HYDROLOGICAL PROCESSES
Hydrol. Process. **24**, 3631–3647 (2010)
Published online 19 July 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.7788

Estimating the deep seepage component of the hillslope and catchment water balance within a measurement uncertainty framework

Chris B. Graham,[1]* Willem van Verseveld,[2] Holly R. Barnard[3] and Jeffrey J. McDonnell[4]

[1] *Crop and Soil Science, Penn State University, University Park, PA, USA*
[2] *Deltares-Delft Hydraulics—Operational Water Management, Delft, The Netherlands*
[3] *Institute of Arctic and Alpine Research—Geography, University of Colorado, Boulder, CO, USA*
[4] *Forest Engineering, Resources and Management, Oregon State University, Corvallis, OR, USA*

Abstract:

Deep seepage is a term in the hillslope and catchment water balance that is rarely measured and usually relegated to a residual in the water balance equation. While recent studies have begun to quantify this important component, we still lack understanding of how deep seepage varies from hillslope to catchment scales and how much uncertainty surrounds its quantification within the overall water balance. Here, we report on a hillslope water balance study from the H. J. Andrews Experimental Forest in Oregon aimed at quantifying the deep seepage component where we irrigated a 172-m^2 section of hillslope for 24·4 days at $3{\cdot}6 \pm 3$ mm/h. The objective of this experiment was to close the water balance, identifying the relative partitioning of, and uncertainties around deep seepage and the other measured water balance components of evaporation, transpiration, lateral subsurface flow, bedrock return flow and fluxes into and out of soil profile storage. We then used this information to determine how the quantification of individual water balance components improves our understanding of key hillslope processes and how uncertainties in individual measurements propagate through the functional uses of the measurements into water balance components (i.e. meteorological measurements propagated through potential evapotranspiration estimates). Our results show that hillslope scale deep seepage composed of $27 \pm 17\%$ of applied water. During and immediately after the irrigation experiment, a significant amount of the irrigation water could not be accounted for. This amount decreased as the measurement time increased, declining from $28 \pm 16\%$ at the end of the irrigation to $20 \pm 21\%$ after 10 days drainage. This water is attributed to deep seepage at the catchment scale. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS hillslope hydrology; water balance; experimental uncertainty; deep seepage; hysteresis; sprinkling experiments

Received 15 September 2009; Accepted 10 May 2010

INTRODUCTION

The role of deep seepage in the hillslope and catchment water balance in humid regions is poorly understood. The 2001 Joint USA–Japan Workshop on Hydrology and Biogeochemistry of Forested Watersheds noted that examination of the infiltration process of subsurface water into bedrock and its effects on hillslope response to rainstorms was one of the most pressing needs in catchment science (McDonnell and Tanaka, 2001). Since then many researchers (Onda *et al*., 2001; Uchida *et al*., 2003; Katsuyama *et al*., 2005) have begun to demonstrate the instances of deep seepage where infiltration into, and fast lateral flow within, the underlying weathered bedrock on steep colluvial mantled slopes contributes significant amounts of bedrock water to streamflow. In an early study, Terajima *et al*. (1993) found that deep seepage was at least 30 and 18% of precipitation in the Obara and Akatsu catchments (Japan), respectively, and that the percentage decreased with increasing catchment size. Other early, but important work by Anderson *et al*. (1997) injected bromide into saturated hillslope colluvium and observed rapid infiltration and flow through the underlying sandstone bedrock to the catchment outlet. More recently, Tromp van Meerveld *et al*. (2006) performed a series of hillslope scale sprinkling experiments and estimated that over 90% of their applied water was 'lost' to deep seepage. Clearly, deep seepage is a major component of the water balance of steep, humid catchments at both the hillslope and catchment scales.

Deep seepage, the transfer of infiltrated precipitation from the surface water zone (soil mantle and hyporheic zone) to bedrock flowpaths occurs at multiple scales, from the plot scale to the subcontinent. We define deep seepage at the hillslope scale as water that moves through the soil profile, and infiltrates into the bedrock, where it can either move laterally towards the stream channel or enter the groundwater. We define deep seepage at the catchment scale as water that does not re-emerge into the stream channel as bedrock return flow, but leaves the catchment as deep groundwater flow. Although current measurement methodologies are unable to directly measure deep seepage, careful water balance measurements are a possible approach for estimating at both the hillslope and catchment scales. Several community science

* Correspondence to: Chris B. Graham, Crop and Soil Science, Penn State University, 453 ASI Building, University Park, PA 16802, USA.
E-mail: cbg12@psu.edu

questions remain: How do deep seepage estimates scale from the hillslope to the catchment? How does deep seepage affect the hydrological function of hillslopes and catchments? How can we quantify the uncertainty associated with each of the water balance components, necessary for computing deep seepage loss?

To estimate deep seepage as the residual of a water balance closure (with the other components directly measured), a thorough estimation of uncertainty must be performed. Uncertainties in initial and boundary conditions such as soil and bedrock water storage, water table height, lateral and vertical subsurface fluxes and evapotranspiration are extremely difficult to measure (Eberhardt and Thomas, 1991) and are often coarsely estimated or simply assumed (Beven, 2006a). Estimates of inputs and internal state conditions are often based on limited point measurements extrapolated to much larger scales, resulting in increased uncertainty in estimates at the scale of interest (Sherlock *et al*., 2000). The heterogeneity of hydrological parameters, such as hydraulic conductivity, soil depth, macroscale soil structure and soil texture, compounds the measurement uncertainty challenge (McDonnell *et al*., 2007). Despite, or perhaps because of these difficulties, a rigorous assessment of measurement uncertainty in the context of closing the hillslope and catchment water balance in conjunction with field experimentation has not yet been attempted (Beven, 2006b).

Here, we present a hillslope irrigation experiment from the well-studied H. J. Andrews (HJA) Watershed 10 (WS10). The overall objective of the experiment was to estimate deep seepage by closing the water balance, identifying the relative partitioning of, and uncertainties around, the measured individual water balance components of evaporation, transpiration, lateral subsurface flow, bedrock return flow and fluxes into and out of storage within the soil profile. We perform a thorough uncertainty analysis of all stores and fluxes in the water balance, including the analysis of the error in all measurements made in the experiment. Within this overall objective, we address the following specific questions:

(1) Can we quantify deep seepage at the hillslope and catchment scales through careful closure of the water balance?
(2) How does deep seepage affect other hydrological processes?
(3) How do measurement uncertainties impact our process conceptualization of deep seepage at the hillslope and catchment scale?

To minimize uncertainty in the measurement of the hillslope water balance components, we used a controlled irrigation experiment rather than passive storm monitoring. Irrigation experiments have the benefit of control of the inputs and directed measurements of the outputs, and have been used effectively in the past for determining hillslope flowpaths (Hornberger *et al*., 1991) and transit times (Nyberg *et al*., 1999), among other applications. Our uncertainty analysis focuses on the measurements themselves and their propagation through rating curves and other functional uses of the data. Random (precision) and systematic (accuracy) uncertainty types are identified for each measurement instrument used and process assumption made.

SITE DESCRIPTION

The study hillslope is located in WS10 of the HJA Experimental Forest in the western Cascades, OR, USA (44·20 °N, 122·25 °W). The HJA is part of the Long-Term Ecological Research program, and has a data record of meteorological and discharge records from 1958 to the present. The climate is Mediterranean, with dry summers and wet winters characterized by long, low-intensity storms: analysis of the HJA long-term data record indicates that dry periods of 25 days during the summer and storms lasting 20 days during the winter have a 1-year return interval. WS10 has been the site of extensive research of hillslope hydrological processes (Ranken, 1974; Harr, 1977; McGuire *et al*., 2005; van Verseveld *et al*., 2009; Barnard *et al*., 2010).

Frequent debris flows at WS10 (most recently 1996) have scoured the stream channel to bedrock removing the riparian area in the lowermost reach. Soils are gravelly clay loams, classified as Typic Dystrochrepts, with poorly developed structure, high-hydraulic conductivities (up to 10 m/h, decreasing rapidly with depth), and high drainable porosity (15–30%) (Ranken, 1974). Surface soils are well aggregated, tending towards massive structure at depth. Soil depths on the study hillslope range from 10 cm adjacent to the stream, to 2·4 m at the upper limit of the irrigated area. Soils are underlain by Saprolite that thins towards the stream. Beneath this bedrock is mainly unweathered andesite and coarse breccias (Swanson and James, 1975; James, 1978). The study hillslope is steep (48°), extending 200 m to the ridge, although the irrigated area is only 20 m upslope from the stream channel. Additional site description can be found in McGuire *et al*. (2007).

METHODS

Hillslope delineation

The irrigated area was chosen so that it would drain downslope into the 10 m wide collecting trench (Figure 1). Both bedrock and surface topography are roughly planar in the area upslope of the collecting trench (van Verseveld *et al*., 2009), hence a trapezoidal area directly upslope of the trench was chosen for irrigation. This trapezoidal area was narrower at the top than the base, due to the uncertainty in subsurface flow paths. During irrigation, the area wetted by the sprinklers was delineated, and then measured after the experiment. The irrigated area was 9·4 m wide at the base, tapering to 8·2 m at the top and extended 20 m upslope of the trench (172 m^2). Wetted area width and length measurements

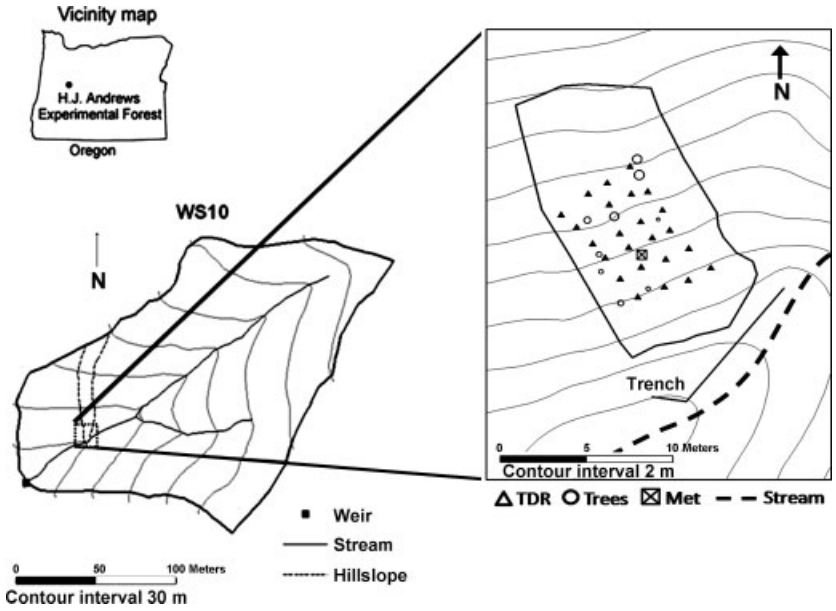

Figure 1. Map of the study site and irrigated area. Twenty-four TDR rods, nine instrumented trees and meteorological station are labelled

are estimated to have an uncertainty of 0·5 m due to uncertainties in measurement and determination of wetted area.

