

# The benefits of gravimeter observations for modelling water storage changes at the field scale

B. Creutzfeldt, A. Güntner, S. Vorogushyn, and B. Merz

Deutsches GeoForschungsZentrum GFZ, Section Hydrology, Potsdam, Germany

Received: 14 March 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 1 April 2010

Revised: 16 July 2010 – Accepted: 12 August 2010 – Published: 1 September 2010

**Abstract.** Water storage is the fundamental state variable of hydrological systems. However, comprehensive data on total water storage changes (WSC) are practically inaccessible by hydrological measurement techniques at the field or catchment scale, and hydrological models are highly uncertain in representing the storage term due to the lack of adequate validation or calibration data. In this study, we assess the benefit of temporal gravimeter measurements for modelling WSC at the field scale. A simple conceptual hydrological model is calibrated and evaluated against records of a superconducting gravimeter (SG), soil moisture, and ground-water time series. The model is validated against independently estimated WSC based on lysimeter measurements. Using gravimeter data as a calibration constraint improves the model results substantially in terms of predictive capability and variation of the behavioural model runs. Thanks to their capacity to integrate over different storage components and a larger area, gravimeters provide information on total WSC that can be used to constrain the overall status of the hydrological system in a model. The general problem of specifying the internal model structure or individual parameter sets can, however, not be solved with gravimeters alone.

## 1 Introduction

In hydrology, measuring the water storage term in the *only hydrological equation* (Blöschl, 2005) – the water balance equation – is still a challenging task at all scales. Therefore, catchments are characterised by the output – in general the discharge – using the storage-output relationship to draw conclusions on the storage of an area. However, as Beven (2005) states “... we do not have the investigative

measurement techniques necessary to be secure about what form these (storage-output; note from the author) relationships should take... except by seeing which functions might be appropriate in reproducing the discharges at the catchment outlet (where we can take a measurement).”

Progress in observation techniques has improved the estimation of water storage at various scales. At the global scale, GRACE (Tapley et al., 2004) gives us the unique opportunity to estimate water storage changes (see Ramillien et al. (2008) for a review) and to improve macro-scale hydrological models (Zaitchik et al., 2008; Güntner, 2008; Werth et al., 2009a; Lo et al., 2010). At the field scale, water storage and its changes are generally estimated by point measurements, but high spatial and temporal variability makes the estimation of water storage difficult. Different techniques and strategies have been developed to overcome these problems, e.g., gathering many soil moisture measurements and inter-/extrapolating them by geostatistics (e.g., Western et al., 2002) or ground penetrating radar measurements (GPR; e.g., Huisman et al., 2003; Huisman et al., 2002), the use of spatial TDR soil moisture measurements (e.g., Graeff et al., 2010) or of high-precision lysimeters (e.g., von Unold and Fank, 2008) and the development of cosmic ray neutron probes (Zreda et al., 2008). In general, these techniques are limited to the estimation of near-surface water storage. Neutron probes, electromagnetic sensors in access tubes, electrical resistivity tomography (ERT) or (cross-)borehole geophysics allow for the estimation of water storage in deeper zones, but the temporal as well as the spatial resolution (depth and area) is limited. Further limitations such as high inaccuracies of electromagnetic sensors in access tubes (e.g., Evett et al., 2009) make the estimation of subsurface WSC at the field scale a challenging task, especially for deeper zones.

Ground-based temporal gravity measurements using absolute or relative gravimeters are influenced by local WSC (e.g., Amalvict et al., 2004; Bonatz, 1967; Abe et al., 2006; Crossley et al., 1998; Longuevergne et al., 2009; Kroner



Correspondence to: B. Creutzfeldt  
(benjamin.creutzfeldt@gfz-potsdam.de)

and Jahr, 2006; Van Camp et al., 2006; Bower and Courtier, 1998; Boy and Hinderer, 2006; Meurers et al., 2007; Llubes et al., 2004; Pool and Eychaner, 1995; Naujoks et al., 2008; Jacob et al., 2008). Within the Bouguer approximation, a one meter water mass change in a flat and infinitely extended plate causes a gravity response of  $42 \mu\text{Gal}$ . Focusing on where this gravity response is generated in this layer, the study of Leirião et al. (2009) showed that 90% of the gravity signal comes from a circular disk of a radius 10 times the vertical distance between the layer and the instrument. Topography determines the distribution of hydrological masses in space and influences the relationship of WSC and gravity response. For the Geodetic Observatory Wettzell, for example, distributing the infinitely extended plate along the topography, a water mass change of 1 m causes a gravity change of  $52 \mu\text{Gal}$  (Creutzfeldt et al., 2008). Hence, the effect of WSC on gravity measurements depends on the topography around the gravity sensor and is also a function of the vertical distribution of mass change below the sensor. Different studies showed that local WSC within a radius of 50 to 150 m around the gravimeter are of primary interest for the local hydrological effect on temporal gravity measurements (e.g., Hasan et al., 2008; Van Camp et al., 2006; Hokkanen et al., 2006; Naujoks et al., 2008; Kazama and Okubo, 2009). The gravity time series thus primarily reflect WSC on the field scale, but the exact sampling volume is difficult to define.

Consequently, the question arises: How can we use temporal gravity measurements for hydrological applications? Different studies focus on the interpretation of the gravity signal by single storage components (e.g., surface water (Lampitelli and Francis, 2010; Bonatz and Sperling, 1995), snow (Breili and Pettersen, 2009), soil moisture (Van Camp et al., 2006), or groundwater (Takemoto et al., 2002; Harnisch and Harnisch, 2006)), or by estimation of different subsurface properties (e.g., porosity (Jacob et al., 2009), fractures (Hokkanen et al., 2007), block content (Van Camp et al., 2006) or specific yield (Pool and Eychaner, 1995)). The unambiguous identification of the exact source of the gravimeter signal is difficult or even impossible if no additional information is available implying that the estimation of single parameters on the storages or properties is associated with a high uncertainty (Pool, 2008; Creutzfeldt et al., 2010a). Blainey et al. (2007), for example, pointed out that the estimation of hydraulic conductivity and specific yield by gravity data alone was likely to be unacceptably inaccurate and imprecise.

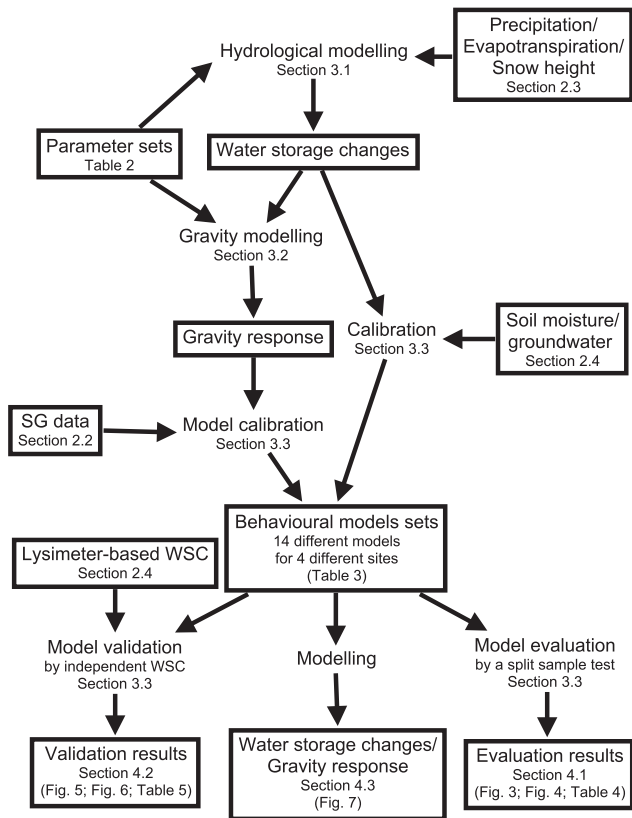
Temporal gravimeter measurements result in an integral signal, integrating over different hydrological storages like snow, soil moisture, and groundwater. Hence, one may adopt a holistic perspective by considering temporal gravimeter measurements as an integral signal of the hydrological system status similar in nature to discharge measurements (Hasan et al., 2008). More precisely, temporal gravity data can be a direct measure of the change of the system status – the change of water storage – whereas discharge is a measure

for the catchment response. The latter requires assumptions about the storage-output relationship to characterise the system status.

In the absence of adequate observation data, the only and frequently used alternative to comprehensively characterise the hydrological system status is by applying hydrological models. Many different hydrological models have been developed ranging from simple, lumped, and conceptual models to complex, distributed, and physically-based ones. Typically, measured input fluxes are used to drive a model. The model parameters are calibrated to match the observed output fluxes, usually river discharge. This approach leaves the model with considerable uncertainty in representing the status of a complex hydrological system because the relationship between the system response and its status may not be unique (hysteresis, e.g., Spence et al., 2010) and/or many different parameter sets may result in similar system responses (equifinality problem after Beven and Binley, 1992).

In this study, instead of calibrating a hydrological model against output fluxes, we use information about the change of the integral system status for model calibration. The aim is to investigate the benefits of temporal gravimeter measurements for hydrological modeling as an integrative measure of the water storage term. Different strategies exist to parameterise/calibrate a hydrological model with geophysical measurements. Frequently, geophysical data are integrated into a hydrological model by inverting the geophysical data to estimate the spatial distribution of geophysical properties. Hydrological quantities are then derived from the estimated geophysical properties and the hydrological model is parameterised/calibrated based on these quantities (uncoupled hydrogeophysical inversion). Contrary to that, a coupled hydrogeophysical inversion framework, as summarised by Ferré et al. (2009), directly infers hydrological quantities from geophysical measurements. Geophysical data are interpreted for hydrological research by coupling hydrological and geophysical models during inversion (Hinnell et al., 2010; Rings et al., 2010; Rucker, 2009).

For this study, this means in practice that we use (1) a hydrological model with different parameter sets to calculate the WSC, (2) a geophysical model to calculate how the gravimeter responds to these WSC, and (3) the SG data to assess the parameter set by comparing them to the modelled gravity response. The value of gravimeter observations is assessed in comparison to classical hydrological point measurements (groundwater and soil moisture) using the different data sets as calibration constraints. We apply a simple conceptual model that comprises a set of connected linear storages with a limited number of free parameters as a typical example of hydrological models. The results of the different calibrated models are evaluated and also validated by independent WSC. The concept and structure of this study is outlined in Fig. 1.



**Fig. 1.** Concept map (Cañas et al., 2005) showing objects (boxes) and processes of this study in combination with the study structure.

