Doctoral School in Environmental Engineering

Hydrological modelling with components:

the OMS3 NewAge-JGrass system

Giuseppe Formetta

Henry Rosseau - The dream, 1920

UNIVERSITY OF TRENTO - Italy
Department of Civil, Environmental
and Mechanical Engineering

February 2013

Doctoral thesis in Environmental Engineering, XIV cycle.
Faculty of Engineering, University of Trento
Academic year 2012/2013.

Supervisors:

- **Riccardo Rigon**,

  University of Trento, Trento (Italy)

- **Olaf David**,

  Colorado State University, Fort Collins, CO (USA)

External Advisor:

- **Roger Moussa**,

  INRA, UMR LISAH, Montpellier, France.

Day of the defense: 19/04/2013

Signature from head of PhD committee:

Abstract

This thesis presents the implementation and the applications of the new hydrological NewAge-JGrass system for forecasting and modelling of water resources in general at the basin scale. As a modern hydrological modelling, it is composed of two parts: (i) the system for data and results visualization based on the Geographic Information System uDig and (ii) the component based modelling system. The latter is built on top of the Object Modelling System v3. Modeling components can be selected, adopted, and connected according to the modeller needs and executed within the GIS uDig.

Different hydrological components were integrated to the system. The semi-distributed hydrological model NewAge represents an attempt to represent and model all the hydrological processes. The workflow start with the subbasins delineation by using the tools within the GIS uDig-JGrass. The system is based on a hillslope-link geometrical partition of the landscape. The basic unit, for the waterbudget evaluation is the hillslope. Each hillslope drains into a single associated link rather than cells or pixels. This conceptual partition was developed using an informatics with vectorial features for channels and raster data for hillslopes.

Different models were implemented to simulate different hydrological processes. Each model is a component, according to the definitions in OMS3 which can be substituted easily with others components without rewriting the whole model. NewAge requires interpolated meteorological variables (such as air temperature, precipitation, and relative humidity) as input data for each hillslope. They can be computed by a deterministic or geostatistic approaches. The energy model includes both, shortwave and longwave radiation calculation components for each hillslope. The first implements algorithms that take into account shade and complex topography and cloud cover.

Evapotraspiration can be modelled using two different solutions: the Fao-Evapotraspiration model and the Priestley-Taylor model. A snow melting and snow water equivalent model is also part of the system. Duffy's model and Hymod model are the runoff production models implemented in NewAge. In both cases the model is applied for each hillslope. Finally, the discharge generated at each hillslope is routed to each associated stream link.

All modelling components can be calibrated using one of the auto-calibration algorithms such as Particle Swarm Optimization algorithm and LUCA. All the components were verified by comparing modeled with measured data. Five river basins were used to verify the components. They are different in size, climate, and topographic complexity. Shortwave and longwave radiation, evapotranspiration discharge, snow melting and snow water equivalent were simulated.

NewAge is a modern component based hydrological system which allows for models integration, comparison, and substitutition, with one framework. Moreover, the system is able to manage components input output and visualization since it is linked to the GIS uDig-JGrass as a part of the same framework.

To my parents, Antonio and Silvana, and my sisters Achiropita and Grazia, for their unflagging belief in me.

Acknowledgements

First of all, I would like to thank Dr. Riccardo Rigon and Dr. Olaf David for their excellent supervision and support throughout these years. For their consistent guidance, invaluable and constructive advices, patience, and kindness.

My special thanks go to my jury members: Prof. Stuart Lane (Universit de Lausanne), Prof. Giorgio Budillon (Universit di Napoli Parthenope), Prof. Mathias W. Rotach (University of Innsbruck) and Dr. Eng. Michele Larcher (University of Trento) for their comments and suggestions.

I like to thank very much the Department of Civil and Enviromental Engineering in Trento University: all the professors, the administrative and technical staff.

A special thanks to the coauthors of some of the papers drawn from this dissertation: Stephanie K. Kampf and Jose Luis Chavez, Silvia Franceschi, Andrea Antonello and Ricardo Mantilla.

I also would like to thanks Luis A. Garcia, James C. Ascough, Timothy R. Green, and all the USDA-ARS employees for their help and precious suggestions. I want to thank the Colorado State University: all the professors and the administrative staff.

I want to express my gratitude to my friends: the old friends, the friends I met in Trento and the guys I met in the Colorado State University: I will never forget the good times we had togheter. A special thank goes to Alessandra.

I would like to acknowledge Olaf, Andrea, Antonia and Jhonas. When I was in Fort Collins they made me feel at home and part of the family: i will never forget the Easter and the Christmas we had all togheter.

Last but not the least, I would like to thank my family: my parents Antonio and Silvana, my two sisters Achiropita e Grazia. I dont imagine a life without their love.

Contents

List of Figures

List of Tables

1

Introduction

In this chapter, the context, the motivations and the objectives of the present dissertation are presented. The section (1.1) introduces briefly the background of the recent hydrological modelling efforts providing the starting point of this manuscript. The objectives and goals of the research activities are discussed in the section (1.2). Finally the section (1.3) provides the outline of the dissertation.

1.1 Background

The presented thesis is about hydrological modelling. The need of a not only reliable but also accurate hydrologic cycle modelling, of both energy and mass budget, is nowadays of great importance in many fields including engineering:

- Water Control and Risk management were getting more attention with the increasing number of natural disasters due not only to floods but also to drought. According the United Nations International Strategy for Disaster Reduction Secretariat (UNISDR) in one of its report (09-03-2009), floods and storms in Europe account for 40% and 33% of the total economic damages for the period 1989-2008. "Floods and storms explain part of the economic losses as weather related disasters have devastating effects on infrastructures which have on average, a higher value in Europe than in Asia or Africa" says Professor Debarati Guha-Sapir, director of CRED. "The trend will probably continue to rise as floods and storms are expected to become more frequent and severe in the future in Europe." In January 2009, storm "Klaus" affected southern Europe (France, Spain, and Italy), causing at least 25 deaths and over 750 million US dollar of estimated economic damages. Final costs and losses from the cold wave are still unavailable. The problem is worldwide, according to statistics from the United Nations [1], during 1970-2005 over 30% of natural disasters were floods and nearly 15% were droughts or drought-related natural disasters (wild fires and extreme high temperatures). During the 30-year period 1980-2010, floods accounted for more deaths in the United States than hurricanes, tornados, or lightning, ranking first among weather fatalities. Droughts are the main cause of

agricultural distress, accounting for over 11 billion of US dollars in damage in the United States during the first decade of this century;

- Water Use is meant to supply enough water not only for municipial use but also for irragation, agricultural use and energy supply by using hydroelectric-power;

- Water Quality Management in which is elevating the issues of pollutions and contaminants and their transport in groundwater lakes and rivers.

Determining the discharge of rivers during flood events has been a central topic since more than a century; firstly through the rational model of (108), later through the use of instantaneous unit hydrograph models ((138), (39)), and more recently by including the geomorphological approach (i.e. GIUH) ((133), (64), (136), (38)). Even runoff generation such as Topmodel ((11); (10), (53)) have mainly been used for this purpose.

Models developed to reproduce a whole set of hydrological quantities for operational purposes originated in water resource management and their development was driven agriculture needs. In this context large modeling systems were developed which of the precursors are the Stanford watershed model ((32)), the Sacramento model (e.g. (26)), and the PRMS model ((82)). They were based on the metaphor of intercommunicating compartments (reservoirs), each representing a process domain with its proper residence time. The recent "Distributed Model Intercomparison Project", DMIP, (123), revealed some of the many differences among the models reported above, and provided a first set of tentative comparisons. Despite the major emphasis of the project was the reproducing discharges, a more prominent focus to the prediction of the entire hydrograph, instead of the hydrograph peak was evident: a necessary element for the overall management of basins and particularly for the management of droughts.

To look at the topic from a different point of view, there exist an even larger variety of models, with varying degrees of complexity, and simplifications in the literature. The two extremes of modeling are offered by fully distributed models (for a recent review, see (74) and (126)), and lumped models (e.g (10)). In the first case, the physics is modeled at grid (pixels) level using the fundamental laws of conservation of energy, mass, and momentum, in the second case, the governing equations are simplified in order to obtain some statistics of the hydrological budget without the unnecessary representation of the entire spatial variability.

A simplification is offered by the theory of the geomorphological unit hydrograph which provides flow values at a single point of the river network (i.e. at the outlet of the basin). In this case, many models with few parameters are able to reproduce the expected result with an acceptable degree of confidence. This is possible because the outlet discharge is an additive stochastic process, e.g. (130), in which the topology and the geometry of the river network is more important than the details of the local dispersive dynamics e.g. (131). In addition (85) observed that the action of hydrological and geomorphological forces acts in maintaining approximately constant the flow velocity. This simplification is not appropriate when spatial

prediction is required, e.g. discharge at intranet location. To this end, it is necessary to make use of detailed information like topography (as derived from modern LIDAR or SAR sensors), and a large variety of remote sensed information, which provide new tools for the representation of the physics of flow transport along the channels of the river network and processes into the hillslopes.

The hydrological cycle is hard to simulate because of the many interactive processes and those that stimulated this dissertation.

Different hydrological processes apply to a different hydrological scale in time and space. Processes such as a thunderstorms occurs over the course of minutes to hours and spatial scales of a few kilometers or less to the formation of major river basins taking place over millions to tens of millions of years and spatial scales of 1000-10000 km.

One needs to simulate the hydrological cycle in order to model such as floods forecasting, landslide understanding, water managments water supply.

After choosing the temporal and spatial scale the target of a particular class of hydrological problem is defined. To solve a problem belonging to one of these class a given amount of prior information need to be known. The amount of prior information to solve the problem of simulating a the pick flood is lower than the amount of prior information to solve the problem of the whole hydrological cycle simulation. In the first case the physic of the problem allows to make some hypotesis that in a general case are not valid (such as constant velocity in the stream).

Because of this reason different type of hydrological model are able to solve different type of hydrological problems and the user, according the aim and the scale of his problem has to select the appropriate model.

Our research group in Trento University proposed different kinds of hydrological models. The fully distributed GEOtop solves energy and water balance at pixel scale, the Boussinesque Model solves at a regular mesh the Boussinesque equation. It is based on Darcys law for groundwater flow, and finally PeakFlow models the peak during a floods event.

The amount of prior information required to solve the problem decrases from GeoTop to PeakFlow.

In this dissertation the NewAge-JGrass is presented and below the motivation and goals are presented that justify this novel hydrological model.

Nowadays another important issue that is associated with hydrological modelling is the uncertainty estimate. Every hydrological problem and every kind of hydrological model has to deals with many sources of uncertainty:

- Model structural uncertainty due to a simplified model hypotesis or due to the discrepancies between the model and the real world processes;

- Uncertainty due to the model parameters estimation;

- Data uncertainty: although great improvements in experimental field measurements have been made during the last decades, data errors are source of uncertainty. This is obvious for some hydrological processes, such as the runoff production, where the model's forcing data (precipitation) is strictly related to the model output (discharge).

- Uncertainty of initial conditions that is due to the fact that the user is not able to correctly define the system states at the beginning of the model run.

Therefore, a novel infrastructure for hydrological modelling must deal with calibration issues and has to include a solid calibration algorithm if it wants to provide reliable results.

1.2 Motivation and Objectives

The NewAge-JGrass hydro-informatic infrastructure is presented in this dissertation. By contributing a novel model to the hydrological community a demand is addressed by going beyond the classical hydrological model.

NewAge is not an ordinary hydrological model but an attempt to provide an infrastructure where a user can build and use hydrological components and assemble them according to the application requirements.

An hydrological components simulates a particular hydrological process; the infrastructure allows the user to use an existing component or to implement a new one following a particular programming standard. The user is able to create his model according to his aim, input data, and the physics of the problem.

The second important aim is to provide a modern hydrological tool which allows the user to:

- store and manage all the data to run each component using a database in which the results of the simulations can also be stored;

- set his custom model (set of different components) for different sized basin (starting from the small catchemnts moving to bigger basins);

- set up his custom model (set of different components) for different temporal scale (moving from minutely simulation to monthly);

- take into account of the uncertainty within model parameters estimates by providing different calibration methods algorithm;

This dissertation contributes to the hydrological science a dynamic modelling tool which allows the scientist to directly compare different hydrological components. The NewAge-JGrass hydro-informatic infrastructure was developed to satisfy all the above requirements.

1.3 Outline of the dissertation

This section provides an outline of the whole dissertation.

Chapter (1) contains the introduction and the motivation of the work. Chapter (2) introduces the programming approach used (Object Modelling System version 3.0), the motivation for this selection and the informatic infrastructure and tools on which the hydrologicals are based on. Chapter (3) describes all five river basins which were used to test a single model component or a combination of them.

From Chapter (4) to (9) the NewAge-JGrass model is presented within all of its component and the applications are presented and commented.

Chapter (4) presents the model set up for a generic basin and the sub-basin delineation by using the Horton Machine. It is a GIS JGrass package for geomorphological analysis. The model setup is presented for the Little Washita river basin.

Chapter (5) describes the meteorological interpolation algorithms of the system: the implementation and applications of the krigings tools and the application of the JAMI algorithm.

Chapter (6) presents the energy balance tools. The shortwave radiation balance component and its verification using three different river basin. The description of the long wave radiation component and of two different evapotraspiration methods is also part of this chapter.

Chapter (7) explains the runoff production and routing components. Application for Little Washita and Fort Cobb river basins are presented. Discharge is simulated for both basins and calibration algorithms are used for the model parameters estimation. A comparison between models with and without routing and configurations based on different basin delineations are performed. Finally, a comparison with the SWAT model is presented.

The snow melting and snow water equivalent component is presented and verified in Chapter (8). Finally, Chapter (9) contains a review of the calibration algorithms of the system: Particle Swarm Optimization and LUCA.

2

Project Objective

This thesis describes the structure of NewAge-JGrass: a system for hydrological forecasting and modelling of water resources at the basin scale. It has been designed and implemented to emphasise the comparison of modelling solutions and straightforwardly reproduce hydrological modelling efforts. It is composed of two parts: (i) the system for data and results visualization based on the Geographic Information System uDig and (ii) the modelling system, based on the Object Modelling System v3. The latter supports modeling components that can be assembled according to the modeller needs and executed in the uDig Spatial Toolbox. Therefore, the system provides an ideal and modern integration of models and GIS without invalidating existing solutions. Compared to legacy hydrological models, which are built upon monolithic code, NewAge-JGrass allows for multiple modelling solutions for the same process provided and they share the same inputs and outputs. Components are connected by means of a scripting language.

2.1 Overall objective

Many scientists claim that a models source code should always be available [e.g. Ince et al., 2012] since it has become an integral part of the advancement of science. However, with the traditional approach to modelling, external inspection, analysis, improvement, and contribution is difficult. Even if the source code would be available, the growing complexity of the source code makes progress in the model development difficult. Moreover, the implementation as traditional monolithic source code (e.g. a definition in (132)) of the many environmental processes which are intimately interlinked (as snow modelling, runoff production, evapotranspiration in the hydrology case), becomes difficult to understand, to disentangle, and to verify (121). Traditional software methods as applied in hydrology precludes rapid reuse and improvement of the source code, thus not favouring code readability. It is an obstacle to steady advancement in science. Therefore researchers that rely on computational methods and techniques as part of their day-to-day activities need Reproducible-Research-Systems (RRSs), making it easier to document any step during research from data preparation to output analyses, and to improve collaborative work and third-party verification.

Many of the software infrastructures or frameworks (MF) for modelling were actually designed and built to carry out a sound scientific process (e.g. (156), (3), (132)). Among those that specifically targeting the support of hydrological modelling are the Spatial Modelling Environment (SME, (100)), The Invisible Modelling Environment (TIME) and hydrological derivative tools like, E2 (4), OpenMI (105), and the Object Modelling System (OMS, (36)). However, most of the above MF require a quite significant learning curve that not all scientists, even proficient modellers are willing to make.

Therefore, making the transition to modern programming environments easier, some projects recently tried to reduce the invasiveness of frameworks (88) into the model. Especially the third version of OMS and the BIOMA project reveals promising perspectives. The study presented in (55), among others, emphasize that in order to optimise scientific productivity, a RRS infrastructure should include not only the computational code but also visualisation and data-processing tools necessary to synthesise knowledge from high volumes of inputs and outputs. Indeed, preferable tools for the visualisation of hydrological processes have been for a long time Geographic Information Systems (GIS), (89) and (61). However, traditional GIS are usually designed for managing static information layers. They are not designed to interact with dynamic models (e.g. (27) and(156)). The interaction between models and GIS is often off-line or performed with integration strategies that affected either the functionality of GIS tools or the usability of models.

The MF listed above offer instead the proper abstraction to streamline the interaction with a GIS. They promote the separation of the model into well defined module or component elements, each module with a specified method to interact with others through specified interfaces. Through their interfaces the modules can communicate and exchange data at run-time. Therefore it is also timely for a GIS and hydrological model components to constitute a pool of interoperable tools that can be mixed together for creating a system that is accurately tailored to geosciences.

This thesis describes an open source deployment of such a RRS, based on the GIS uDig, the modelling framework OMS3, and the GIS toolkit GeoTools. The key ideas for the framework, libraries, models, and components are: being open source, promoting the modern object oriented informatics introduced above, being portable to all the main platforms (being platform neutral), supporting at least source code written in C/C++ and FORTRAN, and being based on solid communities of developers.

Open source allows researchers to freely access the code and users to extend, modify, and redistribute the system at no charge and limits the choice among MF candidates for a RRS to a few. Portability excludes some solutions, as those based on the .NET platform. NET is practically tight to the Windows platform even if the Mono environment could have been targeted. Therefore, the above considerations led to the following development choices:

- The use of Java as the system platform: essentially for its portability, the availability of a few open source frameworks that allow adaptation for our task. Furthermore, Java is a

modern and mature language that has features such as multithreading that are essential for building a scalable modelling platform.

- The use of a modeling framework OMS or OpenMI/Java to facilite model construction. Both frameworks were tested and used to implement a first version of NewAGE-JGrass. OpenMI was developed within the EU project HarmonIT ((14)) by influential hydrological institutes in Europe, allowing for chained integration of existing hydrological models. OMS v3 revealed less invasive code in practice, thus producing leaner and more descriptive modelling code. In our experience up to thirty percent smaller. Therefore it became the final choice of the system.

- The use of uDig as visualization/GIS platform, including GIS services. uDig stands for User friendly Desktop Internet GIS. Its integration with the JGrass GIS offers the Spatial Toolbox which contains the features previously offered by JGrass. Using uDig as basis also implied the use of the Eclipse as RCP for part of the project. Moreover, Eclipse was also chosen as the IDE for compiling and developing models, which fulfills the further requirement of using a completely open source tool chain for development, which may promote its adoption among models developers.

Other languages and platforms such as C/C++, Phyton, and FORTRAN, would not offer such complete tool chains and middleware compared to Java and its JVM.

R, as an alternative, was also evaluated but discarded since it is not offering enough flexibility and the efficiency for computational intensive model even if it is powerful for post processing model results, graphics, and automatic calibration tools. However, interoperability with R as external tool was already added to OMS as an experimental feature.

2.2 The NewAge-JGrass and JGrasstools goals

The NewAge-JGrass system represents the implementation of the concepts as described above and is an effort to make a RRS available to hydrologists. The name JGrass reminds of the project history which started with the implemention of the GIS JGrass.

To achieve the needs of a RRS, NewAge-JGrass is therefore built upon two main parts: (i) the GIS uDig which manages geospatial data and models, visualizes inputs and results, and (ii) the OMS uDig Spatial Toolbox, referred to in uDig as the Spatial Toolbox.

The main structure of the NewAge-JGrass model is based on the Eclipse Rich Client RCP interface, services provided by uDig, the extensive use of GeoTools, and a modelling strategy based on OMS3. This structure can easily be connected to a geographic database that provides the appropriate input data, and can handle the storage of the geospatial components' outputs. This requirement, not strictly necessary for the building of a RRS, allows however the use of the entire infrastructure, for the deployment of operational system as this shown in fig.(2.1).

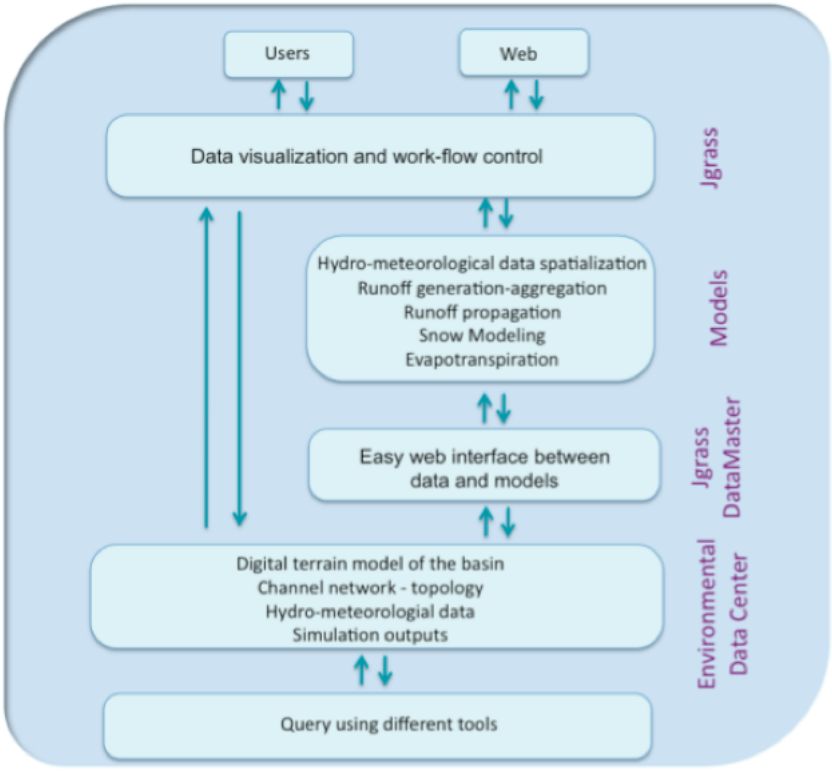

Figure 2.1: Structure of NewAge, as deployed for River Adige basin Authority.

The GIS interface is currently built using uDig 1.3.1 as shown in fig.(2.2). Although if the uDig project is not a key point of this thesis it is appropriate to point out some of its main characteristics. uDig runs as a rich client under different platforms (Windows, Mac OS/X, and Linux) and it is web-service oriented. Hence, it is capable to consume geodata served through standards such as Web Feature Services (WFS), Web Map Services (WMS), and Web Coverage Services (WCS),GeoRSS, and KML. Its interface is built upon the Eclipse rich client platform, guaranteeing a native look-and-feel in any of the operating system. It has a very flexible plugin-in mechanism to add features, to customise the user interface, but most importantly is supported by a very solid industrial foundation. uDig contains not only elements for visualization of maps, but also tools for data manipulation, editing, map printing, and connection with remote databases and servers (Oracle, SDE, Postgres/PostGIS, Teradata, and others). The core GIS functionalities, like data reading, coordinate reprojection, rendering, etc., are provided to uDig by the GeoTools library which allows a convenient management of vector data. It also leverages the ImageIO-ext project providing access to raster formats supported by GDAL. Since version 1.2, uDig is compatiple with the GRASS GIS (110). It can natively manage the GRASS raster data format natively. The Spatial Toolbox (see below) allows the execution of most of GRASS commands. The system is based on Java and, it integrates seamlessly with the other tools that forms the NewAge-JGrass system.

The uDig Spatial Toolbox window is depicted in figure 2.3. Its interface allows the individual

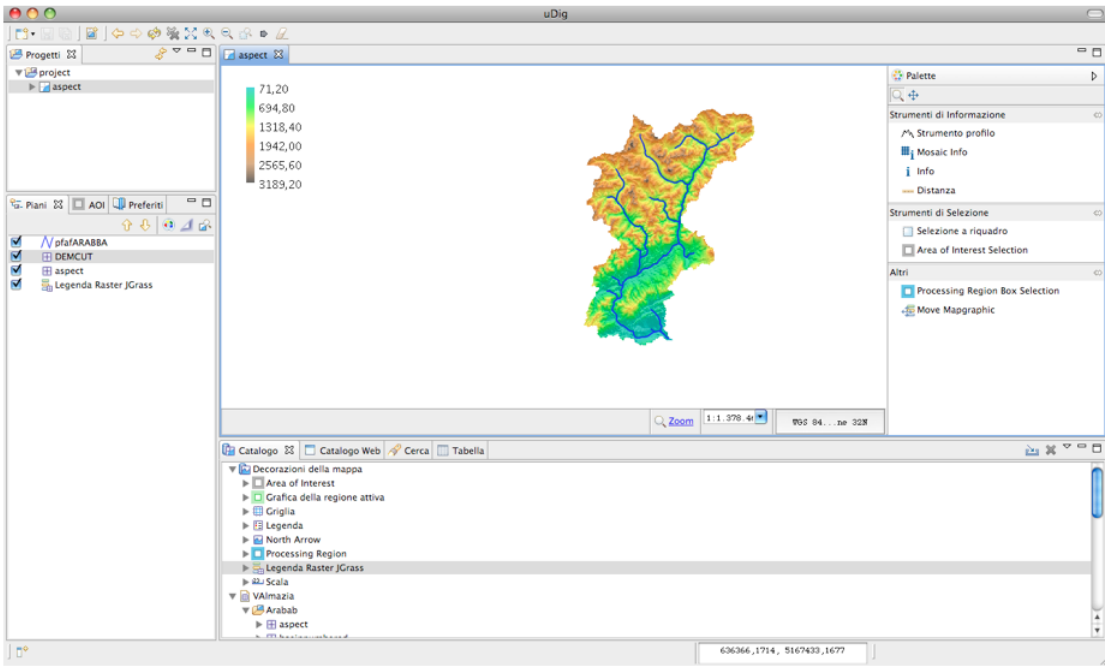

Figure 2.2: The GIS interface of NewAge implemented in uDig. uDig is based on the Eclipse Rich Client Platform.

execution of models and components of NewAge-JGrass. On the left side of the window, folders type icons represent available model components. The uDig Spatial Toolbox contains all needed tools for the management of raster maps (import, export, and manipulation) and provides the geomorphological analysis packages necessary to prepare the input data for NewAge-JGrass, derived from The Horton Machine (125). Tabs on the right side of the window allow access to data input, data output and the description of each field of the selected component. The content as presented in the Spatial Toolbox interface, like the documentation of any spatial modelling component, is automatically generated by obtaining and parsing the metadata attached using annotations provided by the underlying OMS3.

2.3 Object Modelling System v.3 (OMS3)

The Java based, object-oriented modelling framework OMS3 treats models and components as plain objects with meta data provided by means of annotations (36). Creating a modelling object is very easy, there are no interfaces to implement, no classes to extend and polymorphic methods to overwrite; no framework-specific data types need to replace common native language data types. There is only the use of annotations to specify and describe "points of interest" for existing data fields and methods for the framework. The models in OMS3 are components assemblies. Each component is a self-contained unit implemented with a standard, well-defined purpose and interface in mind. Finally, simulations (model applications with data) can be executed individually from the graphical interface or they can be linked together in the uDig

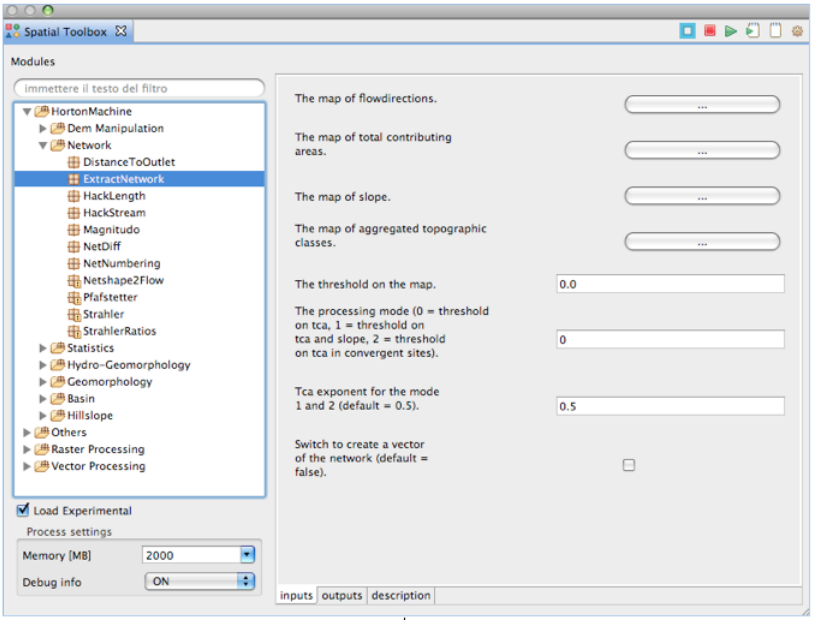

Figure 2.3: The uDig 1.3.1 Spatial Toolbox Interface.

console. They can even run outside the GIS using the OMSConsole. The first use case requires the uDig Spatial Toolbox installation fig.(2.3), the model selection from the modules interface and filled out input field forms. The second option requires the OMS3 scripting knowledge, which allows the user to select and run the models.

3

Watersheds and Datasets Description

In this chapter all the river basins used in this work are described. The selection of the river basins was based on three main reasons: i) a dataset must be free of charge and at least more than one year long; ii) the differences in climatology and topography-geomorphology between basins; iii) different time step of the measurement data (daily and hourly).

3.1 Little Washita River Basin

The Little Washita river basin (611 Km^2), fig.(3.1), is located in southwestern Oklahoma, between Chickasha and Lawton. It was selected because of its complete datasets of meteorological forcings, as provided by the USDA Agricultural Research Service (USDA-ARS), and because of absent relevant snowfall and soil freezing, which is not modeled by the assembly of components used in the specific case.

The climate of the basin can be characterized as moist and sub-humid with a long-term, spatially average, annual precipitation of 760 millimeters and a temperature of 16 degrees Celsius. Winters are typically short and dry but they are usually very cold for a few weeks. Summers are typically long, hot, and relatively dry. The elevation of the basin ranges between about 300 meters and about 500 meters a.s.l. The bedrock exposed in the watershed consists of Permian age sedimentary rocks and soil textures range from fine sand to silty loam.

The meteorological stations used in this study are shown as black dots in fig.(3.1). The hydrometer where the calibration is performed is depicted with a black triangle.

Tab.(3.1) reports the main information (coordinates, and elevations) of the twenty meteorological stations (whose data are available at ARS-MESONET). Five minute measurements of rainfall (P), air temperature (T), and incoming solar radiation (R) were aggregated to hourly time steps and used as input of the modeling system. The stations in bold in tab.(3.1) are removed from the complete dataset and used as verification stations. Measured and modeled time series for precipitation, incoming solar radiation, air temperature are compared at that

stations. This procedure is used in order to verify the goodness of the components presented in the next chapters. The same notation (stations in bold are removed from the complete dataset) is valid for all the river basin presented in this section.

The hydrometer measures discharge at 15 minute resolution http://waterdata.usgs.gov/nwis. The values were aggregated to hourly time steps and used in the automatic calibration procedure.

Table 3.1: List of the meteorological stations used in the simulations performed on Little Washita river basin. ID is the station identificative number, City is the closer city to the station, LAT and LONG stand for latitude and longitude respectively, Elevation and Aspect are the station elevation and aspect respectively. Bold font is used for indicating the stations belonging to the validation set..

ID	City	LAT.	LONG.	Elevation (m)	Aspect (°)
124	**Norge**	**34.9728**	**-98.0581**	**387.0**	**138°**
131	Cyril	34.9503	-98.2336	458.0	245°
133	Cement	34.9492	-98.1281	430.0	116°
134	Cement	34.9367	-98.0753	384.0	65°
135	Cement	34.9272	-98.0197	366.0	182°
136	Ninnekah	34.9278	-97.9656	343.0	270°
144	Agawam	34.8789	-97.9172	388.0	50°
146	Agawam	34.8853	-98.0231	358.0	212°
148	**Cement**	**34.8992**	**-98.1281**	**431.0**	**160°**
149	Cyril	34.8983	-98.1808	420.0	205°
150	**Cyril**	**34.9061**	**-98.2511**	**431.0**	**195°**
153	Cyril	34.8553	-98.2121	414.0	165°
154	Cyril	34.8553	-98.1369	393.0	175°
156	Agawam	34.8431	-97.9583	397.0	290°
159	Rush Springs	34.7967	-97.9933	439.0	235°
162	Sterling	34.8075	-98.1414	405.0	15°
182	**Cement**	**34.845**	**-98.0731**	**370.0**	**245°**

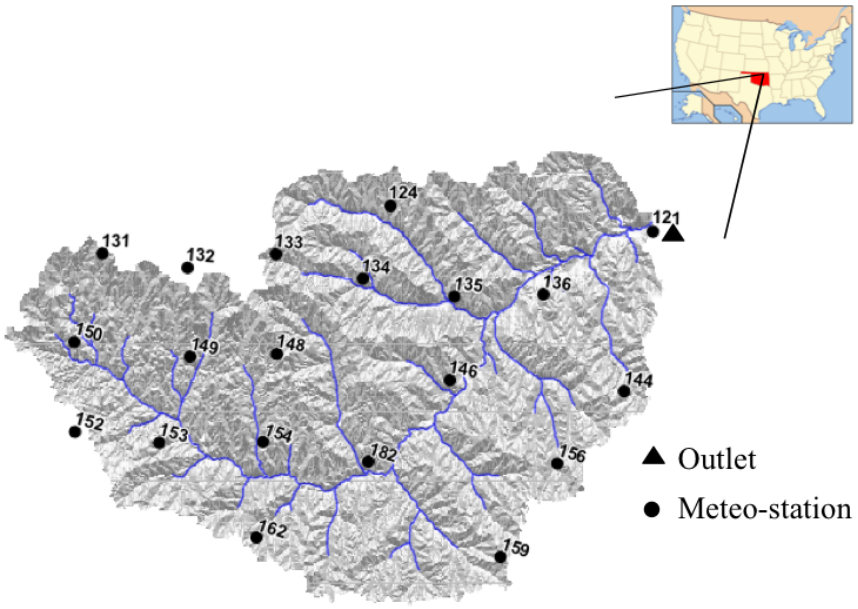

Figure 3.1: The Little Washita river basin, Oklahoma (U.S.A.).

3.2 Fort Cobb river basin

The Fort Cobb Watershed, fig.(3.2), is located in the Central Great Plains Eco-region in south-western Oklahoma in Caddo. It is 813 square kilometres in size. Its elevation ranges between 383 meters and 565 meters a.s.l.. Within the watershed there is the Fort Cobb reservoir, a lake for water supply and recreational use created by the Bureau of Reclamation in 1959 by impounding Cobb Creek three miles north of the town of Fort Cobb. Land use in the watershed includes agricultural fields, cattle operations, rural communities, and one hog feeding operation. Most soils in the watershed are highly erodible, sandy clays and loams underlain primarily by Permian sandstone, siltstone, and claystone.

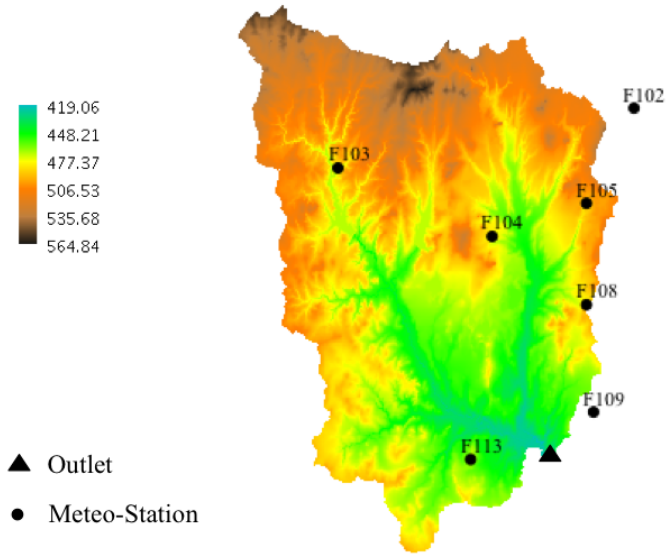

Figure 3.2: The Fort Cobb river basin, Oklahoma (U.S.A.).

