations. To address the problems of equifinality, all models were compared using two methods.

The first method consists of comparing the optimal Nash–Sutcliffe (NS) efficiency of all models. Despite its shortcomings (e.g. McCuen *et al*., 2006; Schaefli and Gupta, 2007; Stedinger *et al*., 2008; Gupta *et al*., 2009), it is a widely used performance measure that is easy to interpret and to compare with other modelling studies. Other studies suggest, however, that the efficiency should only be used as an index of overall performance, as there may be many different parameter sets that achieve equally good performance. Beven (2006a, 2010) has suggested instead, a way of testing models as hypotheses using the limits of acceptability approach. Ideally such limits of acceptability are set before running the model. Here, however, as in many similar modelling studies, we expect that a major source of uncertainty comes from the input data supplied to the model. Since there are no assessments of such input errors independent of the model, here the set of behavioural models was instead determined such that their range of predictions brackets a fixed proportion of the discharge observations (e.g. 90%), the width of the uncertainty bounds can be used directly to compare the quality of different models. Indeed, narrow uncertainty bounds that bracket the observations indicate a good representation of the modelled system.

In this study, the Generalized Likelihood Uncertainty Estimation method (GLUE, see Beven and Binley, 1992; Beven and Freer, 2001b, among many others) is used to generate uncertainty bounds. In GLUE, this is done by calculating weighted quantiles from the predictions of an ensemble of *behavioural* parameter sets. Both the weights and the behavioural limit of parameter sets are generally derived from a performance measure over the model calibration period. Again, the NS efficiency was used here. The behavioural limits were chosen such that the uncertainty bounds bracket 90% of all the observations used in the evaluation. For the split sample test, the behavioural parameter sets of period 1 (23/05/2002 to 04/09/2002) were used to generate prediction limits for period 2 (23/05/2004 to 04/09/2004).

The second method has an added advantage. Sometimes, a model may give reasonable NS efficiencies, but be unable to generate 90% uncertainty limits using reasonable parameter sets. This can happen for instance when the model has a strong bias or systematically fails to represent some parts of the hydrograph (e.g. peak flows). This is one way of identifying model failure that would be less obvious when using only the global optimal NS efficiency.

A second issue in model evaluation is that of input error: that the sets of parameter values in any model structure that give acceptable fits to the discharge observations may depend on the particular sequence of (non-stationary) errors in precipitation and (less so) evapotranspiration forcing data in the calibration period. The use of a split record test will, to some extent, guard against overfitting to such errors when a model can be shown to also give acceptable results in the second evaluation period, though it is expected that such epistemic errors will lead to reduced performance in the test period so that the prediction limits will not necessarily bracket the same number of observations. Therefore both the accuracy and the width of the prediction limits are reported.

A third issue, that of the influence of model structure error, is of less significance here in that the success of the different model structures is being compared and evaluated.

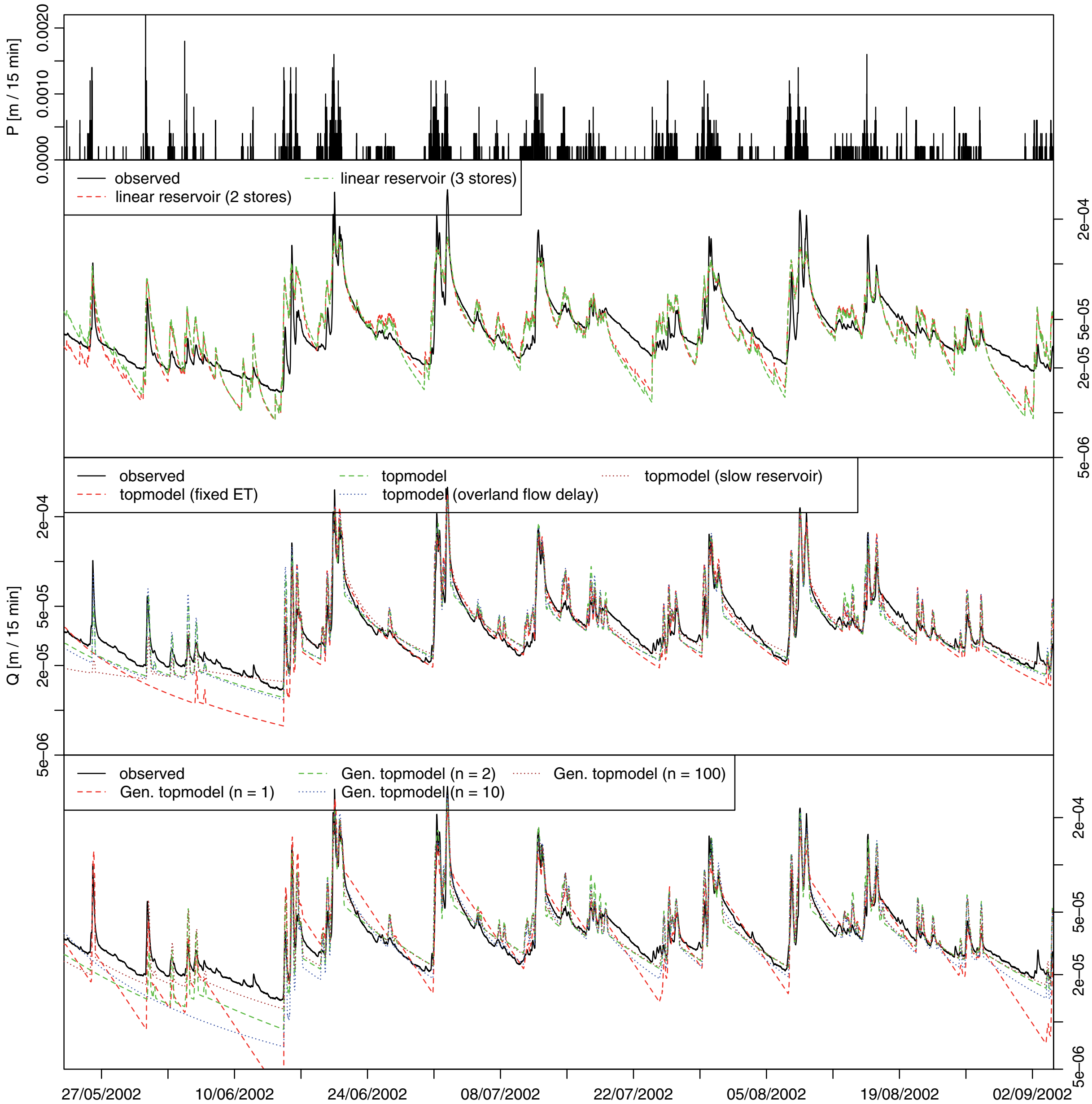

Figure 4. Observed and simulated discharge from the Huagrahuma catchment during the period 23/05/2002 to 04/09/2002