## 2 Study area and data

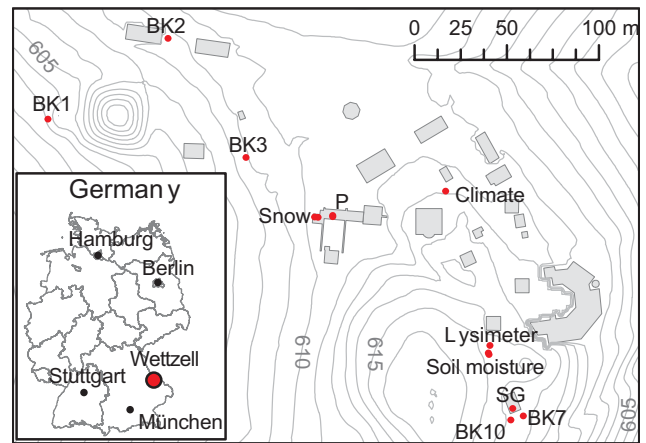
### 2.1 Study area

The study area is located in the Bavarian Forest, a mid mountain range in the Southeast of Germany (Fig. 2). The area is characterised by flat highlands with grassland and fields and steep long slopes dominated by forestry. The study area surrounds the Geodetic Observatory Wettzell operated by the Federal Agency for Cartography and Geodesy (Schlüter et al., 2007).

The observatory is mainly surrounded by grassland with single bushes. The geology is made up of gneiss, and the bedrock seamlessly merges into the weathered saprolite layer. Creutzfeldt et al. (2010a) classified the underground of the gravimeter surrounding into the following four different zones: (1) soil zone with mainly loamy-sandy brown soils (Cambisols), (2) saprolite zone consisting of grus (weathered gneiss), (3) fractured zone, and (4) the basement zone.

### 2.2 Gravity data

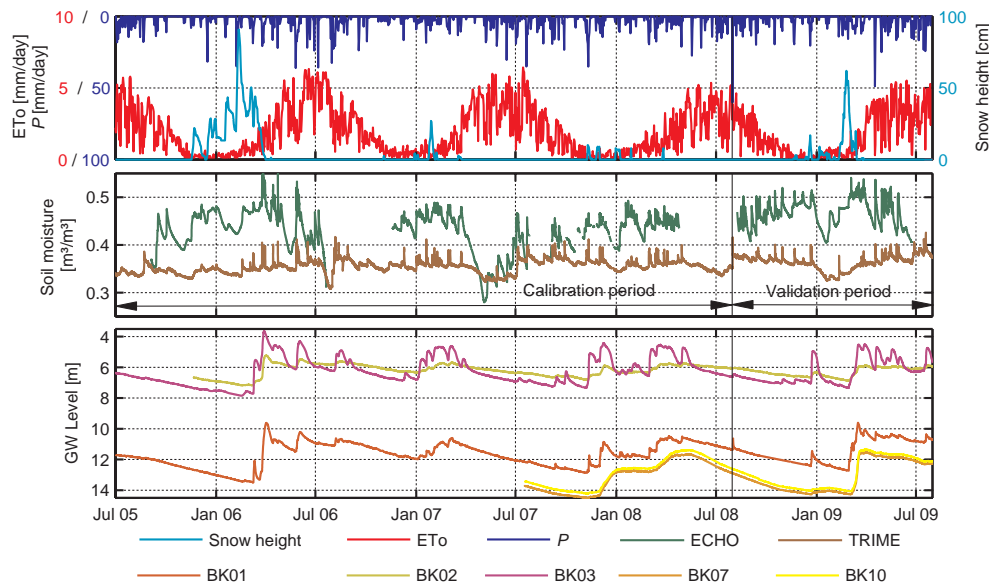
The dual-sphere SG CD029 of the Geodetic Observatory Wettzell, which is part of the Global Geodynamics Project (GGP) network (Crossley et al., 1999; Crossley and Hin-



**Fig. 2.** Study area with the different hydrological sensors and the topography represented by contour lines (distance 1 m). Location of Wettzell in Germany and some major cities are displayed in the inset map.

derer, 2009), measures the temporal variation of the Earth's gravity field. The scale factor and the instrumental drift of the SG were determined by absolute gravity measurements (Wziontek et al., 2009a). Temporal variations of the Earth's gravity field are mainly influenced by tides of the solid Earth, ocean loading effects, mass changes in the atmosphere and polar motion. These gravity effects have to be removed to reveal the hydrological signal in gravimeter measurements. A tidal analysis was performed to remove the solid Earth tides and ocean loading effects. Atmospheric effects were removed by three-dimensional modelling of atmospheric mass changes (Klügel and Wziontek, 2009). The pole coordinates as provided by the International Earth Rotation and Reference Systems Service (IERS) were used to calculate the effect of polar motion. For details on the SG instrument and data processing, the interested reader is referred to Hinderer et al. (2007).

SG residuals were derived by removing the different gravity effects from the SG signal. These residuals are considered to be caused by hydrological mass variations because all other possible effects on temporal gravimeter measurements are assumed to be negligible for Wettzell (e.g., post-glacial rebound or processes in the Earth's mantle and core). For the SG Wettzell, Creutzfeldt et al. (2008) showed that between 52% and 80% of the local hydrological gravity signal is generated within a radius of 50 m around the SG, and 90% of the signal comes from an area within a radius of around 1000 m. A high correlation of independently estimated WSC in this area and SG residuals (coefficient of determination: 0.97; corresponding slope: 1.06) proved that a major part of the gravity residuals is generated by WSC in this area (Creutzfeldt et al., 2010b). In the present study, the large-scale hydrological effect on gravimeters (e.g., Llubes et al., 2004; Weise et al., 2009) is not considered due to



**Fig. 3.** Time series of input and calibration data. Time series of daily precipitation ( $P$ ), daily reference evapotranspiration ( $ET_o$ ) and snow height (top). Time series of soil moisture (middle) and groundwater data (bottom).

the dominant local hydrological influence and high uncertainties in the modelling of large-scale WSC (e.g., Werth et al., 2009b). The SG signal was not corrected for the global hydrological effect because different hydrological models show that the estimated gravity effect due to large-scale WSC lies in the same order of magnitude as differences between different models (Neumeier et al., 2008; Wziontek et al., 2009b).

### 2.3 Meteorological data

Meteorological data – air temperature, relative humidity, wind speed, and global radiation – were recorded at the Geodetic Observatory Wettzell (Table 1 and Fig. 2) and were processed to hourly time series for the whole study period from 1 July 2005 to 31 July 2009. A few data gaps in the time series were filled using the nearby climate station Allmannsdorf at a distance of 6 km and an altitude of 557 m (LfL, 2009). The reference evapotranspiration for short grass canopy was calculated from the climate data based on the Penman-Monteith equation of the American Society of Civil Engineering (Allen et al., 2005).

Precipitation was measured by two heated tipping bucket rain gauges (Fig. 2, Table 1). The differences of the total precipitation were less than 1% for both gauges for the study period, so the mean of both gauges was used for further analysis. The precipitation was corrected for wind effect and wetting losses, acknowledging the well-known undercatch of unshielded heated tipping gauges (Allerup, 1997; Richter, 1995).

Since August 2007, a snow monitoring system consisting of a snow pillow and an ultrasonic snow depth sensor is

**Table 1.** Measured variable and the corresponding devices/sensors at the Geodetic Observatory Wettzell.

Measured variable	Unit	Sensor/Device
Relative gravity	$\mu\text{Gal}$	GWR SG CD029
Absolute gravity	$\mu\text{Gal}$	Micro-g LaCoste FG5
Wind direction	degrees	Lambrecht 14512 G3
Wind speed	$\text{m s}^{-1}$	Lambrecht 14512 G3
Air temperature	$^{\circ}\text{C}$	Lambrecht 809 MU
Relative air humidity	%	Lambrecht 809 MU
Precipitation	mm	L-Tec tipping gauge
Global radiation	$\text{W/m}^2$	Kipp&Zonen CM11
Snow depth	mm	Sommer USH-8
Snow water equivalent	mm	Sommer snow pillow
Soil moisture (TRIME)	$\text{m}^3/\text{m}^3$	IMKO TRIME-EZ
Soil moisture (ECHO)	$\text{m}^3/\text{m}^3$	Decagon EC-10
Groundwater head	m	SEBA MDS-Dipper-3
Groundwater head	m	SEBA MDS-Dipper-Tec

measuring snow depth and the snow water equivalent (SWE) (Fig. 2, Table 1). Before the installation, the snow depth was derived from two snow depth gauge stations close to the observatory (Prackebach-Moosbach: distance 8 km, altitude 505 m; Viechtach-Bühling: distance 3 km, altitude 662 m). Fig. 3 shows the records of precipitation, reference evapotranspiration and snow height.

## 2.4 Hydrological/water storage data

For the whole study period from 1 July 2005 to 31 July 2009, soil moisture and groundwater data were recorded. Soil moisture was measured with a capacitance (ECHO) and ‘pseudo TDR’ sensor (TRIME) at a depth of 0.5 m. Groundwater level data were available from 3 different boreholes (BK1, BK2, BK3) using a relative pressure transducer. Since mid-2007, additional boreholes have been drilled and equipped with multi-parameter sensors, but of these, only boreholes BK7 and BK10 close to the SG are relevant for this study (Figs. 2 and 3, Table 1).

For the period from 30 July 2008 to 30 July 2009, independently estimated WSC were available from Creutzfeldt et al. (2010b). In this study, WSC were derived from lysimeter measurements in combination with complementary hydrological observations and a hydrological 1-D model for the SG site. WSC up to a depth of 1.5 m, precipitation, actual evapotranspiration and deep drainage were estimated by a monolith-filled, suction-controlled, and weighable lysimeter (von Unold and Fank, 2008). WSC below the lysimeter in the deep vadose zones, the saprolite zone, and in the groundwater were estimated by the deep drainage of the lysimeter and the groundwater level. Water redistribution in the saprolite and groundwater zone and the groundwater discharge were calculated using the physically-based hydrological model HYDRUS 1-D (Šimůnek et al., 2008). The underground was classified into the saprolite (thickness 9.5 m) and the fractured (thickness 4.5 m) zone and was parameterised based on measurements of water retention and of saturated hydraulic conductivity and on pump tests (Creutzfeldt et al., 2010a). Deep drainage measurements from the lysimeter were used to define the upper boundary flux of the hydrological model. The lower boundary was defined by the groundwater level (BK07 and BK10) as variable head conditions. This approach developed for the SG site was transferred to the other groundwater sites (BK1, BK2, BK3). The underground model was adjusted and the corresponding groundwater level data were used as the lower model boundary. The underground was classified based on the cores from the corresponding boreholes. At the BK1 site, the thickness of the saprolite (9.0 m) and fractured (3.0 m) zone was comparable to the SG site. For the BK2 and BK3 sites, the thickness of the saprolite zone was only 4.5 m and 3 m respectively. The fractured zone thickness was estimated to be 5.5 m for BK2 and 3.0 m for BK3. Finally, for the four sites, namely SG, BK1, BK2 and BK3, WSC were estimated for the period from 30 July 2008 to 30 July 2009 (Fig. 2).

