Postprint of: Berezowski T., Chormański J., Batelaan O., Skill of remote sensing snow products for distributed runoff prediction, Journal of Hydrology, Vol. 524 (2015), pp. 718-732, DOI: 10.1016/j.jhydrol.2015.03.025

Skill of remote sensing snow products for distributed runoff prediction

Tomasz Berezowski (1, 2), Jarosław Chormański (1) and Okke Batelaan (2,3)

1. Department of Hydraulic Structures, Warsaw University of Life Sciences, Nowoursynowska 166, 02-787 Warsaw, Poland, e-mail: t.berezowski@levis.sggw.pl, corresponding author

2. Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

3. School of the Environment, Flinders University, GPO Box 2100, Adelaide SA 5001, Australia

---

Abstract

With increasing availability of remote sensing snow cover products we aim to evaluate the skill of these datasets with regard to hydrological discharge simulation. In this paper ten model variants using different snow cover data (MOD10A1, IMS, AMSR-E SWE, GLOBSNOW SWE and observed in situ snow depth) and two different model structures for snow accumulation and snowmelt switching (based on snow cover data time series or temperature time series) are calibrated with a global optimization algorithm. The simulated discharge is subjected to five criteria for validation, while the GLUE methodology is used for uncertainty analysis of the ten model variants. The skill of the datasets is tested for the Biebrza River catchment, which has a hydrological regime dominated by snowmelt. The discharge simulations are conducted with the distributed rainfall-runoff model WetSpa. MOD10A1 was the only data source which improved the validation Nash-Sutcliffe (NS) scores in reference to a standard model. However, other evaluation measures indicate that the following data sources performed better than the standard model: MOD10A1, observed snow depth and GLOBSNOW for Kling-Gupta efficiency and for high flows; IMS and MOD10A1 for bias; GLOBSNOW and MOD10A1 for coefficient of determination. MOD10A1 has the highest spatial resolution of all analysed data sources which might contribute to the high skill of this data. The use of the data-based switching model structure generally narrowed the behavioural parameter sets during the uncertainty analysis when compared to the temperature-based switching. However, no clear relation was observed between the prediction confidence interval and the two model structures. It is concluded that the skill of the remote sensing snow cover data for the model is positive, although, strongly varying with the data source used.

Keywords: remote sensing, snow products, snow, catchment hydrology, calibration, uncertainty analysis

---

1. Introduction

With the increasing availability of remote sensing based snow cover products the number of studies using these data in hydrological models are growing. Certainly the most popular remote sensing snow products are derived from the Moderate Resolution Imaging Spectroradiometer (MODIS)\Terra and the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) sensors, but the multi-sensor products like the Interactive Multisensor Snow and Ice Mapping System (IMS) and the relatively new Global Snow Monitoring for Climate Research (GLOBSNOW) are gaining interest. The quality of these data sources is assessed against observations in meteorological stations (Parajka and Blöschl, 2006; Chen et al., 2012; Byun and Choi, 2014). Some studies intercompare two snow products with the ground truth (Şorman et al., 2009; Gao et al., 2010a; Hancock et al., 2013). However, a comparison of all available remote sensing products at the same time is methodically difficult, since they contain different variables e.g.: snow cover fraction (SCF), snow water equivalent (SWE) or snow cover extent.

Hydrological models, however, are flexible in using various quantitative snow variables, because they use different model concepts for simulating snow processes. Most relevant studies use one particular snow cover dataset as input data in a hydrological model (Yan et al., 2009; Butt and Bilal, 2011; Bavera et al., 2012). More interesting results are, however, obtained when a remote sensing snow product is compared in a hydrological model with other datasets or with measurements from meteorological stations (Udnaes et al., 2007; Şensoy and Uysal, 2012; Yatheendradas et al., 2012). These studies reveal the influence of different data sources on modelling results. Hydrological models are thus a good framework for quality assessment of remote sensing snow cover data.

Several studies show how remote sensing snow cover data can aid in hydrological modelling. Molotch and Margulis (2008) used SCF data from various sensors in order to simulate SWE with a spatially distributed snowmelt model. Parajka and Blöschl (2008) used MODIS snow cover data in combination with discharge data for hydrological model calibration in a number of catchments in Austria. The models calibrated with use of the MODIS data improved the simulation of snow cover, but slightly decreased efficiency of discharge simulation when compared to models calibrated with discharge data only. These findings were in agreement with Udnaes et al. (2007) and Şorman et al. (2009). Another approach was presented by Shrestha et al. (2014) who used MODIS snow cover data in order correct snowfall in a distributed hydrological model. The model using the corrected snowfall improved discharge and snow cover simulation when compared to models using uncorrected data. However, so far a study performing a multi-data-source intercomparison with different remote sensing snow products (obtained from microwave and optical sensors at different spatial resolutions) directly using the data as input for a hydrological model is still lacking. It is important to mention that these experiments are indirect assessments, i.e. the snow cover data quality is evaluated in regard to the skill to simulate the discharge, and is not compared to the snow ground truth in meteorological stations.

Because of this indirect evaluation of the snow cover data, the comparison of different snow products should be conducted with an appropriate hydrological model. The model should allow using remote sensing input data, hence be distributed and physically based, because only in this case both the spatial distribution and the states of the snow variables may be evaluated. Of the available hydrological models fulfilling these criteria, the most popular are VIC, DHSVM, WEB-DHM-S, MIKE SHE, SWAT or WetSpa. The GIS, grid-based structure of the WetSpa model allows straightforward implementation of remote sensing input data (Chormański et al., 2008; Berezowski et al., 2012; Verbeiren et al., 2013). Moreover, WetSpa was proven to be sensitive to the spatial distribution of snow cover in particular (Berezowski et al., 2014). An open question is the comparison method for the simulation results of models using different snow products.

Verbeiren et al. (2013) compared WetSpa modelling scenarios using different distributed data. In their study each model variant was calibrated with a local method (PEST; Doherty, 2010) and the simulation results were compared in terms of several evaluation criteria. This framework could be improved by using a global optimization algorithm e.g. Shuffled Complex Evolution (SCE; Duan et al., 1992), which should give more reliable parameter estimates. Another improvement could be to subject the different models to uncertainty analysis. For this purpose Younger et al. (2009) used the Generalized Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992). GLUE was used in their study to show how the rainfall data perturbed by different factors influenced the uncertainty in hydrological modelling scenarios.

The aim of this paper is to assess the influence of different snow cover data on discharge simulations with a distributed hydrological model. The influence is assessed by means of global calibration and uncertainty analysis of ten hydrological model variants using different snow cover data sources and different model structures. The paper also answers the question: Can remote sensing snow cover data be used as a direct driver for snow processes in order to improve the discharge simulation in comparison to a standard model which uses only in situ data? In the section Methods we describe the study area, data and the hydrological modelling experiment. The latter gives insight into the hydrological model with its variants compared in the study and the methods of calibration and uncertainty analysis. In the Results and discussion a detailed description is given of the performed analyses and comparison with other studies. In the

Table 1:

Put Table 1 here.

Conclusions the most important findings of the study are presented.

2. Methods

2.1. Study area

The study area is the Biebrza River catchment located in the north-eastern part of Poland (Fig. 1). The catchment is of medium size (6845 km^2) dominated by agricultural land-use (54%) with a big share of forests (26%) and grasslands (17%); a minor part is covered with water (2%) and buildings (1%). Despite the majority of agricultural land-use, the catchment is considered as natural with very low human impact during last centuries. The indicators of the naturalness are well preserved organic soils in the Biebrza valley, which cover 16% of the whole catchment. The remaining parts of the catchment are covered with mineral soils, mostly: sand, loamy sand and sandy loam. The elevation ranges from 102 to 298 m ASL with low average slope of 1.03%. The Biebrza National Park (592 km^2) covers most of the river valley and is one of Poland's most eminent nature reserves. An important feature of the Biebrza River valley is its water storage capability during the snowmelt-fed spring floods (Grygoruk et al., 2013).

Biebrza River catchment was selected as a study area because the yearly snow cover period is the longest among the medium size catchments in lowland Poland. The meteorological data, obtained from the Olecko station located within the catchment borders is presented in Table 1. The cold winters are characterized by a long period with snow cover. The spring floods, with considerable snowmelt (Chormański et al., 2011), dominate the hydrological behaviour of the Biebrza River and have produced a peak discharge at the catchment outlet of 517 m^3/s (3^{rd} April 1979). In the 1979 hydrological year (in Poland 1^{st} November to 31^{st} October) the snow cover extent (Brodzik and Armstrong, 2013) and discharge at the catchment outlet are maximum observed for the pe-

Put Figure 1 here.

Figure 1:

Put Figure 2 here.

Figure 2:

riod 1966-2012. Each year the spring floods cover 12 to 140 km^2 of the lower Biebrza basin (Ignar et al., 2011), this large flooded zone has an important impact on the landscape and environmental processes (Chormański et al., 2009; Fig. 2). In contrast, the minimum observed discharge is only 4.33 m^3/s (11th July 1969). The mean discharge (1951-2012) is 34.9 m^3/s, while the winter and summer discharge is respectively 43.9 m^3/s, and 26.0 m^3/s.

2.2. Data

2.2.1. Hydrometeorological data

The meteorological data was provided by the Polish Institute of Meteorology and Water Management - National Research Institute (IMGW). Precipitation time series were available for 25 stations, but daily temperature and snow depth (SD [cm]) were available for only 5 stations (Fig. 1). The potential evapotranspiration (PET) time series were approximated by mean monthly evaporation from free water surface (Stachý, 1987). The monthly values were uniformly disaggregated into daily values and used as input data. The PET data was not crucial in this study as it is focused on snow related processes, for which PET is not highly influential. Daily Biebrza River discharge was obtained from the Burzyn gauging station which is located near the Biebrza River confluence with the Narew River (Fig. 1).

2.2.2. Spatial data

The land-use, soil and elevation data are the three main inputs for calculating distributed parameters in WetSpa. The land-use map was obtained from the Corine Land Cover 2006 Project. The source for soil data was the Soil Map of

Table 2:

Put Table 2 here

Poland in scales 1:50 000 for agricultural areas and 1:500 000 for the other land-use classes. These data sources have missing data in the part of the catchment located in Belarus (<1% of the study area) where sand soil and agricultural land-use were assigned. The elevation data was obtained from the Digital Elevation Model of Poland (1:26 000), contours of the Topographic Map of Poland (1:25 000) and geodetic field survey in the river valley, where the lowest slopes occurs. All of these data were used to construct a digital elevation model using TOPO to Raster algorithm in the ArcGIS 10 system with a spatial resolution of 20 m.

2.2.3. Remote sensing data

In this study, the distributed snow time series were obtained from four satellite remote sensing snow products: AE_DySno (AMSR-E/Aqua Daily L3 Global Snow Water Equivalent EASE-Grids; Tedesco et al., 2004), GLOBSNOW (GLOBSNOW Snow Water Equivalent; Takala et al., 2011); IMS (Interactive Multisensor Snow and Ice Mapping System - Daily Northern Hemisphere Snow and Ice Analysis at 4 km; Helfrich et al., 2007; National Ice Center, 2008); and MOD10A1 (MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 5; Hall et al., 2002, 2006).The products may be divided into three groups for the scope of data used in the algorithms: passive microwave, visual/near infra-red and a combination of both.

The products that use passive microwave data for retrieving snow information do not have missing data due to cloud cover, as the atmospheric effect on microwave radiation is assumed to be insignificant. However, it should be noted, that Wang and Tedesco (2007) identified an impact of cloud absorption on AMSR-E SWE retrievals. The snow pack characteristics from microwave data are obtained from differences in brightness temperature registered at different frequencies. Chang et al. (1987) presented an early example of such an approach for SD mapping:

$$SD = 1.59\,(18\mathrm{GHz} - 37\mathrm{GHz}) \tag{1}$$

where 18GHz and 37GHz are brightness temperatures at these frequencies at horizontal polarization. The AE_DySno data extend the Chang et al. (1987) approach by dynamically modelling snow density and grain size with additional quality assessment steps (Kelly et al., 2003). The SWE retrieval algorithm for GLOBSNOW uses a different model which includes a data assimilation technique (Pulliainen, 2006; Takala et al., 2011). However, the brightness temperature input data for GLOBSNOW is obtained at similar frequencies as in the AE_DySno data.

Contrary to the previous products the MOD10A1 data estimate SCF based on visible and near-infrared radiation, which is vulnerable to atmospheric conditions. In order to overcome the cloud contamination problem in the MODIS snow products, a blending methodology with microwave-based snow cover data can be used (e.g. Liang et al., 2008; Gao et al., 2010b). In this study, however, the MOD10A1 data is used as it is provided by the producer. In MOD10A1, the SCF is estimated based on Normalized Difference Snow Index (Hall et al., 1995):

$$NDSI = \frac{\rho_{vis} - \rho_{nir}}{\rho_{vis} + \rho_{nir}} \tag{2}$$

where ρ_{vis} and ρ_{nir} are reflectance at visible and near-infrared wavelengths, which for the MOD10A1 data are 0.545 - 0.565 μm and 1.628 - 1.652 μm respectively. In the MOD10A1 algorithm a pixel with $NDSI > 0.4$ is assumed to be snow covered, however, it also has to pass other tests involving MODIS bands at different wavelengths (Hall et al., 2002) and be corrected in forested areas (Klein et al., 1998).