Irrigation application

A rectangular grid of 36 (9 rows of 4) micro-sprinklers (with ∼1 m irrigation radius) was installed on the hillslope, with sprinkler heads spaced 2 m apart. Sprinklers were controlled with an automatic timer to maintain a consistent application rate throughout the experiment with the exception of four minor malfunction periods. Sprinkler rate was measured by an array of 72 cups (42 cups of 0·05 m and 30 cups of 0·1 m diameter) that were sampled every 4–12 h during days 12–19 of the experiment. In addition, three tipping bucket rain gauges (Trutrack, Rain-SYS-1 mm, Christchurch, NZ) recorded irrigation rates throughout the experiment. The cups and tipping buckets were placed randomly in the sprinkled area between 0·1 and 0·8 m from the sprinkler heads. The uncertainty in the rainfall application was determined by the propagation of the uncertainty in the surface area of the measuring cups (cup radius ±0·001 m) and the volumetric measurements (±0·001 l). Approximately, 16 000 l water was irrigated each day, for a total of over 400 000 l for the duration of the experiment.

Lateral subsurface flow

Hillslope lateral subsurface flow was measured with a 10-m wide trench consisting of sheet metal anchored 0·05 m into bedrock and sealed with cement, installed at the intersection of the study hillslope and the exposed bedrock stream channel (McGuire *et al.*, 2007). The trench system is assumed to be nearly water tight, as no evidence of leakage was seen during the experiment. Bulk lateral subsurface flow was routed to a stilling well with a 30° V-Notch Weir, where a 0·25-m capacitance water level recorder (Trutrack, Model PLUT-HR, measurement ±0·0025 m) measured stage height at 10-min intervals.

A rating curve for the stage–discharge relationship was developed using 32 manual measurements of discharge covering the range of values experienced during the irrigation experiment ($R^2 = 0{\cdot}97$). The relative error between the manual measurements and the stage predicted discharge measurement averaged 8·76%, ranging from 1 to 20%. The relative error was negatively related to stage ($R^2 = 0{\cdot}34$). The absolute error averaged 12 l/s and was not related to stage. The absolute error was used as the systematic uncertainty in lateral subsurface flow, whereas the instrument precision (0·0025 m) was the random uncertainty.

Watershed discharge

Discharge from the second-order stream draining WS10 has been monitored with a broad crested weir 100 m downstream of the hillslope since 1969 as part of the long-term monitoring at the HJA. Watershed discharge is used to estimate water that bypassed the hillslope trench, either beneath via deep seepage or downstream via shallow flowpaths. We assume this water emerges in the stream channel downstream of the hillslope trench. A 90° V-Notch Weir had been installed for higher precision measurement of summer low flows, with the stage measured with a Model 2 Position Analog Transmitter (Model 2 Stevens Instruments, Protland, OR, USA; ±0·0003 m) recorder controlled by a data logger (Campbell Scientific CR10X, Logan, Utah, USA). A stage–discharge relationship was established based on 31 manual measurements of discharge taken over zero to two times the range of discharge seen during the irrigation experiment (Don Henshaw, personal communication). The absolute percent difference between measured and stage estimated discharge averaged 3·6%, with no

correlation between relative error and stage. The absolute error was positively related to stage ($R^2 = 0{\cdot}43$). The discharge during the experiment did not exceed the calibration range, hence the problems of rating curve indefinition are not expected (Clarke, 1999). The percent error in the rating curve was used as the systematic error, whereas the precision of the PAT was defined as the random error for uncertainty analysis.

During the course of the experiment, WS10 discharge receded, as the previous rainfall at the site was 10 days prior to the experiment. To determine the increase in watershed discharge due to the irrigation experiment, we created a master recession for WS10 using data from the summers of 2002 through 2004 from the WS10 gauging (data record available at www.andrewsforest.oregonstate.edu). Due to the variation in timing of the spring rainfall cessation, WS10 drainage began at different dates in different years, ranging from mid-June to late-July. Discharge recessions from the three summers were temporally aligned to begin with similar discharge rates. Discharge ($Q_M(t)$) was then modelled by:

$$Q_M(t) = Q_0 e^{-(t-t_0)/T_c} \quad (1)$$

where t is the day of year (DOY), T_c is the recession coefficient and Q_0 is discharge at the time t_0 (Chapman, 1999; Sujono *et al*., 2004). The recession coefficient $T_c = 28{\cdot}5$ days led to a very good fit to the average of the 3-year recession ($R^2 = 0{\cdot}97$). We applied this function to our 24·4-day experiment period to determine the increase in WS10 discharge due to hillslope irrigation, using t_0 as Julian Day 200 and Q_0 as 1118 l/h.

Uncertainty in the master recession has two components, uncertainty in initial discharge, Q_0, and uncertainty in the recession coefficient T_c. The uncertainty in Q_0 was taken as the uncertainty in the measurement of watershed discharge at time zero, or $3{\cdot}5\% \times 1118$ l/h = 39 l/h. The uncertainty in T_c was determined by fitting an exponential to the 3-year average watershed recession, then varying the uncertainty in T_c until 80% of the average recession readings fell within the error bounds. Due to fluctuations in the individual recessions caused by small rainfall events each summer, a relatively lenient uncertainty bound was chosen, rather than attempt to bound 90 or 95% of the recession. This leads to an estimate of the uncertainty in T_c of 18·6%. The uncertainties in T_c and Q_0 were considered systematic, as they were not based on measurements made during the experiment.

Transpiration and canopy reference evapotranspiration

Transpiration was estimated from xylem water flux measurements of the dominant (diameter base height >5 cm) trees located within or bordering the sprinkled area ($n = 9$) beginning 10 days prior to irrigation (DOY 199) and continuing for 60 days after irrigation stopped (DOY 293). Of the mature trees located within the irrigation area Douglas-fir ($n = 6$), western hemlock (*Tsuga heterophylla*) ($n = 2$) and cascara (*Rhamnus purshiana*) ($n = 1$) represented 67, 27 and 6% of the total basal area, respectively. Dominant understory species include: western swordfern (*Polystichum munitum*), bear grass (*Xerophyllum tenax*), Oregon grape (*Berberis nervosa*) and salal (*Gaultheria shallon*). Sap flux was measured using the constant-heat method (Granier, 1987):

$$Q_n = 0{\cdot}0119 \mathrm{sa}_n \left(\frac{\Delta T_{M,n} - \Delta T_n}{\Delta T_n} \right)^{1{\cdot}231} \quad (2)$$

where sa_n is the sapwood area of the nth tree, ΔT_M is the maximum daily temperature difference between probes and thermocouples installed into the sapwood 10–15 cm apart, and ΔT is the instantaneous temperature difference.

Temperature difference was measured every 15 s using copper–constantan thermocouples hooked in series to measure temperature difference ($\pm 0{\cdot}2\,^{\circ}$C) and stored in a CR-10x data logger (Campbell Scientific, Logan, UT, USA) as 15 min means. We used 0·02 m probes for the sap flux measurements. Sapwood depths were determined by visually examining and measuring tree cores from the height of the sap flux sensors on each tree ($\pm 0{\cdot}001$ m). For trees with sapwood depths greater than 2 cm, corrections for radial variations in sap flux were estimated from measured radial sap flux profiles of trees of the same species and age at another location following methods outlined in Domec *et al*. (2006) and Moore *et al*. (2004). Analysis of the sapflow data *a posteriori* indicated that the daily maximum temperature difference between the sensors was constant and night-time vapour pressure deficits were low (<450 Pa), indicating that the assumption of zero flux at night was valid (for more details on the sapflow measurements and results, see Barnard *et al*., 2010). Uncertainties in the transpiration estimates were treated as systematic, as onsite calibration of the sapflow equation and thermocouples were not performed.

Meteorological measurements for calculating canopy reference evapotranspiration (CRET) were taken at a weather station located in the irrigated area (Figure 1). The weather station was a 6-m tower with sensors located at the standard heights (see below). CRET was estimated using the standard Penman–Monteith Monteith and Unsworth, (2008) equation for CRET, estimated using measured meteorological data taken at the site:

$$\mathrm{CRET} = \frac{\Delta(R_n - G) + \rho c_p \dfrac{\mathrm{VPD}}{r_a}}{\lambda\left(\Delta + \gamma\left(1 + \dfrac{r_c}{r_a}\right)\right)} \quad (3)$$

where Δ is the partial derivative of the saturated vapour pressure curve with respect to temperature, R_n is the net incoming radiation, G is the ground heat flux, ρ is the dry air density, c_p is the specific heat capacity of air, VPD is the vapour pressure deficit, r_c is the canopy resistance, r_a is atmospheric resistance, λ is the latent heat of vaporization and γ is the psychrometer constant. The parameters ρ, c_p λ and γ were assumed to be

constant, and values presented by Monteith and Unsworth (2008) were used. The VPD is the product of the relative humidity (RH) and saturated vapour pressure, $e_s(T)$, and Δ is the first derivative of the $e_s(T)$ curve with respect to temperature. The saturated vapour pressure $e_s(T)$ was calculated using an empirically derived exponential function of temperature (Tetens, 1930; Murray, 1967). The atmospheric resistance, r_a is a function of wind speed (Monteith and Unsworth, 2008):

$$r_a = \frac{\ln((z-d)/z_0)}{ku} \quad (4)$$

where z is the height of the canopy (22 m), d is the zone of zero displacement ($0{\cdot}65z$), k is von Karman's constant, z_0 is the roughness length ($0{\cdot}1z$) and u is the wind speed measured at the hillslope. The canopy resistance, r_c, generally ranging from 100 to 250 s/m (Tan and Black, 1976) is a function of the forest type and structure, vapour pressure deficit and the water potential gradient from the soil to the leaves. Because we lacked water potential measurements needed to calculate site-specific r_c during the experiment, r_c was assumed to be 150 s/m based on measurements from a nearby watershed of similar forest age and structure (Pypker, unpublished data), and the uncertainty was assigned to encompass the wide range observed by Tan and Black (1976), ±75 s/m. Ground heat flux was expected to be small, and estimated as 10% of net radiation, with a similarly large uncertainty, in this case 100% ($G = 0$–$20\%R_n$).

Net radiation (R_n; Campbell Scientific Inc., Logan, Utah, USA, model Q-7·1, ±6%), RH (Campbell Scientific Inc., model HMP 35C, ±2–3%), air temperature (T; Campbell Scientific Inc., Logan, Utah, USA, model HMP 35C, ±0·4 °C) and wind speed (u; R. M. Young Wind Monitors, Traverse City, Michigan, USA, model 05305, ±0·2 m/s) were measured at 15-min intervals throughout the experiment. As a calibration of the meteorological equipment was not performed, the measurement uncertainty presented by the manufacturers was propagated as systematic error through the functional uses of the measured data.