RESULTS AND DISCUSSION

The maximum NS efficiency and the width of the 90% prediction limits are given for each model in Tables II and III. The total modelling period, the two subperiods and the split sample test are reported. Figure 4 shows the predicted discharges for period 1 of the model runs with maximum NS efficiencies. Figure 5 shows the 90% uncertainty bounds for selected models for the same period. For reasons of conciseness, the three-reservoir model, and the alternative and generalized topmodel versions were left out as they do not perform better than the shown models.

Remarkably little difference in performance is observed between the different model structures. For the total period, the NS efficiency ranges from 0·58 to 0·69. Most of the model structures perform better during the short periods, with efficiencies up to 0·89 for TOPMODEL enhanced with an additional slow reservoir. All the models perform less well in the evaluation period of the split sample test, but the linear models are more robust in this respect than the versions of TOPMODEL in both efficiency and bias. The performance in the uncertainty analysis is broadly compatible with the optimal NS values. Models with a high maximum NS tend to generate the tightest uncertainty boundaries. The inability of many models to generate 90% prediction boundaries for the entire observation period is probably related to the lack of proper representation of seasonal variability, particularly in the evapotranspiration data, and taking no explicit account of the (unknown) rainfall input errors.

The following sections discuss to which extent the different hypotheses about the hydrological behaviour of the páramo can be verified by analysing the performance of the different model structures. The performance during

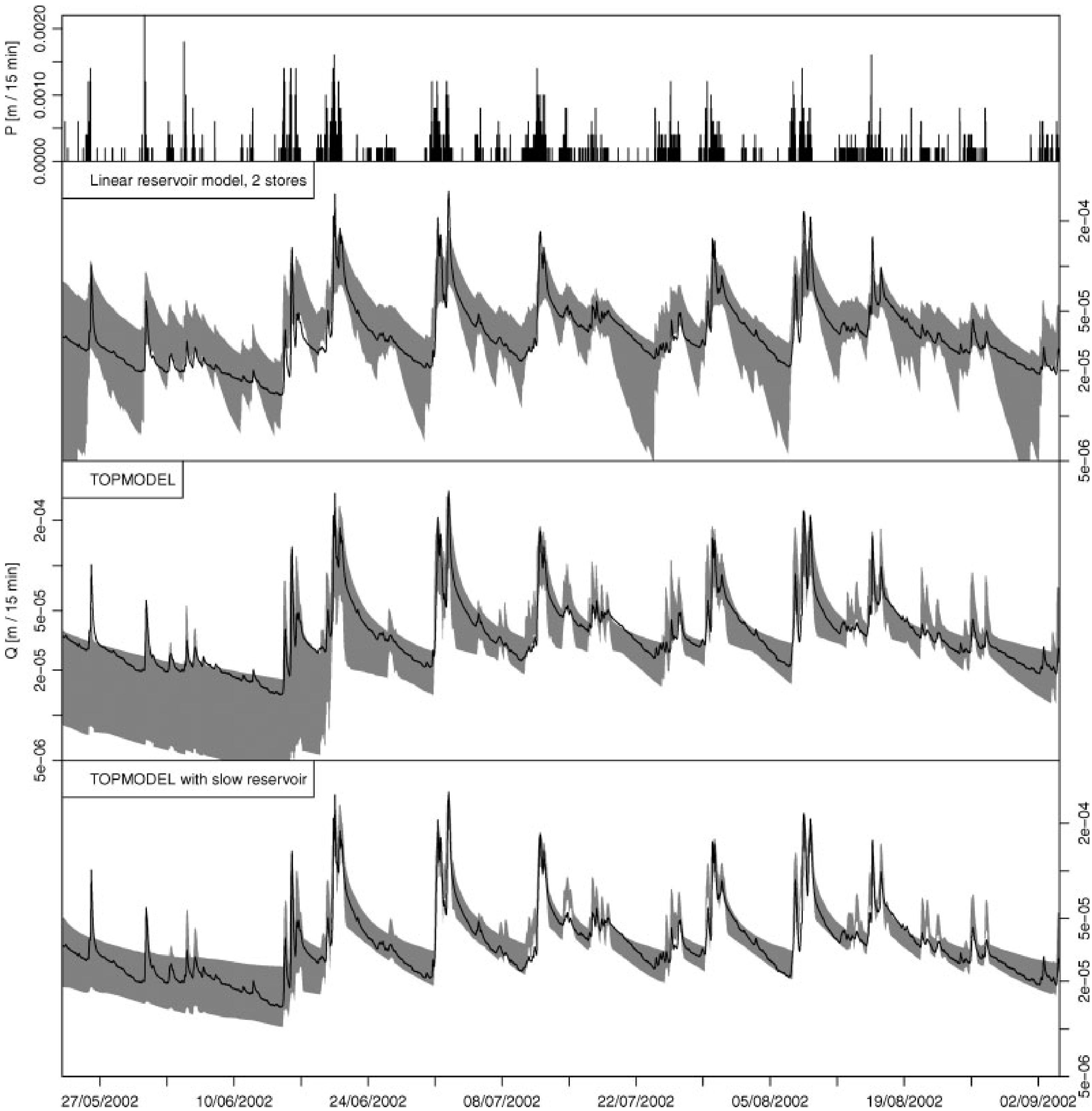

Figure 5. Observed discharge and 90% uncertainty limits of selected model structures over the period 23/05/2002 to 04/09/2002

the evaluation period is discussed in the section on 'Model Predictive Capacity'.