The IMS product benefits from both visible/near-infrared and microwave data. In this case, the snow and ice extent output maps are manually derived by analysts; such a process may take from 1 to 5 hours depending on a season (Helfrich et al., 2007). In this approach a first guess of snow extent is obtained from the previous day map; next an analyst interprets the visible bands and microwave data, but may also rely on automated snow products e.g. for cloud cover areas.

2.3. The snow products comparison with the observed snow depth

The snow products were compared with the daily SD records form the three meteorological stations located in the catchment: Olecko, Biebrza-Pieńczykówek and Różanystok (Fig. 1). The SD from the stations was compared with the values of the remote sensing snow products in the grid cells containing the station. For this analysis, the available snow datasets (remote sensing snow products and the SD in the stations) were reclassified to binary, dimensionless values, representing snow presence and absence, according to:

$$\begin{cases} 0 & \text{if } Sv = 0 \\ 1 & \text{if } Sv > 0 \end{cases} \tag{3}$$

where Sv is the value of an analysed snow dataset in the units of the datasets (e.g. %, cm or mm). The data after reclassification was further cross-tabulated and used as input for the Receiver Operating Characteristic (ROC) curve plots (Brown and Davis, 2006). The ROC curves are able to compare the performance of different binary models with respect to true positive signal rate and false positive signal rate; both rates ranges between 0 and 1. The ROC curves are obtained by plotting the true positive rate against false positive rate with additional points at (0,0) and (1,1). The higher the integral value of a ROC curve, the better the model performs; a random guess case on a ROC plot would be a 1:1 line with the integral equal to 0.5. For an ideal case the true positive rate would be 1 and the false positive rate would be 0 giving the integral equal to 1.

2.4. Snow products preprocessing

A simple preprocessing methodology was setup, because the idea of this study was to compare the datasets in the state as provided by the producer without using any data assimilation or missing-data simulation algorithms (except linear interpolation). Despite the variability of used snow products in this study, all data were preprocessed in the same way:

1. Identification of the snow product grid cells within the study area. The original extents of the grid cells were further used as a snow field map in the model.
2. Extraction of time series from the grid cells identified at point 1 into a tabular format.
3. Linear interpolation of the snow products over time for removing missing data and to achieve a constant 1 day temporal resolution.
4. Assignment of no snow from May to September since no snow cover is observed in that period in lowland Poland (see Tab. 1).

In order to decrease the WetSpa computation time the MOD10A1 data was aggregated from 500 m to 4 km grid in the first step of the preprocessing. The aggregated data allowed decreasing by 64 times the number of SCF values which has to be read at each time step of the model in reference to the data in the original resolution. The observed SD data was provided as point measurements which were interpolated with the Thiessen polygon method in order to obtain a distribution of SD over whole study area.

2.5. Hydrological modelling

2.5.1. The WetSpa model

The hydrological simulations were performed with the WetSpa model, which is an acronym for Water and Energy Transfer between Soil, Plants and Atmosphere (Wang et al., 1996; De Smedt et al., 2000; Liu et al., 2003). WetSpa is a distributed model, which divides an investigated catchment into a regular grid of computational cells. In each grid cell the water balance is calculated based on physical and empirical equations. The computations are based on two types of input data: (1) the time series of precipitation, temperature and PET distributed over the modelled area with Thiessen polygons for the meteorological stations; (2) maps of physical parameters calculated at the preprocessing stage from soil, land-use and elevation data.

A simulation time step begins with processing the precipitation, which can be either stored as interception or as depression storage. From both storages the water can evaporate, but it can infiltrate in the soil from the depression storage. surplus of depression storage is transformed into surface runoff. Based on the porosity and infiltration the soil moisture is calculated. Part of the water in the soil moisture can be taken up by roots and be transpired while the remaining part will move laterally as interflow runoff or flow vertically further to the transition zone. However, if the root zone is fully saturated the surplus of the water will be transformed to interflow runoff. The water in transition zone recharges the saturated zone where it is stored and transformed to groundwater runoff. WetSpa does not consider soil freezing and thawing, these processes are not expected to have a considerable impact on the discharge simulation at the catchment outlet of our study area.

The surface runoff is routed along flow paths from each a cell to the catchment outlet with the instantaneous geomorphological unit hydrograph (IUH). The IUH is an analytical solution of the diffusive wave approximation proposed by Liu et al. (2003):

$$U(\tau) = \frac{1}{\sqrt{2\pi\sigma^2\tau^3/\tau_0^3}} \exp\left[-\frac{(\tau-\tau_0)^2}{2\sigma^2\tau/\tau_0}\right] \tag{4}$$

where: $U(\tau)$ [T^{-1}] is the response function of the flow path at time τ [T], τ_0 [T] is the mean flow time from a grid cell to the catchment outlet, and σ [T] is the variation of the flow time. In the standard WetSpa the interflow runoff is routed the same as surface runoff. However, in this study the interflow runoff is calculated in the same way as the groundwater runoff, i.e. it is integrated on a sub-catchment level and transported to the river using the linear reservoir method, next, through the river it is transported to the catchment outlet with the IUH. This manipulation was made to delay the interflow response in the catchment, which is conceptually more appropriate for this lowland, non urbanized catchment. For this reason an additional recession coefficient (k_{i2} [m^2/s]) was introduced to WetSpa as a global parameter.

In this study the model was set up for 250 m grid cells and a daily time step. Based on previous WetSpa experiences these values are appropriate for a medium sized catchment with slow response time. As a result the model has 467×458 cells and a mean travel time from the grid cells to the catchment outlet of 122 h.

2.5.2. Snow in the WetSpa model

In the standard WetSpa version snow accumulation and snowmelt are calculated based on snow pack modelling. The precipitation (v_{pre} [mm]) is considered to be snowfall if the temperature in a grid cell (t [°C]) is below or equal to the threshold temperature (t_0). Snowfall is accumulated in the cell's snow pack (s [mm]) as a water equivalent which will be released as snowmelt when $t > t_0$, i.e. snowmelt / snow accumulation switching is based on t_0, further referred to as temperature-based switching. The snowmelt amount in a cell (v_{sm} [mm]) is calculated with the degree-day and the rainfall degree-day methods:

$$\begin{cases} k_{snow}(t-t_0) + k_{rain}v_{pre}(t-t_0) & \text{if } t > t_0 \wedge s \geqslant v_{sm} \\ s & \text{if } t > t_0 \wedge s < v_{sm} \\ 0 & \text{if } t \leqslant t_0 \end{cases} \tag{5}$$

where: k_{snow} [mm/°C/day] is the degree-day coefficient (amount of snowmelt caused by 1°C temperature rise above t_0) and k_{rain} [mm/mm/°C/day] is the rainfall-degree-day coefficient (amount of snowmelt caused by 1 mm of rainfall varied by the t rise above t_0). However, in this study, next to the standard model, other models of calculating snowmelt are tested as well. These models involve using the remote sensing snow products and the observed SD in the meteorological stations as input time series in WetSpa.

In these non-standard models, the snow accumulation in s is not calculated, because the distribution or amount of snow in the study area is obtained directly from the datasets. Moreover, modelling the snowmelt from the datasets time series allows using the time series itself instead of temperature to function as

Table 3:

Put Table 3 here

the governing variable for snowmelt / snow accumulation switching, which is further referred to as data-based switching. For the data-based switching, v_{sm} is generated only if Sv (definition in Sect. 2.3) value at time step i is smaller than at time step $i-1$. Both temperature-based and data-based switching are tested in this study, if possible for a given snow product. A summary of the changes in the WetSpa model for different snow products is listed in Table 3.

In total ten model variants were tested, i.e.: four with both temperature and data-based switching and two with temperature-based switching only (Tab. 3). The data-based switching was obviously not possible to implement in the standard model, which does not use any Sv variable. For the IMS model, the data-based switching was not implemented because the IMS data has only discrete values of snow presence and absence (Tab. 3). Hence, snowmelt would occur in a time step when snow cover completely disappears in an IMS grid cell.

The model variants presented in Table 3 use different equations for snowmelt calculation. The SD and SWE models benefit from physical values in the time series and thus have relatively direct calculations which introduced two new parameters: ρ [mm water/cm] is the snow density factor for calculating water equivalent from SD; k_{cor} [-] is the correction factor for the SWE data sources.

The IMS and MOD10A1 model variants (Tab. 3) use Liston (1999) approach for calculating v_{sm}, which provides a link between a dimensionless SCF and snowmelt depth expressed in physical units. The link is achieved by weighting the potential snowmelt (i.e. maximum that could occur under given meteorological conditions) with the value of SCF in a cell. As a result v_{sm} is generated on a sub-grid-cell level with the potential snowmelt simply estimated with the degree-day methods (as in the standard WetSpa).

2.5.3. Calibration and validation

Global calibration and independent validation were used to test the performance of the WetSpa model using various data sources and model structures (switching methods) for snow processes simulation (listed in Tab. 3). The ten model variants were calibrated using a same procedure with the SCE algorithm (Duan et al., 1992). SCE performs a search in the defined parameter space in several complexes that shuffle information between each other after each iteration. This strategy allows to find the global optimum in a complex parameter space. However, the SCE calibration may not always succeed in finding the optimum. In order to increase the probability of finding the global optimum the calibration was repeated three times with a different set of initial parameter values; the number of complexes in the SCE algorithm was set to 5. The calibration period of three hydrological years was from 1^{st} November 2004 to 31^{st} October 2007, preceded by a two month warm-up; validation was performed in the subsequent three hydrological years: 1^{st} November 2007 to 31^{st} October 2010. The objective function was to maximize the Nash-Sutcliffe efficiency [-] (Nash and Sutcliffe, 1970):

$$NS = 1 - \frac{\sum_{n=1}^{N} \left(Q_n - \hat{Q}_n\right)^2}{\sum_{n=1}^{N} \left(Q_n - \bar{Q}\right)^2} \tag{6}$$

where: Q_n and $\hat{Q}_n$ are observed and simulated discharges at time n, $\bar{Q}$ is the mean observed discharge, N is the number of time steps. The parameter space is defined in Table 4.

The model efficiency was additionally evaluated with the logarithmic version of Nash-Sutcliffe efficiency [-] (Smakhtin et al., 1998):

$$NS_{low} = 1 - \frac{\sum_{n=1}^{N} \left(\ln\left(Q_n\right) - \ln\left(\hat{Q}_n\right)\right)^2}{\sum_{n=1}^{N} \left(\ln\left(Q_n\right) - \overline{\ln\left(Q\right)}\right)^2} \tag{7}$$

which put emphasis on the low discharge simulation; the $\overline{\ln(Q)}$ is the mean of natural logarithms: $\sum_{n=1}^{N} (\ln(Q_n))/N$. Another adapted version of Nash-Sutcliffe

Table 4:

Put Table 4 here.

efficiency [-] was used for high discharge evaluation (Liu and De Smedt, 2004):

$$NS_{high} = 1 - \frac{\sum\limits_{n=1}^{N} \left(Q_n + \bar{Q}\right)\left(\hat{Q}_n - Q_n\right)^2}{\sum\limits_{n=1}^{N} \left(Q_n + \bar{Q}\right)\left(Q_n - \bar{Q}\right)^2} \tag{8}$$

The model bias, or the mean error [m^3/s] of the simulation was calculated as:

$$ME = \frac{\sum\limits_{n=1}^{N} \left(\hat{Q}_n - Q_n\right)}{N} \tag{9}$$

and was used to indicate weather a simulation was over or underestimating the observed discharge. Another measure of accuracy was the coefficient of determination, defined as a square of the Pearson's correlation coefficient [-]:

$$r^2 = \left(\frac{\sum_{i=1}^{n}(Q_i - \bar{Q})(\hat{Q}_i - \tilde{Q})}{\sqrt{\sum_{i=1}^{n}(Q_i - \bar{Q})^2}\sqrt{\sum_{i=1}^{n}(\hat{Q}_i - \tilde{Q})^2}} \right)^2 \tag{10}$$

where $\tilde{Q}$ is the mean simulated discharge. The r^2 is sensitive to the collinearity of the variables (Legates and McCabe, 1999), thus gives an idea about timing and behaviour of the simulated discharge in reference to the observed discharge. The model performance was also quantified with Kling-Gupta efficiency (KG, Gupta et al., 2009) [-], which calculates euclidean distance from the optimal point using: ME (Eq. 9), r (Eq. 10) and variability (α). Thus is considered a better overall of eficiency measure than NS.

$$KG = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (ME - 1)^2} \tag{11}$$

where $\alpha = {}^{\sigma_s}/_{\sigma_o}$, σ_s and σ_o are standard deviations of simulated and observed discharge.

2.5.4. Uncertainty analysis

The performance of the WetSpa model variants presented in Table 3 was also evaluated in the scope of uncertainty analysis (UA). For each model variant a

Put Figure 3 here.