Soil moisture

Soil moisture (volumetric water content) was measured at 24 locations within the irrigated area, at 5 depths in each location (0–15, 15–30, 30–60, 60–90 and 90–120 cm) with a time domain reflectometry (TDR) array (Environmental Sensors, Inc., Sidney, British Columbia, CA, model PRB-A, ±3%; Figure 1). The TDR array consisted of a 4 × 6 grid (parallel and perpendicular to the stream channel, respectively), with a TDR sensor spacing of 2 m in each direction. Soil moisture was measured hourly through the experiment. Of the 120 measurement sites (locations and depth), 57 of the probe segments gave consistent results. The remaining 63 measurement segments had data recording problems due to probes that were installed incompletely in the soil profile, poor electrical connection or poor contact between the probe and soil. Only data from consistently working probes were analysed. To determine the total soil storage, the depth weighted average measured soil moisture was multiplied by the estimated soil depth (1·2 ± 0·1 m) and the irrigated area (172 ± 8 m^2).

Two significant sources of error lie in the soil moisture data. The first is the uncertainty in the soil moisture measurements. This uncertainty is estimated to be 3% for each measurement based on error estimates from the manufacturer. The uncertainty in the background, pre-experiment water content was similarly assumed to be 3%. The second source of the uncertainty is the subsurface volume represented by the soil moisture measurements themselves. We assumed that soil moisture outside the TDR grid, but within the sprinkled area, reacted similarly to the area measured by the probes. There is a possibility of some flux of water outside of the sprinkled area due to capillary effects and subsurface flow paths diverting water from the sprinkled area. Additional storage could have occurred in the bedrock itself, which was likely unsaturated prior to the experiment. Our computed subsurface storage volumes were considered to represent a minimum value of total subsurface storage. As calibration of the soil moisture probes was not conducted in the field, the factory calibration uncertainty in soil moisture readings were treated as systematic.

UNCERTAINTY ACCOUNTING AND ESTIMATION

We subdivided our uncertainty analysis into three categories: identification and quantification of measurement uncertainty of the instruments, propagation of the measurement uncertainty through the functional uses of the data and propagation of measurement uncertainty through aggregated measures. We define and describe the mathematical treatment of these terms below.

Individual measurement uncertainty

Measurement error is the uncertainty in the precision and accuracy of the field instrument. The uncertainty in field measurements can be determined in a number of ways, including field calibration, manufacturer calibration and expert opinion. We included all the above ways to determine the uncertainty for this experiment. Hillslope and watershed discharge were determined from field calibration of the stage discharge relationship. Much of the meteorological data were not calibrated in the field, hence the factory calibration uncertainties were used for each of the individual readings. Some variables used in the extended analysis (pre-irrigation hillslope and watershed discharge) were not measured directly throughout the experiment, but were based on historical data. We estimated the uncertainty of these variables based on expert opinion.

Of these sources of error, there are two types, i.e. random and systematic errors, expressed as the precision and accuracy of the measurements. Random errors include measurement errors that deviate randomly from

Table I. Measurements used in the calculation of the water balance components and their estimated random and systematic errors

Water balance component	Measurement	Random uncertainty	Systematic uncertainty
Precipitation	Precipitation rate	—	±1 mm/h (SD rainfall rate)
	Rainfall volume	±1 ml (estimate)	—
	Wetted area (A)	±8 m^2 (estimate)	±8 m^2 (estimate)
Lateral subsurface flow	Stage (s)	±0·25 mm (factory)	±0·0034 l/s (calibration)
WS10 discharge	Stage (s)	±0·3 mm (factory)	±3·6% (calibration)
Transpiration	Temperature (T)	±0·2 °C (estimate)	±0·2 °C (estimate)
	Sapwood depth (sa)	±1 mm (estimate)	±1 mm (estimate)
CRET	Temperature (T)	±0·4 °C (factory)	±0·4 °C (factory)
	Wind speed (u)	±0·2 m/s (factory)	±0·2 m/s (factory)
	Ground heat flux coefficient (a)	±100% (estimate)	±100% (estimate)
	Incoming net radiation (R_n)	±6% (factory)	±6% (factory)
	RH	±2–3% (factory)	±2–3% (factory)
	Wetted area (A)	±8 m^2 (estimate)	±8 m^2 (estimate)
	Canopy resistance (r_c)	—	±75 s/m (estimate)
Soil moisture	Volumetric water content (S)	±3% (factory)	±3% (factory)
	Wetted area (A)	±8 m^2 (estimate)	±8 m^2 (estimate)

The sources of the uncertainty estimates are either factory reported accuracy and precision or our estimates of measurement uncertainty.

the true observed value. These errors are assumed to be evenly distributed above and below the true value and to some extent cancel each other out when aggregated to longer time periods and spatial scales. Systematic errors, on the other hand, can affect all measurements in the same direction (i.e. under or over prediction), and thus do not diminish with increasing the length of the data set. For most of the field instruments, the manufacturer presents only one uncertainty estimate. In this case, when a field calibration has not been made, this value is treated in the uncertainty analysis as both the systematic and random errors. In cases where a calibration has occurred, such as hillslope and catchment discharge, the systematic error (accuracy) is taken from the uncertainty of the calibration, whereas the random error (precision) is taken from the equipment measurement uncertainty. Table I lists the source and type of uncertainty for each measurement.

Error propagation

Uncertainty in field measurements needs to be transferred through the functional uses of the data. Because many of the measured variables, such as temperature and wind speed, are incorporated into nonlinear equations such as the Penman–Monteith equation for CRET, the uncertainty in the calculated water balance component is also nonlinear. To account for this nonlinearity, uncertainties in measurements are propagated using the standard error propagation formula (Taylor, 1997).

When propagating error, we first assumed that the individual instruments were independent from each other. If q is a function of N variables:

$$q = f(x_1, \ldots x_N) \tag{5}$$

where x has some random uncertainty $\delta_r x_n$ and some systematic uncertainty $\delta_s x_n$, then uncertainty in each of the measured values is propagated through q by:

$$\delta q = \sqrt{(\delta_s q)^2 + (\delta_r q)^2} \tag{6}$$

where δq is the propagated error in q, and

$$\delta_r q = \sqrt{\sum_{n=1}^{N} \left(\frac{\partial q}{\partial x_n} \delta_r x_n \right)^2} \tag{7}$$

and

$$\delta_s q = \sqrt{\sum_{n=1}^{N} \left(\frac{\partial q}{\partial x_n} \delta_s x_n \right)^2} \tag{8}$$

For example, transpiration is calculated following the empirical relationship developed by Granier (1987) in Equation (2). Measured variables include the sapwood area (sa), the temperature difference between two thermocouples inserted into the sapwood (ΔT) and the maximum daily temperature difference (ΔT_M). The uncertainty in transpiration is then:

$$\delta Q = \sqrt{(\delta_r Q)^2 + (\delta_s Q)^2} \tag{9}$$

where

$$\delta_r Q = \sqrt{\left(\frac{\partial Q}{\partial s} \delta_r \mathrm{sa} \right)^2 + \left(\frac{\partial Q}{\partial \Delta T} \delta_r \Delta T \right)^2 + \left(\frac{\partial Q}{\partial \Delta T_M} \delta_r \Delta T_M \right)^2} \tag{10}$$

$$\delta_s Q = \sqrt{\left(\frac{\partial Q}{\partial s} \delta_s \mathrm{sa} \right)^2 + \left(\frac{\partial Q}{\partial \Delta T} \delta_s \Delta T \right)^2 + \left(\frac{\partial Q}{\partial \Delta T_M} \delta_s \Delta T_M \right)^2} \tag{11}$$

Complete propagation of error through the various formulae used in the calculation of the water balance components is presented in Graham (2008).

Aggregated error

When aggregating measurements from individual time steps to longer time scales (i.e. daily averages, whole

experiment total fluxes) the type of measurement error, whether random or systematic, determines how the aggregation of the error is performed. Random errors, when aggregated, diminish with increasing size of the data set according to $1/\sqrt{T}$, where T is the number of data points (Barlow, 1989). For random errors, the error of the aggregate is the sum of the squares. If, for instance

$$\hat{q} = \sum_{t=1}^{T} q(x_t) \tag{12}$$

and the error in x is random, then the aggregated error is:

$$\delta_r \hat{q} = \sqrt{\sum_{t=1}^{T} \left(\frac{\partial q}{\partial x_t} \delta_r x_t \right)^2} \tag{13}$$

for measurand x from time $1{:}T$.

Systematic errors must be aggregated differently. As persistent offsets or multipliers to the data, they act in an additive manner (Moncrieff *et al*., 1996) and do not diminish with increasing data set size. When propagated, the error of the aggregate is the square of the sum:

$$\delta_s \hat{q} = \sqrt{\left(\sum_{t=1}^{T} \frac{\partial q}{\partial x_t} \delta_s x_t \right)^2} \tag{14}$$

for measurand x from time $1{:}T$. For values that aggregate over long-time periods, with large T, the aggregated random error is dwarfed by the systematic error, as T becomes much larger than $\sqrt{T}$. More detailed description, including formulas for the error propagation formulas for each data source can be found in Graham (2008).

RESULTS

Water balance components

Inputs. Irrigation application was relatively constant for the 24·4-day experiment with the exception of four malfunctions in the timer apparatus that caused the irrigation to remain either on or off for a short period of time. Irrigation began at 05 : 30 h on Julian Day 208 (27 July 2005), and ended at 14 : 12 h, Julian Day 232 (20 August 2005). On midnight, day 210 irrigation turned off for 9 h. The irrigation rate was constant for the next 18 days. Sprinkler malfunctions also occurred on days 228, 229 and 230.

The weighted irrigation rate based on the 72 collection cups was $3{\cdot}6 \pm 0{\cdot}2$ mm/h. With a measured irrigated area of 172 ± 8 m^2, the corresponding total application was 659 ± 33 l/h for a total applied water volume of $394\,000 \pm 19\,700$ l. Irrigation rates varied spatially due to both variations in the application rates of individual sprinkler heads and temporary obstructions (including vegetation and equipment) between sprinklers and measuring cups (standard deviation = 3·3 mm/h). This variability is more a measure of the spatial variability of application than a measure of application rate uncertainty.