In order to quantify near-surface hydrological flux and storage processes, lysimeters are considered to be a very accurate method (e.g., Tolk and Evett, 2009; Howell, 2004; Yang et al., 2000). In combination with a well-constrained physically-based model and complementary data for the deeper zones, this suggests that the derived WSC are as close as we can get to reality nowadays in terms of estimating

total WSC. In this context, we assume that the estimated WSC from the multi-method and multi-site approach presented above can henceforth serve as validation data at the field scale (hereinafter referred to as “measured WSC”).

## 3 Modelling

### 3.1 Hydrological modelling

For the estimation of WSC, a simple conceptual hydrological model was set up with the prerequisite to account for both, parameter parsimony and adequate representation of hydrological processes. On the one hand, the model should be as simple as possible with a few parameters only. On the other hand, the model must represent the different hydrological storage components and water fluxes between them, because the gravity response depends on where and in which storage WSC occur in relation to the gravimeter (Creutzfeldt et al., 2008). As a simplifying assumption to approximate the complex and open hydrological system, we consider water storages to vary over depth, neglecting the lateral variability of water storages. This assumption was motivated by the fact that at the scale relevant for the gravimeter, the variability of WSC over depth is much more important than the lateral variability of WSC. This is given because water storages are controlled by the driving processes like infiltration, evaporation, plant water uptake, deep drainage, groundwater recharge or groundwater discharge, as well as by internal properties of the system such as soil hydraulic properties or macropores. At the scale relevant for the gravimeter, these first order controls of water storages differ significantly over depth, whereas a lateral continuity is given for most of the processes and landscape features.

The model is based on the HBV model (Bergström, 1992; Seibert, 2005) but has been adopted and modified to reflect storages and fundamental mechanisms of the study area. Based on the underground classification, the model considers WSC in the snow, soil, saprolite and groundwater storage. It consists of a snow (SS), top soil (SM) (depth: 0.0–0.5 m), soil ( $V_{\text{Soil}}$ ) (depth: 0.0–1.0 m), saprolite ( $V_{\text{Saprolite}}$ ) (depth: 1.0–3.0 m), and groundwater module ( $V_{\text{GW}}$ ) (depth: 13.0–16.0 m). The model uses hourly precipitation, reference evapotranspiration and snow height as input data and estimates the WSC in the different storages.

The snow water equivalent was computed based on the snow depth and precipitation data. During periods with a snow depth greater than zero, we assumed that all precipitation had fallen as snow (SIn). We also assumed that a decline of snow depth was caused only by snowmelt (SOut) neglecting snow compaction. The snowmelt amount was proportionally estimated in relation to the snow depth decline. For each time step, the snow storage is

$$SS_{(t)} = SS_{(t_0)} + SIn_{(t)} - SOut_{(t)} \quad (1)$$

where  $t_0$  is the time step preceding  $t$ . Precipitation and snowmelt ( $P$ ) were divided into the top soil and soil module based on the factor alpha. The top soil storage was a simple bucket storage with a maximum storage capacity of FC. The top soil moisture was calculated as

$$SM_{(t)} = SM_{(t_0)} + P_{(t)} \times (1 - \alpha) - ETa_{(t)} \quad (2)$$

Excess water ( $q_{\text{excess}}$ ) was directly routed into the soil storage  $V_{\text{Soil}}$ . The actual evapotranspiration (ETa) was calculated based on the reference evapotranspiration (ETo) as follows

$$ETa_{(t)} = \begin{cases} ETo_{(t)} & \text{for } \frac{SM_{(t_0)}}{FC} > LP \\ ETo_{(t)} \frac{SM_{(t_0)}}{FC \cdot LP} & \text{for } \frac{SM_{(t_0)}}{FC} \leq LP \end{cases} \quad (3)$$

where LP is the threshold reducing the reference evapotranspiration depending on the soil moisture. The input into the soil storage was determined by

$$I_{(t)} = P_{(t)} \cdot \alpha + q_{\text{excess}} \quad (4)$$

and the outflow by

$$Q_{\text{Soil}(t)} = Q_{\text{Soil}(t_0)} e^{-(t-t_0)/k_{\text{Soil}}} + I_{(t)} \left[ 1 - e^{-(t-t_0)/k_{\text{Soil}}} \right] \quad (5)$$

where  $k_{\text{Soil}}$  is the storage coefficient of the soil storage [h]. The soil storage ( $V_{\text{Soil}}$ ) was calculated by

$$V_{\text{Soil}(t)} = k_{\text{Soil}} \cdot Q_{\text{Soil}(t)}. \quad (6)$$

The outflow and water storage of the saprolite ( $V_{\text{Saprolite}}$ ) and groundwater ( $V_{\text{GW}}$ ) were estimated analogously to Eqs. (5) and (6) using the outflow of the upper storage as input.

Three model parameters represented the interaction of atmosphere and soil (FC, LP, and alpha). The other three parameters controlled the water storage in soil, saprolite and groundwater ( $k_{\text{Soil}}$ ,  $k_{\text{Saprolite}}$ , and  $k_{\text{GW}}$ ). A multiplication factor for precipitation correction (Pcorr) was introduced to account for possible differences of precipitation measured by the tipping bucket rain gauge and lysimeter.

### 3.2 Gravity modelling

The gravity response was calculated based on a geophysical model presented by Creutzfeldt et al. (2008) for a square with a side length of 4 km and the SG located in its centre. In this approach, a spatially nested discretisation domain was developed. A high-precision DEM was used to distribute the estimated WSC along the topography and to discretise the continuous landscape into elementary bodies. For each elementary body the gravity effect was calculated based on a modified point mass equation (MacMillan, 1958; Leirião et al., 2009). The gravity response for each different storage component was derived by summation of all gravity changes in each elementary body in the corresponding storage zone of the model domain. By doing this, we derived a “WSC to

gravity response conversion factor” for each storage component.

The surrounding and subsurface structures in the vicinity of the gravimeter have a major influence on the relationship between WSC and gravity residuals. However, we do not exactly know what happens below the gravimeter building, which prevents infiltration of water into the soil (umbrella effect). Hence, uncertainties arise for the physical solution of the forward problem. For each storage component, we estimated the physically possible upper and lower bounds of the “WSC to gravity response conversion factor” to take into account these uncertainties. Therefore, we looked at both possibilities in that we first calculated the gravity effect assuming that WSC can occur below the gravimeter building and then, as a second possibility, excluded mass variations below the base plate. These uncertainties only apply for the storage components SM,  $V_{\text{Soil}}$  and  $V_{\text{Saprolite}}$ , because snow accumulates on the roof of the SG building and the free gravity-driven groundwater flow is not affected by the SG building. Furthermore, we assumed for SM and  $V_{\text{Soil}}$  storages that WSC occur neither in the concrete foundation nor in the base plate of the SG building (Creutzfeldt et al., 2008). This implies that three additional parameters have to be estimated to derive the gravity response from WSC in the SM,  $V_{\text{Soil}}$  and  $V_{\text{Saprolite}}$  storage (Table 2). These parameters were considered to also account for the precipitation redistribution from the SG roof to the drainage tank at a distance of  $\sim 20$  m from the SG.

### 3.3 Assessment of model performance

We distinguished between calibration, evaluation and validation process. The automated calibration of the hydrological model was based on the Generalized Likelihood Uncertainty Estimation (GLUE) method developed by Beven and Binley (1992). 50 000 Monte Carlo runs were performed with different parameter sets. The parameter sets were sampled assuming uniform distribution between the lower and upper bounds. For the initial model runs, the parameter range was chosen based on previous studies (Seibert, 1996; Merz et al., 2009), but the range was adjusted in such a way that the parameters for the behavioural model runs were limited only by physical properties.

In the GLUE approach, the definition of behavioural model runs is based on a threshold value for the performance indices. Here, the correlation coefficient ( $R$ ) was used as a performance index of the relative temporal dynamics in the simulated time series. Using  $R$  avoids the need to get absolute water storage data from the observations by deriving the specific yield of the aquifer and estimating the field capacity of the soil or calibrating the soil moisture sensors. In this study, we defined the top 0.1% of the model runs as behavioural model runs. This allows for a better quantitative comparison of the different calibrated models (Juston et al., 2009).

**Table 2.** Parameters of the hydrological model and for the estimation of the gravity response and their lower and upper bounds.

Parameter	Description	Unit	Min.	Max.
Hydrological parameters				
Pcorr	Precipitation correction parameter	–	0.9	1.15
FC	Field capacity	mm	50	300
LP	Threshold for reduction of ETo	–	0.5	1.0
alpha	Shape coefficient	–	0	0.2
$k_{Soil}$	Recession coefficient for top soil storage	h	0	3000
$k_{Saprolite}$	Recession coefficient for saprolite storage	h	50	5500
$k_{GW}$	Recession coefficient for groundwater	h	50	5500
Gravity response parameters				
$g_{SWE}$	Factor for snow gravity response	$\mu Gal$	$-3.9 \times 10^{-6} \times SWE^2 - 0.0009 \times SWE$	
$g_{SM}$	Factor for top soil gravity response	$\mu Gal$	0.026	0.015
$g_{Soil}$	Factor for soil gravity response	$\mu Gal$	0.035	0.017
$g_{Saprolite}$	Factor for saprolite gravity response	$\mu Gal$	0.050	0.014
$g_{GW}$	Factor for groundwater gravity response	$\mu Gal$		0.049

**Table 3.** The different calibrated models based on the different data sources and the site type.

Models	Data set						Site type
	SG	Groundwater			Soil moisture		
		BK1	BK2	BK3	ECHO	TRIME	
SG	x						Deep vadose zone
SGECHO	x				x		
SGTRIME	x					x	
BK1		x					
BK1ECHO		x			x		
BK1TRIME		x				x	
BK2			x				Shallow vadose zone
BK2ECHO			x		x		
BK2TRIME			x			x	
BK3				x			
BK3ECHO				x	x		
BK3TRIME				x		x	
ECHO					x		
TRIME						x	

The performance of each single model run was evaluated by comparing modelled to measured data. First, we compared the modelled gravity response to the SG residuals. Second, the performance of each model run was evaluated by comparing WSC in the SM storage and the  $V_{GW}$  storage to the soil moisture (ECHO, TRIME) and groundwater measurements (BK1, BK2, BK3). Third, a multi-criteria calibration was performed based on soil moisture and groundwater head on the one hand and soil moisture and gravity data on the other. The mean of the different performance indices was used to allow a direct comparison to single-criteria calibrated

models. In total, 14 different calibrated models were derived (Table 3). Each of these models consisted of 50 behavioural model runs.