The climate of the basin can be characterised as moist with a spatially average, annual precipitation of 816 millimetres and a temperature of 16 degrees Celsius. The NewAGE-JGrass modelling solution is applied for the Fort Cobb river basin at the Eakly outlet, before the river enters the reservoir. The DEM of the basin is available for download at http://seamless.usgs.gov/ with a resolution of 1/3 arc-second. Five minutes meteorological measurements of rainfall, air temperature, and incoming solar radiation are available at http://ars.mesonet.org/ for the watershed. The data was aggregated to hourly time steps and used as input for the modelling system. Seven meteorological stations were used in this study. Tab.(3.3) lists their main features.

Table 3.2: List of the meteorological stations used in the simulations performed on Fort Cobb river basin. ID is the station identificative number, City is the closer city to the station. Bold font is used for indicating the stations belonging to the validation set..

ID	City	LAT.	LONG.	Elevation (m)	Aspect (°)
101	**Hydro**	**35.4551**	**-98.6064**	**504.0**	**120°**
104	Colony	35.3923	-98.6233	484.0	35°
105	**Colony**	**35.4072**	**-98.571**	**493.0**	**300°**
106	Eakly	35.3915	-98.5138	472.0	295°
108	Eakly	35.3611	-98.5712	492.0	40°
109	**Eakly**	**35.3123**	**-98.5675**	**466.0**	**90°**
110	Eakly	35.3303	-98.5202	430.0	115°
113	Colony	35.291	-98.6357	465.0	155°

3.3 Piave river basin

The Piave river basin area (3460 km^2), fig.(3.3), is located in the North-East part of the Italian peninsula. The elevation range is between 700 and 3160 m a. s.l., the main soil uses are: i) crops up to 500 m a.s.l, ii) evergreen and deciduous forests at elevation between 500 and 1800 m a.s.l and iii) alpine pasture and rocks at higher elevations. The mean annual precipitation is about 1500 mm and the mean air temperature is 10 degrees Celsius.

ID	City	LAT.	LONG.	Elevation (m)	Aspect (°)
1	Arabba	46.4999	11.8761	1825	180°
2	**Caprile**	**46.4404**	**11.9900**	**1025**	**170°**
3	Agordo	46.2780	12.0331	602	5°
8	Villanova	46.4433	12.2062	972	71°
9	**Auronzo**	**46.5562**	**12.4258**	**940**	**223°**
11	Campo di Zoldo	46.3466	12.1841	915	160°
12	Domegge di Cadore	46.4609	12.4103	802	148°
14	Monte Avena	46.0321	11.8271	761	55°
18	Passo Pordoi	46.4834	11.8224	357	55°
21	Passo Monte Croce	46.6521	12.4239	1612	120°
22	Col Indes	46.1191	12.4401	1119	210°
23	**Torch**	**46.1515**	**12.3629**	**602**	**177°**
26	Sappada	46.5706	12.7080	1275	156°
29	Feltre	46.0162	11.8946	273	190°
31	Falcade	46.3554	11.8694	1151	50°
32	Cortina	46.536	12.1273	1244	88°
35	Belluno	46.1643	12.2450	378	157°

Seventeen meteorological stations are used for the simulations. Five minutes measurements are available for each stations: air temperature at a heigth of 1.5 m, relative humidity at a heigth of 1.5 m and incoming global solar radiation. The data for the year 2010 were aggregated to an hourly time step and were used in the simulations. The meteo stations main features are reported in tab.(6.4) and fig.(6.5) which shows their positions.

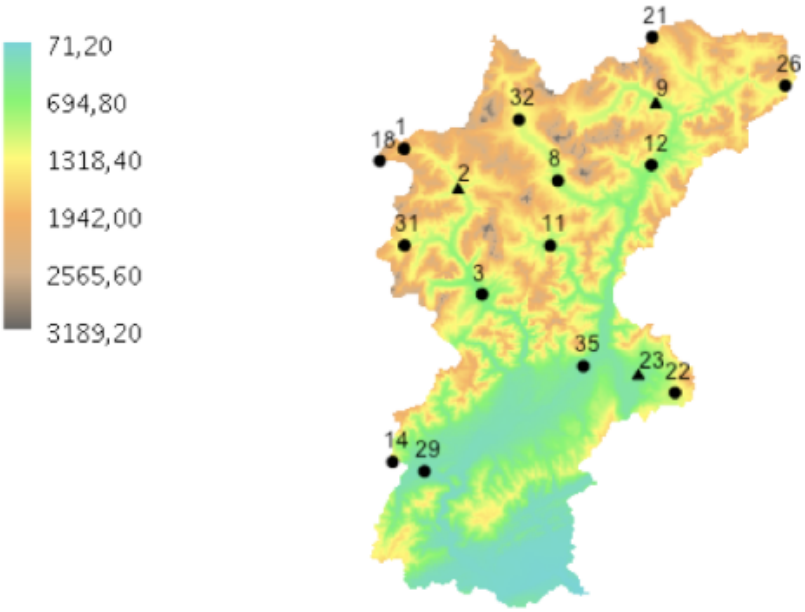

Figure 3.3: River Piave area, (Italy).

3.4 Teton river basin

The Teton river basin is one of three watersheds that comprise the Henry's Fork Basin. The Teton River drains an area of around 2.900 km^2: 2000 km^2 in Idaho and 900 km^2 in Wyoming. Accurate river basin descriptions are presented in (114) and in (122). The Teton Subbasin is physically and biologically diverse. Climate varies within the subbasin according to elevation. Precipitation varies from 330 mm per year at Sugar to more then 100 mm per year in the mountains. Most of the annual precipitation occurs in the form of snow. The runoff is snowmelt dominated with peak flows typically occurring in late May or early June. Railfall interpolation and comparison between different interpolation algorithms where performed in this thesis using Teton river basin rainfall dataset.

Table 3.3: List of the meteorological stations used in the simulations performed on Teton river basin. ID is the station identificative number, City is the closer city to the station, LAT and LONG stand for latitude and longitude respectively, Elevation is the station elevation. Bold font is used for indicating the stations belonging to the validation set..

ID	City	Elevation (m)	LONG.	LAT.
1	Ashton	1589.0	-111.2739	44.04250
2	Palisades	1641.0	-111.2167	43.35000
3	Idaho Fall	1776.0	-111.7847	43.34556
4	**Swan valley**	**1634.0**	**-111.2939**	**43.44472**
5	Driggs	1865.0	-111.1125	43.73056
6	Rexburg Idaho	1526.0	-111.7892	43.80833
7	**Tetonia exper. station**	**1881.0**	**-111.2769**	**43.85639**
8	St. Anthony	1509.0	-111.7128	43.96944
9	Moose	1972.0	-110.7164	43.65528
10	Snake river 1	2098.0	-110.6658	44.13333
11	Pine Creek	2048.0	-111.2116	43.56998
12	Sheep Mtn	2003.0	-111.6878	43.20933
13	Sedgewick Peack	2393.0	-111.9561	42.52470
14	Somsen ranch	2073.0	-111.3589	42.95250
15	**Grassy lake**	**2214.0**	**-110.8344**	**44.12612**
16	Phillips bench	2499.0	-110.9110	43.51948
17	Snake River 2	2109.0	-110.6692	44.13361
18	Lewis lake	2393.0	-110.6664	44.20860

Daily rainfall data for three years (from 1980-10-01 to 1983-10-01) were available in 18 locations, fig.(3.4). Tab.(3.3) contains the meteorological stations' main features. It presents in bold the station excluded from the dataset and used as verification stations where comparisons between interpolated and measured rainfall is performed.

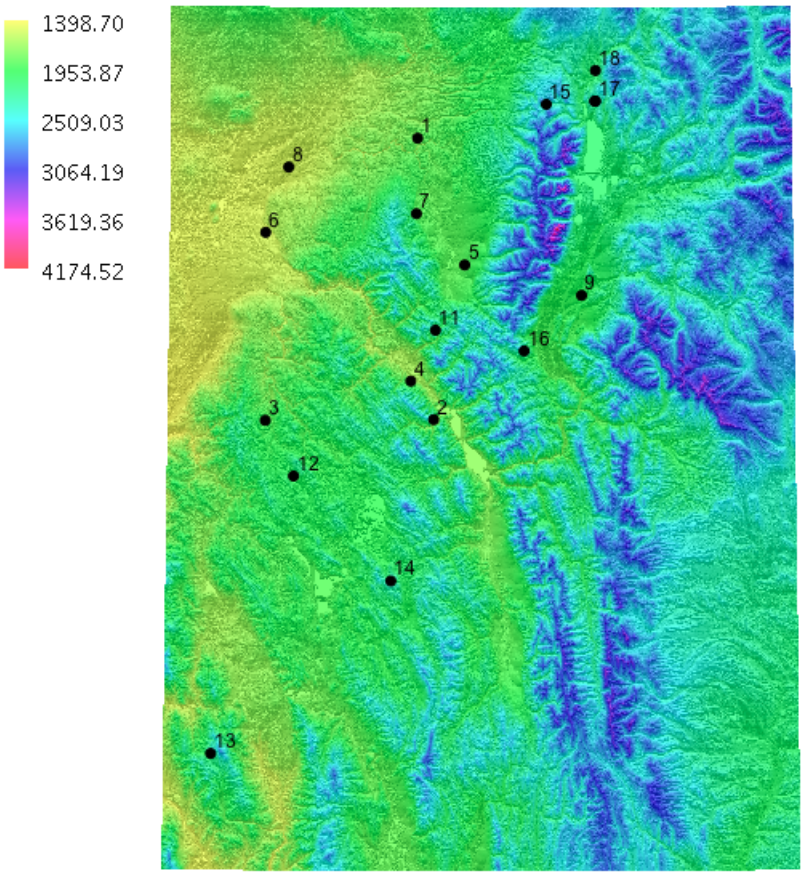

Figure 3.4: Teton area DEM and measurements stations.

3.5 Cache la Poudre river basin

The NewAge snow model was tested on the upper Cache la Poudre basin in the Rocky Mountains of northern Colorado and southern Wyoming, USA. This 2700 km^2 basin has elevations ranging from 1590-4125 m, with mean annual precipitation ranging from 330 mm at lower elevations to 1350 mm at the highest elevations. Three snow telemetry (SNOTEL) stations operated by the US Natural Resource Conservation Service monitor air temperature, precipitation, and snow water equivalent in this basin. At these stations, snow accumulation typically begins in October or November, with peak snow water equivalent in May.

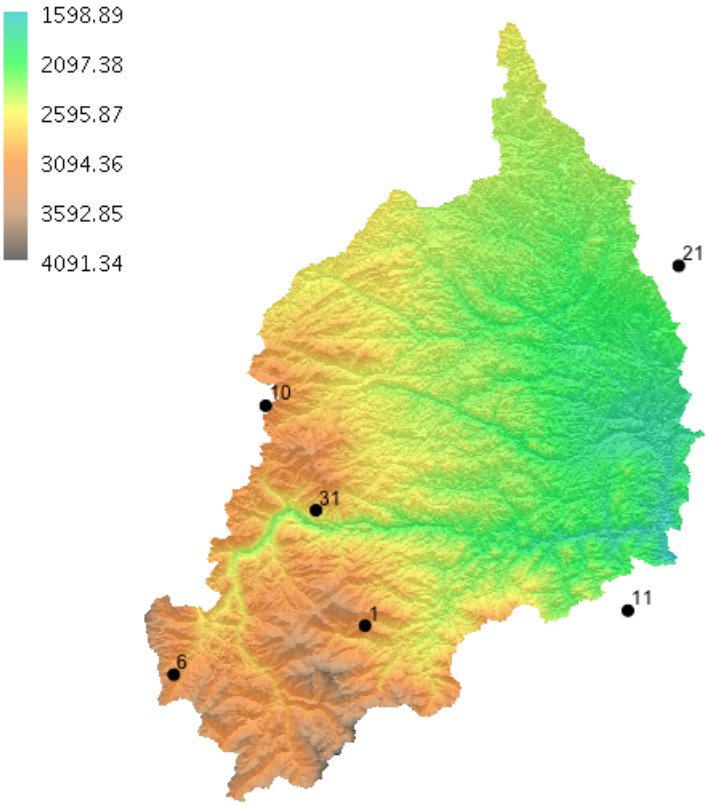

Figure 3.5: Poudre river basin digital elevation model.

The model is applied in the Cache la Poudre River basin whose data are available on http://www.wcc.nrcs.usda.gov/snotel/Colorado/colorado.html. Six are the meteorological stations available on the river basin.

They are presented in fig.(3.5) and tab.(3.4) shows their main features.

Hourglass, Deadman Hill and Joe Wright belong to the Natural Resource Conservation Survey Snow Telemetry (SNOOTEL) meteorological stations. They provide data (precipitation, air temperature and snow water equivalent) at a daily time step. For Hourglass station the data available start on 01-10-2008 and ends on 01-05-2012 (the first year is used as calibration period and the last 3 years are used as validation period); for Joe Wright and Deadman Hill stations

Table 3.4: List of the meteorological stations used in the simulations performed in the Cache la Poudre river basin.

ID	City	LAT.	LONG.	Elevation (m)
1	Hourglass	40.25	105.38	2814.0
6	Joe Wright	40.32	105.53	3085.0
10	Deadman Hill	40.40	105.46	3115.0
11	Buckhorn Mountain	40.60	-105.28	2256.0
21	Virginia Dale	40.95	-105.21	2138.0
31	Rustic	40.70	-105.70	2347.0

they go from 01-10-1999 to 01-10-2009 (the first year is used as calibration period and the last 9 years are used as validation period).

Buckhorn Mountain, Rustic and Virginia Dale belong to Service Cooperative Observer Program (COOP) meteorological stations. They only provide precipitation and air temperature. For the three stations the data available start on 01-10-2008 to 01-10-2009. Those data integrated the SNOOTEL stations measurements. They were used for air temperature and precipitation interpolations in the fully distributed application of the snow melting and snow water equivalent component.

4

The NewAge-JGrass infrastructure.

This chapter provides the rational of the geometrical and topological structure of the watershed model used by NewAge-JGrass. The layout of the catchment derived from a suitable modification of Pfafstetter scheme is presented. Furthermore, all the steps for building the geographical objects necessary to run the NewAge-JGrass model are explained. In sections 2 and 3 a geomorphological analysis is performed with the appropriate tools of the Horton Machine. Issues arisen from the application of the geographical components are discussed throughout the chapter.

4.1 Introduction

Before modelling a basin it is necessary to conceptualise its description in formal terms. This has been described in either explicitly or implicitly in several ways. These formal descriptions are usually named Digital Watershed Model (DWM). A DWM is an electronic representation of a watershed's spatial characteristics and time-series hydrologic information. "It can include elevation, water features, land use, point observation data, and/or gridded data, e.g. remote sensing, climate products, where the data are related and usable for investigations in a participatory environment to promote a broad collaboration among many types of scientists and engineers" (from: http://www.cuahsi.org).

One of the most mature digital watershed schematisations is the one initially encoded by (90), and subsquently endorsed by CUASHI, http://www.cuahsi.org, and described for instance in (58). It distinguishes the basic units that compose a watershed and identifies an appropriate data model and storage formats for them. For instance, Arc Hydro, (90), distinguishes the following as part of the basins:

- the stream network subdivided in links
- the basins subdivided in sub basins
- the lakes and superficial water bodies
- the monitoring points

Recently, the same authors added to Arch Hydro also specifications for the groundwater resources specifications, (144), which are not covered here. These units are usually stored as simple features according to the OGC standard that can be processed by all the major GIS systems. These formats have a correspondent storage format within all major data base systems which cover geographic features. CUASHI has built its Hydrologic Information System (HIS) around these concepts. The effective use of this information in models requires usually a further refinement which is very much modelling dependent, and is usually based on the definition of the hydrologic response units (HRU) (135), (48),(80) and (152). They are the elementary part of the basins that are treated as black box units. (107) represents an example of a detailed partition of a basin for agricultural use. Within AGE (6) sub-catchments are subdivided in many more detailed functional parts which refer to different treatment of soil use and/or land-cover which is thought to influence the hydrological fluxes. In NewAge-JGrass, the basin is partitioned into hillslopes and channels (presented to the model a hillslope-link, HL, structure) where the hillslopes are the basic hydrologic units at least for rainfall-runoff. At this scale energy and water mass budgets statistics are estimated after appropriate averaging. Channels are described as vector features that are topologically interconnected in a simple directed graph. For computational reasons, the partitioning of the area is not usually designed to identify all the physical hillslopes present in the system, but to define the dimensions of small watersheds. In the current applications, of 2-10 Km^2 on average. HRUs can either be represented as vector features or rasters.

Within a model any element of the river network can include anthropogenic structures that regulate the flow regimes, thus make it possible to simulate intakes, management of dams, artificial channels, and water abstractions for example irrigation.

However, hillslopes can be further dissected depending on the processes to be analyse. For instance, when temperature is the concern, each hillslope can be further subdivided in altimetric bands, each one with its own temperature, that is eventually averaged to obtain a single value for the whole hillslope. Regarding the estimation of radiation, or snow, the specific model component of NewAge-JGrass can use information at pixel scale, which is subsequently averaged according to the needs, or information in selected representative points within the hillslope. Therefore, besides, a generic delineation of the basin, each module component can process the data and the geometries it requires.

One critical issue is how single units exchange the main hydrological fluxes. MYTHAS and AGEs HRUs, for instance, exchange runoff and subsurface fluxes in multiple direction and therefore have procedures to manage this complexity. NewAge-JGrass at the present stage just allows one hillslope to discharge into its channel link. In any case the river network constitute a hierarchy in which sources flow into the internal links and these, into larger streams.

To account for this hierarchical simulation various strategies can be used. The most modern scheme is probably (87), which, builds on the knowledge obtained by analysing the (144) generalised Pfafstetter's scheme that is used here, and is described below.

Whatever the conceptualisation, the challenge, is to deploy the ideas in robust and correct code. This is accomplished in NewAge-JGrass by using the GEOtools libraries and their implementation of the geographic features which seamlessly integrate with OMS3 programming and uDig.

The Horton Machine (127) and (128) is built on top of these libraries which are the modelling components that are actually being used.

To obtain this hierarchical structure it is necessary to first process the raster data from a digital elevation model which is summarised below.

4.2 Catchment analysis

The analysis of the catchment, starts with the acquisition of a Digital Terrain Model (DTM) of the catchment, e.g. (159). It is performed as illustrated in fig.(4.1) and summarized for the reader below.

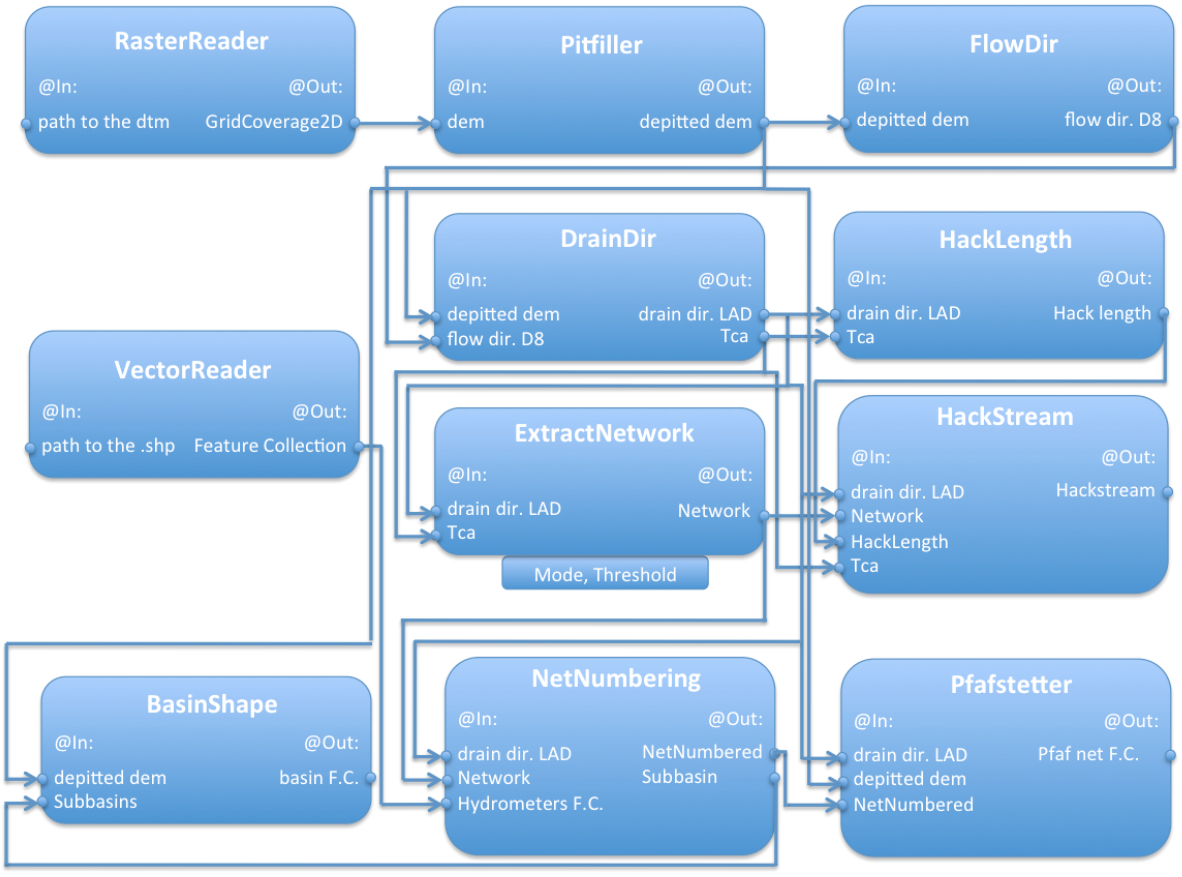

Figure 4.1: The workflow for the basin delineation in NewAge-JGrass -

4.2.1 Geomorphological analysis

Starting from the digital terrain model (DTM), the "Horton Machines" (128) components as provided by the GIS uDig-JGrass are used. In sequence, those are:

- Pitfiller: this tools fills the depression points in the DTM and defines the drainage directions at each point; the input for Pitfiller is the DTM previously imported in the GIS. An example of the output is shown in fig.(4.2,a)

- FlowDirections: it calculates the drainage directions with the method of the maximal steepest descent slope, selecting one out of 8 possible directions (D8); the input of the command is the map output of the Pitfiller component.

- DrainDir: this tools provides the drainage directions minimising the deviation from the real flow. The deviation, calculated using a triangular construction, could be given in degrees (D8 LAD method) or as transversal distance (D8 LTD method), (113). The input raster maps are: the map in output of Pitfiller and Flowdir. The outputs are the raster maps of the drainage directions, fig.(4.2,c), and of the total contributing area, fig.(4.3,b).

- Slope: it estimates the slope at every pixel by employing the drainage directions. The input raster maps of Slope are the map in output of Pitfiller and Flowdir. The output is shown in fig.(4.2,d)

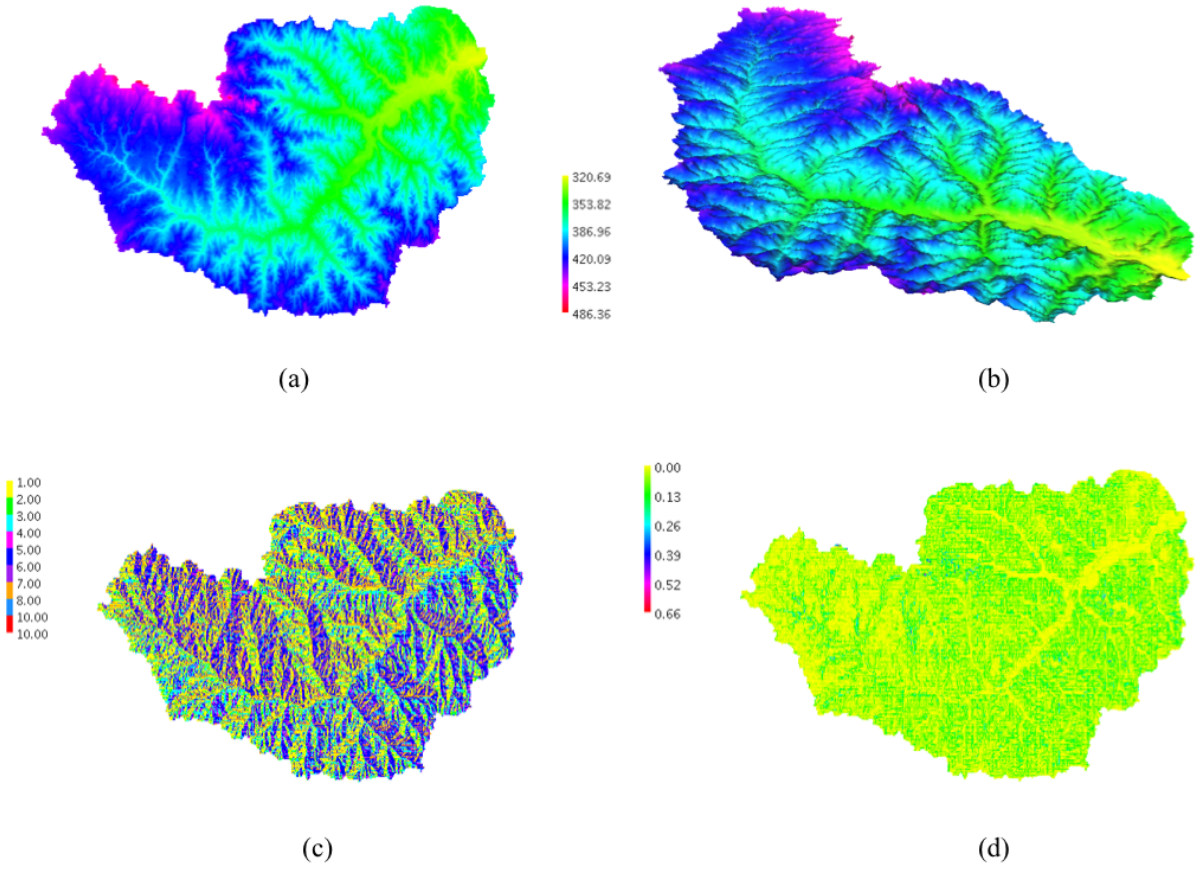

Figure 4.2: The Little Washita basin: output of Pitfiller, Draindir and Slope - The figure shows the output of Pitfiller Horton Machine in a 2D view (a) and 3D view (b) and the outputs of Draindir (c) and Slope (d).

- Aspect: it estimates the aspect (the inclination angle of the gradient) by considering a reference system which puts the zero towards the east and the rotation angle counter-clockwise. The aspect is 0 in the the south direction and then increases clockwise. The output is shown in fig.(4.3,a). The input raster map of Aspect is the map in output of Pitfiller.

- ExtractNetwork: it extracts the channel network from the drainage directions. Three operational modes are implemented. They differs in the way in which the start of the channel is modelled:

  - mode 0: by using a threshold value of the contributing areas (then only the pixels with contributing area greater than the threshold are the channel heads);
  - mode 1: by using a threshold value of the parameter: equivalent to a threshold value of the stress tangential to the bottom;
  - mode 2: by using a threshold value on the stress tangential to the bottom;

  After identifing the beginning of the channel, the points downhill are considered as channel.

  If "mode 0" is used, the inputs of ExtractNetwork are the map outputs of Pitfiller and DrainDir (both drainage directions and TCA). The output will be the raster map and, if the user needs also a vector file an output of the river network, fig.(4.3,c).

- NetNumbering: it assigns different numbers to each networks channel and labels it with the corresponding hillslope number which connects to the link . The input maps are: the file containing the flow directions (generated by DrainDir) and the map containing the channel network (generated by extractnetwork). There are two output raster maps: the network map with the numbered streams and the map containing the labelled sub-basins.

- HackLength: at a given point in a basin, it calculates the distance from the watershed along the network proceeding upstream along the maximal slope length. The input raster maps are: the drainage directions map (obtained with DrainDir) and the contributing areas map. The output is the raster map of the Hack distances.

- HackStream: it arranges a channel network starting from the branch according to Hack. The main stream is of order 1 and its tributaries of order, the sub-tributaries are of order 3, and so on. The input raster maps are: the drainage directions map (obtained with DrainDir), the total contributing areas, the Hack lengths map (obtained with hacklength), and the channel network (obtained with extractnetwork). The output raster map is the network ordered according the Hack lengths (129).

Now we are ready to obtain the features necessary to run the NewAge-JGrass components.

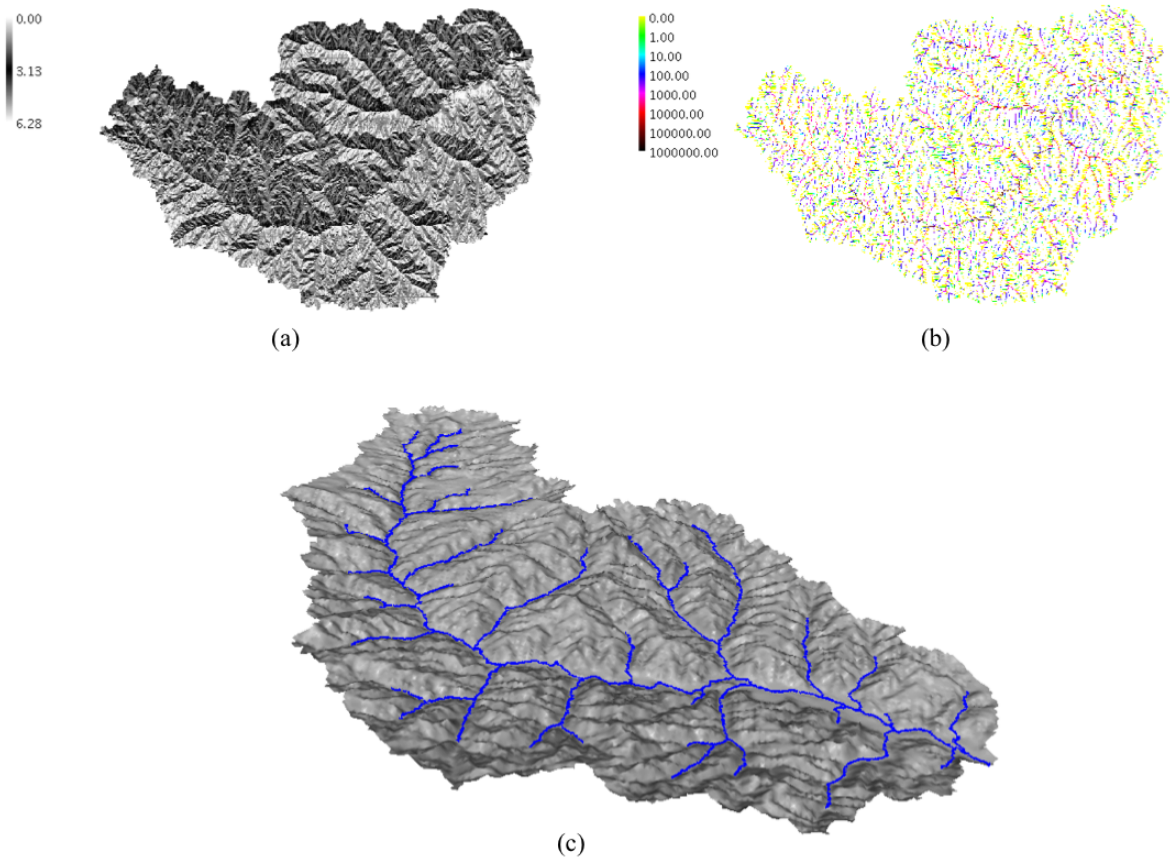

Figure 4.3: The Little Washita basin:output of Aspect, TCA and Extract Network - The figure shows the outputs of Aspect Horton Machine (a), TCA (b) and Extract Network (c).

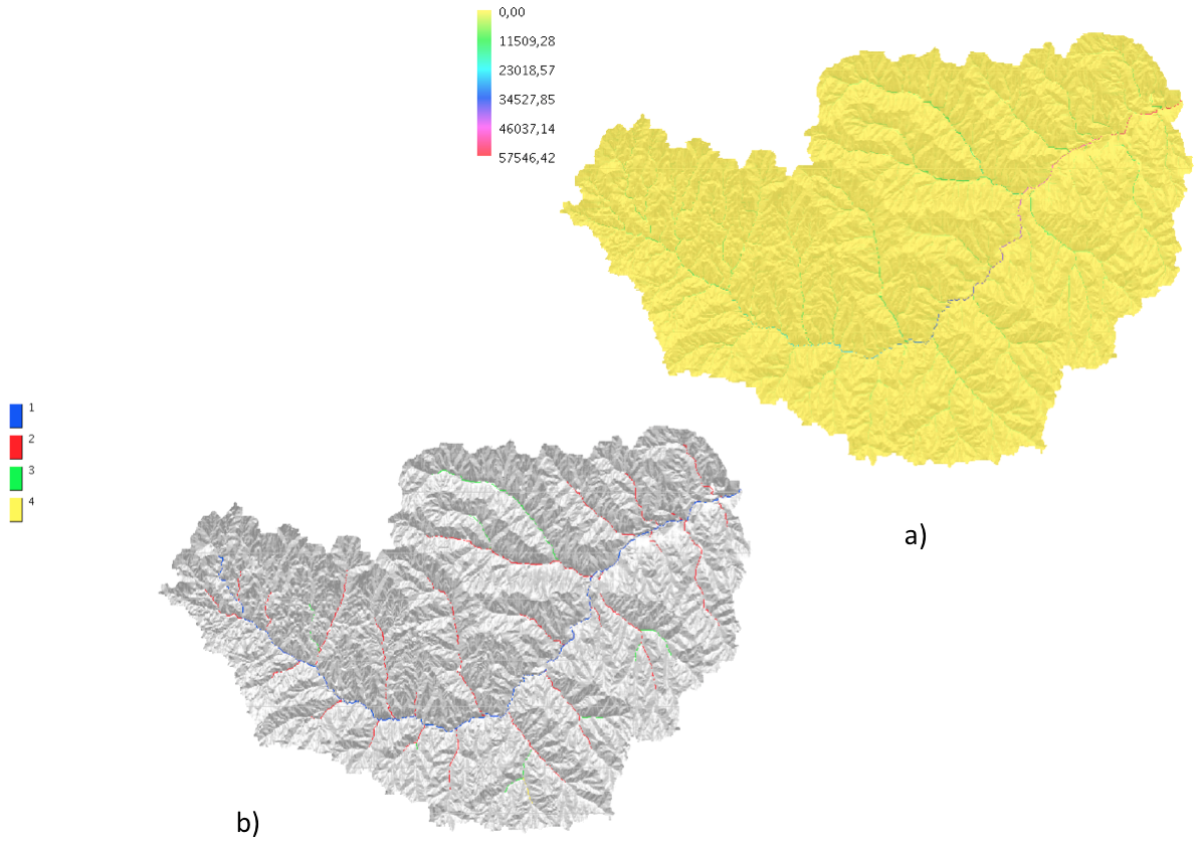

Figure 4.4: The Little Washita basin: output of HackLength, (a), and Hack Stream, (b). -

4.2.2 Delineation of the basin

The tools for the basin delineation are:

- BasinShape: it computes the basin split in hillslopes based on a shape file. The BasinShape inputs are the map output of Pitfiller and the map containing the labeled sub-basins (output of NetNumbering). The resulting shapefile contains the basin split for each hillslope whit some features such as:
  - Area: the hillslope area [m^2];
  - Perimeter: the hillslope perimeter [m];
  - netnum: the hillslope ID;
  - maxZ: the hillslope maximum elevation [m];
  - minZ: the hillslope minimum elevation [m];
  - avgZ: the hillslope average elevation [m];
  - height: the hillslope centroid elevation [m];
- The Pfafstetter algorithm: it builds the topology of the network by enumerating the river network structure according to a generalisation of the Pfafstetters numbering scheme (PNS) (e.g. (151), (54)).