Outputs—lateral subsurface flow. Lateral subsurface flow measured at the trench responded quickly to irrigation, with a detectable increase in discharge within 1 h of irrigation initiation (Figure 2). Lateral subsurface flow increased from a pre-irrigation daily average rate of 30 ± 1 l/h to a steady-state daily average value of 284 ± 20 l/h within 5 days. Before, during and after the experiment, a clear diel pattern in flow was evident. Steady-state discharge was maintained for 13 days, after which a series of sprinkler malfunctions increased discharge by over 30% for 3 days. At the end of the irrigation, on Julian Day 232, the instantaneous lateral subsurface flow was 270 ± 16 l/h. After irrigation ceased, lateral subsurface flow returned to within 200% of pre-experiment levels within 24 h. Total lateral subsurface flow for the duration of the experiment was $102\,543 \pm 7451$ l. Total lateral subsurface flow for the periods of the experiment plus 5 and 10 days was $106\,156 \pm 8979$ l and $107\,760 \pm 10\,507$ l, respectively (Table II). A rainfall

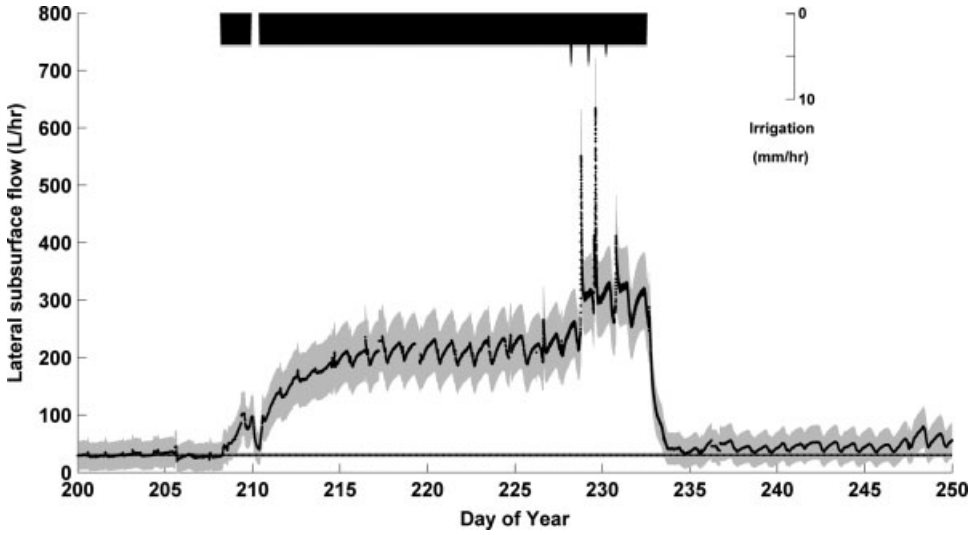

Figure 2. Irrigation and lateral subsurface flow (black) with uncertainty bounds (grey) measured at the hillslope trench. Dashed line is background flow rate projected through time to show how the irrigation-induced flow compares to background

Table II. Water balance components with propagated uncertainty

Water balance component	Steady state (l/h)	Entire irrigation experiment (l)	Irrigation + 5 days (l)	Irrigation + 10 days (l)
Irrigation	659 ± 33	394 000 ± 19 700	394 000 ± 19 700	394 000 ± 19 700
Hill	254 ± 20	102 543 ± 7451	106 156 ± 8979	107 760 ± 10 507
WS10	461 ± 115	252 125 ± 48 035	275 523 ± 56 935	295 781 ± 65 578
Transpiration	9 ± 1	5448 ± 343	6456 ± 409	7318 ± 470
CRET	35 ± 13	30 055 ± 12 692	35 767 ± 15 170	41 114 ± 17 721
ΔS	4 ± 1	25 837 ± 1565	15 718 ± 1560	11 438 ± 1559
Hillslope deep seepage	207 ± 117	149 582 ± 48 563	169 567 ± 57 639	188 021 ± 66 423
Catchment deep seepage	144 ± 121	85 983 ± 53 469	66 992 ± 62 147	45 667 ± 70 746

Hillslope deep seepage is defined as the difference between the WS10 and hill fluxes. Catchment deep seepage is defined as the residual of the water balance ($P - \text{WS10} - E - \Delta S$).

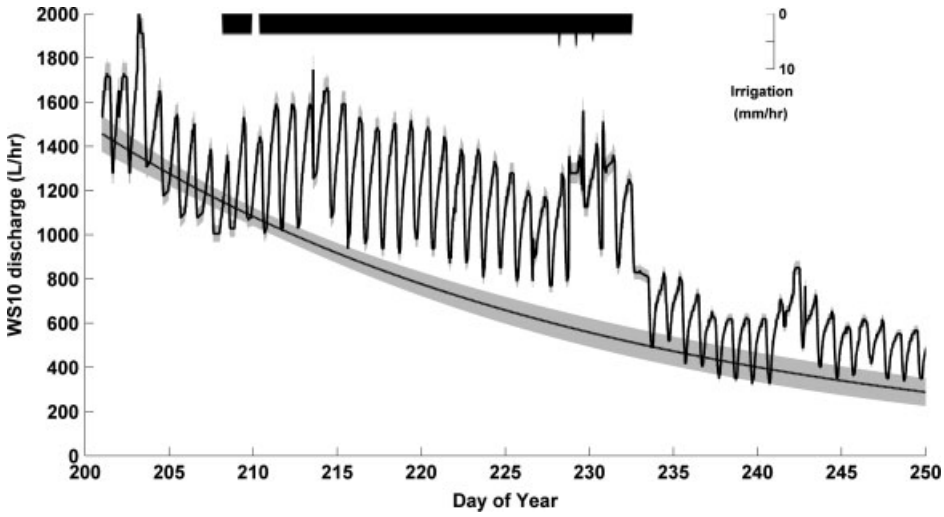

Figure 3. Irrigation water, WS10 discharge (solid black) and WS10 master recession (dash), with uncertainty bounds (grey). Note increased uncertainty in the master recession at late time as irrigation progresses

event 10·5 days after the end of the experiment prevented longer analysis of the irrigation based water balance.

Outputs—WS10 discharge. WS10 discharge responded to irrigation similarly to the lateral subsurface flow measured at the hillslope trench (Figure 3). The pre-irrigation WS10 recession slowed within 1 h after the onset of irrigation. After 5 h, discharge began to increase and continued to increase for the next 6 days of the experiment. After day 6, WS10 discharge recession resumed, now parallel to the master recession curve. This was due to the combined steady input from the irrigated hillslope and continued recession from the remaining area of the watershed. Comparison to the master recession indicated an increase in discharge due to the sprinkling of 461 ± 115 l/h during the period of steady-state input. The recession remained parallel to the master recession until the series of sprinkler malfunctions caused an increase in discharge similar to that seen at the hillslope. After cessation of the sprinkling, WS10 drainage decreased slower than that observed at the trenched hillslope; whereas hillslope discharge returned to pre-event levels within 1 day of the end of irrigation, WS10 discharge did not return to the master recession prior to a rain event 10 days after the end of irrigation. Total increased discharge measured at the watershed outlet for the duration of the irrigation was 252 125 ± 48 035 l with 275 523 ± 56 935 and 295 781 ± 65 578 l for the irrigation plus 5 and 10 days drainage, respectively (Table II). The uncertainty in the aggregated measures increased due to increased uncertainty in background watershed discharge at late time.

Outputs—transpiration and CRET. Transpiration from the dominant trees in both the sprinkled area and CRET showed a strong diel pattern, during and after the irrigation experiment (Figure 4). Sap flux averaged 0·8 ± 0·1 l/h for the nine instrumented trees, for a total sap flux of 9 ± 1 l/h for the stand of trees on the instrumented hillslope. These values of total daily transpiration are comparable to those measured in Douglas-fir of similar age at the HJA during high soil moisture conditions (Barnard, 2009). The maximum instantaneous stand flux rate of 25 ± 2 l/h typically occurred around early afternoon (14 : 00 h). The maximum nightly temperature difference measured by the individual sapflow sensors was relatively stable (range <0·5 °C), supporting our assumptions of negligible sap flux at night. While the mean and maximum flux rates remained constant before, during

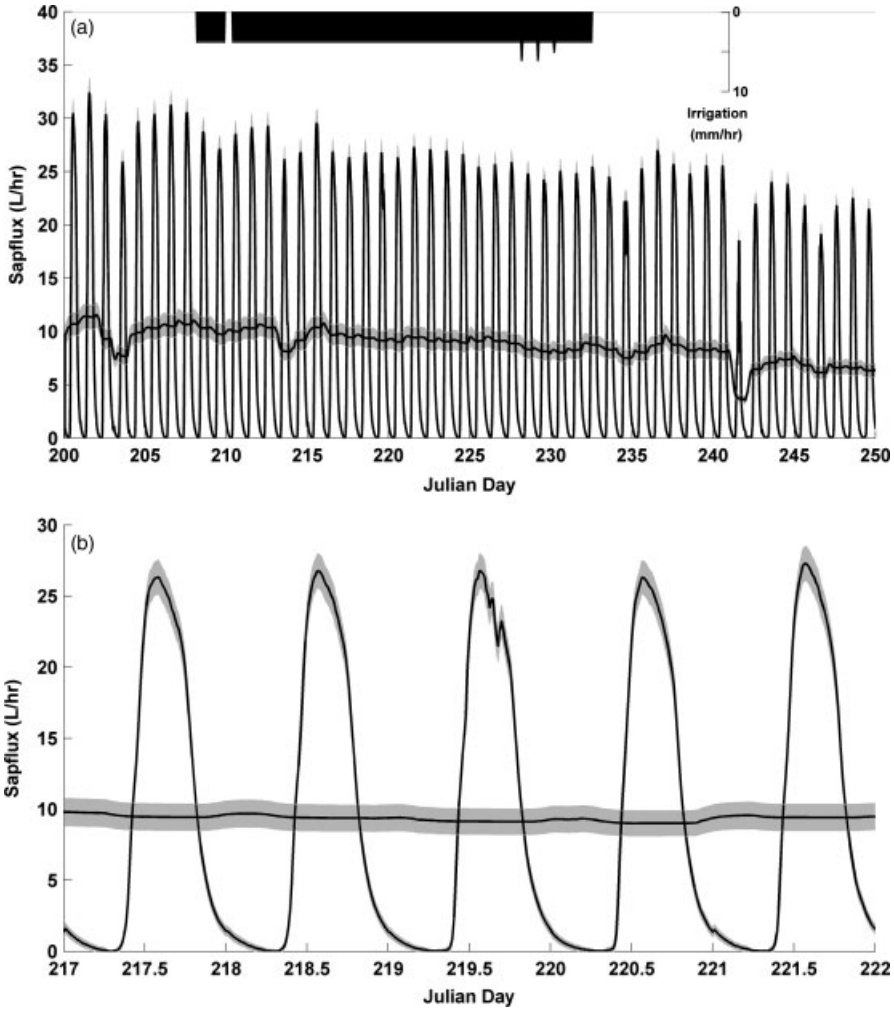

Figure 4. Instantaneous and daily average transpiration (black) with uncertainty bounds (grey) measured from nine trees on the site for the duration of the experiment: (a) throughout the experiment and (b) during steady state. Transpiration declines throughout the experiment

and after the experiment, suggesting that trees in the plot were not water stressed at the onset of irrigation, monitored trees outside of the irrigated area (not presented) showed a decrease in transpiration throughout the sampled time period. This indicates that the irrigation did result in increased transpiration response, although not an increase in absolute transpiration. A detailed analysis of transpiration and water use patterns of vegetation during the experiment is discussed in Barnard *et al.* (2010). Transpiration for the duration of the experiment totalled 5448 ± 343 l. Transpiration for the period of the experiment +5 and +10 days was 6456 ± 409 and 7318 ± 470 l, respectively (Table II).