The model performance was tested using two different strategies: evaluation and validation of the model. For model evaluation, we applied a split-sample test according to Klemes (1986). The record was split into two parts of equal duration from 1 July 2005 to 31 December 2006 and from 1 January 2007 to 1 July 2008. The model was calibrated for the first period and the performance was evaluated using the data from the second period. Then, the periods were swapped and the model was calibrated for the second period and evaluated for the first period. A warm-up period of 2 years was used prior to every simulation. Finally, the model performance was evaluated by comparing the performance indices for the different calibration/evaluation periods. The model can be considered acceptable if the model performs similarly well for both periods.

The split-sample test is a classical hydrological model test, which is a necessary rather than a sufficient testing scheme allowing to assess the capability of the model to make accurate predictions also for periods outside the calibration period (Refsgaard and Knudsen, 1996). Using SG as calibration constraint, this test can prove the model adequacy to represent the SG residuals also outside the calibration period. However, temporal gravity data do not directly measure the WSC in mm, but express the influence of WSC in change of gravity. Hence, a second strategy for testing the model with independent data, the model validation, was implemented. The WSC from the lysimeter approach for the different sites (see Sect. 2.4) were used as the validation data. Based on the available data, the study period was divided into a calibration period (from 1 July 2005 to 30 July 2008) and a validation period (from 30 July 2008 to 31 July 2009).

**Table 4.** Range of performance indices for the different calibration/evaluation periods for the models calibrated against the different data sources.

	Calibration	Evaluation	Calibration	Evaluation
Period	2005–2006	2007–2008	2007–2008	2005–2006
SG	0.98–0.99	0.94–0.99	0.99–0.99	0.84–0.98
BK1	0.90–0.92	0.53–0.75	0.83–0.86	0.55–0.76
BK2	0.91–0.94	0.42–0.63	0.80–0.81	0.31–0.71
BK3	0.86–0.88	0.46–0.71	0.67–0.81	0.56–0.89
ECHO	0.80–0.81	0.76–0.82	0.88–0.89	0.81–0.81
TRIME	0.51–0.52	0.53–0.60	0.53–0.60	0.51–0.52
SGECHO	0.87–0.88	0.76–0.93	0.95–0.95	0.71–0.80
BK1ECHO	0.84–0.85	0.63–0.82	0.87–0.88	0.67–0.77
BK2ECHO	0.84–0.85	0.57–0.77	0.83–0.84	0.51–0.75
BK3ECHO	0.81–0.83	0.63–0.81	0.79–0.84	0.60–0.82
SGTRIME	0.77–0.79	0.72–0.81	0.87–0.87	0.64–0.70
BK1TRIME	0.72–0.80	0.50–0.74	0.77–0.78	0.54–0.66
BK2TRIME	0.72–0.79	0.49–0.69	0.73–0.76	0.40–0.65
BK3TRIME	0.70–0.80	0.51–0.73	0.69–0.74	0.52–0.68

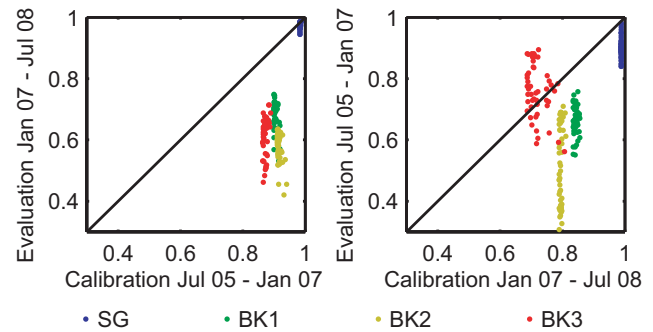
Different models were calibrated against gravimeter, groundwater and/or soil moisture data as described above. The models were validated by independently measured WSC. The modelled and measured results were compared using  $R$ , the standard deviation and the centred root-mean-square differences (RMSD) (Taylor, 2001). The model validation with independent data allows gaining credibility in the novel measurement method to serve as calibration/validation data for hydrological modelling.

Finally, the modelled hydrological gravity response (SG model) is compared to the SG residuals for the whole study period. For this comparison, the Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970) was used as a performance index to constrain the “WSC to gravity response conversion factors”.

## 4 Results and discussion

### 4.1 Model evaluation

Focusing on the performance of the behavioural model sets (top 0.1% simulations) during the calibration/evaluation period, differences between the models calibrated against different data sources could be identified. For the calibration period, using groundwater and/or soil moisture as calibration constraints, the maximum achieved performance indices were smaller and the range of the performance indices was larger than for models calibrated against SG data (Table 4). As a reason for this, one could argue that more parameters are available to match the SG record than to fit the model to the other observation data. The larger number of degrees of freedom may cause better calibration performance. This could be true for soil moisture where only four parameters can be calibrated to match the observation. But all model pa-



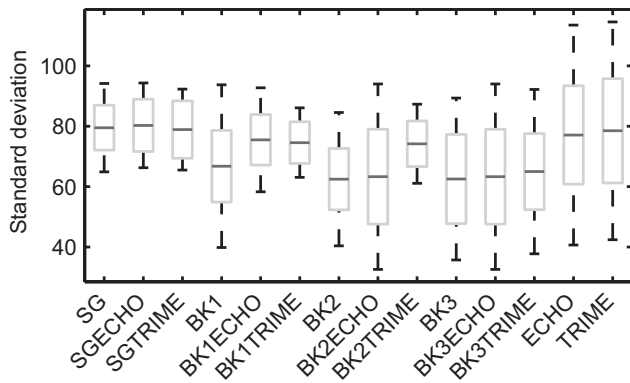
**Fig. 4.** Model performance of four models for the calibration versus the evaluation periods. Here,  $R$  is used as a performance index.

rameters influence the groundwater part of the model because  $V_{GW}$  is the last component in the storage cascade. Hence, the same parameter amount is available to fit the model to the groundwater record as to match the SG observations. The three “WSC to gravity response conversion factors” are of minor importance for the temporal reproduction of the SG residuals because they only influence the amplitude of the signal.

Figure 4 and Table 4 summarise the comparison of the model performance during the calibration versus the evaluation periods. The differences of the performance index between the calibration and evaluation period are higher for models calibrated against groundwater or soil moisture than for models using SG data as calibration constraint. This pattern is persistent also for the multi-objective calibrated models. The model evaluation shows that the model predicts the temporal behaviour of the SG residuals in a better way than the temporal variation of the groundwater or soil moisture for the calibration and evaluation period. For the independent evaluation period, the model performance for groundwater and/or soil moisture models deteriorates more than for SG models.

One explanation for the difference in model performance and predictive capability is that point measurements of WSC are a product of complex processes such as preferential flow, root water uptake, soil freezing/thawing or lateral flow. Furthermore, WSC vary in space due to spatial heterogeneity of landscape features. Hence, differences of modelled and measured records exist because these detailed processes or the spatial variability of WSC could not be represented by the generalised and simplified conceptual model. Since they integrate over different storages and a larger area, SG measurements can resolve neither the detailed and complex processes nor the high spatial variability of WSC. SGs capture a generalised and simplified signal, which is in accordance with the nature of conceptual models. Not surprisingly, the performance and predictive capability is better for the generalised and simplified signal than for a complex and variable signal.





**Fig. 5.** Performance of the different models for the calibration period. The median of the box plot is a measure of the signal amplitude (the standard deviation of the mean signal). The box and whiskers represent the scattering of the standard deviation of the behavioural model runs computed at each time step (the standard deviation of the mean signal was added to the standard deviation of the behavioural model runs at each time step).

Due to the integral character of SG measurements, it remains difficult to make statements about internal model structures or to differentiate between single parameter sets. Soil moisture and/or groundwater data permit the evaluation of internal model components. For example, the model performance of BK2 reaches up to 0.94 for one calibration period, whereas the maximum model performance is only 0.81 for the other calibration. Soil moisture measurements are another example. For one soil moisture sensor, the model performance is as high as 0.89, whereas for the other sensor, the maximum performance index was only 0.60. Still, it remains difficult to evaluate whether the differences are due to parameterisation problems, structural model errors or spatial variability (neglecting observation data errors). SG data, on the contrary, permit the evaluation of the total model because they represent the water storage status instead of evaluating single model parameters or the internal structure.

The model evaluation shows that using SG data as calibration constraints improves the model performance and the predictive capability. In this context, SG measurements can substantially improve the evaluation of the model results. Nonetheless, different parameter sets can give the identical fit to the calibration data, raising the issue of getting the right answers for the wrong reasons.

#### 4.2 Model validation

The model was calibrated for the period from 1 July 2005 to 30 July 2008 using again SG, groundwater and soil moisture data to constrain model parameters. For the calibration period, the performance of each model was assessed in terms of the variation of the behavioural model runs. Here, this variation is expressed as the standard deviation of the behavioural model runs computed at each time step. In gen-

eral, the variation of the behavioural model runs correlates positively with the signal amplitude. Here, the signal amplitude is expressed as the standard deviation of the mean time series. Fig. 5 summarises the performance of the different models in a box plot for the calibration period. Using SG data to constrain the model parameters reveals that the variation of the behavioural model runs is relatively small in comparison to the other groundwater/soil moisture data. Neither variation nor amplitude change when additional information is included.

