Hydrological field data from a modeller's perspective: Part 2. Process-based evaluation of model hypotheses

Journal:	*Hydrological Processes*
Manuscript ID:	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Clark, Martyn; National Institute of Water and Atmospheric Research Ltd, Hydrological Processes McMillan, Hilary; National Institute of Water and Atmospheric Research Ltd, Hydrological Processes Collins, Daniel; National Institute of Water and Atmospheric Research Ltd, Applied Hydrology Kavetski, Dmitri; University of Newcastle, Environmental Engineering Woods, Ross; National Institute of Water and Atmospheric Research Ltd, Hydrological Processes
Keywords:	Hydrology, Model Structure, Multi-model, Process-based, Diagnostic, Evaluation

HYDROLOGICAL FIELD DATA FROM A MODELLER'S PERSPECTIVE:

PART 2: PROCESS-BASED EVALUATION OF MODEL HYPOTHESES

Martyn P. Clark[1*], Hilary K. McMillan[1], Daniel B. G. Collins[1],

Dmitri Kavetski[2] and Ross A. Woods[1]

1.National Institute for Water and Atmospheric Research (NIWA), Christchurch, New Zealand

2.Environmental Engineering, University of Newcastle, Callaghan, NSW, Australia

To be submitted to

Hydrological Processes, April 2010

[*]Corresponding Author
NIWA, 10 Kyle Street, Riccarton, Christchurch, New Zealand
Tel: +64-3-343-7881
Email: mp.clark@niwa.co.nz

Abstract

The current generation of hydrological models has been widely criticized for their inability to adequately simulate hydrological processes. The purpose of this paper is to evaluate competing model representations of hydrological processes with respect to their capability to simulate observed processes in the Mahurangi River basin in Northland, New Zealand. This paper is the second in a two-part series, and builds upon the first paper which analysed precipitation, soil moisture, and flow data in the Mahurangi and identified the dominant hydrological processes and options for their suitable mathematical representation. The diagnostic tests presented in this study offer several insights for model selection: They reveal the importance of different hydrological processes (e.g., the importance of vertical drainage and baseflow), they provide guidance for the choice of modeling approaches (e.g., accounting for horizontal heterogeneity in soils), and they help infer appropriate values for model parameters. The approach used in this paper to evaluate competing hydrological hypotheses demonstrates the need for flexible model structures, both because a single model structure is not applicable everywhere, and because limitations in data and process understanding lead to considerable ambiguity in structural identification. The challenge for the hydrological community is to make better use of the available data to not only estimate parameter values, but also to diagnostically identify scientifically defensible model structures.

1 Introduction

The current generation of hydrological models has been widely criticized for their inability to adequately simulate hydrological processes (e.g., Grayson et al., 1992; Beven, 2002; Kirchner, 2006). The problem is multi-fold: (1) the equations in existing hydrological models often do not represent the dominant hydrological processes demonstrated to be important in experimental watersheds (McDonnell et al., 2007); (2) soil properties and vegetation characteristics have tremendous spatial variability, both within and between basins, and it is extremely difficult to relate model parameters to the available spatial information on vegetation and soils (Koren et al., 2002; Duan et al., 2006); (3) the data used to force the model (e.g., rainfall) and evaluate its performance (e.g., runoff) can contain considerable sampling and measurement errors. Few studies have attempted to disentangle data and model structural errors (see Renard et al., 2010 for a recent development), making meaningful hypothesis-testing problematic; and (4) the numerical implementation of conceptual hydrological models is often of dubious quality, frequently resulting in confounding numerical artifacts (Kavetski et al., 2006; Clark and Kavetski, 2010; Kavetski and Clark, 2010).

It is hence generally recognized that the current approach to hydrological modelling is far from satisfactory. For example, the dominant approach, both in practice and research, is to assume a model structure *a priori*, and then use the observed data solely to determine

("calibrate") the model parameters. However, most current calibration approaches scrutinise neither the assumptions underlying the model hypotheses and its constitutive relations, nor, when the model is calibrated, the assumptions underlying the selected objective function (Kavetski et al., 2006; Clark and Kavetski, 2010). This fundamentally limits insights into model error, making it difficult to discriminate among alternative hypotheses, understand sources of uncertainty and develop strategies for reducing them. The traditional approach of assuming a fixed model structure is reaching the limits of its capability and usefulness, with more flexible and adaptive approaches being of interest.

Alternative approaches have been used to improve the consistency between models and observed processes. One popular approach is the "top-down" methodology (e.g., Sivapalan et al., 2003), which uses a collection of (usually lumped) hydrological models with increasing complexity to simulate hydrological signatures across multiple temporal scales. This approach provides useful insights, but, at least in its current form, has some limitations. The top-down approach normally involves comparing models of differing complexity, rather than multiple plausible models of equivalent complexity (Klemes, 1983; Littlewood et al., 2003; Sivapalan et al., 2003). It may hence overlook model structures that are of comparable complexity, but more suitable than those examined. Put differently, the top-down approach is not designed to evaluate competing hypotheses of hydrological behaviour. Moreover, most applications of the top-down approach (e.g., Atkinson et al., 2003) do not attempt to explicitly link the hydrological signatures to different perceptual models of the dominant storages and fluxes of water within the basin, making it difficult to rigorously evaluate the extent to which the identified models actually represent the dominant hydrological processes.

Clark et al. (2008) recently introduced the Framework for Understanding Structural Errors (FUSE), a new modular approach for hydrological modelling that forces the modeller to identify the major set of decisions (necessarily subjective) for representing the specific catchment of interest, and provides multiple options for each such decision. The FUSE approach is more general than previous studies of model complexity (e.g., Desborough, 1999; Atkinson et al., 2003) because it allows comparing competing hypotheses of observed processes. The FUSE approach also differs from other modular "multi-physics" modelling approaches (e.g., Leavesley et al., 1996; Kumar et al., 2006), because it provides modularity at the level of individual *modelling decisions* rather than at the level of individual models. It therefore provides a more systematic and comprehensive support for hypothesis-testing, allowing different sub-components to be tested under different overall model structures, including varying the number of stores and their connectivity.

The purpose of this paper is to evaluate competing model representations (model hypotheses) with respect to their capability to simulate observed hydrological processes. Here, we focus on two experimental sub-catchments of the Mahurangi River basin in Northland, New Zealand. The first paper in this series analyses the precipitation, soil moisture, and flow data in this sub-catchment, identifying the dominant hydrological processes and several alternative options for their suitable mathematical representation (McMillan et al., under review). It presents a succinct set of diagnostic tests that can be applied to data in a diverse range or river basins. This second paper tests the capability of different model sub-components, integrated into several alternative "complete" model structures, to represent the hydrological processes documented in the companion paper. Consistent with Gupta et al. (2008), the focus of this paper is on evaluating sub-components of the model, rather than on analysis of overall model performance through standard metrics related to the sum-of-squared errors. These

suggestions for diagnostic tests for model structure are intended to foster a wider acceptance of the need to tailor hydrological models for each unique catchment, and the need to vary the model structure over larger modeling domains.

More generally, this two-part study illustrates the use of experimental data to guide the selection of appropriate hydrological model structures, their implementation within a modular framework, and their stringent comparison within this framework. The diagnostic tests described in this and the companion paper can be applied in other experimental basins, and the discussion of the applicability of different model structural choices will help guide model selection in poorly gauged and un-gauged basins.

The paper is organized as follows. The description of the study site and data in Section 2 is followed by a summary of the FUSE approach in Section 3. Section 4 evaluates the major modelling decisions for the Satellite-Rite Mahurangi sub-catchment and Section 5 presents streamflow simulations from the selected models. Section 6 concludes the paper with a summary of the main results and a discussion of the outstanding research issues.

2 Study site and data

2.1 The MARVEX experiment

The data used in this study were collected as part of the Mahurangi River Variability Experiment (MARVEX; Woods et al., 2001), which investigated the space-time variability of the catchment's water balance. Typical annual rainfalls over Mahurangi catchment are about 1600 mm (with convective activity in the austral summer and larger frontal systems in winter) and annual pan evapotranspiration is about 1300 mm. Catchment elevations range from sea level to 250 m, and the land use is a mixture of pasture and plantation forestry. Most

soils in the catchment are clay loams less than a meter deep. MARVEX ran from 1997-2001, and included data from a network of 29 nested stream gauges and 13 raingauges, as well as detailed measurements in different sub-areas of the basin.

As discussed in the companion paper, intensive observations during MARVEX were made in the vicinity of Satellite Station, concentrated in two headwater catchments of the Mahurangi: Satellite Left (57.3 ha) and Satellite Right (25.1 ha). These intensive observations included high resolution spatial snapshots of soil moisture variability across the hillslope, as well as high resolution time series of soil moisture at six locations throughout the two Satellite catchments (Wilson et al., 2003; Western et al., 2004). The observations also included tracer experiments, soil sampling and analysis, and measurements of rainfall, water table depth, and flow in the smaller 1.61 ha "Kauri Tree Catchment", nested within the Satellite Right catchment (Bowden et al., under review), as well as evapotranspiration data throughout the Satellite Left and Satellite Right catchments (Fong, 2001).

2.2 Forcing/response data

All model simulations in this paper are produced for the Satellite sub-catchments. The evaluation of modelling decisions is done using data from the smaller Satellite Right sub-catchment, and then the selected models are applied to both the Satellite Left and Satellite Right sub-catchments.

The model forcing/response data used are:

(i) Rainfall measurements at Toovey's, located approximately 1 km Northwest of Satellite Station;

(ii) Potential evapotranspiration (PET) estimates, calculated using the parameterization of Priestley and Taylor (1972) using data on temperature, relative humidity, and

solar radiation from New Zealand's standard network of observing stations (Tait and Woods, 2007);

(iii) Streamflow measurements from the weirs at Satellite Left and Satellite Right; and

(iv) Soil moisture time series from the three sites on the hillslope transect in Satellite Right, where soil moisture measurements are made at two depths: the first at 0-300 mm, and the second over the 200 mm of soil at the bottom of the soil profile – the deeper vertical measurement of soil moisture was made at 300-500 mm at the lower hillslope site, 320-520 mm at the middle hillslope site, and 600-800 mm at the upper hillslope site.

The rainfall and flow measurements, collected at 5-minute intervals, were averaged to hourly values. Data were examined for three water years over the period 1998-2001. The companion paper further details the rainfall, flow and soil moisture data.

One challenge for model evaluation is that observations have a different meaning to model state variables. To address this issue, the observed soil moisture data were adjusted to improve the correspondence between the point soil moisture measurements and the model soil moisture in the unsaturated zone. First, the data from the shallow and deeper measurements were averaged, with equal weight assigned to the shallow and deeper measurements. Note that the shallow sensors measure soil moisture over a depth of 300 mm, while the deeper sensors measure soil moisture over a depth of 200 mm (see the companion paper for more details). Despite this, equal weighting of the shallow and deep soil moisture measurements was deemed appropriate because the deeper measurements are likely to be effectively representative of active soil moisture variability over a larger range of soil depths

– there is arguably insufficient data to support more elaborate weighting of the shallow and deep sensors.

Second, soil moisture measurements were scaled using the following relation

$$\theta_{scal} = \frac{\theta - \theta_{wp}}{\theta_{fc} - \theta_{wp}} \quad (1)$$

where θ is the measured volumetric water content, θ_{wp} is the volumetric water content at wilting point, θ_{fc} is the volumetric wilting point at field capacity, and θ_{scal} is the scaled soil moisture.

The model soil moisture is similarly scaled as

$$S_{scal} = \frac{S_1}{\phi_{tens} S_{1,\max}} \quad (2)$$

where S_1 is the total water content in the unsaturated zone, ϕ_{tens} is the fraction of total storage as tension storage, $S_{1,\max}$ is the maximum storage in the unsaturated zone, and S_{scal} is the scaled model soil moisture. Here, the model storage in the unsaturated zone represents soil moisture over the depth $S_{1,\max} / \left(\theta_{sat} - \theta_{wp}\right)$, where θ_{sat} is the volumetric soil moisture at saturation (i.e., the total porosity), and the "tension storage" $\phi_{tens} S_{1,\max}$ used for scaling represents storage between wilting point and field capacity. Similar adjustments were used by Schaake et al. (2004), but without the normalization by the total tension storage.