General representation of the hydrological response

H1: The TOPMODEL structure is a good over-all representation of the hydrology of the páramo. TOPMODEL was chosen as the basis for the current modelling effort because of the importance of saturated overland flow, the high gradients in slope and the shallow soils. Calibrating TOPMODEL yields a maximum NS efficiency of 0·67 for the entire modelling period, and respectively 0·87 and 0·77 for the sub-periods (Table II). In view of the large number of time steps (132 650 for the entire period), this can be considered a reasonable performance. The combination of a small catchment with a quick response and a short time step of 15 min makes the NS measure sensitive to errors in the timing of the peakflow, which can have a strong negative impact on the model performance. Also, the internal storage capacity of hydrological systems has generally a smoothing effect on errors in the input data, but this is far less the case in small catchments such as Huagrahuma. In such catchments, epistemic errors in the input propagate to the discharge predictions, especially in peak flows. In irregular mountain areas the resulting input errors may be large and different in the calibration and evaluation periods of the split record test.

In the next sections, the behaviour of TOPMODEL is discussed in detail, and compared to the linear reservoir models.

Representation of baseflow

H2: Recession curves can be represented by one or more linear reservoirs. On the basis of catchment observations, we hypothesized in the introduction that hydrological flows in the páramo consists of different contributions, such as surface runoff, subsurface response from hill slopes, and small, continuously wet and hydrologically disconnected swamps and lakes. Such processes are often represented by linear stores (e.g. (Clark *et al*., 2009)).

The linear reservoirs indeed perform reasonable well with very little difference between the two model structures. Both structures represent the recession curves of most major precipitation events well in the first day(s) after the event (Figure 4).

A first observation is the small difference between the two- and three-store models, suggesting a bimodal response, at least during the steep part of the recession curve. Although there is little justification for assigning processes to the stores, field observation suggests that surface runoff is minimal. The dense vegetation cover of the catchment allows very little runoff to enter the drainage system without passing through the litter layer. This may particularly be the case for rock outcrops, which tend to be located at the higher parts of the catchment.

A second observation is the marked divergence of the linear model predictions from the observed discharge more than a few days after precipitation events. This suggests that different mechanisms dominate before and after this time span, which may be related to slow hill slope contribution and/or lakes and swamps. This may imply that the linear model requires a component with a time delay, or a Nash cascade rather than a simple linear store.

A manual calibration of the three-store model forcing larger residence times (results not shown) reveals that the recession curve and base flow can be represented using average residence times of around 6·3, 103 and 2500 h. But this calibration comes at the expense of a proper peak flow and water balance representation and a lower efficiency of the model (0·68). This suggests that a more physically based representation is required.

H3: A generalized power law transmissivity profile is a better representation than an exponential profile. As explained in the introduction, páramo soils are very porous, and prone to the formation of preferential flow paths and hysteresis. This may invalidate the exponential transmissivity profile of the original TOPMODEL for the páramo region.

However, alternative hypotheses of the transmissivity profile as represented by the generalized TOPMODEL versions are also unable to improve the representation of the baseflow (Figure 4). The linear models already confirmed that a linear representation of subsurface flow ($n = 1$) is not a good representation of subsurface runoff. A Generalized TOPMODEL with exponents two and ten performs equally well, but in both cases the uncertainty limits have problems with bracketing the observed observations (Table III) suggesting systematic bias in the predictions. We conclude that an additional component is required in the model (H4).

H4: Adding a linear reservoir to TOPMODEL to represent lakes and swamps improves predictions. Compared to the linear reservoir models, TOPMODEL represent baseflow in a very different way (Figure 4). Standard TOPMODEL shows a dichotomous behaviour with a fast response that drops quickly and a sustained baseflow that mimics well the observed recession curve from a few days after the precipitation event. Adding a slow reservoir increases the model efficiency only slightly (Table II).

However, the addition of an extra slow reservoir markedly improves the performance of TOPMODEL in the split sample test. One disadvantage of adding complexity (and therefore more parameters) to a model is the risk for overparameterization (Beven, 2006a). More model parameters increases the risk that parameter optimization converges towards unrealistic parameter values due to parameter interaction and compensation of errors. Such compensation will only be beneficial to the model if the errors are stationary. A split sample test can provide some insight, as errors are likely not stationary in time between the calibration and validation period. The observation that the improved TOPMODEL performs well in the split sample test compared to simpler versions adds weight to the assumption that the model is a better representation of the catchment rather than being overparameterized. However, it does not compare well with the linear stores in the split sample test (Tables II and III). The latter are more robust in terms of NS efficiency, bias and accuracy. This is discussed in the section on 'Model Predictive Capacity'.

In the optimized model, the fraction of rainfall routed to the slow reservoir is 0·2, and the average residence time is 1929 h, or around 80 days. This would mean that 20% of rainfall is routed to lakes and swamps that contribute to river discharge as a near constant baseflow (Figure 6). The residence time is compatible with the residence time of the manually calibrated slow pathway of the three-store linear model (2500 h). Such large time constants result in a very slow hydrograph recession and are therefore not very sensitive parameters.

It is difficult to verify such a process by observation in the field. An analysis of the topography suggests that a large part of the catchment is routed through the lakes and swamps, many of which are disconnected during at least part of the year. In view of this, an area of 20% of the catchment supplying runoff to lakes and swamps seems realistic. However, it is not known whether the swamps and lakes themselves are responsible for the large residence time, or whether slow drainage on the hillslopes towards these areas, or from the upper slopes via bedrock fractures, is the main process. Additional experiments such as chemical tracer experiments could yield more insight in the response time of different locations of the catchment.