Figure 3:

discharge simulation uncertainty was estimated with the GLUE method (Beven and Binley, 1992). GLUE assumes model equifinality, what means that a model does not have a single optimal combination of parameter values, but a set of parameter combinations exists with equally good model performance. Such parameter sets are called behavioural, i.e. they properly represent the modelled system. The population of behavioural parameters sets is found by running the model many times with randomly sampled parameters values. Next, the simulations corresponding to the random parameters are quantified with formal or informal likelihood functions. Finally, a threshold for the likelihood function is defined above which the given simulations and corresponding to them parameters ranges are called behavioural.

In this study the UA was performed for the ten models listed in Table 3 for one hydrological year (1^{st} November 2008 to 31^{st} October 2009) with a four months warm-up period. The random 100,000 set of parameters were generated with the latin-hypercube algorithm (McKay et al., 1979), which is used in GLUE for sampling the whole parameter space with a minimum correlation (Beven and Freer, 2001). The parameter space is defined in Table 4. The used likelihood function was NS (Eq. 6), which is one of the suggested by Smith et al. (2008); the threshold for behavioural simulations was $NS > 0.60$.

3. Results and discussion

3.1. Snow product comparison with observed snow depth

All remote sensing snow products have satisfying accuracy with respect to the observed SD data before the preprocessing described in Section 2.4. The ROC curves, presented in Fig. 3, have the following integral values: GLOBSNOW = 0.71, AE_DySno = 0.85, IMS = 0.95, MOD10A1 = 0.97. The highest accuracy was obtained, based on these results, for the MOD10A1 data and the worst for GLOBSNOW. The presented integral values depict also a clear difference between SCF data (MOD10A1 and IMS) and SWE data (AE_DySno and GLOBSNOW). The SWE data have lower true positive rates than SCF data. This can be explained by the difference in spatial resolution that is higher for SCF data and thus represents better the local measurements of the meteorological stations. Also, the sources of uncertainty in SWE data are related to other factors, such as topographic characteristics (forests, complex terrain) and snow properties (grain size, density) (Byun and Choi, 2014).

Despite the differences in true positive rates between the snow products, the false negative rate is at similar level for each dataset. Hence, all the investigated remote sensing datasets performed well when no snow cover was observed.

The good accuracy of the MODIS data (Fig. 3) is also confirmed in other studies, which compared MOD10A1 with meteorological stations readings in the Rio Grande Basin, USA, (Klein and Barnett, 2003), Austria (Parajka and Blöschl, 2006) and in Turkey (Şorman et al., 2007). A comparison of MOD10A1 and AE_DySno was conducted by Gao et al. (2010a) in Alaska. Their results showed, similarly like in this paper, higher accuracy of MOD10A1 data than AE_DySno due to the differences in spatial resolution.

The IMS accuracy was evaluated by Chen et al. (2012) for continental USA but the authors claim that the results are also representative for the mid-latitude region of Eurasia. They reported an accuracy of 80% to 100% varying by season, what despite the different measure of accuracy, agrees with the good accuracy observed in this study (Fig. 3).

The GLOBSNOW product was compared with AE_DySno by Hancock et al. (2013) in terms of timing, seasonal patterns and peak SWE accumulations. The timing analysis showed that start and end dates of snow events was erroneous for both of the SWE products because passive microwave retrievals are insensitive to shallow or wet snow. This can be a reason also of relatively low true positive rate of GLOBSNOW and AE_DySno (Fig. 3). Nevertheless, Hancock et al. (2013) reported also that GLOBSNOW was superior to AE_DySno in seasonal patterns and peak value retrievals of SWE. The superiority of GLOBSNOW is

Put Figure 4 here.

Figure 4:

in contradiction to the accuracy results (Fig. 3), however, the differences may be due to different measures of comparison used in both studies.

Time series of the snow products used as input data to the model variants and simulated snow cover data by the WetSpa model (i.e. after the preprocessing described in Sect. 2.4) are presented in Figure 4. The remote sensing and WetSpa-simulated time series resemble the seasonal pattern of the observed SD data. The differences in magnitude of the variables are, however, clearly visible especially with the SWE data. In the case of the SCF data a clear difference is visible between continuous MOD10A1 and discrete IMS data.

3.2. Calibration and validation of WetSpa with variety of snow products

The calibrations with the global SCE algorithm and validation results are presented in Tab. 5. Each model variant was calibrated with similar, high NS scores at least for 2 of the 3 calibration runs meaning that the calibration results converges to the global optimum. The differences between the optimal calibration results (Tab. 5), gives evidence for equifinality of the models.

The best model variant in terms of validation NS score was MOD10A1 with the data-based switching; the temperature-based switching variant for MOD10A1 was the second best (Tab. 5). The Standard WetSpa, which did not use remote sensing snow cover data, but the dense representation of meteorological stations (Fig. 1) to model the snow accumulation and snowmelt, resulted in only a slightly worse NS score. Nevertheless, the use of MOD10A1 data improved the NS scores when compared to the Standard WetSpa variant. Similar findings were presented by Parajka and Blöschl (2008) who reported improved simulation efficiency of snow cover when MODIS snow cover data was included additionally to discharge in the optimization function. However, the inclusion of MODIS data did not improve the discharge simulation. On the other hand, the discharge simulation was improved when snowfall was corrected

Table 5:

Put Table 5 here.

by MODIS data, as reported by Shrestha et al. (2014). Both mentioned studies, did not use MODIS data to drive directly the snow processes in a hydrological model presented in this research.

The model variants with data sources that could implement both data-based and temperature-based switching performed better, with respect to the NS validation scores for the data-based switching than for the temperature-based switching (Tab. 5). This suggests that the variance of the snow products tested in this study reflects the variance in observed phenomena and do not have to be adjusted with the temperature. The decrease in NS validation scores for the temperature-based switching in comparison to the data-based switching is also noted for the Observed SD model variant, which does not use remote sensing data. This confirms that day to day snow variations represent better snowmelt phenomena than the daily averaged temperature variations.

The best NS calibration scores were obtained by the Observed SD model variants (Tab. 5). However, the NS scores were much lower for the validation period. The reason could be the use of a constant snow density factor (ρ), while the true snow density is nonstationary during the snow season and between the seasons, moreover, it is geographically dependent (Onuchin and Burenina, 1996; Bormann et al., 2013). Hence, the ρ value that fit well the calibration period did not represent the snow densities in the validation period. Another reason could be a too sparse network of meteorological stations with SD and temperature measurements. Of the five stations which registered the SD and temperature series, only three were within the catchment borders, and none was in the south-eastern part of the catchment (Fig. 1).

3.2.1. Hydrograph evaluation

The simulation results for all model variants described in Table 3 are presented in Figure 5. The simulations follow the observed hydrographs well for most of the high and low discharge events, however, none of the model variants fit the highest peaks with discharges above 120 m^3/s. The erroneous peaks are above the 98% quantile of the data; nonetheless, these peaks are very important features from an eco-hydrological point of view. The underestimation of the highest peaks was most probably due to rating curve uncertainty for the highest water level range, when the measurement profile may be 200-1000 m wide on a densely vegetated floodplain (Fig. 2). This is also justified by the fact that highest SWE in the catchment may not always correspond to the highest peak discharges (cfr. SWE in Fig. 4 with the observed discharge in Fig. 5).

The model variants using the SWE data (AE_DySno and GLOBSNOW) clearly underestimated the observed discharge (Fig. 5). The underestimation was also visible in the Observed SD model which use ρ to recalculate SD to SWE, but only during the 2008 flood event. In contrary, the SCF model variants (IMS and MOD10A1) and the Standard WetSpa simulated the observed discharge much better, especially during the falling and rising limbs and the low flow periods. The hydrograph of the IMS model variant presented a very good match with the observed discharge, what is in contrast to the low NS scores for the validation period (Tab. 5). This dichotomy between the IMS hydrographs and NS scores is due to the large overestimation of the flood event in 2010. The simulated peak discharge in 2010 was 429 m^3/s (not plotted in Fig. 5 for clarity), i.e. 2.9 times higher than the observed peak. Note that, in the calibration period, when no erroneous peaks were simulated, the IMS model variant has nearly the highest NS score (cfr. Fig. 5 and Tab. 5). The huge overestimation of the event in 2010 was because the snow presence in the IMS data is assumed as 100% SCF. As a result, during the winter 2009/2010 the snow cover was constantly 100% in the catchment, while SWE was decreasing (Fig. 4). This produced v_{sm} at the highest potential rate and overestimated the discharge.

The use of IMS and MOD10A1 snow products in the WetSpa model improved visually the simulated discharge when compared to the standard WetSpa. Moreover, the IMS and MOD10A1 with data-based switching model variants

Put Figure 5 here.

Figure 5:

performed better than Standard WetSpa in the peak discharge estimation and simulated fewer peaks that did not occur in the observed data (Fig. 5).

Needless to say, the drawback of using the data-based switching with MOD10A1 data is that the v_{sm} estimation is not fully independent from t and t_0(Tab. 3). Thus in this model structure the data-based switching gives only additional fuse that prevents generating v_{sm} if no decrease in SCF was observed between the consecutive time steps.

3.2.2. *Extended validation*

Although NS is one of the most popular efficiency measures used in hydrological modelling, it is subject to criticism (e.g. Schaefli and Gupta, 2007) mostly because of over-sensitivity to high discharge values (Legates and McCabe, 1999). In order to extend the validation of the discharge simulation additional efficiency measures were calculated (Fig. 6).

The overall model efficiency was quantified with KG, which showed a similar pattern of scores as NS (cfr. Fig. 6 and Tab. 5). Similarly like NS the data-based switching model variants had higher KG scores than the temperature-based variants given the same data source. However, three differences between KG and NS scores pattern can be distinguished. The IMS model variant (although still the lowest KG score) was evaluated much better in reference to the other model variants than it was the case for NS scores. This could be because KG do not normalize ME to σ_o as it is the case for NS (Gupta et al., 2009). Second difference is that MOD10A1 data-based switching model variant (altough still the highest KG score) was evaluated much better in reference to the other model variants than it was the case for NS scores. This confirms the good skill for runoff prediction of MOD10A1 data when data-based switching is used. The last difference is that GLOBSNOW with data-based switching and both Observed SD model variants had higher KG scores than the Standard WetSpa model, what was not the case when these models were compared in terms of NS scores. This points to the importance of the selection of a validation measure for model calibration and comparison as well as shows limitations of the here used methodology.

The low discharge error was quantified with the NS_{low}. The highest NS_{low} efficiencies were obtained by the Standard WetSpa, but the MOD10A1 with temperature-based switching has a similar score (Fig. 6, NS_{low} values). The model variants using remote sensing data performed similarly with NS_{low} in the range of 0.3 - 0.5. The worst performance with respect to NS_{low} was obtained by the Observed SD model variant. The NS_{low} scores demonstrate that the model variants updated with the remote sensing snow cover data performed worse than Standard WetSpa for the low flow simulation. However, the low flows occur in the study area in the summer half-year, i.e. during the snow-free season.

The error for high discharge simulation was quantified with NS_{high}. The highest NS_{high} efficiencies were obtained by the MOD10A1 with the data-based switching (Fig. 6, NS_{high} values). Similar NS_{high} scores were obtained by MOD10A1 (temperature-based switching), Observed SD, GLOBSNOW (both with data-based switching) and Standard WetSpa model variants. The NS_{high} were higher for the data-based switching than for the temperature-based switching in the model variants using external snow data sources. This is a strong indication that for simulation of snowmelt driven high discharges the data-based switching is more suited than the temperature-based approach when the degree-day method is used. This would not necessarily be the case for simulations with energy balance models, where energy fluxes would drive physical processes occurring in the snow pack. The IMS model variant received the lowest $NS_{high} = -0.99$ (not plotted in Fig. 6 for clarity), which is obviously due to the erroneous peak in the 2010 flood (Fig. 5). The NS_{high} values indicate that the high discharge simulation may be improved in comparison to Standard WetSpa when remote sensing, or in situ measurements of snow cover data is used. This was, however, not the case for the AE_DySno model variants and the temperature-based switching variant of GLOBSNOW and Observed SD.

All model variants, except IMS, had a negative bias (Fig. 6, ME values). Reason for this behaviour is that the models failed to simulate properly the peaks of the flood events above 120 m^3/s. The MOD10A1 data-based switching variant underestimated the observed discharge only for 6 m^3/s which is 4.3% of the data range. An even closer to zero bias was achieved by the IMS variant: 1.2 m^3/s (0.8%). The MOD10A1 with temperature-based switching and Standard WetSpa underestimated the observed discharge at similar level, about 7 m^3/s (5.0%). The SWE and SD model variants performed worse than Standard WetSpa in terms of bias (Fig. 6). The relatively high bias values suggest that remote sensing SWE and observed SD recalculated to water equivalent based on k_{cor} are not adequate for snowmelt modelling in WetSpa. Another reason for the negative bias could be that the passive microwave data underestimate snow cover compared to visible / near infra-red data (Armstrong and Brodzik, 2001, 2002).