An additional challenge for evaluating model simulations of soil moisture is the limited spatial representativeness of point soil moisture observations. Since the differences among individual measurement sites cannot be resolved by a lumped hydrological model, all

comparisons between model soil moisture simulations and scaled soil moisture observations are performed using the three hillslope sites simultaneously. That is, the inter-site differences among soil moisture are explicitly included in the comparisons to account for the scale mis-match between the observations and the model simulations.

3 Hydrological models

3.1 The FUSE modelling approach

This paper uses the Framework for Understanding Structural Errors (FUSE) hydrological toolkit (Clark et al., 2008) to investigate fundamental model-building decisions, including

(i) the choice of state variables in the unsaturated and saturated zones;

(ii) the choice of flux equations describing surface runoff, interflow, vertical drainage between the soil layers, baseflow and evapotranspiration.

Figure 1 depicts these models diagrammatically, while Table 1 summarizes their key properties. The state and flux equations are listed in Table 2, the model state variables and fluxes are listed in Table 3, and the model parameters are listed in Table 4. Appendix A provides further details.

All FUSE models are formulated in state-space form as ordinary differential equations $\mathrm{d}\mathbf{S}/\mathrm{d}t = \mathbf{g}(\mathbf{S})$, where $\mathbf{S}$ represents storage in the various conceptual compartments of the model, and $\mathbf{g}(\mathbf{S})$ is assembled using the hypothesized connectivity of the stores and formulation of flux expressions (Clark et al., 2008). Formulating hydrological models in state-space form is beneficial in several respects. First, it clearly delineates the model hypotheses from their numerical approximations, which is more satisfying scientifically and

simplifies numerical error control (Clark et al., 2008; Clark and Kavetski, 2010). Second, and more important for this study, it simplifies the design of flexible hydrological software: the FUSE approach facilitates "plug-and-play" of different fluxes and state equations, facilitating the comparison of alternative process representations and paving the way for a more systematic model improvement strategy.

The model simulations described in this paper are performed at an hourly time step, using the fixed-step implicit Euler approximation of the model state equations. Clark and Kavetski (2010) and Kavetski and Clark (2010) provide an extensive discussion of the time stepping schemes currently implemented in FUSE and their impact on model analysis and prediction.

3.2 Model naming convention

The specific models used in this paper are named using letters that refer to each of the modelling decisions (Table 1) and equations (Table 2). For example, the first model trialed uses option A for the unsaturated zone architecture, option A for the saturated zone architecture, option A for evapotranspiration, option B for vertical drainage, option B for interflow, option D for baseflow, and option B for surface runoff. The model is hence named "AAABBDB". Table 5 documents the names and default parameter values for each of the models trialed in this paper.

3.3 Rationale for model selection

The FUSE models used in this study are carefully selected to facilitate testing different model hypotheses.

The first three models, AAABBDB (FUSE-016), CABBBDB (FUSE-014) and CACBBDB (FUSE-170) are selected to evaluate both (i) the state variables used for the unsaturated zone

architecture (option A denotes a single state variable and option C denotes two cascading buckets; Table 1), and (ii) the choice of evaporation parameterization (option A denotes single layer evaporation, option B denotes the sequential parameterization, and option C denotes the root weighting parameterization). All other components of these models are kept constant.

The fourth model, AAAABDB (FUSE-072) is used to evaluate different vertical drainage parameterizations, specifically contrasting the configurations in the first three models that do not allow any vertical drainage below field capacity (option B), with a model where vertical drainage is as a non-linear function of total storage in the unsaturated zone (option A). The approach used to simulate vertical drainage is the only difference between model AAABBDB (FUSE-016) and AAAABDB (FUSE-072). The model AAAABDB also has some additional utility – since vertical drainage does not depend on tension storage, there is more flexibility to modify tension storage and evaluate the impact on evaporation.

The fifth model, ACCBBBB (FUSE-160) is included to evaluate different baseflow parameterizations, here contrasting the single non-linear reservoir used in the first four models with two parallel linear reservoirs. To facilitate the "two parallel linear reservoir" formulation in model ACCBBBB, the unsaturated zone architecture is modified to include two state variables for free storage. The approach used to simulate baseflow is therefore the only difference between models AAABBDB (FUSE-016) and ACCBBBB (FUSE-160).

Some other model components are not evaluated. For example, none of the models allow interflow (option B), and all of the models simulate saturated areas and hence surface runoff as a non-linear function of storage in the unsaturated zone based on the Pareto distribution (option B). The "zero interflow" hypothesis is consistent with the field evidence presented

by Bowden et al. (under review), as described in the companion paper, and expressing saturated areas as a non-linear function of storage in the unsaturated zone is consistent with field evidence that saturated areas are located in upland areas of the catchment and connected to the stream via ephemeral channels (see also the spatial patterns of soil moisture presented by Wilson et al., 2003) –the alternative saturated area function, where saturated areas are controlled by storage in the saturated zone is considered less likely. Use of constant interflow and saturated area parameterizations also makes the analysis of different modeling options more manageable.

3.4 Default model parameters

Many of the diagnostic tests presented in this paper are sensitive to only a subset of model parameters – for example, simulations of soil moisture in the unsaturated zone do not depend on the recession parameters in the saturated zone – and hence the dimensionality of the parameter space is often fairly small. This is one of the major benefits for using diagnostic tests, rather than aggregate assessments of model performance using, for example, the sum of squared errors between simulated and observed streamflow, as diagnostic tests reduce compensatory effects among parameters in different parts of the model (Gupta et al., 2008).

The default values for the model parameters (Table 5) are carefully selected to illustrate behavior in different sub-components of the model. This is done by varying parameters over reasonable ranges (Table 4), and selecting parameters that reproduce the signature indices. Since many of the diagnostic tests are sensitive to only a subset of parameters, conclusions are not expected to be overly sensitive to specific values of individual model parameters.

4 Evaluation of modelling decisions

4.1 Unsaturated zone architecture

To evaluate model simulations of the vertical variation in soil moisture documented in the companion paper, Figure 2 illustrates simulations of fractional tension storage in each of the tension buckets for models AAABBDB (single layer architecture for the unsaturated zone), CABBBDB (cascading buckets architecture and sequential evapotranspiration) and CACBBDB (cascading buckets architecture and root weighting evapotranspiration), using parameter values defined in Table 5. Results show that, for both ET models, the cascading bucket architecture reproduces both the lower variability of soil moisture at greater depth and the delayed wetting at the bottom of the unsaturated zone, as depicted in the companion paper. By contrast, the single layer representation in model AAABBDB is unable to reproduce the observed vertical variations in soil moisture. The impact of the simplifications in the single layer representation will be revisited in Section 4.2 in the context of evaluating evapotranspiration representations – at this stage there is insufficient evidence to reject the single-layer representation.

Figure 3 illustrates the interplay between the selected model structure and appropriate parameter values. It shows simulations of soil moisture for different unsaturated zone architectures using different maximum storage parameters $S_{1,\max}$. As discussed in the companion paper, qualitative analysis suggests that the maximum water content should be approximately 300 mm. Figure 3 illustrates simulations of scaled total soil moisture for models AAABBDB (single state variable in the unsaturated zone), and for the CABBBDB and CACBBDB models used in Figure 2. Results clearly demonstrate that, for all model structures, soil moisture exhibits excessive variability for maximum storage parameters of 100 mm and 200 mm. Also (although less obvious in Figure 3), the maximum storage

parameter value of 500 mm appears to produce insufficient variability, hence model results confirm field data evaluation. There is also a tendency for the sequential evapotranspiration to produce lower soil moisture, and this is examined in the next section.

4.2 Evapotranspiration

To evaluate different evapotranspiration models, Figure 4 (top row) plots the time series of scaled evapotranspiration for the three different schemes tested. For consistency with the neutron probe data presented in the companion paper, evapotranspiration in all of the schemes is restricted to the unsaturated zone (Option A is used for the saturated zone architecture). Figure 4 shows that the sequential evapotranspiration model predicts much higher evapotranspiration in summer months, when soil moisture in the lower layer is low (compare Figure 4 with Figure 2). Indeed, at some times in summer months the soil evaporates at close to the potential rate (Figure 4). This is consistent with the lower soil moisture values in Figure 3. By contrast, the single-layer and root-weighting evapotranspiration models predict lower evapotranspiration in summer months because in these schemes the evapotranspiration is controlled by soil moisture throughout the soil profile. Figure 4 (bottom row) plots the fractional evapotranspiration against fractional tension storage (the predictions of the single-layer model lie on the 1:1 line), and shows that the sequential evapotranspiration parameterization is capable of producing much higher evapotranspiration rates for a given tension storage than the single-layer and root weighting schemes.

It is worth noting that the functional dependence of evapotranspiration on field capacity differs from the approaches used in ecohydrology (Rodriguez-Iturbe et al., 1999; Rodriguez-Iturbe, 2000; Porporato et al., 2002). In ecohydrological models, plants do not generally

become stressed (i.e. transpire at a rate lower than potential evapo-transpiration) until soil moisture drops below a "stress point", which is much lower than field capacity (e.g., Porporato et al., 2002). To represent the lower stress point, we avoid using the "field capacity" parameter in the vertical drainage model component by modeling vertical drainage as a non-linear function of total storage in the unsaturated zone (Option A in Table 2). This means that the parameter that defines the fraction of tension water (ϕ_{tens} in Table 3), or field capacity, can now be used to represent the (lower) stress point for plants. Table 5 defines the parameters used for this new model (AAAABDB). Note that we halved the value of ϕ_{tens} [consistent with the conceptual model of Laio et al. (2001)] and increase the maximum water storage to 400 mm to account for the reduction in tension storage associated with the use of the nonlinear drainage model. Figure 5 compares the normalized evapotranspiration predicted by the new model (AAAABDB) with the single-layer evapotranspiration model used in Figure 4 (model AAABBDB). As expected, the lower stress point causes both a higher evapotranspiration in the wet season and a more rapid depletion of soil moisture in the dry season.

Selecting among these evapotranspiration parameterization schemes requires independent information on evapotranspiration. Assuming negligible change in storage, or flow that by-passes the weir, mean evapotranspiration can be estimated from the long-term water balance as the difference between precipitation and runoff. Figure 6 compares mean evapotranspiration rates for the four schemes, using different maximum unsaturated zone storage parameters ($S_{1,\max}$), against the water-balance estimates of the mean evapotranspiration rate. Recall that all of these schemes restrict evapotranspiration to the unsaturated zone. Results show that mean evapotranspiration is lowest for both the two-layer

root-weighting scheme and the single-layer scheme where evapotranspiration is parameterized based on field capacity. In these cases, evapotranspiration only approaches the water-balance evapotranspiration estimate when the $S_{1,\max}$ parameter is close to 500 mm. By contrast, evapotranspiration in the sequential representation is most similar to the water balance evapotranspiration estimate when $S_{1,\max}$ is approximately 275 mm, which, perhaps coincidentally, is almost identical to the water holding capacity of the soil inferred from both the neutron probe data (McMillan et al., under review) and the parameter perturbation experiments for soil moisture (Figure 3). Also, as expected from Figure 5 and field evidence, Figure 6 shows that mean evapotranspiration is higher when evapotranspiration is modeled as a function of the stress point.

Other information on evapotranspiration is available from the intensive field campaigns. Fong (2001) describes evapotranspiration measurements in the Satellite catchments during three intensive observation periods in spring-summer 1999-2000, where they deployed weighing lysimeters on different aspects. On some summer days, lysimeter-based estimates of evapotranspiration were close to their estimates of potential evapotranspiration, while on many spring days evapotranspiration estimates equaled or exceed the estimated potential evapotranspiration. These observations lend credence to the sequential evapotranspiration hypothesis, which predicted summertime evapotranspiration close to the potential evapotranspiration rate.

