

– MANUSCRIPT rev. January 5, 2017 –

**lumpR: An R package facilitating hillslope-based landscape
discretisation for large-scale hydrological models**

Tobias Pilz, Till Francke, and Axel Bronstert
Institute of Earth and Environmental Science, University of Potsdam, Germany
Correspondence to: Tobias Pilz (tpilz@uni-potsdam.de)

Todo list

	Tobias:	
	Sufficient definition of term "semi-distributed"?	2
	Tobias:	
5	Definition of term "semi-distributed" in Introduction sufficient?	4
	Tobias:	
	I defined all these abbreviations but maybe this is a bit too much for "non-computer-nerds"?	7
	Tobias:	
	I still need to make an improved version of this figure. Suggestions?	10
10	Tobias:	
	I guess this figure could still be improved as well. And it needs a north arrow.	13
	Tobias:	
	Other or more studies to be cited?	13
	Tobias:	
15	Does the Discussion section already contain too much conclusions?	25
	Tobias:	
	Figures herein are still preliminary and need adjustments (legend, larger font size etc.).	26
	Tobias:	
	Any more acknowledgements from your side?	27

1 Introduction

Hydrological simulation models are important tools for gaining process understanding, forecasting of streamflow, supporting water managers, and climate and/or land-use change impact studies. However, the lack of a unified theory in catchment hydrology led to a large number of competing models (e.g., Weiler and Beven, 2015). These differ in terms of aim and scope, process representation and parametrisation, temporal resolution, and spatial discretisation. The latter includes the size and shape of units where the model's underlying mathematical equations are solved.

The most straightforward procedure is the discretisation of the landscape into quadratic grid cells of equal size, commonly referred to as *fully distributed* approach. This, however, suffers from several drawbacks such as a high computational burden and large memory demands as well as limitations in representing landscape variability and processes at smaller scales than given by the grid's resolution. Another strategy is to treat a hydrological catchment as a single unit, known as *lumped* approach. While being easy to implement along with simple and fast computable conceptual process representations this approach fails to adequately represent complex landscapes and process interactions leading to a deteriorated simulation of hydrological catchment behaviour. Therefore, *semi-distributed* approaches are usually the preferred choice in practical applications (e.g., Kumar et al., 2010; Euser et al., 2015). Although no clear definition exists and sometimes termed as "distributed" approach, this family of discretisation typically consists of a hierarchical multi-level scheme dividing a hydrological basin into several subbasins which in turn contain irregularly shaped computational elements being hydrologically uniform entities (e.g., Krysanova et al., 1998). It is therefore distributed in a way that hydrological processes are simulated at different locations in the study area taking into account distributed model input (such as meteorological forcing and landscape parameters) and producing spatially variable output (such as lateral and vertical water flows) but lumped as naturally heterogeneous though hydrologically similar areas are aggregated and parameterised in the same manner.

In the last decades a large number of discretisation schemes for semi-distributed models has been developed tailored to various model structures and adapted to specific objectives and needs of modellers. Along with that an even larger number of software solutions exists and it is a difficult task to decide for a specific model and discretisation approach and a programme for landscape pre-processing. Therefore, the first objective of this paper is to provide an overview of landscape discretisation algorithms and software implementations for semi-distributed hydrological model application.

As will be further discussed in section 2, hillslope-based modelling is an efficient semi-distributed strategy for the investigation of semi-arid landscapes and allows to represent heterogeneous runoff generation processes preserving hydrological connectivity. However, so far only few computer programmes exist aiding in the pre-processing of hillslope-based models. The second objective therefore is to present a new software package for the discretisation of the landscape into representative hillslopes. The package aims to be model independent, user-

Tobias:
Sufficient definition of term "semi-distributed"?

friendly, free and open-source, applicable to large scales, and where the underlying algorithm should preserve hydrological connectivity and landscape characteristics along the hillslopes. To facilitate implementation and be able directly connect to a *Geographical Information System* (GIS) it was decided to build upon the free and in scientific contexts widely used scripting language R (R Core Team, 2015). The package is introduced and described in section 3.

The third objective of this study is to investigate the role of subjective user decisions during landscape discretisation on simulated streamflow. To achieve this, an example application of the developed R package in combination with the *Water Availability in Semi-Arid landscapes* (WASA) model (Güntner and Bronstert, 2004) in a semi-arid landscape in northeastern Brazil is presented. A sensitivity analysis is employed to analyse the impact of user decisions and accompanying discretisation complexity on the model results (section 4). Eventually, the findings of this study are discussed and conclusions are inferred (sections 5 and 6, respectively).

2 Review of landscape discretisation schemes and software implementations

2.1 Discretisation approaches in semi-distributed hydrological modelling

Landscapes are typically mapped using Digital Elevation Models (DEMs). *Contour-based* DEMs store terrain information as contour lines (or x, y coordinate pairs) of specific elevation (Moore et al., 1988). Moore and Grayson (1991) and Maurer (1997) show example applications of contour-based terrain analysis for hydrological model application where this type of DEM proved to be powerful as its structure is based on the way how water flows on (albeit not necessarily below) the land surface. However, although they have been further investigated for hydrological application (e.g., Maunder, 1999; Moretti and Orlandini, 2008) contour-based DEMs come along with some limitations. They have a relatively high data storage demand, topographic attributes are complicated to derive, and they provide no computational advantages (Moore et al., 1991). *Triangulated Irregular Networks* (TINs) form a type of DEM sampling elevation points at specific landscape features, such as peaks or ridges, and form an irregular network of x, y, and z coordinates. They are very flexible as, due to their irregular structure, they are able to map regions of high heterogeneity with more data points than smooth terrain and thus avoid redundancy and increase data storage efficiency (Moore et al., 1991; DeVantier and Feldman, 1993). TINs also proved to be useful in a number of hydrological applications (e.g., Tucker et al., 2001; Vivoni et al., 2004; Ivanov et al., 2004; Freitas et al., 2016). Their irregularity, however, makes the computation of topographic attributes more difficult and there can be problems when determining upslope connections for watershed derivation (Moore et al., 1991). *Grid-based* DEMs store elevation information as a regularly spaced mesh. There are a number of drawbacks as the regular structure might impose artefacts and discontinuities while sub-resolution landscape features cannot be captured limiting the applicability for hydrological purposes. Furthermore, when increasing the grid's resolution to reduce these problems, computational burden and memory requirements are increased reducing their suitability for large scale applications. Nevertheless, grid-based DEMs are the most widely used data structures due to their

straightforward generation from remote sensing data, direct applicability for further investigations, and efficient calculation of geomorphological characteristics (Moore et al., 1991; DeVantier and Feldman, 1993).

As as basis for watershed delineation and further landscape discretisation, a number of pre-processing steps have to be performed. This typically includes (i) pit-filling of the DEM to remove sink areas and ensure a proper drainage of water from the catchment; (ii) computation of flow directions; (iii) computing upslope contributing area (i.e., the accumulated flow for each DEM unit); (iv) derivation of the river network, typically based on flow accumulation from the previous step; (v) delineation of the hydrological basin and subbasins. For each step different algorithms have been proposed whose selection depends on the type of DEM, model to be employed, aim and scope of the study, or just personal preferences, e.g., in terms of software to be used (e.g., O'Callaghan and Mark, 1984; Tarboton et al., 1991; Costa-Cabral and Burges, 1994; Lacroix et al., 2002; Vivoni et al., 2004; Moretti and Orlandini, 2008).

There exists a large number of landscape discretisation schemes for semi-distributed hydrological model application. Some models explicitly use TINs, such as tRIBS (Ivanov et al., 2004), or modelling units derived from contour-based DEMs, e.g., THALES (Grayson et al., 1992). The majority, however, delineate irregularly shaped polygons as computational model units derived from grid-based DEMs.

Tobias:
Definition of term
"semi-distributed"
in Introduction sufficient?

Wood et al. (1988) defined the smallest discernible averaging watershed where statistics of runoff generation did not further change as *Representative Elementary Area* (REA) and applied simple conceptual equations for simulation of runoff generation. They found the size of their REA to be primarily influenced by topography. The concept of *Hydrological Response Units* (HRUs), introduced by Leavesley et al. (1983) for their Precipitation-Runoff Modeling System (PRMS) and further elaborated by Flügel (1995), is maybe the most prominent landscape discretisation scheme. An HRU is assumed to be a homogeneous set of hydrological process dynamics formed by a pedo-topo-geological association with specific land-use and as such controlled by land-use management and physical landscape properties. The conception has been adopted for a number of models such as SWAT (Manguerra and Engel, 1998), SWIM (therein termed *hydrotopes*) (Krysanova et al., 1998), PREVAH (Viviroli et al., 2009), or GSFLOW (Markstrom et al., 2008). MHYDAS is a process-based hydrological model for which the HRU concept has been further pursued for application in agricultural management contexts by including man-made hydrological discontinuities such as ditches and field boundaries (Moussa et al., 2002). Kouwen et al. (1993) were looking for an alternative approach to better represent watershed heterogeneity over large basins. They defined *Grouped Response Units* (GRUs), small watersheds as computational units with uniform meteorological forcing that are considered to be hydrologically heterogeneous by consisting of a range of land cover (or some other attribute's) characteristics where only percent cover is used for characterisation instead of an explicit spatial reference. Topography and the other relevant attribute (e.g., land cover) are assumed to be the major factors influencing runoff generation. A similar conception is utilised by the *Aggregated Simulation Area* (ASA) approach developed for the SLURP model (Kite, 1995). Winter (2001) aimed for an integrated inspection of the complete hydrological system in different terrain types by introducing *hydrologic landscapes*. These consist of variations and multiples of fundamental hy-

hydrologic landscape units as building blocks characterised by landsurface form, geologic framework, and climatic setting to describe movements of surface water, groundwater, and atmospheric water, respectively.

The formerly introduced approaches have been further evolved into conceptions focusing on *functional units* rather than mere spatial units which always came along with the assumption of homogeneous process dynamics in a certain area. Reggiani et al. (1998) delineated the basin into autonomous functional subbasins, the *Representative Elementary Watersheds* (REWs), using the drainage network as basic organising structure. The REW is then divided into five functional sub-regions whereas micro-scale physical conservation equations for each sub-region are simplified and averaged, further accounting for thermodynamic exchange between sub-regions and REWs. Zehe et al. (2014) propose three *functional response units* separating radiation-driven vertical flow from rainfall-driven lateral flow processes on similar functional entities within a hydro-geomorphic homogeneous subbasin.

Other approaches divide the watershed into *representative hillslopes* as one- or two-dimensional approximation of a three-dimensional soil catena separated by drainage network and ridges. For instance, Troch et al. (2003) developed the hillslope-storage Boussinesq equation to simulate drainage and soil moisture storage dynamics along a hillslope described by a polynomial function. Flanagan and Nearing (1995) introduced WEPP, a complex process-based soil erosion prediction model applicable over hillslopes or small watersheds comprised of multiple hillslopes, channels, and impoundments. They lump individual hillslopes by calculating and averaging quantitative hillslope characteristics. Examples for other hillslope-based models include KINEROS (Woolhiser et al., 1990), IHDM (Beven et al., 1987), or CATFLOW (Maurer, 1997). Several studies focussed on how to delineate and describe hillslopes from a DEM (e.g., Cochrane and Flanagan, 2003; Noël et al., 2014), discussing morphometric controls on hillslope parameters (Bogaart and Troch, 2006), or investigating the role of hydrologic connectivity (Smith et al., 2013). Overall, hillslope hydrological modelling has been proven to be useful in regions with steeply sloping landscapes and heterogeneous runoff generation mechanisms where lateral water fluxes are relevant (Bronstert, 1999).

For the application at large scales it is necessary to find a compromise between sufficient detail in landscape representation and computational feasibility. Thus, several studies examined the impact of discretisation complexity on model performance. It could be seen that (semi-)distributed models usually perform better than lumped ones (Kumar et al., 2010; Euser et al., 2015) but there also is a threshold of subdivision level above which no improvements can be achieved (Wood et al., 1988; Han et al., 2014; Haghnegahdar et al., 2015). On the other hand, sub-grid variability is an important characteristic of individual hydrological landscape response. This can be respected by acknowledging different spatial scales in a model application (e.g., the scales of meteorological forcing, model application, and basin characteristics) and combining them using parameter regionalisation (e.g., Samaniego et al., 2010).

Defining a hierarchical multi-scale discretisation scheme for the model is a common strategy in semi-distributed modelling. This method was formally introduced with the GRU approach but is being done by many of the aforementioned models, in a varying degree of detail, for other discretisation approaches as well. It commonly includes

the basin of investigation being discretised into subbasins and hydrologically homogeneous modelling units. For the modelling system RHESSys the landscape is partitioned into a hierarchy of progressively finer units modelling different processes associated with a particular scale. A given spatial level is represented as object type with a set of states, fluxes, process representations, and corresponding model parameters (Band et al., 2000; Tague and Band, 2004).