  The PNS algorithm is defined as follows: starting from the outlet of the watershed, the main stream is delineated first. It uses the river streams extracted from the tree-like network of the drainage directions according to the algorithm presented in (113). This river network presents links (channel segments) separated by junctions, where tributaries meet. Each stream is characterized by an total upslope area, which is the total area of the basin attached in that stream. The main stream is obtained by following the river network, starting from the outlet going upstream. When finding a junction, the selected direction follows the channel link with the largest upslope area. In the case of equal areas a random direction is chosen. Each junction separates the main stream in links, which are numbered with the series of odd numbers starting with 1 at the outlet, (see fig.(4.5), below). Tributaries of the main stream are numbered with increasing even numbers while going upstream (assuming that two tributaries do not flow into the main stream at the same point), fig.(4.5). Tributaries can have sub-tributaries. As shown in fig.(4.5) one tributary has sub-tributaries, and therefore a second order numbering is used, represented by two digits separated by a point. The main stream of 8 is split into links 8.1 and 8.3 and two order 2 headwater basins are delineated: 8.2 and 8.3. As it appears, PNS uniquely identifies the watershed channels downstream of a point of interest (i.e. a droplet fallen into the 8.3 sub-catchment is guaranteed to flow into links 8.1, 7,5,3,1) allowing (i) an easy navigation through the river network, and (ii) to determine if two links are connected. PNS allows to prune out the smaller channels to analyse the basin at different scale of

resolution, therefore allowing to deal with of basins at multiple scales according the scope of the analysis performed. Clearly, the above topological and geometrical information can easily be translated into interconnected tables of a SQL database, as ishown as prototype implementation for the River Adige Basin Authority.

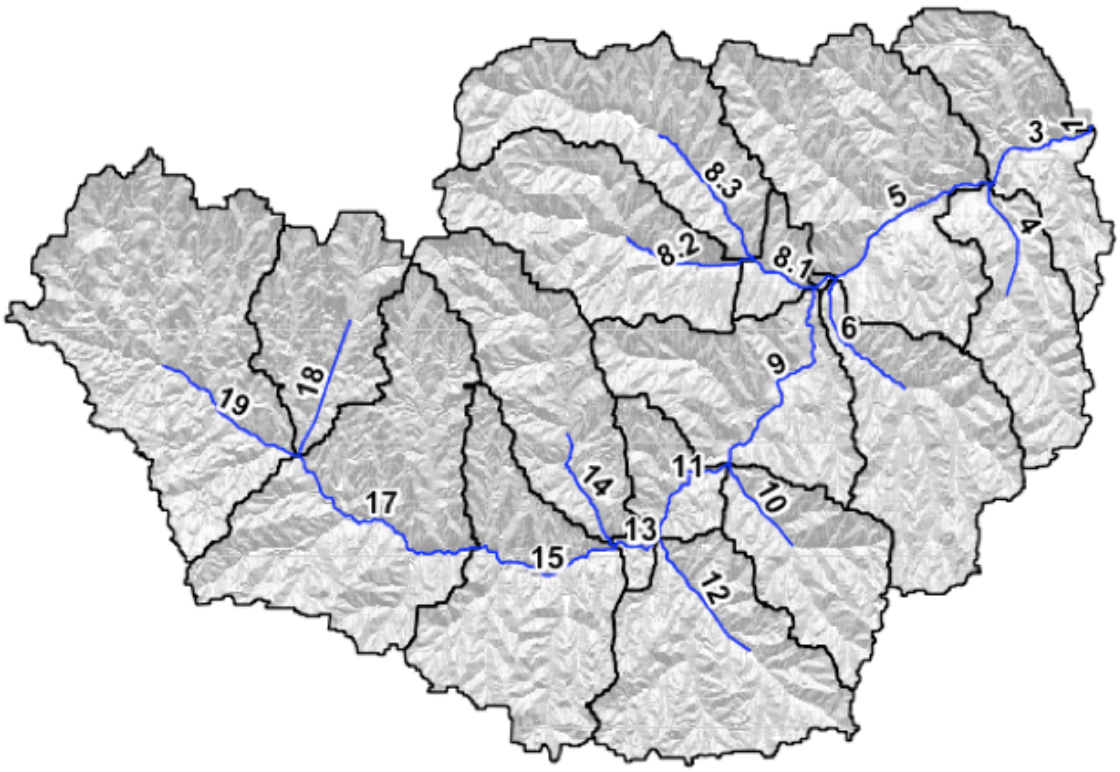

Figure 4.5: Pfafstetter numbering scheme for the Little Washita watershed, Oklahoma (U.S.) - The figure shows an example of Pfafstetter river network numbering.

The structure of a DWM is a crucial aspect of any model since it determines the components that can be used.

4.3 Discussion

As an identified requirement of the NewAge-JGrass system a formal definition of a DWM is needed. The topic is of particular interest because spatial information is significant for modelling of hydrological resources, even if remarkable results were obtained in hydrological science just by using point models (e.g (134)). One of the goals of NewAge-JGrass is to advance this philosophy and to allow the construction of minimalist model solutions which are aware of the spatial information and of its topological connections. Model deployment is based on the integration of geospatial modeling libraries by means of which the geographical information is encoded and becomes an integral part of the modelling process itself. OMS3 does the rest of the job by allowing the connection of all parts as well designed modelling solutions. For the application presented in this thesis simple storage techniques were used for all the information

(e.g. CSV files or other ASCII formats). However, it is clear that the system can be expanded to easily communicate with databases and servers, thus it therefore could be scaled to real operational situations where such systems are necessary.

5

Meteorological Interpolation Algorithms in NewAge-JGrass

This chapter contains all the information about the meteorological forcings interpolation algorithms implemented in the NewAge-JGrass. A common hydrological modelling problem is solved: the information transfer from meteorological measurement points to each element of the model: the pixels in fully distributed hydrological models case and the hillslopes or HRUs in the case of semidistributed hydrological models. This model perfectly matches with NewAge-JGrass: the interpolation component is able to provide outputs indipendently of the component which will consume them and of the simulation timestep which can be set by the user.

5.1 Introduction

Meteorological data (precipitation, temperature, and relative humidity) is usually available only in some few locations with respect to the number of points required by a common hydrological model (number of pixels for a fully distributed model or number of HRUs for a semidistributed model). Let assume that the number of measurement points is O and the number of points in which the meteorological variables are estimated is I with $I > O$. In the case of fully distributed hydrological models O is the number of cells in the study area which correspond to the center of each cells of a raster map. In the case of semidistributed hydrological models O is the number of HRUs in which the studied area is subdivided and may correspond to the centroids of each HRU. In the case of a fully distrubuted hydrological model, the Meteo Interpolation Components (MIC) require the digital elevation model of the river basin and performs the interpolations for each cell. In the case of the NewAge-JGrass hydrological model (51) the geomorphologic analysis components automatic splits the basin into HRUs and provides the shapefile of the HRU centroids in which the MIC performs the interpolations. Capturing the spatial and temporal variation of the meteorological forcing data is crucial for every hydrological model. Precipitation is considered the most important input in hydrological modelling ((8) and(9)), but air temperature and air humidity are also crucial to capture the physical and

biological processes that control the water cycle. Variability, in time and space, as an intrinsic feature of these stochastic variables, is amplified in mountain areas because of the more complex topography.

Many interpolation algorithms are presented in the literature to solve the spatial interpolation problem of meteorological variables. They have different grades of complexity ranging between deterministic approaches (areal mean, Thiessen Poligon (147), inverse distance (34)) and statistic approachs (multiple regression (111), optimal interpolation, geostatistic methods (59), (56)). Here, krigings interpolation algorithms were chosen for many reasons. First of all they are able to provide not only a measure of the interpolated variables but also an estimate of their variance (or prediction error) (59). Krigings are applied to different kind of meteorological variables(rainfall ((60), (116)), air temperature ((142), (69))). Moreover, krigings offers the possibility to include secondary information (auxiliary) in order to add more information and get robust estimations.

However, because the model infrastructure NewAge-JGrass is supposed also to run in real time a different interpolation algorithm, a simpler and more robust of the Kriging, was implemented previously by Hydrologis and revised in the current work. This method is called JAMI, abbreviation of Just Another Meteo Interpolation, which is explained and used in the next section.

5.2 Kriging(s)

As a geostatistical interpolation method, Ordinary Kriging (OK), (59) uses the semivariogram function (59) to specify spatial data variability. The experimental semivariogram measures the statistical correlation as a function of distance. It is computed by considering the difference between observations separated by a distance $\mathbf{h}$:

$$2 \cdot \gamma(\mathbf{h}) = \frac{\sum_{j=1}^{N(\mathbf{h})} (P_j - P_{j+h})^2}{N(\mathbf{h})} \tag{5.1}$$

where $2 \cdot \gamma(\mathbf{h})$ is the experimental variogram, $N(\mathbf{h})$ is the number of points separated by a distance $\mathbf{h}$, and P_j is the meteorological variable measured at the point j. Usually, $\gamma(\mathbf{h})$ is then fit by a theoretical model in a way to derive semivariogram values for any possible lag $\mathbf{h}$ are required by the interpolation. As presented in (59) and (33) not all functions are permitted. The semivariogram model must be conditionally negative. The estimated theoretical semivariogram model will be used for the resolution of the ordinary kriging system to compute the unknowns weights α_k (59), (155):

$$\begin{cases} \sum_{i=1}^{O} \alpha_i \cdot \gamma_{i,j} - \mu = \gamma_{i,0} \quad j = 1, ..., O \\ \sum_{i=1}^{O} \alpha_i = 1 \end{cases} \tag{5.2}$$

where μ is the Lagrange parameter used to constraint the weights to unity, $\gamma_{i,j}$ is the semivariance of the meteorological variable between locations i and j, and the term $\gamma_{i,0}$ represent the semivariance between the measurement point i and the interpolation point 0.

A variant of ordinary kriging is the local ordinary kriging (LOK). Here, a user defined number of nearest observations should be used for the interpolation. In this case the estimate is only influenced by the measurements belonging to the neighbor. It can be defined in two ways: by specifying the maximum searching radius or the maximum number or closer stations to the point in which interpolate.

Detrended Ordinary Krgiging (DOK), (57) takes into account not only of the horizontal spatial variability but also the vertical spatial variability of the meteorological variables. This is very important for interpolating the meteorological variables such as rainfall and temperature which correlate to elevation. In many cases this cannot be neglected (56).

In summary, in order to capture the vertical spatial variability a linear regression model is used and it is computed for each simulation timestep. If there is a trend for meteorological variables and elevation and it is statistically significant then the residuals from this linear trend are computed for each meteorological stations. The residual semivariogram is estimated and the OK of the residual is performed. The final interpolated field of the meteo variable is given by adding two terms: i) the residuals' spatial field which takes into account of the horizontal component variability; ii) the prevously estimated linear trend that depends on the elevation of the interpolation point considering the vertical component variability.

In the case in which the meteorological variable-elevation trend is statistically insignificant a classical ordinary kriging is performed.

5.3 Motivation for Semivariogram modelling and providing krigings tools in NewAge-JGrass.

There are many reasons for implementing krigings tools in NewAge-JGrass:

- For each time step, krigings tools outputs can be linked to all hydrological components implemented in NewAge-JGrass (short wave energy component, runoff-routing component, and snow melting model). The linkage between different components can be realized by a simple scripting language and can be executed in the OMS3 console or SpatialToolbox.

- For a selected model solution (meteorological interpolation, radiation, evapotraspiration, snow melting, runoff production, and routing component), the influence of different types of interpolation algorithms can be easily investigated by substituting the interpolation component while preserving the same model solution.

- The GIS based structure of the NewAge-JGrass system facilitates the input/output process. JGrass is able to manage in a efficient way both raster and vectorial data by using

concepts like GridCoverage and feature collections. Moreover, it allows for visualization and management of interpolated raster outputs.

- For the semivariogram parameter estimation the calibration algorithms implemented in NewAge-JGrass can be used.
- Within the NewAge-JGrass applications the computational time is minimized by avoiding writing results to files. Data between the components is exchanged by HashMaps (collection of value such as (HRU-id, value)) or GridCoverage (GeoTools object used to store raster data).

The next subsections presents the krigings tools. The process is splitted into three parts: experimental variogram computation, theoretical variogram model estimate, and krigings component application.

5.3.1 The Experimental Variogram Component (EVC).

The variogram is the foundation for the geostatistic and kriging theory are founded. As a necessary tool for krigings applications an OMS3 component for the experimental variogram calculation is implemented within NewAge-JGrass infrastructure. The module, named "Variogram" is quite similar to the Gstat Variogram (115) algorithm implemented in the software environment for statistical computing R, (13). The algorithm's flowchart is presented in fig.(5.1).

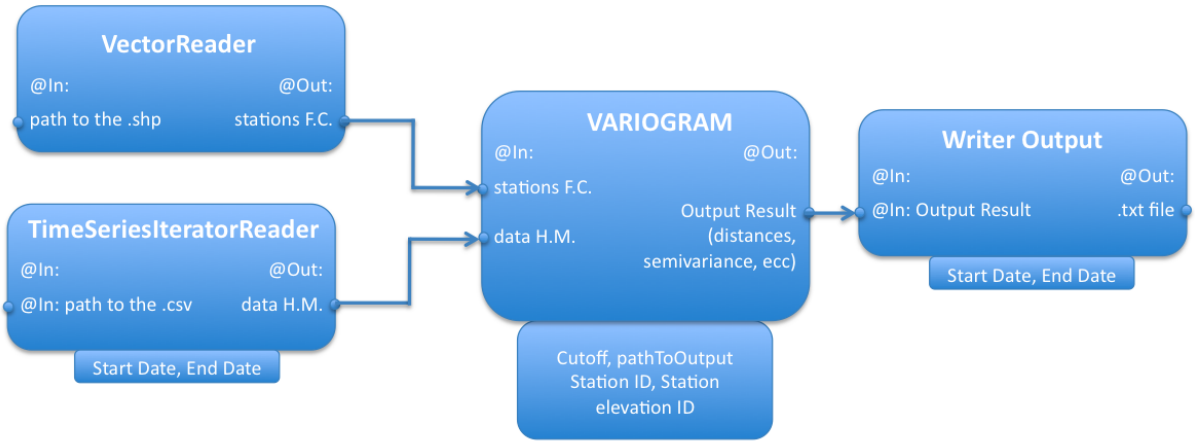

Figure 5.1: Variogram workflow.

After reading the input data (the measured values as .csv file and the shapefile of the measurement stations) the Variogram component computes not only the experimental variance but also some other informations about the correlations such as the Moran and Geary autocorrelation.

Moreover, the user is able to define the cutoff value (the spatial separation distance which includes point pairs in semivariance estimates). As a default, cutoff is equal to the length of the

diagonal of the box spanning the data divided by three. The algorithm is tested on the JURA dataset and the results are presented in Appendix 2.

5.3.2 Vgm: Theoretical Variogram Computation

Prediction of the spatial process at unsampled locations by techniques such as ordinary kriging requires the assessment of a theoretical semivariogram or covariance model because the krigings need to know the semivariogram values for lag distances other than those used in the empirical semivariogram. The theoretical variogram is a fitting of the experimental variogram by using a theoretical variogram functions which has peculiar properties (29). In practise, the semivariogram model needs to be non-negative definite, in order for the system of kriging equations to be non-singular.

The theoretical variogram model will be used in the kriging algorithms to compute the weigths reported in eq.(11.9).

The JGrass-VGM component is able to compute some of the more common theoretical semivariogram models. All these models are reported in tab.(5.1). All theoretical models have

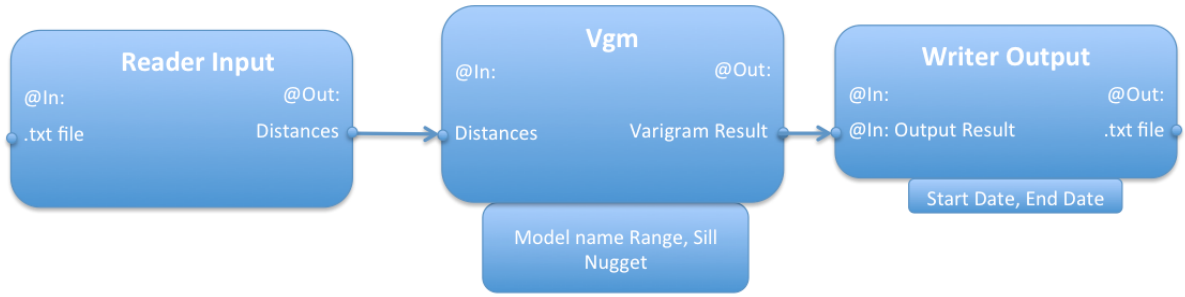

Figure 5.2: The VGM flowchart.

three parameters:

- Nugget (N): is the not zero value the semivariance may have at infinite small separation distance; in tab.(5.1) a nugget N could be added to any presented semivariogram model.
- Sill (s): The semivariance value at which the variogram levels off.
- Range (r): is the lag distance at which the threshold is reached.

Those parameters are tuned in order to find the best theoretical model which fits the experimental variogram.

Fig.(5.2) shows the flowchart of the VGM component: the input is the distances vector corresponding the theoretical semivariance. The output is the .txt file containing the theoretical semivariances.

Table 5.1: Theoretical semivariogram implemented in the VGM OMS3 component

Name	Formula
Spherical	$\begin{cases} \gamma(h) = s \cdot \left[1.5 \cdot \frac{h}{r} - 0.5 \cdot \left(\frac{h}{r}\right)^3\right] & h < r \\ \gamma(h) = s & h \geq r \end{cases}$
Exponential	$\gamma(h) = s \cdot \left[1 - e^{-\frac{\lvert h \rvert}{r}}\right]$
Linear	$\begin{cases} \gamma(h) = s \cdot \left(\frac{h}{r}\right) & h < r \\ \gamma(h) = s & h \geq r \end{cases}$
Power	$\gamma(h) = s \cdot (h^r)$
Gaussian	$\gamma(h) = s \cdot \left(1 - e^{-\frac{h}{r}}\right)$
Circular	$\begin{cases} \gamma(h) = s \cdot \frac{2}{\pi} \cdot \left(\frac{h}{r} \cdot \sqrt{1 - (\frac{h}{r})^2} + \arcsin \frac{h}{r}\right) & h < r \\ \gamma(h) = s & h \geq r \end{cases}$
Bessel	$\gamma(h) = s \cdot \left(1 - \frac{h}{r} \cdot k_1\left(\frac{h}{r}\right)\right)$
Pentaspherical	$\begin{cases} \gamma(h) = s \cdot \left(\frac{15}{8}\frac{h}{r} - \frac{5}{4}\frac{h^3}{r^3} + \frac{3}{8}\frac{h^5}{r^5}\right) & h < r \\ \gamma(h) = s & h \geq r \end{cases}$
Periodic	$\gamma(h) = s \cdot \left(1 - cos(2\pi \cdot \frac{h}{a})\right)$

5.3.3 The krigings tools in the NewAge-JGrass system

After the variogram assessment, we are able to apply it for kriging interpolation of a dataset. The flow chart of the kriging algorithm is presented in fig.(5.3). The input data are: i) the shape file of the measurement stations, ii) the .csv file of the measured data, iii) the shape file or the raster map of the interpolations points, iv) the semivariogram model to use for the interpolation. The model parameters are: a flag to specify the working mode (raster or vector), the semivariogram model parameter, a flag to specify the kriging type (ordinary, local, or detrended) and some control parameters related to the selected kriging algorithm (maximum distance for local kriging, threshold of the correlation between elevation and measurements for detrended kriging). Within kriging model configuration, different variogram models can be used for different time steps. The outputs could be or a .csv file or a raster map with the interpolated values.

Comparisons with the R-package Gstat (115) are presented in Appendix 1 in order to test the implemented algorithms (ordinary and local kriging).

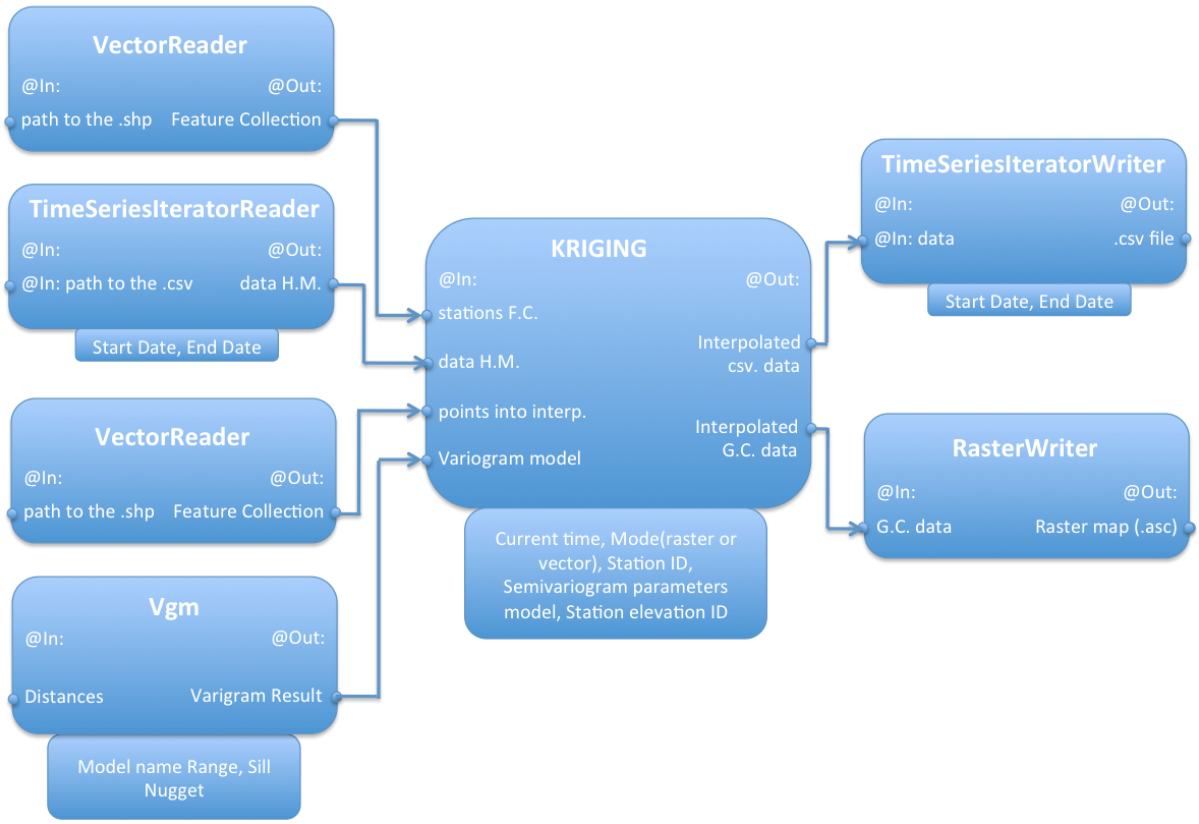

Figure 5.3: The Kriging flowchart.

5.4 Theoretical semivariogram fitting method: OMS3 components

There exist different methods for fitting semivariogram models, such as least squares, maximum likelihood, and other robust methods presented in (33). By using different OMS3 NewAge-JGrass components it is possible to use the least squares procedure. Thes is summarized in a few steps:

- run the Variogram component in order to compute the experimental semivariance values;
- select a theoretical semivariogram from those presented in tab.(5.1);
- tune the model parametersl by using the OMS3 optimization algorithms;
- store the results which are the best model parameter set and the objective function's optimum value.

After repeating the procedure for a user defined number of theoretical semivariogram, the algorithm selects the model and parameters which give the best values. This will be the theoretical model used for the kriging application.

The procedure which includes the experimental semivariograms estimate, optimal theoretical model parameters estimate and the kriging interpolator algorithm application is implemented in a OMS3 script which is provided in Appendix 3. The flowchart is presented in fig.(5.4). Each block in fig.(5.4) is a OMS3 model. Firstly, the EVC component runs and provides as output the experimental variogram estimate. It will be the input for the Particle Swarm Optimization (PSO) algorithm which calibrates one of the VGM theoretical semivariograms. PSO tunes the model parameters to minimize a function of theoretical and experimental semivariance values. In fig.(5.4), the red dashed lines represent the connections between the OMS3 models and the blue dashes represent the connection between components in each model.

The theoretical semivariogram model estimate procedure is verified by comparison with the R-Gstat package. The comparison results are presented in Appendix 4.

5.5 Semivariogram estimate and Krigings application and verification

The complete procedure presented in the previous section is now tested on the Teton river basin, for the precipitation interpolation. By working with the dataset, some stations are excluded and the interpolation is performed at the removed stations. This allows the comparison between interpolated and measured time series. The algorithms applied for the interpolations are: Ordinary Kriging using all the available stations, Local Ordinary Kriging, and Detrended Ordinary Kriging using different neighbors. For each simulation time step a best semivariogram estimate is performed according to the methodology presented in the previous section.

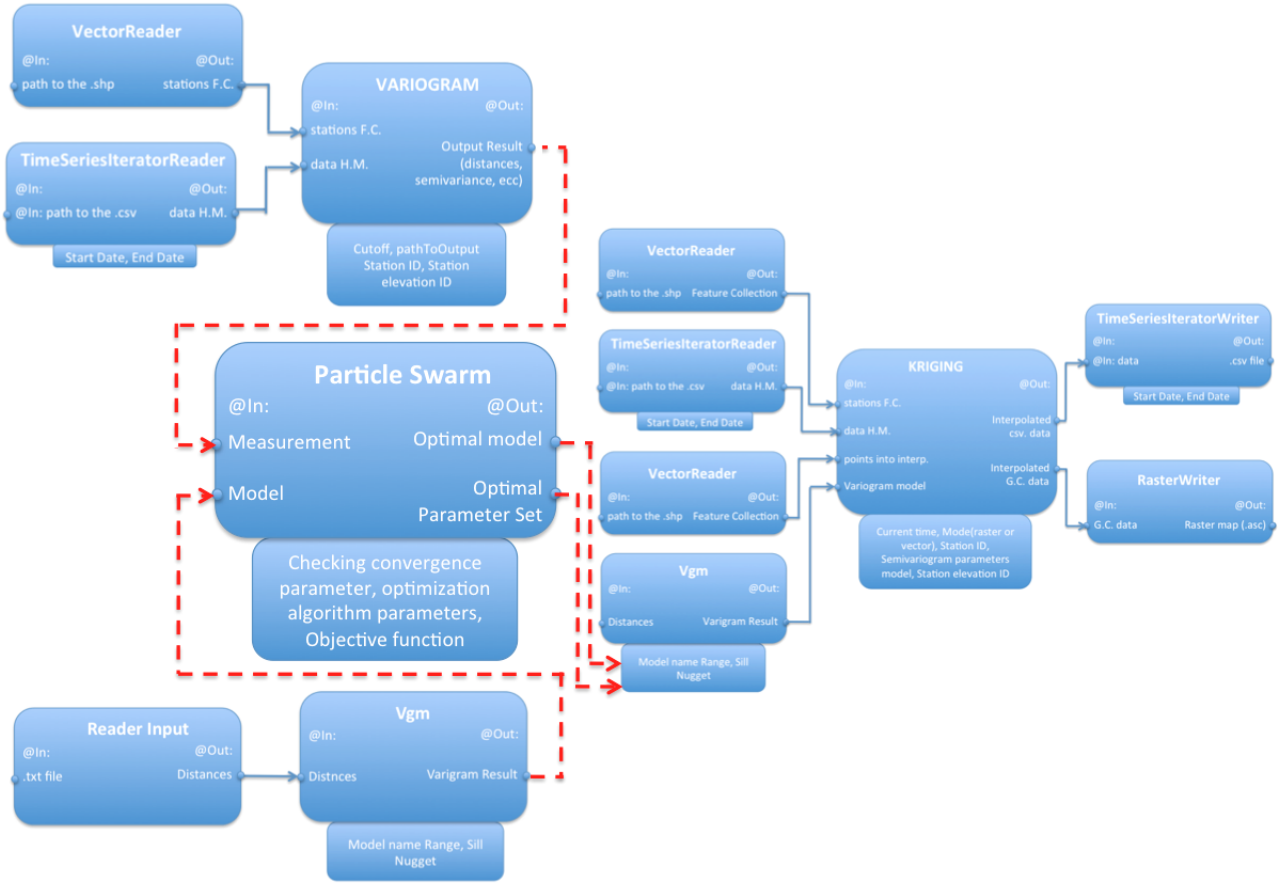

Figure 5.4: Workflow of kriging parameter estimation and interpolation. The red dashed lines represent the connections between the OMS3 models. The blue dashed lines represent the connection between components in each model. After computing the experimental variogram, the Particle Swarm algorithm is used to estimate the theoretical model parameters. Finally the kriging algorithm runs.

5.5.1 Teton river basin daily precipitation interpolation.

The Teton river basin dataset is used to verify the interpolation algorithms performances on a river basin with complex topography for daily time step rainfall. For the verification procedure three stations of the complete dataset as presented in chapter 3 are removed and at these positions the interpolated time series are computed. This allows the comparison between measured and interpolated time series. Daily rainfall data for three years (from 1980-10-01 to 1983-10-01) was interpolated. Different interpolations algorithms were applied. Local ordinary kriging was used with a radius equal to 25,50 and 75 km. Two types of detrended ordinary kriging were applied: in the first case the trend is computed by considering all the available stations and in the second case by considering just the neighboring stations. Fig.(5.5) shows the correlation coefficient for the three validation stations and for the different interpolation algorithms. The results shows that detrended ordinary kriging in general performs very well compared to ordinary and local kriging even if the correlation coefficient which measure the goodness of the fit between measured and interpolated values does not increase significally.

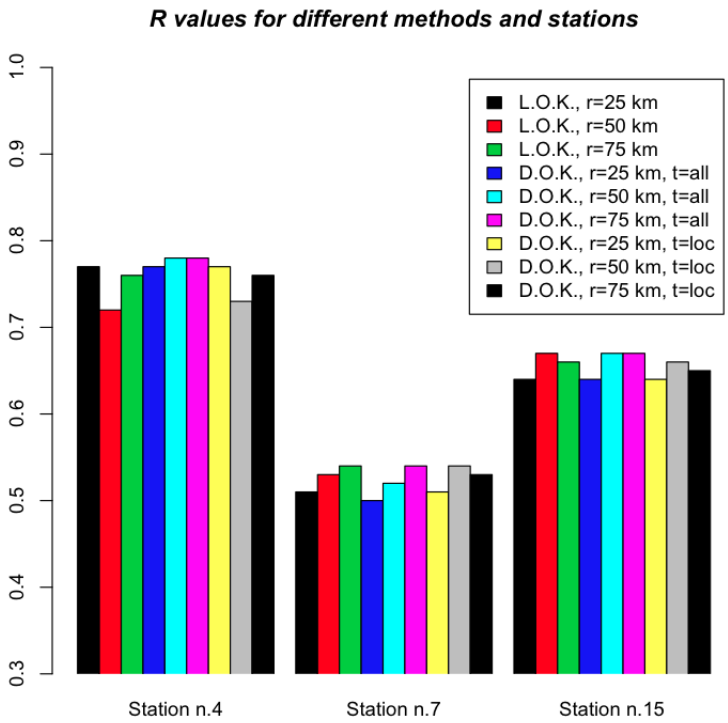

Figure 5.5: Teton river basin rainfall interpolation algorithms comparison. For each validation station and for each interpolation algorithm, the correlation coefficient between measured and interpolated time series is presented.

5.6 JAMI

JAMI (Just Another Meteo Interpolator) is the meteorological spatialization algorithm implemeted for the meteorological variables depending on the elevation, such as air temperature, air pressure, and relative humidity. JAMI is fully integrated into the NewAge-JGrass infrastructure. It is only able to interpolate the meteo-station measurement for the centroid of each hillslope but it also able to preserve the meteo forcing data variability at each hillslope. For this reason JAMI uses the concept of altimetric band. The user can define the number of altimetric bands, each splitting the elevation range of each hillslope. The algorithm provides the interpolated time series of the meteo-variable for each centroids of each altimetric band, as presented in fig.(5.6). JAMI's theoretical background is explained in Appendix 5. The next sections present JAMI applications and verifications for interpolating temperature and air humidity.

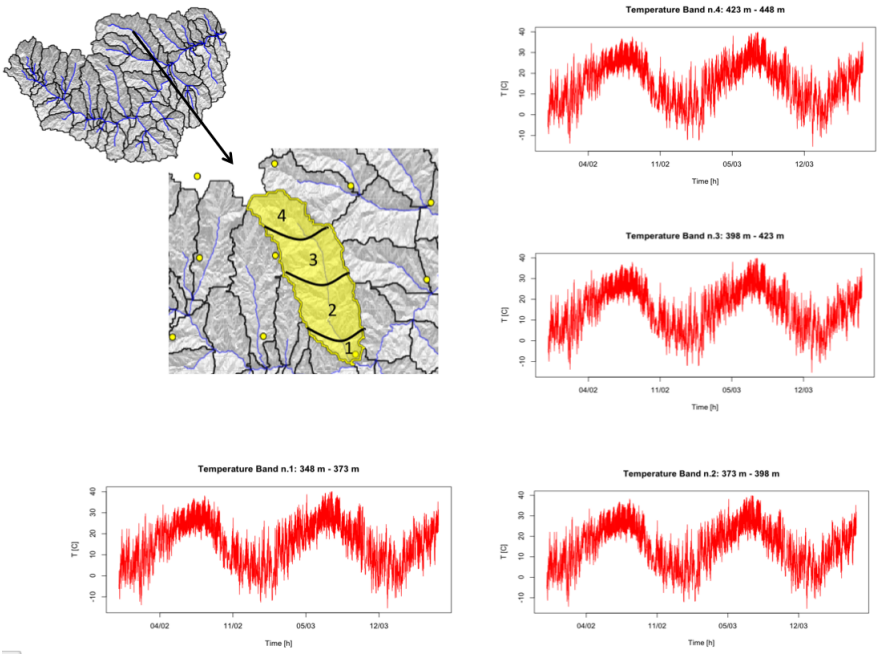

Figure 5.6: JAMI temperature interpolation in Little Washita watershed, Oklahoma (U.S.)

5.7 JAMI: applications and validations

The JAMI interpolator is applied and validated on two different river basins: Little Washita and Piave river basin. The air temperature interpolation is performed on both the river basin. The relative humidity interpolation is performed only on the Little Washita river basin. Some of the measurement stations for each basin were left out. The algorithm was applied and the interpolated time series is computed for the measurement station that were left out. The comparison between interpolated and measured meteorological variables is performed from a qualitative point of view (scatterplot) and from a quantitive point of view (using classical

goodness of fit measure: Index of agreement (IOA), Pearson Correlation coefficient (R) and the percentual bias (Pbias).

5.7.1 JAMI Air Temperature interpolation: applications and validations

In the Piave river basin the algorithm was applied for temperature by using 19 measurement stations represented by circles in fig.(5.7). The simulation period is one year (2010) and the simulation time step is hourly. The algorithm is verified at 3 stations (ids: 29, 3 and 9) represented by triangles in fig.(5.7). Fig.(5.7) shows the scatter plot of the measured and interpolated values. Finally, tab.(5.2) show the goodness of fit for the three validation stations.

Table 5.2: Index of goodness of fit for the three validation stations within Arabba river basin

Station ID	R	IOA	PBIAS
29	0.97	0.98	4.5
3	0.95	0.94	14.4
9	0.94	0.92	16.2

Within the Little Washita river basin JAMI was applied for temperature by using 17 measurement stations represented by circles in fig.(5.8). The simulation period is one year (2002) and the simulation time step is houly. The algorithm is validated at 3 stations (ids: 12, 8 and 16) represented by triangles in fig.(5.8). Fig.(5.8) shows the scatter plot of the measured and interpolated values. Finally, tab.(5.3) shows the goodness of fit for the three validation stations.