4.3 Vertical Drainage

To evaluate alternative vertical drainage model hypotheses, Figure 7 compares the shape of the drainage-storage relationship underlying Options A (gravity drainage) and B (drainage

above field capacity). Figure 6 suggests the two relations are most similar when exponent $c = 5$. The third drainage representation available in FUSE, Option C where drainage is controlled by storage in the saturated zone, is not evaluated because the variability in soil moisture at the hillslope sites in the Satellite Right catchment does not appear to be influenced by the water table (e.g., see the neutron probe depth profiles presented in the companion paper). Similar results were found by Bowden et al. (under review), where the five neutron probe profiles across the hillslope did not show any impact from the water table, but the neutron probe profile in the riparian zone did show evidence of interaction with the water table at depths greater than 1000 mm below the surface. It appears that storage in the saturated zone has a limited impact on the vertical drainage across the catchment as a whole, supporting the a priori exclusion of this hypothesis from the empirical comparison on independent physical grounds.

Figure 8 compares soil moisture simulations for the "gravity drainage" model (Option A) and the "drainage above field capacity" model (Option B), for several parameter sets. The tension storage used to normalize the results in the gravity drainage model is taken as the point where the relative drainage rate is 0.01, so that lowering the exponent c has the effect of reducing tension storage. Model AAAABDB, which uses the gravity drainage parameterization (top plot in Figure 8), predicts excessive variability of soil moisture for low values of c (e.g., $c = 2$) when compared to the point soil moisture observations. This is consistent with the simulations in Figure 3 with lower $S_{1,\max}$ parameters. Variability in soil moisture seems to be more similar to the point observations when $c = 5$ and $c = 10$ (Figure 8), but the scaled soil moisture in these simulations tends to be below field capacity during the wet season. Model CABBBDB (bottom plot in Figure 8), where drainage is restricted to occur above field

capacity, shows reasonable behavior during dry spells, but indicates that the vertical drainage parameter must be at least 100 mm/day to allow sufficient drainage of water. This emphasizes the importance of the vertical drainage flux, and, ultimately, the importance of baseflow from the saturated zone.

The simulations in Figure 8 emphasize the importance of rapid vertical drainage from the unsaturated zone to the saturated zone. This is consistent with the soil moisture time series in the companion paper, which indicate that, at a point, the vertical drainage parameterization must severely restrict drainage below field capacity and allow for rapid drainage above field capacity. The simulations of rapid vertical drainage (and substantial recharge to the saturated zone) are also consistent with the relatively low runoff ratios during individual storms: the companion paper shows that runoff ratios are almost always below 0.5, even in winter when the soils are at field capacity. On its own, the point soil moisture data appears to favor the hypothesis that restricts vertical drainage to occur above field capacity (Option B), provided a sufficiently large vertical drainage rate can be maintained otherwise. However, the importance of relaxing the fixed threshold at field capacity will be re-visited in Section 4.5 in the evaluation of quickflow processes.

4.4 Saturated zone architecture and baseflow parameterizations

To evaluate different hypotheses describing the saturated zone architecture and baseflow, we compare the recession behaviors of a model with a single non-linear storage-discharge relationship (model AAABBDB) and a model with two linear baseflow reservoirs (model ACCBBBB). The two-reservoir model ACCBBBB uses the Sacramento model architecture (Burnash et al., 1973), which includes tension storage and evaporative fluxes in the saturated zone ("Option C" is used for both the Saturated Zone Architecture and evapotranspiration).

To prevent evapotranspiration from the saturated zone in this model, we specify that a very small fraction $\kappa = 0.001$ of the vertical drainage enters the tension storage in the saturated zone and 99.9% of the root weights are in the unsaturated zone ($r_1 = 0.999$).

Figure 9 compares the recession behavior of models AAABBDB and ACCBBBB. As expected from the discussion in the companion paper, model AAABBDB (a single non-linear reservoir) is unable to reproduce the highly nonlinear recessions, where the gradient of $-dQ/dt$ versus Q exceeds 2. Moreover, the single nonlinear reservoir is characterized by a unique storage-discharge relationship, and is unable to reproduce the seasonal differences between individual recessions. By contrast, model ACCBBBB (two linear reservoirs) is able to reproduce the non-unique relationship between total storage and discharge, because different initial storages in the two reservoirs will produce different recessions even if the total storage is the same. However, it does not reproduce the strongly nonlinear behavior at the start and end of the recessions. It is possible that additional linear reservoirs may enable a better fit to the data: for example, Clark et al. (2009) demonstrated that a third linear reservoir was necessary to mimic the highly nonlinear recession behavior at the 41-ha Panola catchment. However, as argued by Harman et al. (2009a), it is more likely that the recession behavior represents a combination of horizontal heterogeneity and hydraulic complexity (i.e., multiple non-linear reservoirs with different transmissivity). The evidence here suggests that neither the single non-linear reservoir model nor the parallel linear reservoir model adequately simulate baseflow behavior in the Satellite Right catchment. We will return to this point later in the paper.

4.5 Quickflow

To evaluate the importance of surface runoff, Figure 10 shows the runoff ratio and the runoff timing for individual storms, for model AAABBDB (a single layer in both the saturated and unsaturated zone, with gravity drainage restricted to times when soil moisture is above field capacity, and a single non-linear reservoir for baseflow). The storm runoff ratios were predicted reasonably accurately when surface runoff exponents of either $b = 0.5$ or $b = 0.9$ were used (top row of Figure 10), emphasizing the importance of saturation-excess runoff in the Satellite Right catchment. However, the timing of runoff in model AAABBDB is, on average, about half a day earlier than in the observations (bottom row of Figure 10), suggesting that the runoff response is too flashy.

Experimentation with different model structures suggests that the flashy response in Figure 10 is caused by inadequate partitioning of precipitation between surface runoff and baseflow. It appears that while the gravity drainage parameterization with a threshold at field capacity provides reasonable simulations of soil moisture at the point scale (e.g., as shown in Figure 8), the threshold is inappropriate at the catchment scale because the simulated vertical drainage is zero when catchment-average storage is less than field capacity. To illustrate this, Figure 11 shows the runoff ratio and the runoff timing for model AAAABDB – the only difference between model AAABBDB (used for the simulations in Figure 10) and AAAABDB (used for the simulations in Figure 11) is that the latter allows drainage below field capacity. Figure 11 suggests that vertical drainage below field capacity improves simulations of the runoff timing for individual storms. Clearly, the runoff timing metric may also be improved by invoking other delay mechanisms (e.g., re-infiltration of saturation-excess runoff). However, the partitioning of water between surface runoff and baseflow clearly affects the dynamics of the runoff response and analysis of this partitioning should be

a key criterion in the evaluation of hydrological models. For example, Vache and McDonnell (2006) used the mean residence time to evaluate the heterogeneity of flow paths.

5 Streamflow simulations

Figure 12 illustrates streamflow simulations for one year period for each of the five models assessed in this paper (refer to Tables 1 and 2 for model descriptions and equations respectively). The streamflow simulations are plotted on a log-scale to emphasize some of the problems with the baseflow parameterizations. Two main features are apparent.

First, all models tend to over-estimate baseflow in wet periods and under-estimate baseflow in dry periods. This is consistent with the recession analysis in Figure 9, where neither the single non-linear reservoir nor the two parallel linear reservoirs were able to adequate represent baseflow dynamics. Note that in the fifth model, ACCBBBB, the "slow" linear reservoir dominates the recession during the majority of the simulation period, as evidenced by the straight line in the bottom plot of Figure 12.

Second, there are clear differences in the wetting-up periods, for example, in May 1999, where there are differences in the amount of baseflow. Models that use a fixed threshold for vertical drainage are characterized by a step increase in baseflow (models AAABBDB, CABBBDB, CACBBDB, and ACCBBBB), and model AAAABDB that does not impose a fixed drainage threshold, has a smoother increase in baseflow. Note also that the size of the step increase in baseflow can be linked to the evapotranspiration parameterization, with the sequential parameterization (model CABBBDB) having a smaller step increase than models AAABBDB and CACBBDB, most likely because of the lower soil moisture (see Figure 3).

There is considerable scope to improve these streamflow simulations. It is of course possible to calibrate the models, and when this is done all models have a Nash-Sutcliffe score of around 0.9 (not shown), which is similar to the Nash-Sutcliffe scores obtained in previous modelling studies in the Mahurangi (2006). However, even with these high Nash-Sutcliffe scores, the models fail to adequately mimic observed processes – as discussed by Schaefli and Gupta (2007) and Gupta et al. (2008), a high objective measure of model performance does not necessarily indicate close correspondence between model simulations and observed processes. Some general suggestions for model improvement are detailed in the next section.

6 The case for increased model complexity?

6.1 Potential areas for model improvement

Results from this paper expose numerous areas for model improvement:

1. It is necessary in these catchments to resolve vertical variations in soil moisture in the unsaturated zone, as simulations in this study demonstrated that the vertical variability in soil moisture has a strong impact on total evapotranspiration. Interestingly, Liang et al. (1996) demonstrated improvements in simulations of both shallow soil moisture and total evapotranspiration by adding a 10 cm layer on top of the original upper layer used to represent the unsaturated zone in the VIC model. The importance of vertical variability in soil moisture will be explored further in subsequent work, where several alternative hypotheses describing transpiration will be investigated.

2. The representation of evapotranspiration needs further scrutiny with respect to the incorporation of the stress point parameter for different vegetation types. This study demonstrated that the use of vegetation stress points produced more realistic estimates

of total evapotranspiration. However, the common approach of using field capacity as the stress point for evapotranspiration is inconsistent with insights from ecological studies (e.g., Porporato et al., 2002) and should be revised.

3. The magnitude and timing of vertical drainage from the unsaturated zone to the saturated zone needs further attention, as the simulations in this study suggested that vertical drainage affects both the total evapotranspiration and recession behavior. Recharge to the saturated zone is a key component of the complex coupling between surface water and shallow groundwater in different environments (Peters et al., 2003; Buttle et al., 2004; McGlynn et al., 2004; Katsuyama et al., 2005; Jencso et al., 2009), and, in many environments, saturated zone recharge must be examined jointly with transpiration losses from shallow groundwater (Goodrich et al., 2000; Tromp-van Meerveld and McDonnell, 2006).

4. The baseflow representation also needs further attention, with this study demonstrating that baseflow models based on either a single non-linear reservoir or two parallel linear reservoirs (typical of widely-used conceptual models) were unable to adequately reproduce observed recession characteristics. As discussed by Harman et al. (2009a), Szilagyi (2009), and Harman et al. (2009b), it is likely that the complexity of recessions results from a mixture of horizontal heterogeneity and hydraulic complexity (i.e., multiple non-linear reservoirs with different transmissivity).

6.2 Adequacy, appropriateness, and ambiguity

In the context of identifying new strategies for model improvement, it is natural to ponder if simple 'bucket-style' conceptual hydrological models have passed their "use-by" date.

Indeed, in most cases, our process-based evaluations of different model structures recommended the most complex structure from the options available in the FUSE framework (e.g., multiple soil layers, multiple baseflow reservoirs), moreover, often suggesting that further complexity is justified. While some of these conclusions may be tainted by data errors, model improvement does appear to require more complex representations of dominant hydrological processes. Adequate hydrological simulations may therefore require a shift from 'bucket-style' conceptual models to process-based models (Ivanov et al., 2004). Justification for this increase in complexity may come from the inclusion of multiple data streams and associated model performance diagnostics.