Güntner and Bronstert (2004), for their WASA model developed an even more complex scheme of six spatial levels. Starting at the watershed, subbasins are delineated, sub-divided into representative hillslopes termed landscape units which are further separated into specific parts of the hillslope, the terrain components to account for lateral redistribution of water flows, whereas vertical processes and runoff generation are simulated over individual homogeneous *Soil-Vegetation Components* (SVCs) for which a representative soil profile with respective soil layers has to be given. As for the GRU approach, at the smaller scales an explicit spatial representation is omitted in favour of a percentage cover representation in order to better capture landscape heterogeneity while keeping storage and computational demands at a minimum. This conception has been proven to be efficient and successful for the simulation of heterogeneous semi-arid landscapes with complex hillslopes and patchy vegetation over large scales dominated by Hortonian overland flow and runoff redistribution mechanisms. As hydrologic connectivity of the landscape could be represented in a realistic manner the model has been used for a number of studies investigating runoff redistribution and erosion processes (e.g., Güntner and Bronstert, 2004; Mueller et al., 2010; Medeiros et al., 2010; Bronstert et al., 2014).

2.2 Software for model pre-processing and landscape discretisation

Together with improving computer facilities and increasingly available DEMs and processing algorithms software for terrain analysis and discretisation started to evolve. Already in the early 1990s, DeVantier and colleagues published a review of applications of GIS in hydrological modelling (DeVantier and Feldman, 1993). Among the earliest and still most prominent GIS for hydrological purposes are Esri's commercial ArcGIS, launched in 1982, and the *Geographic Resource Analysis Support System* (GRASS), started in 1982 by the U.S. Army Corps of Engineers and since 1999 developed and distributed as *free and open-source software* (FOSS) (Mitasova and Neteler, 2004).

However, many researchers started writing their own scripts tailored to their needs and sometimes later on published or distributed their solutions both commercially and in non-profit manners. TAPES-G is an early terrain analysis programme written in FORTRAN-77 and C for use on Unix machines already including several algorithms for specific tasks, e.g., five methods alone to calculate flow accumulation (Gallant and Wilson, 1996). A prominent and widely used example is the TOPAZ software package for automated analysis of digital landscape topography addressed to guide farmers, engineers, and scientists in both research and practical application. The programme is free of charge and available on request but the development stopped in 1999 (Garbrecht and Martz, 1999). Tarboton (2003) introduced TauDEM, a freely usable terrain analysis programme for Windows that can be applied

independently from the command line or comes along with a *graphical user interface* (GUI) as extension for the commercial ArcMap software and is still being further developed. More information on hydrologically relevant software of the commercial ArcGIS family, such as ArcHydro, can be found at the software's community web pages: <http://resources.arcgis.com/en/communities/hydro/> (accessed 23 June 2016). GRASS provides a number of

5 elaborated and still evolving FOSS tools for hydrological model pre-processing, such as *r.watershed*, that have been successfully applied in a number of studies (e.g., Kinner et al., 2005; Metz et al., 2011; Neteler et al., 2012). Another recent example of stand-alone software is GeoNet, a tool for automatic channel head, channel network, and channel morphological parameter extraction from high resolution topography data that can be employed within MATLAB or, in a more recent version, as Python programme (Passalacqua et al., 2010; Sangireddy et al., 2016).

10 As being a frequently employed discretisation scheme used in well-known hydrological models such as SWAT, many software solutions exist implementing the HRU concept. Typically, the delineation process is based on the overlay of spatial raster data including land cover, soil, and/or geology. The programmes basically differ in terms of additional processing steps (such as terrain analysis), used algorithms, supported data formats and whether they are tailored to a specific model or are stand-alone, GIS back end (mostly ArcGIS or GRASS), supported operating

15 systems, and whether they provide a GUI or have to be run from command line. Sanzana et al. (2013) developed a number of stand-alone and model independent Python scripts for terrain analysis and HRU mesh generation making use of GRASS functionalities. Also relying on GRASS and Python, Schwartze (2008) created an extension for QGIS (a FOSS GIS with user-friendly GUI) as HRU delineation tool. Other software is model specific such as IOSWAT (Haverkamp et al., 2005), AGWA (Miller et al., 2002, 2007), or AVSWAT (Di Luzio et al., 2004)

20 being addressed to SWAT, WINHRU (Viviroli et al., 2009) written for PREVAH, or Geo-MHYDAS (Lagacherie et al., 2010) which is a collection of SHELL and PERL scripts using GRASS to help users of MHYDAS with the model pre-processing. Some of these are mere wraps around individual programmes to guide through the whole modelling process, including terrain analysis, HRU delineation, preparation of input files, model execution and parameter calibration, and graphical and/or statistical analysis of simulation results.

25 The pool of software packages for other landscape discretisation schemes is less rich. Lacroix et al. (2002) presented SLURPAZ, an interface between the TOPAZ terrain analysis tool and the SLURP model for delineation of ASAs. Around the hillslope-based WEPP model the geo-spatial assessment software GeoWEPP has been developed integrating TOPAZ, WEPP, and other tools for detailed analysis of spatially and temporally variable environmental and management scenarios (Renschler, 2003). It is nowadays integrated into the ArcGIS project and there

30 exists a lightweight web-based interface for less advanced users and ad hoc model application (Flanagan et al., 2013). RHESSys comes with interfaces for both GRASS and ArcGIS to assist in landscape pre-processing (Band et al., 2000).

Francke et al. (2008) published an algorithm for the semi-automated delineation of representative hillslopes. This *Landscape Unit Mapping Program* (LUMP) first discretises the landscape into various hillslopes, the *Elementary Hillslope Areas* (EHAs), and computes a representative catena for each of them. As a next step, similar

35

Tobias:
I defined all these abbreviations but maybe this is a bit too much for "non-computer-nerds"?

EHAs are grouped into *Landscape Units* (LUs) whereas for the classification several variables can be taken into account such as horizontal and vertical catena length, shape of the profile, and sets of supplemental attributes further characterising the hillslope, e.g., qualitative data such as soil type and land cover class, or quantitative attributes such as LAI. The LUs are eventually sub-divided into *Terrain Components* (TCs), e.g., into upslope, middle, and downslope parts. The software is semi-automated in way that the hillslope-based landscape parameterisation is largely automated generating reproducible results and reducing required user decisions to a minimum whereas, on the other hand, expert knowledge can be easily incorporated to improve the discretisation outcome. Their software is basically a collection of freely available scripts written in Matlab and SHELL using GRASS functionalities available via <http://brandenburg.geoecology.uni-potsdam.de/projekte/sesam/download/software/lump.zip> (accessed 4 July 2016). LUMP was directed to the pre-processing of the WASA model (Güntner and Bronstert, 2004) but is stand-alone and the output can as well be used for other hillslope-based models. For more information see their publication.

In order to meet the second objective of this study to develop a user-friendly and efficient tool for hillslope-based landscape discretisation it was decided to build upon the LUMP algorithm due to several reasons. It is possible to generate a reproducible hillslope-based landscape discretisation which is, due to the efficient implementation and contrary to other hillslope approaches, also applicable to large areas. As the approach is semi-automated it relies only on a small number of user decisions and thus reduces subjectivity to a minimum. Furthermore, in contrast to simple GIS overlay techniques often used for HRU delineation, it preserves information on the distribution of landscape parameters including their relative topographic position and makes it possible to reproduce hydrological connectivity within simulation studies.

3 lumpR: R package description

The computer software *lumpR* was developed with the aim to obtain a lightweight user-friendly and efficient tool for hillslope-based landscape discretisation and serving as pre-processing tool for the WASA model (Güntner and Bronstert, 2004, cf. section 2). In a more general sense, however, it should meet the objectives of being (i) platform independent, (ii) largely model independent, (iii) free and open-source, (iv) automated as far as possible reducing subjectivity but (v) allowing to include expert knowledge. In order to produce an easily applicable software avoiding specific programming skills to potential end-users and to meet objectives (i) and (iii) in particular, it was decided to use the scripting language R (R Core Team, 2015) and assemble *lumpR* as a software package for this environment.

Figure 1 gives an overview over structure and functionalities of the package. In the following these shall be explained in more detail. For more detailed information on how to install and use the package and to inform about updates the reader is referred to the package's documentation (see section 7).

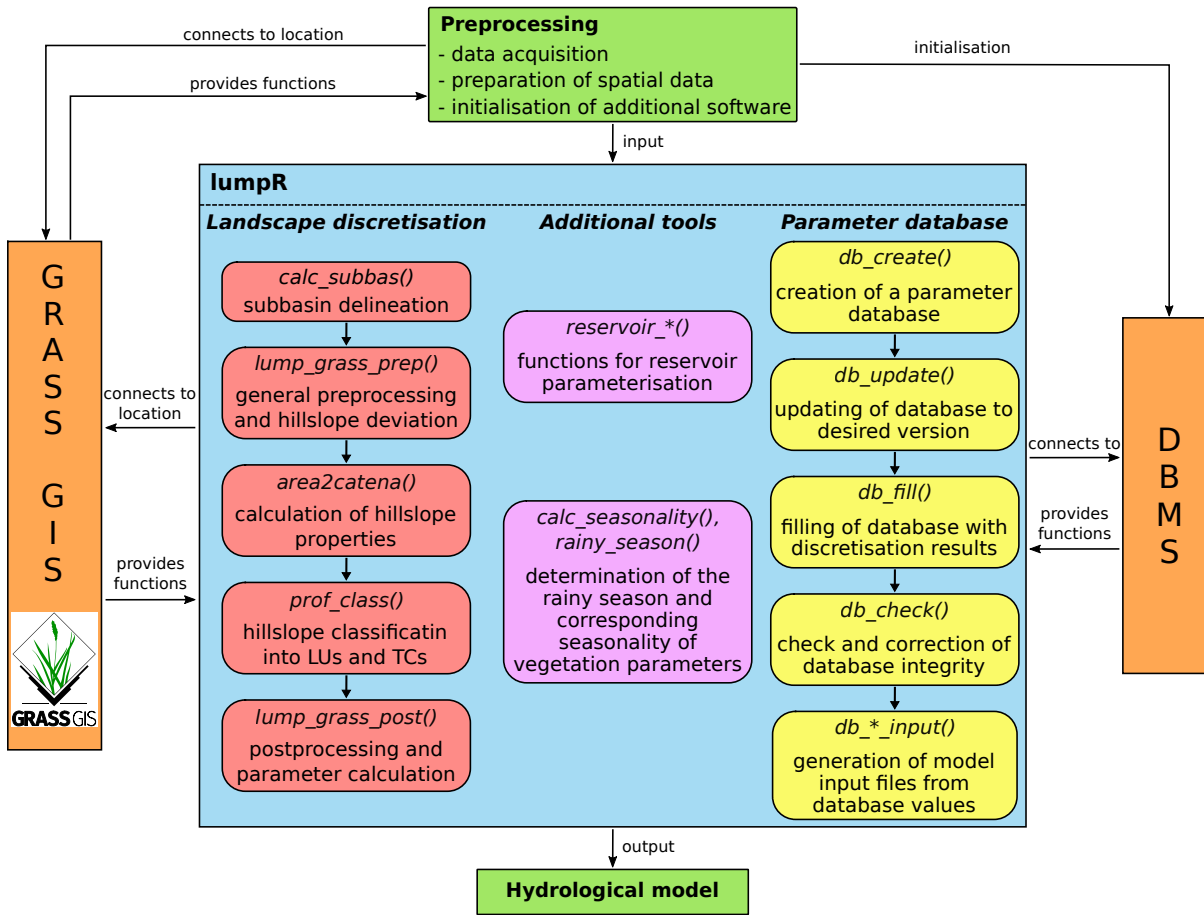


Figure 1. Schematic overview of the functionalities of the lumpR package. Shown are all available functions (*in italics*) with a short explanation, interactions with external software, and flowcharts of typical workflows during application. Note that the acronym DBMS refers to *DataBase Management System*.