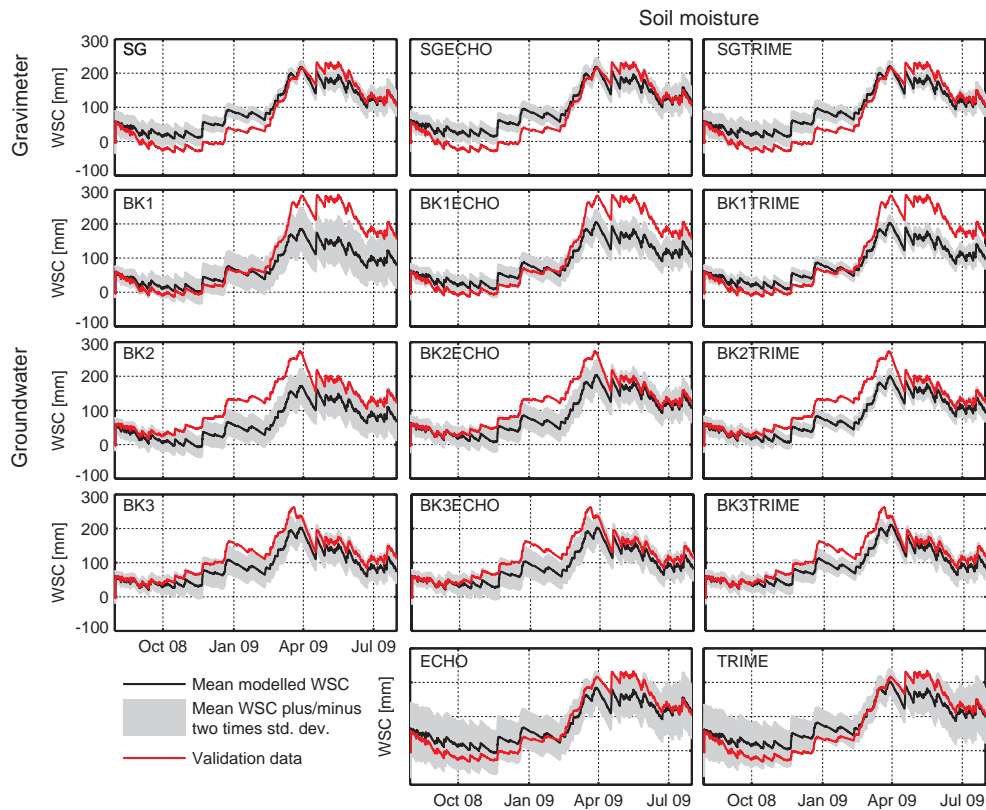
The model based on BK1 data shows that including soil moisture data into the calibration process reduces the variation of the model runs and increases the total amplitude. For the BK2 model, soil moisture data can increase (BK2ECHO) or decrease (BK2TRIME) the variation of the behavioural model runs. Soil moisture data do not affect the variation or amplitude of the BK3 model significantly, but the scattering of the behavioural model runs is relatively large. The behavioural model runs show the maximum variation for models using soil moisture data as the only calibration constraint.

SG data can characterise the whole hydrological system because the inclusion of additional data does not change the model results in terms of variation of the behavioural model runs and total signal amplitude. In contrast to this, soil moisture or groundwater data can be used to calibrate single model components directly, whereas including additional data can have a significant effect on the model results.

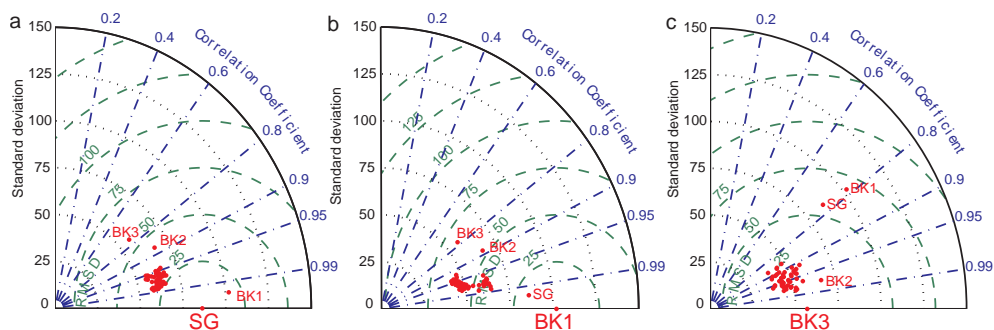
Focusing on the validation data in Fig. 6, two different site types can be distinguished (Table 3). The seasonal amplitude of WSC of the sites SG and BK1 is larger than that of sites BK2 and BK3, something which is also reflected by the standard deviation of the measured time series which amounts to 86 mm for the SG site and to 102 mm for the BK1 site, but is only as high as 67 mm and 57 mm for BK2 and BK3 respectively. The two site types differ not only in the seasonal amplitude but also in temporal dynamics. At the sites SG and BK1, we can identify a later and stronger increase of water storage during the snowmelt event from February to March 2009, whereas the recession of water storages is faster for the sites BK2 and BK3 (Fig. 6). These differences are caused by the varying thickness of the vadose zone at the different sites. For BK2 and BK3, the groundwater depth varies between 4 and 8 m, whereas for SG and BK1, the groundwater depth amounts to up to 14.5 m (Fig. 3).

Figure 6 shows the modelled WSC in comparison to the validation data estimated based on the lysimeter approach. The same picture applies to the validation and the calibration period. In general, the variation of the behavioural models runs is larger for models calibrated against groundwater than for the SG models. All models underpredict the seasonal amplitude of the measured WSC. Most of the models reproduce the temporal variations of WSC well and agree on the event scale as well as on the seasonal scale.

The variation of the behavioural model sets, the differences in amplitude and temporal variation, are graphically



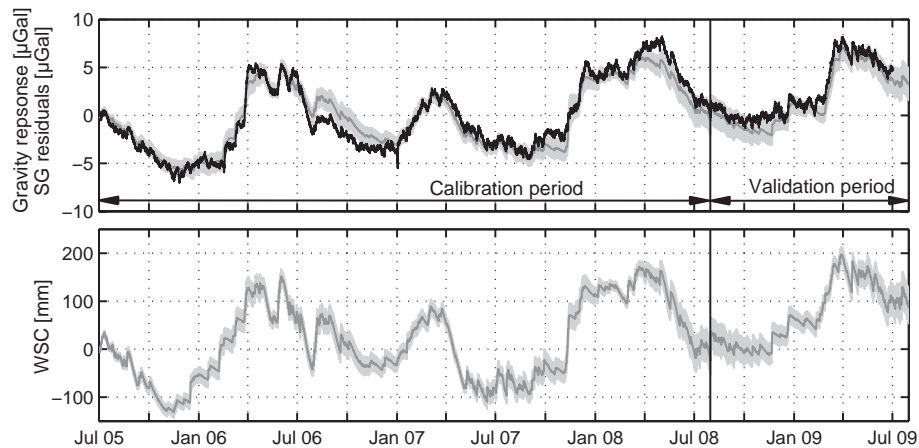
**Fig. 6.** Modelled WSC in comparison to the corresponding validation data estimated based on the lysimeter approach for the period of July 2008 to July 2009. The modelled WSC are presented as the mean of the behavioural model runs in combination with the mean plus/minus two times the standard deviations of the behavioural model runs (see Table 3 for explanation of different models).



**Fig. 7.** Taylor diagrams (Taylor, 2001) comparing measured and modelled WSC for the models SG (a), BK1 (b) and BK3 (c). Each figure also contains the measured WSC for the other sites.

summarised in Taylor diagrams (Taylor, 2001). This is illustrated for the SG model and for two different site types (BK1 and BK3) (Fig. 7). The Taylor diagram can show how well the modelled pattern matches the validation data in terms of  $R$ , RMSD and standard deviation. The standard deviation of the modelled WSC for BK1 ranges between 43 and 65 mm and is clearly smaller than the observed one, but  $R$  can be as high as 0.99. Contrary to model BK1, the differences for BK3 are smaller in terms of observed and modelled standard

deviation, whereas the observed and modelled WSC have a smaller  $R$ . This pattern is consistent for the different site types. The modelled WSC for deeper vadose zone sites agree better with the validation data in terms of temporal dynamics (higher  $R$ ). For shallower sites, the models fit the total signal amplitude in a better way (smaller RMSD). The results of the SG model lie between these two different characteristics. For the sake of completeness, Table 5 also shows the Taylor statistics for the other models. The validation shows



**Fig. 8.** SG residuals (black line) and modelled gravity response (grey band) (top). Modelled water storage change (bottom). The model was calibrated against the SG residuals for the period of 1 July 2005 to 30 July 2008.

**Table 5.** Statistics of the model validation against WSC data.

Models	Std. dev.	<i>R</i>	RMSD
SG	55.45–67.41	0.94–0.99	25.31–37.29
SGECHO	53.80–66.10	0.95–0.99	25.16–38.53
SGTRIME	52.85–67.16	0.91–0.99	26.35–41.21
BK1	43.34–64.80	0.92–0.99	39.57–63.06
BK1ECHO	48.69–64.80	0.96–0.99	39.41–55.86
BK1TRIME	47.53–62.81	0.96–0.99	41.34–57.37
BK2	42.92–64.80	0.87–0.97	17.29–34.53
BK2ECHO	48.69–66.55	0.89–0.96	19.41–32.13
BK2TRIME	47.53–62.80	0.88–0.96	19.81–32.87
BK3	38.43–57.18	0.87–0.98	11.89–29.93
BK3ECHO	38.43–57.17	0.86–0.98	12.74–28.93
BK3TRIME	39.21–56.28	0.89–0.98	11.99–27.73
ECHO	52.23–61.43	0.91–0.98	27.13–41.81
TRIME	47.03–65.35	0.78–0.99	26.72–58.01

that hydrological models constrained by temporal gravimeter data only, can reasonably predict the measured WSC in terms of amplitude and temporal dynamics. Hence, temporal gravimeter data can be used to estimate WSC, even though WSC are measured in change of gravity and not in millimetre of water.

### 4.3 Water storage changes

When we compare the modelled hydrological gravity response (SG model) to the SG residuals for the whole study period, we find that both signals show similarities in terms of amplitude, interannual, seasonal, and short-term variations (Fig. 8). The maximum amplitudes of the SG residuals and the gravity response amount to 15.24 and 14.35  $\mu\text{Gal}$  respectively. This is caused by a maximum WSC of 342 mm.

These numbers are in line with the seasonal gravity variations of 10 to 15  $\mu\text{Gal}$  for the Durzon karst system in France estimated by Jacob (2008). They are caused by a seasonal WSC of 240–360 mm. The RMSD varies between 0.89 and 1.16  $\mu\text{Gal}$ . For the SG residuals and the gravity response, the regression slope of 0.96–1.25 and a corresponding coefficient of determination of 0.90–0.95 reflect a good agreement in phase and amplitude of both time series. The *R* of SG residuals and gravity response ranges between 0.95–0.97.