Nevertheless, even with detailed process-based models, it is unlikely that a single model structure is applicable everywhere (Savenije, 2009). Indeed, the next generation of models will likely have a spatially varying structure to provide an appropriate representation of regional differences in dominant hydrological processes, just as the current generation of hydrological models has spatially varying parameters to account for regional differences in soil properties and vegetation characteristics. Achieving the vision of appropriate model structures requires solving some pressing research problems. First, it is necessary to "map" the dominance of hydrological processes across the landscape (Winter, 2001; McDonnell and Woods, 2004; Wolock et al., 2004; Wagener et al., 2007); and second, it is necessary to identify, *a priori*, the set of model structures that are best suited to different hydrological environments (Beven, 2001; Buttle, 2006). Such research is critical to facilitate the development and application of models with spatially varying structure.

Invariably limitations in both data and process understanding result in considerable ambiguity in the mapping between model structure and hydrological landscapes. Such uncertainty can

be reduced through both thoughtful analysis of experimental observations and innovative modelling approaches, as attempted in this two-part paper. However, ambiguity is unavoidable, and will need to be represented using carefully constructed ensembles of model structures.

7 Summary and Conclusions

This two-part paper examines model representations of hydrological processes in a small headwater sub-catchment (Satellite Right) within the Mahurangi River basin in Northland (New Zealand), with the goal of providing guidance for model selection. In the first paper, precipitation, soil moisture and flow data in the Satellite Right sub-catchment, are used to provide estimates of the dominant hydrological processes and recommendations for representing those processes in conceptual models. In this paper, the focus shifts to an evaluation of several distinct model hypotheses and their ability to represent observed hydrological processes in the Satellite Right sub-catchment.

The data analysis and model sensitivity experiments presented in this study offer several insights for model selection. They demonstrate both the importance of different hydrological processes, for example, the importance of vertical drainage and baseflow, and guidance for the choice of modeling approaches, for example, representing horizontal heterogeneity in soils. The sensitivity experiments also help identify appropriate values for model parameters, such as soil depth and drainage rates, and, notably, focusing attention on signatures that explain behavior in sub-components of the model helps resolve some of the compensatory effects between the choice of model components and the choice of model parameters. This study demonstrates that it is possible to identify the model components that are suitable for simulating the dominant processes in a specific catchment – when such components are

integrated to form a complete model, the resultant model has a structure that is quite different to existing hydrological models.

There are clearly some limitations with the models currently included in the FUSE framework – indeed, the model sensitivity experiments invariably suggest increasing model complexity. However, the quest for adequacy (as obtained through increased complexity) must be tempered with considerations of appropriateness (model structures should vary regionally to represent differences in dominant processes) and considerations of ambiguity (gaps in data and process understanding make it impossible to identify a single model structure). As such, hydrological modeling problems will not be solved by developing an increasingly complex "super-model"; rather it will be necessary to use flexible modeling approaches that facilitate selecting among competing hydrological hypotheses where the data and process understanding allow it, and using an ensemble of hydrological hypotheses to represent ambiguity in model selection.

The challenge for the hydrological community – illustrated in this two-part paper – is to make better use of the available data to not only estimate parameter values, but also to diagnostically identify an appropriate model structure or range of plausible structures. As discussed in this paper, this requires that individual components of the model meaningfully approximate the processes that they are intended to represent, while sufficient data must be available and a suitable parameterization selected to avoid the confounding problem of compensatory parameters both within and between different model components. We anticipate that the synthesis of experimental fieldwork in a variety of hydroclimatically diverse basins with more powerful model and parameter identification techniques, necessarily accounting for the uncertainty in both the available data and process insights and allowing

very stringent checks of all assumptions and hypotheses describing the model and the observations system, will lead to more scientifically defensible model hypotheses and, as a consequence, more operationally reliable predictions of observed hydrological processes.

Appendix A

Bucket overflow fluxes

All fluxes in FUSE are smoothed to ensure continuous and smooth constitutive relations (Kavetski and Kuczera, 2007). In particular, the bucket overflow fluxes q_{utof}, q_{ufof}, q_{stof}, q_{sfof}, q_{sfofa} and q_{sfofb} (see Table 3 for definitions) are computed as

$$q_{(xx)of} = q_{in}\,\Phi\left(S_i^{(XX)} \mid S_{i,\max}^{(XX)}, \omega\right) \tag{3}$$

where $q_{(xx)of}$ is the overflow from storage component *xx* (described by state variable $S_i^{(XX)}$, with upper bound $S_{i,\max}^{(XX)}$), $q_{(xx)in}$ is the infiltration into this storage and $\Phi\left(S_i^{(XX)} \mid S_{i,\max}^{(XX)}, \omega\right)$ is a smooth function controlling the fraction of infiltration that overflows. Note that the infiltration q_{in} into a given store depends on both the model architecture and the specific selection of flux functions (see Table 2).

A logistic smoothing function $\Phi\left(S_i^{(XX)} \mid S_{i,\max}^{(XX)}, \omega\right)$ is used,

$$\Phi\left(S_i^{(XX)} \mid S_{i,\max}^{(XX)}, \omega\right) = \left[1+\exp\left(\frac{S_i^{(XX)} - S_{i,\max}^{(XX)} - \omega\varepsilon}{\omega}\right)\right]^{-1} \tag{4}$$

where ω is a smoothing parameter and $\varepsilon = 5$ ensures storage never exceeds capacity. We set $\omega = \rho S_i^{(XX)}$ and use the parameter ρ to control the strength of smoothing as a function of total storage of each state variable (here, $\rho = 0.01$ in all cases).

Basin routing

The runoff delay due to basin routing is represented in all FUSE models using a Gamma distribution

$$P(\tau \mid a, \mu_\tau) = \frac{\gamma(a, x)}{\Gamma(a)} \quad (5)$$

where τ is the time delay, $\gamma(a, x)$ is the incomplete Gamma function (Press et al., 1992), a is a shape parameter, μ_τ is the mean time delay parameter and $x = \tau a / \mu_\tau$ is a scaled time variable. Equation (5) is used to estimate the fraction of runoff in a future time step $t + \tau$. We fix $a = 3$ and only calibrate μ_τ.

References

Atkinson, S.E., Sivapalan, M., Woods, R.A. and Viney, N.R., 2003. Dominant physical controls on hourly flow predictions and the role of spatial variability: Mahurangi catchment, New Zealand. Advances in Water Resources, 26(3): 219-235.

Beven, K., 2001. On landscape space to model space mapping. Hydrological Processes, 15(2): 323-324.

Beven, K., 2002. Towards a coherent philosophy for modelling the environment. Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences, 458(2026): 2465-2484.

Bowden, W.B., Fahey, B., Basher, L., Stewart, M. and Bidwell, V.J., under review. Subsurface flow dynamics in a pasture hillslope, North Island, New Zealand. XXXXX, xx: xxxx-xxxx.

Buttle, J., 2006. Mapping first-order controls on streamflow from drainage basins: the T-3 template. Hydrological Processes, 20(15): 3415-3422.

Buttle, J.M., Dillon, P.J. and Eerkes, G.R., 2004. Hydrologic coupling of slopes, riparian zones and streams: an example from the Canadian Shield. Journal of Hydrology, 287(1-4): 161-177.

Clark, M.P. and Kavetski, D., 2010. The ancient numerical daemons of conceptual hydrological models: Part 1: Fidelity and efficiency of time stepping schemes. Water Resources Research, submitted.

Clark, M.P. et al., 2009. Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale. Hydrological Processes, 23(2): 311-319.

Clark, M.P. et al., 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44.

Desborough, C.E., 1999. Surface energy balance complexity in GCM land surface models. Climate Dynamics, 15(5): 389-403.

Duan, Q. et al., 2006. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. Journal of Hydrology, 320(1-2): 3-17.

Fong, A., 2001. Evaporation from hillslope pasture at Mahurangi, Northland, University of Waikato, 144 pp.

Goodrich, D.C. et al., 2000. Seasonal estimates of riparian evapotranspiration using remote and in situ measurements. Agricultural and Forest Meteorology, 105(1-3): 281-309.

Grayson, R., Moore, I.D. and McMahon, T.A., 1992. Physically-based hydrologic modeling. 2. Is the concept realistic? Water Resources Research, 26: 2659-.

Gupta, H.V., Wagener, T. and Liu, Y.Q., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes, 22(18): 3802-3813.

Harman, C.J., Sivapalan, M. and Kumar, P., 2009a. Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics. Water Resources Research, 45.

Harman, C.J., Sivapalan, M. and Kumar, P., 2009b. Reply to comment by J. Szilagyi on "Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics". Water Resources Research, 45.

Ivanov, V.Y., Vivoni, E.R., Bras, R.L. and Entekhabi, D., 2004. Catchment hydrologic response with a fully distributed triangulated irregular network model. Water Resources Research, 40(11).

Jencso, K.G. et al., 2009. Hydrologic connectivity between landscapes and streams: Transferring reach-and plot-scale understanding to the catchment scale. Water Resources Research, 45.

Katsuyama, M., Ohte, N. and Kabeya, N., 2005. Effects of bedrock permeability on hillslope and riparian groundwater dynamics in a weathered granite catchment. Water Resources Research, 41(1).

Kavetski, D. and Clark, M.P., 2010. The ancient numerical daemons of conceptual hydrological models: Part 2: Impact of the time stepping scheme on model analysis and prediction. Water Resources Research, in press.

Kavetski, D. and Kuczera, G., 2007. Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration. Water Resources Research, 43(3).

Kavetski, D., Kuczera, G. and Franks, S.W., 2006. Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. Journal of Hydrology, 320(1-2): 173-186.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3).

Klemes, V., 1983. Conceptualisation and scale in hydrology. Journal of Hydrology, 65: 1-23.

Koren, V., Smith, M. and Duan, Q., 2002. Use of a Priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models. In: Q.

Duan, H. Gupta, S. Sorooshian, A.N. Rousseau and R. Turcotte (Editors), Calibration of Watershed Models. American Geophysical Union, Washington D.C.

Kumar, S.V. et al., 2006. Land information system: An interoperable framework for high resolution land surface modeling. Environmental Modelling & Software, 21(10): 1402-1415.

Laio, F., Porporato, A., Ridolfi, L. and Rodriguez-Iturbe, I., 2001. Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress - II. Probabilistic soil moisture dynamics. Advances in Water Resources, 24(7): 707-723.

Leavesley, G., Restrepo, P.J., Markstrom, S.L., Dixon, M. and Stannard, L.G., 1996. The modular modeling system--MMS: User's manual.

Liang, X., Wood, E.F. and Lettenmaier, D.P., 1996. Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. Global and Planetary Change, 13(1-4): 195-206.

Littlewood, I.G., Croke, B.F.W., Jakeman, A.J. and Sivapalan, M., 2003. The role of 'top-down' modelling for Prediction in Ungauged Basins (PUB). Hydrological Processes, 17(8): 1673-1679.

McDonnell, J.J. et al., 2007. Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. Water Resources Research, 43.

McDonnell, J.J. and Woods, R., 2004. On the need for catchment classification. Journal of Hydrology, 299(1-2): 2-3.

McGlynn, B.L., McDonnell, J.J., Seibert, J. and Kendall, C., 2004. Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations. Water Resources Research, 40(7).

McMillan, H., Clark, M., Bowden, W.B., Duncan, M. and Woods, R., under review. Hydrological field data from a modeller's perspective. Part 1. Diagnostic tests for model structure. Hydrological Processes.

Peters, N.E., Freer, J. and Aulenbach, B.T., 2003. Hydrological Dynamics of the Panola Mountain Research Watershed, Georgia. Ground Water, 41(7): 973-988.

Porporato, A., D'Odorico, P., Laio, F., Ridolfi, L. and Rodriguez-Iturbe, I., 2002. Ecohydrology of water-controlled ecosystems. Advances in Water Resources, 25(8-12): 1335-1348.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., 1992. Numerical Recipes in Fortran 77: The Art of Scientific Computing (Second Edition). Cambridge University Press.

Priestley, C.H.B. and Taylor, R.J., 1972. On the assessment of surface evaporation using largescale parameters. Monthly Weather Review, 100(81-92).

Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resources Research, in press.

Rodriguez-Iturbe, I., 2000. Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamics. Water Resources Research, 36(1): 3-9.

Rodriguez-Iturbe, I., D'Odorico, P., Porporato, A. and Ridolfi, L., 1999. On the spatial and temporal links between vegetation, climate, and soil moisture. Water Resources Research, 35(12): 3709-3722.

Schaake, J.C. et al., 2004. An intercomparison of soil moisture fields in the North American land data assimilation system (NLDAS). Journal of Geophysical Research-Atmospheres, 109(D1).

Schaefli, B. and Gupta, H.V., 2007. Do Nash values have value? Hydrological Processes, 21(15): 2075-2080.

Sivapalan, M., Bloschl, G., Zhang, L. and Vertessy, R., 2003. Downward approach to hydrological prediction. Hydrological Processes, 17(11): 2101-2111.

Szilagyi, J., 2009. Comment on "Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics" by C. J. Harman, M. Sivapalan, and P. Kumar. Water Resources Research, 45.

Tait, A. and Woods, R., 2007. Spatial interpolation of daily potential evapotranspiration for New Zealand using a spline model. Journal of Hydrometeorology, 8(3): 430-438.

Tromp-van Meerveld, H.J. and McDonnell, J.J., 2006. On the interrelations between topography, soil depth, soil moisture, transpiration rates and species distribution at the hillslope scale. Advances in Water Resources, 29(2): 293-310.

Vache, K.B. and McDonnell, J.J., 2006. A process-based rejectionist framework for evaluating catchment runoff model structure. Water Resources Research, 42(2).

Wagener, T., Sivapalan, M., Troch, P. and Woods, R., 2007. Catchment classification and hydrologic similarity. Geography Compass, 1: 901-931.

Western, A.W. et al., 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. Journal of Hydrology, 286(1-4): 113-134.

Wilson, D.J. et al., 2003. Spatial distribution of soil moisture over 6 and 30 cm depth, Mahurangi river catchment, New Zealand. Journal of Hydrology, 276(1-4): 254-274.

Winter, T., 2001. The concept of hydrologic landscapes. Journal of the American Water Resources Association, 37: 335-349.

Wolock, D.M., Winter, T.C. and McMahon, G., 2004. Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. Environmental Management, 34: S71-S88.

Woods, R.A. et al., 2001. Experimental Design and Initial Results from the Mahurangi River Variability Experiment: MARVEX. In: V. Lakshmi, J.D. Albertson and J. Schaake (Editors), Observations and Modelling of Land Surface Hydrological Processes. American Geophysical Union, Washington, DC, pp. 201-213.

Table 1. Summary of FUSE modelling decisions.

Model Decisions	Parameterization Name	Reference
Upper layer architecture		
A. Single state	Single State	Wood et al., 1992 Beven, 1997
B. Separate tension and free storage	Separate Tension Storage	Burnash et al., 1973
C. Separate tension and free storage, with tension storage further disaggregated into cascading buckets	Cascading Buckets	Leavesley et al., 1983
Lower layer architecture		
A. Single state, unlimited storage, no lower layer evapotranspiration	Single state – without evapotranspiration	Beven, 1997 Leavesley et al., 1983[#]
B. Single state, fixed storage, allow lower layer evapotranspiration	Single state – with evapotranspiration	Liang et al., 1994
C. Tension storage plus two parallel baseflow reservoirs	Parallel baseflow reservoirs – with evapotranspiration	Burnash et al., 1973
Evapotranspiration		
A. Evapotranspiration restricted to the upper layer, and is a linear function of storage between wilting point and field capacity.	Single layer evapotranspiration	Beven, 1997
B. Evapotranspiration in the upper and lower layers, where evapotranspiration in the lower layer is restricted by the potential evapotranspiration satisfied in the upper layer	Sequential	Burnash et al., 1973 Leavesley et al., 1983[#]
C. Evapotranspiration in the upper and lower layers, where evapotranspiration in each soil layer depends on relative root fraction in the upper and lower soil layers.	Root weighting	Liang et al., 1994
Drainage		
A. Nonlinear function of total storage in upper layer	Gravity drainage	Liang et al., 1994
B. Nonlinear function of free storage in upper layer	Drainage above field capacity	Beven, 1997 Leavesley et al., 1983
C. Linear function of upper layer free storage and non-linear function of total lower layer storage	Saturated zone control	Burnash et al., 1973
Interflow		
A. Linear function of free storage in the upper layer	Interflow allowed	Burnash et al., 1973 Leavesley et al., 1983
B. No interflow	Interflow denied	Liang et al., 1994 Beven et al., 1997
Baseflow		
A. Single linear reservoir	Single linear reservoir	Leavesley et al., 1983
B. Two parallel linear reservoirs	Two parallel linear reservoirs	Burnash et al., 1973
C. Single non-linear reservoir	Single non-linear reservoir	Liang et al., 1994
D. Single non-linear reservoir, transmissivity determined based on topographic index	Single non-linear reservoir, topographic index	Beven, 1997
Surface Runoff		
A. Saturated area is a linear function of tension storage in the unsaturated zone	Unsaturated zone linear	Leavesley et al., 1983
B. Saturated area is related to storage in the unsaturated zone via a Pareto distribution	Unsaturated zone Pareto	Wood et al., 1992
C. Saturated area is related to storage in the saturated zone via the topographic index	Saturated zone topographic	Beven, 1997

Leavesley et al. (1983) apply the two-layer sequential evapotranspiration parameterization to the two cascading buckets in the unsaturated zone.

Table 2. State and flux equations currently available in FUSE.

Decision	Option A*	Option B*	Option C*	Option D
Unsaturated Zone Architecture	$\frac{\mathrm{d}S_1}{\mathrm{d}t} = (p - q_{sx}) - e_1 - q_{12} - q_{if} - q_{ufof}$	$\frac{\mathrm{d}S_1^{Tens}}{\mathrm{d}t} = (p - q_{sx}) - e_1 - q_{utof}$ $\frac{\mathrm{d}S_1^{Free}}{\mathrm{d}t} = q_{utof} - q_{12} - q_{if} - q_{ufof}$	$\frac{\mathrm{d}S_1^{TensA}}{\mathrm{d}t} = (p - q_{sx}) - e_1^A - q_{urof}$ $\frac{\mathrm{d}S_1^{TensB}}{\mathrm{d}t} = q_{urof} - e_1^B - q_{utof}$ $\frac{\mathrm{d}S_1^{Free}}{\mathrm{d}t} = q_{utof} - q_{12} - q_{if} - q_{ufof}$	*not available*
Saturated Zone Architecture	$\frac{\mathrm{d}S_2}{\mathrm{d}t} = q_{12} - q_b$	$\frac{\mathrm{d}S_2}{\mathrm{d}t} = q_{12} - e_2 - q_b - q_{sfof}$	$\frac{\mathrm{d}S_2^{Tens}}{\mathrm{d}t} = \kappa q_{12} - e_2 - q_{stof}$ $\frac{\mathrm{d}S_2^{FreeA}}{\mathrm{d}t} = \frac{(1-\kappa)q_{12}}{2} + \frac{q_{stof}}{2} - q_b^A - q_{sfofa}$ $\frac{\mathrm{d}S_2^{FreeB}}{\mathrm{d}t} = \frac{(1-\kappa)q_{12}}{2} + \frac{q_{stof}}{2} - q_b^B - q_{sfofb}$	*not available*
Evapotranspiration	$e_1 = pet\ \min\left(\frac{S_1}{\phi_{tens} S_{1,\max}}, 1\right)$	$e_1 = pet\ \min\left(\frac{S_1}{\phi_{tens} S_{1,\max}}, 1\right)$ $e_2 = (pet - e_1)\ \min\left(\frac{S_2}{\phi_{tens} S_{2,\max}}, 1\right)$	$e_1 = pet\ r_1 \min\left(\frac{S_1}{\phi_{tens} S_{1,\max}}, 1\right)$ $e_2 = pet\ r_2 \min\left(\frac{S_1}{\phi_{tens} S_{1,\max}}, 1\right)$	*not available*
Drainage	$q_{12} = k_u \left(\frac{S_1}{S_{1,\max}}\right)^c$	$q_{12} = k_u \left(\frac{S_1^{Free}}{S_{1,\max}^{Free}}\right)^c$	$q_{12} = q_0\, d_{lz} \left(\frac{S_1^{Free}}{S_{1,\max}^{Free}}\right)$ $d_{lz} = 1 + \alpha \left(1 - \frac{S_2}{S_{2,\max}}\right)^{\psi}$	*not available*
Interflow	$q_{if} = k_i \left(\frac{S_1^{Free}}{S_{1,\max}^{Free}}\right)$	$q_{if} = 0$	*not available*	*not available*
Baseflow	$q_b = \nu S_2$	$q_b = \nu_A S_2^{FreeA} + \nu_B S_2^{FreeB}$	$q_b = k_s \left(\frac{S_2}{S_{2,\max}}\right)^n$	$q_b = \frac{k_s m}{\lambda_n^n} \left(\frac{S_2}{m\, n}\right)^n$
Surface Runoff	$q_{sx} = p\ \min\left(\frac{S_1^{Tens}}{\phi_{tens} S_{1,\max}}, 1\right) A_{c,\max}$	$q_{sx} = p \left[1 - \left(1 - \frac{S_1}{S_{1,\max}}\right)^b\right]$	$q_{sx} = p \int_{\zeta_{crit}}^{\infty} f(\zeta)\, d\zeta$ $\zeta_{crit} = \lambda_n \left(\frac{S_2}{S_{2,\max}}\right)^{-1}$	*not available*

* The superscripts *Tens* and *Free* denote tension storage and free storage respectively.

** The subsurface depth scaling parameter for Option D of the baseflow parameterization is $m = S_{2,\max} / n$.

*** The variable ζ for Option C of the surface runoff parameterization describes the spatial distribution of the topographic index (Beven and Kirkby, 1979; Beven, 1997).

Table 3. FUSE state variables and fluxes

Variable	Description	Units
S_1	Total water content in the upper soil layer	mm
S_1^{Tens}	Tension water content in the upper soil layer	mm
S_1^{TensA}	Primary tension water content in the upper soil layer	mm
S_1^{TensB}	Secondary tension water content in the upper soil layer	mm
S_1^{Free}	Free water content in the upper soil layer	mm
S_2	Total water content in the lower soil layer	mm
S_2^{Tens}	Tension water content in the lower soil layer	mm
S_2^{FreeA}	Free water content in the primary baseflow reservoir	mm
S_2^{FreeB}	Free water content in the secondary baseflow reservoir	mm
p	Precipitation	mm day^{-1}
pet	Potential evapotranspiration	mm day^{-1}
e_1	Evapotranspiration from the upper soil layer	mm day^{-1}
e_2	Evapotranspiration from the lower soil layer	mm day^{-1}
e_1^A	Evapotranspiration from the primary tension store	mm day^{-1}
e_1^B	Evapotranspiration from the secondary tension store	mm day^{-1}
q_{sx}	Surface runoff	mm day^{-1}
q_{12}	Drainage of water from the upper to the lower layer	mm day^{-1}
q_{if}	Interflow	mm day^{-1}
q_b	Baseflow	mm day^{-1}
q_b^A	Baseflow from the primary reservoir	mm day^{-1}
q_b^B	Baseflow from the secondary reservoir	mm day^{-1}
q_{urof}	Overflow of water from the primary tension store in the upper soil layer	mm day^{-1}
q_{utof}	Overflow of water from tension storage in the upper soil layer	mm day^{-1}
q_{ufof}	Overflow of water from free storage in the upper soil layer	mm day^{-1}
q_{stof}	Overflow of water from tension storage in the lower soil layer	mm day^{-1}
q_{sfof}	Overflow of water from free storage in the lower soil layer	mm day^{-1}
q_{sfofa}	Overflow of water from primary baseflow storage in the lower soil layer	mm day^{-1}
q_{sfofb}	Overflow of water from secondary baseflow storage in the lower soil layer	mm day^{-1}