3.1 Landscape discretisation

After installing the package and all additionally needed software the user has to acquire and prepare all needed spatial data in a location in GRASS GIS. Internally the package's functions connect to that location, use the given data for processing while partly employing GRASS functions, and finally stores the spatial output in the GRASS location and/or text files in a specified directory for immediate inspection after each function call. As a first step in a new R session, before executing any of the package's function R has to be connected to the GRASS location by the user as shown in the following example:

```
> library(lumpR)
> initGRASS(
```

```

> gisBase = "/path/to/grass",
> home = tempdir(),
> location = "example_location",
> mapset = "example_mapset",
5 > gisDbase = "/path/to/grassdata")

```

As is sketched in Fig. 1 the process of landscape discretisation involves five functions that should be applied in the following order. (i) *calc_subbasin()* sub-divides the hydrological basin into subbasins using a given grid-based DEM. Subbasin size can be influenced by the user by either giving a set of coordinates of drainage locations inferred beforehand or by specifying the parameter `thresh_sub` being the minimum size of a subbasin in number of grid cells internally used by GRASS function `r.watershed`. (ii) *lump_grass_prep()* does several pre-processing steps needed for later use such as computing SVCs as simple overlay of soil and vegetation raster maps, inferring Horton stream order, and the calculation of hydrologically relevant landscape characteristics such as a flow direction, flow accumulation, relative elevation (i.e., elevations above next downstream river grid cell), and distance to next river grid cell put as raster maps into the initialised GRASS location. Furthermore, the function infers EHAs based on the crucial size parameter `eha_thres` that has to be given by the user. (iii) *area2catena()* takes generated data from step (ii) and supplemental raster maps of quantitative and/or qualitative variables to calculate a representative catena for every EHA characterised by horizontal and vertical length, shape (in terms of cumulated elevation along the hillslope), slope width (approximated by taking the number of grid cells at a profile point divided by the total number of grid cells representing the whole hillslope), and optional information from supplemental data. The algorithm is based on the work of Cochrane and Flanagan (2003) for the WEPP model. (iv) *prof_class()* takes the representative catenas and classifies them into LUs. In this step similar catenae are identified and lumped together based on the calculated properties employing an unsupervised K-means clustering method. The user has to give a number of parameters affecting the importance of individual catena characteristics during the clustering. Each LU is then further sub-divided into TCs, e.g., representing upper, middle, and downslope parts of the lumped catena. The number of TCs to be generated for each LU is another input parameter. For visual inspection of results the user has the option to let the functions generate plots during steps (iii) and (iv). The employed algorithms for steps (iii) and (iv) are explained in more detail by Francke et al. (2008). Figure 2 gives an illustrative example for the outcome after applying steps (i) to (iv).

Finally, (v) *lump_grass_post()* prepares subbasin and LU specific parameters as needed by the WASA model thereby employing rather simple relationships or typical standard values. The output of this function should therefore be carefully revised. See the function's documentation for more details.

Tobias:
I still need to make
an improved ver-
sion of this figure.
Suggestions?

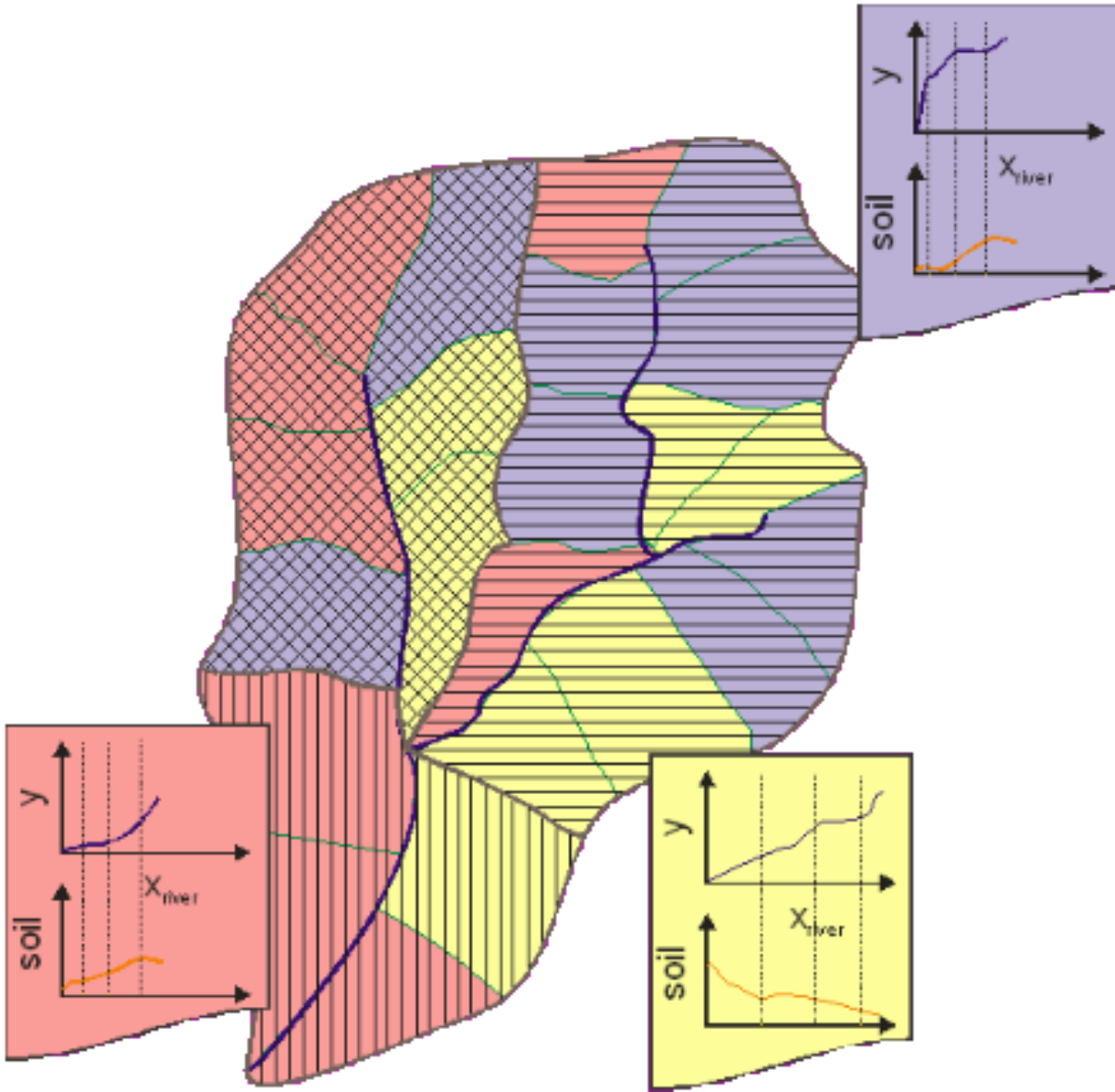


Figure 2. Illustrative example of the outcome of steps (i) to (iv) of landscape discretisation. Shown is a hydrological basin sub-divided into subbasins (hatchures), EHAs (thin green lines), and LUs (colours). Average catena properties for the three LUs are exemplified by the coloured diagrams. TCs are indicated by vertical dashed lines in the diagrams.

3.2 Additional tools

In order to meet further capabilities of the WASA model functions *reservoir_**(*)* have been introduced. They facilitate the pre-processing of reservoir-specific input files for the model using spatial reservoir data and pre-compiled

parameterisations. Additional function *rainy_season()* calculates start and end dates of the rainy season for every year based on a time series of daily precipitation values using a statistical approach described by Gerstengarbe and Werner (1999). In WASA this information is needed as parameterisation of intra-annual variations of vegetation parameters. *calc_seasonality()* takes output of the former function and information about seasonal variation of a vegetation parameter to calculate a daily time series of that parameter by linear interpolation of the parameter's node points depending on the current start and end dates of the rainy season for a specific year. For more detailed information on vegetation parameterisation the reader may consult Güntner (2002) and the function's documentation. Although being rather specific to WASA the former functions may as well be of use for other models.

3.3 Parameter database

- 10 To store the results of landscape discretisation in a clearly arranged manner for readily generation of necessary model input files the package was enhanced by a set of database functions. Guaranteeing platform independence and as much freedom of choice as possible to the user different frequently used *DataBase Management Systems* (DBMS) are supported by lumpR currently including MS Access, MySQL / MariaDB, and SQLite. As interface between R and the DBMS the *Open DataBase Connectivity* (ODBC) in terms of the R package RODB is employed.
- 15 To use a specific DMBS a number of pre-processing steps are necessary, such as installing required software and registering a new empty database as ODBC source, before applying the package's database functions. For inexperienced users we prepared some general information on that topic in the Wiki of lumpR's web page (see section 7). As for landscape discretisation the database functions might be applied in a defined order as indicated in Fig. 1 and will be further explained in the following.
- 20 (i) *db_create()* creates the latest implemented version of all tables in the registered parameter database. As the package underlies continuous further development which might also affect the parameter database (ii) *db_update()* was written to ensure compatibility between different database versions. The user can specify to which version the current database shall be updated to include the latest changes in the database scheme to an existing database or to go back to an earlier version. (iii) *db_fill()* takes text files produced as output of landscape discretisation and other pre-processing not included in the package (e.g., readily prepared soil and vegetation parameters) and fills the values into the respective tables of the parameter database. (iv) *db_check()* performs a number of checks to identify and (if possible and desired) automatically resolve inconsistencies and/or missing information in the database. These include filtering of tiny and insignificant areas (e.g., SVCs) to reduce computational overhead, checking that all TCs have a slope larger than zero, defining special areas for separate treatment (this currently includes SVCs marked as water or impervious), removing special areas marked was water, computing fractions of impervious surfaces at TC level, removing impervious surfaces, estimating a storage coefficient for groundwater delay (for a simple linear bucket approach of groundwater simulation) at LU level, deleting obsolete datasets (special area SVCs, LUs not in any subbasin, TCs not in any LU, and SVCs not in any TC), checking for completeness (all IDs in the *_contains_* tables exist within the respective referenced tables), and computing the subbasin order
- 30

from upstream (largest number) to downstream (one). The user can decide which checks to perform, how to deal with inconsistencies, and define thresholds for certain checks. To keep an overview any changes to the database will be tracked in table `meta_info`.

At any point during the processing the user can freely inspect and adjust the parameter database by means other than the functions provided by lumpR. In the current version the package provides the function `db_wasa_input()` to take the values of the parameter database and generate most of the generally necessary input files for WASA except for meteorological forcing and files related to sediment routing and reservoirs. However, the user may as well export the values needed and compile the input files for any other model that makes use of the values assembled in the parameter database. Furthermore, it is a straightforward task to extend the package and include export functions for other models if requested.

4 Example application and sensitivity analysis

4.1 Study site

To demonstrate the functionalities of lumpR the Benguê catchment was selected (Fig. 3). It is part of the upper Jaguaribe river catchment in the northeast of Brazil within the federal state of Ceará. The area has been investigated in a number of studies, in many cases employing the WASA model, and was thus selected to ensure the ability of the model for the catchment (Medeiros et al., 2010; Krol et al., 2011; de Araújo and Medeiros, 2013; Bronstert et al., 2014; Medeiros et al., 2014; Medeiros and de Araújo, 2014; de Figueiredo et al., 2016).

The Benguê catchment drains an area of about 926 km². At the outlet the ephemeral Umbuzeiro river is constrained by a reservoir built in 2000 with a storage capacity of 19.6 million m³ to enhance water supply and reliability in the region. As the area is located within the "drought polygon" of Brazil annual average precipitation is low with about 600 mm in comparison to a potential evapotranspiration of more than 2000 mm. The mean annual temperature is 25 °C with little variation. Climate is further characterised by a strong intra-annual variation of precipitation leading to distinct rainy (January to May with more than 80 % of annual rainfall) and dry seasons. Rainfall is mostly convective and concentrated in only a few events a year with high intensity. Inter-annual variation of precipitation, however, is also high causing recurrent droughts which in cases may last over several consecutive years. The dominant natural vegetation is Caatinga consisting of deciduous bushland with xerophytic species ranging from dense dry forests to almost desert-like sites. The population density is low (6.4 inhabitants per km²) with rural lifestyle. Parts of the area are used for small-scale farming including cattle-breeding and growing of maize and beans in particular. The environment is further characterised by mainly sedimentary plateaus in the southern and western parts of the study site with steep terrain and deep (> 1 m) permeable soils, predominantly Latosols. In the North and East crystalline bedrock is prevailing with shallow Luvisols (< 1 m) causing high runoff coefficients. In alluvial zones Planosols are dominant.

Tobias:
I guess this figure
could still be im-
proved as well.
And it needs a
north arrow.

Tobias:
Other or more
studies to be cited?