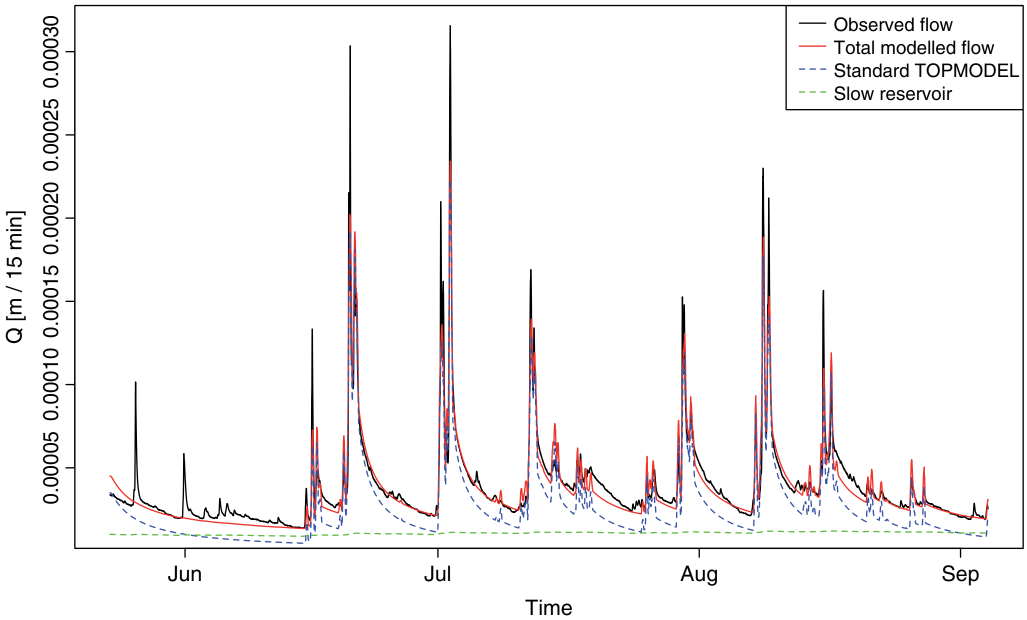

Figure 6. Performance and dissection of the flows generated by the TOPMODEL version enhanced with an additional slow reservoir

Such experiments are necessary to improve the model, since a linear reservoir with an input proportional to the rainfall inputs is still a very rudimentary representation of the actual process. For instance, one of the processes that is not taken into account explicitly is hydrological connectivity. It is likely that connectivity varies during the year. During the wet season, the water table in the disconnected swamps rises, connecting some of them to the river network. This process is currently not represented by the linear reservoir, and may be a reason why the model performance is less satisfactory during the dry season (Figure 4).

A better representation of the dynamics of connectivity would be required to improve the model. Such indices are the topic of ongoing research (e.g. (Lane *et al*., 2009)), but can only be applied if a sufficiently high-resolution digital elevation model is available and the soil surface provides a good representation of the bedrock surface connectivity (Freer *et al*., 2002).

Overland flow

H5: Antecedent moisture conditions are an important factor in predicting peak flows. Although infiltration excess overland flow is nearly absent in the páramo ecosystem (Buytaert *et al*., 2008), surface runoff can be an important contributor to river discharge due to the numerous saturated areas in a typical páramo landscape. These areas expand and shrink depending on the overall catchment wetness, and therefore strongly regulate the amount of surface runoff.

Figures 4 and 5 show that the linear reservoir models overestimate peak flows during the drier periods (e.g. June and September in Figure 4) while underestimating peak flows during wetter periods. This is not surprising. The linear reservoir assumes a constant ratio between peak and base flows, without taking into account antecedent soil moisture conditions. This deficiency generates a more constant pattern than observed in the field (Figure 5).

Despite the strong emphasis of TOPMODEL structure on the relation between soil moisture content and overland flow production, prediction of peak flows are only marginally improved. Wet season peak flows are well represented but dry season peak flows are still overestimated.

Hydrological connectivity is also here a likely explanation. During drier periods, more areas prone to saturation become hydrologically disconnected, trapping overland flow and reducing peak flows. Either a dynamic implementation of the contributing area as in Dynamic Topmodel (Beven and Freer, 2001a), or data collection on the spatial distribution of soil properties (Lamb *et al*., 1998; Blazkova *et al*., 2002) may improve the predictions.

H6: An overland flow delay function improves the timing of modelled peak flows. In none of the three calibration periods does the TOPMODEL version with an additional overland flow delay function perform better than the standard TOPMODEL (Table II). This is somewhat surprising. With a timestep of only 15 min, and a small residence time in the channel network, the delay of overland flow would be expected to have a notable impact.

Closer scrutiny of the parameter behaviour reveals a significant negative correlation between the value of the overland flow velocity and channel velocity in the behavioural models. It is clear that an interaction between routing the overland flow and channel flow occurs. Indeed, the optimal channel velocity for the standard TOPMODEL is around 500 m h^{-1}, depending on the calibration period. This value is very low for torrential streams in high-mountain environments. Clearly the low value compensates for the delay due to overland flow.

Therefore, although separation of both processes makes sense from a physical viewpoint, it does not improve model performance because of the high tendency for parameter interaction.

Water balance

H7: Actual evapotranspiration is independent of soil water contents. Keeping the evapotranspiration mechanism constant in TOPMODEL has a small degradational effect on model performance. This observation rejects the hypothesis that actual evapotranspiration is independent of soil water contents. The most likely explanation is that the impact of saturated areas and wet vegetation on evaporation, which seems to have a stronger effect on evapotranspiration than hypothesized in the introduction.

The potentially large errors in precipitation caused by the mountainous terrain may generate erroneous peak flows in the simulations that are given a strong weight in the performance measure and skew the final prediction. The linear reservoirs tend to underpredict peak flows (Figure 4), particularly during the wet period, and are therefore less affected by errors in rainfall inputs.

Additionally, occult precipitation and fog interception are thought to have a non-negligible impact in many páramos and may further complicate the water balance (Bruijnzeel and Veneklaas, 1998; Buytaert *et al*., 2006a).

Model predictive capacity

Although the seasonality of the páramo climate is very low (Buytaert *et al*., 2006b), model parameterization may be affected by temporal variability in the dominant hydrological processes. The split sample test (Tables II and III) gives an indication of the robustness of optimized model structures and parameter sets for prediction purposes.

Models with a higher number of parameters tend to fit the data better, but their performance may be affected by over-fitting and parameter interaction. These effects may cause a poor model predictive capacity in the split record evaluation test since they will degrade the model's representation of the hydrological system. The split sample test shows indeed a larger degradation of the performance of the TOPMODEL versions compared to the linear reservoir models (Tables II and III). Even TOPMODEL enhanced with an additional slow reservoir, although outperforming the other TOPMODEL versions, has a higher bias, a lower NS efficiency and a lower accuracy than the linear reservoirs.

So are all the versions of TOPMODEL over-fitted? After all, if models are to be used for prediction, it is more important that they are robust, rather than simply that they can reproduce adequately the calibration data. This may then come at an expense of wider prediction limits (Table III and Figure 7).