Table 4. FUSE model parameters

Parameter	Description	Units	Lower Limit	Upper Limit
$S_{1,\max}$	Maximum storage in the unsaturated zone	mm	25.000	500.000
$S_{2,\max}$	Maximum storage in the saturated zone	mm	50.000	5000.000
ϕ_{tens}	Fraction total storage as tension storage	-	0.050	0.950
ϕ_{rchr}	Fraction of tension storage in primary zone (unsaturated zone)	-	0.050	0.950
ϕ_{base}	Fraction of free storage in primary reservoir (saturated zone)	-	0.050	0.950
r_1	Fraction of roots in the upper soil layer	-	0.050	0.950
k_u	Vertical drainage rate	mm day^{-1}	0.010	1000.000
c	Vertical drainage exponent	-	1.000	20.000
α	Vertical drainage multiplier for the lower layer	-	1.000	250.000
ψ	Vertical drainage exponent for the lower layer	-	1.000	5.000
κ	Fraction of drainage to tension storage in the lower layer	-	0.050	0.950
k_i	Interflow rate	mm day^{-1}	0.010	1000.000
k_s	Baseflow rate	mm day^{-1}	0.001	1000.000
n	Baseflow exponent	-	1.000	10.000
v	Baseflow depletion rate for single reservoir	day^{-1}	0.001	0.250
v_A	Baseflow depletion rate for primary reservoir	day^{-1}	0.001	0.250
v_B	Baseflow depletion rate for secondary reservoir	day^{-1}	0.001	0.250
$A_{c,\max}$	Maximum saturated area (fraction)	-	0.050	0.950
b	ARNO/VIC "b" exponent	-	0.001	3.000
λ	Mean of the log-transformed topographic index distribution[#]	m	5.000	10.000
χ	Shape parameter defining the topographic index distribution[#]	-	2.000	5.000
μ_τ	Time delay in runoff	day	0.010	5.000

[#] The probability distribution of the TOPMODEL topographic index (Beven and Kirkby, 1979; Beven, 1997) can be defined using a parametric probability distribution, which requires at least two additional model parameters. The mean and shape parameters used here are for the Gamma distribution (Clark et al., 2008).

Table 5. Default parameters used for different models

Parameter	AAABBDB (FUSE-016)	CABBBDB (FUSE-014)	CACBBDB (FUSE-170)	AAAABDB (FUSE-072)	ACCBBBB (FUSE-160)
$S_{1,\max}$	300.000	300.000	300.000	400.000	300.000
$S_{2,\max}$	100.000	100.000	100.000	100.000	1000.000
ϕ_{tens}	0.500	0.500	0.500	0.200	0.500
ϕ_{rchr}	N/A	0.250	0.250	N/A	N/A
ϕ_{base}	N/A	N/A	N/A	N/A	0.250
r_1	N/A	N/A	0.500	N/A	0.999
k_u	750.000	750.000	750.000	750.000	750.000
c	1.000	1.000	1.000	1.000	1.000
α	N/A	N/A	N/A	N/A	N/A
ψ	N/A	N/A	N/A	N/A	N/A
κ	N/A	N/A	N/A	N/A	0.001
k_i	N/A	N/A	N/A	N/A	N/A
k_s	1000.000	1000.000	1000.000	1000.000	N/A
n	10.000	10.000	10.000	10.000	N/A
v	N/A	N/A	N/A	N/A	N/A
v_A	N/A	N/A	N/A	N/A	1.000
v_B	N/A	N/A	N/A	N/A	0.025
$A_{c,\max}$	N/A	N/A	N/A	N/A	N/A
b	0.500	0.500	0.500	0.500	0.500
λ	7.500	7.500	7.500	7.500	7.500
χ	3.000	3.000	3.000	3.000	3.000
μ_τ	0.100	0.100	0.100	0.100	0.100

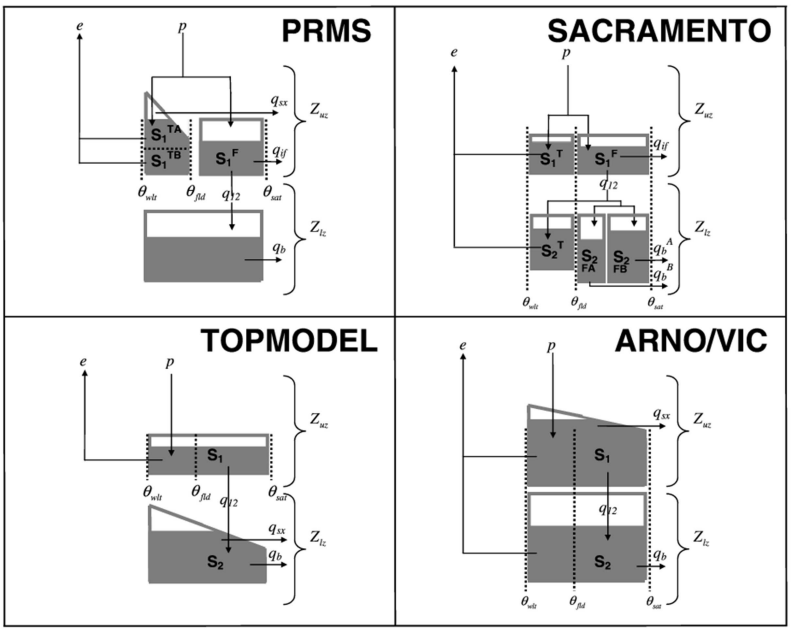

Figure 1. Simplified wiring diagrams for each of the parent models used in FUSE (the state variables and fluxes are defined in Table 3). Here Z_{uz} and Z_{lz} denote the depth of the upper (unsaturated) and lower (saturated) zones, and θ_{wlt}, θ_{fld}, and θ_{sat} denote the soil moisture at wilting point, field capacity, and saturation. Saturation excess runoff (q_{sx}) is defined as the fraction of precipitation that falls on saturated areas of the basin and does not infiltrate into the soil; q_{sx} is shown as originating from the lower zone storage in TOPMODEL because lower zone storage in TOPMODEL controls the saturated area.
274x190mm (284 x 284 DPI)

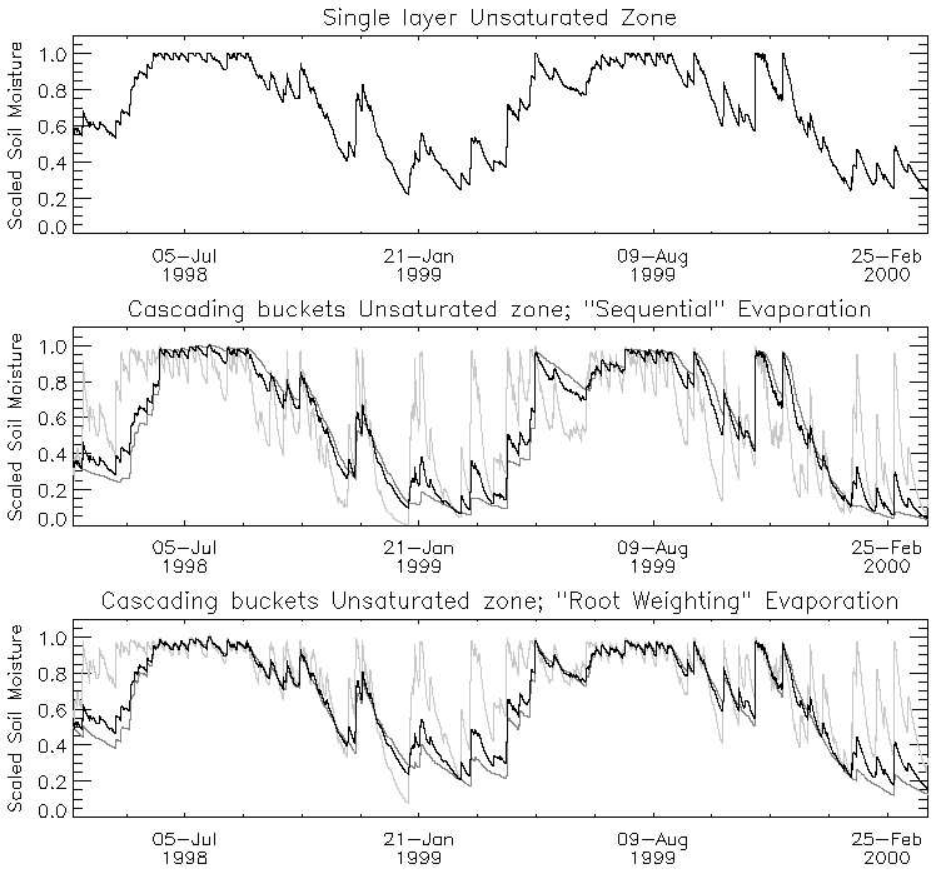

Figure 2. Time series of scaled tension storage in the "single layer" (top row) and two "cascading bucket" models, using the "sequential" (middle row) and "root weighting" (bottom row) evapotranspiration representations. The black line denotes total scaled tension storage, and the pale and mid-tone lines depict tension storage in the upper and lower buckets respectively.
203x177mm (100 x 100 DPI)

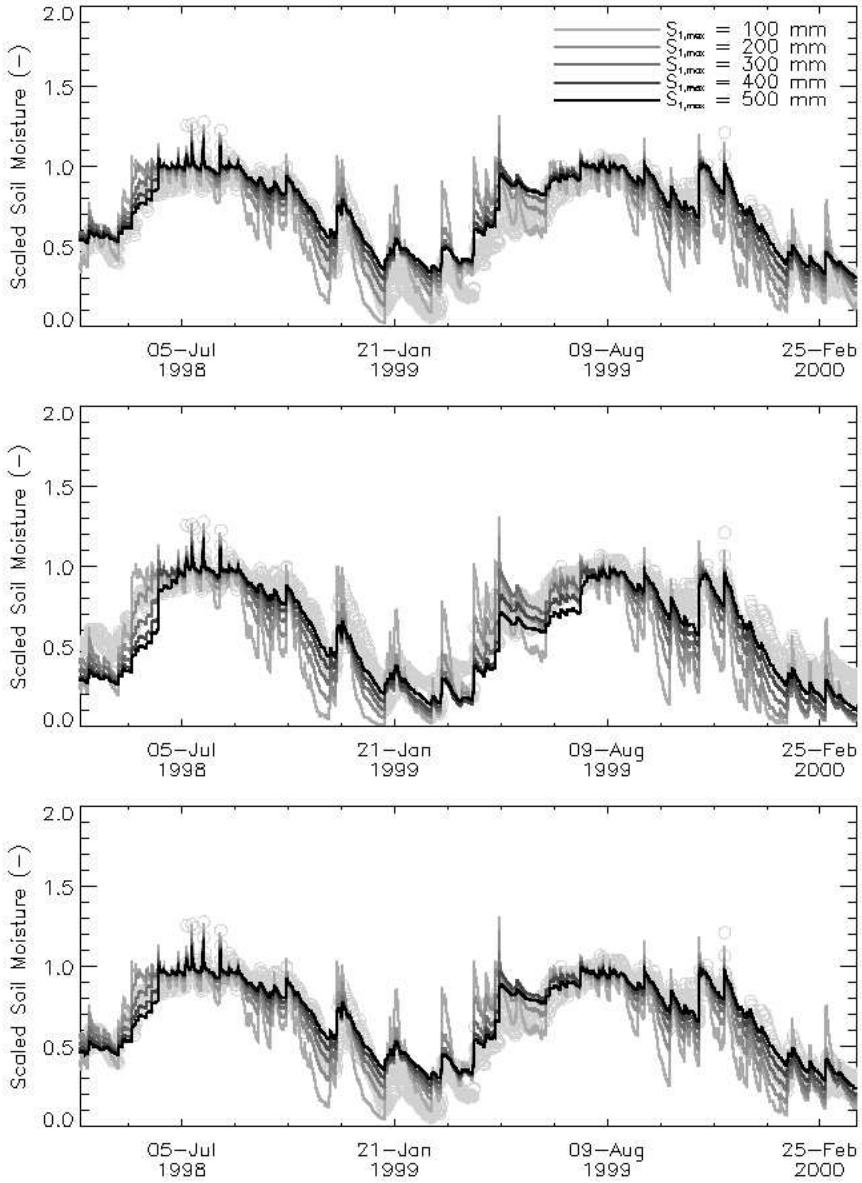

Figure 3. Soil moisture predicted under three different hypotheses describing the unsaturated zone architecture and evapotranspiration schemes, showing: a single store for the unsaturated zone (top) ; the "cascading buckets" representation of the unsaturated zone using (i) the "sequential" evapotranspiration scheme (middle) versus (ii) the "root weighting" evapotranspiration scheme. The pale circles denote the observed soil moisture from all three sites; the lines are for different maximum storage, ranging from 100 mm (lightest) to 500 mm (darkest).
177x228mm (100 x 100 DPI)