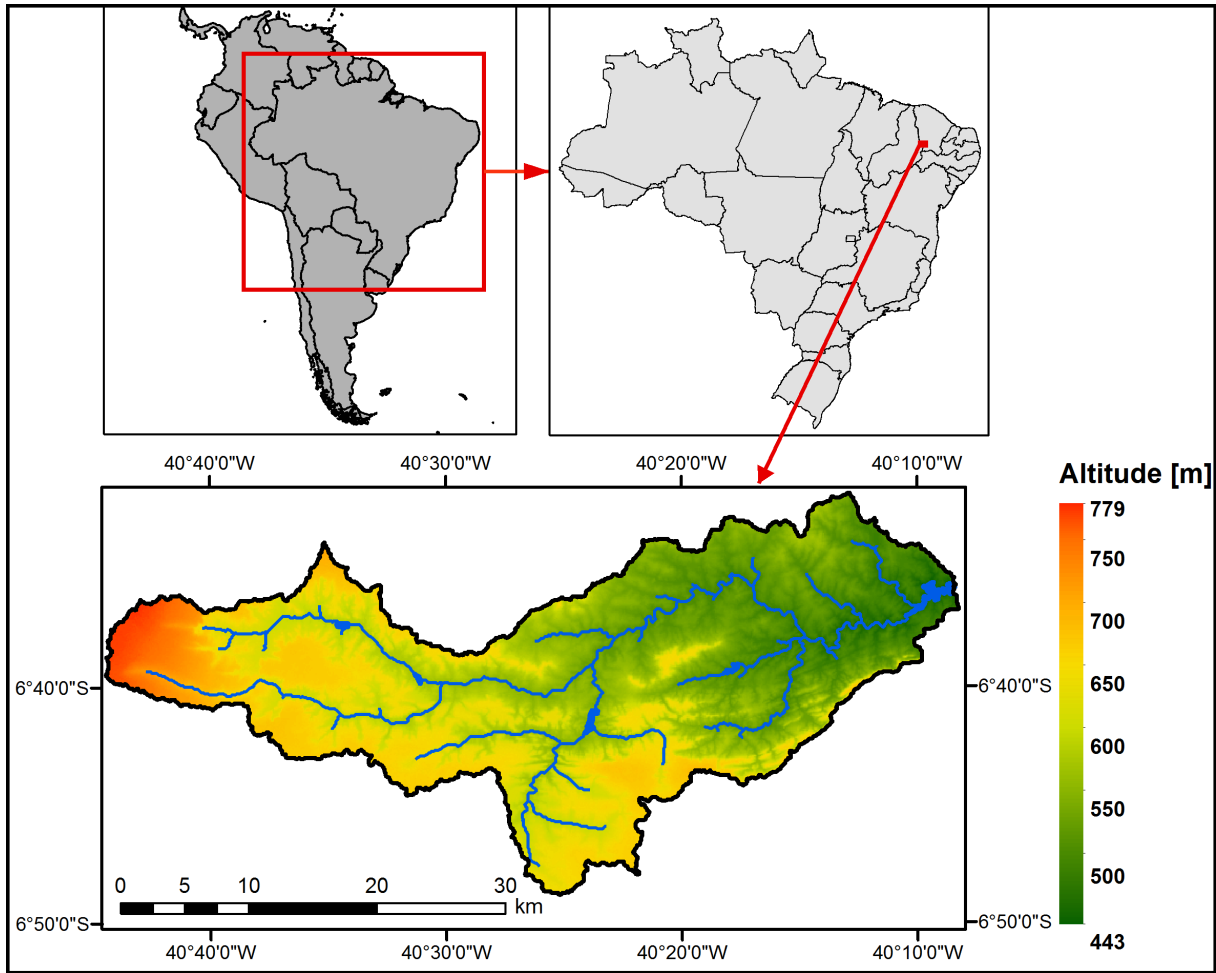


Figure 3. Overview over the Benguê catchment (lower panel) and location within Brazil (upper right) and South America (upper left), respectively.

4.2 WASA model

The WASA model is used to study the effects of different landscape discretisations realised with lumpR on simulated streamflow dynamics. The model was first introduced by and is described in detail within Güntner (2002). It has been frequently employed in semi-arid areas such as northeastern Brazil (including the Benguê catchment; for references see section 4.1), India (Jackisch et al., 2014) and Spain (Mueller et al., 2009, 2010; Bronstert et al., 2014). For information on model availability the reader is referred to section 7.

WASA is a deterministic, process-based, semi-distributed, time-continuous hydrological model. It is dedicated to natural semi-arid environments. The model incorporates the Shuttleworth-Wallace approach for evapotranspiration calculation over sparsely vegetated surfaces and an infiltration approach based on Green-Ampt accounting

for Horton-type infiltration. Via the complex hierarchical spatial disaggregation scheme (cf. section 2.1) lateral re-distribution processes as well as ex- and re-infiltration along a hillslope are considered while the model can still be applied over large scales (up to the order of magnitude of 100,000 km²). Large strategic reservoirs can be represented in an explicit manner while smaller ones are treated as lumped lakes of different size classes to efficiently account for water retention of many small reservoirs in a study region. The model has been subsequently expanded to account for sediment dynamics with several optional USLE-based approaches (Mueller et al., 2010), whereas it was henceforth called WASA-SED. As in this study no erosional processes are considered the model shall herein furthermore termed WASA.

4.3 Data and model setup

Meteorological data in daily resolution to drive WASA, in particular precipitation, air temperature (obtained by minimum and maximum temperatures by simple averaging), relative humidity, and incoming short-wave radiation, have been derived from the dataset described by Xavier et al. (2016). The gridded dataset is based on station information that have been checked, corrected, and interpolated to a grid with 0.25° x 0.25° resolution. Extraterrestrial radiation as a further driver that has been calculated from astronomical relationships employing the R package *sirad*. Meteorological data have been interpolated to the locations of subbasin centroids using an inverse-distance weighting approach by employing the R package *geostat*.¹

The basis of terrain analyses was the 90m x 90m grid-based SRTM DEM. As pre-processing step the raw raster has been sink-filled employing the GRASS function *r.terraflow*. Soil information have been derived from the database of Jacomine et al. (1973). The parameters needed by WASA have been inferred by pedo-transfer functions based on soil texture information using the R packages *soilwaterfun* and *soilwaterptf* available via <http://soilwater.r-forge.r-project.org/>. Vegetation parameters for the types occurring within the study area have been elaborated during the development of WASA (Güntner, 2002) and have been adopted for this study. An updated shapefile of landcover distribution could be obtained from the Brazilian ministry for the environment and the land cover classes have been reclassified to those used in (Güntner, 2002). Data on reservoirs and the geology of the area have been collected and processed within the SESAM project, see <http://www.uni-potsdam.de/sesam>.

It should be stressed that WASA has been initialised by using only independent observational data. No calibration of any model parameters has been done as a comparison with observational data shall not be part of this study. As has been mentioned, however, the model has been used in and proved its ability for the catchment (see references given in section 4.1).

4.4 Sensitivity analysis of user decisions

During the process of model initialisation and landscape discretisation a user commonly has to make a number of often subjective decisions. These are directly influencing the complexity of discretisation and thereby affecting

¹This package is not on CRAN but available via the ECHSE tools library from https://github.com/echse/echse_tools.

Table 1. User decisions affecting landscape discretisation complexity and their realisations used for sensitivity analysis.

Identifier	Meaning	Realisations
SUB_thresh	Minimum size of subbasins in number of grid cells	1000, 2000, 5000, 10000, 30000
EHA_thresh	Minimum size of EHAs in number of grid cells	25, 50, 100, 200, 500, 750, 1000
LU_no	Maximum number of LUs to be classified	5, 10, 20, 50, 75, 100, 150, 200, 250, 300
LU_atts	Number of attributes to be considered during LU classification	1...7
TC_no	Number of TCs to be deviated for every LU	1...5

computational efforts and possibly model results. Due to the complexity of the software, the diversity of links, and potentially non-linear relationships the effects of these user decisions shall be analysed by conducting a number of numerical experiments. User decisions are thereby treated as parameters within a model sensitivity analysis. Each experiment within the analysis stands for a specific realisation of parameters in turn referring to a particular

5 landscape discretisation setup used to drive WASA.

4.4.1 Experimental setup

For landscape discretisation using lumpR five parameters reflecting the most important user decisions have been identified and are summarised in Tab. 1. Their presented realisations are based on a priori knowledge and should span a reasonable value range at the one hand and, at the other hand, result in as much as possible variation in
10 generated spatial units. Therefore, for some parameters such as SUB_thresh, also non-uniform distributions of realisations have been taken into account. All in all it should be noted that a general statement on which parameter realisation is the most plausible and suitable one can hardly be given as the choices are strongly case study specific and partly also depend on available data. What follows is a reasoning on selected parameter realisations for the experiments that can be used as guidelines for lumpR applications.

15 SUB_thresh and EHA_thresh are size thresholds internally directed to the GRASS function *r.watershed* affecting the size and thereby the number of delineated subbasins and EHAs, respectively. As their realisations are given in number of grid cells their choice depends on the resolution of the GRASS location (herein the SRTM DEM resolution of 90m x 90m was used), and catchment size. LU_no and LU_atts control the process of clustering EHAs into LUs. The maximum possible value for LU_atts depends on the available information that can be
20 used for classification. This generally includes the shape of EHAs, their horizontal and vertical extension, and a proxy for hillslope width which are all inferred from a DEM. Further supplemental attributes can be added which in this study included raster maps of soil types, land cover, geology, and SVCs that resulted in a total of seven attributes. For the experiments different realisations of LU_atts shall reflect a differing degree of information

available for the deviation of LUs. Which of the aforementioned attributes are used for a certain experiment is sampled randomly. Each attribute then gets assigned a factor meaning that for every attribute that many LUs shall be deviated. LU_no defines the maximum number of LUs to be generated. Hence, due to the multiplicative nature of the number of LUs to be generated, a factorisation approach is employed to derive the factors for each attribute.

- 5 Herein, one of the sampled attributes is randomly selected and its factor increased by one. This procedure is repeated until the product of factors, i.e., the actual number of LUs, is greater than or equal to LU_no. Finally, TC_no is the number of TCs that will be delineated for every LU. The minimum number of one would effectively result in treating a single LU without further sub-discretisation into TCs. A number larger than five, although this decision is somewhat subjective, seemed rather futile due to increasing computational burden and being harder to interpret in a geomorphologically meaningful way.

- For the sensitivity analysis all possible combinations of parameter realisations have been employed which resulted in a total of 12,250 experiments. This comprises varying complexities within all spatial levels and different degrees of data availability. All discretisations have been realised using lumpR as explained in section 3.1 employing always the same framework and varying only the aforementioned parameters. Finally, WASA has been run with each realisation over a 13 year period whereas the first five years were considered as warm-up and thus have been excluded from the analysis.

4.4.2 Scalar model output

- The target variable of the analysis is the time series of simulated daily river contributions to the Benguê reservoir located at the catchment outlet. However, for conduction of sensitivity analyses a scalar target function is generally needed. As it is impossible to summarise all important characteristics of a streamflow time series in a single scalar value it was decided to employ a whole set of indices and perform the sensitivity analysis for each index separately. The indices are presented and described in Tab. 2. As it can be seen the set of indices describes different aspects of streamflow behaviour ranging from the magnitude of flow (RR , P_{flow} , Q_{avmax}) over flow regime ($SFDC$)² to frequency (f_{low} , f_{high}) and runoff concentration time (RC_{rise} , RC_{fall}).

4.4.3 Analysis method

- There exists a large number of approaches for sensitivity analysis. The selection basically depends on the objective of the study, the nature and complexity of the model, its parameters and outputs, and available computing resources. For the latest review of methods and guidelines for practical application in environmental studies see Pianosi et al. (2016).

- The goals of this analysis were, first, a *ranking* of the described discretisation parameters (sometimes also referred to as *priorisation*) and, second, the identification of those parameters with negligible influence on the respective streamflow index (also referred to as *screening* or *fixing*). As has been described above, model parameters were

²High values stand for a more variable whereas low values represent a more damped flow regime (Sawicz et al., 2011).

Table 2. Streamflow indices used as scalar response functions for sensitivity analysis.

Symbol	Index	Calculation	Unit
RR	Runoff ratio	Sum of daily streamflow values divided by sum of daily precipitation over the whole period of analysis multiplied by 100	%
P_{flow}	Probability for significant streamflow	Number of days with significant ¹ streamflow divided by total number of values multiplied by 100	%
Q_{avmax}	Average annual maximum flow	Average over all annual maximum streamflow values	m^3s^{-1}
$SFDC$	Slope of flow duration curve	Average slope of the flow duration curve for significant ¹ medium ranged ² streamflow values	dimensionless
f_{low}	Frequency of low flows	Average number of insignificant ¹ flow events ³ per year	year^{-1}
f_{high}	Frequency of high flows	Average number of high flow ⁴ events ³ per year	year^{-1}
RC_{rise}	Rate of change during rise	Average rate of change of the rising limbs of high flow ⁴ events ³	$\text{m}^3\text{s}^{-1}\text{day}^{-1}$
RC_{fall}	Rate of change during fall	Average rate of change of the falling limbs of high flow ⁴ events ³	$\text{m}^3\text{s}^{-1}\text{day}^{-1}$

(1) (In-) significant streamflow defined as those values (less than or equal to) larger than $0.01 \text{ m}^3\text{s}^{-1}$.