There is, however, a fundamental difficulty testing models of this type as hypotheses about catchment response. As noted earlier, there will always be epistemic non-stationarities when a model is migrated in space or time (Buytaert and Beven, 2009). Errors in the input data may have changed, for instance due to recalibration, ageing, or replacement of instruments, or due to different precipitation patterns. Evolution of the catchment vegetation may alter various hydrological processes including evapotranspiration and infiltration. Finally, some hydrological processes, particularly those controlled by thresholds, may have been inactive during the calibration or the validation period. Thus, we should *expect* models to perform less well in a split record evaluation test. This is the case in nearly all published hydrological model applications that show such results. When using models in a hypothesis testing framework, it is therefore important to identify such errors, since they are a primary source of information for model improvement and might be disinformative in model calibration (Beven, 2008, 2010).

Figure 7 indeed shows valuable insights in the behaviour of the enhanced TOPMODEL. A main observation is the fact that the TOPMODEL version enhanced with a slow reservoir represents the form of the recession curve at least as adequately as the classic TOPMODEL and the linear stores. However, the model consistently overestimates the observed flow. This is also observed for the other two models (e.g. at the end of June) but most of the time the wide uncertainty bounds mask the problem. Model failure seems therefore a result of problems with the water balance. We identify several hypotheses for this failure:

- The model failure emerges after a major storm event at the beginning of June, of which only part of the recession curve was monitored. Such an event did not occur in the calibration period (Table I). During the event, lakes and swamps may rise above normal levels, and shortcut overland flow pathways. This would result in a larger proportion of overland flow than normally observed and proportionally less infiltration. The model would therefore overestimate internal storage, thus overestimating flows in the period after the event. The enhanced TOPMODEL is particularly vulnerable to wrong internal states because of the large persistence of the slow reservoir. In general, the evaluation period (period 2) experienced a wider range of flows than the calibration period (period 1, Table I), which may result in a model not properly calibrated for both low flow and high flow extremes.
- Spatial variability of precipitation is often very large in mountain areas. A constant bias in the precipitation pattern during the wet period, e.g. due to a dominant wind direction, may have resulted in large model input errors. Again, such error would persist longer in the enhanced TOPMODEL.
- Finally, the sensor was replaced between the calibration and evaluation period for one of lower quality. This may have lead to errors in the calibration curve and bias in the observation data.

It is clear that further research is needed before a decisive conclusion can be drawn about the predictive

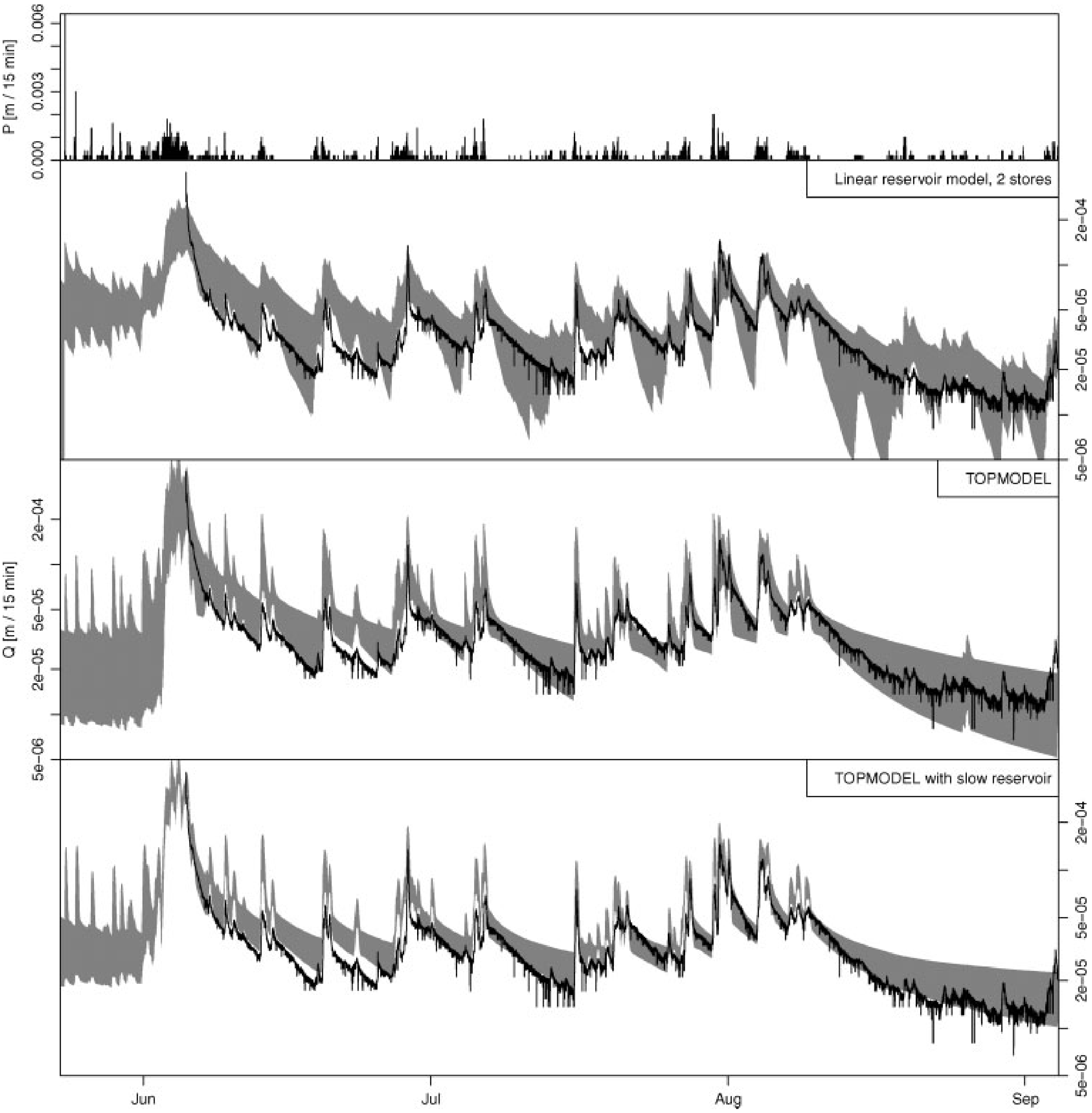

Figure 7. Observed discharge and 90% prediction limits of selected model structures over the period 07/02/2003 to 13/08/2003 using parameter sets from the period 23/05/2002 to 04/09/2002

capacity of the tested models. In the current study, insufficient information is available for further investigation of the hypotheses about this specific event. However, they provide directions for further research and model improvement when new data will be available for similar events in the future.