By focusing on the system state in comparison with the meteorological driving forces, a clear response of WSC can be observed in relation to the input/output fluxes (compare Fig. 3). The different time series show weather-related characteristics and a seasonal course. Similar temporal characteristics can be identified in the hydrological gravity response, the SG residuals and the modelled WSC. High deviations in absolute value as well as in temporal dynamics for groundwater and soil moisture data make it difficult to identify the system response to the meteorological conditions. In Fig. 3, a high variability between the different groundwater levels highlights the problem of single point measurements. It raises the issue of choosing “representative” sites for hydrological measurements, in particular for these complex geological settings. The differences of soil moisture measurements may not only reflect the spatial variability of soil moisture but can also be due to the soil moisture technique used highlighting general problems in measuring soil moisture in the vadose zone (e.g., Evett et al., 2009; Chow et al., 2009; Saito et al., 2009).

Furthermore, WSC in the deep vadose zone may differ significantly from the top soil moisture, but no measurements are available for this zone. Gravity measurements integrate over the different hydrological storage components and the sampling volume is several orders of magnitude larger than that for the point measurements. Gravity observations allow for the identification of whole hydrological system responses

to the driving forces, and gravimeters can serve as a novel measurement instrument for hydrology.

Still, practical aspects limit the application of gravimeters for hydrology. SGs are the state-of-the-art relative gravimeters with a temporal resolution of  $\sim 1$  sec and an accuracy of  $\sim 0.1$   $\mu\text{Gal}$ . However, they are cost-intensive in acquisition and operation. In general, they need a good infrastructure and are operated at a fixed location, although first attempts have been made to take SGs into the field (Wilson et al., 2007). The new SG generation – the *iGrav*<sup>TM</sup> SG – will improve the applicability of SGs in terms of portability, low drift and usability (GWR, 2009). Absolute gravimeters (FG5 and A10; Micro-g LaCoste, 2010a, b) are stable concerning the temporal drift and have the advantage of being portable. The accuracy and temporal resolution is not as high as for SGs (Schmerge and Francis, 2006), but they have already been used to study the relationship of gravity and hydrology (Jacob et al., 2008, 2009). Spring-based gravimeters are relative gravimeters, portable, and relatively inexpensive. In the context of WSC, they are used on a campaign-basis to map the spatial variation of gravity changes in comparison to a reference point. In general, gravity changes above 10–15  $\mu\text{Gal}$  can be detected by these gravimeters, and with very high effort, the detection limit can be lowered to  $\sim 2$   $\mu\text{Gal}$  (Brady et al., 2008; Chapman et al., 2008; Gettings et al., 2008; Pool, 2008; Naujoks et al., 2008). For the sake of completeness, we would like to mention that advances in atom interferometry promise to improve the reliability of absolute gravity measurements and will be available to the geophysical community in the future (de Angelis et al., 2009; Peters et al., 2001). Hence, technical advances in gravimeter technology are necessary in terms of portability, precision and cost-efficiency to tap the full potential of gravimeter measurements for hydrological applications and to make them routinely available to the hydrological community.

## 5 Conclusions

This study investigates the use of temporal gravity measurements as an integrative measure of the hydrological system state. The benefits of gravimeters when it comes to measuring WSC were assessed also in comparison to classical hydrological point measurements (groundwater and soil moisture). To estimate local WSC, a simple conceptual hydrological model was set up. This is the first study in which a model has been calibrated and evaluated using temporal gravimeter data as the only calibration/evaluation constraint. For the sake of comparison, the model was also calibrated against groundwater and soil moisture data and combinations of observation data sets. Using SG measurements as calibration constraints, improved the model results substantially in terms of the model fit to the calibration data, the predictive capability, and the variation of the behavioural model runs. For the SG model, the variations of the behavioural model runs and

the amplitude do not change when additional calibration data are included. They do however change for models calibrated against groundwater data when soil moisture is included.

SG observations are generalised and simplified measurements because they integrate over different storages and a larger area. In this context, they are in accordance with the nature of strongly generalised and simplified models (conceptual models). Furthermore, SG data can help hydrologists find out which simplifications and generalisations are the right ones to describe the overall system state (Kirchner, 2009). SG time series can characterise the hydrological system as a whole, whereas groundwater and soil moisture only permit the evaluation of model components. In this context, the ‘right answers for wrong reasons’ issue remains because it is difficult to assess the internal model structure or single parameter sets using gravimeter data only. Gravimeter records can help finding the right answer, in this case total WSC, instead of evaluating whether the reasons (model structures/parameters) are right or wrong but not knowing the right answer.

The results of different models were validated using independently estimated WSC based on a state-of-the-art lysimeter and complementary observations. Some models predicted the amplitude of measured WSC in a better way and others showed a higher agreement with temporal dynamics. The results of SG models lie between these two different characteristics. In principle, the model validation with independent data proves that gravimeters can serve as a novel measurement method to observe WSC. Rather than solving the inverse problem, WSC are derived from a hydrological model of which the gravity response is calibrated against the SG (forward problem).

The high variability of groundwater and soil moisture data raises the issue of representativeness of point measurements. SG measurements integrate over different hydrological storages and larger volumes and thus permit the identification of the system response to the driving meteorological forces. Hence, temporal gravimeter observations may reveal some system characteristics like maximum total storage capacity, which could not be observed in soil moisture and/or groundwater data.

In this context, gravimeters might contribute to upscale point measurements to the field scale and can narrow the gap to the catchment scale. Hence, temporal gravity measurements also should be investigated in the context of the lateral variability of water storages. For example, as a next step at the Geodetic Observatory Wettzell, the spatial variability of water storages will be investigated along the hill-slope using a physically-based hydrological model in a coupled hydrogeophysical inversion framework. Additionally, different concepts of spatio-temporal variability and stability (e.g., Western et al., 2004; Vereecken et al., 2007; Teuling and Troch, 2005; Brocca et al., 2010; Grayson and Western, 1998; Kachanoski and de Jong, 1988; Vachaud et al., 1985; Famiglietti et al., 2008) should be evaluated in the context

of gravity observations (e.g., Glegola et al., 2009). These theories were developed and tested based mainly on near-surface water storage, but only very few studies used data from deeper zones (e.g., Pachepsky et al., 2005; Kachanoski and de Jong, 1988). So, it might be problematic to apply them directly to gravity measurements. At the same time, this reveals the potential of gravity measurements to test the developed theories of spatio-temporal variability in combination with different spatial scales not only for near-surface water storages but also for the whole hydrological system.

Gravity measurements provide an integral signal which makes them comparable to discharge measurements (Hasan et al., 2008). The disadvantages of gravimeters are that it is difficult to unambiguously identify the signal source and that the sampling volume and the radius of influence change over time (Creutzfeldt et al., 2010a; Creutzfeldt et al., 2008). These downsides also apply to a certain extent to discharge measurements where the area contributing to runoff may change over time or the source is difficult to define (e.g., event/pre-event water). This study shows additional similarities between gravimeter and discharge measurements because due to the integral character of gravimeters, it is difficult to constrain internal model structures or single parameters solely based on one method as already highlighted by Mroczkowski et al. (1997) for discharge measurements. Nonetheless, gravimeter measurements can complement discharge observations. They can help to characterise the catchment status above the outlet point and thus to define storage-output relationships. This provides a valuable contribution towards a better general understanding of catchment dynamics and towards constraining hydrological models.

*Acknowledgements.* The authors are sincerely grateful to the section National Reference Systems for Gravity and to the section Geodetic Observatory Wettzell of the Federal Agency for Cartography and Geodesy (BKG) for providing us with the data. We would also like to thank Ty Ferré, Peter Bauer-Gottwein and Ewald Brückl for their constructive and helpful comments for improving the manuscript. The financial support by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the Priority Program SPP 1257 “Mass transport and mass distribution in the system Earth”, project TASMAGOG is greatly appreciated.

Edited by: W. Wagner

## References

- Abe, M., Takemoto, S., Fukuda, Y., Higashi, T., Imanishi, Y., Iwano, S., Ogasawara, S., Kobayashi, Y., Dwipa, S., and Kusuma, D. S.: Hydrological effects on the superconducting gravimeter observation in Bandung, *J. Geodynam.*, 41, 288–295, 2006.
- Allen, R. G., Walter, I. A., Elliott, R. L., Howell, T. A., Itenfisu, D., Jensen, M. E., and Snyder, R. L.: The ASCE standardized reference evapotranspiration equation, American Society of Civil Engineers, Virginia, USA, 2005.
- Allerup, P.: A comprehensive model for correcting point precipitation, *Nordic Hydrol.*, 28, 1–20, 1997.
- Amalvict, M., Hinderer, J., Makinen, J., Rosat, S., and Rogister, Y.: Long-term and seasonal gravity changes at the Strasbourg station and their relation to crustal deformation and hydrology, *J. Geodynam.*, 38, 343–353, 2004.
- Bergström, S.: The HBV model – its structure and applications., Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, 1992.
- Beven, K., and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Proc.*, 6, 279–298, 1992.
- Beven, K. J.: Rainfall-runoff modeling: Introduction, in: Encyclopedia of hydrological sciences, edited by: Anderson, M. G., John Wiley & Sons Ltd, Chichester, UK, 1857–1868, 2005.
- Blainey, J. B., Ferré, T. P. A., and Cordova, J. T.: Assessing the likely value of gravity and drawdown measurements to constrain estimates of hydraulic conductivity and specific yield during unconfined aquifer testing, *Water Resour. Res.*, 43, W12408, doi:10.1029/2006WR005678, 2007.
- Blöschl, G.: On the fundamentals of hydrological sciences, in: Encyclopedia of hydrological sciences, edited by: Anderson, M. G., John Wiley & Sons Ltd, Chichester, UK, 3–12, 2005.
- Bonatz, M.: Der Gravitationseinfluss der Bodenfeuchte, *Zeitschrift für Vermessungswesen*, 92, 135–139, 1967.
- Bonatz, M. and Sperling, D.: Gravitation effects at the Vanden storage power station, *Cahiers du Centre Européen de Géodynamique et de Séismologie*, Walferdange, 169–180, 1995.
- Bower, D. R. and Courtier, N.: Precipitation effects on gravity measurements at the Canadian Absolute Gravity Site, *Phys. Earth Planet. Int.*, 106, 353–369, 1998.
- Boy, J. P. and Hinderer, J.: Study of the seasonal gravity signal in superconducting gravimeter data, *J. Geodynam.*, 41, 227–233, 2006.
- Brady, J. L., Hare, J. L., Ferguson, J., Seibert, J. E., Klopping, F. J., Chen, T., and Niebauer, T.: Results of the world’s first 4D microgravity surveillance of a waterflood – Prudhoe Bay, Alaska, *SPE Reservoir Evaluation & Engineering*, 11, 824–831, 2008.
- Breili, K. and Pettersen, B. R.: Effects of surface snow cover on gravimetric observations, *J. Geodynam.*, 48, 16–22, 2009.
- Brocca, L., Melone, F., Moramarco, T., and Morbidelli, R.: Spatial-temporal variability of soil moisture and its estimation across scales, *Water Resour. Res.*, 46, W02516, doi:10.1029/2009wr008016, 2010.
- Cañas, A. J., Carff, R., Hill, G., Carvalho, M., Arguedas, M., Eskridge, T. C., Lott, J., and Carvajal, R.: Concept maps: Integrating knowledge and information visualization, in: Knowledge and Information Visualization: Searching for Synergies, edited by: Tergan, S. O., and Keller, T., Lecture Notes in Computer Science, Springer-Verlag Berlin, Berlin, Germany, 205–219, 2005.
- Chapman, D. S., Sahn, E., and Gettings, P.: Monitoring aquifer recharge using repeated high-precision gravity measurements: A pilot study in South Weber, Utah, *Geophysics*, 73, WA83, doi:10.1190/1.2992507, 2008.
- Chow, L., Xing, Z., Rees, H., Meng, F., Monteith, J., and Stevens, L.: Field Performance of Nine Soil Water Content Sensors on a Sandy Loam Soil in New Brunswick, Maritime Region, Canada, *Sensors*, 9, 9398–9413, 2009.
- Creutzfeldt, B., Güntner, A., Klügel, T., and Wziontek, H.: Simulat-