(2) Values between the 33 % and 66 % percentiles.

(3) An event is defined as a period of consecutive days a certain condition is fulfilled.

(4) High flows are those values being larger than a flow threshold which is herein defined as the 90 % percentile of all significant¹ flow values from all 12,250 model realisations during the analysis period.

sampled over their entire (a priori as meaningful defined) value ranges and varied at the same time within individual runs which leads to a *global* sensitivity analysis with *all-at-a-time* sampling. Based on these considerations Pianosi et al. (2016) suggest a variance- or density-based approach. The former is based on the calculation of sensitivity indices based on the variance of the response function. This, however, requires the assumption that the variance is a good proxy for describing the variation of the value range. For multi-modal or highly-skewed distributions this cannot be guaranteed. In such a case Pianosi et al. (2016) recommend density-based methods. Rather than the variance alone this family of sensitivity analyses considers the probability density function of the response surface. We decided on using the recently introduced PAWN method by Pianosi and Wagener (2015) as this density-based method can cope with skewed distributions and is relatively easy and straightforward to implement.

- 10 PAWN uses empirical approximations of the unconditional cumulative distribution function $F_{y_i}(y_i)$, with y_i being one of the streamflow indices selected as scalar response functions over all 12,250 realisations, and the conditional cumulative distribution functions $F_{y_i|p_j}(y_i)$ where a certain parameter p_j is fixed at a specific value.

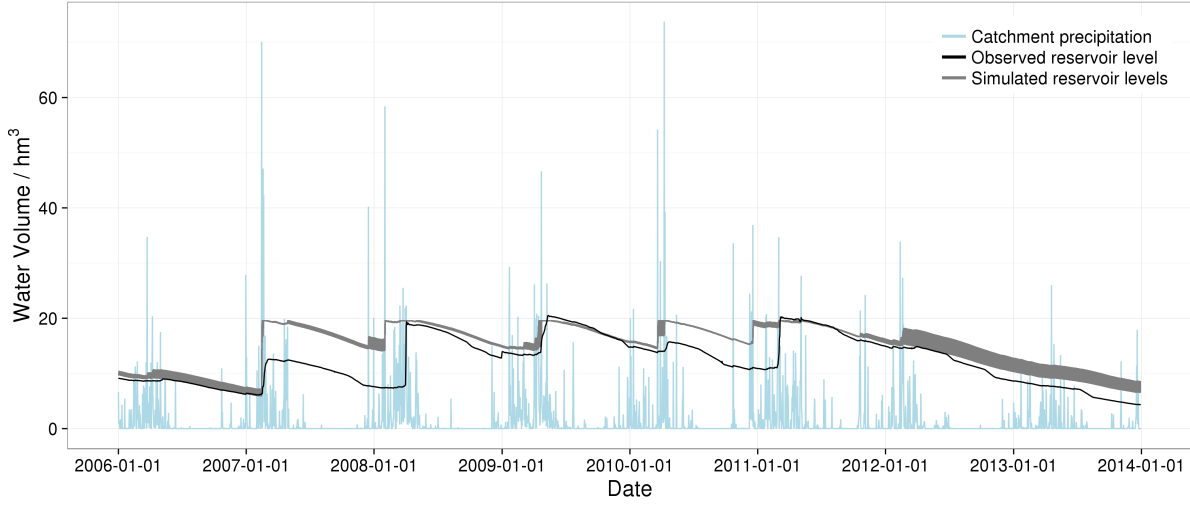


Figure 4. Time series plot of daily resolution of simulated (all 12,250 experiment runs) and observed Benguê reservoir volume and areal precipitation.

The PAWN index T_j as a sensitivity measure can then be calculated for each parameter employing a numerical approximation of the Kolmogorov-Smirnov statistic KS following

$$KS(p_j) = \max_{y_i} \left| F_{y_i}(y_i) - F_{y_i|p_j}(y_i) \right| \quad (1)$$

and

$$5 \quad T_j = \text{median}_{p_j} [KS(p_j)] \quad (2)$$

T_j varies between 0 and 1. It was used to rank the parameters in which lower values of T_j identify the less influential ones. For parameter screening the two-sample Kolmogorov-Smirnov test was employed. It calculates a critical value KS_{crit} below or equal to which a parameter can be judged insignificant as its conditional cumulative distribution function differs significantly from the unconditional one at a certain confidence level α :

$$10 \quad KS_{crit} = c(\alpha) \sqrt{\frac{n+m}{nm}} \quad (3)$$

with n and m being the number of samples to estimate $F_{y_i}(y_i)$ and $F_{y_i|p_j}(y_i)$, respectively, and $c(\alpha)$ being a literature value. Herein we set $\alpha = 0.05$ and thus $c(\alpha) = 1.36$.

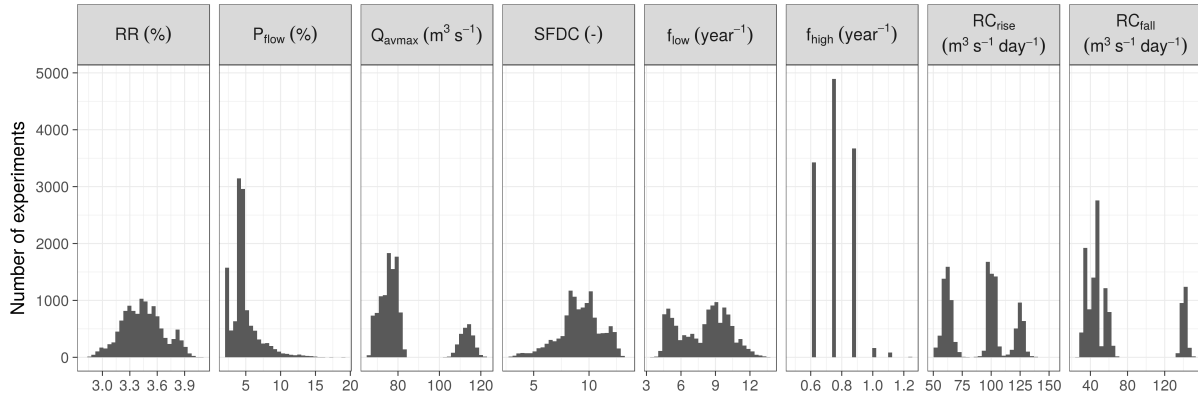


Figure 5. Histograms showing the value distributions of streamflow indices over the 12,250 experiments.

4.5 Results

Figure 4 provides an overview over the daily model results in comparison to observations for the Benguê reservoir at the catchment outlet together with the catchment's areal precipitation used to drive the model. Absolute deviations of uncalibrated model runs from observations are large whereas qualitative behaviour is matched well.

- 5 However, in some years (2008 and 2011) the model simulates reservoir filling within the rainy season much earlier than observed. For the simulations, a rainfall volume of at least about 35 hm³ seems to be necessary to produce significant reservoir inflow. Simulated reservoir depletion is often slower than it can be observed. Overall it should be noted that variability within the experiments is rather small whereas deviations seem to be originated from differing sensitivities of streamflow to a certain rainfall input.
- 10 Figure 5 gives an overview of the streamflow index value distributions from the 12,250 experiments. It can be seen that some indices, namely Q_{avmax} , f_{high} , RC_{rise} , and RC_{fall} , show distinct multi-modal value distributions. Distributions for the other indices are slightly bi-modal (f_{low}) or skewed (P_{flow} and $SFDC$) with RR as the only index being more or less normal distributed. In general, the runoff coefficient RR is consistently low ranging between 3 % and 4 % and streamflow can be characterised as ephemeral due to a low probability of days showing
- 15 significant streamflow (P_{flow} about 5 % for most of the experiments). Contrary to low flow periods, high flow events do rarely occur whereas the variability between experiments is high (f_{low} and f_{high} , respectively). Peak flows (Q_{avmax}) as well as runoff concentrations (characterised by RC_{rise} and RC_{fall}) vary considerably between experiments also exhibiting multi-modal distributions.

- Screening and ranking of landscape discretisation parameters is exhibited with Fig. 6. The size of subbasins
- 20 (SUB_thresh) is the most influential parameter for all indices except for those being related to low flow characterisation (f_{low} and P_{flow}) which are dominated by the size of EHAs (EHA_thresh). The number of LUs (LU_no) can be regarded as the third important parameter being especially of relevance for high flow related

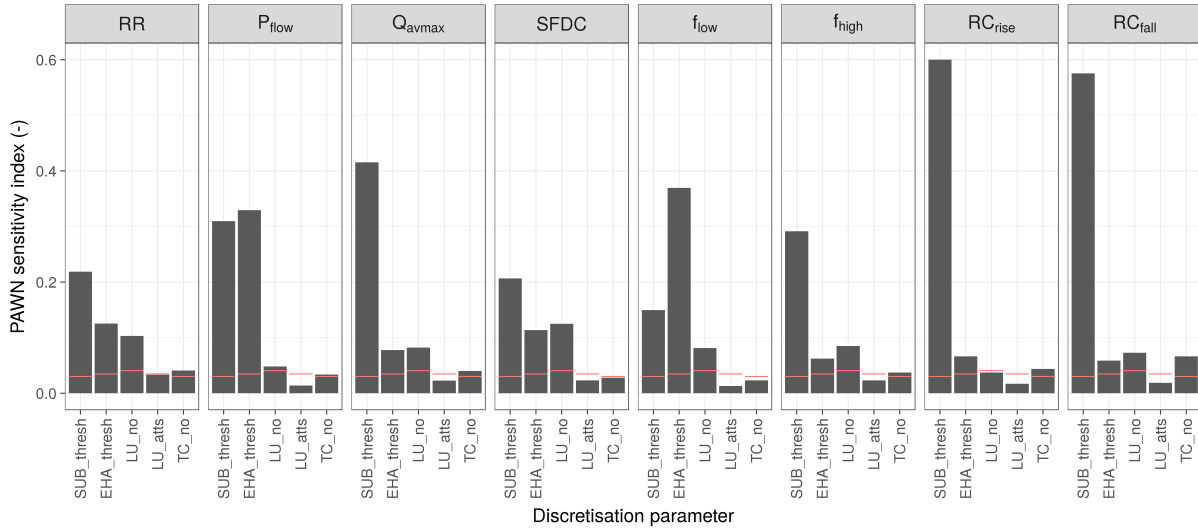


Figure 6. Barplots of PAWN sensitivity indices for each streamflow index and landscape discretisation parameter. Red lines indicate critical Kolmogorov-Smirnov values (KS_{crit}), i.e., parameters for which the PAWN index is less than or equal to this value can be regarded as insignificant for a specific streamflow index.

indices (f_{high} and Q_{avmax}) and, to some extent, flow regime ($SFDC$) and runoff concentration (RC_{fall}). The least important parameters are the number of TCs (TC_{no}), except for runoff concentration (RC_{rise} and RC_{fall}), and the number of attributes considered for LU classification (LU_{atts}) which is insignificant for all streamflow indices.