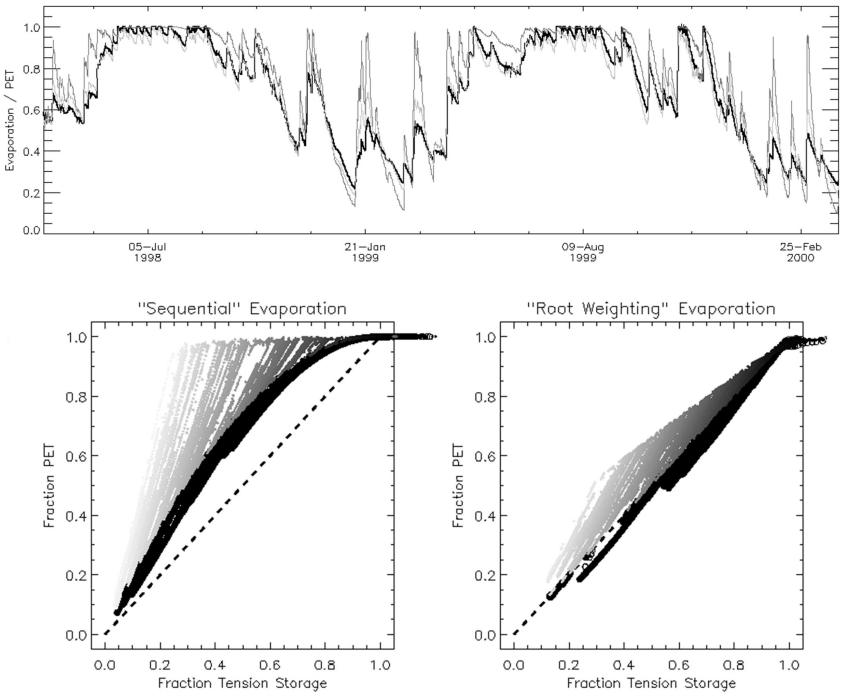

Figure 4. Internal behavior of the evapotranspiration models. The top plot illustrates the ratio of actual evapotranspiration to potential evapotranspiration, for model AAABBDB (single-layer evapotranspiration; black line), model CABBBDB (cascading buckets unsaturated zone architecture using sequential evapotranspiration; mid-tone line), and model CACBBDB (the cascading buckets unsaturated zone architecture using the root weighting evapotranspiration; pale line). The bottom plots illustrate the relationship between scaled total tension storage and scaled evapotranspiration, for the "sequential" (left plot) and "root weighting" (right plot) evapotranspiration models, with symbols plotted in grey when the scaled storage in the lower tank is less than the scaled storage in the upper tank (the tone denotes the fraction of storage in the lower tank, where pale grey denotes relatively low storage in the lower tank and dark grey denotes relatively high storage in the lower tank).
356x267mm (219 x 219 DPI)

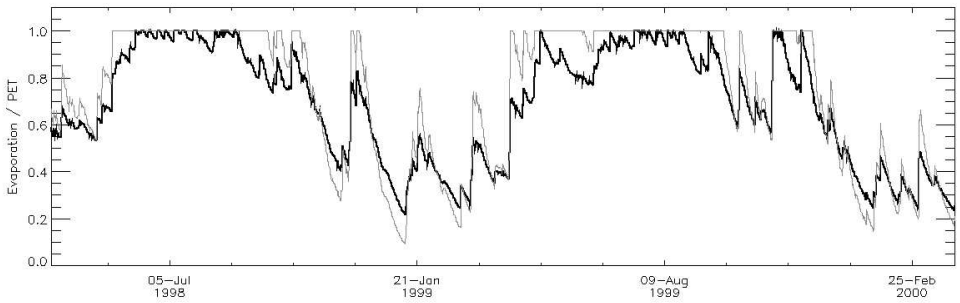

Figure 5. The ratio of actual evapotranspiration to potential evapotranspiration for model AAABBDB (black line, where transpiration is controlled by field capacity) and model AAAABDB (grey line, where transpiration is controlled by plant "stress point").
304x101mm (100 x 100 DPI)

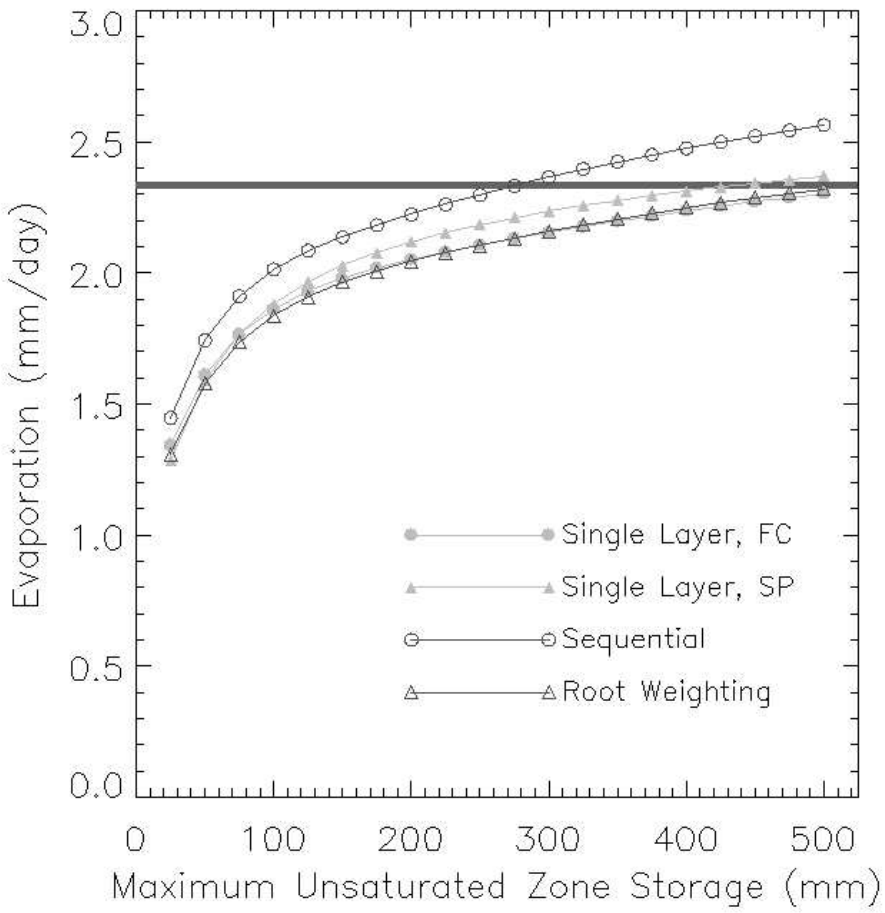

Figure 6. Comparison of mean evapotranspiration predicted under different model hypotheses with a water-balance-based estimate (horizontal line), calculated as the difference between precipitation and runoff averaged over the simulation period.
177x177mm (100 x 100 DPI)

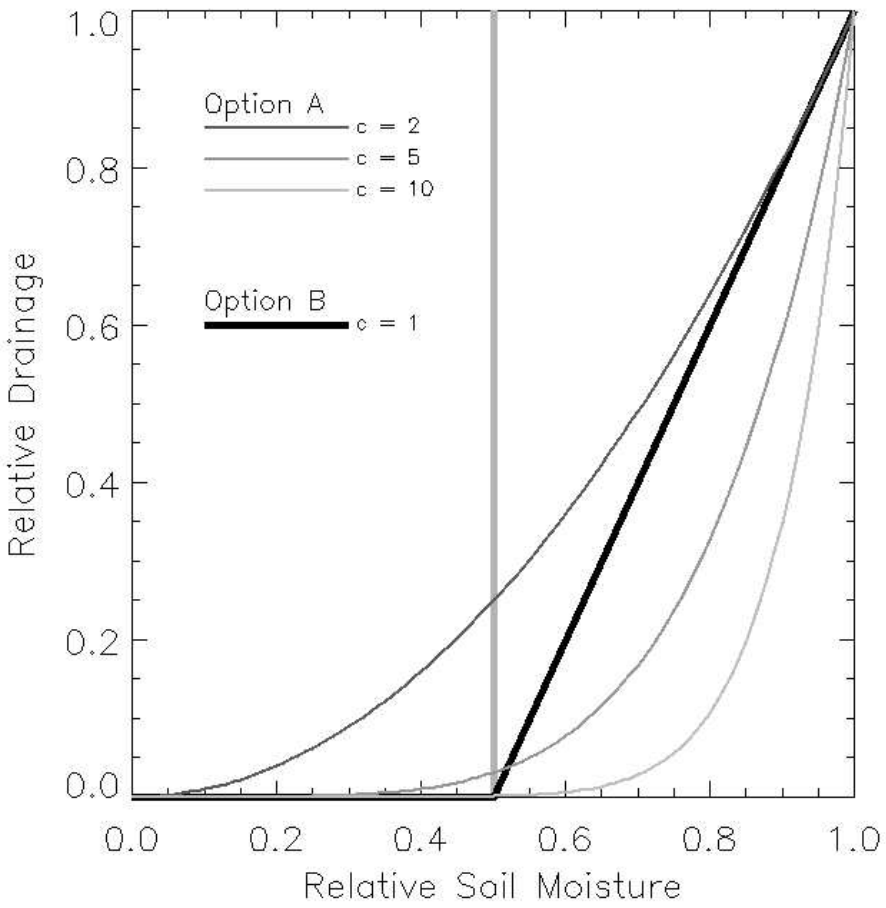

Figure 7. Comparison of the vertical drainage-storage relationship arising in Option A (gravity drainage) and Option B (drainage above field capacity). See Table 2 for the different drainage formulations.
177x177mm (100 x 100 DPI)