CRET declined from a high at the initiation of irrigation ($36{\cdot}0 \pm 13$ l/h—daily average) through the end of the experiment ($33{\cdot}5 \pm 12$ l/h) and to 5 and 10 days after the end of irrigation ($30{\cdot}4 \pm 12$ and $26{\cdot}4 \pm 10$ l/h, respectively; Figure 5). This decline was influenced primarily by incoming net radiation, which declined throughout the monitoring period, due to a reduction in the daylight hours. As the soil remained wet and water supply was not likely the limiting factor for evapotranspiration, we assume that actual evapotranspiration equals CRET during the experiment and afterwards (Rodriguez-Iturbe, 2000). Total evapotranspiration for the duration of the experiment was estimated as $30\,055 \pm 12\,692$ l. Evapotranspiration for the period extending 5 and 10 days afterwards was $35\,767 \pm 15\,170$ and $41\,114 \pm 17\,721$ l, respectively (Table II).

Change in storage—soil moisture. Soil moisture followed the same general pattern as the hillslope and WS10 discharge: a quick response to irrigation, then near steady-state conditions, and a recession after the irrigation ended on day 232 (Figure 6). The change in soil moisture storage (Q_S) was calculated as:

$$Q_S = \Delta S A d \tag{15}$$

where ΔS is the difference in average soil column water content before the experiment and at the measurement time, A is the area sprinkled and d is the average soil depth. The average soil depth on the sprinkled hillslope is $1{\cdot}2 \pm 0{\cdot}1$ m, and the area sprinkled was 172 ± 8 m^2. This formula assumes that the depth of soil storage is equivalent to the soil depth (i.e. the bedrock is saturated), and the aerial extent is equivalent to the sprinkled area (i.e. no lateral spreading parallel to the trench).

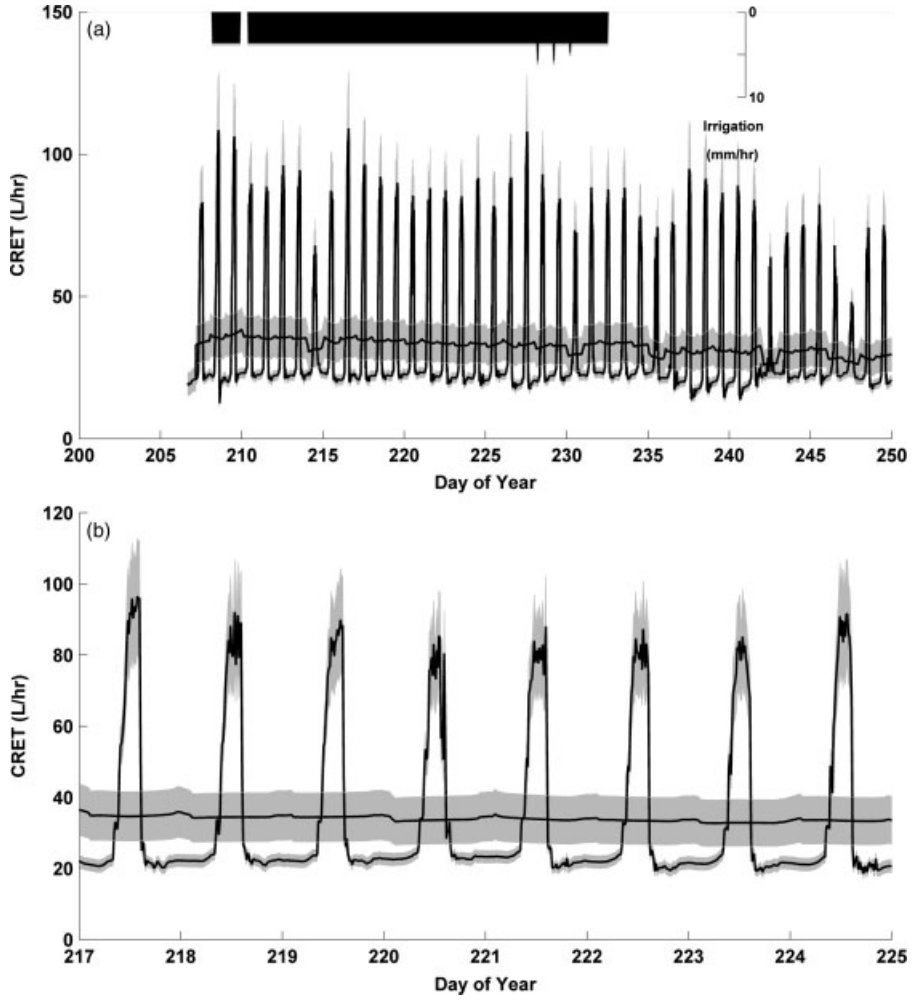

Figure 5. Instantaneous and daily average CRET (black) with uncertainty bounds (grey) for (a) the experiment duration and (b) during steady-state conditions

Volumetric water content averaged $8{\cdot}6 \pm 0{\cdot}3\%$ at the onset of irrigation. TDR readings showed an initial increase in soil moisture in the upper 0·6 m in the first 30 min of irrigation. Soil moisture at 0·6–0·9 m increased after 90 min, and sensors below 0·9 m increased after 150 min. Soil moisture reached near steady-state conditions within 5–6 days (days 213–214), with the shallower depths reaching steady state more quickly than at depth. From days 215 through 228, soil moisture was nearly constant, increasing 0·6% in 13 days, from 18·4 to 19·0%, representing a flux of 4 ± 1 l/s. Near steady-state conditions persisted until day 228, when the first of the sprinkler malfunctions caused an increase in soil moisture. After irrigation ceased on day 232, the soil profile drained quickly for the first 8–12 h from an average volumetric water content of $21{\cdot}5 \pm 0{\cdot}6\%$, followed by a slower, more sustained drainage for the duration of monitoring. The upper soil layers drained most rapidly with slower drainage at depth. The average profile soil volumetric water content dropped to $16{\cdot}6 \pm 0{\cdot}5\%$ within 5 days, and $14{\cdot}4 \pm 0{\cdot}4\%$ within 10 days. None of the five soil profiles returned to pre-irrigation levels by day 250, over 3 weeks after irrigation ceased.

At steady state, the total soil storage is estimated at $41\,693 \pm 4208$ l. At the end of irrigation, after the series of sprinkler malfunctions, total storage was $44\,376 \pm 4479$ l. Total profile storage declined to $34\,324 \pm 3465$ and $29\,784 \pm 3006$ l after 5 and 10 days, respectively (Table II).

Deep seepage at the hillslope scale. Deep seepage at the hillslope scale was estimated as the difference between the measured lateral subsurface flow at the hillslope trench and the increase in catchment discharge. This defines deep seepage at the hillslope scale as water that infiltrates into the bedrock at the hillslope, to re-emerge relatively quickly as bedrock return flow in the stream channel. At steady state, the estimated deep seepage at the hillslope scale was 207 ± 117 l/s. Estimates of deep seepage increased with the integrated interval from $149\,582 \pm 48\,609$ (duration of irrigation), $169\,367 \pm 57\,639$ (+5 days) and $188\,021 \pm 66\,414$ l (10 days).

Catchment scale deep seepage. Catchment scale deep seepage, subsurface flow that was not measured at the catchment outlet, was estimated as the residual from

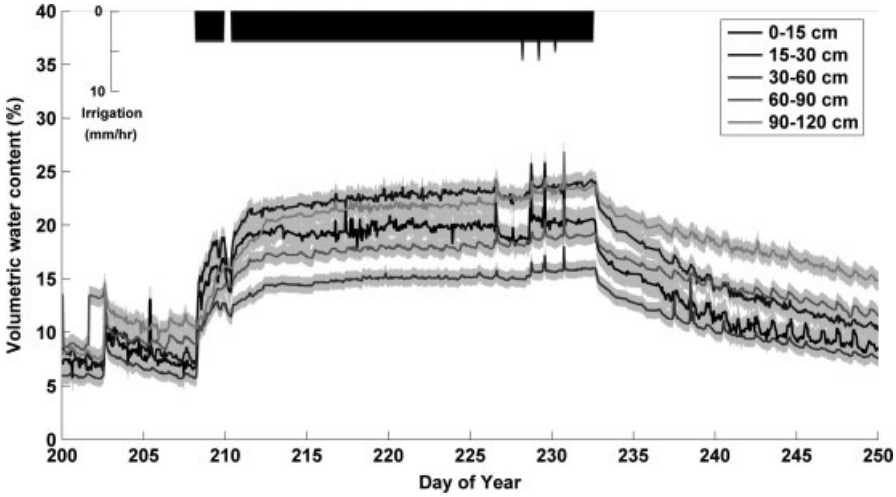

Figure 6. Volumetric soil water content averaged for five measurement depths (black lines) with uncertainty bounds (grey). Note the rapid recession of the shallower depths (0–30 cm) and the rapid response to sprinkler malfunction on day 210. At steady state, the TDR rods at 15–30 cm recorded the highest water content, followed by 90–120, 0–15, 60–90 and 30–60 cm

the water balance equation, using measured precipitation, discharge, ET and changes in soil moisture storage. Catchment scale deep seepage was estimated as the difference between precipitation and three measured components of the water balance: (1) the increase in catchment discharge due to irrigation, which included lateral subsurface flow measured at the hillslope trench and hillslope scale deep seepage that re-emerged in the stream channel; (2) evapotranspiration that includes transpiration measured as sapflow in the instrumented trees and (3) changes in soil moisture storage. The uncertainty in each of the averaged values was calculated using the error propagation approach in the section on Error propagation.

Steady-state rates were calculated as the average values for the period DOY 219 through 226. During this period, the system was close to steady state, as evidenced by relatively steady lateral subsurface flow, soil moisture storage and WS10 discharge, and did not include the period of sprinkler malfunctions (Figures 2–4). During this period, deep seepage accounted for 144 ± 121 l/s (listed as the residual in Table II).

Cumulative deep seepage was estimated as the difference of the total irrigated volume and the cumulative flow volumes for each component for the duration of irrigation, the duration +5 days and the duration +10 days (Table II). Estimated integrated deep seepage volumes are $85\,983 \pm 53\,469$, $66\,992 \pm 62\,147$ and $45\,667 \pm 70\,746$ l for the three time periods.

DISCUSSION

Developing process understanding of deep seepage

Deep seepage at the hillslope scale. Estimated deep seepage at the hillslope scale was 207 ± 117 l/s at steady state (or $45 \pm 25\%$ of the increase in catchment discharge). This component consisted of water that was observed at the catchment outlet, but not at the trench. Somehow this water bypassed the trench system, and was expressed in the stream channel. We have two hypotheses for this bypassing: (1) flow routing around (likely down valley) of the hillslope trench and (2) leakage underneath the hillslope trench. While the planar, delineated irrigated area was assumed to drain downslope into the trench, there is uncertainty in the exact locations of the dominant subsurface flowpaths. Studies elsewhere have shown that bedrock topography is often a first-order control on flow routing at similar, forested hillslopes (Freer *et al*., 2002; Graham *et al*., 2010). Analysis of soil depth measurements at our site indicates a relatively planar bedrock surface, parallel to the soil surface. However, the spatial scale or grain of the bedrock features that can control such routing can be very small, on the order of centimetres (Graham *et al*., 2010). Our map of soil depth was made on a 1-m grid, and likely did not capture these small scale features. Some evidence of down valley flow routing around the hillslope trench was observed in the form of bank seepage downstream of the trench, although the observed seepage was a small fraction (estimated $<10\%$) of the measured lateral subsurface flow, and could not account for the large amount of water unaccounted for by the trench.