- 5 More information on parameter influences on the various streamflow indices can be obtained from Fig. 7. The more influential a parameter the more colourful a diagram, i.e., the larger the deviations of the conditional empirical cumulative distribution functions from the unconditional one. Many of the diagrams show a clear relationship between parameter realisation and streamflow index value. The larger the subbasins (i.e., the larger SUB_thresh and the less the number of subbasins in the area) the smaller the generated amount of runoff (RR gets smaller) and
- 10 the smaller the probability of significant runoff (P_{flow}). On the other hand, peak discharges (Q_{avmax}) increase and the catchment appears to produce more rapid runoff responses (higher values of RC_{fall} and RC_{rise}). Furthermore it can be seen that SUB_thresh is responsible for the multi-modal distributions of Q_{avmax} , f_{high} , RC_{fall} , and RC_{rise} as the conditional distribution functions show a less stepped shape than the unconditional functions. The influence on f_{low} appears to be less clear. It can be seen, however, that larger values of SUB_thresh result in
- 15 a more pronounced bimodal distribution of that index. Low flows are otherwise more dominated by the size of EHAs (EHA_thresh). The larger EHA_thresh (i.e., the larger the EHAs and the lower their number) the less the probability for significant streamflow (P_{flow}) and the less the number of low flow events per year (f_{low}) while the flow regime becomes more variable (higher values of $SFDC$). Higher numbers of LUs (LU_{no}) result in more

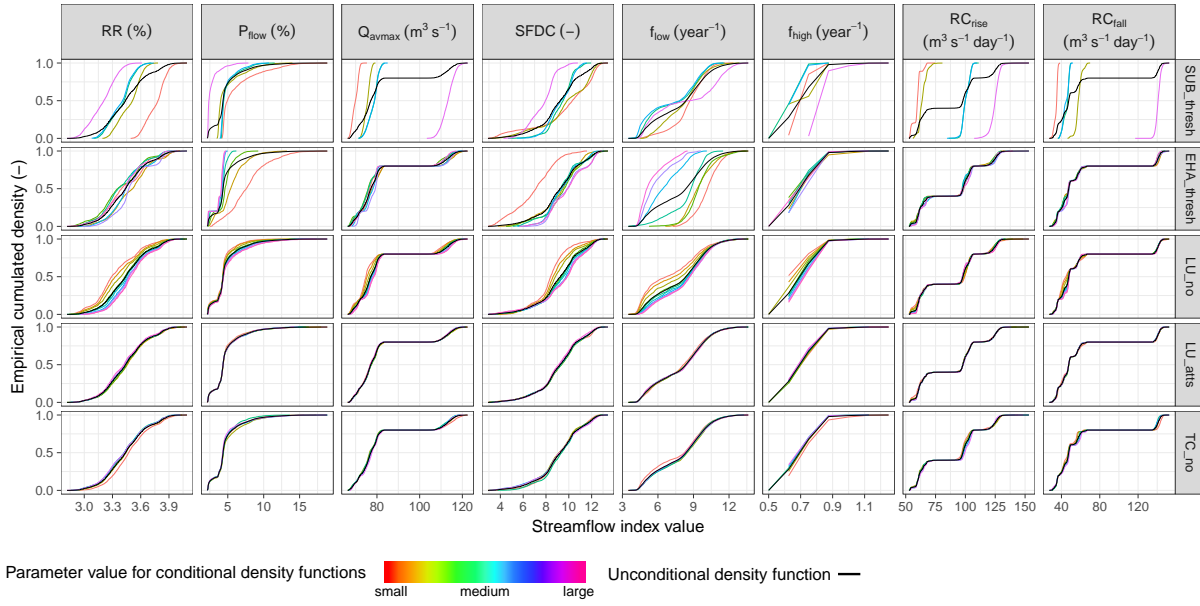


Figure 7. Unconditional (black lines) and conditional (rainbow coloured lines) empirical cumulative distribution functions for each streamflow index–discretisation parameter combination.

generated runoff (RR increases) with a tendency to both higher frequencies of low flow and high flow events, and generate more variable streamflow regimes.

5 Discussion

5.1 lumpR: features, benefits, limitations

- 5 It is still common practise for many researchers in the field of hydrological modelling to not automate their pre-processing steps and even if they do they write their own scripts which they do not distribute openly. Furthermore, common limitations of existing software are that they are often model specific, and/or perform only certain steps of pre-processing, and/or are commercial or can only be used along with commercial software (e.g., ArcGIS or MATLAB). On the other hand, the importance of FOSS GIS is increasing (e.g., Steiniger and Bocher, 2009;
- 10 Branger et al., 2012). Our work tries to address these limitations and builds on the philosophy of FOSS.

With lumpR a supplemental package for the free and open-source programming language R has been developed. It is a tool for the discretisation of landscapes for use along with hydrological computer models employing hierarchical spatial aggregation schemes with hillslopes as the fundamental terrain element. The software has been so far tested under Windows as well Linux-based operating systems such as openSUSE and Ubuntu. It is user-friendly

- 15 in a way that it interacts with the GIS GRASS and thus allows graphical investigation and manual correction of

outcomes. Furthermore, it hides unnecessary detail from the user while at the same time ensuring a certain level of control over the discretisation process. As R is a scripting language, model pre-processing in that way can easily be automated and reproduced. Via the database tools the software allows to keep the output directories clearly arranged by putting all information into a database while supporting several database systems to do so. On the other hand, depending on the user's needs, graphical outputs and files with intermediate results can be used for investigation or debugging.

Although in this study lumpR was tested together with the model WASA and the package provides a function to create the model's input files, it can be regarded as model independent. First tests employing the simulation environment ECHSE (Kneis, 2015) revealed the package to be easily adaptable for producing the input files needed by a specific model. Furthermore, it is easily possible to manually export and arrange the values from the databases as needed. Other candidate models are, in principal, all hydrological models with similarly complex spatial aggregation schemes as used by WASA that can make use of the information provided by lumpR, such as the WEPP model (Flanagan and Nearing, 1995).³

The centrepiece of the package is the LUMP algorithm introduced by Francke et al. (2008) for the calculation of hillslope properties and the delineation of representative LUs together with the subdivision into TCs. In contrast to simple GIS overlay techniques, as usually employed for the delineation of HRUs, it preserves information on the distribution of hillslope parameters and their relative topographic position. In that way the task of landscape discretisation and parametrisation can be purposefully directed to landscape properties dominating hydrological response. By the decision on certain parameter realisations it has been shown in the example application that discretisations of varying complexity can be easily produced. A user can thus experiment and find out the optimal degree of complexity for a certain catchment and a specific objective, e.g., by systematically employing a multiple hypothesis framework. Via its integration into lumpR the application of the algorithm could be further simplified and harmonised as the original LUMP consisted of a loose collection of less user-friendly scripts, partly also relying on non-free software such as MATLAB. Furthermore, lumpR comes with a lot more additional functionalities than the mere LU and TC deviation.

Even so the package addresses a number of common limitations within existing software it is not free of constraints. First of all its GRASS support is already deprecated as it builds, by the time of writing this paper, on GRASS version 6.4 whereas the first stable release of version 7 with a lot of improvements has been published in 2015. However, due limited backwards compatibility the adaptation of lumpR demands a profound revision of the source code and is thus still planned as future enhancement. This fact, however, comes with no further restriction except that a potential lumpR user will need to have the old GRASS 6.4 to be installed. With regards to the applicability over large datasets, i.e., when applying the package to large areas in the order of $>100,000 \text{ km}^2$ and/or when employing high-resolution DEMs, time consumption might pose a restriction. Therefore, future enhancements also need to include improvements regarding computational efficiency.

³Readers considering lumpR for use with their model are encouraged to contact the corresponding author of this paper for support.

A limitation more related to the algorithm is the restricted use of morphological parameters, currently limited to shape, horizontal and vertical extension, and hillslope width. Future improvements should include the use of other factors relevant for hillslopes characterisation such as, for instance, contour curvature being related to the convergence or divergence of flowpaths and as such being of hydrological relevance (Bogaart and Troch, 2006).

5 Furthermore, the software is not able to automatically distinguish and account for artificial hydrological discontinuities such as ditches and field boundaries or other problematic formations such as large flat areas in a DEM as produced by lakes. Such elements may negatively influence the delineation of hillslopes and LUs and should, if being of significant influence, be masked in GRASS before the analysis. In addition, the pre- and post-processing steps within lumpR (i.e., functions *lump_grass_prep()* and *lump_grass_post()*) still partly suffer from imperfect

10 or crude approaches. This affects in particular the deviation of the river network, subbasin delineation, and the approximation of streamflow routing parameters (the latter tailored to the simplistic unit hydrograph approach of WASA). In this respect future enhancements should also include a review on latest advancements of terrain analysis and parameterisation and the refinement of employed algorithms.

5.2 On user decisions and sensitivity analysis

15 As in science reproducibility and objectivity are primary criteria for any investigation it has to be noted that the LUMP algorithm, while providing some freedom of choices to the user, at the same time suffers from a certain degree of subjectivity. This has been explicitly addressed via the sensitivity analysis within the example application and its results shall be discussed in the following.

The results from the 12,250 realisations of landscape discretisation show only little difference with respect to

20 water storage dynamics of the Banguê reservoir (Fig. 4). The small variation of the runoff coefficient (cf. Fig. 5) further supports the conclusion that decisions on landscape discretisation parametrisation only have a minor impact on simulated runoff volume. On the other hand, the influence on other indices describing runoff concentrations and dynamics, and the frequency of flood or drought events is much more obvious.

The hydrological regime of the study area is primarily influenced by precipitation which is characterised by a

25 high temporal concentration and a large temporal and spatial variability. A comparison with the data reported in Medeiros and de Araújo (2014) further supports that uncertainties regarding the precipitation input to the model have a much larger impact on simulation results than user decisions on LUMP parameters. In their study Medeiros and de Araújo (2014) used a set of raw station data in contrast to the preprocessed and gridded dataset by Xavier et al. (2016) used for our experiments and their runoff values have been assessed by taking the Benguê reservoir

30 inflows computed from water balance calculations. Their reported runoff coefficients are mostly lower than ours, even when precipitation is higher, and shows less inter-annual variation (see Fig. A1 in Appendix). The precipitation forcing for the experiments was implicitly influenced by variable subbasin sizes and numbers as the WASA model derived rainfall input at the subbasin level. The variability in precipitation input among the experiments,

however, appears to be negligible as it is generally less than 6 mm for daily values and less than 10 mm for yearly sums (see Fig. A2 in Appendix).

Regarding the method of sensitivity analysis it can be concluded that the choice for a density-based approach was reasonable as most of the analysed streamflow indices exhibit multi-modal or skewed value distributions.

5 Furthermore, the interpretation of the PAWN index proofed to be straightforward. On the other hand, also the whole parameter range has been sampled, the definition of reasonable ranges was based on subjective decisions and data availability. Another drawback on the analysis approach is that only first order effects of parameter sensitivities have been quantified while interactions among parameters have been neglected. It might thus be that insignificant parameters (i.e., the parameter `LU_atts`) have significant higher order effects due to parameter correlations.

10 Overall, the example study and sensitivity analysis was strongly case study specific and conclusions are thus limited to regions with similar hydro-climatic conditions, i.e., semi-arid areas without substantial groundwater influences mainly characterised by spatially and temporally heterogeneous precipitation patterns.

Tobias:

Does the Discussion section already contain too much conclusions?

6 Conclusions

15 TODO

7 Code availability

TODO

8 Data availability

Meteorological data are available from <http://careyking.com/data-downloads/>. DEM raw data can be obtained via

20 <http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp> by selecting tile 28/14 (horizontal/vertical). Reservoir data and the geology map have been processed within the SESAM project and are not publicly available. For more information and contact details see <http://www.uni-potsdam.de/sesam>. Land cover and soil raster maps are not publicly available.

Appendix A: Precipitation uncertainty and comparisons with other studies

Tobias:
Figures herein are still preliminary and need adjustments (legend, larger font size etc.).

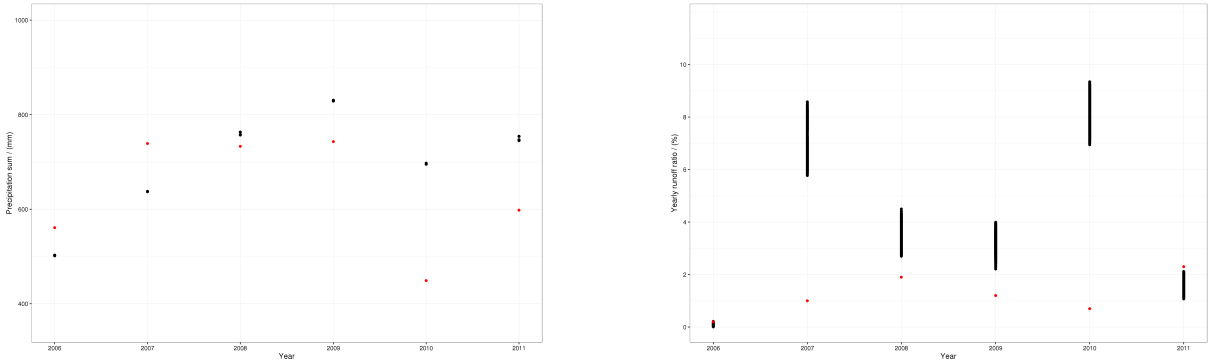


Figure A1. Comparison of precipitation forcing (left) and simulated yearly runoff coefficients (right) with values reported by Medeiros and de Araújo (2014) (red dots).