Table 5.3: Index of goodness of fit for the three validation stations within Little Washita river basin

Station ID	R	IOA	PBIAS
12	0.99	0.99	1.1
8	0.99	0.99	1.2
16	0.99	0.99	1.9

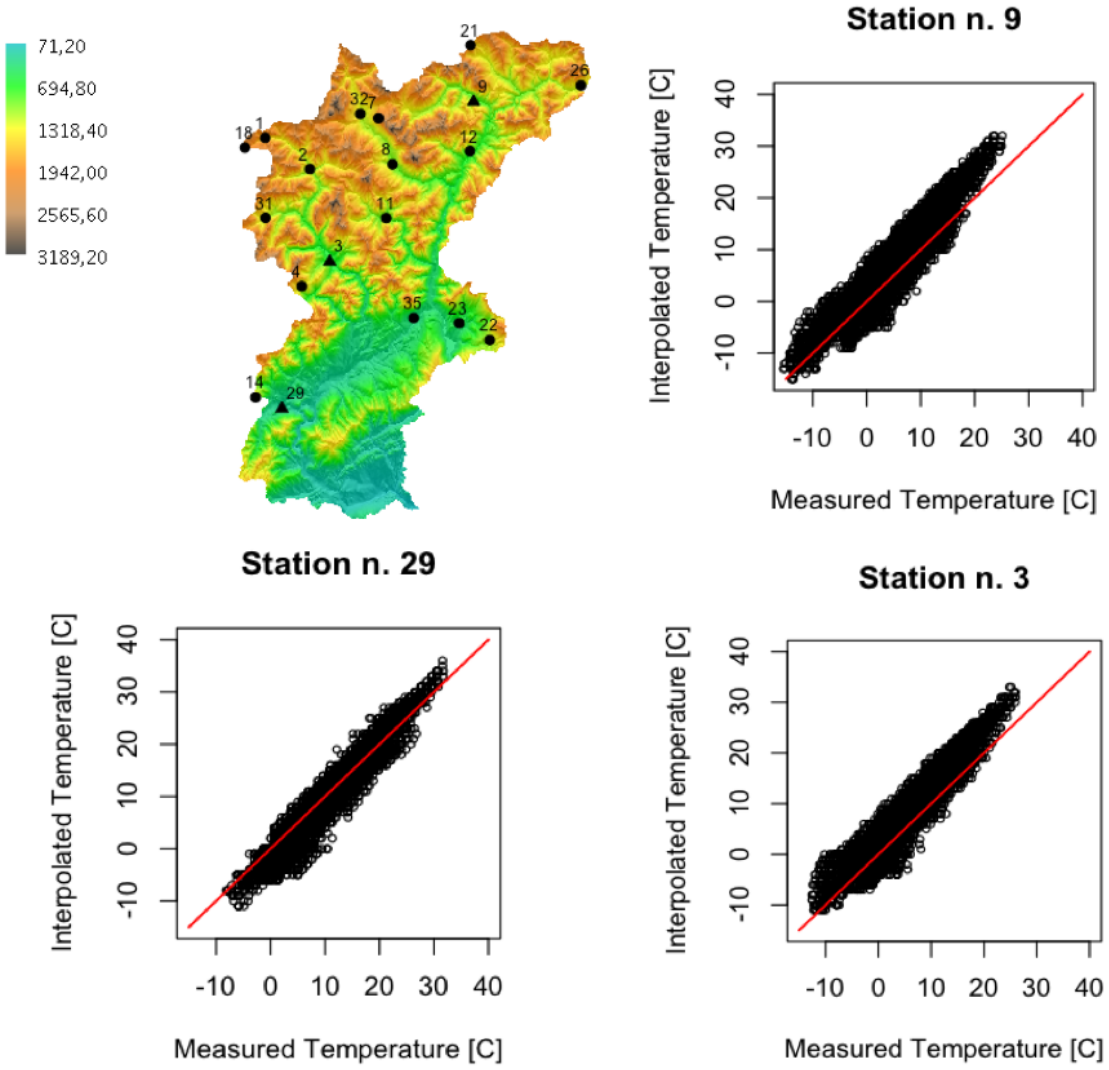

Figure 5.7: Arabba river basin a the top: measuremet (circles) and validation (triangles) stations. Scatter plot measured vs. interpolated temperature for the validation stations.

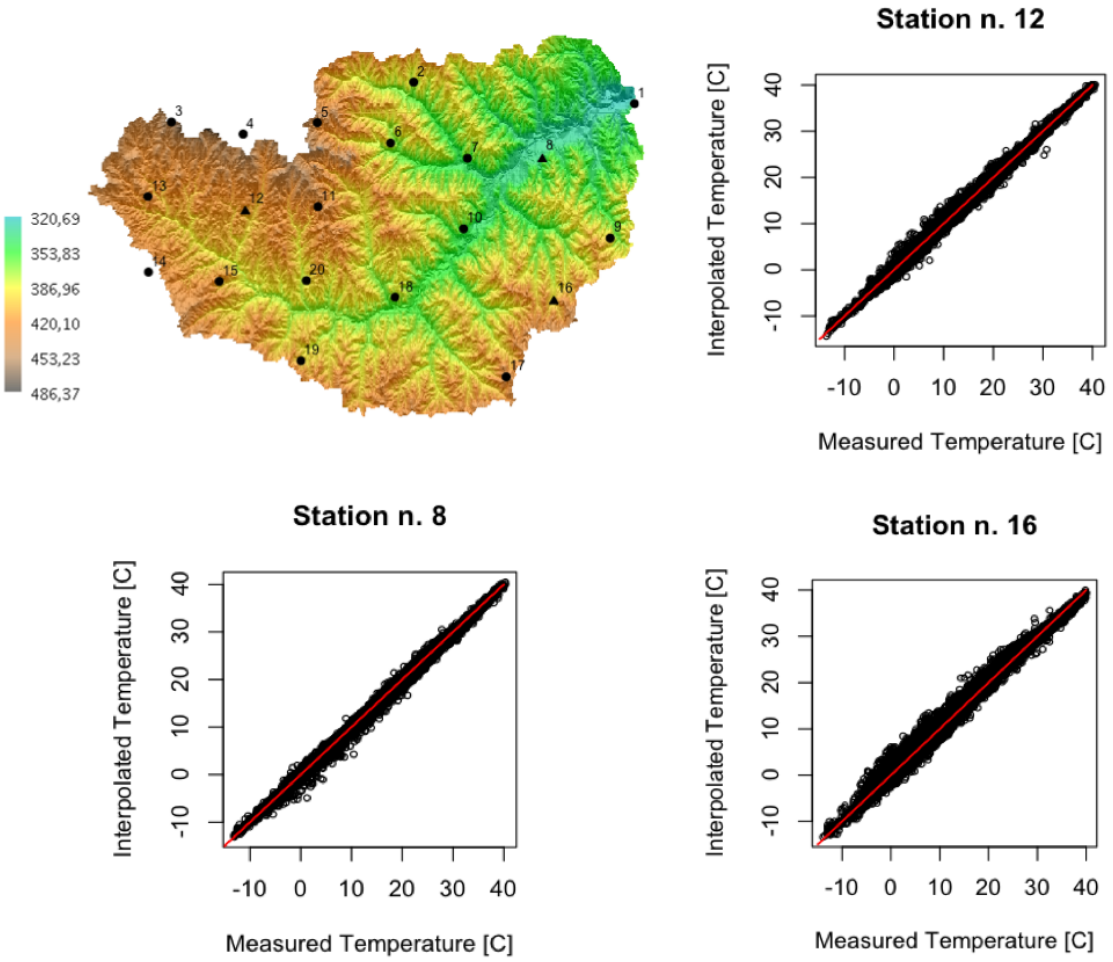

Figure 5.8: Little Washita river basin a the top: measuremet (circles) and validation (triangles) stations. Scatter plot measured vs. interpolated temperature for the validation stations.

5.7.2 JAMI Relative Humidity interpolation: applications and validations

The same methodology as applied for the temperature interpolation is repeated for the relative humidity interpolation. Because measured relative humidity data was only available for Little Washita river basin, the model is not applied for the Arabba river basin.

For the Little Washita river basin, fig.(5.9) shows the scatter plot of the measured and interpolated values and tab.(5.4) shows the goodness of fit for the three validation stations.

Table 5.4: Index of goodness of fit for the three validation stations within Little Washita river basin

Station ID	R	IOA	PBIAS
12	0.98	0.97	1.6
8	0.99	0.96	1.4
16	0.98	0.96	1.7

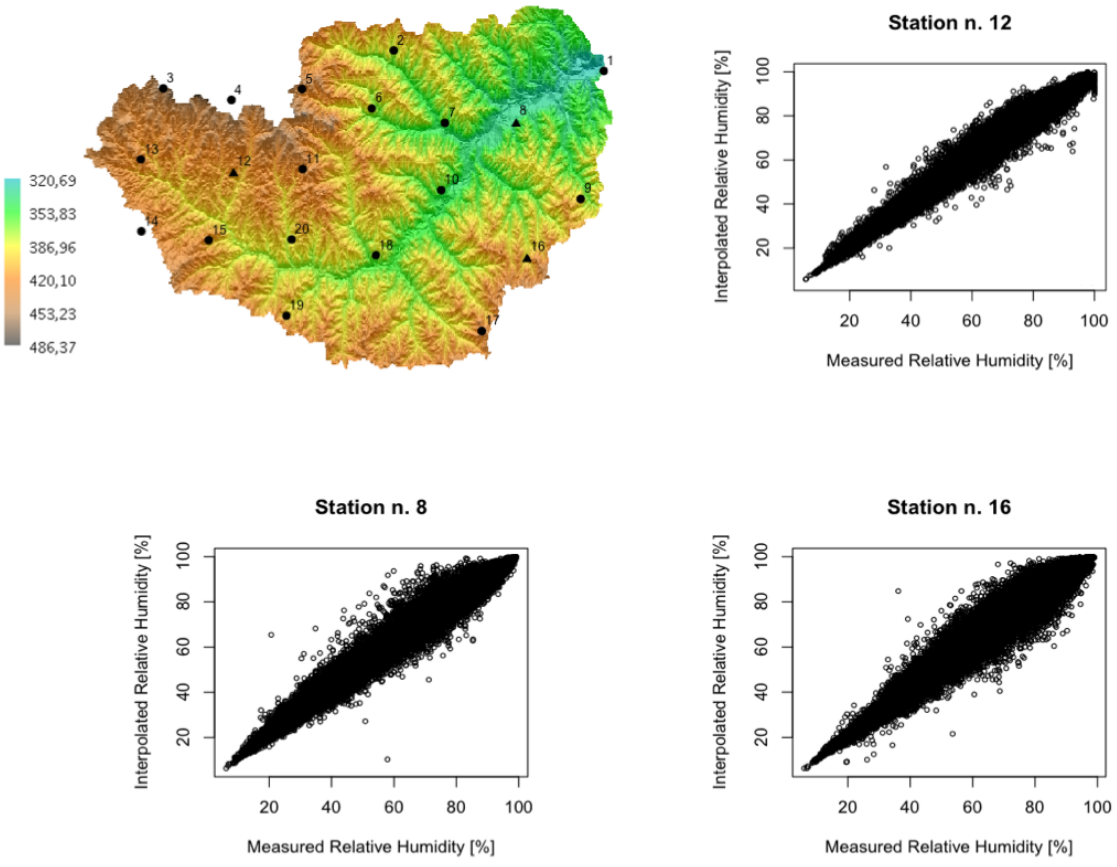

Figure 5.9: Little Washita river basin a the top: measuremet (circles) and validation (triangles) stations. Scatter plot measured vs. interpolated temperature for the validation stations.

5.8 Conclusion

This chapter presented the meteorological interpolation algorithms of the NewAge system. Theoretical aspects of geostatistic and deterministic methods are discussed. Applications on different river basins with different topographic complexities and density of the measurement stations are shown. The NewAge system allows the user to easily compare all the algorithms as presented and to visualize the results. The methods were able to operate in two modes: raster based method, providing the raster map of the interpolated variable and as vector method, providing its point time series. For this reason they are helpfull both for semi-distributed and for fully distributed hydrological models. The algorithms, can be easily integrated or substituted with other interpolation algorithms from in the OMS3 core (e.g. inverse distance weighted). All the tools will be very helpful for many other components of the system such as the shortwave radiation, the rainfall runoff, and the snow water equivalent component.

6

NewAge-JGrass Shortwave radiation model

This chapter presents two modelling components based on the OMS3 for the calculation of the shortwave incident radiation on a complex topography settings. The first component, NewAGE-SwRB, accounts for slope, aspect, shadow, and the topographical information of the sites. It uses a suitable parametrisation in order to obtaining the cloudless irradiance. A second component, NewAGE-DEC-MOD's is implemented to estimate the irradiance reduction due to the presence of clouds, according to three parameterisations. After a short introduction on shortwave radiation balance in section 1, section 2 and 3 deals with models equations for NewAGE-SwRB and NewAGE-DEC-MOD's, respectively. Section 4 presents the models applications on three different river basins with complex climate and topography. The last two sections present the long wave radiation model and two evapotranspirations model implemented in the system.

6.1 Shortwave radiation budget

Solar radiation at the top of the atmosphere is function of sun activity. In the case of hydrological studies, the solar constant, $I_{sc} \sim 1367$ [W m^{-2}] is used as a suitable approximation of the irradiance at the top of the atmosphere. This value represents the maximum irradiance when the solar beam hit orthogonally the Earth. Reduction of irradiance due to latitude and longitude, the day of the year, and the hour, is necessary and can be easily calculated with the desired approximation, e.g. (72) and (86).

In the absence of clouds, solar radiation arrives at the Earth's ground surface in two classes. Direct radiation ($S^* \downarrow$) is the part of the solar beam which arrives at the surface without any interaction with the Earth's atmosphere. Diffuse radiation ($d^* \downarrow$) is shortwave radiation scattered down back to the Earth's surface after hitting molecules of the atmospheric gases and aerosols or after upwards reflection by the Earth's surface and atmospheric components. We will call the sum of $S^* \downarrow$ and $d^* \downarrow$, total Shortwave Radiation at the ground ($R^* \downarrow_{sw}$). NewAGE-SwRB (or simply SwRB) was developed to be able to simulate the direct shortwave radiation budget

at multiple locations in a landscape, and to provide inputs to hydrological components independently of their geographical structure (either implementing fully distributed, semi-distributed, or lumped concepts). From a spatial point of view, the output of SwRB can be a raster (the results are provided for each pixel of the computational domain) or vector (the results are provided only at some points of the computational domain) according to the the modeller's needs, and in Open GIS Consortium standard formats (as GridCoverage and shapefiles respectively). For the various use, the component was implemented to be able to provide results using a generic hourly, sub-hourly, and daily time step, according to the users' specifications.

While not trivial to obtain, the geometrical dissemination of the radiation that returns the incoming solar radiation on a tilted plane, is estimated according to the elegant solution provided by Corripio's algorithms, (30) and (31). Therefore, it is assumed that the solar constant, I_{sc} has been spatially corrected to account for the geometry and the position of the landscape in order to provide a "corrected" solar constant, $\hat{I}_{sc}$.

In the next two subsections direct and diffuse solar radiation model equations are presented.

6.1.1 Direct Solar Radiation under cloudless sky conditions

Therefore, the incident $R\downarrow_{sw}$, on an arbitrary slope surface in a point, under cloudless sky condition is given by (30):

$$R\downarrow_{sw} = C_1 \cdot \hat{I}_{sc} \cdot E_0 \cdot \cos(\theta_s) \cdot (T_s + \beta_s) \cdot \psi \tag{6.1}$$

in which:

- $C_1 = 0.9751$ is the fraction of solar radiation that is included between 0.3 μm and 3.0 μm wavelength
- E_0 [-] is a correction factor related to the Earth's orbit eccentricity computed according to (141):

$$E_0 = 1.00011 + 0.034221cos(\kappa) + 0.00128sin(\kappa) + \\ +0.000719cos(2\kappa) + 0.000077sin(2\kappa) \tag{6.2}$$

$$\kappa := 2\pi \cdot \left(\frac{N-1}{365}\right) \tag{6.3}$$

  where κ is the day angle [rad] and N is the day number of the year (N=1 on 1 January, N=365 on 31 December);

- T_s [-] the product of the atmospheric transmittances, is defined as:

$$T_s := \tau_r \cdot \tau_0 \cdot \tau_g \cdot \tau_w \cdot \tau_a \tag{6.4}$$

where the τ functions are the transmittance functions for Rayleigh scattering, ozone, uniformly mixed gases, water vapour, and aerosols, respectively. They are computed for each point as defined in the last part of this section.

- β_s [m] is a correction factor for increased transmittance with elevation $z[m]$, defined according to (30):

$$\beta_s = \begin{cases} 2.210^{-5} \cdot z_p & if \;\; z \leq 3000\ m \\ 2.210^{-5} \cdot 3000.0 & if \;\; z > 3000\ m \end{cases} \tag{6.5}$$

- θ_s [rad] is the angle between the Sun vector and the surface plane ((31)); for a horizontal surface $\theta_s = \theta_z$ where θ_z is the zenith angle.

- ψ_s is the shade index that accounts for the sun or shade of the analysed point, and is modelled according to (31):

$$\psi_s = \begin{cases} 1 & if \;\; \text{the point p is in the sun} \\ 0 & if \;\; \text{the point p is in the shadow} \end{cases} \tag{6.6}$$

The atmospheric transmittances in (11.9) are estimated according to (12) and (72). They are functions of the atmospheric pressure, the ozone layer thickness, the precipitable water amount, the zenith angle and visibility, which are eventually taken assumed fixed values, according to the literature values reported in tab.(6.1).

Table 6.1: List of the SwRB component parameter used in simulations.

Symbol	Parameter description	Dimension	Values
l_{oz}	vertical ozone layer thickness	[cm]	0.30
V	visibility, (30)	[km]	80.0
ω_0	single-scattering albedo fraction of incident energy scattered to total attenuation by aerosols	[-]	0.9
F_s	fraction of forward scattering to total scattering	[-]	0.84

The transmittance function for Rayleigh scattering τ_r [-] is estimated as:

$$\tau_r = \exp\left[-0.0903 \cdot m_a^{0.84} \cdot (1 + m_a - m_a^{1.01})\right] \tag{6.7}$$

where m_a [-] is the relative air mass at actual pressure defined as:

$$m_a := m_r \cdot \left(\frac{p}{1013.25}\right) \tag{6.8}$$

in which p[mbar] is the local atmospheric pressure and m_r[-] is the relative optical air mass:

$$m_r = \frac{1.0}{\cos(\theta_s) + 0.15(93.885 - (180/2\pi)\,\theta_s)^{-1.253}} \tag{6.9}$$

The transmittance by ozone τ_o [-] is defined as:

$$\tau_o = 1.0 - \left[0.1611 l_{oz} m_r (1.0 + 139.48\, l_{oz} m_r)^{-0.3035} + - \frac{0.002715\, l_{oz} m_r}{1.0 + 0.044\, l_{oz} m_r + 0.0003\, (l_{oz} m_r)^2}\right] \tag{6.10}$$

where l_{oz} [cm] is the vertical ozone layer thickness, and the coefficients have the appropriate dimensionality to make τ_0 dimensionless.

Transmittance by uniformly mixed gases τ_g [-] is modelled as:

$$\tau_g = \exp\left[-0.0127 \cdot m_a^{0.26}\right] \tag{6.11}$$

Transmittance by water vapour τ_w is estimated as:

$$\tau_w = 1.0 - \frac{2.4959\, w\, m_r}{(1.0 + 79.034\, w\, m_r)^{0.6828} + 6.385\, w\, m_r} \tag{6.12}$$

where w [cm] is precipitable water in cm calculated according to (117).

Finally, the transmittance by aerosols τ_a [-] is evaluated as:

$$\tau_a = \left[0.97 - 1.265 \cdot V^{-0.66}\right]^{ma^{0.9}} \tag{6.13}$$

where V [km] is the visibility, i.e. an estimation of the visibility extent as in (30).

6.1.2 Diffuse solar radiation under cloudless sky conditions

The diffuse component of solar radiation, $d\downarrow$ is modelled as (72):

$$d\downarrow = (d\downarrow_r + d\downarrow_a + d\downarrow_m) \cdot V_s \tag{6.14}$$

where $d\downarrow_r$, $d\downarrow_a$ and $d\downarrow_m$ are the diffuse irradiance components after the first pass through the atmosphere due to the Rayleigh-scattering, the aerosol-scattering, and multiple-reflection respectively.

The Rayleigh-scattered diffuse irradiance is computed as:

$$d\downarrow_r = \frac{0.79 \cdot cos(\theta_z) \cdot I_{sc} \cdot E_0 \cdot \tau_o \cdot \tau_g \cdot \tau_w \cdot \tau_{aa} \cdot (1 - \tau_r)}{2.0 \cdot (1.0 - m_a + m_a^{1.02})} \tag{6.15}$$

where τ_{aa} is the transmittance of direct radiation due to aerosol absorptance modelled as:

$$\tau_{aa} = 1.0 - (1 - \omega_0) \cdot (1 - m_a + m_a^{1.06}) \cdot (1.0 - \tau_a) \tag{6.16}$$

where $\omega_0 = 0.9$ [-] is the single-scattering albedo fraction of incident energy scattered to total attenuation by aerosols ((68)).

The aerosol-scattered diffuse irradiance component is defined as:

$$d\downarrow_a = \frac{0.79 \cdot I_{sc} \cdot cos(\theta_z) \cdot E_0 \cdot \tau_o \cdot \tau_g \cdot \tau_w \cdot \tau_{aa} \cdot F_c \cdot (1 - \tau_{as})}{1 - m_a + m_a^{1.02}} \tag{6.17}$$

where $\tau_{as} := \tau_a \tau_{aa}^{-1}$ and F_c is the fraction of forward scattering to total scattering ($F_s = 0.84$ if no information about the aerosols are available, (72)).

The diffuse irradiance from multiple reflections between the earth and the atmosphere is computed as:

$$d\downarrow_m = \frac{\left(R\downarrow sw + d\downarrow_r + d\downarrow_a\right) \cdot \alpha_g \cdot \alpha_a}{1.0 - \alpha_g \cdot \alpha_a} \tag{6.18}$$

where α_g is the albedo of the ground and α_a is the albedo of the cloudless sky computed as:

$$\alpha_a = 0.0685 + (1.0 - F_c) \cdot (1 - \tau_{as}) \tag{6.19}$$

Finally, V_s is the sky view factor, i.e. the fraction of sky visible at a point, computed using the algorithm presented in (30).

6.2 DEC-MOD's: The shortwave radiation correction for cloudy sky

The radiation components presented in the previous subsections are computed under the assumption of cloudless sky conditions. To account for the presence of clouds some models were developed, denominated decomposition models. The procedure described in this thesis is in line with (67). It corrects the clear sky direct and diffuse irradiance by means of adjustment coefficients and the clear sky irradiances so that, for any point:

$$S\downarrow^* := c_s \cdot S\downarrow \tag{6.20}$$

is the corrected irradiance for direct shortwave radiation (and c_s is the correction coefficient for $S\downarrow$) , and

$$d\downarrow^* := c_d \cdot d\downarrow \tag{6.21}$$

is the corrected irradiance for the diffuse shortwave radiation (and c_d is the correction coefficient for $d\downarrow$). The reduction coefficients depend upon the global shortwave irradiance measured at

the available stations as suggested in (112), (46) and (124). For any station, i:

$$\hat{R}_{sw}\downarrow_i = S^*\downarrow_i + d^*\downarrow_i \tag{6.22}$$

$$d^*\downarrow_i = (k_d)_i \hat{R}_{sw}\downarrow_i \tag{6.23}$$

where k_d is the ratio between the diffuse shortwave irradiance and the shortwave total irradiance. Therefore, at stations:

$$(c_d)_i = \frac{\hat{R}_{sw}\downarrow_i \cdot k_d}{d\downarrow_i} \tag{6.24}$$

and

$$(c_s)_i = \frac{\hat{R}_{sw}\downarrow_i \cdot (1-(k_d)_i)}{S\downarrow_i} \tag{6.25}$$

Clearly k_d becomes the key parameter for the stations, in estimating the cloudy irradiances.Three different parameterisations are proposed.

- (46) estimated k_d for latitudes between 31 and 42 degrees North, using hourly data from five irradiances measurement stations within the USA:

$$k_d = \begin{cases} 1.0 - 0.09\,k_t & if \quad k_t \leq 0.22 \\ 0.951 - 0.1604 k_t + 4.388\,k_t^2 + & \\ -16.638\,k_t^3 + 12.336\,k_t^4 & if \quad 0.22 < k_t \leq 0.80 \\ 0.165 & if \quad k_t > 0.80 \end{cases} \tag{6.26}$$

- (124) estimated the diffuse fraction k_d known k_t using measured data in the USA and Europe (latitude between 28-60 degrees North.) and developed this relations:

$$k_d = \begin{cases} 1.02 - 0.248 \cdot k_t & if \quad k_t \leq 0.30 \\ 1.45 - 1.67 k_t & if \quad 0.30 < k_t \leq 0.78 \\ 0.147 & if \quad k_t > 0.78 \end{cases} \tag{6.27}$$

- using data from Victoria, Australia, (15) provided the exponential relation:

$$k_d = \frac{1.0}{1.0 + e^{7.997(k_t - 0.586)}} \tag{6.28}$$

Equation (9.18) above is completely driven wiht the knowledge of the clearness sky index, k_t [-], which is defined as:

$$k_t := \frac{\hat{R}_{sw}\downarrow}{\hat{I}_{sc} \cdot E_0 \cdot cos(\theta_s)} \tag{6.29}$$

Using the above equations a set of adjustment coefficients, c_s and c_d for beam and diffuse radiation component are obtained for the measurements station. To extended this to any spatial point, extrapolation is required to all the points of interest, where incoming shortwave

solar radiation is not measured. This is accomplished in NewAGE-DEC-MOD's component by using the NewAge-JGrass Kriging component (51) with simple kriging algorithm (59).

6.3 Applications

The SwRB component estimates for any point of a basin the incoming radiation. It does not require any calibration, once the four parameters in tab.(6.1) are assigned according to literature values. These outputs, however, do not correspond to a measured quantity but to an intermediate step of the calculations. Only the data from the DEC-MOD's component corresponds to measured quantities, and, this component uses the measured quantity to estimate the attenuation coefficients. Therefore, to allow some validation, we divided any of the group of measurements stations into two subgroups: one used for the estimation of the coefficients, (say C-set), and the other one for the verification of the results, (say V-set). Stations used for verification are shown in bold font in tab.(6.2),(6.3) and (6.4). More complex verification strategies could be used, as described in the discussion section, but their application is beyond the scope of the presented work.

Therefore, for any of the three basins the following components are applied:

- the SwRB against a subset of the measurement stations. The result for this step is the computation of the clear sky surface shortwave radiation. Inputs and outputs of the model are reported in fig.(6.1). The main parameters values used in the simulations are reported in tab.(6.1) according to (72) and (30);
- the DEC-MOD component as in the previous section and estimation of the coefficients c_s and c_d. Inputs and outputs of the model are reported in fig.(6.1);
- the ordinary Kriging component (51) to extrapolate the coefficients c_s and c_d for the set of stations left for verification (shown bold in tab.(6.2),(6.3) and (6.4));
- an estimate of the shortwave incoming solar radiation under generic sky condition in the V-set (SwRB-Allsky, which multiplies the SwRB output by the kriging output for the V-Set stations);
- the verification component NewAGE-V (51) to evaluate the performance of the model.

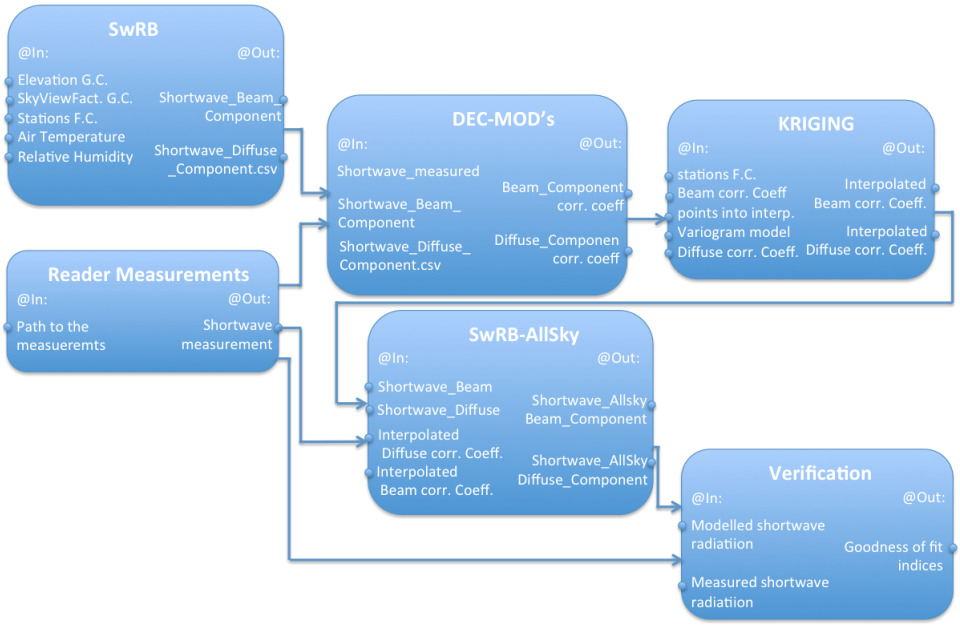

Figure 6.1: OMS3 SWRB components of NewAge-JGrass and the flowchart to model shortwave radiation at the terrain surface with generic sky conditions. Where not specified, quantity for input or output must be a spatial field for any instant of simulation time. "Measured" refers to a quantity that is measured at a meteorological station. Geomorphic features refer to the hilllslope and channel delineation, slope and aspect. The components, besides the specfied files received in input, include an appropriate set of parameter values.

For the simulations of this thesis, the Erbs model was used in the case of Piave river basin and Reindl model was used for Little Washita and Fort Cobb catchments.

For verification we used three performances indices:

- mean absolute error (MAE):

$$MAE = \frac{1}{N} \cdot \sum_{i}^{N} |S_i - O_i| \tag{6.30}$$

where N is the number of records of the time-series, O are the observed values and S are the simulated values. MAE is expressed in the same units of O and S, and is zero for perfect agreement between observations and estimates.

- Percentual bias (PBIAS):

$$PBIAS = 100 \cdot \frac{\sum_i^N (S_i - O_i)}{\sum_i^N O_i} \tag{6.31}$$

PBIAS measures the average tendency of the simulated values to be larger or smaller than their observed ones. The optimal value of PBIAS is zero, with low-magnitude values indicating accurate model simulation

- Kling Gupta Efficiency (KGE) as reported in (63):

$$KGE = 1 - \sqrt{(R-1)^2 + (A-1)^2 + (B-1)^2} \tag{6.32}$$

in which R represents the linear correlation coefficient between the simulated (S) and measured (O) values, A and B are respectively expressed in equations (9.17) and (9.18):

$$A = \frac{\sigma_o}{\sigma_s} \tag{6.33}$$

where σ_o is the observed standard deviation value and σ_s is the simulated standard deviation;

$$B = \frac{\mu_s - \mu_o}{\sigma_o} \tag{6.34}$$

where μ_s and μ_o are the means of simulated (S) and measured (O) values. For this index, the best agreement is obtained with the value 1.

The Kriging package can utilise the most common variogram models (spheric, linear, exponential, and gaussian). However, for these cases below, a linear model was used.

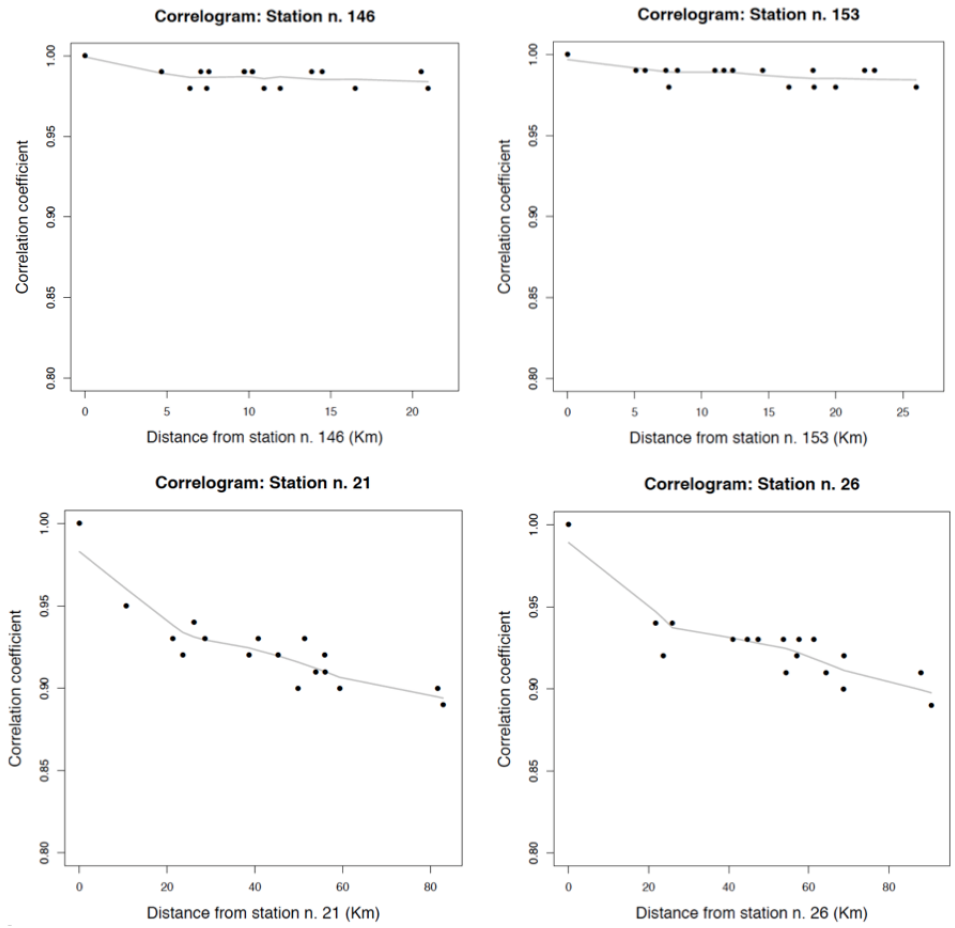

Figure 6.2: Correlogram between station 146 and 159 of the Little Washita river basin, at the top. Correlogram for station 21 and 26 of the Piave river basin, at the bottom.

Table 6.2: List of the meteorological stations used in the simulations performed on Little Washita river basin. Bold font is used for indicating the stations belonging to the validation set.

ID	City	LAT.	LONG.	Elevation (m)	Aspect (°)
124	**Norge**	**34.9728**	**-98.0581**	**387.0**	**138°**
131	Cyril	34.9503	-98.2336	458.0	245°
133	Cement	34.9492	-98.1281	430.0	116°
134	Cement	34.9367	-98.0753	384.0	65°
135	Cement	34.9272	-98.0197	366.0	182°
136	Ninnekah	34.9278	-97.9656	343.0	270°
144	Agawam	34.8789	-97.9172	388.0	50°
146	Agawam	34.8853	-98.0231	358.0	212°
148	**Cement**	**34.8992**	**-98.1281**	**431.0**	**160°**
149	Cyril	34.8983	-98.1808	420.0	205°
150	**Cyril**	**34.9061**	**-98.2511**	**431.0**	**195°**
153	Cyril	34.8553	-98.2121	414.0	165°
154	Cyril	34.8553	-98.1369	393.0	175°
156	Agawam	34.8431	-97.9583	397.0	290°
159	Rush Springs	34.7967	-97.9933	439.0	235°
162	Sterling	34.8075	-98.1414	405.0	15°
182	**Cement**	**34.845**	**-98.0731**	**370.0**	**245°**

Table 6.3: List of the meteorological stations used in the simulations performed on Fort Cobb river basin. Bold font is used for indicating the stations belonging to the validation set.

ID	City	LAT.	LONG.	Elevation (m)	Aspect (°)
101	**Hydro**	**35.4551**	**-98.6064**	**504.0**	**120°**
104	Colony	35.3923	-98.6233	484.0	35°
105	**Colony**	**35.4072**	**-98.571**	**493.0**	**300°**
106	Eakly	35.3915	-98.5138	472.0	295°
108	Eakly	35.3611	-98.5712	492.0	40°
109	**Eakly**	**35.3123**	**-98.5675**	**466.0**	**90°**
110	Eakly	35.3303	-98.5202	430.0	115°
113	Colony	35.291	-98.6357	465.0	155°

Table 6.4: List of the meteorological stations used in the simulations performed on Arabba river basin. Bold font is used for indicating the stations belonging to the validation set.