The r^2 values presented in Fig. 6 do not account for the magnitude of error between simulated and observed daily values, but shows the collinearity of the two variables. The collinearity, in this case, informs weather the day to day variations in observed discharge are followed by the simulated discharge (numerator in Eq. 10). In other words, r^2 is related to the accuracy of timing in the simulated discharge.

The r^2 values, similarly like NS_{high}, were higher for the data-based switching variants of the data source than for the corresponding temperature-based switching (Fig. 6). Note that between the MOD10A1 model variants the difference was only marginal in advantage of data-based switching. This decrease of r^2 means that the temperature-based switching deforms the snowmelt pattern that is achieved from the data alone.

The highest r^2 was obtained for the GLOBSNOW data-based switching model variant. The discrepancy between high r^2 and lower values of the other accuracy measures may be due to strong underestimation of simulated discharge in this model variant (Fig. 6). It is worth mentioning that the r^2 values were

Put Figure 6 here.

Figure 6:

considerably higher for GLOBSNOW than for the second SWE model variant - AE_DySno. This is in agreement with Hancock et al. (2013), who demonstrated that GLOBSNOW performs better in SWE timing than AE_DySno. Additional correction of GLOBSNOW or AE_DySno data could allow obtaining better simulation results than currently presented. This could be achieved by assimilation techniques, as a Kalman filter presented by Andreadis and Lettenmaier (2006) for AMSR-E and MODIS data.

3.3. Uncertainty analysis

Figure 7 present the 95% confidence intervals, under the assumption that $NS > 0.6$, i.e. behavioural models, for the one year simulation period. Observed discharge is within the confidence interval for more than 90% of the time for all model variants except the MOD10A1 with the temperature-based switching (Tab. 6). This model variant has relatively narrower confidence interval than other variants, in particular for low discharge periods from August to October (Fig. 7). This means that MOD10A1 with temperature-based switching simulates low discharges with heavy negative bias resulting in unrealistic confidence interval estimation. Similar, but not as big bias is observed in MOD10A1 with data-based switching. Another example of substantially different confidence intervals than the observed discharge is Standard WetSpa (Tab. 6). The confidence intervals estimated in Standard WetSpa shows large deviation from the observed discharge in the low flow period: December to January 2009. Remaining model variants had the lower confidence bound close to the observed discharge suggesting that this event was problematic to simulate in each case. Nonetheless, the deviation from the observed discharge was not as big as in Standard WetSpa.

The model variants which implemented both switching algorithms had the narrower confidence intervals at the peak of flood events simulated for data-

Table 6:

Put Table 6 here.

based switching (Fig. 7). This was also the case for MOD10A1 temperature-based switching model variant which has very narrow confidence interval. Nevertheless, the confidence interval during the peak discharge was narrower in the MOD10A1 data-based switching variant. Beside that no relation between the model structure (switching method) and the confidence interval fitting the observation was observed (Tab. 6).

The snow cover uncertainty results are presented in Figure 8. The 95% confidence interval for SWE simulations in the Standard WetSpa covers in majority the SWE series obtained with other data sources. This suggests that the snow accumulation and snowmelt algorithm in the Standard WetSpa (Sect. 2.5.2) works properly. The 95% confidence interval for SCF from the Standard WetSpa also covers well the other SCF series (Fig. 8). In this case the confidence interval fit the IMS SCF much better than MOD10A1 SCF. This is due to the SCF calculation method in WetSpa, which gives 100% SCF when in a Thiessen polygon (i.e. in a representative area for a meteorological station) SWE$>$0 . Even though 25 meteorological stations are used in the model, the simulated SCF at the catchment scale is not continuous and resembles better the discrete series of IMS than MOD10A1 SCF.

The uncertainty analysis could have been conducted using another efficient method for uncertainty analysis, such as Differential Evolution Adaptive Metropolis (DREAM; Vrugt et al., 2008). DREAM, however, was not used due two reasons. First, the GLUE method is easy to apply while giving reasonable uncertainty estimation (Vrugt et al., 2009). Second, the purpose of the uncertainty analysis in this paper was only to highlight the interference of the different data sources and model structures with the uncertainty. The latter was also a reason for relatively short, one year simulation period selection.

Put Figure 7 here.

Figure 7:

Put Figure 8 here.

Figure 8:

3.3.1. Behavioural parameters distribution

In this section the distributions of behavioural parameters related to snow processes (k_{snow}, k_{rain}, t_0, k_{cor} and ρ) and the groundwater recession coefficient (k_g) are discussed. The k_g is selected for the analysis, because it is used in all model variants and is related to groundwater, which is the dominant discharge component in the study catchment. The distributions of the behavioural parameters, presented in Figure 9, highlights two general issues: (1) different data sources give different distributions for the same model structure, and (2) different model structures change the distributions within the same data source.

The first issue is clearly visible on the example of AE_DySno GLOBSNOW model variants. The variants use exactly the same model structures, but the snow data sources are different. As a result the behavioural distributions for k_{cor} are completely different between the data sources, while similar between the switching variants for the same data source (Fig. 9). This need not be the cases for all parameters, since the t_0 distributions are similar for AE_DySno and GLOBSNOW variants with temperature-based switching. On the other hand, the Standard WetSpa and the SCF models (MOD10A1, IMS) had different model structures but produce similar, wide behavioural parameters distributions (k_{snow}, k_{rain} in Fig. 9). This may be because of high uncertainty related to these parameters or model structures or data sources.

The second issue is clear when comparing the temperature- and data-based switching model variants for the same data source. For the example of ρ, k_{cor} and k_g (except MOD10A1) presented in Figure 9, it is clear that the data-based switching narrows the behavioural parameters distributions in comparison to the temperature-based switching, i.e. different model structure influence the uncertainty with the same data source. This effect is best visible with the k_g parameter, which is used in all model variants. Median of the behavioural parameters distributions is at similar level in all model variants, but the distributions are much narrower for the data-based switching. The exception is MOD10A1, which has a wider distribution with the data-based switching than with t_0. This is a result of over-parametrization in the model structure, i.e. both temperature and SCF are responsible for the generation of snowmelt (as described in Sect. 2.5.2). Thus, wide representation of global WetSpa parameters can produce behavioural models in the MOD10A1 data-based switching variant.

The Standard WetSpa model variant have wide behavioural parameters distribution for the snow related parameters used in the degree-day method: t_0, k_{snow}, k_{rain} (Fig. 9). Other model variants using the degree-day method (and the same parameters) have narrower behavioural distributions, especially for t_0. This may be due to the constant values of the degree-day method parameters. As a result the degree-day parameters of the Standard WetSpa model may not represent physical values, but just the best fit that distributes snowmelt over whole simulation period. On the other hand, Standard WetSpa has narrow behavioural distribution of k_g when compared to most of the variants. So the uncertainty of the degree-day parameters is balanced by the behavioural distributions of other global parameters, as in this case k_g.

Finally, worth noticing is the behavioural distribution of t_0 for IMS. The t_0 parameter is responsible for timing and magnitude of snowmelt and with the high quality of the IMS data (Fig. 3) the resultant behavioural distribution is the narrowest of all compared model variants.

It is important to notice, that the behavioural distribution of parameters (as well as the values of parameters obtained during the calibration) responsible for simulation of snow processes in the degree-day based modes (i.e. k_{snow}, k_{rain} and t_0) may be affected by snowfall undercatch. The rain gauge measurements were not corrected for this phenomenon. Hence, these parameters may be biased in order to reflect underestimated snowfall measurements.

Put Figure 9 here.

Figure 9: .

4. Conclusions

Nowadays, more and more distributed remote sensing time series are becoming available and replace the standard, in-situ measurements. It is of interest to see whether hydrological models can benefit (e.g. by improved calibration) from the remote sensing data when compared to the in-situ data. The aim of this paper was to assess how snow cover data (AE_DySno, GLOBSNOW, IMS, MOD10A1 and observed SD) can influence the model in terms of calibration quality and uncertainty analysis, i.e. to assess the skill of the snow products including insights into the role of the model structure. The tested model structures had two different rules for snowmelt and accumulation switching. The data-based switching was based on the day to day changes in the snow cover, while the temperature-based switching was based on the threshold for mean daily temperature.

The high resolution data (MODIS and IMS) had higher agreement with observed SD than the low resolution, passive microwave data (GLOBSNOW and AE_DySno). The accuracy of the snow products was reflected on the hydrological model calibration results.

The calibration results have demonstrated that using MOD10A1 remote sensing snow cover data as a driver of snow processes in the WetSpa hydrological model improves the validation NS scores in comparison to the standard WetSpa model, not using remote sensing data. Potential improvements could be also achieved with the use of IMS data. In this case, an additional processing would be needed to prevent simulating erroneous peak discharges when SCF in a catchment is 100%, while the corresponding SWE already started to decrease. With the undertaken methodology the GLOBSNOW and AE_DySno SWE products could not achieve the validation NS scores any better than Standard WetSpa.

The validation of the calibration results was extended with the use of additional measures quantifying: overall efficiency (KG), low flows (NS_{low}), high flows (NS_{high}), bias (ME) and collinearity (r^2). The overall efficiency depicts better performance of GLOBSNOW, MOD10A1 and Observed SD than the standard WetSpa model. Moreover, use of the GLOBSNOW, MOD10A1 and Observed SD data sources results in superior simulations of high flows in comparison to the Standard WetSpa, but at the cost of reduced low flow performance. Indeed, the Standard WetSpa was the best in low flow simulations of all compared model variants. However, it has to be mentioned, that low flows in the study area occur in summer half-year, when the influence of antecedent snow is negligible. In terms of bias, the best results were obtained by the model variants using MOD10A1 and IMS data. The SWE and observed SD model variants were strongly biased. This suggests the chosen method for SWE data assimilation using a stationary correction factor (k_{cor}) was too simplistic to achieve correct simulations. However, the GLOBSNOW data has a potential for providing better hydrological simulations since the simulated discharge has the strongest collinearity with the observed discharge of all model variants.

This study demonstrated also the role of model structure in the WetSpa behaviour using different sources of snow cover data. The model variants which implemented both switching methods performed better for the data-based switching in terms of KG, NS, NS_{high} and r^2. This means that using temperature as a tuning variable for snowmelt timing from snow cover data results in decreased model performance. Furthermore, the role of the model structure is best visible in the uncertainty analysis results. Not only the model variants using the same data source had different behavioural parameters distributions with different switching algorithm (mostly narrower for the data-based switching; Fig. 9), but the width of the simulation confidence intervals at the highest flow event was narrowed when the data-based switching was used instead of the temperature-based switching (Fig. 7).

The uncertainty analysis reveals also that under the selected threshold for behavioural models ($NS > 0.6$) the confidence interval in all variants covered the observations similarly. The pattern of the confidence interval was similar in the summer storms periods, while it was different in the snow accumulation periods: models using SCF data sources simulated more false positive peaks than the remaining model variants.

To conclude, each of the tested remote sensing data sources has positive and negative influences on the results of discharge simulations in the WetSpa model. Some data sources, like AE_DySno (low resolution, passive microwave), have more negative influence, which possibly could be overcome when other model structures or additional pre-processing would be used. While other data sources, like MOD10A1 (high resolution, visual / near-infrared), show more positive influences on the modelling results and proved to give better calibration than the standard WetSpa model. Needless to say, the model structure has a huge influence on the calibration and the uncertainty of the simulations as it can considerably change the behavioural parameters distributions and the widths of the confidence intervals.

For future research it would be interesting to conduct similar experiments with other hydrological models like SWAT or MIKE SHE and to extend the uncertainty analysis. It would be also of interest to conduct similar analysis with using a set of various efficiency measures for calibrating the models. Another possibility would be to repeat the experiment on a broader range of study sites covering different catchment areas and climatic conditions. Finally, interactions between frozen soil and snowmelt runoff as well as a distributed degree-day coefficient could be tested in the future since these features are currently lacking in WetSpa.

5. Acknowledgement

First author acknowledge Jiri Nossent for his inspiring discussions on the uncertainty analysis topics and Ignacy Kardel for sharing the DEM and sources for the soil map used in this study. First author also acknowledges the Flemish Government for supporting his research visit to the Vrije Universiteit Brussel.

The Hydro-meteorological data was provided by Institute of Meteorology and Water Management - National Research Institute (IMGW).

Andreadis, K. M., Lettenmaier, D. P., 2006. Assimilating remotely sensed snow observations into a macroscale hydrology model. Advances in Water Resources 29 (6), 872–886.

Armstrong, R. L., Brodzik, M. J., 2001. Recent northern hemisphere snow extent: A comparison of data derived from visible and microwave satellite sensors. Geophys. Res. Lett. 28 (19), 3673–3676.

Armstrong, R. L., Brodzik, M. J., 2002. Hemispheric-scale comparison and evaluation of passive-microwave snow algorithms. Annals of Glaciology 34 (1), 38–44.

Bavera, D., De Michele, C., Pepe, M., Rampini, A., 2012. Melted snow volume control in the snowmelt runoff model using a snow water equivalent statistically based model. Hydrol. Process. 26 (22), 3405–3415.

Berezowski, T., Chormański, J., Batelaan, O., Canters, F., Van de Voorde, T., 2012. Impact of remotely sensed land-cover proportions on urban runoff prediction. International Journal of Applied Earth Observation and Geoinformation 16, 54–65.