- ing the influence of water storage changes on the superconducting gravimeter of the Geodetic Observatory Wettzell, Germany, *Geophysics*, 73, WA95, doi:10.1190/1.2992508, 2008.
- Creutzfeldt, B., Güntner, A., Thoss, H., Merz, B., and Wziontek, H.: Measuring the effect of local water storage changes on in-situ gravity observations: Case study of the Geodetic Observatory Wettzell, Germany, *Water Resour. Res.*, 46, W08531, doi:10.1029/2009WR008359, 2010a.
- Creutzfeldt, B., Güntner, A., Wziontek, H., and Merz, B.: Reducing local hydrology from high precision gravity measurements: a lysimeter-based approach, *Geophys. J. Int.*, in press, doi:10.1111/j.1365-246X.2010.04742.x, 2010b.
- Crossley, D., Xu, S., and Dam, T. v.: Comprehensive analysis of 2 years of SG data from Table Mountain, Colorado, 13th International Symposium on Earth Tides, Brussels, 1998.
- Crossley, D., Hinderer, J., Casula, G., Francis, O., Hsu, H.-T., Imanishi, Y., Jentzsch, G., Kääriäinen, J., Merriam, J., Meurers, B., Neumeier, J., Richter, B., Shibuya, K., Sato, T., and Dam, T. v.: Network of superconducting gravimeters benefits a number of disciplines, *EOS Trans. Am. Geophys. Union*, 80, 121–126, doi:10.1029/99EO00079, 1999.
- Crossley, D. and Hinderer, J.: A review of the GGP network and scientific challenges, *J. Geodynam.*, 48, 299–304, 2009.
- de Angelis, M., Bertoldi, A., Cacciapuotii, L., Giorgini, A., Lamporesi, G., Prevedelli, M., Saccorotti, G., Sorrentino, F., and Tino, G. M.: Precision gravimetry with atomic sensors, *Meas. Sci. Technol.*, 20, 022001, doi:10.1088/0957-0233/20/2/022001, 2009.
- Evelt, S. R., Schwartz, R. C., Tolk, J. A., and Howell, T. A.: Soil profile water content determination: spatiotemporal variability of electromagnetic and neutron probe sensors in access tubes, *Vadose Zone Journal*, 8, 926–941, doi:10.2136/vzj2008.0146, 2009.
- Famiglietti, J. S., Ryu, D., Berg, A. A., Rodell, M., and Jackson, T. J.: Field observations of soil moisture variability across scales, *Water Resour. Res.*, 44, W01423, doi:10.1029/2006WR005804, 2008.
- Ferré, T. P. A., Bentley, L., Binley, A., Linde, N., Kemna, A., Singha, K., Holliger, K., Huisman, J. A., and Minsley, B.: Critical steps for the continuing advancement of hydrogeophysics, *EOS Trans. Am. Geophys. Union*, 90, doi:10.1029/2009EO230004, 2009.
- Gettings, P., Chapman, D. S., and Allis, R.: Techniques, analysis, and noise in a Salt Lake Valley 4D gravity experiment, *Geophysics*, 73, WA71, doi:10.1190/1.2996303, 2008.
- Glegola, M., Ditmar, P., Bierkens, M. F. P., Arts, R., and Vossepoel, F.: Estimation of the time-lapse gravity errors due to water table and soil moisture variations, *SEG Technical Program Expanded Abstracts*, 28, 976–980, 2009.
- Graeff, T., Zehe, E., Schlaeger, S., Morgner, M., Bauer, A., Becker, R., Creutzfeldt, B., and Bronstert, A.: A quality assessment of spatial TDR soil moisture measurements in homogenous and heterogeneous media with laboratory experiments, *Hydrology and Earth System Sciences*, 14, 1007–1020, doi:10.5194/hess-14-1007-2010, 2010.
- Grayson, R. B. and Western, A. W.: Towards areal estimation of soil water content from point measurements: time and space stability of mean response, *J. Hydrol.*, 207, 68–82, 1998.
- Güntner, A.: Improvement of global hydrological models using GRACE data, *Surv. Geophys.*, 29, 375–397, 2008.
- GWR: *iGrav<sup>TM</sup> SG*: Simplified superconducting gravimeter for portable operation, GWR Instruments, Inc., San Diego, USA, 2009.
- Harnisch, G. and Harnisch, M.: Hydrological influences in long gravimetric data series, *J. Geodynam.*, 41, 276–287, 2006.
- Hasan, S., Troch, P. A., Bogaart, P. W., and Kroner, C.: Evaluating catchment-scale hydrological modeling by means of terrestrial gravity observations, *Water Resour. Res.*, 44, W08416, doi:10.1029/2007wr006321, 2008.
- Hinderer, J., Crossley, D., and Warburton, R. J.: Gravimetric methods – superconducting gravity meters, *Treatise on Geophysics*, 3, 65–122, 2007.
- Hinnell, A. C., Ferré, T. P. A., Vrugt, J. A., Huisman, J. A., Moysey, S., Rings, J., and Kowalsky, M. B.: Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion, *Water Resour. Res.*, 46, W00D40, doi:10.1029/2008WR007060, 2010.
- Hokkanen, T., Korhonen, K., and Virtanen, H.: Hydrogeological effects on superconducting gravimeter measurements at Metsähovi in Finland, *J. Environ. Eng. Geophys.*, 11, 261–267, doi:10.2113/jeeg11.4.261, 2006.
- Hokkanen, T., Korhonen, K., Virtanen, H., and Laine, E. L.: Effects of the fracture water of bedrock on superconducting gravimeter data, *Near Surf. Geophys.*, 5, 133–139, 2007.
- Howell, T. A.: Lysimetry, in: *Encyclopedia of soils in the environment*, edited by: Hillel, D., Elsevier Press, Oxford, UK, 379–386, 2004.
- Huisman, J. A., Snepvangers, J. J. J. C., Bouten, W., and Heuvelink, G. B. M.: Mapping spatial variation in surface soil water content: Comparison of ground-penetrating radar and time domain reflectometry, *J. Hydrol.*, 269, 194–207, 2002.
- Huisman, J. A., Hubbard, S., Redman, J. D., and Annan, A. P.: Measuring soil water content with ground penetrating radar: A review, *Vadose Zone J.*, 2, 476–491, 2003.
- Jacob, T., Bayer, R., Chery, J., Jourde, H., Moigne, N. L., Boy, J.-P., Hinderer, J., Luck, B., and Brunet, P.: Absolute gravity monitoring of water storage variation in a karst aquifer on the larzac plateau (Southern France), *J. Hydrol.*, 359, 105–117, 2008.
- Jacob, T., Chery, J., Bayer, R., Moigne, N. L., Boy, J.-P., Vernant, P., and Boudin, F.: Time-lapse surface to depth gravity measurements on a karst system reveal the dominant role of the epikarst as a water storage entity, *Geophys. J. Int.*, 17, 347–360, 2009.
- Juston, J., Seibert, J., and Johansson, P.-O.: Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment, *Hydrol. Proc.*, 23, 3093–3109, 2009.
- Kachanoski, R. G. and de Jong, E.: Scale dependence and the temporal persistence of spatial patterns of soil water storage, *Water Resour. Res.*, 24, 85–91, doi:10.1029/WR024i001p00085, 1988.
- Kazama, T. and Okubo, S.: Hydrological modeling of groundwater disturbances to observed gravity: Theory and application to Asama Volcano, Central Japan, *J. Geophys. Res.*, 114, B08402, doi:10.1029/2009JB006391, 2009.
- Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, 45, W02429, doi:10.1029/2008WR006912, 2009.
- Klemes, V.: Operational testing of hydrological simulation models,