ID	City	LAT.	LONG.	Elevation (m)	Aspect (°)
1	Arabba	46.4999	11.8761	1825	180°
2	**Caprile**	**46.4404**	**11.9900**	**1025**	**170°**
3	Agordo	46.2780	12.0331	602	5°
8	Villanova	46.4433	12.2062	972	71°
9	**Auronzo**	**46.5562**	**12.4258**	**940**	**223°**
11	Campo di Zoldo	46.3466	12.1841	915	160°
12	Domegge di Cadore	46.4609	12.4103	802	148°
14	Monte Avena	46.0321	11.8271	761	55°
18	Passo Pordoi	46.4834	11.8224	357	55°
21	Passo Monte Croce	46.6521	12.4239	1612	120°
22	Col Indes	46.1191	12.4401	1119	210°
23	**Torch**	**46.1515**	**12.3629**	**602**	**177°**
26	Sappada	46.5706	12.7080	1275	156°
29	Feltre	46.0162	11.8946	273	190°
31	Falcade	46.3554	11.8694	1151	50°
32	Cortina	46.536	12.1273	1244	88°
35	Belluno	46.1643	12.2450	378	157°

6.3.1 Results and discussion

The models were applied on three different river basin: Little Washita, and Fort Cobb and Piave river basin. Their main hydrological features and datasets used are illustrated in the chapter 4. Results are presented separately for the three case studies. They confirm the results found in literature, and reveal a reasonable agreement between measured and simulated data.

Fig.(6.3) shows the scatter plot between the simulated and the measured total incoming solar radiation at the four stations of the V-set.

Tab.(6.5) shows the result of the NewAge-V which provides as output the user defined goodness of fit indexes.

For the Fort Cobb river, the same procedure presented for the Little Washita river basin was performed.

Fig.(6.4) shows the scatter plot between the modelled and the measured total incoming solar radiation at the four V-set stations. Tab.(6.6) shows the results in term of goodness of fit indices for the V-set.

For the Arabba river the same procedure is applied as presented for the Little Washita river basin. The decomposition model used in this case is (124). Fig.(6.5) shows the scatter plot of the modelled and the measured total incoming solar radiation in the four V-set. Tab.(6.7) show the results in term of goodness of fit indexes for the same set of stations.

The model was applied in various case studies with various characteristics in topography: a) two cases presented a gentle topography and a high density measurement network (Little

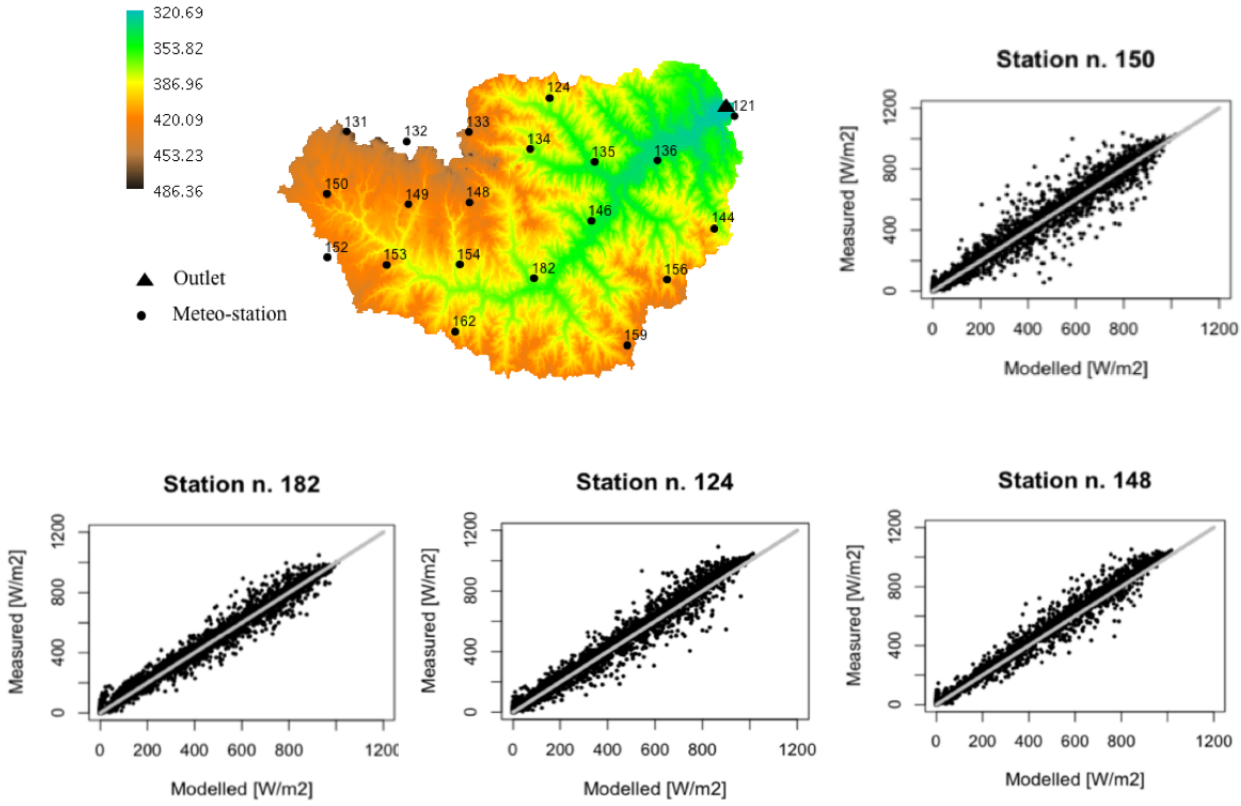

Figure 6.3: The Little Washita river basin, Oklahoma (U.S.A.).

Washita and Fort Cobb watersheds) b) the other case presented a typical hydrological basin with complex topography, high elevation range, and few monitoring stations.

In all cases the model was able to simulate the global shortwave solar radiation showing relatively good goodness of fit indices presented in tab.(6.5) and tab.(6.6) for Little Washita and Fort Cobb respectively and in tab.(6.7) for the Arabba river basin.

The model performs with similar and acceptable accuracy for the Little Washita and Fort Cobb river basin. The result is confirmed by the goodness of fit indices and by the graphical analysis.

The model performance degrades in the Arabba case study. This could be due to the effect of the complex topography on the computation of the clear sky solar radiation but also to the lower measurements stations density in high elevation zones.

Because of this topographic condition the increasing measurement data uncertainty of the temperature and humidity influenced the atmospheric transmittance computations. This is also confirmed by the data analysis: for the Piave river basin measurements shows lower correlation respect, for example, the correlation between measurements on the Little Washita river basin, where the gentle topography does not play a crucial rule.

The model, was able to reproduce well the shortwave solar radiation also for a watershed with complex topography. The PBIAS index was 14.80 in the worst case. According the hydrological

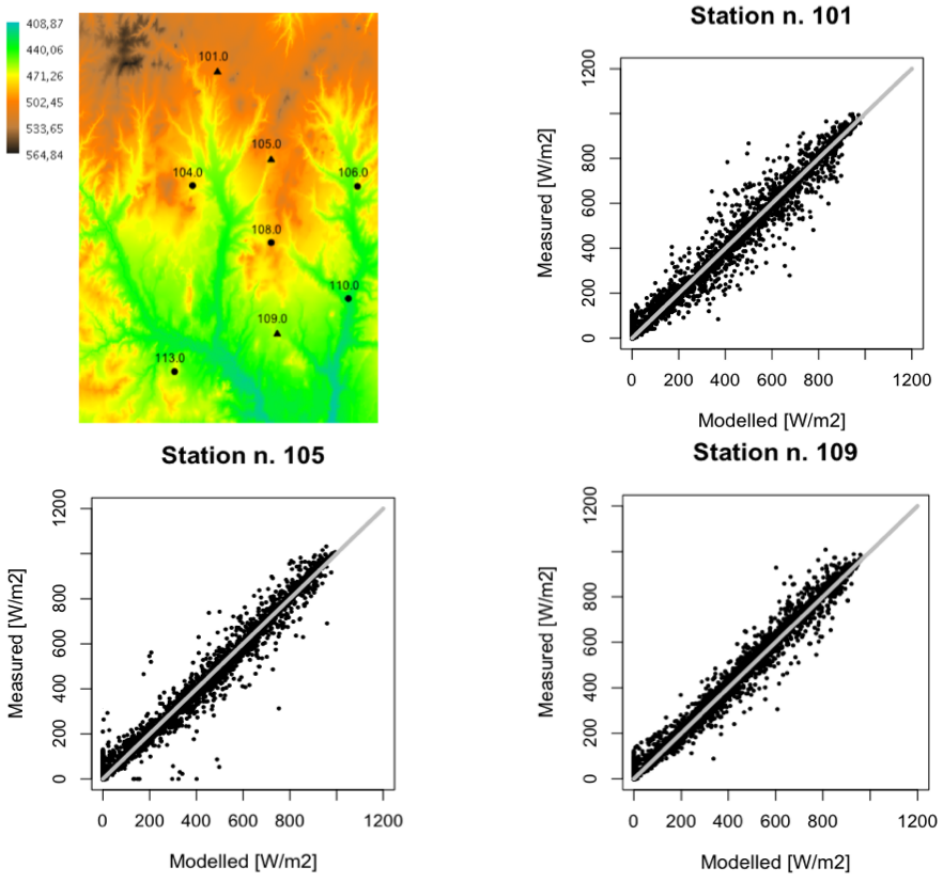

Figure 6.4: The Fort Cobb river basin results.

model classification based on the PBIAS index, (150) and (143), the results achieved in our study are classified as "good". Therefore the solar radiation model is suitable to be used to estimate incoming shortwave solar radiation.

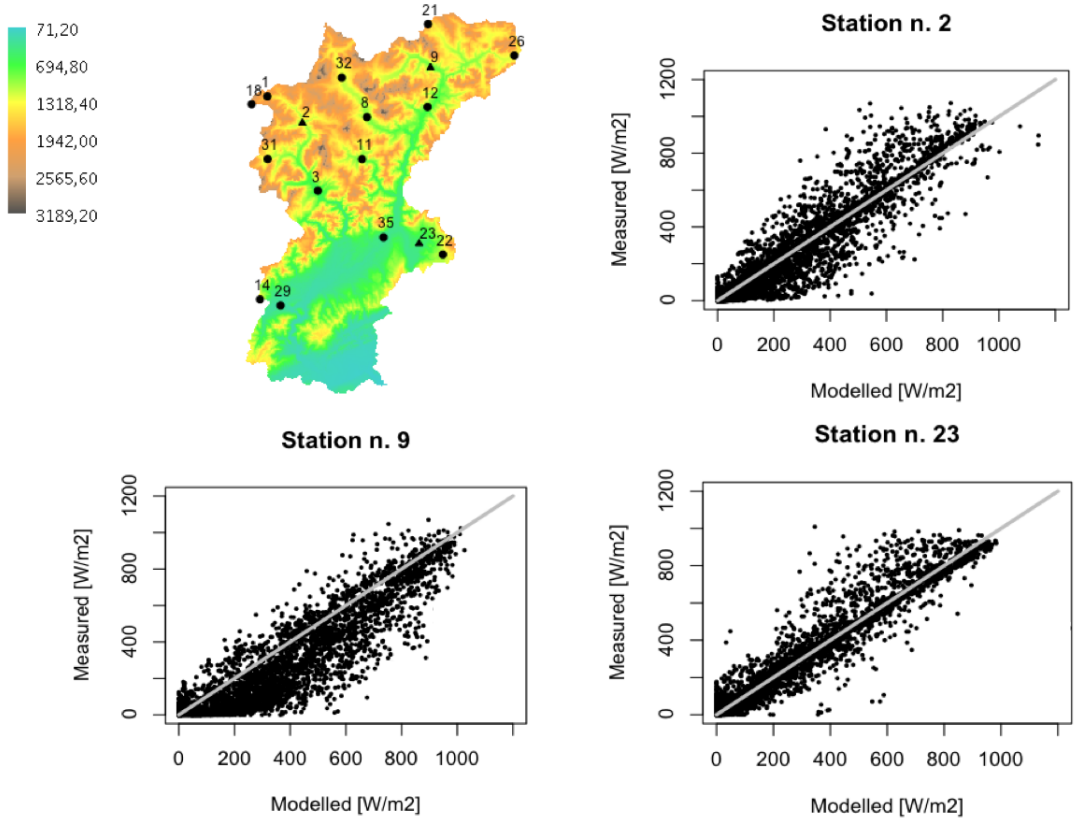

Figure 6.5: River Piave area, (Italy).

Table 6.5: Index of goodness of fit between modelled and measured solar radiation in Little Washita river basin.

STATION ID	KGE	MAE [W/m^2]	PBIAS [%]
148	0.94	16.65	4.90
124	0.95	17.50	3.80
182	0.98	16.50	1.80
150	0.97	17.90	2.10

Table 6.6: Index of goodness of fit between modelled and measured solar radiation in Fort Cobb river basin

STATION ID	KGE	MAE [W/m^2]	PBIAS [%]
101	0.96	15.6	5.5
105	0.95	13.50	2.80
109	0.97	14.07	2.70

Table 6.7: Index of goodness of fit between modelled and measured solar radiation in Arabba river basin

STATION ID	KGE	MAE	PBIAS
2	0.92	4.53	2.7
9	0.89	22.10	14.80
23	0.95	3.58	2.1

6.4 Longwave radiation component (LW-C)

The JGrass NewAge LW-Component estimates the net long wave solar radiation for a certain point under cloudy sky conditions. As longwave radiation is more difficult and expensive to measure than shortwave radiation, many efforts were made to estimate it climate variables such as air temperature and vapor pressur (106) or to model it (47), (145), (30). The net long wave radiation L is the algebraic sum between incoming or downwelling long wave radiation (L_d) and outgoing or upwelling long wave radiation (L_u).

Downwelling longwave radiation is dependent on the atmospheric emissivity and temperature. Because it is difficult to estimate these quantities, parameterizing the longwave downwelling radiation based upon near-surface measurements of temperature and/or vapor pressure was developed, (22), (1), etc. (23) presents the derivation of an equation to compute downward longwave radiation under clear skies and nearly standard atmospheric conditions:

$$L_d = 1.24 \cdot \left(\frac{10 \cdot e}{T}\right)^{\frac{1}{7}} \cdot \sigma \cdot T^4 \tag{6.35}$$

where L_d is expressed in $W \cdot m^{-2}$, T is the air temperature in K and e is the vapor pressure in kPa and σ is the Stephan Boltzman constant. The upwelling long wave radiation depends on the surface temperature T_s and surface emissivity ϵ_s:

$$L_u = \epsilon_s \cdot \sigma \cdot T_s^4 \tag{6.36}$$

If the surface temperature is not available, it can be replaced by the air temperature, (24). In order to extend this formula for generic sky conditions, the equation presented in (25) was implemented:

$$L = (1 + a \cdot c^b) \cdot (L_d + L_u) \tag{6.37}$$

where a and b are coefficients (b=1 in (25)) and c is the cloud cover.

6.5 Evapotranspiration

Evapotranspiration is the flux through which the water, in liquid phase, changes it phase and returns back into the atmosphere in the gas form. It includes two phenomena:

- evaporation: the process in which water is transferred from free surface water to the atmosphere;
- transpiration: process in which water evaporates from the air spaces in plant leaves through the stomata.

Quantification of actual evapotranspiration (ET) is a difficult and very important task for water resources management. In (7) it has been shown that ET returns about 64% of land precipitation to the atmosphere. As explained in (2) the main factors controlling the the evapotranspiration

process depends on are: climate parameters such as precipitation, radiation, and air humidity, crop characteristics such as species, age, height, roughness; management, and environmental aspects.

In order to study the evapotranspiration independently from crop species and management practices the concept of reference evapotranspiration (ET_0) was introduced (2) which depend on climate variables. ET_0 as referred to a specific surface, is computed and it provides a reference which can be related to ET of other surfaces. The crop evapotranspiration is then computed by multiplying ET_0 by a coefficient which takes into account the differences between field crops and the reference grass crop.

In the next subsection the NewAge-JGrass components for evapotranspiration modeling is presented.

6.5.1 The JGrass NewAge evapotranspiration component (ET-C).

The NewAge-JGrass ET-CTwo offers two different formulations for the evapotraspiration modeling: the FAO Penman-Monteith model (2), eq. (6.38), and the PriestleyTaylor model((119), (139), (120)), eq. (6.39).

$$ET_0 = \frac{0.408 \cdot \Delta \cdot (R_n - G) + \gamma \cdot u_2 \cdot (e_s - e) \cdot \frac{C_p}{T+273}}{\Delta + \gamma \cdot (1 + C_d \cdot u_2)} \tag{6.38}$$

$$ET = \alpha \cdot \frac{\Delta \cdot (R_n - G)}{\Delta + \gamma} \tag{6.39}$$

where ET_0 or ET are expressed in $mm \cdot day^{-1}$ or $mm \cdot hour^{-1}$; R_n is the net radiation expressed in $MJ \cdot m^{-2} \cdot day^{-1}$ or $MJ \cdot m^{-2} \cdot h^{-1}$; G is the soil heat flux at the soil surface $MJ \cdot m^{-2} \cdot day^{-1}$ or $MJ \cdot m^{-2} \cdot h^{-1}$. It is considered zero at the daily time step and it is considered a fraction of the net radiation at the hourly time step as proposed in (2); T is the mean daily or hourly air temperature expressed in $^{\circ}C$; e_s is the mean saturation vapor-pressure expressed in kPa; e is the mean actual vapor-pressure; Δ is the slope of the saturation vapor-pressure curve expressed in $kPa \cdot^{\circ} C^{-1}$; γ is the psychometric constant expressed in $kPa \cdot^{\circ} C^{-1}$; C_d is a coefficient equal to 0.34 and C_p is a coefficient equal to 900 in the case of a daily time step and equal to 37 in the case of a hourly time step.

The NewAge-JGrass ET-C component uses the shortwave and long wave energy component in order to estimate the net radiation and the meteorological interpolation algorithms in order to estimate the climate variable if it is necessary. The quantity e, e_s, γ, and Δ are computed according to (2).

7

NewAge-JGrass Rainfall Runoff model

This chapter focuses on the hydrological budgets for modelling of medium scale to large scale basins as the result of the processes at the hillslope scale with the interaction of the river network. The part of the modeling system presented here deals with the: (i) estimation of the space-time structure of precipitation, (ii) estimation of runoff production, (iii) aggregation and propagation of flows in the channel, (v) estimation of evapotranspiration, and (vi) the automatic calibration of the discharge with the particle swarming. The semi-distributed rainfall runoff-production and channel routing components are presented and verified in two river basins: Fort Cobb and Little Washita. The second applications presents a comparison between two modeling solutions: with and without channel routing model. Finally a comparison of the model with the SWAT model is presented.

7.1 Preliminary analysis

The first step for applying the NewAge-JGrass runoff production and channel routing components is the basin delineation. As presented in Chapter 4, the model partitions the basin into hillslopes and channels (passed to the model as hillslope-link structure), where the hillslopes are the basic hydrological units. At this scale that the energy and water mass budgets statistics are estimated. The channels are represented as vector elements (features) that are topologically interconnected in a simple directed graph. This concept could be mistaken with the concept of hydrological runoff units (HRUs) promoted in (135), (48), and used, for instance in (80), and in (152).

Thus HRUs can be seen as sub-partitions of the hillslope, and in NewAge-JGrass these sub-classes provide statistical models at the hillslope (or small watershed) level, rather than single estimates of the hydrological quantities. For computational reasons, the partitioning of the area is not usually designed to identify all the physical hillslopes present in the system, but to define

small watersheds with dimensions, in the current application, of 5-20 Km^2 on average. HRUs can be either treated as vector feature, or raster, according to convenience.

The second step is to implement suitable algorithms for spatially distributing the precipitation measured at ground level at meteorological station sites. The hydrological budgets estimation can be inaccurate if the atmospheric forcings (i.e., spatial field time series of precipitation, air temperature, and solar and thermal radiation) are not properly accounted for. These variables have particular characteristics and levels of data availability that make it necessary to use a variety of procedures to develop spatial fields for each. However, to keep the modeling chain simple, we limit the present investigation to the use of simple kriging (59), and a detrended kriging for the estimation of precipitation and for interpolating air temperatures. An internal component of the Kriging module provides, for each time step, the best theoretical semivariogram model able to fit the experimental one (Gaussian, Exponential, Spheric and Linear models (see Chapter 5)).

We used these techniques mainly because they can easily account for topographic and other features as discussed for instance in (73), where it was argued that modelers should be aware of the influence of many topographic characteristics besides elevation. Also, (56) provides an insightful discussion of this topic with particular focus of the modeling of snow, which we do not replicate here.

Solar radiation was estimated implementing (31) models which allow for simulating the incoming shortwave radiation according to topography inclination and aspect and estimating shadowing and the angle of view, which lessens the portion of sky visible from any point. Therefore, the amount of radiation received in a valley is reduced. Long wave radiation was estimated using air temperature as a proxy for the terrain and canopy temperatures, and using the (23) and (24) parametrizations which revealed to be effective in other studies e.g (126).

7.2 Runoff generation

The third step was the selection of the runoff generation mechanism. While a more physically based choice, built for instance on an estimation of the flow paths structure derived from a detailed digital elevation model, was preferable, in this worked it is opted to use a standard model already published in literature. First, a modelling solution (42) was looked at for adaptation. However, we later had to accept that its conceptual simplicity did not allow for to an easy of application, since the parameters' range provided in (42) was not extensible to catchments with different soil types and soil hydraulic properties. Therefore, once the parameters that were constant in the original paper and the tunable parameters were considered all together in this case, calibrating them became a computationally overwhelming task. Hence, we decided to use the Hymod model (104) and (18), which is outlined below. The Hymod runoff component has only five parameters which can be automatically calibrated.

The Hymod model was not used for the whole catchment, providing for only water storage and the streamflow for each time-step for any single hillslope. The rationale of using several Hymods instances, one for each hillslope, was twofold: firstly, to preserve the geometrical and topological structure of the river network, which proved to embed significant information about the shape of discharge hydrograph, (38), and secondly, to allow the use as input of spatially varying rainfall and evapotranspiration fields.

Hymod is based on the ideas presented in (104) and (16) and consists of two main parts: a non-linear component that partitions precipitation into precipitation excess, and two series of linear routing reservoirs that models quick and slow flow. The original model uses one linear reservoir that models the slow flow component and a series of three identical linear reservoirs modelling the quick flow.

The conceptualization in Hymod considers a catchment in which water storage capacity is partially filled up with water (as in fig.(7.1a)) to C_{max}(L), until it reaches the maximum water storage capacity. The water storage capacity between different points varies and it is assumed to be represented by the reflected power distribution function F(C):

$$F(C) = 1 - \left(1 - \frac{C}{C_{max}}\right)^{B_{exp}} \tag{7.1}$$

in which C [L], $0 \leq C \leq C_{max}$, is the water storage capacity (104), C_{max} [L] is the maximum value of the water storage capacity of the basin and B_{exp} accounts for the degree of spatial variability in the water storage capacities as modeled in the Arno (148) model, which uses a different mechanism for separating slow and quick flows.

The precipitation P [L] that falls and exceeds C_{max} (as in fig.(7.1b)) directly flows along the quick flow paths, into the river.

The precipitation that exceeds the water storage capacity C of points with a lower capacity than C_{max} (as in fig.(7.1c)) is instead divided into quick and slow flows according to a partition parameter Alpha. Finally, some water evaporates according to the water stored in the slow reservoirs and the potential evapotranspiration, given by an external model.

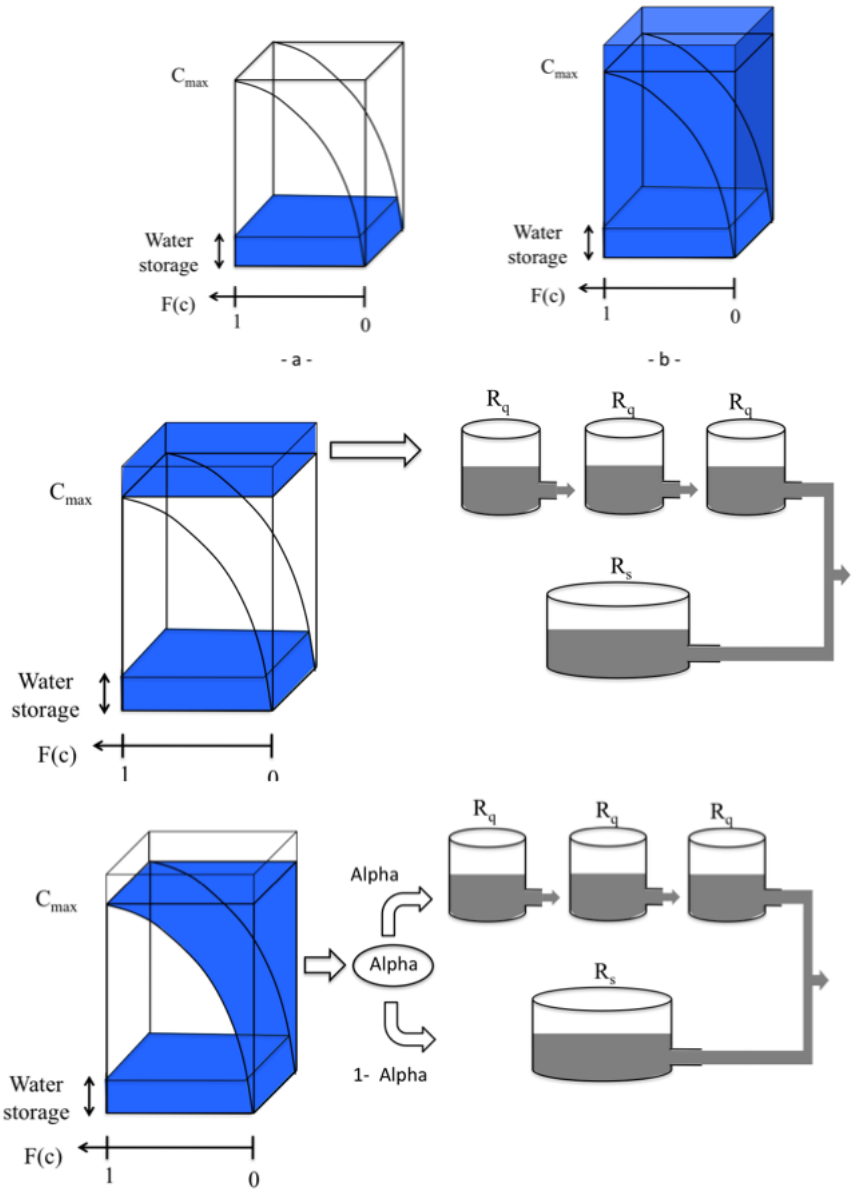

Figure 7.1: Runoff generation in NewAge-JGrass System. (Top) On the left a representation of partially filled reservoirs; on the right the case of precipitation exceeding the storage capacity. (Center) When the total storage is exceeded, the precipitation excess is directly routed as overland flow by using three linear reservoirs. (Bottom) For precipitation not exceeding C_{max} the volume of precipitation above the curve is divided into overland flow and subsurface flow according to a coefficient of partition Alpha.

7.3 Flow Routing

The flow generation model along hillslopes delivers discharge to the channel network, which is conceptualized in the model as an directed tree graph, and is kinematically propagated downstream through a simplified model derived from the CUENCAS model (93). It is a non linear variant of the Saint Venant equation e.g. (19) integrated in each channel link.

The resulting system of equations allows an estimate of the varying discharge value in each link of the river network, with flow velocities varying with stage and positions. For each link the continuity equation, as presented in (94), is:

$$\frac{dS_i(t)}{dt} = \Big[Q_{gen}(t) + \sum_{trib} Q_{trib}(t) - Q_i(t)\Big] \qquad i = 1, 2, ..., H \tag{7.2}$$

where is the $S_i(t)$ is storage in the i-th link at time t, H is the total number of network links, $Q_i(t)$ [L^3 T^{-1}] is the output discharge from i-th link, Q_{trib} [L^3 T^{-1}] is the flow of upstream links, and $Q_{gen}(t)$ [L^3 T^{-1}] is the discharge generated at the hillslope of the link in question. Under the hypothesis that the link has a rectangular cross-section, so that the width, w, does not change in time. The channel storage and the discharge can be expressed as:

$$S_i(t) = l_i \cdot w_i \cdot d_i(t) \tag{7.3}$$

and:

$$Q_i(t) = v_i(t) \cdot w_i(t) \cdot d_i(t) \tag{7.4}$$

where $v_i(t)$ [L T^{-1}] is the flow velocity, $w_i(t)$ [L] is the mean width of the link, $d_i(t)$ [L] is the mean channel depth and l_t [L] is the link length.

Combining the equations (7.3) and (7.4) gives $S_i(t)$ in function of $Q_i(t)$. Finally, using the Chezy equation:

$$v = C \cdot R^{0.5} \cdot i_b^{0.5} \tag{7.5}$$

where v [L T^{-1}] is the mean velocity , C [$L^{0.5}T^{-1}$] is the Chézy coefficient , R [L] is the hydraulic radius, and i_b [-] is the bottom slope, $S_i(t)$ can be expressed as:

$$S(t) = Q(t)^{\frac{2}{3}} \cdot C^{-\frac{2}{3}} \cdot w^{\frac{1}{3}} \cdot l \cdot i_b^{-\frac{1}{3}} \tag{7.6}$$

The left hand side of the eq. (7.9) is expressed by the derivative of the eq. 7.6. After some

algebra, eq. 7.7 gives the non-linear ordinary differential equation in the unknown $Q_i(t)$:

$$\frac{dQ_i(t)}{dt} = K\Big(Q_i(t)\Big) \cdot \Big[Q_{gen}(t) + \sum_{trib} Q_{trib}(t) - Q_i(t)\Big] \qquad i = 1, 2, ..., H \tag{7.7}$$

The coefficient $K\Big(Q_i(t)\Big)$ is equal to:

$$K_Q = \frac{3}{2} \cdot Q^{\frac{1}{3}} \cdot C^{\frac{2}{3}} \cdot b^{-\frac{1}{3}} \cdot l^{-1} \cdot i_b^{\frac{1}{3}} \tag{7.8}$$

where C [$L^{1/3}$ T^{-1}] is the Chezy coefficient, b [L] and l [L] represent the width and average length of the link respectively, i_b [-] is the average slope of the link, and Q [L^3 T^{-1}] is the channel discharge. For a more detailed discussion of the terms in eq. (7.8) see (101), and (93) which provide also a description of how the parameters can be estimated by using geomorphic information.

7.4 An Application to Little Washita (OK, USA) river basin

To test the capabilities of the NewAge-JGrass system, we applied it to the Little Washita river basin. Two applications are presented in this thesis. In Test A an application of the runoff generation component Hymod is performed for the whole basin. The input precipitation time series is the spatial mean of all the measurements and the evaporation time series represents the global mean potential evapotranspiration. This configuration can be seen as the null hypothesis against the other model setups are tested.

For the second application (Test B) the Little Washita river basin is divided into 75 sub-basins and Hymod is executed for each of them, with its own estimate of evapotranspiration and rainfall by preserving the total volumes of the quantities for comparison with the lumped case. Furthermore, the generated discharge of each hillslope is routed and the parameters were evaluated according to the network geomorphology (93).

In both Test A and Test B the automatic calibration was performed at the outlet of the basin and the simulation period ranged from 01/01/2002 to 31/12/2003. The first year was used for the calibration, while the second year was used for the validation of the results.

The number of "particles" used in the calibration process was 50 and the algorithm parameter values were set according to literature hints (43) as: ω=0.6, c_1=1.8, c_2=2.0, s_1 and s_2 randomly distributed between 0 and 1.

Tab.(7.1) shows the parameter values obtained by the calibration component for Test A and Test B, respectively. The top line of tab.(7.2) shows the values of the test of fit obtained by the calibration for the Test A and the bottom line of tab.(7.2) shows the values of test of fit of Test B.

Table 7.1: Parameter values used in simulation

	C_{max}	B_{exp}	$alpha$	R_s	R_q
Test A	603.45	0.31	0.35	0.0098	0.13
Test B	572.52	0.85	0.39	0.001	0.12

Table 7.2: Index of goodness of fit for calibration and validation period

		IOA	RMSE	PBIAS
Test A	Calibration	0.76	0.96	18.7
Test A	Validation	0.71	1.06	24.8
Test B	Calibration	0.88	0.76	3.5
Test B	Validation	0.81	0.80	5.8

The simulation results of Test A and Test B are presented in fig.(7.2) and fig.(7.3). The gray dots represent the measured discharge and the black solid line represents the simulated discharge. These results will be discussed in the next section.

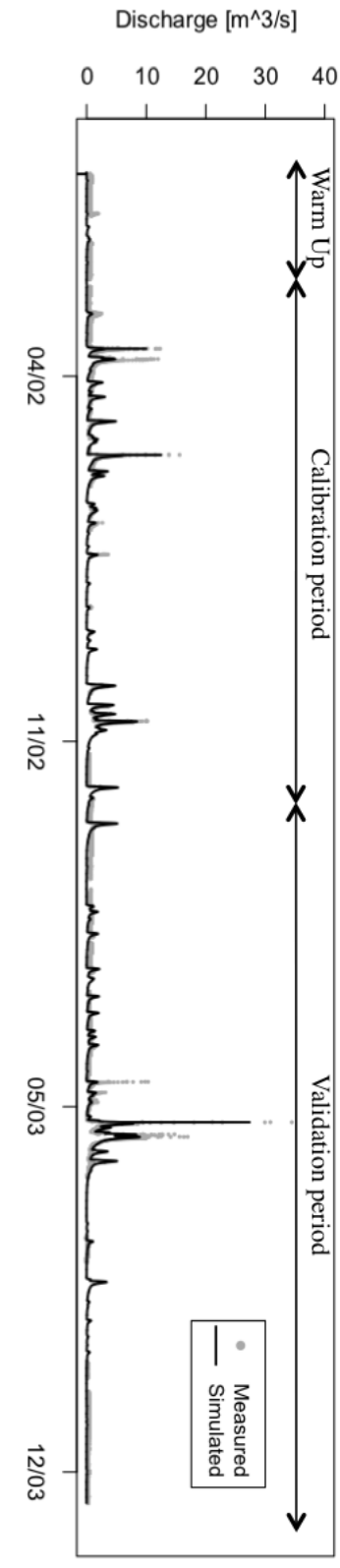

Figure 7.2: Test A: application of the NewAge-JGrass model for the period 01/01/2002 to 31/12/2003: the solid curve represents the simulated discharge, while the dots represents the measured one.

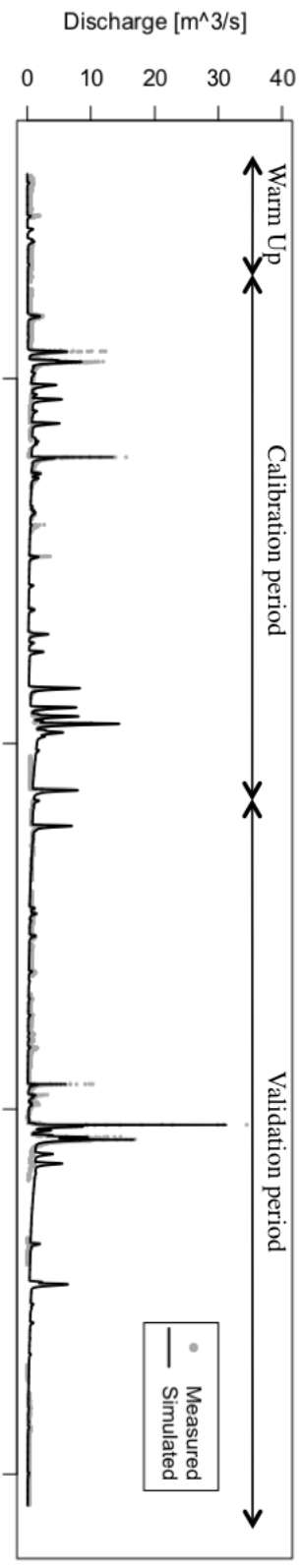

Figure 7.3: Test B: application of the NewAge-JGrass model for the period 01/01/2002 to 31/12/2003: the solid curve represents the simulated discharge, while the dots represents the measured one.