Berezowski, T., Nossent, J., Chormański, J., Batelaan, O., 2014. Spatial sensitivity analysis of snow cover data in a distributed rainfall-runoff model. Hydrol. Earth Syst. Sci. Discuss. 11, 11987–12025.

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Process. 6 (3), 279–298.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. Journal of Hydrology 249 (1-4), 11–29.

Bormann, K. J., Westra, S., Evans, J. P., McCabe, M. F., 2013. Spatial and temporal variability in seasonal snow density. Journal of Hydrology 484 (0), 63–73.

Brodzik, M. J., Armstrong, R. L., 2013. Northern Hemisphere EASE-Grid 2.0 Weekly Snow Cover and Sea Ice Extent. Version 4.

Brown, C. D., Davis, H. T., 2006. Receiver operating characteristics curves and related decision measures: A tutorial. Chemometrics and Intelligent Laboratory Systems 80 (1), 24–38.

Butt, M. J., Bilal, M., 2011. Application of Snowmelt Runoff Model for water resource management. Hydrol. Process. 25 (24), 3735–3747.

Byun, K., Choi, M., 2014. Uncertainty of snow water equivalent retrieved from AMSR-E brightness temperature in northeast Asia. Hydrol. Process. 28 (7), 3173–3184.

Chang, A., Foster, J., Hall, D. K., 1987. Nimbus 7 SMMR derived global snow cover parameters. Ann. Glaciology 9, 39–44.

Chen, C., Lakhankar, T., Romanov, P., Helfrich, S., Powell, A., Khanbilvardi, R., 2012. Validation of NOAA-Interactive Multisensor Snow and Ice Mapping System (IMS) by comparison with ground-based measurements over continental United States. Remote Sensing 4 (5), 1134–1145.

Chormański, J., Miroslaw-Swiatek, D., Michalowski, R., 2009. A hydrodynamic model coupled with gis for flood characteristics analysis in the biebrza riparian wetland. Oceanological and Hydrobiological Studies 38 (1), 65–73.

Chormański, J., Okruszko, T., Ignar, S., Batelaan, O., Rebel, K., Wassen, M., 2011. Flood mapping with remote sensing and hydrochemistry: a new method to distinguish the origin of flood water during floods. Ecological Engineering 37 (9), 1334–1349.

Chormański, J., Van de Voorde, T., De Roeck, T., Batelaan, O., Canters, F., 2008. Improving distributed runoff prediction in urbanized catchments with remote sensing based estimates of impervious surface cover. Sensors 8 (2), 910–932.

De Smedt, F., Liu, Y. B., Gebremeskel, S., 2000. Risk Analysis II. WTI press, Southampton, Boston, Ch. Hydrologic modeling on a catchment scale using GIS and remote sensed land use information, pp. 295 – 304.

Doherty, J., 2010. PEST, Model-Independent Parameter Estimation-User Manual, 5th ed., with slight additions. Watermark Numerical Computing, Brisbane, Australia.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28 (4), 1015–1031.

Gao, Y., Xie, H., Lu, N., Yao, T., Liang, T., 2010a. Toward advanced daily cloud-free snow cover and snow water equivalent products from Terra-Aqua MODIS and Aqua AMSR-E measurements. Journal of Hydrology 385 (1-4), 23–35.

Gao, Y., Xie, H., Yao, T., Xue, C., 2010b. Integrated assessment on multi-temporal and multi-sensor combinations for reducing cloud obscuration of MODIS snow cover products of the Pacific Northwest USA. Remote Sensing of Environment 114 (8), 1662–1675.

Grygoruk, M., Mirosław-Świątek, D., Chrzanowska, W., Ignar, S., 2013. How much for water? economic assessment and mapping of floodplain water storage as a catchment-scale ecosystem service of wetlands. Water 5 (4), 1760–1779.

Gupta, H. V., Kling, H., Yilmaz, K. K., Martinez, G. F., 2009. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377 (1-2), 80–91.

Hall, D. K., Riggs, G. A., Salomonson, V. V., 1995. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. Remote Sensing of Environment 54 (2), 127–140.

Hall, D. K., Riggs, G. A., Salomonson, V. V., 2006. MODIS/Terra Snow Cover Daily L3 Global 500m Grid V005. Digital media.

Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., Bayr, K. J., 2002. MODIS snow-cover products. Remote Sensing of Environment 83 (1-2), 181–194.

Hancock, S., Baxter, R., Evans, J., Huntley, B., 2013. Evaluating global snow water equivalent products for testing land surface models. Remote Sensing of Environment 128 (0), 107–117.

Helfrich, S. R., McNamara, D., Ramsay, B. H., Baldwin, T., Kasheta, T., 2007. Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). Hydrol. Process. 21 (12), 1576–1586.

Ignar, S., Maksymiuk-Dziuban, A., Miroslaw-Siatek, D., Chormański, J., Okruszko, T., Wysocki, P., 2011. Temporal variability of the selected flood parameters in the biebrza river valley. Annals of Warsaw University of Life Sciences - SGGW, Land Reclamation 2 (43), 135–142.

Kelly, R., Chang, A., Tsang, L., Foster, J., 2003. A prototype AMSR-E global snow area and snow depth algorithm. Geoscience and Remote Sensing, IEEE Transactions on 41 (2), 230–242.

Klein, A. G., Barnett, A. C., 2003. Validation of daily MODIS snow cover maps of the Upper Rio Grande River Basin for the 2000-2001 snow year. Remote Sensing of Environment 86 (2), 162 – 176.

Klein, A. G., Hall, D. K., Riggs, G. A., 1998. Improving snow cover mapping in forests through the use of a canopy reflectance model. Hydrol. Process. 12 (10-11), 1723–1744.

Legates, D. R., McCabe, G. J., 1999. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. Water Resources Research 35 (1), 233–241.

Liang, T., Zhang, X., Xie, H., Wu, C., Feng, Q., Huang, X., Chen, Q., 2008. Toward improved daily snow cover mapping with advanced combination of MODIS and AMSR-E measurements. Remote Sensing of Environment 112 (10), 3750–3761.

Liston, G. E., 1999. Interrelationships among snow distribution, snowmelt, and snow cover depletion: Implications for atmospheric, hydrologic, and ecologic modeling. Journal of Applied Meteorology 38 (10), 1474–1487.

Liu, Y. B., De Smedt, F., 2004. WetSpa Extension, A GIS-based Hydrologic Model for Flood Prediction and Watershed Management. Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel.

Liu, Y. B., Gebremeskel, S., De Smedt, F., Hoffmann, L., Pfister, L., 2003. A diffusive transport approach for flow routing in GIS-based flood modeling. Journal of Hydrology 283 (1-4), 91 – 106.

McKay, M., Beckman, R., Conover, W., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.

Molotch, N. P., Margulis, S. A., 2008. Estimating the distribution of snow water equivalent using remotely sensed snow cover data and a spatially distributed snowmelt model: A multi-resolution, multi-sensor comparison. Advances in Water Resources 31 (11), 1503–1514.

Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part I - A discussion of principles. Journal of Hydrology 10 (3), 282–290.

National Ice Center, 2008. IMS daily Northern Hemisphere snow and ice analysis at 4 km and 24 km resolution.

Onuchin, A., Burenina, T., 1996. Climatic and geographic patterns in snow density dynamics, northern eurasia. Arctic and Alpine Research 28 (1), 99–103.

Parajka, J., Blöschl, G., 2006. Validation of MODIS snow cover images over Austria. Hydrology and Earth System Sciences 10 (5), 679–689.

Parajka, J., Blöschl, G., 2008. The value of modis snow cover data in validating and calibrating conceptual hydrologic models. Journal of Hydrology 358 (3-4), 240–258.

Pulliainen, J., 2006. Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. Remote Sensing of Environment 101 (2), 257–269.

Schaefli, B., Gupta, H., 2007. Do Nash values have value? Hydrological Processes 21, 2075–2080.

Şensoy, A., Uysal, G., 2012. The value of snow depletion forecasting methods towards operational snowmelt runoff estimation using MODIS and numerical weather prediction data. Water Resources Management 26 (12), 3415–3440.

Shrestha, M., Wang, L., Koike, T., Tsutsui, H., Xue, Y., Hirabayashi, Y., 2014. Correcting basin-scale snowfall in a mountainous basin using a distributed snowmelt model and remote-sensing data. Hydrology and Earth System Sciences 18 (2), 747–761.

Smakhtin, V. Y., Sami, K., Hughes, D. A., 1998. Evaluating the performance of a deterministic daily rainfall-runoff model in a low-flow context. Hydrol. Process. 12 (5), 797–812.

Smith, P., Beven, K. J., Tawn, J. A., 2008. Informal likelihood measures in model assessment: Theoretic development and investigation. Advances in Water Resources 31 (8), 1087–1100.

Şorman, A. A., Şensoy, A., Tekeli, A. E., Şorman, A. U., Akyürek, Z., 2009. Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of Turkey. Hydrol. Process. 23 (7), 1031–1040.

Şorman, A. U., Akyürek, Z., Şensoy, A., Şorman, A. A., Tekeli, A. E., 2007. Commentary on comparison of MODIS snow cover and albedo products with ground observations over the mountainous terrain of Turkey. Hydrology and Earth System Sciences 11 (4), 1353–1360.

Stachý, J., 1987. Hydrological Atlas of Poland. Vol. 1. Wydawnictwo Geologiczne, in Polish with English translation.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., Koskinen, J., Bojkov, B., 2011. Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. Remote Sensing of Environment 115 (12), 3517–3529.

Tedesco, M., Kelly, R. E. J., Foster, J. L., Chang, A. T. C., 2004. AMSR-E/Aqua Daily L3 Global Snow Water Equivalent EASE-Grids V002.

Udnaes, H. C., Alfnes, E., Andreassen, L. M., 2007. Improving runoff modelling using satellite-derived snow covered area? Nordic Hydrology 38 (1), 21–32.

Verbeiren, B., Van De Voorde, T., Canters, F., Binard, M., Cornet, Y., Batelaan, O., 2013. Assessing urbanisation effects on rainfall-runoff using a remote sensing supported modelling strategy. International Journal of Applied Earth Observation and Geoinformation 21 (0), 92–102.

Vrugt, J. A., Braak, C. J. F., Gupta, H. V., Robinson, B. A., 2009. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? 23 (7), 1011–1026.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., Robinson, B. A., 2008. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resour. Res. 44 (12), W00B09.

Wang, J., Tedesco, M., 2007. Identification of atmospheric influences on the estimation of snow water equivalent from AMSR-E measurements. Remote Sensing of Environment 111 (2-3), 398–408.

Wang, Z.-M., Batelaan, O., Smedt, F. D., 1996. A distributed model for water and energy transfer between soil, plants and atmosphere (WetSpa). Physics and Chemistry of The Earth 21 (3), 189 – 193.

Yan, F., Ramage, J., McKenney, R., 2009. Modeling of high-latitude spring freshet from amsr-e passive microwave observations. Water Resour. Res. 45 (11), W11408.

Yatheendradas, S., Lidard, C. D. P., Koren, V., Cosgrove, B. A., De Goncalves, L. G. G., Smith, M., Geiger, J., Cui, Z., Borak, J., Kumar, S. V., Toll, D. L., Riggs, G., Mizukami, N., 2012. Distributed assimilation of satellite-based snow extent for improving simulated streamflow in mountainous, dense forests: An example over the DMIP2 western basins. Water Resour. Res. 48 (9), W09557.

Younger, P. M., Freer, J. E., Beven, K. J., 2009. Detecting the effects of spatial variability of rainfall on hydrological modelling within an uncertainty analysis framework. Hydrological Processes 23 (14), 1988–2003.

- Figure1: Map showing the true colour satellite image of the study area. The red and green points indicate the meteorological stations from which the data was used in this study. At the climatological stations (red) the precipitation, temperature and snow depth data were measured, while the precipitation stations provide the precipitation data only. The three labelled stations were used to conduct the snow data accuracy assessment described in Sect. 2.3.

- Figure 2: A look over the lower basin of the Biebrza River: during the spring flood (left) and no flood moment (right) in the year 2007. Courtesy of Sylwia Szporak-Wasilewska.

- Figure 3: ROC curves presenting the snow products quality with respect to the observed SD after reclassification to binary values (Eq. 3).

- Figure 4: Comparison of the catchment averaged snow products used with the simulated SWE (top) and SCF (bottom) by the Standard WetSpa model variant.