An alternative hypothesis to the one above is that water seeped into near surface bedrock and re-emerged in the stream channel downstream of the trench. While irrigation was confined to the area near the trench (<20 m upslope), and a 1-m buffer was placed on each side to minimize flow bypassing to the right or left of the trench, 39% of water measured in the stream at the WS10 outlet was not observed in the trench. The hillslope trench, while not assumed to be entirely watertight, is designed to minimize leakage, and thus is expected to capture the vast majority of the lateral subsurface flow at the hillslope base. If we assume that this water bypassed the hillslope trench by infiltrating into the bedrock across the wetted cross-sectional area, this would correspond to a leakage rate of $1{\cdot}1 \pm 0{\cdot}6$ mm/h. While significant, these rates

are well below the measured hydraulic conductivities of other steep, forested hillslopes that had previously been assumed impermeable (~5 mm/h at sites underlain by a conglomerate of sandstone, granite and schist in a clay matrix (Maimai, New Zealand; Graham *et al*., 2010) and Saprolite derived from granite and Granodiorite (Panola, Georgia; Tromp-van Meerveld *et al*., 2006).

Previous WS10 hillslope storm monitoring of lateral subsurface flow by McGuire *et al*. (2007) reported that the hillslope area defined by the collection trench placement (upslope contributing area of 1·7% of the watershed), contributes 2% of the annual catchment discharge. This indicates that the majority of the water falling on the hillslope is observed in the trench, and losses to deep seepage at the hillslope scale were minimal. However, their estimate of the runoff ratio for the hillslope was dependent on an accurate assessment of the upslope contributing area, an easily calculated but very imprecise measure at the hillslope scale. Woods and Rowe (1997) demonstrated that small uncertainties in topography measurements greatly affected the upslope contributing area for a hillslope trench system. In addition, if bedrock topography rather than surface topography controls flow routing, further uncertainties arise.

Transient lateral flow through the near surface bedrock has been observed at a number of field sites (Montgomery *et al*., 1997; Katsuyama *et al*., 2005; Tromp-van Meerveld *et al*., 2006; Katsura *et al*., 2008). While the re-emergence of water lost to bedrock at the hillslope scale has been observed downstream (Montgomery *et al*., 1997); generally, the fate of this water, its interaction with shallow lateral subsurface flow paths, and the travel to the stream channel are unclear and poorly understood. Although it is often assumed that the bedrock is effectively impermeable on the timescale of event-based hillslope experimentation and monitoring (Mosley, 1979; Freer *et al*., 2002), recent evidence has shown that significantly permeable bedrock is the rule rather than the exception at steep, forested hillslopes (Montgomery *et al*., 1997; Katsuyama *et al*., 2005; Tromp-van Meerveld *et al*., 2006; Katsura *et al*., 2008). Leakage at the hillslope scale and re-emergence at the catchment scale are a possible explanation for the high discrepancies often seen between hillslope and catchment runoff ratios (Woods and Rowe, 1996).

Deep seepage at the catchment scale. During and immediately after the irrigation experiment, a significant amount of the irrigation was not accounted for at the catchment scale. The amount of unaccounted water decreased as the measurement time increased, declining from $22 \pm 14\%$ at the end of the experiment, to $17 \pm 16\%$ after 5 days drainage, to $12 \pm 18\%$ after 10 days drainage. At steady state, the unaccounted water was $22 \pm 18\%$ of the irrigated water. The increase in the residual uncertainty with increasing integrated time was due primarily to increased uncertainty in the WS10 master recession. A natural rainfall event occurred 10 days after the experiment, preventing further monitoring. Two possible explanations for this missing water are discussed: recharge to storage in the bedrock and catchment scale deep seepage that leaves the catchment via flow paths that do not include the WS10 weir.

In the calculation of the storage component of the water balance above, it was assumed that the storage was confined to the soil horizon. When accounting for the discrepancy between the hillslope and catchment recovery, however, we hypothesize that leakage through the bedrock was a significant flowpath. If we assume that flow through the bedrock is significant, and further assume that the bedrock was unsaturated prior to irrigation (due to 10+ days of antecedent drainage time before irrigation), it is reasonable to argue that increased water storage in the bedrock during the irrigation experiment was non-trivial. Storage in the bedrock that would drain and contribute to WS10 discharge, but not to measurable lateral subsurface flow (as that was dominated by shallow flow above the soil bedrock interface). This would further account for the rapid decline in lateral subsurface flow and the slow recession in the WS10 discharge. When the bedrock drains, some of this water is released into the stream channel, whereas the rest moves as catchment scale deep seepage. The water volume entering the stream channel would result in a component that is not measured (bedrock storage) decreasing, whereas a component that is measured (WS10 discharge) remains high. This would lead to an increased mass recovery through time. However, the sheer volume of water unaccounted for precludes storage in the bedrock as the sole explanation. Immediately after the experiment, $85\,983 \pm 53\,469$ l was unaccounted for. If it is assumed that this water is all held in bedrock storage, distributed evenly across the wetted area (172 ± 8 m^2), this would represent $0{\cdot}5 \pm 0{\cdot}3$ m of water, a large amount of water to be stored in low porosity, low-permeability bedrock.

Deep seepage to groundwater and flow in the bedrock aquifer underneath the WS10 gauging station could account for the remaining water missing from the water balance. The continued baseflow throughout the summer, even during prolonged (90+ days) droughts, suggests a significant groundwater system at the HJA. The catchment weirs, although constructed directly on the exposed bedrock in the stream channel, are not thought to be completely water tight. Recent modelling work in other HJA watersheds has suggested that deep seepage is a significant part of the water balance. Waichler *et al*. (2005) modelled the nearby Watersheds 1, 2 and 3 at the HJA using a distributed conceptual model of hillslope processes (distributed hydrology soil vegetation model), and concluded that evapotranspiration could not account for the differences between measured inputs and outflows. The discrepancy was attributed to deep seepage bypass flow around the weirs at the catchment outlet. This bypass was a significant portion, 12%, of the annual water balance, and especially concentrated during the wet, winter months. These estimates are similar to the observed missing water in this experiment after 10 days drainage.

The possibility that 12% of precipitation is bypassing the weirs via deep bedrock flowpaths has considerable implications for catchment scale water transit times. If water held as groundwater is subsequently draining below weirs at watershed outlets, the isotopic and chemical signature of this old water is likely not being expressed in the stream discharge, resulting in an underestimate of water age at this scale. Consequently, reported estimates of a mean transit time of around 1 year at this site are likely skewed as they only consider surface waters (McGuire *et al*., 2005). In addition, the subsurface flow under the WS10 weir would result in persistent underestimate of the flows from the watershed, as suggested by Waichler *et al*. (2005).

Hillslope scale storage–discharge relationship. The hysteretic nature of the soil moisture release curve has been acknowledged for nearly 80 years (Jaynes, 1990). Hysteretic loops also exist in the storage–discharge relationship at the hillslope scale (Kendall *et al*., 1999; Seibert *et al*., 2003; Beven, 2006a; Ewen and Birkinshaw, 2007). This hysteresis, a signal of the non-singular relationship between hillslope storage and hillslope discharge, has been attributed to the connection–disconnection of subsurface saturated areas (Tromp-van Meerveld and McDonnell, 2006) aggregated hysteresis in the core scale soil characteristics (Beven, 2006a), the activation of preferential flow pathways (McDonnell, 1990), and the transition between different flow processes (Ewen and Birkinshaw, 2007).

A counterclockwise hysteretic relationship between storage and lateral subsurface flow was observed during the irrigation experiment (Figure 7a). During wetup, there was a linear relationship between soil moisture and lateral subsurface flow, as both responded quickly to irrigation. After irrigation ceased, lateral subsurface flow quickly returned to pre-irrigation levels (within 50 min), while soil moisture showed a much more gradual recession, resulting in the counterclockwise hysteresis. At the catchment scale, however, no clear hysteretic relationship was observed (Figure 7b), as the slow recession increase in catchment discharge followed the recession of soil moisture, in line with the increases in both during wetup. Although somewhat masked by the strong diel signal seen in the discharge, there appears to be a singular relationship between catchment discharge and hillslope storage. This singular relationship is seen in a plot of daily average catchment discharge versus soil moisture (shown as an insert to Figure 7b) that removes the diel signal in the catchment discharge.

One possible explanation for this behaviour is the transition between vertical and lateral subsurface flow at the soil bedrock interface. During irrigation, the infiltration capacity of the underlying bedrock is quickly reached, and lateral subsurface flow is initiated as infiltration excess flow at the soil bedrock interface. Later, when irrigation ceases, lateral subsurface flow ceases as vertical fluxes through the soil profile drop lower than the infiltration rate of the bedrock. This would result in a quick reduction in lateral subsurface flow as observed at the hillslope during the sprinkler malfunction and after irrigation had ceased, whereas soil moisture would remain elevated as it slowly drains. Catchment discharge, which is a combination of both lateral subsurface flow at the soil bedrock interface (as measured by the hillslope trench) and flow through the bedrock, remains elevated as storage in the soil and bedrock drained, now predominantly vertically in the soil profile and laterally through the bedrock. This would result in the observed singular relationship between storage and stream discharge. The difference in response between the catchment discharge and lateral subsurface flow indicates that soil moisture drainage after the end of irrigation is not contributing to the two fluxes in the same way, and aggregated hysteresis at the core scale cannot solely explain the observed larger scale behaviour.

This interpretation suggests that the observed counterclockwise hysteretic relationship between storage and flow here and elsewhere in the literature could be a measure of the relative contributions of lateral and vertical flow. If catchment or hillslope subsurface flow is dominated by lateral flow through the soil profile, or at the soil bedrock interface, with minimal bedrock leakage at the site of monitoring (either a system underlain by relatively impermeable bedrock or at larger scales, where return flow is significant), then little hysteresis would be expected between soil moisture storage and discharge. On the other hand, a system where leakage is a significant component of the water balance (such as this hillslope, see above), a hysteretic pattern would be expected, as high bedrock infiltration rates are exceeded only during large events or high-intensity rainfall. The observed hysteretic response observed by others at the watershed scale (Beven, 2006a; Ewen and Birkinshaw, 2007) suggests that their watersheds are not watertight, and deep seepage may be a significant component of the water balance.

Measurement uncertainty and the dialog between experimentalist and modeller. Reporting of model uncertainty, whether due to equifinality issues, model parameterization and structure uncertainties (Beven, 2002) or the problems associated with extrapolating models beyond the calibration ranges, has become standard in catchment and hillslope hydrology (Beven, 2006b; Mantovan and Todini, 2006; Andréassian *et al*., 2007; Hall *et al*., 2007; Montanari, 2007; Sivakumar, 2008). While this work is important, the lack of clear and thoughtful analysis of uncertainty in field studies forces the catchment modellers to either estimate on their own the error structure of field data or ignore it entirely. Some attempts to incorporate input uncertainty have been developed, such as the use of fuzzy measures (Bárdossy, 1996; Özelkan and Duckstein, 2001) and soft data (Seibert and McDonnell, 2002). These efforts, however, are still dependent on field experimentalist reports of measurement uncertainty. Nevertheless, uncertainties in flux rates, mass balances, field parameter measurements and other potential inputs into numerical models are rarely reported.