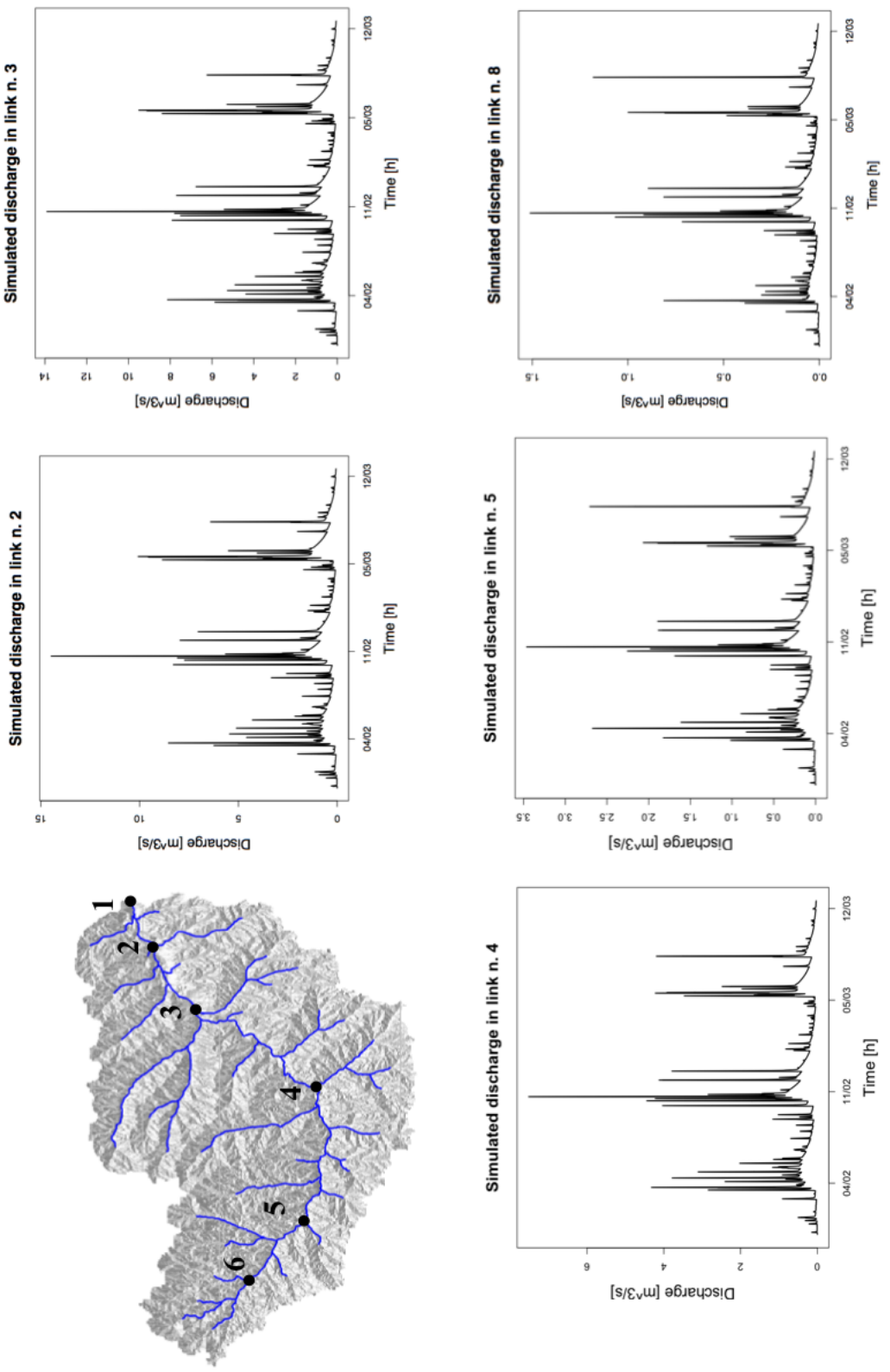

Figure 7.4: Application of the NewAge-JGrass model for the period 01/01/2002 to 31/12/2003.

7.4.1 Results

The visual inspection of simulated hydrographs provided by Test A and Test B show an acceptable agreement with the measured ones. However, an apparent tendency towards under prediction of flow peak values is evident in both tests, even though it is more prominent in Test A. While the largest peaks are usually underestimated, the secondary peak flows are sometimes overestimated. The underestimation for the largest peak is around the 0.32% in the case of Test A and around the 0.10% in the case of Test B. To make this more clear, single events where the same behavior is evident are plotted in fig. (7.5) and (7.6). The indices of goodness are reported in tab.(7.3). The recession curves are usually well reproduced with an apparent tendency of underestimation.

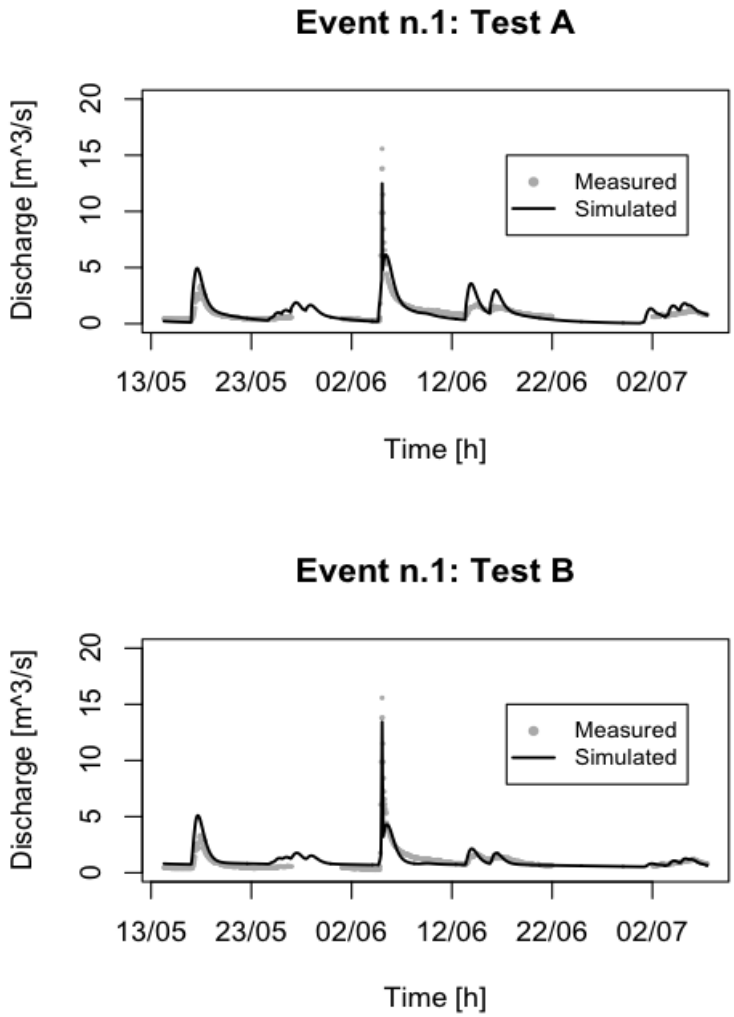

Figure 7.5: Event No.1: test case A at the top and test case B at the bottom. The year of the events registered is 2003.

The values of all three indices of goodness confirm the suspicion that came from the visual inspection. In Test B all the values are significantly better performing than in Test A. It can be observed that the values of the parameters obtained for the test case are similar, with better performances for the test B, greater than 10% for IOA and 20% for the RMSE. However, the

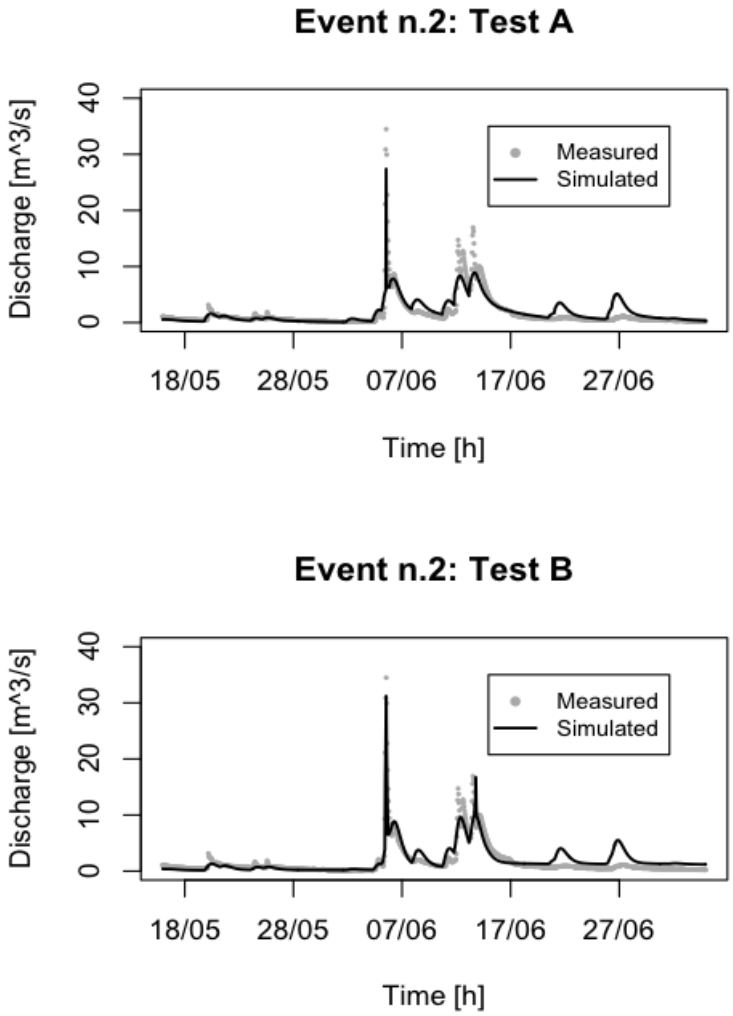

Figure 7.6: Event No.2: test case A at the top and test case B at the bottom.

the PBIAS greatly differentiates the better performances of test B, since test B has a very good performance. Therefore, all the test indices suggest that using the spatial information available and the increased complexity of the distributed model it is obvious to achieve a significantly better forecasting, at least for the case presented. The result is much more significant since both Test A and Test B were obtained not only with equal meteorological forcings but with the forcings computed by the same code components plugged with the core model assembly at run-time.

The residual plot, for test case B (in fig.(7.7)), gives a different view of the errors. They are as large as ten cubic meter per second, which is a quite large fraction of the peak discharge, and larger than the differences of the peaks of corresponding events because the simulated and measured peaks are recorded at slightly different times. Therefore, when the simulated peak occurs the measured discharge is already in the recession limb.

The analysis of the histogram of the residuals presented in fig.(7.8) for the validation period shows an almost normal distribution of the residual with the mean close to zero (0.0102 m^3/s) and standard deviation less then 0.71 m^3/s. The low value of the bias confirms the goodness of

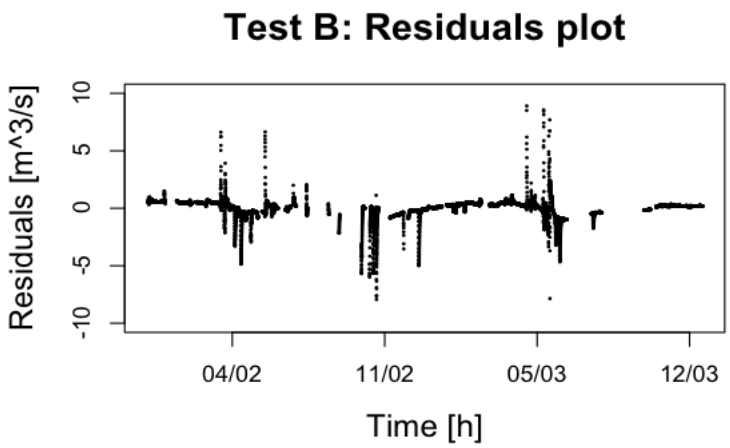

Figure 7.7: Test B: plot of the residuals for the validation period. Large values are usually due to time shifts between the measured and simulated discharge.

Table 7.3: Test B: indices of goodness of fit for the calibration and validation periods

	IOA	NSE
Event n.1: Test A	0.86	0.50
Event n.1: Test B	0.92	0.64
Event n.2: Test A	0.93	0.78
Event n.2: Test B	0.87	0.65

the calibration procedure and the goodness of the model assembly with respect of the simulations of the hydrological behavior of the basin analysed.

Estimations of inner values of the discharge in the basin have been provided in fig.(7.4) in order to visualize this capability of the model. However, these estimations cannot be compared against measured data and have only a demonstration value. From the figure is visible that: i) discharge decreases with contributing areas closely linear for major events (where the precipitation affects the whole catchment); ii) during minor events the local distribution of rainfall can produce uneven behaviors.

In many cases the error in forecasting is small and even if more accurate studies of this aspect should be necessary. It can be stated that they are contained within an confidence interval depending on the uncertainty (e.g. (154)) that can be due to inaccurate estimation of local precipitation and of relation between the storage and discharge.

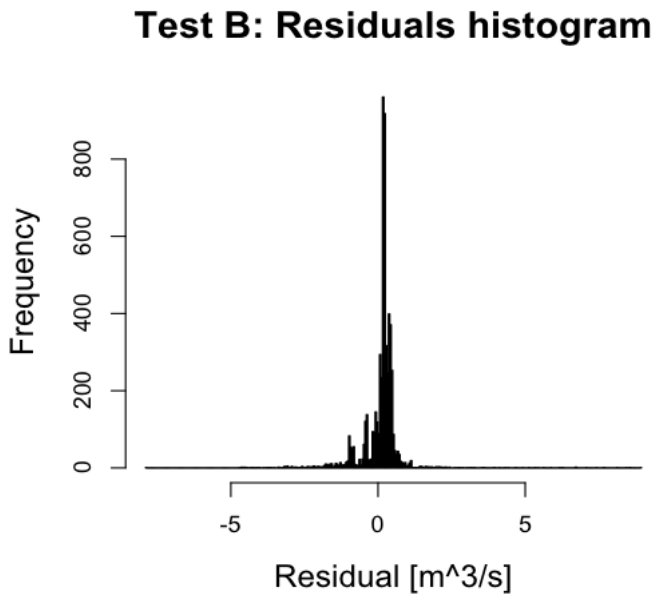

Figure 7.8: Test B: histogram of the residuals of the simulated discharge with respect to the measured ones.

7.5 Experimenting different modeling solutions.

The Hymod component is applied for each HRU and the runoff production is then propagated in the channel network. A new runoff propagation components is implemented and presented in the next subsection. To study the role and the importance of the channel routing component a test is performed. Two river basins are used for the test and modeled in a three different delineations by using one (DL1), three (DL3) and twenty (DL20) HRU's. Two modeling solutions were set up: Hymod and RHymod in fig.(7.9).

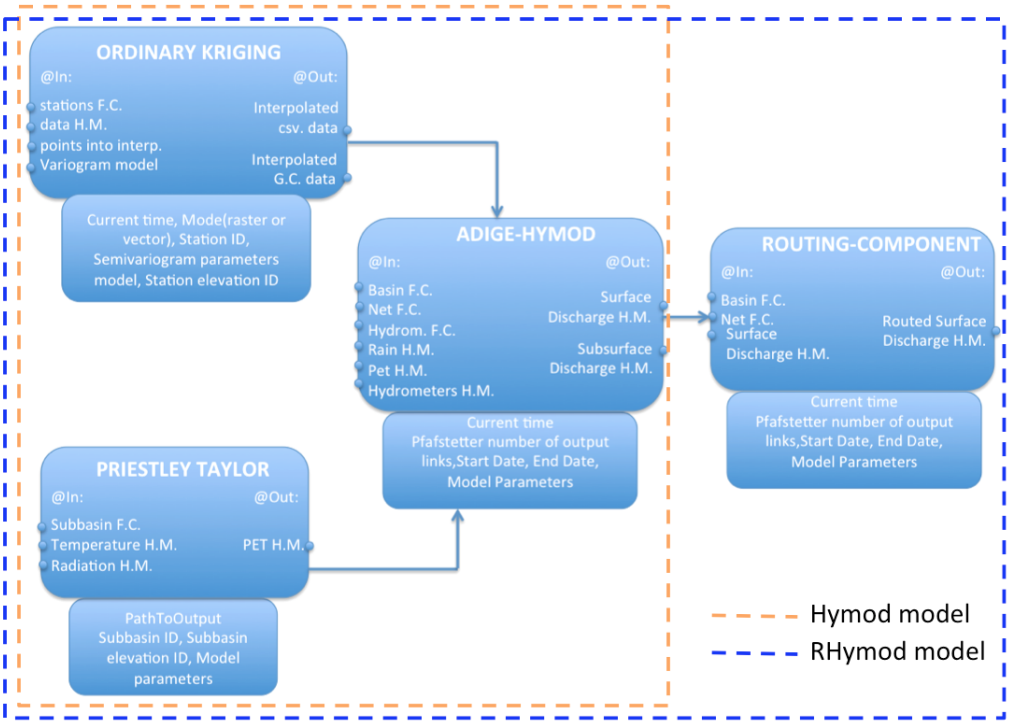

Figure 7.9: Modelling solutions: Hymod (in red dashed line) and RHymod (in blued dashed line).

The modeling solution RHymod includes: the Pristley-Taylor component for the evapotraspiration estimate, the ordinary kriging algorithm for the rainfall spatialization, the hymod model for the runoff production of the hillslope, and finally the new channel routing component presented in the next section. The modeling solution Hymod differs from the model solution RHymod by only turning off the channel routing component and the discharge for each HRU are just added downstream. LUCA (66) was selected as calibration component for both the modeling solutions. The objective function is the Kling-Gupta efficiency (KGE) function as presented in (63).

The test is performed on two different river basin: Fort Cobb and Little Washita. The simulation period covered 2006-2007 in the case Fort Cobb and 2002-2003 in the case of Little Washita river basin; one year was used for calibration and one year for verification. The simulations time step was hourly.

7.5.1 The flow routing component.

As presented in (51) the flow generated for each hillslope is kinematically propagated downstream in the channel network by integrating a non linear variant of the Saint Venant equation at each channel link (e.g. (19)).

For each link the continuity equation, is:

$$\frac{dS_i(t)}{dt} = \left[Q_{gen}(t) + \sum_{trib} Q_{trib}(t) - Q_i(t)\right] \qquad i = 1, 2, ..., H \tag{7.9}$$

where is the $S_i(t)$ is storage in the i-th link at time t, H is the total number of network links, $Q_i(t)$ [L^3 T^{-1}] is the output discharge from i-th link, Q_{trib} [L^3 T^{-1}] is the flow of upstream links, and $Q_{gen}(t)$ [L^3 T^{-1}] is the discharge generated at the hillslope of the link in question.

Differently from (51) the routing component is modified taking into account the novel approach proposed in (91).

Considering a generic cross section, the relation between the storage and the output discharge from a generic i-th link is presented in eq. 7.10:

$$S_i(t) = \frac{Q_i(t) \cdot l_i}{v_i(t)} \tag{7.10}$$

in which l_i [L] and v_i [L T^{-1}] indicate respectively the length and the velocity in the channel i-th. The velocity is estimated as presented in (92):

$$v_i(t) = v_r \cdot \left(\frac{Q_i(t)}{Q_R}\right)^{\lambda_1} \cdot \left(\frac{A_i}{A_R}\right)^{\lambda_2} \tag{7.11}$$

where A_i [L^2] is the upstream area of the link, v_r [L^2], Q_r [L^3 T^{-1}] and A_r [L^2], are reference velocity, discharge and area, λ_1 and λ_2 are the scaling exponents of velocity for discharge and upstream area, respectively.

Replacing in eq. 7.10 the velocity as proposed in eq. 7.11 gives:

$$S_i(t) = \frac{Q_i(t) \cdot l_i}{v_r \cdot Q_i(t)^{\lambda_1} \cdot {A_i}^{\lambda_2}} \tag{7.12}$$

where Q_r and A_r are taken to be 1 [L^3 T^{-1}] and 1 [L^2], respectively.

Deriving in time eq. 7.12, the left hand side of eq. 7.9 becomes:

$$\frac{dS_i(t)}{dt} = \frac{l_i \cdot (1 - \lambda_1)}{v_R \cdot Q_i(t)^{\lambda_1} \cdot {A_i}^{\lambda_2}} \cdot \frac{dQ_i(t)}{dt} \tag{7.13}$$

Finally, replacing eq. 7.13 in eq. 7.9, the continuity equation for the link i-th the ordinary

becomes a non linear ordinary differential equation:

$$\frac{dQ_i(t)}{dt} = K\Big(Q_i(t)\Big) \cdot \Big[Q_{gen}(t) + \sum_{trib} Q_{trib}(t) - Q_i(t)\Big] \tag{7.14}$$

where:

$$K\Big(Q_i(t)\Big) = \frac{v_R \cdot Q_i(t)^{\lambda_1} \cdot {A_i}^{\lambda_2}}{l_i \cdot (1-\lambda_1)} \tag{7.15}$$

Eq. 7.14 has to be solved for each link i, i=1,2,..., H of the channel network.

Using Pfafstetter scheme as described in Chapter 4 the network is represented. The resolution of the set of routing equations starts from the upstream hillslopes (where the term $\sum_{trib} Q_{trib}(t)$ is null) and goes dowstream (where the $\sum_{trib} Q_{trib}(t)$ becames known term) according to the numbering rules.

The procedure allows to provide for each link i both the outgoing discharge and the mean velocity.

7.5.2 Applications and results.

Differents components of the framework NewAge-JGrass are applied in sequence accordingly the methology presented in (51). They are: the geomorphological analysis tools to extract the HRU, the river network and the geomorphological features used in the other components, the meteorological interpolator components for the spatialization of the meteo variables (air temperature and rainfall), the potential evapotraspiration component, the runoff production and eventually the routing component to compute discharge, the automatic calibration component to estimate the best set of model parameters, the validation package component to compute some goodness of fit indexes and to measure quantitatively the performance of the model.

For each delineation (DL1, DL3 and DL20) the Hymod and RHymod modeling solution were applied. One year calibration was performed by using the LUCA algorithm by optimizing the KGE objective function. Finally, the simulation results are presented qualitatively by comparing measured and simulated hydrograph and quantitatively, by computing two indices of goodness of fit: the index of agreement, IOA, (158) and the percentage model bias (PBIAS).

The Fort Cobb and the Little Washita river basin results are presented in tab.(7.4) and (7.5), respectively. Each row contains: i) the delineation type (DL1, DL3 and DL20); ii) the model solution (Hymod and RHymod); iii) the optimized objective function (KGE) value and the goodness of fit indices (IOA and PBIAS) for all the simulation period.

Tab.(7.6) and (7.7) present the optimum values of the model parameters for both the model configurations (Hymod and RHymod) and for all the delineations (DL1, DL3 and DL20) for Fort Cobb and Little Washita river basin.

From the quantitive analysis it can be concluded that both modeling solutions, Hymod and RHymod, are able to simulate the discharge in a reliable way.

Table 7.4: Fort Cobb simulation results for different delineations and for different model configurations.

Delineation	Modeling solution	KGE	IOA	PBIAS
DL1	Hymod	0.53	0.79	24.40
DL1	RHymod	0.70	0.83	9.2
DL3	Hymod	0.65	0.81	13.01
DL3	RHymod	0.81	0.89	3.9
DL20	Hymod	0.63	0.80	18.40
DL20	RHymod	0.65	0.83	17.20

Table 7.5: Little Washita simulation results for different delineations and for different model configurations.

Delineation	Modeling solution	KGE	IOA	PBIAS
DL1	Hymod	0.69	0.81	16.50
DL1	RHymod	0.74	0.85	7.3
DL3	Hymod	0.76	0.84	9.01
DL3	RHymod	0.82	0.89	3.2
DL20	Hymod	0.76	0.85	8.40
DL20	RHymod	0.77	0.84	7.60

Based on the goodness of fit indices, the RHymod simulates the total volume actually better than the Hymod model. The RHymod shows lower PBIAS values both for the Fort Cobb and for the Little Washita river basin.

Moreover, the RHymod model is able to simulate the peak values and the peak time well, as is confirmed by better KGE and IOA values compared to the Hymod case in both study cases.

Furthermore, the RHymod model, has three more parameters compared to the Hymod model. It brings to increase the time required by the calibration to convergence compared to the Hymod model which present anyway acceptable results.

For both river basins, the RHymod model provides the better performances in the delineation DL3. The models performances decrease in the case of delineation DL1. Even if RHymod outperforms the Hymod model this could be due to the lack of spatial rainfall spatial variability: a spatially uniform rainfall is applied in this case. In the case of the DL20 delineation, the use of RHymod and the explicit routing model does not provide any model performance improvement. This applies for both river basins. This result confirms the findings of in (**?**), (38) and (17) where it is evidently showns that the hillslope and not the channel contribute with the largest part of the residence time. The DL20 delineation shows the smallest HRU's size. Moreover for both the basins up to 15-25 km^2, non linearity in the process of runoff production could not be well simulated by a model based on a linear reservoir.

The two presented models slightly underestimate the highest peak flow values. This could be

Table 7.6: Fort Cobb river basin: parameter sets used in the simulations for RHymod (RH) and Hymod (H) model for different delineations (DL1, DL3 and DL3).

Symbol	DL1-H	DL1-RH	DL3-H	DL3-RH	DL20-H	DL20-RH
C_{max}	906.0546	403.8039	141.3387	813.2981	696.2530	595.0249
B_{exp}	1.7392	0.7204	2.1321	1.2697	1.1917	2.2070
$Alpha$	0.2193	0.2487	0.1294	0.2300	0.3616	0.3619
R_q	0.2261	0.3587	0.2088	0.2099	0.2288	0.2207
R_s	0.0008	0.0007	0.0001	0.0009	0.0012	0.0011
v_r	-	0.3132	-	0.6594	-	0.5806
$\lambda 1$	-	0.7978	-	-0.3915	-	-0.5298
$\lambda 2$	-	-0.0750	-	0.8423	-	0.3720

Table 7.7: Little Washita river basin: parameter sets used in the simulations for RHymod (RH) and Hymod (H) model for different delineations (DL1, DL3 and DL3).

Symbol	DL1-H	DL1-RH	DL3-H	DL3-RH	DL20-H	DL20-RH
C_{max}	841.8321	520.7016	155.8308	635.1629	998.9327	743.3361
B_{exp}	1.2597	1.2449	1.9143	5.6479	3.6370	2.5734
$Alpha$	0.2675	0.4501	0.2122	0.2526	0.2669	0.2540
R_q	0.1259	0.4357	0.1202	0.4882	0.1338	0.1360
R_s	0.0039	0.0006	0.0003	0.0002	0.0048	0.0048
v_r	-	1.1870	-	0.5449	-	0.5964
$\lambda 1$	-	0.0900	-	-0.0323	-	0.2262
$\lambda 2$	-	-0.0703	-	-0.0215	-	0.1150

due to the fact that not specific calibrations were performed for these events and to the implicit assumptions made in the classical formulation of Hymod model (104) where residence time in the hillslopes does not depending on soil moisture conditions. This is not in total agreement with the common knowledge for this hydrological problem. A possible solution could be the use of non linear runoff generation models instead the linear model presented in this study.

7.6 Soil and Water Assessment Tool (SWAT) vs NewAge-JGrass

A daily application of the NewAge-JGrass model and results comparison with the SWAT (5)model are presented. The test case is the Little Washita river basin.

7.6.1 The Soil and Water Assessment Tool (SWAT)

SWAT (5) was used to simulate hydrologic/water quality fluxes. Hydrologic processes simulated by SWAT include snow accumulation and melt, evapotranspiration, infiltration, percolation losses, surface runoff, and groundwater flows ((109)).

SWAT is a physically-based watershed-scale, distributed-parameter, continuous time, and

long-term, model that runs on a daily time step. It subdivides a watershed into subbasins connected by a stream network, and further delineates hydrologic response units (HRUs) consisting of unique combinations of land cover and soils in each subbasin.

SWAT uses a modification of the SCS curve number method (149) or Green and Ampt (62) infiltration model to compute surface runoff volume for each HRU. Peak runoff rate is estimated using a modification of the Rational Method (10). For evapotranspiration estimation, three methods are available in SWAT: Penman-Monteith (103), Priestley-Taylor (118), and Hargreaves (65). Daily or sub-daily precipitation and temperature data are used for calculations. Flow is routed through a channel using a variable storage coefficient method developed by Williams (157) or the Muskingum routing method. Outflow from a channel is adjusted for transmission losses, evaporation, diversions, and return flow. A kinematic storage model is used to predict lateral flow, whereas return flow is simulated by creating a shallow aquifer (5).

7.6.2 Results and Comments

SWAT and NewAge models were calibrated int the 2002-2003. The warm-up period ranges between 01-01-2002 and 01-06-2002 and the calibration period ranges between 01-06-2002 and 31-12-2003. Finally, the year 2004 is used as verification period. Shuffled Complex Evolution is used for the calibration of the models. The objective function used is KGE. Moreover, the models performances are measured considering three indices of goodness of fit: Index of Agreement (IOA), Percentage Bias (PBIAS) and Low Flow Function (FLF). Tab.(7.8) contains the optimized model parameters for the model NewAge-JGrass, a brief description, the minimum and maximum value, and the optimized value.

Fig.(7.10) presents the comparison between NewAge-JGrass simulated and measuered hydrographs; tab.(7.9) shows the objective function (in bold) and the indexes of goodness of fitness for calibration and entire simulation period.

Table 7.8: List of NewAge parameters

Symbol	Description	Min	Max	KGE
C_{max}	maximum storage in watershed, L	10.0	1000.0	476.33
B_{exp}	spatial variability of soil moisture storage, $-$	1.0	2.0	0.64
$Alpha$	distribution factor between two reservoirs , $-$	0.1	0.99	0.75
R_q	quick linear reservoir coefficient, T^{-1}	0.01	10000.0	0.66
R_s	slow linear reservoir coefficient, T^{-1}	0.001	1000.0	0.01
v_r	reference velocity in the routing process, LT^{-1}	0.1	2.0	0.75
λ_1	velocity scaling exponents for discharge, $-$	-0.9	0.9	0.66
λ_2	velocity scaling exponents for area, $-$	-0.9	0.9	0.03

Table 7.9: NewAge model Calibration results

	Test Period			
Period	KGE	FLF	IOA	PBIAS
2003	**0.75**	2.7	0.90	6.7
2003 – 2004	**0.65**	2.5	0.83	11.9

Tab.(7.11) contains the optimized SWAT model parameters for the considered objective functions (KGE), the minumum and maximum value and the optimum value. The simulation

Table 7.10: SWAT model Calibration results

	Test Period			
Period	KGE	FLF	IOA	PBIAS
2003	**0.71**	2.4	0.88	6.1
2003 – 2004	**0.58**	2.3	0.81	10.2

results are presented both from a qualitative point of view: fig.(7.11) show the comparison between measured and SWAT simulated discharge. Tab.(7.10) shows the optimized objective function (in bold) and the indexes of goodness of fitness for calibration and entire simulation period.

Both models are able to simulate the observed flow in the calibration and in the validation period, qualitatively, as presented in fig.(7.10), (7.11) and quantitatively speaking as presented in tab.(7.9) and (7.10).

The NewAge model was able to capture peaks flow slightly more accurately compared to SWAT model. However, SWAT is able to better reproduce the time series trends due to the more accurate spatial distribution of the soil physical features. Looking at simulations which use KGEas objective function, the NewAge model has better KGE and IOA values whereas SWAT produces better PBIAS and FLF best values.

Table 7.11: List of SWAT parameters

Symbol	Description	Min	Max	KGE
ALPHA_BF	base flow alpha factor for recession constant, *day*	0	2	0.39
CH_K(1)	fraction change in hydraulic conductivity in tributary channels, mm/hr	0	300	170.0
CH_K(2)	fraction change in hydraulic conductivity in the main channel, mm/hr	-0.01	500	23.07
CH_N(1)	Manning's n value for tributary channels	0.01	0.3	0.007
CH_N(2)	Manning's n value for the main channel	-0.05	0.05	
CN_F	fraction change in SCS runoff curve number, %	-0.25	0.25	-0.037
DDRAIN	depth of tile drains, mm	100	2500	1538
EPCO	plant uptake compensation factor	0.01	1	0.62
ESCO	soil evaporation compensation factor	0.001	1	0.62
GDRAIN	drain tile lag time, hr	0	48	34.8
GW_DELAY	groundwater delay, *day*	0	20	8.02
GW_REVAP	groundwater "revap" coefficient	0.02	0.2	0.033
GW_SPYLD	Specific yield of shallow aquifer, %	-0.5	1	0.24
GWQMN	threshold depth of water in the shallow aquifer for return flow, mm	0	5000	3151
OV_N	Manning's "n" value for overland flow, $s/m^{1/3}$	0.01	0.6	0.14
RCHRG_DP	deep aquifer percolation fraction	0	1	0.64
RSDIN	initial residue cover, kg/ha	0	10000	7359
SFTMP	snowfall temperature, C	-5	5	-3.172
SLOPE	fraction change in average slope steepness, %	-0.1	0.1	-0.078
SLSUBBSN	fraction change in average slope length, %	10	250	211.3
SMFMN	minimum melt rate for snow, $mm/C-day$	0	10	2.367
SMTMP	snow melt base temperature, C	-5	5	0.82
SNO50COV	snow water equivalent that correspond to 50% snow cover, mm	0	2	1.81
SNOCOVMX	minimum snow water content that corresponds to 100% snow cover mm	0	650	270.2
SOL_ALB	moist soil albedo, %	-0.5	1	0.14
SOL_AWC	available soil water capacity, %	-0.5	2	1.605
SOL_K	fraction change in saturated hydraulic conductivity, %	-0.5	5	2.848
SOL_Z	soil depth, %	-0.5	1	0.178
SURLAG	surface runoff lag time, *day*	1	12	4.22
TIMP	Snow pack temperature lag factor	0.01	1	0.434

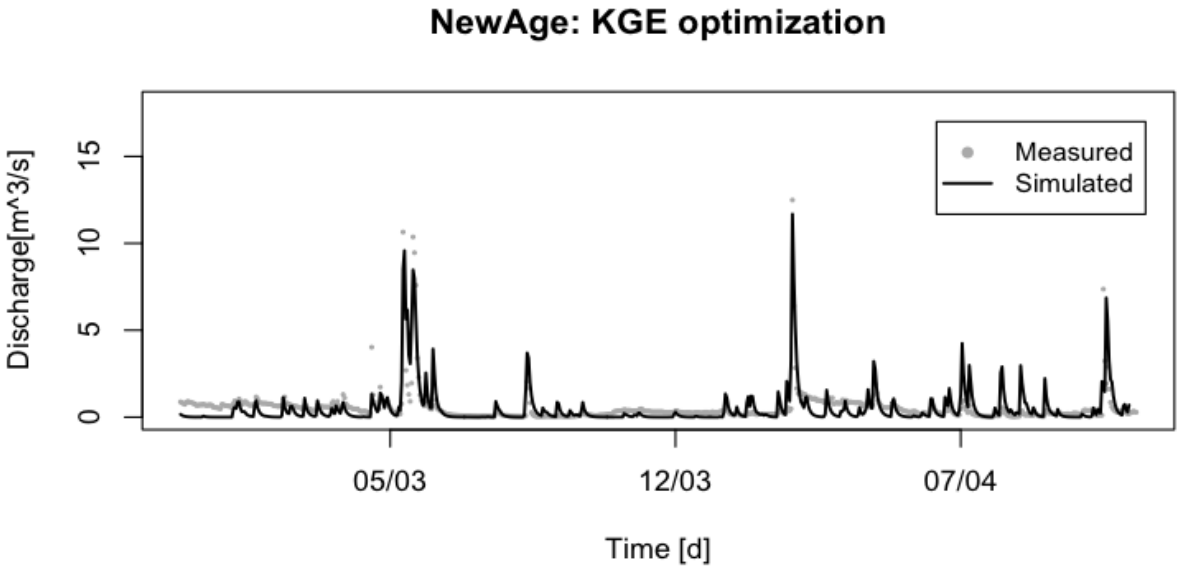

Figure 7.10: NewAge daily discharge simulation 2003-2004: KGE optimization.