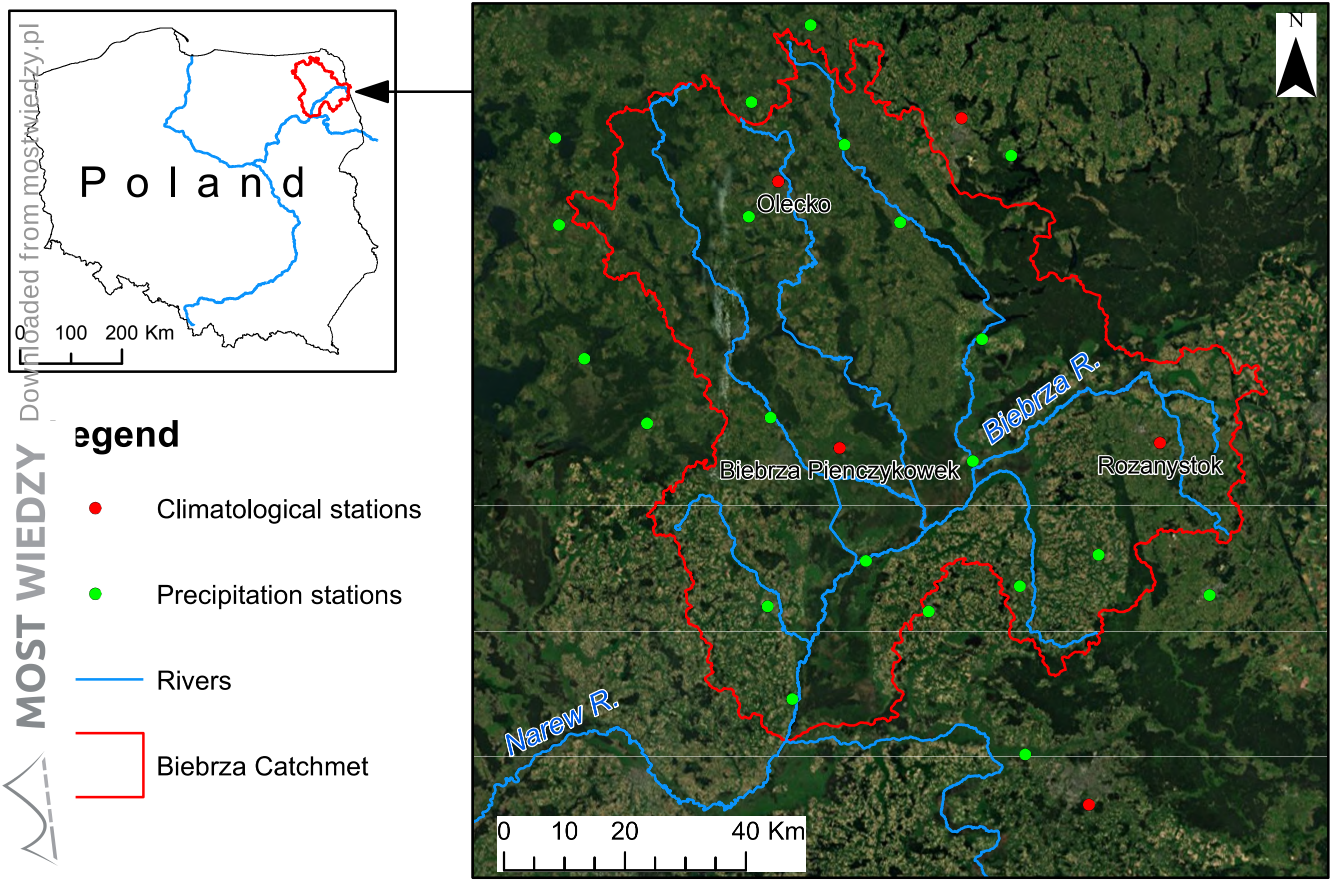
- Figure 5: The best simulated discharge hydrographs for each model variant selected based on validation results (see Tab. 5) of the three calibration runs. The hydrographs present the whole simulation period (calibration and validation) from 1^{st} November 2004 to 31^{st} December 2010.

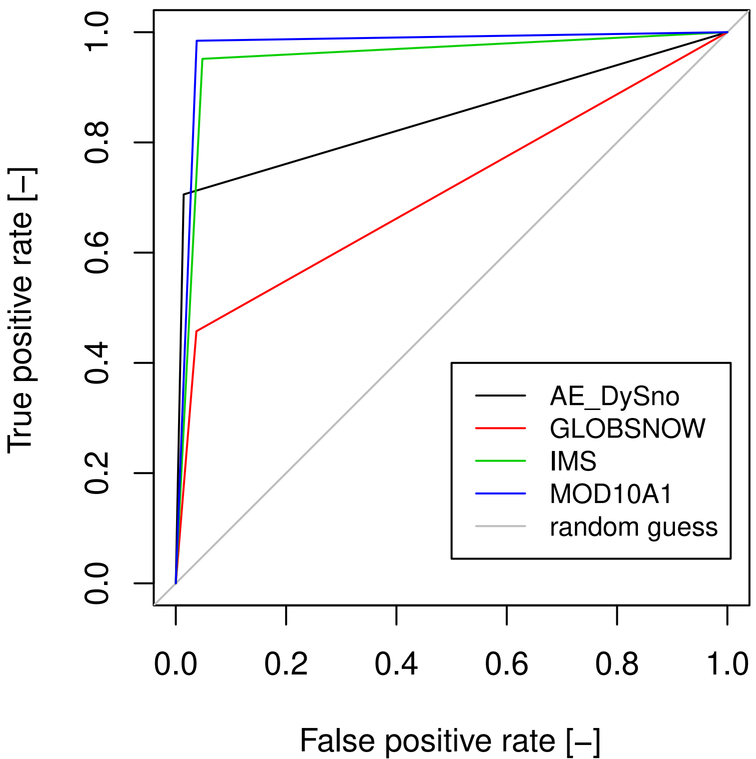
- Figure 6: Validation of the best hydrological model variants calibrated with NS using different criteria, from top to bottom: KG, NS_{low}, NS_{high}, ME and r^2. The dashed vertical lines separates the different model variants; the dotted horizontal line in the ME plot is $ME = 0$ m^3/s, i.e. no bias. The value of IMS $NS_{high} = -0.99$ is not plotted for clarity. The labels in the bottom axis presents the dataset used in the model variant and the switching method (temperature- or data-based)

- Figure 7: Discharge uncertainty obtained wit the GLUE method. The threshold for behavioural parameters set was $NS > 0.6$ and the confidence interval was 95% i.e. lower and upper dashed lines show the 2.5^{th} and 97.5^{th} quantiles.

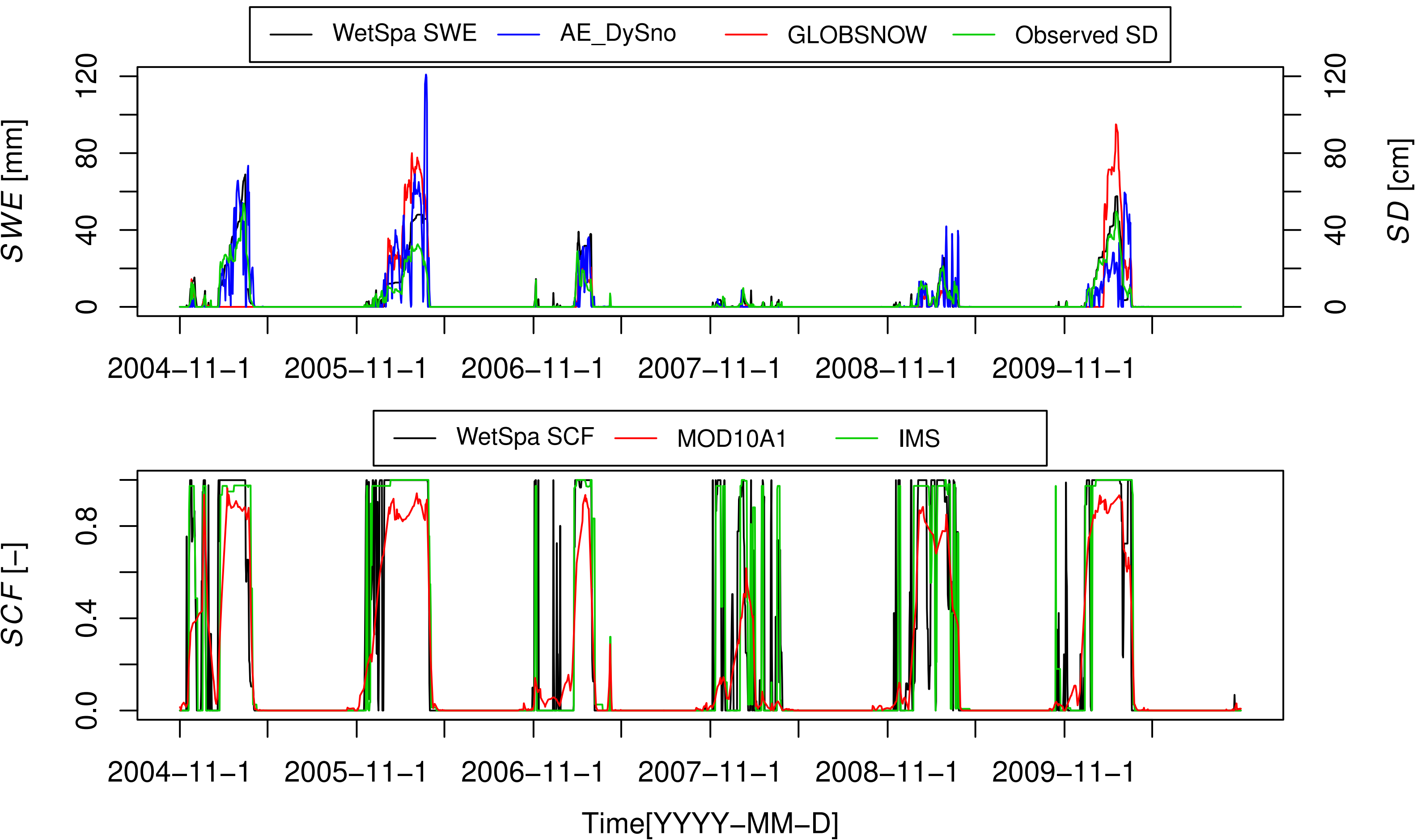
- Figure 8: Comparison of the 95% confidence interval estimated with the GLUE method for SWE (top) and SCF (bottom) in the Standard WetSpa model variant with the SWE and SCF from the snow cover data used in the study. The presented time series are catchment averaged. The observed SD from the meteorological stations was recalculated to SWE with $\rho = 0.95$ mm water/cm, i.e. the median ρ from the behavioural distribution in the Observed SD data-based switching model variant (Fig. 9).

- Figure 9: Behavioural parameters distribution for, from top to bottom: k_g, t_0, k_{snow}, k_{rain}, k_{cor} and ρ for the model variants which use these parameters (Tab. 3); k_g is the global WetSpa parameter used in all model variants. The labels in the bottom axis presents the dataset used in the model variant and the switching method (temperature- or data-based).

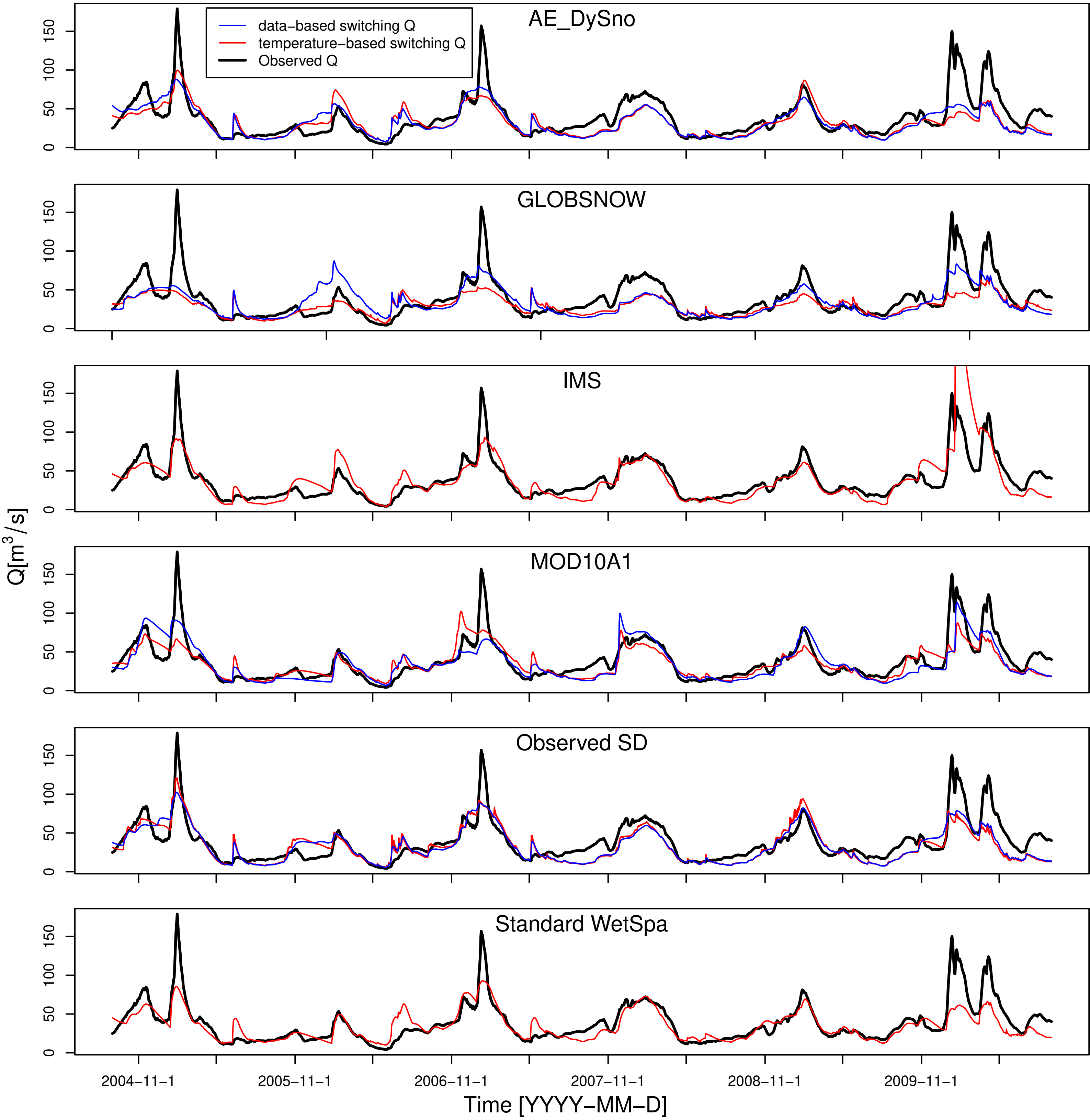

Poland
0 100 200 Km
egend
Climatological stations
Precipitation stations
Rivers
Biebrza Catchmet
N
Olecko
Biebrza R.
Biebrza Pienczykowek
Rozanystok
Narew R.
0 10 20 40 Km