CONCLUSIONS

The tropical alpine wetlands of the northern Andes are a unique hydrological system. Hydrological modelling of these areas is an essential tool to improve local water resources management. It is also scientifically challenging because of the large spatial variability of precipitation, specific soil properties and a very irregular topography.

This study uses a comparison of nine optimized model structures to test seven hypothesis about the hydrological behaviour of the system at a catchment scale. Scrutinizing the model performance in different parts of the hydrograph reveals important insights in the hydrological behaviour of the ecosystem:

- The hydrological response of the páramo, with its consistently high runoff ratios, can be decomposed in three response modes, which are related to (1) a combination of overland and interflow, (2) hillslope response and (3) a very slow response of disconnected wetlands and depressions. However, a model consisting of three parallel linear reservoirs does not properly capture the non-linear response of the system, and also fails to capture the seasonal variability of overland flow generation.

- Saturated overland flow is a dominant process in the páramo ecosystem. The fact that TOPMODEL is able to capture the seasonal variability of peak flows better than the linear reservoir models suggests that there is a relation between catchment wetness and dynamic contributing areas for fast runoff. However, a correct estimation of peak flow remains a challenge. These problems are attributed to the high spatial variability in precipitation in the region, leading to errors in the model's boundary conditions.
- Because of the observations of preferential flow paths and pipeflow in hillslopes of the páramo, TOPMODEL's assumption of an exponential decline of transmissivity with depth appears a reasonable representation of subsurface flow in reproducing the hydrograph recession characteristics. This may indicate that locally observed heterogeneity is averaged out at the catchment scale.

This paper has underscored the difficulties of doing hypothesis testing in the representation of runoff processes by models when the data available to characterize catchment processes are limited and the forcing data required to run a model are uncertain (see also Piñol *et al*., (1997) and Beven (2001a, 2010)). Different model representations are different in different ways, and the way in which epistemic forcing error interacts with epistemic model structural error is manifest in the way that performance in calibration periods is nearly always better than in split record evaluation periods, even though the results for both might deemed to be reasonable or acceptable for use in prediction.

It is interesting to note that in the original TOPMODEL paper (Beven and Kirkby, 1979) optimization of the model parameters resulted in a set of parameter values that forced the model to use the subsurface storage to represent a fast saturation overland flow response (albeit in a catchment where the response was flashier than those considered here). Field estimated parameter values, were sub-optimal in terms of the NS efficiency but resulted in the structure of the model being used in the way originally intended. This is an indication that, in testing model hypotheses, we must be careful about getting the right results for the right reasons (see Kirchner (2006)).

Nevertheless, this study has revealed that certain model structures perform better than others and that the addition of a slow parallel store to the original TOPMODEL appears the most realistic representation of the system to date. In this formulation, about 20% of the total areal precipitation is re-routed to a reservoir with a residence time in the order of months, providing a nearly constant baseflow. This reservoir is thought to represent slowly draining closed depressions and small ponds that are typical for the páramo. The resulting model performs satisfactorily with NS efficiencies up to 0·89. Remaining errors are attributed to errors in the model inputs, and the rudimentary representation of inferred effect of the ponds and swamps that are not connected to the channel network.

This is, however, only an inference and suggests that more types of data should be used in evaluating the process representations (Lamb *et al*., 1998; Blazkova *et al*., 2002). Mapping of saturated areas and their connectivity might provide more information with which to test the dynamics of the fast responses. Tracer experiments might reveal more about the connectivity and time scales of water fluxes from different types of saturated area to the stream channel. Further data collection might also reduce the uncertainties associated with the forcing data, or allow different realizations of the input errors to be generated for use in hypothesis testing, an issue that will be pursued in future work.

ACKNOWLEDGEMENTS

Buytaert was funded by a Marie Curie EIF Fellowship at Lancaster University during part of the research. All models were developed using the R data analysis and computing environment. We would like to thank the R community for developing the R software and the additional packages. The models and related computer code are available at http://www.paramo.be. W.B. acknowledges the Universidad de Cuenca for support during the field work.

REFERENCES

Allen RG, Pereira LS, Raes D, Smith M. 1998. Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56. FAO, Rome.

Andréassian V, Lerat J, Loumagne C, Mathevet T, Michel C, Oudin L, Perrin C. 2007. What is really undermining hydrologic science today? *Hydrological Processes* **21**: 2819–2822.

Bastola S, Ishidaira H, Takeuchi K. 2008. Regionalisation of hydrological model parameters under parameter uncertainty: a case study involving topmodel and basins across the globe. *Journal of Hydrology* **357**: 188–206.

Beven KJ. 1997. Topmodel: a critique. *Hydrological Processes* **11**: 1069–1086.

Beven KJ. 2001a. On hypothesis testing in hydrology. *Hydrological Processes* **15**: 1655–1657.

Beven KJ. 2001b. *Rainfall-runoff Modelling: The Primer*. John Wiley & sons, Ltd.: Chichester.

Beven KJ. 2006a. A manifesto for the equifinality thesis. *Journal of Hydrology* **320**: 18–36.

Beven KJ. 2006b. On undermining the science? *Hydrological Processes* **20**: 3141–3146.

Beven KJ. 2006c. Searching for the holy grail of scientific hydrology: $Q_t = H(\underleftarrow{S}\ \underleftarrow{R})A$ as a closure. *Hydrology and Earth System Sciences Discussions* **3**: 769–792.

Beven KJ. 2008. On doing better hydrological science. *Hydrological Processes* **22**: 3549–3553.

Beven KJ. 2009. *Environmental Modelling: An Uncertain Future?* Routeledge: London.

Beven KJ. 2010. Preferential flows and residence time distributions: some issues in defining adequate hypothesis tests for hydrological models. *Hydrological Processes* **24**: 1537–1547.

Beven KJ, Binley A. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**: 279–298.

Buytaert W, De Bièvre B, Célleri R, Cisneros F, Wyseure G, Deckers S. 2008. Comment on "Human impacts on headwater fluvial systems in the northern and central Andes". (Carol P. Harden. Geomorphology 79: 249–263). *Geomorphology* **96**: 239–242.