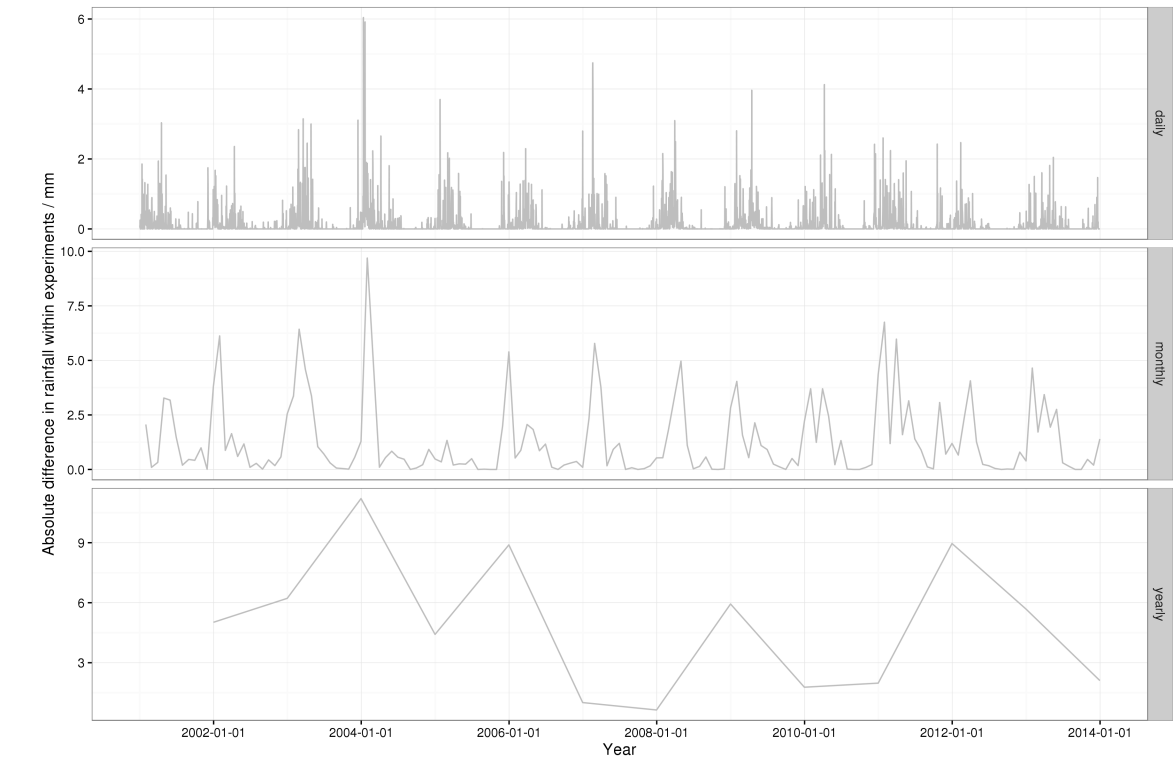


Figure A2. Maximum of absolute differences in rainfall between the 12,250 experiments for every time step.

Author contributions. TODO

Acknowledgements. Authors thank Fernanda Scholz (former M.Sc. student at the Leibniz University of Hannover) for the preparation of soil data. M. Sc. student Lisa Berghäuser from the University of Potsdam is acknowledged for testing lumpR within her Thesis and giving valuable feedback and bug reports. Tobias Pilz acknowledges funding by the Helmholtz graduate

5 school GeoSim.

Tobias:
Any more ac-
knowledgements
from your side?

References

- Band, L. E., Tague, C. L., Brun, S. E., Tenenbaum, D. E., and Fernandes, R. A.: Modelling Watersheds as Spatial Object Hierarchies: Structure and Dynamics, *Transactions in GIS*, 4, 181–196, doi:10.1111/1467-9671.00048, 2000.
- Beven, K., Calver, A., and Morris, E. M.: The Institute of Hydrology distributed model, IH Report 98, Institute of Hydrology, Wallingford, UK, 1987.
- Bogaart, P. W. and Troch, P. A.: Curvature distribution within hillslopes and catchments and its effect on the hydrological response, *Hydrol. Earth Syst. Sc.*, 10, 925–936, doi:10.5194/hess-10-925-2006, 2006.
- Branger, F., Jankowsky, S., Vannier, O., Viallet, P., Debionne, S., and Braud, I.: Use of Open-Source GIS for the Pre-processing of Distributed Hydrological Models, in: *Geospatial Free and Open Source Software in the 21st Century: Proceedings of the first Open Source Geospatial Research Symposium, OGRS 2009*, edited by Bocher, E. and Neteler, M., chap. 3, pp. 35–48, Springer Berlin Heidelberg, doi:10.1007/978-3-642-10595-1_3, 2012.
- Bronstert, A.: Capabilities and limitations of detailed hillslope hydrological modelling, *Hydrol. Process.*, 13, 21–48, doi:10.1002/(SICI)1099-1085(199901)13:1<21::AID-HYP702>3.0.CO;2-4, 1999.
- Bronstert, A., de Araújo, J.-C., Batalla, R. J., Costa, A. C., Delgado, J. M., Francke, T., Foerster, S., Guentner, A., López-Tarazón, J. A., Mamede, G. L., Medeiros, P. H., Mueller, E., and Vericat, D.: Process-based modelling of erosion, sediment transport and reservoir siltation in mesoscale semi-arid catchments, *J. Soils Sediments*, 14, 2001–2018, doi:10.1007/s11368-014-0994-1, 2014.
- Cochrane, T. and Flanagan, D.: Representative hillslope methods for applying the WEPP model with DEMs and GIS, *T. ASAE*, 46, 1041–1049, 2003.
- Costa-Cabral, M. C. and Burges, S. J.: Digital Elevation Model Networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas, *Water Resour. Res.*, 30, 1681–1692, doi:10.1029/93WR03512, 1994.
- de Araújo, J. C. and Medeiros, P. H. A.: Impact of Dense Reservoir Networks on Water Resources in Semiarid Environments, *Australian Journal of Water Resources*, 17, 87–100, doi:10.7158/13241583.2013.11465422, 2013.
- de Figueiredo, J. V., de Araújo, J. C., Medeiros, P. H. A., and Costa, A. C.: Runoff initiation in a preserved semiarid Caatinga small watershed, Northeastern Brazil, *Hydrol. Process.*, 30, 2390–2400, doi:10.1002/hyp.10801, 2016.
- DeVantier, B. A. and Feldman, A. D.: Review of GIS Applications in Hydrologic Modeling, *J. Water Res. Pl.-ASCE*, 119, 246–261, doi:10.1061/(ASCE)0733-9496(1993)119:2(246), 1993.
- Di Luzio, M., Srinivasan, R., and Arnold, J. G.: A GIS-Coupled Hydrological Model System for the Watershed Assessment of Agricultural Nonpoint and Point Sources of Pollution, *Transactions in GIS*, 8, 113–136, doi:10.1111/j.1467-9671.2004.00170.x, 2004.
- Euser, T., Hrachowitz, M., Winsemius, H. C., and Savenije, H. H.: The effect of forcing and landscape distribution on performance and consistency of model structures, *Hydrol. Process.*, 29, 3727–3743, doi:10.1002/hyp.10445, 2015.
- Flanagan, D. C. and Nearing, M. A.: USDA Water Erosion Prediction Project – Hillslope Profile and Watershed Model Documentation, NSERL Report 10, USDA-ARS National Soil Erosion Research Laboratory, West Lafayette, Indiana, USA, 1995.
- Flanagan, D. C., Frankenberger, J. R., Cochrane, T. A., Renschler, C. S., and Elliot, W. J.: Geospatial Application of the Water Erosion Prediction Project (WEPP) Model, *T. ASABE*, 56, 591–601, doi:10.13031/2013.42681, 2013.

- Flügel, W.-A.: Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany, *Hydrol. Process.*, 9, 423–436, doi:10.1002/hyp.3360090313, 1995.
- Francke, T., Güntner, A., Mamede, G., Müller, E. N., and Bronstert, A.: Automated catena-based discretization of landscapes for the derivation of hydrological modelling units, *Int. J. Geogr. Inf. Sci.*, 22, 111–132, doi:10.1080/13658810701300873, 2008.
- Freitas, H. R. d. A., Freitas, C. d. C., Rosim, S., and Oliveira, J. R. d. F.: Drainage networks and watersheds delineation derived from TIN-based digital elevation models, *Comput. Geosci.*, 92, 21–37, doi:10.1016/j.cageo.2016.04.003, 2016.
- Gallant, J. C. and Wilson, J. P.: TAPES-G: A grid-based terrain analysis program for the environmental sciences, *Comput. Geosci.*, 22, 713–722, doi:10.1016/0098-3004(96)00002-7, 1996.
- Garbrecht, J. and Martz, L. W.: An Overview of TOPAZ: An Automated Digital Landscape Analysis Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation, and Subcatchment Parameterization, *Grazinglands Research Laboratory, USDA, Agricultural Research Service, El Reno, Oklahoma, USA, report no. GRL 99-1*, 1999.
- Gerstengarbe, F.-W. and Werner, P. C.: Estimation of the beginning and end of recurrent events within a climate regime, *Clim. Res.*, 11, 97–107, 1999.
- Grayson, R. B., Moore, I. D., and McMahon, T. A.: Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes, *Water Resour. Res.*, 28, 2639–2658, doi:10.1029/92WR01258, 1992.
- Güntner, A.: Large-scale hydrological modelling in the semi-arid North-East of Brazil, *PIK Report 77*, Potsdam Institute for Climate Impact Research, Potsdam, Germany, 2002.
- Güntner, A. and Bronstert, A.: Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas, *J. Hydrol.*, 297, 136–161, doi:10.1016/j.jhydrol.2004.04.008, 2004.
- Haghnegahdar, A., Tolson, B. A., Craig, J. R., and Paya, K. T.: Assessing the performance of a semi-distributed hydrological model under various watershed discretization schemes, *Hydrol. Process.*, 29, 4018–4031, doi:10.1002/hyp.10550, 2015.
- Han, J.-C., Huang, G.-H., Zhang, H., Li, Z., and Li, Y.-P.: Effects of watershed subdivision level on semi-distributed hydrological simulations: case study of the SLURP model applied to the Xiangxi River watershed, China, *Hydrolog. Sci. J.*, 59, 108–125, doi:10.1080/02626667.2013.854368, 2014.
- Haverkamp, S., Fohrer, N., and Frede, H.-G.: Assessment of the effect of land use patterns on hydrologic landscape functions: a comprehensive GIS-based tool to minimize model uncertainty resulting from spatial aggregation, *Hydrol. Process.*, 19, 715–727, doi:10.1002/hyp.5626, 2005.
- Ivanov, V. Y., Vivoni, E. R., Bras, R. L., and Entekhabi, D.: Catchment hydrologic response with a fully distributed triangulated irregular network model, *Water Resour. Res.*, 40, W11 102, doi:10.1029/2004WR003218, w11102, 2004.
- Jackisch, C., Zehe, E., Samaniego, L., and Singh, A. K.: An experiment to gauge an ungauged catchment: rapid data assessment and eco-hydrological modelling in a data-scarce rural catchment, *Hydrolog. Sci. J.*, 59, 2103–2125, doi:10.1080/02626667.2013.870662, 2014.
- Jacomine, P. K. T., Almeida, J. C., and Medeiros, L. A. R.: Levantamento exploratorio – Reconhecimento de solos do Estado do Ceará, vol. 1, *DNPEA, DRN-SUDENE, Recife, Brazil*, 1973.