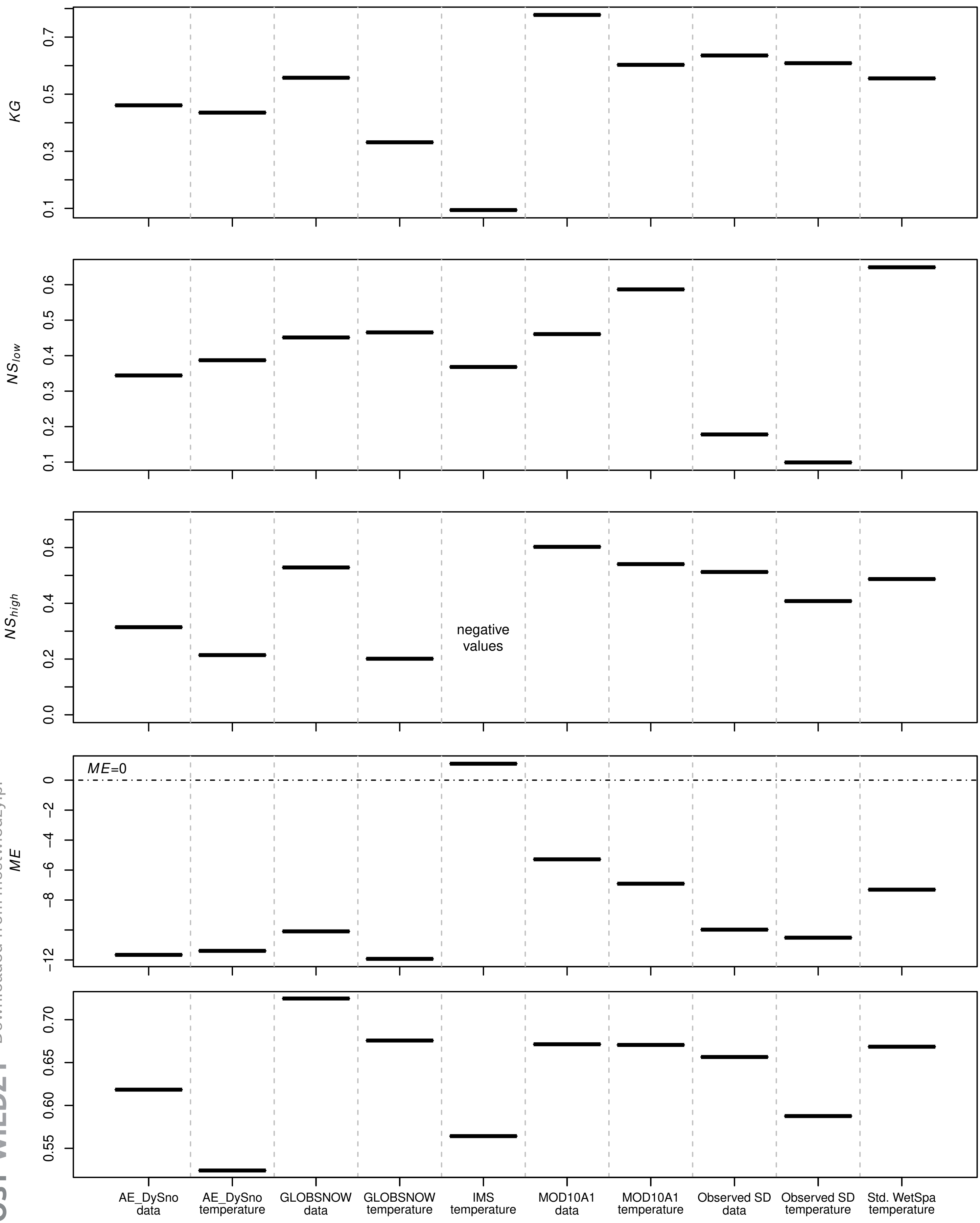
1.0
0.8
0.6
0.4
0.2
0.0
True positive rate [–]
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate [–]
AE_DySno
GLOBSNOW
IMS
MOD10A1
random guess

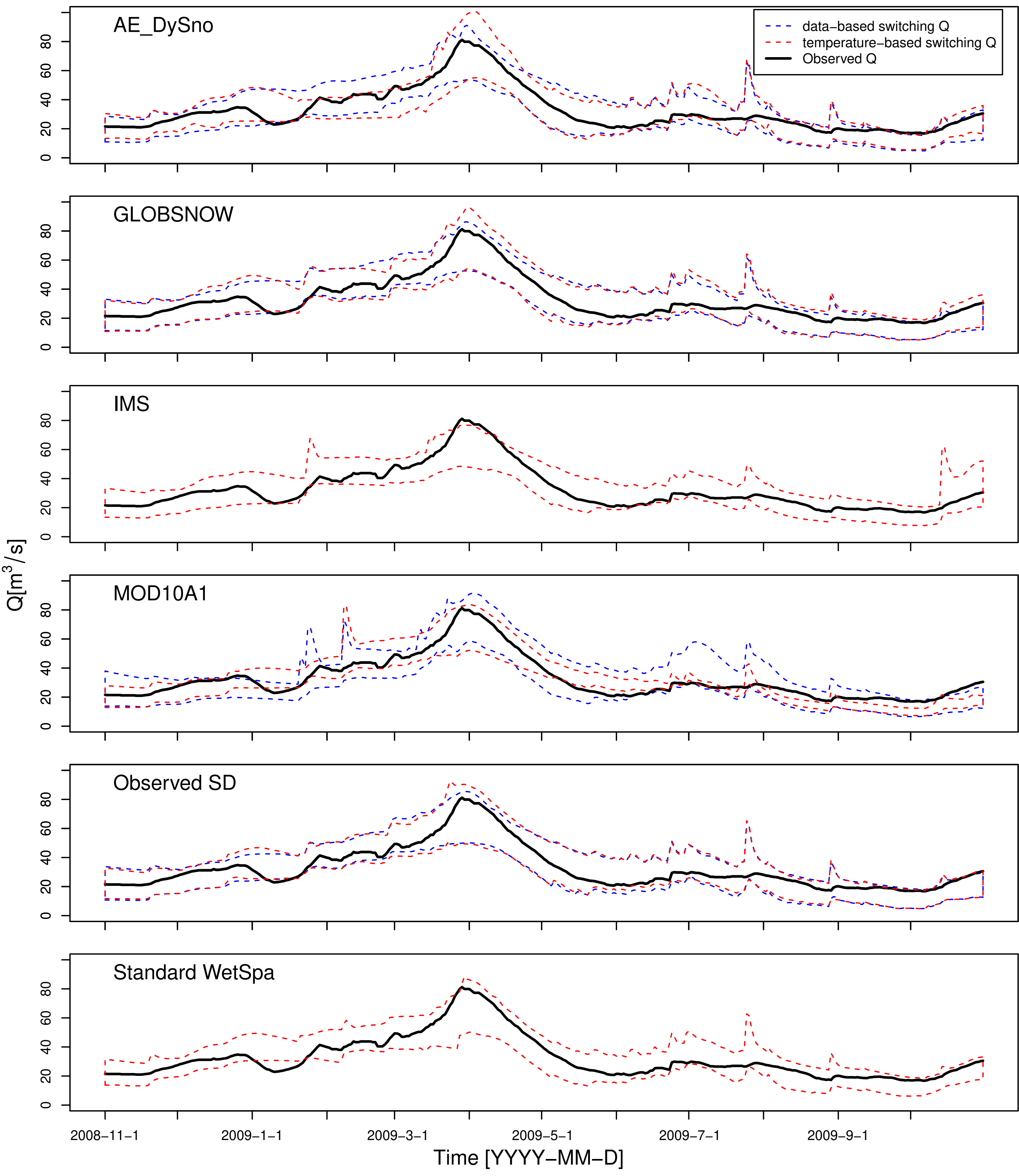


WetSpa SWE
AE_DySno
GLOBSNOW
Observed SD
SWE [mm]
SD [cm]
0
40
80
120
2004−11−1
2005−11−1
2006−11−1
2007−11−1
2008−11−1
2009−11−1
WetSpa SCF
MOD10A1
IMS
SCF [−]
0.0
0.4
0.8
Time[YYYY−MM−D]