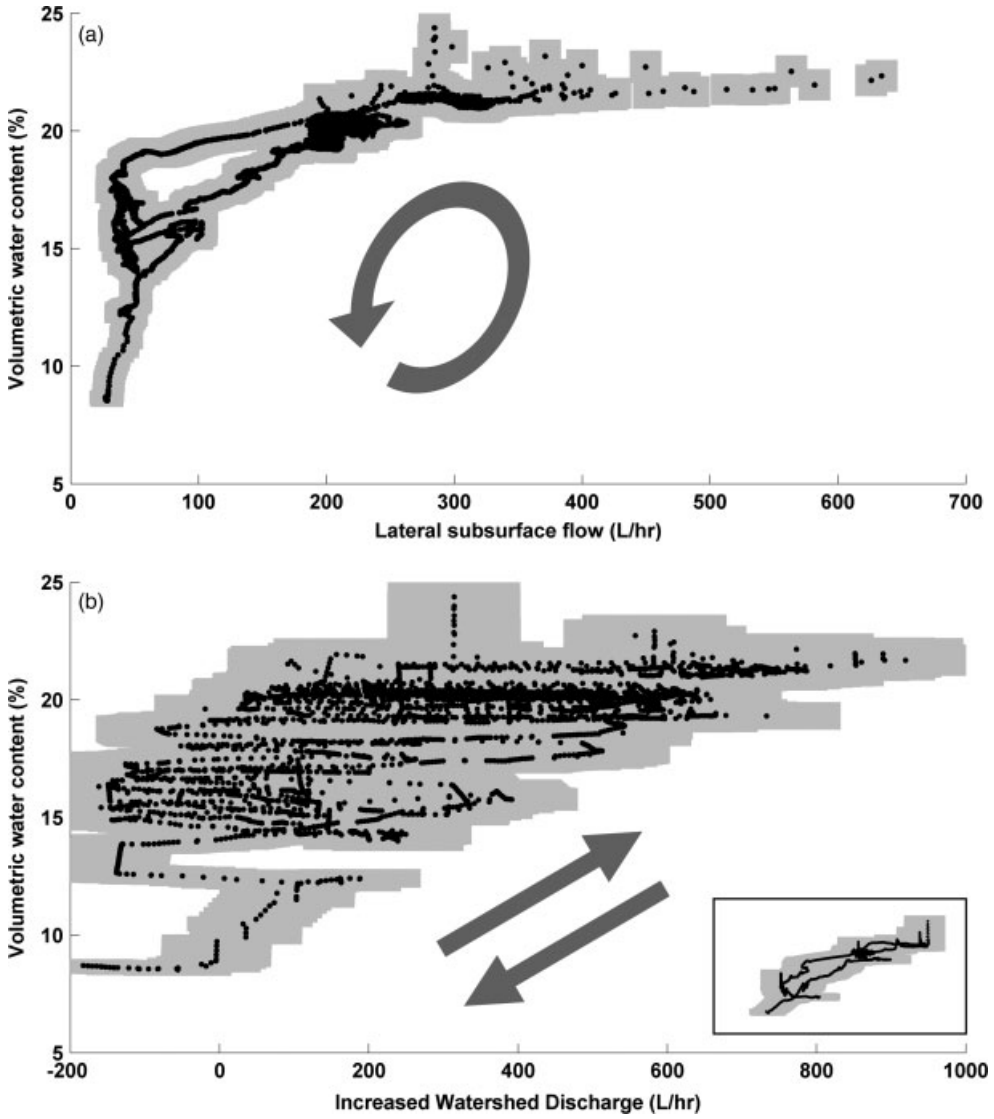

Figure 7. Storage–discharge relations for (a) the hillslope—where hillslope average volumetric water content is plotted against lateral subsurface flow measured at the hillslope trench and (b) the catchment—where water content is plotted against increase in WS10 discharge (inset shows daily average WS discharge vs soil moisture). Uncertainty bounds are shown in grey. As indicated by the grey arrows, the hysteresis is in the counterclockwise direction for the hillslope and a singular relationship for the catchment with marked diurnal variations

After propagation from the various sources of measurement, the uncertainty in many of the water balance components was large, up to 20% of the total component flux for the duration of the experiment. While large, these are due to real uncertainties in the assumptions and measurements used in the calculations. For instance, the uncertainty in the exponent in the WS10 master recession curve alone was 15%. This value was chosen to encompass 80% of the 3-year watershed discharge measurements. To encompass 90%, the increase in the uncertainty of the recession coefficient would have to increase to 50%, with a corresponding increase in overall WS10 discharge uncertainty.

The evaporation measures showed a similarly large uncertainty, up to 43% of the measured value (experiment + 5 days drainage). This high uncertainty is due to two factors: measurement uncertainty and the way we classify the uncertainty. The measurement uncertainty was small for most of the sensors, between 1 and 5% of the readings. For wind speed, however, the instrument accuracy was within ±0·2 m/s. Wind speeds on the hillslope were less than 0·4 m/s for 94% of the monitoring period. This leads to large uncertainty in the evaporation measurements, although the effect was somewhat mitigated by the relative unimportance of wind speed in the final calculations (see Graham (2008) for the wind speed contribution to uncertainty).

The other, larger factor in the uncertainty is the treatment of the errors. As a calibration of the meteorological sensors was not performed in the field, it was impossible to independently determine the accuracy of the sensors. Therefore, the accuracy of the sensors was taken as the factory level uncertainty, and propagated as systematic error. Systematic errors are propagated as the square of sums, rather than the sum of squares. For a given measurement, aggregation of N measurements, the relationship between the random error (on the left) and the

systematic error (on the right in parentheses) is

$$\sum_{n=1}^{N}(\delta x_n)^2 \approx \frac{1}{\sqrt{N}}\left(\sum_{n=1}^{N}\delta x_n\right)^2 \qquad (16)$$

where δx_n is the uncertainty in measurement x_n. For evaporation, where measurements were taken every 15 min, or 96 times daily, 2304 readings were aggregated to determine the total experiment flux. For an instrument with equal systematic and random error, the aggregated systematic error for the duration of the experiment would be $\sqrt{2304}$ or 48 times larger than the random error. While some systematic errors are unavoidable, such as the wetted area measurements or the calculation of the background watershed discharge, elimination or reduction of systematic errors should be the focus of experimental design. Random errors, while still a concern, are shown to be a much smaller component of the uncertainty.

Measurement uncertainty: undermining field hydrology?. Beven (2006b) has been challenged about undermining hydrological science by overemphasizing model uncertainty. Nevertheless, the generally positive response to his article suggests that a rigorous, honest assessment of model uncertainty is considered a positive development by the scientific hydrological community, despite the concerns it may confuse or discourage stakeholders (Mantovan and Todini, 2006; Andréassian *et al*., 2007; Hall *et al*., 2007; Montanari, 2007; Sivakumar, 2008). A similar concern might be raised for a rigorous analysis of uncertainty in experimental hydrology, especially in field campaigns, where measurement and process uncertainties have the potential to be large compared to the measurements. Indeed, in this experiment, the uncertainties in the residual of the water balance are of greater magnitude than the residual itself in some cases (especially 10 days after the experiment), calling into question whether the residual exists at all. Did we measure all the water and not notice it? How can our measured fluxes have a total uncertainty of over 82 000 l (82 m^3), or 20% of the application? How does this impact our conclusions—namely that the system responds quickly; that flow through the bedrock is a significant component at the hillslope scale; and that deep seepage and bypass flow through the bedrock may be a significant component at the watershed scale?

The presented uncertainty impacts some of our conclusions of observed hillslope scale hydrological processes. The dynamics of hillslope and catchment response with respect to storage have little to do with the uncertainty in the measurements, as the uncertainty does not include the possibility that no response occurred. The extent of the observed hysteresis in the hillslope discharge relationship is greater than the uncertainty in the individual measurement. However, the evidence for deep seepage at the catchment scale, and hydrological bypass of the weir is tempered by the uncertainty analysis. The uncertainty in the deep seepage/deep storage (the residual of the water balance) after 10 days drainage encompasses the estimate, which precludes the conclusion that we have strong evidence that deep seepage/deep storage exists. However, the uncertainty also allows for the possibility that deep seepage is a much larger proportion of the water balance (up to 43%) than previously predicted (12%; (Waichler *et al*., 2005)).

The rigorous analysis of the uncertainties allows for the identification of weaknesses in study design and implementation. Although it is unpleasant to identify the weaknesses in one's experiment and to quantify the uncertainties in the results, this analysis allows for a better understanding of the strengths and weaknesses of the work. In this case, the need for on-site calibration of field instrumentation is highlighted, to turn the potentially systematic measurement errors into less significant random errors. The development of better methodology for determining the master recession, perhaps using correlations developed with nearby instrumented catchments, would also serve greatly to reduce the uncertainty in the increase in catchment discharge, and increase the strengths of the observations.

Finally, the explicit presentation of the uncertainty will help in model development and evaluation. As the identification and incorporation of input uncertainty into hydrological models become standard practice, the uncertainties in reference data sets will be required. Without this explicit uncertainty analysis, modellers are often required to either completely trust the data sets that they calibrate their models to or arbitrarily assign an uncertainty bound based on estimates of measurement precision, typical bounds for similar systems, stochastic assignment of errors or the modellers inherent level of trust of the experimentalist. Although these methods are all valid for certain circumstances, a quantitative analysis is likely preferred.

CONCLUSIONS

Our analysis of hillslope and catchment scale deep seepage shows that flow through the near surface bedrock that re-emerges at the stream channel is a significant flowpath at the hillslope scale, nearly equal in volume to the lateral subsurface flow (interflow) in the soil. Deep seepage at the catchment scale (defined as water that enters the groundwater system and does not re-emerge in the stream channel) comprised a large portion of the water balance, averaging approximately 21% of precipitation at steady state. Our analyses of soil moisture and hillslope and catchment discharge dynamics show a common hysteretic relationship seen at many sites. We were able to identify the physical meaning of this behaviour at the hillslope scale via our irrigation work whereby the hysteresis is controlled by the transition between vertical percolation and lateral subsurface flow. At the catchment scale, the hysteretic pattern was not observed when the uncertainties in the soil moisture and watershed discharge were accounted for. This lack of hysteretic relationship

was attributed to the lesser impact of vertical percolation at the catchment scale. We hypothesize that it is likely the permeability of the bedrock (and deep percolation into it) that controls the hysteretic relationship seen in the storage–discharge relationship at other study sites around the world.