- Kinner, D., Mitasova, H., Stallard, R., Harmon, R. S., and Toma, L.: GIS-Based Stream Network Analysis for the Upper Río Chagres Basin, Panama, in: *The Río Chagres, Panama: A Multidisciplinary Profile of a Tropical Watershed*, edited by Harmon, R. S., chap. 6, pp. 83–95, Springer Netherlands, doi:10.1007/1-4020-3297-8_6, 2005.
- Kite, G. W.: Scaling of Input Data for Macroscale Hydrologic Modeling, *Water Resour. Res.*, 31, 2769–2781, doi:10.1029/95WR02102, 1995.
- Kneis, D.: A lightweight framework for rapid development of object-based hydrological model engines, *Environmental Modelling & Software*, 68, 110–121, doi:10.1016/j.envsoft.2015.02.009, 2015.
- Kouwen, N., Soulis, E. D., Pietroniro, A., Donald, J., and Harrington, R. A.: Grouped Response Units for Distributed Hydrologic Modeling, *J. Water Res. Pl.-ASCE*, 119, 289–305, doi:10.1061/(ASCE)0733-9496(1993)119:3(289), 1993.
- 10 Krol, M. S., de Vries, M. J., Oel, P. R., and de Araújo, J. C.: Sustainability of Small Reservoirs and Large Scale Water Availability Under Current Conditions and Climate Change, *Water Resour Manage*, 25, 3017 – 3026, doi:10.1007/s11269-011-9787-0, 2011.
- Kryanova, V., Müller-Wohlfeil, D.-I., and Becker, A.: Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds, *Ecol. Model.*, 106, 261–289, doi:10.1016/S0304-3800(97)00204-4, 1998.
- 15 Kumar, R., Samaniego, L., and Attinger, S.: The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics, *J. Hydrol.*, 392, 54–69, doi:10.1016/j.jhydrol.2010.07.047, 2010.
- Lacroix, M. P., Martz, L. W., Kite, G. W., and Garbrecht, J.: Using digital terrain analysis modeling techniques for the parameterization of a hydrologic model, *Environ. Modell. Softw.*, 17, 125–134, doi:10.1016/S1364-8152(01)00042-1, 2002.
- Lagacherie, P., Rabotin, M., Colin, F., Moussa, R., and Voltz, M.: Geo-MHYDAS: A landscape discretization tool for distributed hydrological modeling of cultivated areas, *Comput. Geosci.*, 36, 1021–1032, doi:10.1016/j.cageo.2009.12.005, 2010.
- 20 Leavesley, G. H., Lichty, R. W., Troutman, B. M., and Saindon, L. G.: *Precipitation-Runoff Modeling System: User’s manual*, Water-resources investigations report 83-4238, United States Department of the Interior, Denver, Colorado, USA, 1983.
- Manguerra, H. B. and Engel, B. A.: Hydrologic parameterization of watersheds for runoff prediction using SWAT, *J. Am. Water Resour. As.*, 34, 1149–1162, doi:10.1111/j.1752-1688.1998.tb04161.x, 1998.
- 25 Markstrom, S., Niswonger, R., Regan, R., Prudic, D., and Barlow, P.: *GSFLOW–Coupled ground-water and surface-water flow model based on the integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005)*, U.S. Geological Survey Techniques and Methods 6-D1, 240 p., 2008.
- Maunder, C. J.: An automated method for constructing contour-based digital elevation models, *Water Resour. Res.*, 35, 3931–3940, doi:10.1029/1999WR900166, 1999.
- 30 Maurer, T.: *Physikalisch begründete, zeitkontinuierliche Modellierung des Wassertransports in kleinen ländlichen Einzugsgebieten*, Dissertation, Mitteilungen Inst. f. Hydrologie u. Wasserwirtschaft, 61, Universität Karlsruhe, 1997.
- Medeiros, P. H. A. and de Araújo, J. C.: Temporal variability of rainfall in a semiarid environment in Brazil and its effect on sediment transport processes, *J. Soils Sediments*, 14, 1216–1223, doi:10.1007/s11368-013-0809-9, 2014.
- Medeiros, P. H. A., Güntner, A., Francke, T., Mamede, G. L., and de Araújo, J. C.: Modelling spatio-temporal patterns of sediment yield and connectivity in a semi-arid catchment with the WASA-SED model, *Hydrolog. Sci. J.*, 55, 636–648, doi:10.1080/02626661003780409, 2010.
- 35

- Medeiros, P. H. A., de Araújo, J. C., Mamede, G. L., Creutzfeldt, B., Güntner, A., and Bronstert, A.: Connectivity of sediment transport in a semiarid environment: a synthesis for the Upper Jaguaribe Basin, Brazil, *J. Soils Sediments*, 14, 1938–1948, doi:10.1007/s11368-014-0988-z, 2014.
- Metz, M., Mitasova, H., and Harmon, R. S.: Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search, *Hydrol. Earth Syst. Sc.*, 15, 667–678, doi:10.5194/hess-15-667-2011, 2011.
- Miller, S., Semmens, D., Miller, R., Hernandez, M., Goodrich, D., Miller, W., Kepner, W., and Ebert, D.: GIS-based Hydrologic Modeling: The Automated Geospatial Watershed Assessment Tool, in: Second Federal Interagency Hydrologic Modeling Conference, Las Vegas, Nevada, USA, 2002.
- Miller, S. N., Semmens, D. J., Goodrich, D. C., Hernandez, M., Miller, R. C., Kepner, W. G., and Guertin, D. P.: The Automated Geospatial Watershed Assessment tool, *Environ. Modell. Softw.*, 22, 365–377, doi:10.1016/j.envsoft.2005.12.004, 2007.
- Mitasova, H. and Neteler, M.: GRASS as Open Source Free Software GIS: Accomplishments and Perspectives, *Transactions in GIS*, 8, 145–154, doi:10.1111/j.1467-9671.2004.00172.x, 2004.
- Moore, I. D. and Grayson, R. B.: Terrain-based catchment partitioning and runoff prediction using vector elevation data, *Water Resour. Res.*, 27, 1177–1191, doi:10.1029/91WR00090, 1991.
- Moore, I. D., O’Loughlin, E. M., and Burch, G. J.: A contour-based topographic model for hydrological and ecological applications, *Earth Surf. Proc. Land.*, 13, 305–320, doi:10.1002/esp.3290130404, 1988.
- Moore, I. D., Grayson, R. B., and Ladson, A. R.: Digital terrain modelling: A review of hydrological, geomorphological, and biological applications, *Hydrol. Process.*, 5, 3–30, doi:10.1002/hyp.3360050103, 1991.
- Moretti, G. and Orlandini, S.: Automatic delineation of drainage basins from contour elevation data using skeleton construction techniques, *Water Resour. Res.*, 44, W05 403, doi:10.1029/2007WR006309, 2008.
- Moussa, R., Voltz, M., and Andrieux, P.: Effects of the spatial organization of agricultural management on the hydrological behaviour of a farmed catchment during flood events, *Hydrol. Process.*, 16, 393–412, doi:10.1002/hyp.333, 2002.
- Mueller, E. N., Francke, T., Batalla, R. J., and Bronstert, A.: Modelling the effects of land-use change on runoff and sediment yield for a meso-scale catchment in the Southern Pyrenees, *CATENA*, 79, 288–296, doi:10.1016/j.catena.2009.06.007, 2009.
- Mueller, E. N., Güntner, A., Francke, T., and Mamede, G.: Modelling sediment export, retention and reservoir sedimentation in drylands with the WASA-SED model, *Geoscientific Model Development*, 3, 275–291, doi:10.5194/gmd-3-275-2010, 2010.
- Neteler, M., Bowman, M. H., Landa, M., and Metz, M.: GRASS GIS: A multi-purpose open source GIS, *Environ. Modell. Softw.*, 31, 124–130, doi:10.1016/j.envsoft.2011.11.014, 2012.
- Noël, P., Rousseau, A. N., Paniconi, C., and Nadeau, D. F.: Algorithm for Delineating and Extracting Hillslopes and Hillslope Width Functions from Gridded Elevation Data, *J. Hydrol. Eng.*, 19, 366–374, doi:10.1061/(ASCE)HE.1943-5584.0000783, 2014.
- O’Callaghan, J. F. and Mark, D. M.: The extraction of drainage networks from digital elevation data, *Comput. Vision Graph.*, 28, 323–344, doi:10.1016/S0734-189X(84)80011-0, 1984.
- Passalacqua, P., Do Trung, T., Foufoula-Georgiou, E., Sapiro, G., and Dietrich, W. E.: A geometric framework for channel network extraction from lidar: Nonlinear diffusion and geodesic paths, *J GEOPHYS RES-EARTH*, 115, F01 002, doi:10.1029/2009JF001254, 2010.
- Pianosi, F. and Wagener, T.: A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, *Environ. Modell. Softw.*, 67, 1–11, doi:10.1016/j.envsoft.2015.01.004, 2015.

- Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, *Environ. Modell. Softw.*, 79, 214–232, doi:10.1016/j.envsoft.2016.02.008, 2016.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>, accessed 5 July 2016, 2015.
- Reggiani, P., Sivapalan, M., and Hassanizadeh, S. M.: A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, *Advances in Water Resources*, 22, 367–398, doi:10.1016/S0309-1708(98)00012-8, 1998.
- Renschler, C. S.: Designing geo-spatial interfaces to scale process models: the GeoWEPP approach, *Hydrol. Process.*, 17, 1005–1017, doi:10.1002/hyp.1177, 2003.
- Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resour. Res.*, 46, W05 523, doi:10.1029/2008WR007327, 2010.
- Sangireddy, H., Stark, C. P., Kladzyk, A., and Passalacqua, P.: GeoNet: An open source software for the automatic and objective extraction of channel heads, channel network, and channel morphology from high resolution topography data, *Environ. Modell. Softw.*, 83, 58–73, doi:10.1016/j.envsoft.2016.04.026, 2016.
- Sanzana, P., Jankowsky, S., Branger, F., Braud, I., Vargas, X., Hitschfeld, N., and Gironás, J.: Computer-assisted mesh generation based on hydrological response units for distributed hydrological modeling, *Comput. Geosci.*, 57, 32–43, doi:10.1016/j.cageo.2013.02.006, 2013.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sc.*, 15, 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.
- Schwartz, C.: Deriving Hydrological Response Units (HRUs) using a Web Processing Service implementation based on GRASS GIS, *Geoinformatics FCE CTU*, 3, 67–78, 2008.
- Smith, T., Marshall, L., McGlynn, B., and Jencso, K.: Using field data to inform and evaluate a new model of catchment hydrologic connectivity, *Water Resour. Res.*, 49, 6834–6846, doi:10.1002/wrcr.20546, 2013.
- Steiniger, S. and Bocher, E.: An overview on current free and open source desktop GIS developments, *Int. J. Geogr. Inf. Sci.*, 23, 1345–1370, doi:10.1080/13658810802634956, 2009.
- Tague, C. and Band, L.: RHESSys: Regional Hydro-ecologic simulation system: An object-oriented approach to spatially distributed modeling of carbon, water and nutrient cycling, *Earth Interact.*, 8, 1–42, 2004.
- Tarboton, D.: Terrain Analysis Using Digital Elevation Models in Hydrology, in: 23rd ESRI International Users Conference, San Diego, California, USA, 2003.
- Tarboton, D. G., Bras, R. L., and Rodriguez-Iturbe, I.: On the extraction of channel networks from digital elevation data, *Hydrol. Process.*, 5, 81–100, doi:10.1002/hyp.3360050107, 1991.
- Troch, P. A., Paniconi, C., and Emiel van Loon, E.: Hillslope-storage Boussinesq model for subsurface flow and variable source areas along complex hillslopes: 1. Formulation and characteristic response, *Water Resour. Res.*, 39, doi:10.1029/2002WR001728, 1316, 2003.

- Tucker, G. E., Lancaster, S. T., Gasparini, N. M., Bras, R. L., and Rybarczyk, S. M.: An object-oriented framework for distributed hydrologic and geomorphic modeling using triangulated irregular networks, *Comput. Geosci.*, 27, 959–973, doi:10.1016/S0098-3004(00)00134-5, 2001.
- Viviroli, D., Zappa, M., Gurtz, J., and Weingartner, R.: An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools, *Environ. Modell. Softw.*, 24, 1209–1222, doi:10.1016/j.envsoft.2009.04.001, 2009.
- Vivoni, E. R., Ivanov, V. Y., Bras, R. L., and Entekhabi, D.: Generation of Triangulated Irregular Networks Based on Hydrological Similarity, *J. Hydrol. Eng.*, 9, 288–302, doi:10.1061/(ASCE)1084-0699(2004)9:4(288), 2004.
- Weiler, M. and Beven, K.: Do we need a Community Hydrological Model?, *Water Resour. Res.*, 51, 7777–7784, doi:10.1002/2014WR016731, 2015.
- 10 Winter, T. C.: The concept of hydrologic landscapes, *J. Am. Water Resour. As.*, 37, 335–349, doi:10.1111/j.1752-1688.2001.tb00973.x, 2001.
- Wood, E. F., Sivapalan, M., Beven, K., and Band, L.: Effects of spatial variability and scale with implications to hydrologic modeling, *J. Hydrol.*, 102, 29–47, doi:10.1016/0022-1694(88)90090-X, 1988.
- Woolhiser, D. A., Smith, R. E., and Goodrich, D. C.: KINEROS, A Kinematic Runoff and Erosion Model: Documentation and User Manual, USDA, Agricultural Research Service, ARS-77, 130 pp., 1990.
- 15 Xavier, A. C., King, C. W., and Scanlon, B. R.: Daily gridded meteorological variables in Brazil (1980–2013), *International Journal of Climatology*, 36, 2644–2659, doi:10.1002/joc.4518, 2016.
- Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A.: HESS
- 20 Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, *Hydrol. Earth Syst. Sc.*, 18, 4635–4655, doi:10.5194/hess-18-4635-2014, 2014.