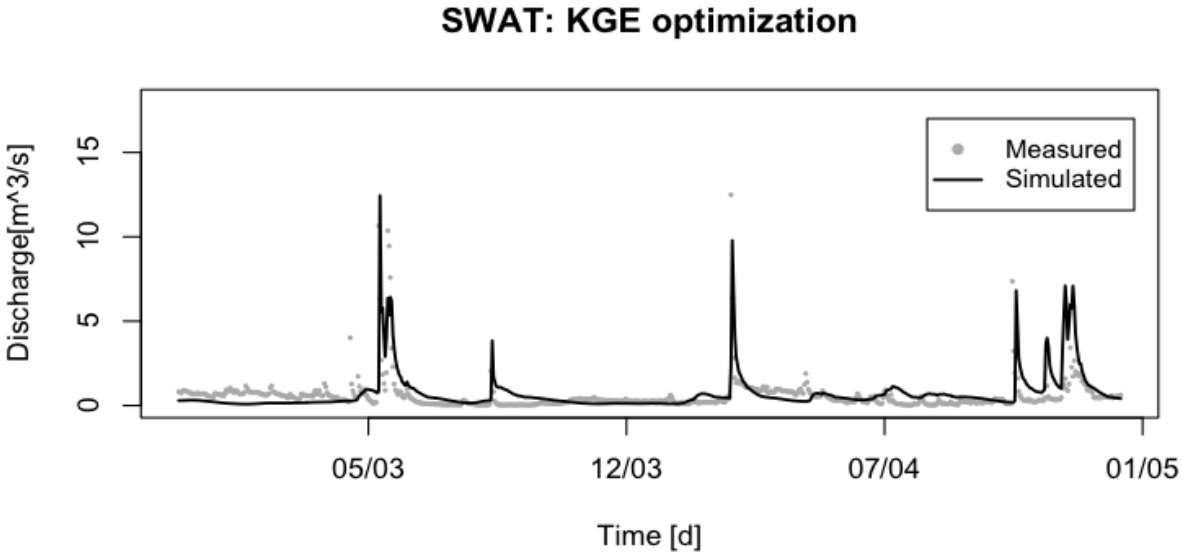

Figure 7.11: SWAT daily discharge simulation 2003-2004: KGE optimization.

7.7 Conclusion

The concept behind the NewAge-JGrass system is to provide a system in which any model can be built based on components that can be independently modified or changed, and which seamlessly work in a GIS environment (through the udig Spatial Toolboox), if they are implemented as OMS 3 components.

The system allows for the verification in single parts of a modelling chain thus making the localization of errors and the testing of alternatives altogether easier. An example of a routing component substitution is presented and applied on two different river basin. The versatility of the modeling approach was also tested by implementing two different modeling chains, one performing simulations with a lumped application of the model, using Hymod for the whole catchment. The other is representing a more distributed version of the same Hymod runoff generating mechanism, connected with a routing scheme. The predictions were tested by the analysis of the residuals and through the estimation of some objective indices, which were also implemented as software components. As a result the performances of the distributed version of the modeling chain was significantly better than the lumped version, thus supporting the idea that the increase in model complexity was worthwhile. The modeling chain was actually implemented using advanced specifications of the geographical objects, as required by OGC, and uses a particular specification of the river network hierarchy and the related hillslopes that were built upon the Pfafstetter ordering scheme.

These comparisons could be made by the same authors or independently by other researchers, since the NewAge-JGrass modeling system is freely available, with only the new component requiring coding. In this sense the infrastructure promotes independent testing and verification of research results with unprecedented easiness. In this perspective a component by component and interoperability comparison of the NewAge-JGrass system with others, such as PRMS, (82), or J2000, (79), that embraced the OMS3 frameworks can be investigated.

Finally a comparison with the SWAT model is presented. Both The models provided good results in term of discharge simulation. SWAT provided more accurate results in simulating trends and NewAge in simulating peak flows. The comparison opens large prospectives in term of models merging tools such as Bayesian model averaging.

8

The NewAge-JGrass Snow melting and Snow water equivalent component

This chapter presents the snow melting/snow water equivalent component, (SWE-C), in NewAge-JGrass. After a brief literature review, the model equations are presented. Moreover, a model application on the Poudre river basin (Colorado, USA) is presented. Model parameter calibrations by using Particle Swarm Optimization and model verification are performed at three different locations. Finally, and application of the model in the raster mode is presented creating raster maps of SWE.

8.1 Introduction

The physically based distributed approach is the best way to simulate the snowpack evolution. This solution has reached maturity and was pursued successfully with many recents models including CROCUS (21), Alpine3D (84), GEOtop (160), (45) and (35), ISNOBAL (96), UEB (146). These models often implement, besides the core energy budget, ancillary modeling of blowing snow, and other features that are required to reproduce the full set of thermodynamic quantities representing the snowpack state. However, performing the snow budget and modelling in its complete variability is not always necessary and requested. In many situations, where the prognostic significant quantity is just the global snow water equivalent in a sub-catchment, more simple models can work better. Also, realtime modelling with data assimilation and parameter calibration require that a whole forecasting cycle is obtained in few minutes for an entire day in advance in order to proceed with all the appropriate operations and testing. In any case, a best practice is to compare the most complete models with the simplest ones in order to assess the degree of complexity that is required for any task. The ancestor of all these simple models is the SRM model by Martinec (97) which was implemented several times and applied to hundreds of basins with reasonable success (98) and (99) . SRM is a linear model in which the independent

variables are an average of the daily temperature and an estimate of the catchment area covered by snow which is called the snow water depletion curve. These are tricky to determine, but possible to be detected by satellites. Therefore the model was largely used together with remote sensing data. In this paper we implement one of the minimalist SWE models based on the idea, partially investigated in (160). Once a good estimate of radiation is available, good spatially distributed estimates of the snow-water equivalent can be obtained.

(81) and (20) introduced simple SWE modelling based on the use of the radiation budget. However, in this dissertation we use the formulation of the problem developed by (28), since it was based simply on the estimate of the direct solar radiation, rather than the total net radiation, which is more rarely used and more difficult to obtain.

The SRM parameters as adopted by the authors were not calibrated or optimised with historical data. They can either be derived from measurements or judged based on hydrological experience taking into account the basin characteristics, physical laws, and theoretical relations or empirical regression relations. In many studies, this hypothesis was weakened, and we adopted a completely opposite strategy, in which we use all available data to assess the model's parameters. Therefore, we make use of data measured "at stations" and use the particle-swarm optimiser (77) to obtain the parameters of the model which can be studied for finding regularities and gaining knowledge about the phenomena.

Another novelty of our model, NewAGE-SWE, is that it is a part of a larger model, NewAge-JGrass (51),(52) and (50), which includes several modelling components.

8.2 The SWE-Component's equations

The snow melting model is based on a modificated approach presented in (78). The snowpack mass balance was simplified as follows. For the water equivalent of ice (M_i[L]):

$$\frac{dM_i}{dt} = P_s + F - M \tag{8.1}$$

and for liquid water (M_w [L]) in the snowpack.

$$\frac{dM_w}{dt} = P_r - F + M \tag{8.2}$$

Eq.(8.1) represents the variation in the time of the ice in the snowpack is equal to the algebraic sum of the snowfall, P_s, freezing, F, and melting, M (all expressed as snow water equivalent). Subsequently Eq.(8.2) represents the variation in time of the liquid water in the snowpack is equal to the algebric sum of the rainfall, P_r, freezing, F, and melting, M. If liquid water M_w exceeds liquid water-retention capacity of the snowpack (M_{max} [mm]), the surplus becomes snowmelt discharge q_m The liquid water retention capacity of a snowpack is related to the ice

content by a linear relationship, eq.(8.3)

$$L_{max} = \alpha_l \cdot I \tag{8.3}$$

Differently from (78), the time step to be used in these two coupled equations is not necessarily a daily timestep. We actually use a hourly timestep. This depends also on the choice of the flux's mathematical description that appears in the second term of equations (8.2,8.1).

8.2.1 The Type of Precipitation

The first hydrological process to be simulated is the discrimination between rainfall and snowfall considering that the two forms of precipitation appears as distinct in equations (8.2) and (8.1). Usually only rain gauge measurements and air temperatures are available. A common procedure is to consider a threshold for the air temperature T_s: all precipitation is considered snow if the air temperature for the time interval is less than or equal to T_s; all precipitation is considered rain if air temperature is greater than T_s. As proposed in (75) to avoid problems for parameter calibration, a smoother filter for thresholds is applied and the algorithm to discriminate between rainfall and snowfall can be described as follows:

$$\begin{cases} P_r = \alpha_r \cdot \left[\frac{P}{\pi} \cdot \arctan\left(\frac{T - T_m}{m_1}\right) + \frac{P}{2} \right] \\ P_s = \alpha_s \cdot \left[P - P_r \right] \end{cases} \tag{8.4}$$

where: P [L/T] is measured precipitation, P_r [L/T] is the rainfall precipitation, P_s [L/T] is the snowfall precipitation, T_m [C] is the threshold temperature, and m_1 [-] is the parameter controlling the degree of smoothing (if $m_1 \to 0$ threshold behaviour is simulated). The two coefficients α_r and α_s adjust for measurement errors of rain and snow. Because different values for different climate region were presented ((49), (137), (102)), in the model the two coefficients are considered parameters and therefore calibrated.

8.2.2 Snow melt fluxes

Based on the approach presented in (28) the melting process, eq.8.5, is a function of both shortwave radiation and air temperature. The two main differences in the presented model compared to (28) are: a new algorithm is used to compute the shortwave radiation (direct plus diffuse component) proposed by (31) and integrated into NewAge-JGrass model (52) which accounts for the complex topography, shadows and the sky view factor (30), and the cloud cover. The equation for the melt process is:

$$M = \begin{cases} \alpha_m \cdot EI \cdot T \cdot V_S & \text{during the day} \\ \alpha_m \cdot min(EI) \cdot T \cdot V_S & \text{during the night} \end{cases} \tag{8.5}$$

where:M [L/T] is the melt rate, α_m [L C^{-1} T^{-1} E] is the combined melting factor, T [C] is the air temperature, EI [E/T] is the energy index and V_s [-] is the sky view factor. The energetic

index is the potential energy accumulated over a given period at a certain point. To compute the energy index the shortwave energetic balance component implemented in NewAge (52) is used. The shortwave beam and diffuse solar radiation is accumulated for each pixel and the result is divided by the given period of the time. As presented in (28) five energetic index maps are computed starting from December 21st (winter solstice) to the middle of each month from February to June. During the night the snow melt is a function of the energetic index minimum value of the considered map, as presented in (28).

8.2.3 Freezing

The rate of freezing F that is compared in the mass budgets is linear related to the air temperature when the air temperature is less then the melting temperature, as presented in eq.(8.6)

$$F = \begin{cases} \alpha_f \cdot (T_m - T) & T < T_m \\ 0 & T \geq T_m \end{cases} \tag{8.6}$$

where F [L/T] is the freezing rate and α_f [L C^{-1} T^{-1}] is the freezing degree-day(hourly) factor. If the model is used with daily time steps temperature is the mean daily temperature. If it is used at hourly scale, temperature is the mean hourly temperature. Accordingly the value of the parameter α_f change values.

8.3 SWE-C integration in NewAge System

The SWE-C is perfectly integrated in the NewAge System as presented in fig.(8.1). Firstly, it uses the meteorological interpolation algorithms: Krigings tools, for temperature and precipitation interpolation, and JAMI for the temperature interpolation. Like the interpolation algorithms, SWE-C is able to work at a raster and a point scale. Secondly it uses the NewAge short wave radiation component in order to estimate the maps of cumulated energy in different periods of the year as explained in the model equations section. This components is able to take into account complex topography, shadow, and clouds cover. Thirdly, the SWE-C outputs could be: raster maps or time-series (one for each hillslopes centroids) of snow water equivalent and snow melt. Those could be used by the rainfall-runoff components in order to model a river basin where the snow contribution is not negligible. Finally, the SWE-C component could be connected to the NewAge and OMS3 calibration algorithm in order to estimate the best model parameters values.

8.4 SWE-C Application and results

The model is applied in the Cache la Poudre River basin as presented in chapter 3. Three applications are presented in this section. Firstly, the model was applied point mode for three stations where snow water equivalent measurement were available the model was calibrated

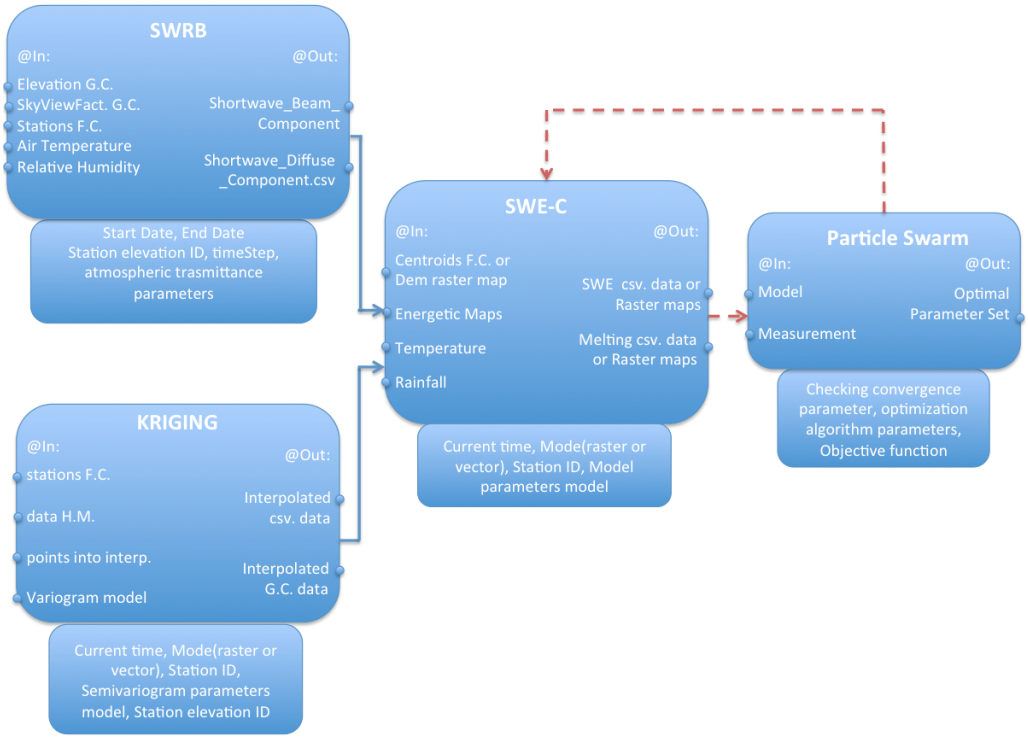

Figure 8.1: The SWE-C integration in the NewAge System. Connections with short wave radiation component and kriging interpolation algorithm. Connections to the Particle Swarm Optimization algorithm is shown as red dashed line.

and verified. Secondly, simulations were performed in order to investigate how representative the optimal parameter sets are for each stations. Finally, the model is applied in the fully distributed mode: raster maps of the snow water equivalent over the entire basin are simulated.

8.4.1 Model calibration and verification

As mentioned in the basin description there are three snow telemetering (SNOTEL) stations, fig. (3.5): Hourglass, Joe Wright and Deadman Hill. Tab.(6.2) shows their main features. They provide daily rainfall, temperature, and snow water equivalent data.

For Hourglass station the available data starts on 01-10-2008 and ends on 01-05-2012 (the first year is used as calibration period and the last 3 years are used as validation period); for the Joe Wright and Deadman Hill stations data goes from 01-10-1999 to 01-10-2009 (the first year is used as calibration period and the last 9 years are used as validation period).

To calibrate the SWE-C the configuration of the NewAge-JGrass components shown in fig.(8.1) was used. For this task, the the calibration algorithm Particle Swarm Optimization was used (77) and (43).

As objective function the Kling-Gupta Efficiency (KGE) presented in(63) was selected.

The model was verified for the three stations in two different ways. In a first approach a different optimal parameters set was estimated at each station and was used to simulate the

Table 8.1: List of the optimal parameters estimated at each of the three considered SNOTEL stations.

ID	City	α_m	α_r	α_s	α_f	T_m	α_l
1.0	Hourglass	0.19	0.91	1.12	0.085	1.00	0.14
6.0	Joe Wright	0.24	1.24	0.90	0.060	-0.48	0.23
10.0	Deadman Hill	0.16	1.32	0.98	0.017	1.55	0.51

validation period. The second method estimated the optimal parameters set in one station to model the simulation period in the other 2 stations and the procedure was repeated for each stations. For the Deadman Hill and Joe Wright stations the calibration period was the year 1999 and for Hourglass was the year 2008.

Theree classical GOF index are computed: Nash-Sutcliffe (NSE), Percentual Bias (PBIAS) and Index of Agreement (IOA). NSE values greater than 0.75 mean that the model can be considered good, values between 0.75 and 0.36 are associated with a satisfactory model and values below 0.36 indicate not a satisfactory model. Looking at the hydrological mode classification, as presented in (143) and (150), a model which presents an absolute PBIAS value less then 20 is considered "good", if the values are between 20 and 40 it is considered satisfactory, and if it is greater than 40 the model is considered "not satisfactory".

Tab.(8.2) shows, for the calibration period, at the top, and for entire simulation period, at the bottom, the indexes of goodness of fit for the three SNOTEL stations.

The model calibrated at each station and validated by using the optimized parameter can be considered "good" in both calibration and validation periods even if the model performance in the validation period is slightly lesser.

Table 8.2: List of the goodness of fit indexes for calibration period at the top and for entire simulation period at the botton, in the three SNOTEL considered stations .

Period	ID	City	KGE	NSE	PBIAS	IOA
Calibration	1.0	Hourglass	0.96	0.97	3.2	0.98
Calibration	6.0	Joe Wright	0.96	0.99	5.1	0.99
Calibration	10.0	Deadman Hill	0.97	0.98	1.9	0.99
Validation	1.0	Hourglass	0.94	0.92	2.8	0.96
Validation	6.0	Joe Wright	0.90	0.82	3.0	0.95
Validation	10.0	Deadman Hill	0.85	0.84	6.3	0.96

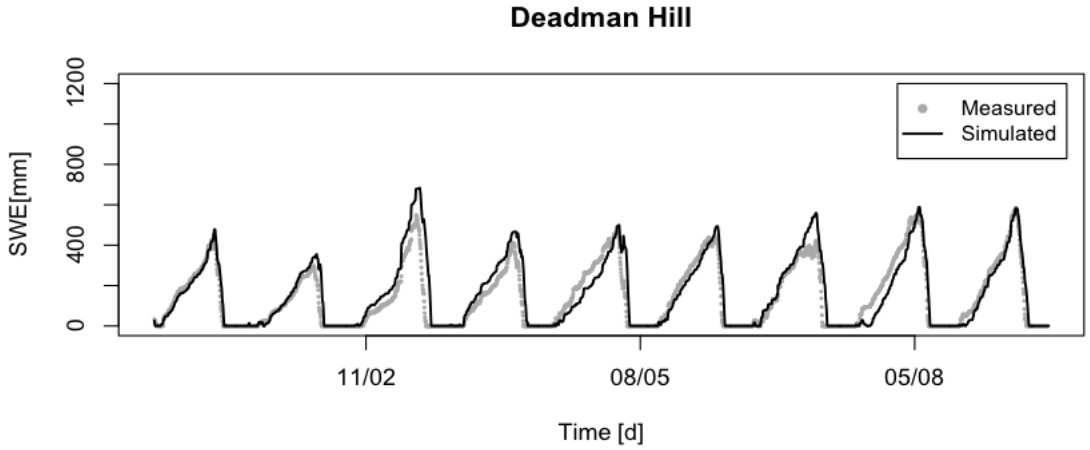

Figure 8.2: Calibration and validation results at Deadman Hill station: the gray dots represent the measured SWE and the solid black line represents the modelled SWE.

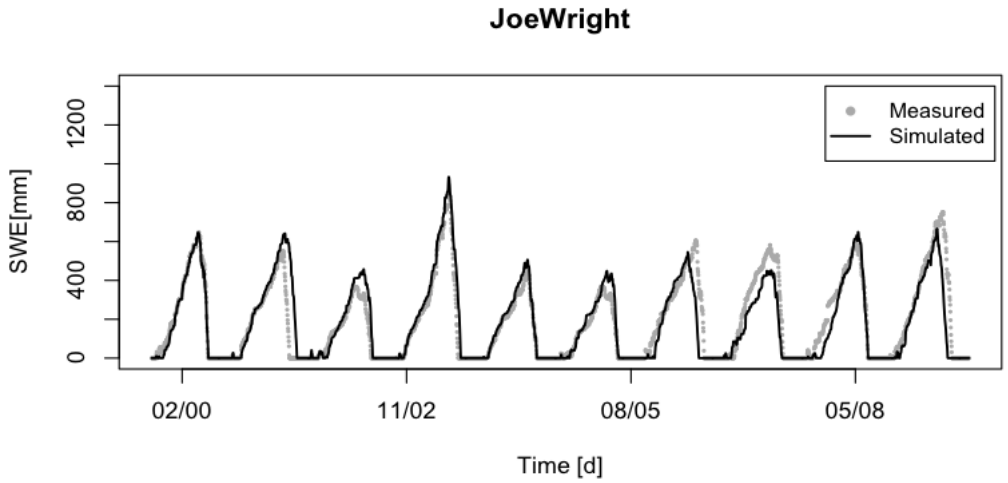

Figure 8.3: Validation results at Joe Wright station: the gray dots represent the measured SWE and the solid black line represents the modelled SWE.

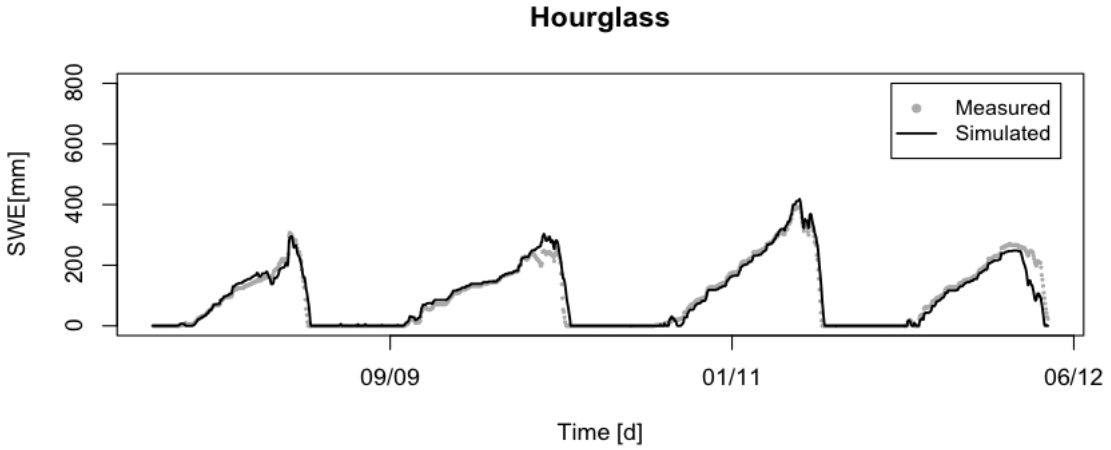

Figure 8.4: Validation results at Hourglass station: the gray dots represent the measured SWE and the solid black line represents the modelled SWE.

8.4.2 How much representative the parameters are?

In order to investigate how representative a parameter set really is a number of simulations are performed. For the entire simulation period the optimal parameter set for the Deadman Hill station was used for estimating the other two stations and the GOF indexes were computed. The same methodology was also applied for the Hourglass and Joe Wright stations, respectively. The simulations results are presented in tab.(8.3): the column "Optimal parameter set" specifies the station's parameter set used in the simulation.

Table 8.3: List of the goodness of fit indexes for entire simulation period ate the three SNOTEL stations: the column Optimal parameter set specifies which parameter set is used in the simulation and the columns ID and City specify the location in which the simulation is performed.

Optimal parameter set	ID	City	KGE	NSE	PBIAS	IOA
Joe Wright	10.0	Deadman Hill	0.77	0.6	11.2	0.91
Joe Wright	1.0	Hourglass	0.38	0.5	34.2	0.80
Hourglass	10.0	Deadman Hill	0.6	0.5	28.0	87.0
Hourglass	6.0	Joe Wright	0.36	0.4	33.8	0.82
Deadman Hill	1.0	Hourglass	0.49	0.41	35.0	0.81
Deadman Hill	6.0	Joe Wright	0.56	0.46	32.1	0.86

As presented in tab.(8.3), the model results are sensible to parameters variations. Even if the model for all the simulations performed can be classified as at least "satisfactory" for the NSE and PBIAS GOF's, this application emphasizes that the modeller has to pay attention to the parameters representativeness expecially at different locations. This becomes more inportant when the parameters are stictly related to measurement site features. For example, T_m could depend on the elevation, aspect of the measurement site, α_s and α_r could be function of the measurement instrument, α_m and α_f could be connected to the causes related to the amount of energy collected at the site (sky view factor, vegetation, or antrophic occlusions).

8.4.3 A distributed application of SWE-C

The SWE-C model is tested in distributed mode for the Poudre river. The simulation period was between 01-10-2008 and 01-10-2009. Daily rainfall and temperature raster maps were computed by using the detrended kriging algorithm. In this case three SNOTEL and three COOP meteorological stations were used. Tab.(3.4) shows their main features.

The mean values of the three optimal parameters set as presented in the previous section were used in this simulation. The results are presented in fig.(8.5). Snow water equivalent maps were plotted for each month starting from 01-11-2008 to 01-04-2009. Fig.(8.5).

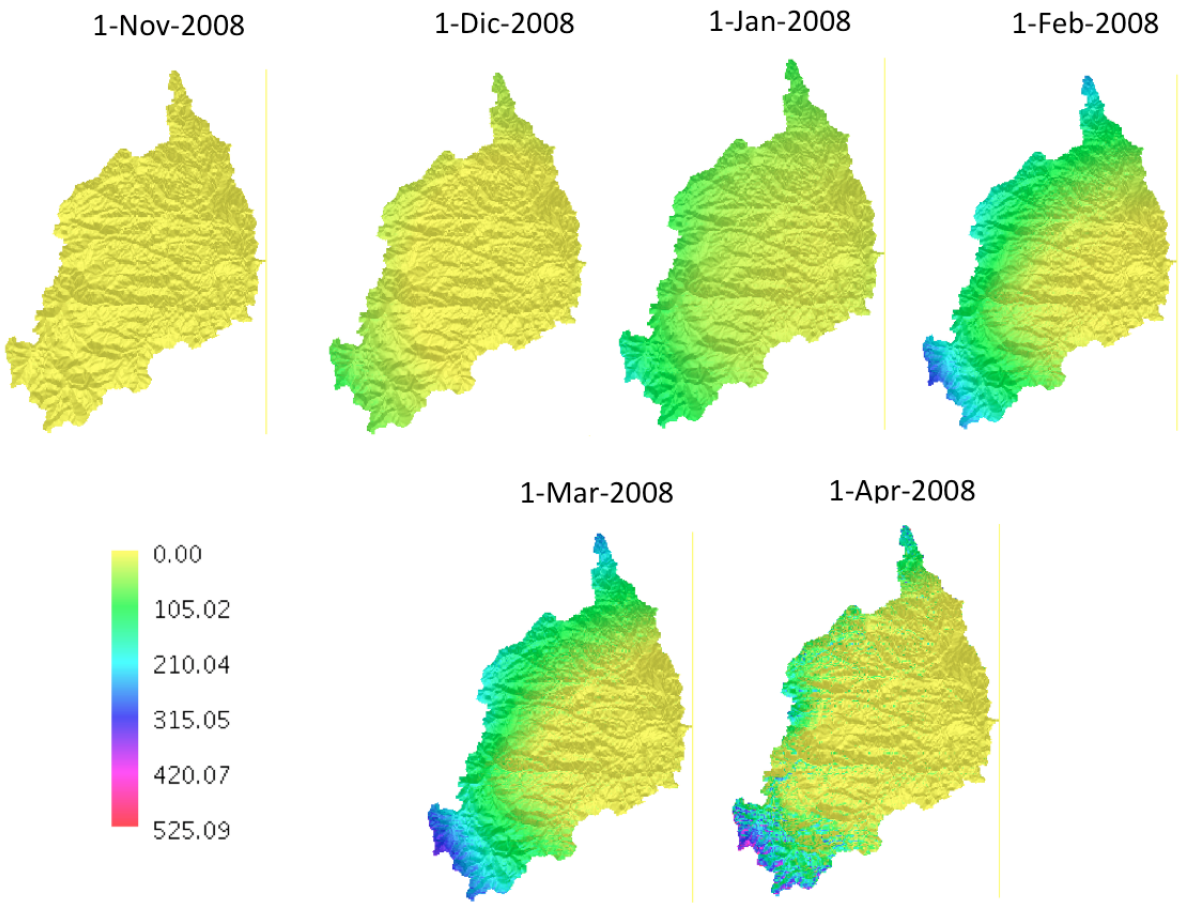

Figure 8.5: The SWE-C application in distributed mode: snow water equivalent maps from November 1st to June 1st .

8.5 Summary

In the chapter a novel snow melting and snow water equivalent model based on water and ice balance is presented. Here, the snow melt takes not only into account the temperature but also the energy received at the simulated point. The model is integrated into the NewAge-JGrass hydrological model as OMS3 component and for this reason it can make use of all the OMS3 components of the system: GIS based visualization, automatic calibration algorithm, and validation package. All these components are applied and verified at three SNOTEL stations located in the Cache la Poudre river basin (Colorado, U.S.) providing satisfactory results at all sites. A second model application focuses on the parameter representativeness. It shows that extending optimal parameter set at some location decreases model performances expecially when the parameters are striclty related to the climate and geomorphological feautures of the site. Finally, the distributed application in the Poudre river basin is presented. Modelling snow water equivalent patterns in a distributed mode provides the possibility to compare them with more physically based snow models and the option to verify them with snow water equivalent remote sensing data.

9

Model Calibration Algorithms in OMS3

This chapter presents selected calibration algorithms available in OMS3 as they were used for NewAge model. The genetic algorithm Particle Swarm Optimization (PSO) (77) was implemented first within NewAge directly and was later included into the OMS3 core as a Domain Specific Lenguage DSL. The OMS3 PSO component was tested in order to minimize analytical test-functions which are used as benchmark for global optimisation problems. For each minimization test a number of function evaluations is presented. The second optimization algorithm is LUCA ((66)). It was originally developed as a calibration method for the Precipitation-Runoff Modeling System model (PRMS) (82) and was later ported to OMS3 as DSL.

PSO and LUCA can be used with any OMS3 component or model.

9.1 What is model calibration and why do we need it?

In hydrology like in many other disciplines such as atmospheric science, business and statistics, models are inherently uncertain due to a variety of reasons:

- Incomplete model structure identification: model equations and hypothesis do not perfectly represent the real world because of highly interrelated water and energy processes;
- Incomplete model parameters identification due to the heterogeneity and non-linear nature of hydrological processes;
- Undetermined or missing initial conditions;
- Errors in the observed data used to drive and evaluate the model.

If all is known about a certain watershed such as meteorological input data, discharge, and energetic fluxes this can be called "truth".

In order to model such a watershed an appropriate model has to be selected that could be based on a lumped, semi distributed, or fully distributed approach. The model complexity, as

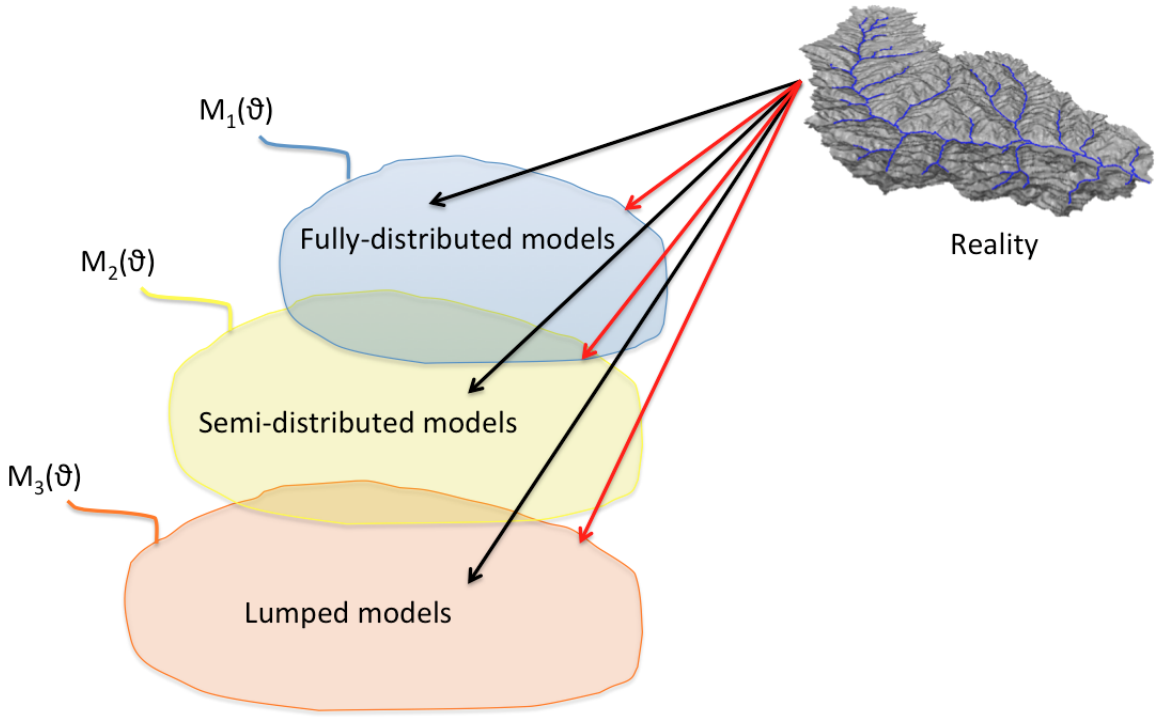

Figure 9.1: Hydrological model parameter calibration: $M_1(\theta)$, $M_2(\theta)$ and $M_3(\theta)$ represent the model parameter space for fully-distributed, semi-distributed and lumped hydrological model respectively.

well as the physical and spatial heterogeneity of the simulated processes, increases from the lumped to the fully distributed approach.

Each approach still represents a parametrized model. The gap between reality (actual discharge) and the modeled variable (simulated discharge) depends on the parameter values choosen within the model parameter space.

This is presented in fig.(9.1) by the length of the black arrows.

A calibration algorithm searches the parameter space for the parameter vector that provides the minimum distance between the observed value and the simulated variable. Such an "optimal" parameter vector is shown in fig.(9.1) by the length of the red arrows. The method by which the model parameters are distributed within the parameter space is defined by the search algorithm.

The problem is even more complicated because the true value is, generally, unknown and the uncertainty is always inherent to simulated values. The distances between the simulated and the observed variables is defined by the objective functions. Many objective functions are used in hydrology and a package was implemented in NewAge in order to allow the user to choose one according to the simulated process. This package is presented in the next sections

Automatic calibration algorithms were developed since 1960s. Two main approaches can be identified: local search (LSO) and global search optimization (GLO). LSO methods ((37)

(70), (71)) minimize the objective function by tuning the initialized parameters according to a pre-defined method such as derivative based or downhill simplex methods. Because LSOs are based on a local improvement of the objective function and because they strongly depend on the initial parameter set, their application has been largely unsuccessful in hydrology.

GSO methods, however, provide a better exploration of the parameter space since they avoid problems that arise from the highly dimensionality of parameters such as local minimum, discontinuous derivatives, and multiple regions of attractions. GSO methods are widely used in hydrology. The SCE-UA algorithm ((140) and (40)) combines the direct search method and the simplex downhill descent procedure. In AMALGAM (153) different search algorithms run concurrently and are learning from each other. Here, algorithms that present the highest reproductive success during the search are favored.

9.2 Particle Swarm Optimization (PSO)

PSO algorithms ((77)) are stochastic, population-based algorithms inspired by social behavior and movement dynamics of insects, birds, and fish.

PSO originated from research by Russ Eberhart, professor at the Purdue School of Engineering and Technology, in Indianapolis and co-author of the book Computational intelligence PC tools (44) and Jim Kennedy who graduated in psychology and is the other co-author of the book Swarm intelligence (76).

The basic concepts of PSO are based on philosophical and socio psychological insights about relationships between mind, intelligence, cooperation etc. The main goal is to take advantage of social studies to design efficient optimisation methods

"Like the common Genetic algorithms (GA's), PSO is a population based method, but unlike GAs, the underlying metaphor is cooperation instead of rivalry" (Maurice Clerc).