ACKNOWLEDGEMENTS

This work was supported through funding from the National Science Foundation (Grant DEB 021-8088 to the Long-Term Ecological Research Program at the H. J. Andrews Experimental Forest) and from the Ecosystem Informatics IGERT program. We thank the helpful comments of three anonymous reviewers, who suggested significant changes to the direction of the article. J. Renee Brooks, Barbara Bond, Kate Lajtha, Bob McKane and Julia Jones provided help in developing the experimental design and research objectives. We thank Matthew Bergen, Aiden Padilla and John Moreau for providing field assistance and John Selker and Sherri Johnson for loan of TDR and meteorological equipment. We especially thank the McKenzie River Ranger District for providing sprinkler water during the experiment, and Kari O'Connell and Cheryl Friesen for coordinating logistics.

REFERENCES

Anderson SP, Dietrich WE, Montgomery DR, Torres R, Conrad ME, Loague K. 1997. Subsurface flow paths in a steep, unchanneled catchment. *Water Resources Research* **33**: 2637–2653.

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.

Bárdossy A. 1996. The use of fuzzy rules for the description of elements of the hydrological cycle. *Ecological Modelling* **85**: 59–65.

Barlow RJ. 1989. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. Wiley: New York.

Barnard HR. 2009. Inter-relationships of vegetation, hydrology and micro-climate in a young, Douglas-fir Forest. PhD thesis, Oregon State University: Corvallis, Oregon. p. 142.

Barnard HR, Graham CB, Van Verseveld W, Brooks JR, Bond BJ, McDonnell JJ. 2010. Examining hillslope transpiration controls of diel sub-surface flow using a steady-state irrigation experiment. *Ecohydrology* **3**: 133–142.

Beven K. 2002. Towards a coherent philosophy for modelling the environment. *Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences* **458**: 2465–2484.

Beven KJ. 2006a. Searching for the Holy Grail of scientific hydrology: $Q_t = (S,R,\Delta t)$ as closure. *Hydrology and Earth System Sciences* **10**: 609–618.

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

Chapman T. 1999. A comparison of algorithms for stream flow recession and baseflow separation. *Hydrological Processes* **13**: 701–714.

Clarke RT. 1999. Uncertainty in the estimation of mean annual flood due to rating-curve indefinition. *Journal of Hydrology* **222**: 185–190.

Domec J-C, Meinzer FC, Gartner BL, Woodruff D. 2006. Transpiration-induced axial and radial tension gradients in trunks of Douglas-fir trees. *Tree Physiology* **26**: 275–284.

Eberhardt LL, Thomas JM. 1991. Designing environmental field studies. *Ecological Monographs* **61**: 53–73.

Ewen J, Birkinshaw SJ. 2007. Lumped hysteretic model for subsurface stormflow developed using downward approach. *Hydrological Processes* **21**: 1496–1505.

Freer J, McDonnell JJ, Beven KJ, Peters NE, Burns DA, Hooper RP, Aulenbach B. 2002. The role of bedrock topography on subsurface storm flow. *Water Resources Research* **38**: DOI: 10.1029/2001WR000872.

Graham CB. 2008. A Macroscale Measurement and Modeling Approach to Improve Understanding of the Hydrology of Steep, Forested Hillslopes. PhD thesis, Oregon State University: Corvallis Oregon. p 174.

Graham CB, Woods RA, McDonnell JJ. 2010. Hillslope threshold response to rainfall: (1) a field based forensic approach. *Journal of Hydrology*. DOI: 10.1016/j.jhydrol.2009.12.015.

Granier A. 1987. Evaluation of transpiration in a Douglas-fir stand by means of sap flow measurements. *Tree Physiology* **3**: 309–320.

Hall J, O'Connell E, Ewen J. 2007. On not undermining the science: coherence, validation and expertise. *Hydrological Processes* **21**: 985–988 [Discussion of Invited Commentary by Keith Beven *Hydrological Processes* **20**: 3141–3146].

Harr RD. 1977. Water flux in soil and subsoil on a steep forested slope. *Journal of Hydrology* **33**: 37–58.

Hornberger GM, Germann PF, Beven KJ. 1991. Throughflow and solute transport in an isolated sloping soil block in a forested catchment. *Journal of Hydrology* **124**: 81–99.

James ME. 1978. *Rock Weathering in the Central Western Cascades*. University of Oregon: Eugene; 119.

Jaynes DB. 1990. Soil water hysteresis: models and implications. In *Process Studies in Hillslope Hydrology*, Anderson MG, Burt TP (eds). Wiley: Chichester; 93–126.

Katsura Sy, Kosugi K, Mizutani T, Okunaka S, Mizuyama T. 2008. Effects of bedrock groundwater on spatial and temporal variations in soil mantle groundwater in a steep granitic headwater catchment. *Water Resources Research* **44**: DOI: 10.1029/2007WR006610.

Katsuyama M, Ohte N, Kabeya N. 2005. Effects of bedrock permeability on hillslope and riparian groundwater dynamics in a weathered granite catchment. *Water Resources Research* **41**: DOI: 10.1029/2004WR003275.

Kendall KA, Shanley JB, McDonnell JJ. 1999. A hydrometric and geochemical approach to test the transmissivity feedback hypothesis during snowmelt. *Journal of Hydrology* **219**: 188–205.

Mantovan P, Todini E. 2006. Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology. *Journal of Hydrology* **330**: 368–381.

McDonnell JJ. 1990. A rationale for old water discharge through macropores in a steep, humid catchment. *Water Resources Research* **26**: 2821–2832.

McDonnell JJ, Sivapalan M, Vache KB, Dunn S, Grant GE, Haggerty R, Hinz C, Hooper RP, Kirchner JW, Roderick ML, Selker JS, Weiler M. 2007. Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. *Water Resources Research* **43**: DOI: 10.1029/2006WR005467.

McDonnell JJ, Tanaka T. 2001. On the future of forest hydrology and biogeochemistry. *Hydrological Processes* **15**(10): 2053–2055.

McGuire KJ, McDonnell JJ, Weiler MH. 2007. Integrating tracer experiments with modeling to infer water transit times. *Advances in Water Resources* **30**: 824–837.

McGuire KJ, McDonnell JJ, Weiler MH, Kendall C, McGlynn BL, Welker JM, Seibert J. 2005. The role of topography on catchment-scale water residence time. *Water Resources Research* **41**: DOI: 10.1029/2004WR003657.

Moncrieff JB, Malhi Y, Leuning R. 1996. The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water. *Global Change Biology* **2**: 231–240.

Montanari A. 2007. What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrological Processes* **21**: 841–845.

Monteith JL, Unsworth MH. 2008. *Principles of Environmental Physics*. Academic Press: London; 418.

Montgomery DR, Dietrich WE, Torres R, Anderson SP, Heffner JT, Loague K. 1997. Hydrologic response of a steep, unchanneled valley to natural and applied rainfall. *Water Resources Research* **33**: 91–109.

Moore GW, Bond BJ, Jones JA, Phillips N, Meinzer FC. 2004. Structural and compositional controls on transpiration in 40- and 450-year-old riparian forests in western Oregon, USA. *Tree Physiology* **24**: 481–491.

Mosley MP. 1979. Streamflow generation in a forested watershed. *Water Resources Research* **15**: 795–806.

Murray FW. 1967. On the computation of saturation vapor pressure. *Journal of Applied Meteorology* **6**: 203–204.

Nyberg L, Rodhe A, Bishop K. 1999. Water transit times and flow paths from two line injections of super(3)H and super(36)Cl in a microcatchment at Gaardsjoen, Sweden. *Hydrological Processes* **13**: 1557–1575.

Onda Y, Komatsu Y, Tsujimura M, Fujihara J-i. 2001. The role of subsurface runoff through bedrock on storm flow generation. *Hydrological Processes* **15**: 1693–1706.

Özelkan EC, Duckstein L. 2001. Fuzzy conceptual rainfall-runoff models. *Journal of Hydrology* **253**: 41–68.

Ranken DW. 1974. *Hydrologic properties of soil and subsoil on a steep, forested slope*. Department of Forest Engineering, Oregon State University: Corvallis; 114.

Rodriguez-Iturbe I. 2000. Ecohydrology: a hydrologic perspective of climate-soil-vegetation dynamics. *Water Resources Research* **36**: 3–9.

Seibert J, Bishop K, Rodhe A, McDonnell JJ. 2003. Groundwater dynamics along a hillslope: a test of the steady state hypothesis. *Water Resources Research* **39**: 1014. DOI: 1010.1029/2002WR001404.

Seibert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resources Research* **38**: DOI: 10.1029/2001WR000978.

Sherlock MD, Chappell NA, McDonnell JJ. 2000. Effects of experimental uncertainty on the calculation of hillslope flow paths. *Hydrological Processes* **14**: 2457–2471.

Sivakumar B. 2008. Undermining the science or undermining Nature? *Hydrological Processes* **22**: 893–897.

Sujono J, Shikasho S, Hiramatsu K. 2004. A comparison of techniques for hydrograph recession analysis. *Hydrological Processes* **18**: 403–413.

Swanson FJ, James ME. 1975. *Geology and geomorphology of the H.J. Andrews Experimental Forest, western Cascades, Oregon*. U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station, Portland, OR; 14.

Tan CS, Black TA. 1976. Factors affecting the canopy resistance of a Douglas-fir forest. *Boundary-Layer Meteorology* **10**: 475–488.

Taylor JR. 1997. *An Introduction to Error Analysis*. University Science Books: Sausalito, California; 328.

Terajima T, Kitahara H, Nakai Y. 1993. Experimental consideration related to the effects of subsurface flow on subsurface hydraulic erosion (I). Relationship between the variety of soil pipe length and the hydraulic action of subsurface flow. *Transaction of the 104th Meeting of the Japanese Forestry Society*: 701–704.

Tetens O. 1930. Uber einige meteorologische begriffe. *Zeitschrift für Geophysik* **6**: 203–204.

Tromp-van Meerveld HJ, McDonnell JJ. 2006. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research* **42**: DOI: 10.1029/2004WR003800.

Tromp-van Meerveld HJ, Peters NE, McDonnell JJ. 2006. Effect of bedrock permeability on subsurface stormflow and the water balance of a trenched hillslope at the Panola Mountain Research Watershed, Georgia, USA. *Hydrological Processes* **21**: 750–769.

Uchida T, Asano Y, Ohte N, Mizuyama T. 2003. Seepage area and rate of bedrock groundwater discharge at a granitic unchanneled hillslope. *Water Resources Research* **39**: DOI: 10.1029/2002WR001298.

van Verseveld WJ, McDonnell JJ, Lajtha K. 2009. The role of hillslope hydrology in controlling nutrient loss. *Journal of Hydrology* **367**: 177–187.

Waichler SR, Wemple BC, Wigmosta MS. 2005. Simulation of water balance and forest treatment effects at the H.J. Andrews Experimental Forest. *Hydrological Processes* **19**: 3177–3199.

Woods R, Rowe L. 1996. The changing spatial variability of subsurface flow across a hillside. *Journal of Hydrology (New Zealand)* **35**: 51–86.

Woods R, Rowe L. 1997. Reply to "Comment on "The changing spatial variability of subsurface flow across a hillside" by Ross Woods and Lindsay Rowe". *Journal of Hydrology (New Zealand)* **36**: 51–86.