- Hydrol. Sci. J., 31, 13–24, 1986.
- Klügel, T. and Wziontek, H.: Correcting gravimeters and tiltmeters for atmospheric mass attraction using operational weather models, *J. Geodynam.*, 48, 204–210, 2009.
- Kroner, C. and Jahr, T.: Hydrological experiments around the superconducting gravimeter at Moxa Observatory, *J. Geodynam.*, 41, 268–275, 2006.
- Lampitelli, C. and Francis, O.: Hydrological effects on gravity and correlations between gravitational variations and level of the Alzette River at the station of Walferdange, Luxembourg, *J. Geodynam.*, 49, 31–38 2010.
- Leirião, S., He, X., Christiansen, L., Andersen, O. B., and Bauer-Gottwein, P.: Calculation of the temporal gravity variation from spatially variable water storage change in soils and aquifers, *J. Hydrol.*, 365, 302–309, 2009.
- LfL: Agrarmeteorologisches Messnetz Bayern: Wetterstation Nr. 127, Allmannsdorf, Bayerische Landesanstalt für Landwirtschaft (Bavarian State Research Center for Agriculture) (LfL), Freising-Weihenstephan, Germany, 2009.
- Llubes, M., Florsch, N., Hinderer, J., Longuevergne, L., and Amalvict, M.: Local hydrology, the Global Geodynamics Project and CHAMP/GRACE perspective: some case studies, *J. Geodynam.*, 38, 355–374, 2004.
- Lo, M.-H., Famiglietti, J. S., Yeh, P. J.-F., and Syed, T. H.: Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated baseflow data, *Water Resour. Res.*, 46, W05517, doi:10.1029/2009WR007855, 2010.
- Longuevergne, L., Boy, J. P., Florsch, N., Viville, D., Ferhat, G., Ulrich, P., Luck, B., and Hinderer, J.: Local and global hydrological contributions to gravity variations observed in Strasbourg, *J. Geodynam.*, 48, 189–194, 2009.
- MacMillan, W. D.: Theoretical mechanics: The theory of the potential, Dover Publications, Inc., New York, USA, 469 pp., 1958.
- Merz, R., Parajka, J., and Blöschl, G.: Scale effects in conceptual hydrological modeling, *Water Resour. Res.*, 45, W09405, doi:10.1029/2009WR007872, 2009.
- Meurers, B., Camp, M. V., and Petermans, T.: Correcting superconducting gravity time-series using rainfall modelling at the Vienna and Membach stations and application to Earth tide analysis, *J. Geodesy*, 81, 703–712, 2007.
- Micro-g LaCoste: FG5 gravity meter, Micro-g LaCoste – A Division of LRS, Lafayette, USA, 2010a.
- Micro-g LaCoste: A10 portable absolute gravity meter, Micro-g LaCoste - A Division of LRS, Lafayette, USA, 2010b.
- Mroczkowski, M., Raper, P. G., and Kuczera, G.: The quest for more powerful validation of conceptual catchment models, *Water Resour. Res.*, 33, 2325–2335, 1997.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Naujoks, M., Weise, A., Kroner, C., and Jahr, T.: Detection of small hydrological variations in gravity by repeated observations with relative gravimeters, *J. Geodesy*, 82, 543–553, 2008.
- Neumeyer, J., Barthelmes, F., Kroner, C., Petrovic, S., Schmidt, R., Virtanen, H., and Wilmes, H.: Analysis of gravity field variations derived from superconducting gravimeter recordings, the GRACE satellite and hydrological models at selected European sites, *Earth Planet. Space*, 60, 505–518, 2008.
- Pachepsky, Y. A., Guber, A. K., and Jacques, D.: Temporal persistence in vertical distribution of soil moisture contents, *Soil Sci. Soc. Am. J.*, 69, 347–352, 2005.
- Peters, A., Chung, K. Y., and Chu, S.: High-precision gravity measurements using atom interferometry, *Metrologia*, 38, 25–61, doi:10.1088/0026-1394/38/1/4, 2001.
- Pool, D. R. and Eychaner, J. H.: Measurements of aquifer-storage change and specific yield using gravity surveys, *Ground Water*, 33, 425–432, 1995.
- Pool, D. R.: The utility of gravity and water-level monitoring at alluvial aquifer wells in southern Arizona, *Geophysics*, 73, WA49, doi:10.1190/1.2980395, 2008.
- Ramillien, G., Famiglietti, J., and Wahr, J.: Detection of continental hydrology and glaciology signals from GRACE: a review, *Surveys Geophys.*, 29, 361–374, 2008.
- Refsgaard, J. C. and Knudsen, J.: Operational validation and intercomparison of different types of hydrological models, *Water Resour. Res.*, 32, 2189–2202, 1996.
- Richter, D.: Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Meßfehlers des Hellmann-Niederschlagsmessers, *Deutscher Wetterdienst, Offenbach am Main, Germany*, 93 pp., 1995.
- Rings, J., Huisman, J. A., and Vereecken, H.: Coupled hydrogeophysical parameter estimation using a sequential Bayesian approach, *Hydrol. Earth Syst. Sci.*, 14, 545–556, doi:10.5194/hess-14-545-2010, 2010.
- Rucker, D.: A Coupled Electrical Resistivity-Infiltration Model for Wetting Front Evaluation, *Vadose Zone J.*, 8, 383–388, doi:10.2136/vzj2008.0080, 2009.
- Saito, T., Fujimaki, H., Yasuda, H., and Inoue, M.: Empirical temperature calibration of capacitance probes to measure soil water, *Soil Sci. Soc. Am. J.*, 73, 1931–1937, doi:10.2136/sssaj2008.0128, 2009.
- Schlüter, W., Brandl, N., Dassing, R., Hase, H., Klügel, T., Kilger, R., Lauber, P., Neidhardt, A., Plötz, C., Riepl, S., and Schreiber, U.: Fundamentalstation Wettzell – ein geodätisches Observatorium, *Zeitschrift für Vermessungswesen*, 132, 158–167, 2007.
- Schmerge, D. and Francis, O.: Set standard deviation, repeatability and offset of absolute gravimeter A10-008, *Metrologia*, 43, 414–418, doi:10.1088/0026-1394/43/5/012, 2006.
- Seibert, J.: Estimation of parameter uncertainty in the HBV model, *Nordic Hydrological Conference, Akureyri, Iceland*, 1996, 247–262.
- Seibert, J.: HBV light version 2, user's manual, Department of Earth Sciences, Uppsala University, Uppsala, 2005.
- Šimunek, J., Šejna, M., Saito, H., Sakai, M., and van Genuchten, M. T.: The HYDRUS-1-D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media, version 4.0, HYDRUS software series 3, Department of Environmental Sciences, University of California, Riverside, California, 2008.
- Spence, C., Guan, X. J., Phillips, R., Hedstrom, N., Granger, R., and Reid, B.: Storage dynamics and streamflow in a catchment with a variable contributing area, *Hydrol. Proc.*, 24, 2209–2221, doi:10.1002/hyp.7492, 2010.
- Takemoto, S., Fukuda, Y., Higashi, T., Abe, M., Ogasawara, S., Dwipa, S., Kusuma, D. S., and Andan, A.: Effect of groundwater changes on SG observations in Kyoto and Bandung, *Bulletin d'Information des Marées Terrestres*, 134, 10839–10848, 2002.

- Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M. M.: GRACE measurements of mass variability in the Earth system, *Science*, 305, 503–505, 2004.
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, 106, 7183–7192, 2001.
- Teuling, A. J. and Troch, P. A.: Improved understanding of soil moisture variability dynamics, *Geophys. Res. Lett.*, 32, L05404, doi:10.1029/2004GL021935, 2005.
- Tolk, J. A. and Evett, S. R.: Lysimetry versus neutron moisture meter for evapotranspiration determination in four soils, *Soil Sci. Soc. Am. J.*, 73, 1693–1698, doi:10.2136/sssaj2009.0037, 2009.
- Vachaud, G., Passerat De Silans, A., Balabanis, P., and Vauclin, M.: Temporal stability of spatially measured soil water probability density function, *Soil Sci. Soc. Am. J.*, 49, 822–828, 1985.
- Van Camp, M., Vanclooster, M., Crommen, O., Petermans, T., Verbeeck, K., Meurers, B., Dam, T. v., and Dassargues, A.: Hydrogeological investigations at the Membach station, Belgium, and application to correct long periodic gravity variations, *J. Geophys. Res.*, 111, B10403, doi:10.1029/2006JB004405, 2006.
- Vereecken, H., Kamai, T., Harter, T., Kasteel, R., Hopmans, J., and Vanderborght, J.: Explaining soil moisture variability as a function of mean soil moisture: A stochastic unsaturated flow perspective, *Geophys. Res. Lett.*, 34, L22402, doi:10.1029/2007GL031813, 2007.
- von Unold, G. and Fank, J.: Modular design of field lysimeters for specific application needs, *Water Air Soil Poll. – Focus*, 8, 233–242, 2008.
- Weise, A., Kroner, C., Abe, M., Ihde, J., Jentsch, G., Naujoks, M., Wilmes, H., and Wziontek, H.: Terrestrial gravity observations with superconducting gravimeters for validation of satellite-derived (GRACE) gravity variations, *J. Geodynam.*, 48, 325–330, 2009.
- Werth, S., Güntner, A., Petrovic, S., and Schmidt, R.: Integration of GRACE mass variations into a global hydrological model, *Earth Planet. Sci. Lett.*, 277, 166–173, 2009a.
- Werth, S., Güntner, A., Schmidt, R., and Kusche, J.: Evaluation of GRACE filter tools from a hydrological perspective, *Geophys. J. Int.*, 179, 1499–1515, 2009b.
- Western, A. W., Grayson, R. B., and Blöschl, G.: Scaling of soil moisture: a hydrologic perspective, *Ann. Rev. Earth Planet. Sci.*, 30, 149–180, 2002.
- Western, A. W., Zhou, S. L., Grayson, R. B., McMahon, T. A., Blöschl, G., and Wilson, D. J.: Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes, *J. Hydrol.*, 286, 113–134, 2004.
- Wilson, C. R., Wu, H., Scanlon, B., and Sharp, J. M.: Taking the superconducting gravimeter to the field – A new tool for hydrologic and other investigations, 2007,
- Wziontek, H., Falk, R., Wilmes, H., and Wolf, P.: Precise gravity time series and instrumental properties from combination of superconducting and absolute gravity measurements, in: *Observing our Changing Earth*, edited by: Sideris, M. G., Springer, Berlin, 301–306, 2009a.
- Wziontek, H., Wilmes, H., Creutzfeldt, B., and Güntner, A.: High precision gravimetric time series in comparison with global and local hydrology models, 2009 AGU Fall Meeting, San Francisco, USA, 2009b.
- Yang, J., Li, B., and Shiping, L.: A large weighing lysimeter for evapotranspiration and soil-water-groundwater exchange studies, *Hydrol. Proc.*, 14, 1887–1897, 2000.
- Zaitchik, B. F., Rodell, M., and Reichle, R. H.: Assimilation of GRACE terrestrial water storage data into a land surface model: results for the mississippi river basin, *J. Hydrometeorol.*, 9, 535–548, 2008.
- Zreda, M., Desilets, D., Ferré, T. P. A., and Scott, R. L.: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons, *Geophys. Res. Lett.*, 35, L21402, doi:10.1029/2008GL035655, 2008.