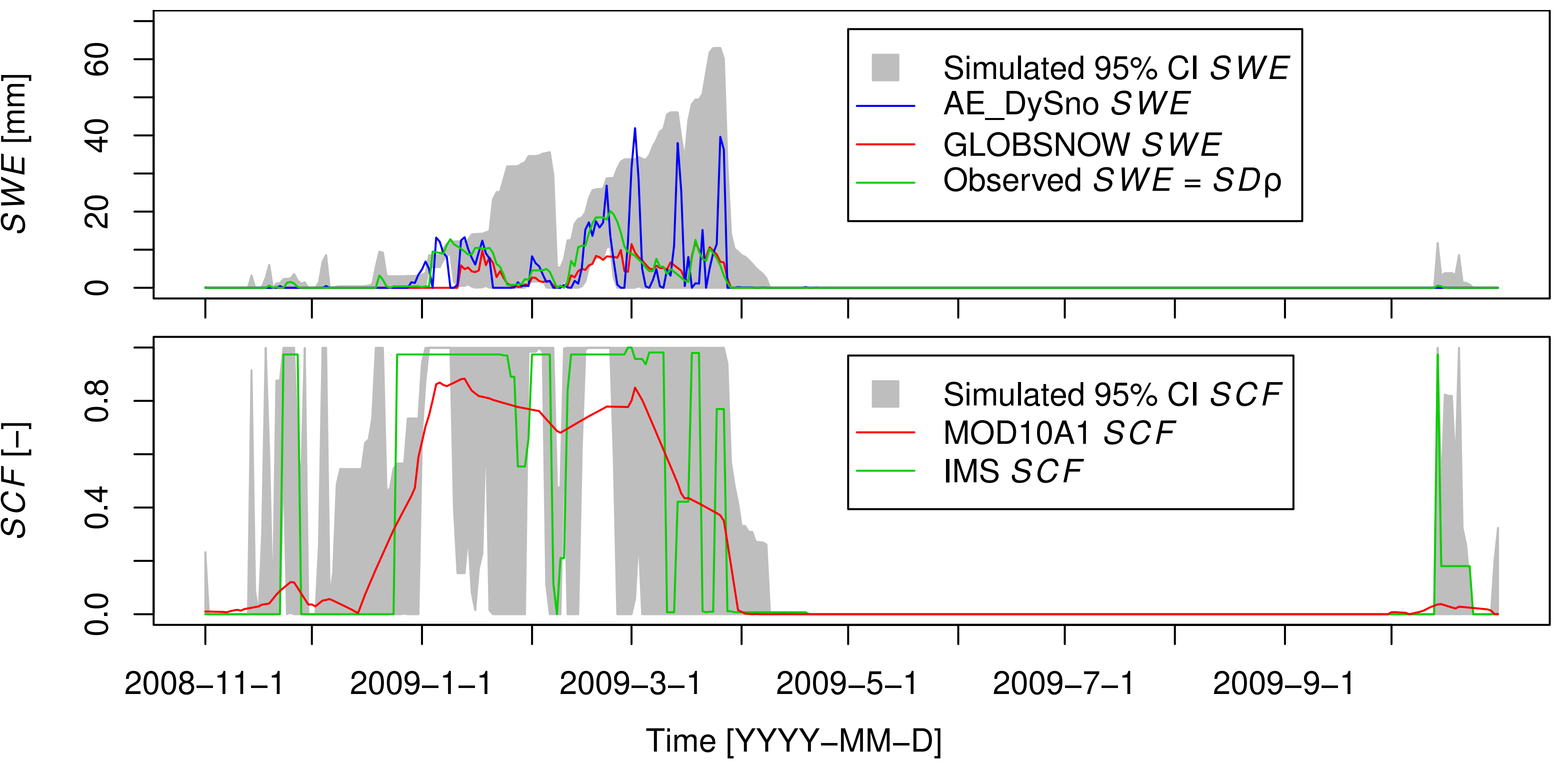


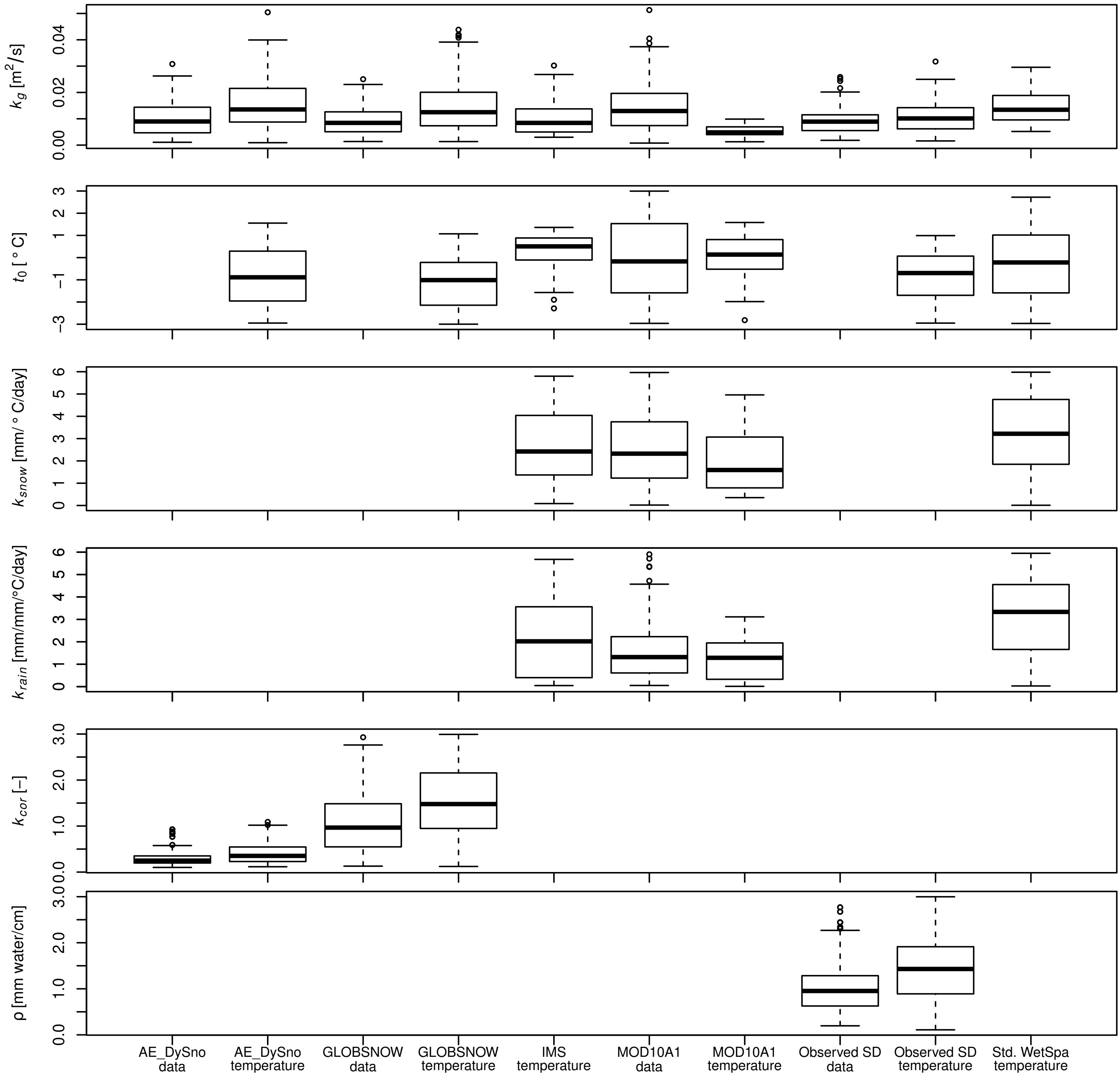
AE_DySno
data−based switching Q
temperature−based switching Q
Observed Q
GLOBSNOW
IMS
MOD10A1
Observed SD
Standard WetSpa
Q[m^3/s]
0
50
100
150
2004−11−1
2005−11−1
2006−11−1
2007−11−1
2008−11−1
2009−11−1
Time [YYYY−MM−D]


KG
0.1
0.3
0.5
0.7
NS_low
0.1
0.2
0.3
0.4
0.5
0.6
NS_high
0.0
0.2
0.4
0.6
negative values
ME
ME=0
0
−2
−4
−6
−8
−12
0.55
0.60
0.65
0.70
AE_DySno data
AE_DySno temperature
GLOBSNOW data
GLOBSNOW temperature
IMS temperature
MOD10A1 data
MOD10A1 temperature
Observed SD data
Observed SD temperature
Std. WetSpa temperature



AE_DySno
data−based switching Q
temperature−based switching Q
Observed Q
GLOBSNOW
IMS
MOD10A1
Observed SD
Standard WetSpa
Q[m^3/s]
0
20
40
60
80
2008−11−1
2009−1−1
2009−3−1
2009−5−1
2009−7−1
2009−9−1
Time [YYYY−MM−D]


Simulated 95% CI SWE
AE_DySno SWE
GLOBSNOW SWE
Observed SWE = SDρ
SWE [mm]
0
20
40
60
Simulated 95% CI SCF
MOD10A1 SCF
IMS SCF
SCF [–]
0.0
0.4
0.8
2008–11–1
2009–1–1
2009–3–1
2009–5–1
2009–7–1
2009–9–1
Time [YYYY–MM–D]



k_g [m^2/s]
t_0 [° C]
k_{snow} [mm/° C/day]
k_{rain} [mm/mm/°C/day]
k_{cor} [–]
ρ [mm water/cm]
AE_DySno data
AE_DySno temperature
GLOBSNOW data
GLOBSNOW temperature
IMS temperature
MOD10A1 data
MOD10A1 temperature
Observed SD data
Observed SD temperature
Std. WetSpa temperature

Table 1: Meteorological data from the Olecko station for the period: 1975 - 2012. Location of the station is presented in Figure 1. Winter is from November to April, Summer is from May to December. For sources of potential evapotranspiration (PET) see section 2.2.1; remaining data was provided by the Polish Institute of Meteorology and Water Management - National Research Institute (IMGW)

Variable	Yearly	Winter	Summer	Maximum / Month	Minimum / Month
Mean temperature [°C]	7.0	0.1	13.7	17.6 / VII	-3.6 / I
Precipitation sum [mm]	619	231	388	82 / VII	27 / II
PET sum [mm]	550	85	465	105 / VII	5 / I, II, XII
Mean Snow depth [cm]	3.7	7.5	0.0	17.8 / II	0 / V to IX
Percentage of days with snow cover	21%	43%	0%	75% / II	0% / V to IX

Table 2: Properties of the four snow products used in the study.

Product	Algorithm	Values	Resolution		Availability	Missing data
			Spatial	Temporal		
AE_DySno	Automatic, based on passive microwave data from AMSR-E sensor with ancillary data.	SWE in range 0-480 [mm]	25 km	day	2002 to 2011 (instrument failure)	Due to not full cover by orbits: 8.7% in the study area.
GLOBSNOW	Automatic, based on passive microwave data from SMMR, SMM/I and AMSR-E sensors and meteorological stations data.	SWE [mm]	25 km	day (in the study period)	1987 to 2010	Due to water, mountains (none in the study area) or missing input data: 58.0% (majority in the snow free period).
IMS	Manual, based on a images form many satellites with visible, infra-red and passive microwave sensors.	Snow and ice cover extent [presence or absence]	4 km	day	2004 to present	Rare, from various reasons: 0.3% of the time series.
MOD10A1	Automatic, based on visible and near infra-red reflectance from MODIS\Terra sensor.	SCF [-]	500 m	day	2000 to present	Frequent, mostly due to cloud cover: 26.0% of the time series.

Table 3: Snowmelt and snow accumulation calculation in the different model variants. The last column presents which snowmelt / snow accumulation switching method was implemented.

Model variant	Sv variable in a dataset	Accumulation	Melt	Switching based on
Standard WetSpa	-	$s = s + v_{rain}$	$v_{sm} = k_{snow}(t - t_0) + k_{rain}v_{pre}(t - t_0)$	temperature
Observed SD	SD [cm]	based on data	$v_{sm} = \rho\,(SD_{i-1} - SD_i)$	temperature, data
MOD10A1	SCF [-] $\in \{0:1\} \cap \mathbb{R}$	based on data	$v_{sm} = SCF\,(k_{snow}(t - t_0) + k_{rain}v_{pre}(t - t_0))$	temperature, data
IMS	SCF [-] $\in \{0,1\} \cap \mathbb{N}$	based on data	$v_{sm} = SCF\,(k_{snow}(t - t_0) + k_{rain}v_{pre}(t - t_0))$	temperature
GLOBSNOW	SWE [mm]	based on data	$v_{sm} = k_{cor}\,(SWE_{i-1} - SWE_i)$	temperature, data
AE_DySno	SWE [mm]	based on data	$v_{sm} = k_{cor}\,(SWE_{i-1} - SWE_i)$	temperature, data

Table 4: Ranges of the WetSpa global parameters optimized with the SCE method and used for the uncertainty analysis with the GLUE method. Only the parameters marked with a star (*) were used in all model variants, other parameters were variant specific (see Tab. 3 and Sect. 2.5.2). The full parameters description is available in Liu et al. (2004).

Parameter	Description	Range
k_i*	interflow scaling factor [-]	0.1 : 6.0
k_g*	groundwater flow recession coefficient [m^2/s]	1×10^{-6} : 1×10^{-1}
k_{i2}*	interflow recession coefficient [m^2/s]	1×10^{-4} : 6
k_{ss}*	initial soil moisture ratio to field capacity [-]	0.1 : 3.0
k_{ep}*	correction factor for evapotranspiration [-]	0.3 : 2.0
G_0*	initial groundwater storage depth [mm]	1 : 400
G_{max}*	maximum groundwater storage depth [mm]	1 : 1000
t_0	threshold temperature [$^\circ C$]	-3 : 3
k_{snow}	degree-day coefficient [mm/$^\circ C$]	0.01 : 6.00
k_{rain}	rainfall degree-day coefficient [mm/mm/$^\circ C$]	0.01 : 6.00
k_{run}*	coefficient reflecting the effect of rainfall intensity on runoff [-]	0.01 : 2.00
P_{max}*	rainfall intensity threshold above which k_{run} is set to 1 [mm/day]	1 : 700
k_{cor}	correction factor for the SWE data [-]	0.01 : 3.00
ρ	snow density factor [mm water/cm]	0.01 : 3.00

Table 5: The NS scores for the calibration, validation and the whole period of the WetSpa model for all tested model variants. Each cell shows the three NS scores as: "calibration / validation / whole". The global SCE calibrations were repeated three times for each model setup in order to make sure that the global optimum was converged. The best NS scores in each snow product of the three calibration runs, for each switching mode, are indicated in bold.

switching:	data-based			temperature-based		
calibration no:	1	2	3	1	2	3
AE_DySno	0.33 / 0.16 / 0.26	0.66 / **0.36** / 0.53	0.66 / 0.34 / 0.52	0.64 / **0.30** / 0.49	0.62 / 0.25 / 0.45	0.64 / 0.28 / 0.48
GLOBSNOW	0.55 / 0.36 / 0.47	0.59 / **0.53** / 0.47	0.60 / 0.50 / 0.49	0.49 / **0.29** / 0.36	0.49 / 0.15 / 0.30	0.51 / 0.26 / 0.34
IMS	-	-	-	0.73 / -0.99 / -0.06	0.73 / -7.99 / -3.29	0.73 / **-0.69** / 0.08
MOD10A1	0.65 / 0.37 / 0.53	0.67 / 0.50 / 0.60	0.64 / **0.60** / 0.63	0.70 / 0.54 / 0.63	0.62 / **0.58** / 0.61	0.70 / 0.49 / 0.61
Observed SD	0.74 / 0.44 / 0.61	0.53 / 0.35 / 0.46	0.73 / **0.50** / 0.63	0.75 / 0.37 / 0.57	0.74 / 0.41 / 0.59	0.77 / **0.41** / 0.60
Standard WetSpa	-	-	-	0.68 / 0.54 / 0.62	0.73 / **0.55** / 0.65	0.69 / 0.53 / 0.62

Table 6: The percentage of time the observed discharge was outside the 95% confidence interval estimated with the GLUE method for all model variants

Model variant	switching	
	data-based	temperature-based
MOD10A1	8.8%	46.0%
IMS	-	4.9%
Observed SD	5.5%	5.5%
GLOBSNOW	5.2%	2.7%
AE_DySno	8.8%	9.9%
Standard WetSpa	-	6.6%