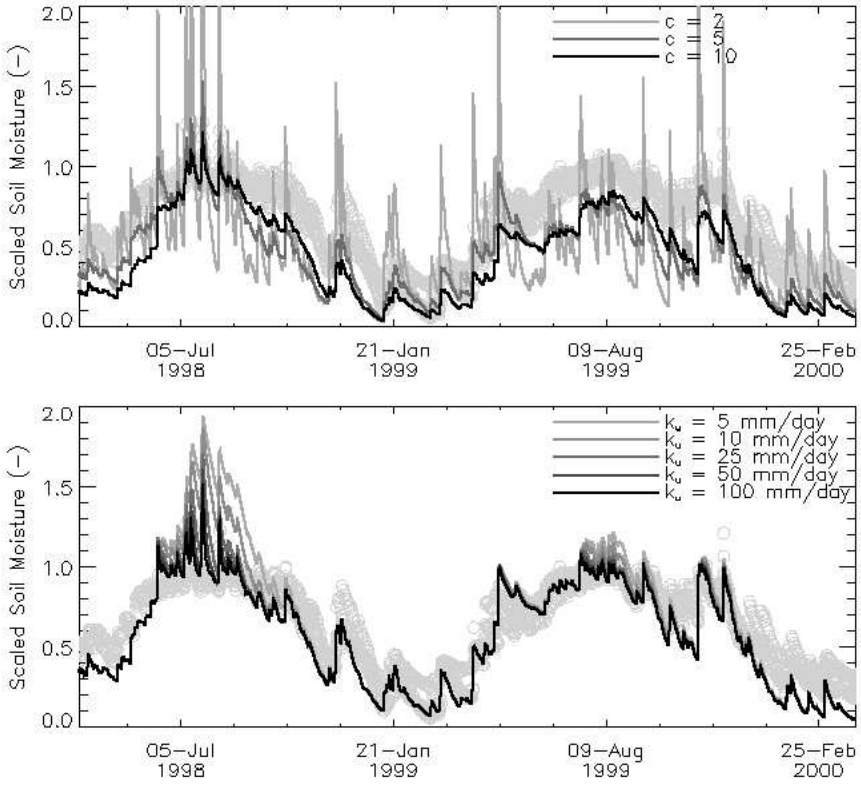

Figure 8. Behavior of different vertical drainage formulations for model AAAABDB that uses the gravity drainage parameterization (top row) and model CABBBDB that uses the drainage above field capacity parameterization (bottom row). Here c is the exponent in the non-linear drainage function and k_u is the parameter that defines the maximum vertical drainage rate (Table 4).
177x228mm (100 x 100 DPI)

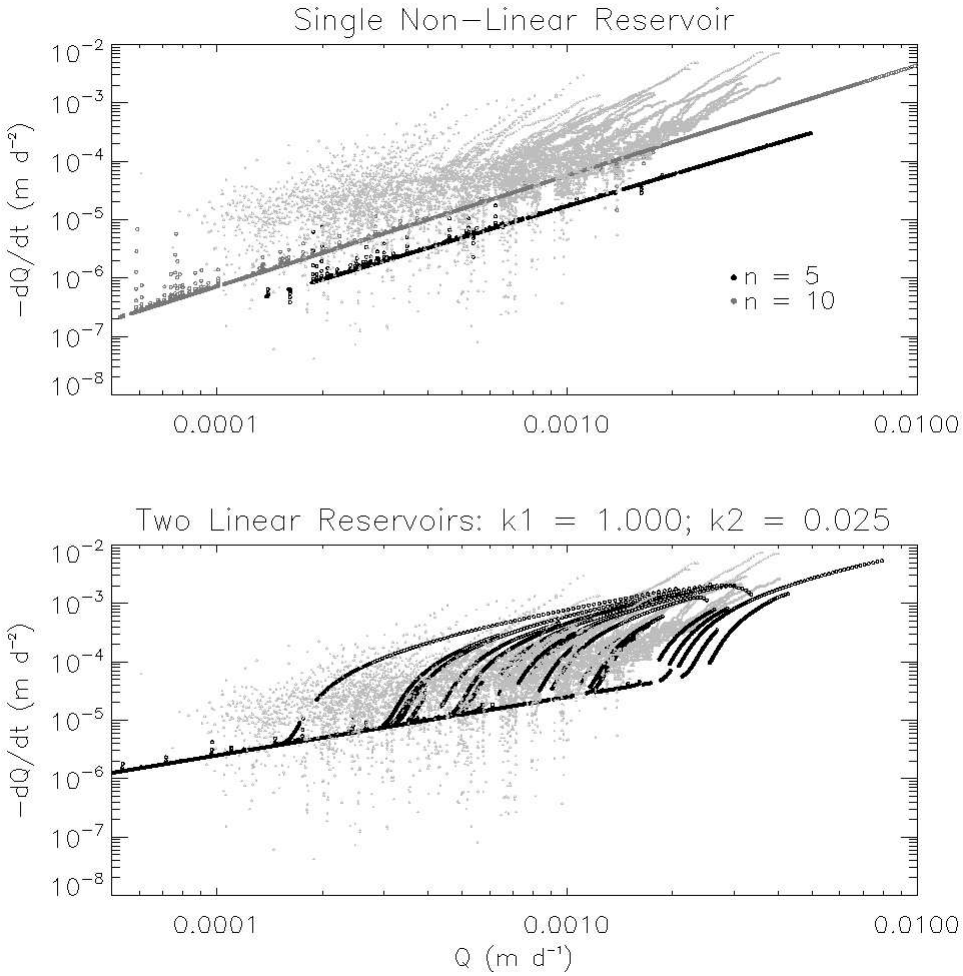

Figure 9. Recession behavior for two FUSE models, showing model AAABBDB with a single non-linear reservoir (top row), and model ACCBBBB with two linear reservoirs (bottom row); both overlaid on the measured recessions behaviour. The top plot illustrates simulations using baseflow exponents n = 5 (black dots) and n = 10 (grey dots).
254x254mm (100 x 100 DPI)

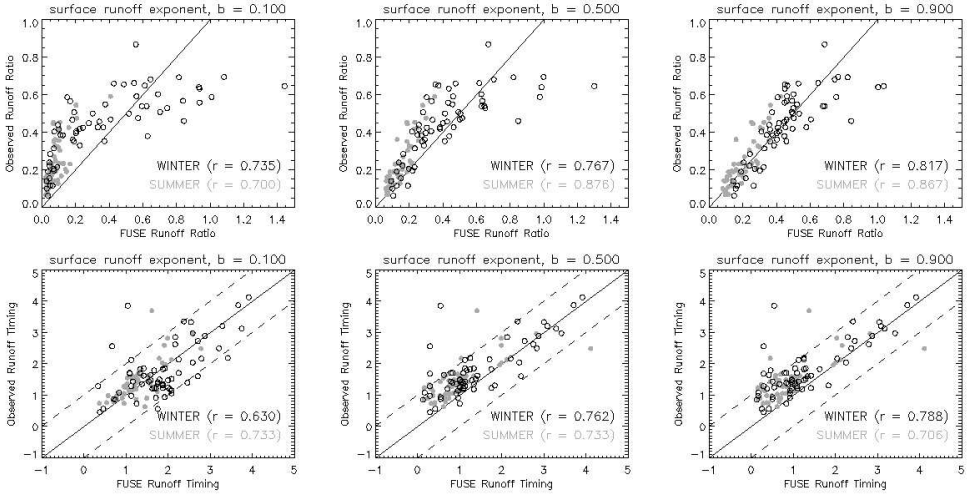

Figure 10. Relationship between the simulated and observed runoff ratio (top row) and runoff timing (bottom row) for individual storms, for model AAABBDB with surface runoff exponents b = 0.1 (left column), b = 0.5 (middle column) and b = 0.9 (right column). The dashed lines in the bottom row of figures depict errors in runoff timing of one day -- errors of timing of one day are included just for reference to facilitate comparing different model hypotheses.
304x152mm (100 x 100 DPI)

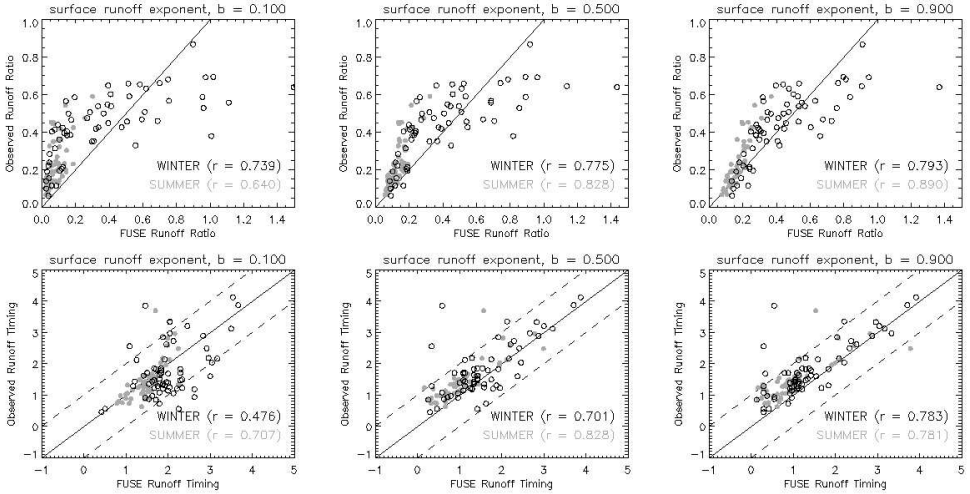

Figure 11. Relationship between the simulated and observed runoff ratio (top row) and runoff timing (bottom row) for individual storms, for model AAAABDB with surface runoff exponents b = 0.1 (left column), b = 0.5 (middle column) and b = 0.9 (right column). The dashed lines in the bottom row of figures depict errors in runoff timing of one day.
304x152mm (100 x 100 DPI)

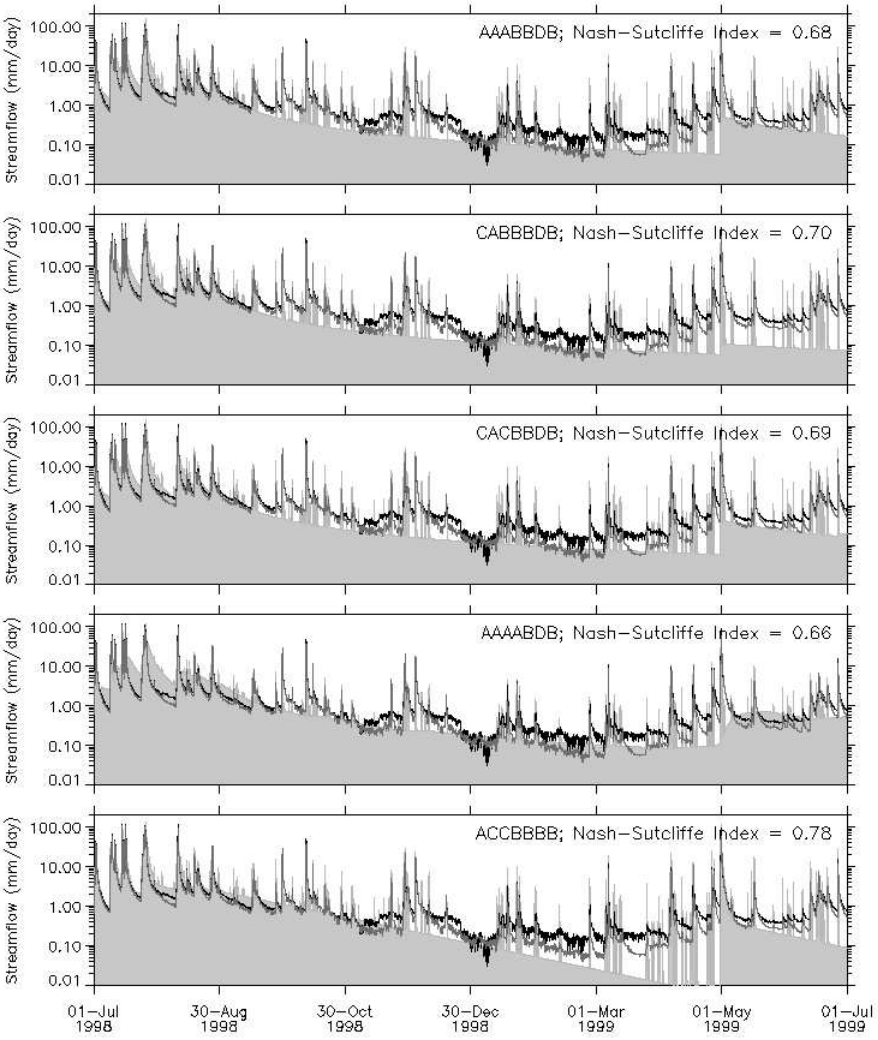

Figure 12. Comparison of streamflow predicted by each of the five models (pale grey) with measurements at the Satellite Right and Satellite Left weirs (black and dark grey respectively). 203x254mm (100 x 100 DPI)