Beven KJ, Freer J. 2001a. A dynamic topmodel. *Hydrological Processes* **15**: 1993–2011.
Beven KJ, Freer J. 2001b. Equifinality, data assimilation and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. *Journal of Hydrology* **249**: 11–29.
Beven KJ, Kirkby MJ. 1979. A physically based variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin* **24**: 43–69.
Beven KJ, Lamb R, Quinn P, Romanowicz R, Freer J. 1995. Topmodel. In *Computer Models of Watershed Hydrology*, Singh VP (ed.). Water Resources Publications: Colorado; 627–668.
Blazkova S, Beven KJ, Tacheci P, Kulasova A. 2002. Testing the distributed water table predictions of topmodel (allowing for uncertainty in model calibration): the death of topmodel? *Water Resources Research* **38**: 1257.
Bragg OM. 2002. Hydrology of peat-forming wetlands in Scotland. *Science of the Total Environment* **294**: 111–129.
Bruijnzeel LA, Veneklaas EJ. 1998. Climatic conditions and tropical montaine forest production: the fog has not lifted yet. *Ecology* **79**: 3–9.
Buytaert W. 2004. The properties of the soils of the south Ecuadorian páramo and the impact of land use changes on their hydrology. PhD thesis, Faculty of Agricultural and Applied Biological Sciences, Katholieke Universiteit Leuven.
Buytaert W, Beven KJ. 2009. Regionalization as a learning process. *Water Resources Research* **45**: W11419.
Buytaert W, Célleri R, De Bièvre B, Hofstede R, Cisneros F, Wyseure G, Deckers J. 2006a. Human impact on the hydrology of the Andean páramos. *Earth-Science Reviews* **79**: 53–72.
Buytaert W, Célleri R, Willems P, De Bièvre B, Wyseure G. 2006b. Spatial and temporal rainfall variability in mountainous areas: a case study from the south Ecuadorian Andes. *Journal of Hydrology* **329**: 413–421.
Buytaert W, Deckers J, Wyseure G. 2006c. Description and classification of highly weathered Andosols in the south Ecuadorian páramo. *Geomorphology* **73**: 207–221.
Buytaert W, De Bièvre B, Wyseure G, Deckers J. 2004. The use of the linear reservoir concept to quantify the impact of land use changes on the hydrology of catchments in the Ecuadorian Andes. *Hydrology and Earth System Sciences* **8**: 108–114.
Buytaert W, De Bièvre B, Wyseure G, Deckers J. 2005. The effect of land use changes on the hydrological behaviour of Histic Andosols in south Ecuador. *Hydrological Processes* **19**: 3985–3997.
Buytaert W, Iñiguez V, De Bièvre B. 2007. The effects of afforestation and cultivation on water yield in the andean páramo. *Forest Ecology and Management* **251**: 22–30.
Célleri R. 2007. Rainfall variability and rainfall-runoff dynamics in the paute river basin—southern ecuadorian andes. PhD thesis, Faculty of Engineering, Katholieke Universiteit Leuven.
Célleri R, Willems P, Buytaert W, Feyen J. 2007. Space-time variability of rainfall in the Paute river basin of South Ecuador. *Hydrological Processes* **21**: 3316–3327.
Clark MP, Kavetski D. 2010. Ancient numerical daemons of conceptual hydrological modeling: 1. fidelity and efficiency of time stepping schemes. *Water Resources Research* **46**: W10510.
Clark MP, Rupp DE, Woods RA, van Meerveld HJT, Peters NE, Freer JE. 2009. Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale. *Hydrological Processes* **23**: 311–319.
Clark MP, Slater AG, Rupp DE, Woods RA, Vrugt JA, Gupta HV, Wagener T, Hay LE. 2008. Framework for understanding structural errors (fuse): a modular framework to diagnose differences between hydrological models. *Water Resources Research* **44**: W00B02.
Coltorti M, Ollier CD. 2000. Geomorphic and tectonic evolution of the Ecuadorian Andes. *Geomorphology* **32**: 1–19.
FAO/ISRIC/ISSS. 1998. World Reference Base for Soil Resources. No. 84 in World Soil Resources Reports. FAO, Rome.
Farley KA, Kelly EF, Hofstede RGM. 2004. Soil organic carbon and water retention after conversion of grasslands to pine plantations in the Ecuadorian Andes. *Ecosystems* **7**: 729–739.
Fenicia F, Savenije HHG, Matgen P, Pfister L. 2008. Understanding catchment behavior through stepwise model concept improvement. *Water Resources Research* **44**: W01402.
Freer J, McDonnell JJ, Beven KJ, Peters NE, Burns DA, Hooper RP, Aulenbach B, Kendall C. 2002. The role of bedrock topography on subsurface storm flow. *Water Resources Research* **38**: 1–16.
Götzinger J, Bárdossy A. 2008. Generic error model for calibration and uncertainty estimation of hydrological models. *Water Resources Research* **44**: W00B07.
Gupta HV, Kling H, Yilmaz KK, Martinez GF. 2009. Decomposition of the mean squared error and nse performance criteria: implications for improving hydrological modelling. *Journal of Hydrology* **377**: 80–91.
Gupta HV, Sorooshian S, Yapo PO. 1998. Towards improved calibration of hydrological models: multiple and noncommensurable measures of information. *Water Resources Research* **34**: 751–763.
Harden CP. 2006. Human impacts on headwater fluvial systems in the northern and central andes. *Geomorphology* **79**: 249–263.
Hofstede RGM, Mondragon MX, Rocha CM. 1995. Biomass of grazed, burned, and undistributed páramo grasslands, Colombia. 1. aboveground vegetation. *Arctic and Alpine Research* **27**: 1–12.
Hofstede R, Segarra P, Mena PV. 2003. Los Páramos del Mundo. Global Peatland Initiative/NC-IUCN/EcoCiencia, Quito.
Holden J, Burt TP. 2002. Laboratory experiments on drought and runoff in blanket peat. *European Journal of Soil Science* **53**: 675–689.
Holden J, Evans MG, Burt TP, Horton M. 2006. Impact of land drainage on peatland hydrology. *Journal of Environmental Quality* **35**: 1764–1778.
Hungerbühler D, Steinmann M, Winkler W, Seward D, Eguez A, Peterson DE, Helg U, Hammer C. 2002. Neogene stratigraphy and Andean geodynamics of southern Ecuador. *Earth-Science Reviews* **57**: 75–124.
Ibbitt RP, Henderson RD, Copeland J, Watt DS. 2001. Simulating mountain runoff with meso-scale weather model rainfall estimates: a New Zealand experience. *Journal of Hydrology* **239**: 19–32.
Iorgulescu I, Musy A. 1997. Generalisation of topmodel for a power law transmissivity profile. *Hydrological Processes* **11**: 1353–1355.
IUCN. 2002. High andean wetlands. Tech. rep., IUCN, Gland, Switzerland.
Jansky L, Ives JD, Furuyashiki K, Watanabe T. 2002. Global mountain research for sustainable development. *Global Environmental Change* **12**: 231–239.
Kappelle M, Horn SP, Chaverri A. 2005. Páramos de Costa Rica. INBio: Herredia.
Kirchner JW. 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* **42**: W03S04.
Klemeš V. 1986. Operational testing of hydrological simulation models. *Hydrological Sciences Journal* **31**: 13–24.
Lamb R, Beven KJ, Myrabø S. 1997. Discharge and water table predictions using a generalised topmodel formulation. *Hydrological Processes* **11**: 1145–1167.
Lamb R, Beven KJ, Myrabø S. 1998. Use of spatially distributed water table observations to constrain uncertainty in a rainfall—runoff model. *Advances in Water Resources* **22**: 305–317.
Lane SN, Brookes CJ, Kirkby MJ, Holden J. 2004. A network-index-based version of topmodel for use with high-resultion digital topographic data. *Hydrological Processes* **18**: 191–2001.
Lane SN, Reaney SM, Heathwaite AL. 2009. Representation of landscape hydrological connectivity using a topographically driven surface flow index. *Water Resources Research* **45**: W08423.
Liniger HP, Weingartner R, Grosjean M. 1998. *Mountains of the world: water towers for the 21st century*. Mountain Agenda, University of Berne: Switzerland.
Liu Y, Gupta HV. 2007. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. *Water Resources Research* **43**: W07401.
Luteyn JL. 1999. *Páramos: A Checklist of Plant Diversity, Geographical Distribution, and Botanical Literature*. The New York Botanical Garden Press: New York.
Luteyn JL, Cleef AM, Rangel O. 1992. Plant diversity in páramo: towards a checklist of páramoplants and generic flora. In *Páramo: An Andean Ecosystem Under Human Influence*, Balslev H, Luteyn JL (eds). Academic Press: London; 71–84.
McCuen RH, Knight Z, Cutter AG. 2006. Evaluation of the nash–sutcliffe efficiency index. *Journal of Hydrologic Engineering* **11**: 597–902.
McMillan H, Clark M. 2009. Rainfall-runoff model calibration using informal likelihood measures within a markov chain monte carlo sampling scheme. *Water Resources Research* **45**: W04418.
Mena P, Medina G. 2001. Los páramos en el Ecuador. In *Los páramos del Ecuador*, Mena P, Medina G, Hofstede R (eds). Proyecto Páramo: Quito; 1–24.
Messerli B, Viviroli D, Weingartner R. 2004. Mountains of the world: vulnerable water towers for the 21st century. *Ambio Special Report* **13**: 29–34.
Mitasova H, Mitas L. 1993. Interpolation by regularized spline with tension: I. theory and implementation. *Mathematical Geology* **25**: 6 641–655.

Piñol J, Beven KJ, Freer J. 1997. Modelling the hydrological response of mediterranean catchments, Prades, Catalonia—the use of distributed models as aids to hypothesis formulation. *Hydrological Processes* **11**: 1287–1306.

Podwojewski P, Poulenard J, Zambrana T, Hofstede R. 2002. Overgrazing effects on vegetation cover and properties of volcanic ash soil in the páramo of Llangahua and La Esperanza (Tungurahua, Ecuador). *Soil Use and Management* **18**: 45–55.

Poulenard J, Podwojewski P, Janeau JL, Collinet J. 2001. Runoff and soil erosion under rainfall simulation of andisols from the Ecuadorian páramo: effect of tillage and burning. *Catena* **45**: 185–207.

Quinn P, Beven KJ, Chevallier P, Planchon O. 1991. The prediction of hill slope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes* **5**: 59–79.

Sánchez-Vega I, Dillon M. 2006. Jalcas. In *Botánica Económica de los Andes Centrales*, Moraes M, Øllgaard RB, Kvist LP, Borchsenius F, Balslev H (eds). Universidad Mayor de San Andrés: La Paz; 77–90.

Savenije HHG. 2008. The art of hydrology. *Hydrology and Earth System Sciences Discussions* **5**: 3157–3168.

Schaefli B, Gupta HV. 2007. Do nash values have value? *Hydrological Processes* **21**: 2075–2080.

Sivapalan M. 2009. The secret to doing better hydrological science: change the question! *Hydrological Processes* **23**: 1391–1396.

Sklenář P, Balslev H. 2005. Superparamo plant species diversity and phytogeography in ecuador. *Flora* **200**: 416–433.

Sklenář P, Bendix J, Balslev H. 2008. Cloud frequency correlates to plant species composition in the high andes of ecuador. *Basic and Applied Ecology* DOI: 10.1016/j.baae.2007.09.007.

Smiles DE. 2000. Hydrology of swelling soils: a review. *Australian Journal of Soil Research* **38**: 501–521.

Stedinger JR, Vogel RM, Lee SU, Batchelder R. 2008. Appraisal of the generalized likelihood uncertainty estimation (glue) method. *Water Resources Research* **44**: W00B06.

Timbe E. 2004. Disgregacion temporal de datos diarios de precipitacion en microcuencas de páramo. Master's thesis, Universidad de Cuenca.

Uchida T, Tromp-van Meerveld I, McDonnell JJ. 2005. The role of lateral pipe flow in hillslope runoff response: an intercomparison of non-linear hillslope response. *Journal of Hydrology* **311**: 117–133.

Viviroli D, Dürr HH, Messerli B, Meybeck M, Weingartner R. 2007. Mountains of the world, water towers for humanity: typology, mapping, and global significance. *Water Resources Research* **43**: W07447.

Vrugt JA, Gupta HV, Bastidas LA, Bouten W, Sorooshian S. 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research* **39**: 1214.

Vuille M, Bradley RS, Keimig F. 2000. Climate variability in the Andes of Ecuador and its relation to tropical Pacific and Atlantic sea surface temperature anomalies. *Journal of Climate* **13**: 2520–2535.

Weigend M. 2004. Additional observations on the biogeography of the Amotape-Huancabamba zone in Northern Peru: defining the south-eastern limits. *Revista Peruana de Biología* **11**: 127–134.