Suppose there is a search space and suppose that for each point in it we are able to compute the fitness f which is numerical value evaluated by a fitness function $F : R^n \rightarrow R$. The goal is to find the global optimum in the search space defined as the point which represent the best fitness (the smallest one). The basic PSO algorithm uses a certain number of particles randomly positioned in the search space. Each particle is able to move within the search space taking into account of the information the other particles provide in order to determine the "global optimum point".

The main concept behind the population-based particle swarm optimizer is the social behavior and movement dynamics of insects, birds, and fish. A group of random "particles" (values of parameters) is initialized randomly. In order to find the global optimum of the objective function each particle in the population adjusts its "flying" (i.e change) according to its own flying experience and that of its companions. The flying experience is determined by the flying velocities, i.e. the rate of change of their position in parameters space.

To draw M particles in a N-dimensional search space at the step t the i-th particle of the swarm and its velocity are represented by N-dimensional vectors respectively: $\mathrm{X}_i^t = \{x_{i,1}^t, x_{i,2}^t, ..., x_{i,N}^t\}$ and $\mathrm{V}_i^t = \{v_{i,1}^t, v_{i,2}^t, ..., v_{i,N}^t\}$. At each time step, the velocity and position of each particles (i.e. of the parameter set) are updated according to the equations:

$$v_{i,n}^{t+1} = \omega \cdot v_{i,n}^t + c_1 \cdot s_1 \cdot (p_{i,n}^t - x_{i,n}^t) + c_2 \cdot s_2 \cdot (g_n^t - x_{i,n}^t) \tag{9.1}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{9.2}$$

where i=1,2,....,D, and n=1,2,...,N, and in which:

- p_i^t is the element of the vector $\mathrm{P}_i^t = \{p_{i,1}^t, p_{i,2}^t, ..., p_{i,N}^t\}$ representing the individual best position of the i-th particle (i.e. the best visited position of the i-th particle);
- g_n^t is the element of the vector $\mathrm{G}^t = \{g_1^t, g_2^t, ..., g_N^t\}$ representing the best individual of the whole swarm

The search space S is defined by a hyperparallelepid defined as the Euclidean product: $S = \bigotimes_{n=1}^{N}[l_n, u_n]$ where l_n and u_n are the lower and upper limits of the hyperparallelepid. At t=0, the initialization process is performed according these equations:

$$x_i^0 = U(, l_n, u_n) \tag{9.3}$$

$$v_{i,n}^0 = \frac{U(l_n, u_n) - x_i^0}{2} \tag{9.4}$$

where U(a,b) is a uniform distributed random generated number between a and b.

The system evolves untill one of these two conditions is reached: i) the maximum number of iteration reached as specifiedby the user; ii) a relative or absolute tolerance between the last two global optima fitness is reached.

There are five parameters of the particle swarming algorithm: s_1 and s_2 are uniformly distributed random numbers between 0 and 1; c_1, the so called self confidence factor, and c_2, the so called swarm confidence factor. Both are acceleration constants ranging between 1.5 and 2. Finally ω is an inertial factor usually ranging between 0.4 and 1.4. All these parameter valuea are set by the users at the beginning of the optimization process.

9.2.1 Testing PSO algorithm

In order to assess the performances of the implemented PSO test functions commonly used benchmarks for global optimisation problems are implemented and optimized. The package of the implemented test functions includes:

- Rastrigin function has several local optima arranged in a regular lattice, but it only has one global optimum located at the point opt=(0,...,0)

$$n * 10 + \sum_{i=1}^{n} (x_i^2 - 10\cos(2\pi x_i)) \ , \quad -5.12 \leq x_i \leq 5.12 \tag{9.5}$$

- Rosenbrock function has only one optimum located at the point opt=(1,...,1)

$$\sum_{i=1}^{n-1} 100\,(x_{i+1} - x_i^2)^2 + (x_i - 1)^2 \ , \quad -30 \leq x_i \leq 30 \tag{9.6}$$

- Sphere model (first De Jong's function) only has one optimum at the point opt=(0,...,0)

$$\sum_{i=1}^{n} x_i^2 \ , \quad -100 \leq x_i \leq 100 \tag{9.7}$$

- Schwefel's function as the global minimum at position opt=(420.9687,...,420.9687)

$$418.982887274338 \cdot n + \sum_{i=1}^{n} -x_i \sin(\sqrt{|x_i|}) \ , \quad -500 \leq x_i \leq 500 \tag{9.8}$$

- Griewangk's function only has one global optimum located at the point opt=(0,...,0)

$$-\prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + \sum_{i=1}^{n} \frac{x_i^2}{4000} + 1 \ , \quad -600 \leq x_i \leq 600 \tag{9.9}$$

PSO was tested by minimizing the five test functions in order to find the analytical minimum. The results are presented in tab.(9.1). It shows the global optimal value and the number of evaluated functions. In all the applications the number of particles was set to 15 and the dimension of the search space was set to 4. The relative tolerance of criteria for termination was 1E-15. To avoid the influence of the numerical random generator, the PSO ran each test function 10 times.

9.3 Objective functions package in NewAge-JGrass

In order to measure the distance between modeled and measured time series a package of the classical objective function was implemented in NewAge-JGrass. The functions are:

Table 9.1: Summary of the test functions minimization

Name	Optim. Val.	Fun. calls
Rastrigin	[2.1E-3, 1.5E-3, 6.65E-4, 1.34E-3]	12120
Rosenbrock	[0.9972, 1.012, 1.0034, 1.0124]	14120
Sphere	[7.30E-6, -5.65E-6, -3.15E-5, 1.90E-5]	10230
$Schwefel's$	[420.58, 419.90, 420.06, 420.36]	25012
$Griewangk's$	[1.30E-2, -2.65E-3, -1.15E-2, 1.9E-2]	28640

- Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(S_i - O_i\right)^2} \tag{9.10}$$

  The RMSE is more sensitive to the occasional large error.

- Percent bias (PBIAS):

$$PBIAS = 100 \cdot \frac{\sum_{i=1}^{N}\left(S_i - O_i\right)}{\sum_{i=1}^{N} O_i} \tag{9.11}$$

  The PBIAS measures the average tendency of the simulated flow to be larger or smaller than their observed values. The optimal PBIAS value is 0.0, positive values indicate an overestimation of the model and negative values represent an underestimation. According to (95) $|PBIAS| < 5$ indicates excellent model performance, $5 < |PBIAS| < 10$ indicates very good model performance, with a $10 < |PBIAS| < 20$ the model performance is good, while a $20 < |PBIAS| < 40$ indicates that they are poor. Finally, a $|PBIAS| > 40$ indicates very poor model performance.

  This useful index indicates whether the model is systematically underestimating or overestimating the observations.

- Nash-Sutcliffe efficiency (NSE):

$$NSE = 1.0 - \frac{\sum_{i=1}^{N}\left(S_i - O_i\right)^2}{\sum_{i=1}^{N}\left(O_i - \bar{O}\right)^2} \tag{9.12}$$

  NSE numerator is the variance of the data that has not been explained by the model and NSE denominator is the total variance of the observed values about the mean. A NSE equals 1.0 means perfect fit; a NSE less than zero means that the mean value is more accurate than the model.

- Index of Agreement (IOA):

$$IOA = 1.0 - \frac{\sum_{i=1}^{N}\left(O_i - S_i\right)^2}{\sum_{i=1}^{N}\left|S_i - \bar{O}\right| + \left|O_i - \bar{O}\right|} \tag{9.13}$$

The IOA ranges between 0 (nocorrelation) and 1 (perfect fit). It represents the ratio of the mean square error and the potential error (158). As presented in (83) one of the advantage of this index is the sensitivity to extreme values due to the squared differences.

- Mean absolute error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|S_i - O_i\right| \tag{9.14}$$

It is the mean of the absolute value of the differences of the measured and simulated values.

- Kling-Gupta efficiency (KGE):

$$KGE = 1 - ED \tag{9.15}$$

where

$$ED = \sqrt{(R-1)^2 + (A-1)^2 + (B-1)^2} \tag{9.16}$$

in which R represents the linear correlation coefficient between the simulated time series S and the observed one O. A (eq.9.17) is the ratio between the observed (σ_o) and modelled (σ_s) standard deviations of the time series and takes account of the relative variability:

$$A = \frac{\sigma_o}{\sigma_s} \tag{9.17}$$

B (eq.9.18) takes account of the bias error.

$$B = \frac{\mu_s - \mu_o}{\sigma_o} \tag{9.18}$$

where μ_s and μ_o are the means of simulated and measured time series.

In all the previous definitions O stands for observed time series and S stands for simulated time series. N is the time series length.

9.4 Let us Calibrate (LUCA)

The LUCA calibration algorithm (66) in OMS3 is a multiple-objective, stepwise, automated procedure for model calibration. Like the particle swarm algorithm LUCA is based on two concepts: a search algorithm and the objective function(s) to evaluate model performance. The LUCA global searching algorithm is the Shuffled Complex Evolution (41). The SCE method is a global optimization algorithm that synthesizes deterministic and probabilistic concepts, controlled random search, competitive evolution, and complex shuffling approaches. For a problem

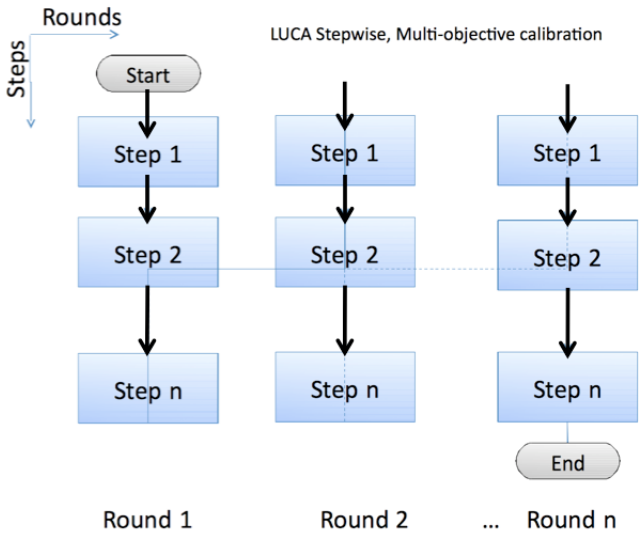

Figure 9.2: LUCA's rounds and steps schematic representation.

with n parameters, p complexes, and m points in each complex, the SCE method starts with random drawings $s = p \cdot m$ for generating the initial population of parameter sets from the feasible parameter space. Each criterion value at a point is evaluated. After sorting SCE then partitions s points into p complexes each containing m points. The following steps are then repeated until the pre-specified termination criteria are met: (i) evolve each complex according to the competitive complex evolution (CCE), (ii) combine the points in the evolved complexes into a single sample population, sort the sample population by increasing values and shuffle the sample population into p complexes. Complex evolution at each complex is independent from other complexes that make SCE well suited for parallelization within each iteration. Most commonly used termination criteria for SCE are maximum number of simulations and convergence criteria.

The LUCA algorithm is based on two important concepts: steps and rounds fig.9.2.

A step is associated with a parameter set, which contains one or more parameter values. A round consists of the execution of one or more steps. The selected parameters are calibrated for each for each calibration step. These calibrated parameter values replace the previous step parameter. Completion of the user-designated number of steps constitutes a round. A LUCA run configured with 1 step and 1 round round represents a classic SCE.

10

Synthesis

This final Chapter presents the overall conclusions derived from this dissertation.

10.1 Conclusions

The dissertation presents a novel hydrological infrastructure where a user can build, use, and assemble hydrological components according to the application requirements. As a modern hydrological model requires, NewAge selected uDig as GIS for visualization and managing of geospatial data and models, the OMS3 Console or the Spatial Toolbox for model execution, and the OMS3 system as framework to create the models in a very easy way. There are no interfaces to implement, no classes to extend and polymorphic methods to overwrite.

Each component of the NewAge is presented, applied, and verified by comparing model results with observed data. As the model is based on the hillslope-link partition of the basin, a formal definition of a DWM is given in Chapter 4. Basin delineation is performed by using the Pfafstatter algorithm and the Horton Machine tools for the geomorphological analysis of the basin

The problem of the interpolation of meteorological variables is presented and solved in Chapter 5. Deterministic and geostatistical algorithms were presented. The methods were able to operate in i) raster mode, providing the raster map of the interpolated variable and in ii) vector mode, providing its point time series. They can be easily integrated or substituted with other interpolation algorithms from in the OMS3 core and they were very helpful for many other components of the system such as the shortwave radiation, the rainfall runoff, and the snow water equivalent component.

The problem of energy balance and evapotranspiration estimate is presented in Chapter 6. Shortwave radiation was represented by using a parametric model which takes into account the effects of complex topography, shadow and cloud cover. Evapotraspiration can be modeled by choosing PenmanMonteith or PriestleyTaylor model according to the meteorological data availability.

The problem of runoff modeling is presented in Chapter 7. Runoff production and channel routing components are explained and applied on two river basins. Many examples were presented: building different modelling solutiions, using different methods for model parameter estimations, and analyzing the effect of different basin delineations. The overall performances of the discharge simulation can be considered very good for both the modeled river basins. Finally, a comparison with the SWAT model is performed in order to verify the performance of the new system against a well known hydrological model. Although SWAT model is able to better reproduce the time series trend NewAge model was able to capture peaks flow slightly more accurately than SWAT. The NewAge modularity feature allows to substitute and enhance single components of the modeling solution in order to improve the discharge trend simulation.

The system allows for verification of single parts of a modelling chain while keeping the constant fixed, thus making the localization of errors and the testing of alternatives altogether easier, as presented for the routing components.

The last NewAge-JGrass component presented is the snow melting-snow water equivalent model. The snow melt functions not only of the temperature as in many degree-day models but also of the energy received at the simulated point. Moreover, it works in raster and in vector mode providing as outputs snow water equivalent maps or point time series. It is perfectly integrated into the system. It uses the GIS visualization, the meteorological interpolation algorithms, the shortwave radiation model, the automatic calibration algorithms, and the validation package. Finally, its output can be used as input for the runoff component in order to model river basin where the snow melting is an important processes. The model is applied and verified for the Cache la Poudre river basin.

Chapter 9 presents the calibration algorithms as are implemented in OMS3: LUCA and Particle Swarm Optimization. All the models presented in this dissertation and all the OMS3 components can use these methods for parameters estimation. Applications of these methods are presented in all the dissertation: semivariogram estimate for the kriging interpolation, runoff and routing component, and the snow melting model.

Last but not least, the NewAge-JGrass is an example of hydrological infrastructure where the objective is not only to provide good performances from an operational point of view, but also development efficiency for researchers who want to build a reproducible-research systems (RRS). RRS means model source codes, data, and results sharing in order to allow the researchers to repeat the simulations using the same conditions while spending more time on scientific improvements.

11

Appendices

11.1 Appendix 1

The dataset Jura is presented in (59). Data of concentrations of seven heavy metals (cadmium, cobalt, chromium, copper, nickel, lead and zinc) were measured at 359 locations. The complete dataset is split into two parts: a validation dataset of 100 measurements and a dataset with 259 predictions, as presented in fig. 11.1. The validation dataset is used to check the results provided by the krigings interpolation algorithms. In particular tests are performed according to this schema: i) the predicted dataset is used as input data for kriging; ii) kriging is used to interpolate the heavy metal concentrations in the validation dataset; iii) the result comparisons of Gstat and NewAge-JGrass krigings are performed.

Two test are performed in order to asses the performances of the NewAge-JGrass krigings algorithms. In the first case the ordinary kriging is tested and all the measurement station are included in the interpolation. Figure 11.2 shows the scatterplot between the interpolated results in Gstat and NewAge for ordinary kriging. In the second test the local ordinary kriging is validated: only observations within a user defined distance (maxdist) were included in the interpolation. Two comparisons are performed between Gstat and Jgrass-NewAge for local ordinary kriging: in the first case maxdist parameter is set to 400m and in the second to 200. Figures 11.3 and 11.4 show the scatterplot and the residuals between the Gstat and NewAge interpolated results for local ordinary kriging, for maxdist=400 m and maxdist=200 m, respectively. The semivariogram for both algorithms is set to "Exponential" with nugget equals to 0.583, range equal to 0.775 and sill equal to 865.144.

11.2 Appendix 2

To test the Variogram component, the experimental variograms for lead (Pb) and cadmium (Cd) contamimants of the JURA dataset are computed and compared with the results provided by the algorithm implemented in (115). The results are presented in tables 11.1 and 11.2, respectively.

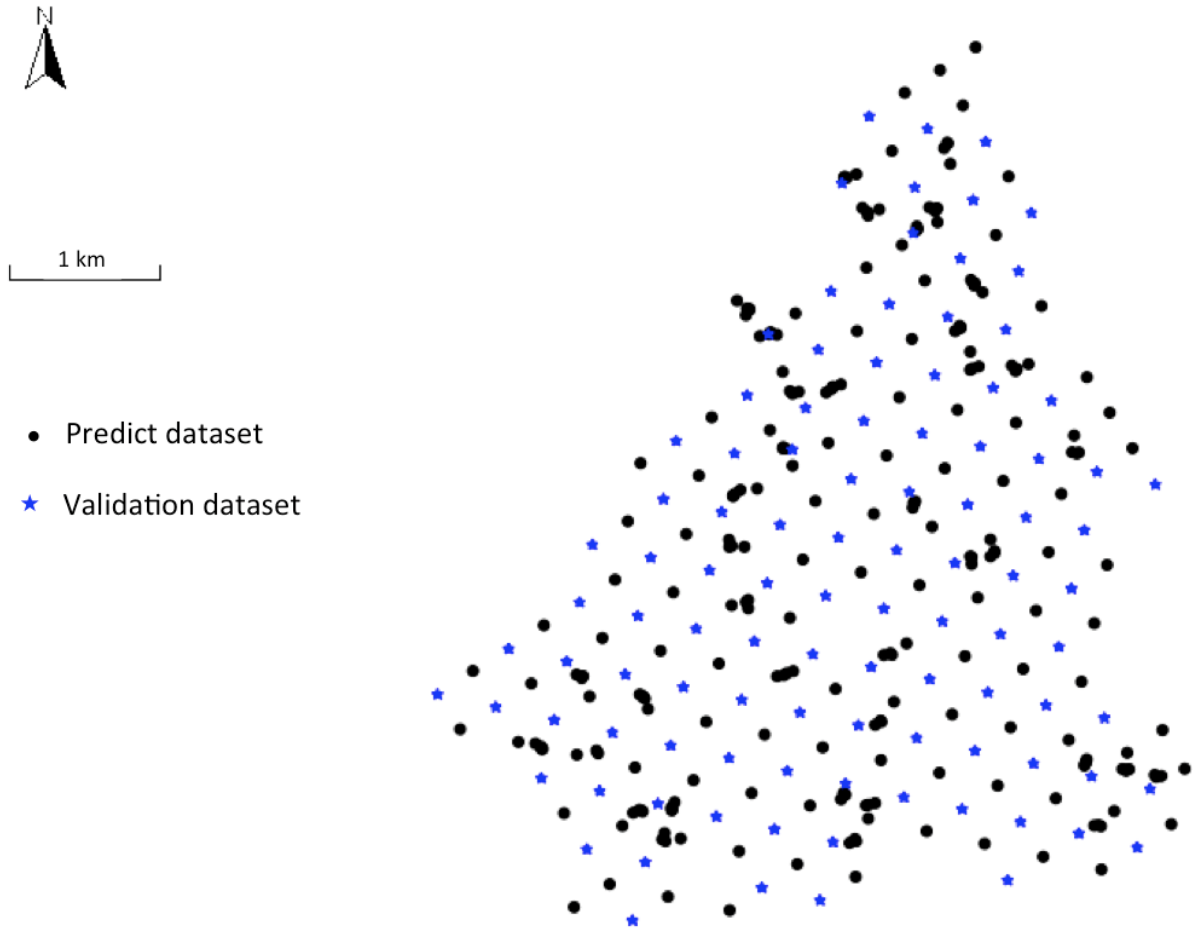

Figure 11.1: Jura dataset - The black dots represent the predict dataset and the blue stars represent the validation dataset.

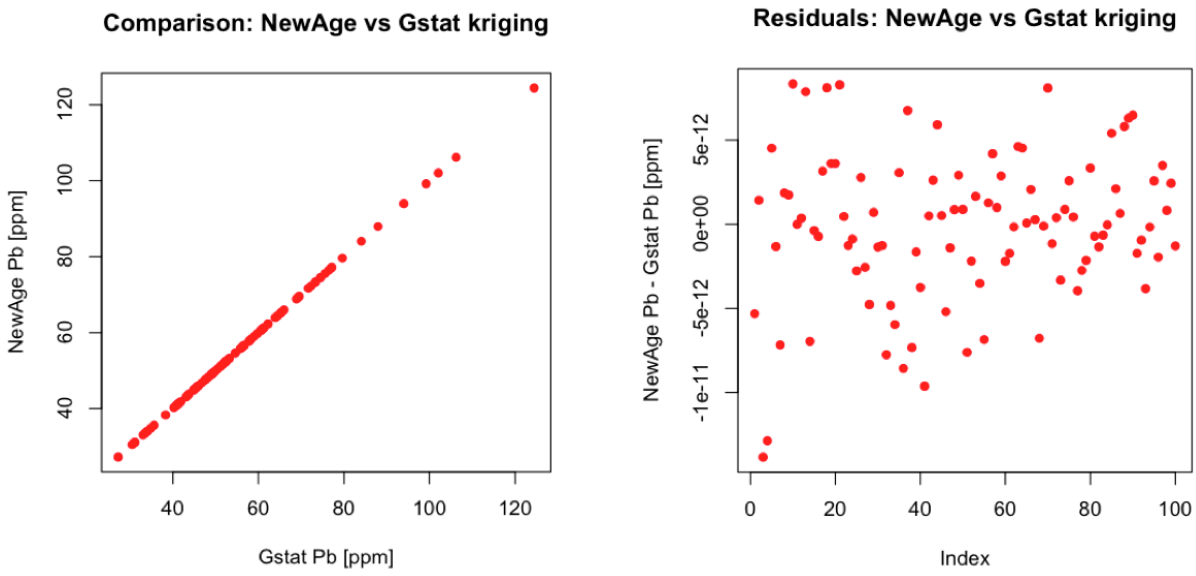

Figure 11.2: Ordinary kriging validation - Gstat and NewAge Ordinary Kriging Interpolated result scatter plot on the left and residuals plot on the right.

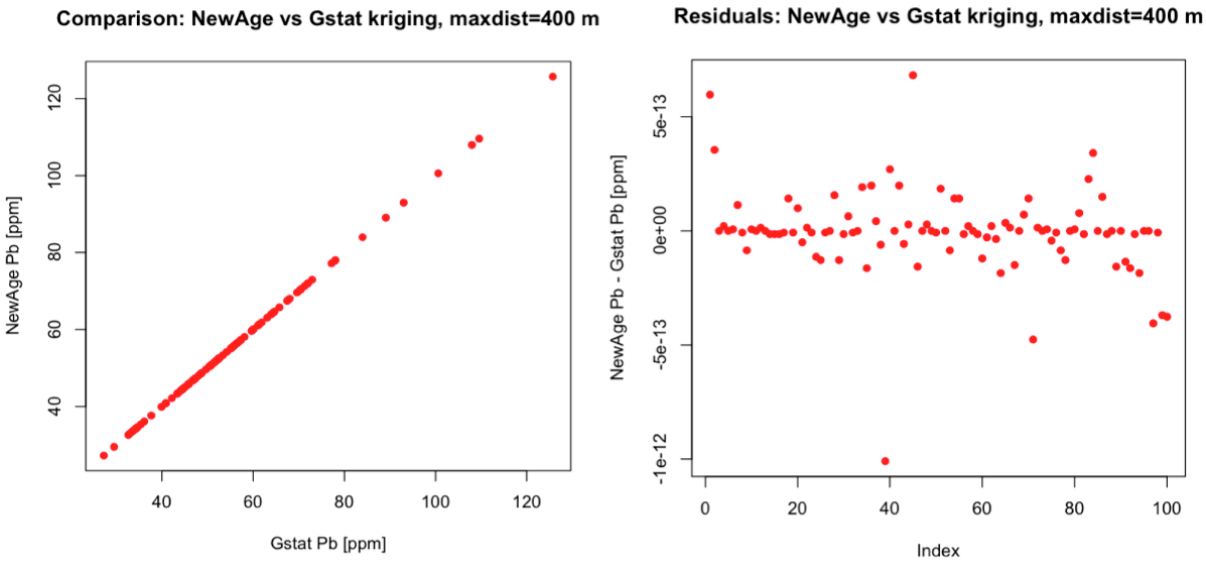

Figure 11.3: Local ordinary kriging validation with maxdist=400m - Gstat and NewAge local ordinary Kriging interpolated results scatter plot on the left and residuals plot on the right.

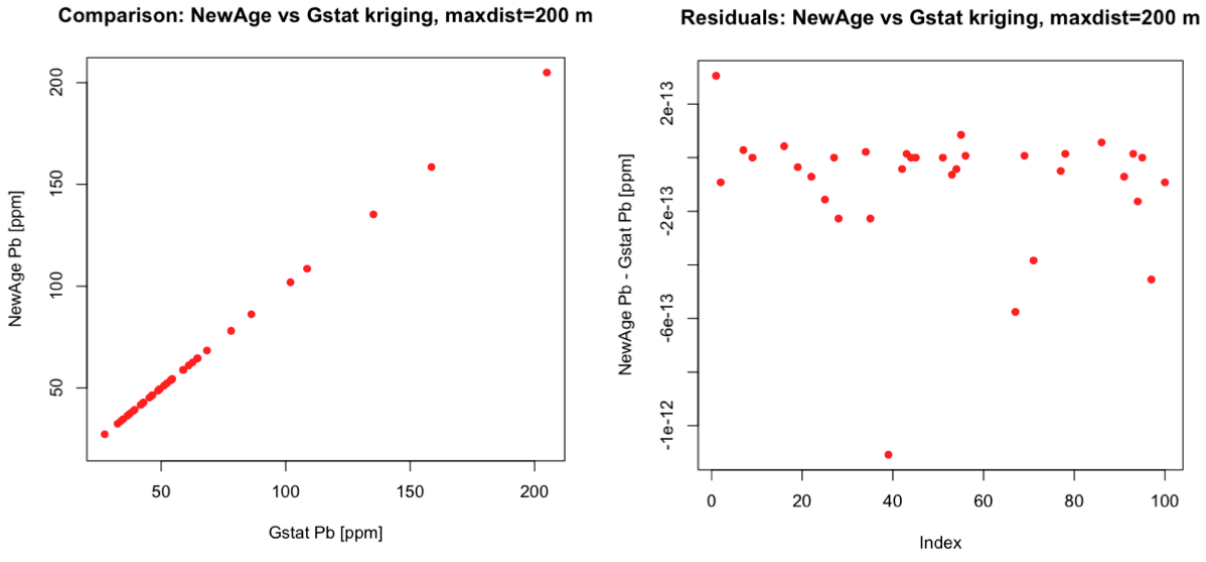

Figure 11.4: Local ordinary kriging validation with maxdist=200m - Gstat and NewAge local ordinary Kriging interpolated results scatter plot on the left and residuals plot on the right.

Table 11.1: Lead contaminant: experimental semivariogram comparison between Variogram and Gstat

Variogram experimental variance [*ppm*]	Gstat experimental variance [*ppm*]
450.531	450.531
825.663	825.663
716.872	716.872
742.012	742.012
920.709	920.710
770.260	770.260
757.003	757.003
865.575	865.575
814.006	814.520
850.888	850.068
852.501	852.220
885.145	885.743
1042.870	1042.870
1030.456	1030.456
871.414	871.414

Table 11.2: Cadmium contaminant: experimental semivariogram comparison between Variogram and Gstat

Variogram experimental variance [*ppm*]	Gstat experimental variance [*ppm*]
0.521	0.521
0.659	0.659
0.685	0.685
0.854	0.854
0.736	0.736
0.813	0.813
0.781	0.781
0.769	0.769
0.896	0.896
0.810	0.811
1.013	1.012
0.811	0.811
0.852	0.852
0.842	0.842
0.745	0.745

11.3 Appendix 3

The OMS3-simulation script for running the theoretical semivariogram estimate and the kriging algorithm in the OMS3 console:

```
1 import oms3.SimBuilder as OMS3
2 def dir = oms_prj
3 // This is variogram
```

```
4  vario = OMS3.sim_run(name:"variogram", {
5      model (while:"reader_data.doProcess") {
6          components {
7              "reader_data"       "org.jgrasstools.gears.io.timedependent.↩
                   TimeSeriesIteratorReader"
8              "vreader_station"   "org.jgrasstools.gears.io.vectorreader.↩
                   VectorReader"
9              "variogram"         "org.jgrasstools.hortonmachine.modules.↩
                   statistics.kriging.Variogram"
10          }
11          parameter {
12              "vreader_station.file"      "${dir}/data/jura.shp"
13              "variogram.fStationsid"     "Id"
14              "variogram.pPath"           "${dir}/output/out.txt"
15              // READERDATA
16              "reader_data.file"          "${dir}/data/variogram_test.csv"
17              "reader_data.idfield"      "ID"
18              "reader_data.tStart"     "2000-01-01 00:00"
19              "reader_data.tTimestep"      60
20              "reader_data.fileNovalue"   "-9999"
21          }
22          connect {
23              "vreader_station.outVector" "variogram.inStations"
24              "reader_data.outData"  "variogram.inData"
25          }
26      }
27 })
28
29 //println vario.model.variogram.outDist
30 //println vario.model.variogram.outVar
31
32 ps = OMS3.ps_run(name:"vgm",{
33          model() {
34              components {
35                  "vgm"  "org.jgrasstools.hortonmachine.modules.statistics.↩
                       kriging.VGM"
36              }
37              parameter {
38               "vgm.modelname" "exponential"
39               "vgm.nugget"    10    // 1..15
40               "vgm.sill"      600   // 100 .. 10000
41               "vgm.range"     500   // 0..10
42               "vgm.distances"  vario.model.variogram.outDist
43               "vgm.inp"       vario.model.variogram.outVar
44              }
45          }
46          // kmax : optional, default 1000
47          kmax     5000
```

```
48          // numPart optional default 10
49          numPart 10
50          // all check variables optional, values
51          // below show the defaults.
52          check_after 175
53          check_last  50
54          check_min   30
55          check_delta 1e-8
56          verbose 0
57          // parameter to optimize
58          parameter {
59              "vgm.range"  (lower:0.0, upper:5.0)
60              "vgm.sill"   (lower:0, upper:2000)
61              "vgm.nugget" (lower:0.0, upper:2000.0)
62          }
63          // objective function
64          objfunc(method:RMSE) {
65              sim(data:"vgm.result")
66              obs(data:"vgm.obs")
67          }
68      })
69  printf('rangeFinal= ')
70  println ps.model.vgm.range
71  printf('sillFinal= ')
72  println ps.model.vgm.sill
73  printf('nuggetFinal= ')
74  println ps.model.vgm.nugget
75
76  krig = OMS3.sim_run(name:"kriging") {
77   model (while:"reader_data.doProcess") {
78     components {
79      "reader_data"              "org.jgrasstools.gears.io.timedependent.↩
           TimeSeriesIteratorReader"
80        "vreader_station"          "org.jgrasstools.gears.io.vectorreader↩
             .VectorReader"
81        "vreader_interpolationpoint" "org.jgrasstools.gears.io.↩
             vectorreader.VectorReader"
82         "writer_data"               "org.jgrasstools.gears.io.↩
             timedependent.TimeSeriesIteratorWriter"
83          "kr"                       "org.jgrasstools.hortonmachine.↩
               modules.statistics.kriging.KrigingRagInf"
84     }
85     parameter {
86      // READER
87          "vreader_station.file"     "${dir}/data/jura.shp"
88          "vreader_interpolationpoint.file"    "${dir}/data/↩
             InterpolarionPoints.shp"
89          "reader_data.file"       "${dir}/data/variogram_test.csv"
```

```
90          "reader_data.idfield"          "ID"
91          "reader_data.tStart"         "2000-01-01 00:00"
92          "reader_data.tTimestep"      60
93          "reader_data.fileNovalue"   "-9999"
94       //KRIGING
95          "kr.fInterpolateid"     "Field2"
96          "kr.fStationsid"     "Id"
97          "kr.doLogarithmic"       false
98          "kr.pA"             ps.model.vgm.range
99          "kr.pNug"          ps.model.vgm.nugget
100         "kr.pS"             ps.model.vgm.sill
101          "kr.pMode"           0
102          "kr.defaultVariogramMode" 1
103          "kr.pSemivariogramType"   "exponential"
104          "kr.doDetrended" "false"
105       //WRITER
106         "writer_data.file"      "${dir}/output/kriging_interpolated.csv"
107         "writer_data.tStart"       "2000-01-01 00:00"
108         "writer_data.tTimestep" 60
109      }
110      connect {
111         "vreader_station.outVector"              "kr.inStations"
112         "vreader_interpolationpoint.outVector"   "kr.inInterpolate"
113         "reader_data.outData"                    "kr.inData"
114         "kr.outData"                             "writer_data.inData"
115      }
116    }
117 }
```

11.4 Applendix 4

In order to verify the theoretical semivariogram estimate as implemented in JGrass-NewAge a comparison with the GSTAT-R package was performed. The Jura dataset was used as test case and semivariogram fitting is performed for Copper, Zinc, and Cadmium. The model parameters nugget, sill, and range, are tuned in order to minimize the root mean square error with weighted respect to the number of pairs in each semi variance class between the theoretical and empirical semivariogram. The OMS3 script presented in Appendix 3 is used for this purpose. In the JGrass-NewAge application, the EVC component computes the experimental variogram. The genetic optimization algorithm Particle Swarm tunes the model parameters of the theoretical Vgm semivariogram in order to fit the experimental values best.

In the Gstat application the Variogram is used. Variogram computes the experimental variogram. The function iterates over (a) a direct (ordinary or weighted least squares) fit of the partial sills and (b) search using gradients for the optimal range value(s), until convergence of after a combined step (a) and (b) is reached.

Tables 11.3, 11.4, and 11.5 report the results provided by the R-Gstat and JGrass-NewAge packages for Lead, Cadmium, and Zinc theoretical semivariogram parameters estimation. The Gstat results are quite well reproduced and the differences between the two packages could be due to the different nature of the minimization algorithms in Gstat.

Table 11.3: Comparison between Gstat and Jgrass-NewAge semivariogram fitting results for Lead heavy metal, Jura dataset.

	Variogram	**Range[m]**	**Sill[ppm]**	**Nugget[ppm]**
Gstat	Exponential	426.94	0.434	482.26
Jgrass-NewAge	Exponential	426.77	0.433	482.24

Table 11.4: Comparison between Gstat and Jgrass-NewAge semivariogram fitting results for Cadmium heavy metal, Jura dataset.

	Variogram	**Range[m]**	**Sill[ppm]**	**Nugget[ppm]**
Gstat	Gaussian	0.319	0.302	0.512
Jgrass-NewAge	Gaussian	0.311	0.296	0.517

Table 11.5: Comparison between Gstat and Jgrass-NewAge semivariogram fitting results for Zinc heavy metal, Jura dataset.

	Variogram	**Range[m]**	**Sill[ppm]**	**Nugget[ppm]**
Gstat	Spherical	0.62	610.72	218.76
Jgrass-NewAge	Spherical	0.60	610.78	218.71

11.5 Appendix 5: JAMI

The JAMI algorithm can be explained as followed. As fist step input files are read in: the meteo stations shapefile; the meteo stations measurements .csv file; the shape file of the basin splitted in hillslope; the Areas.csv file which contains for each hillslopes and for each altimetric band the area in [km^2]; the Altimetry.csv file which contains for each hillslope the maximum elevation of each altimetric band and the hillslope centroid elevation.

The interpolation of the meteo variable for each hillslope and for each altimetric band is based on so called "Active stations": a group of stations, in the hillslope or in its neighborhood which provide data at the current time step and are sorted by distance ascending from the hillslope centroid. Fig. 11.5 presents JAMI as OMS3 component by showing all the inputs and the outputs JAMI require to run. The interpolation algorithms depends on the meteo

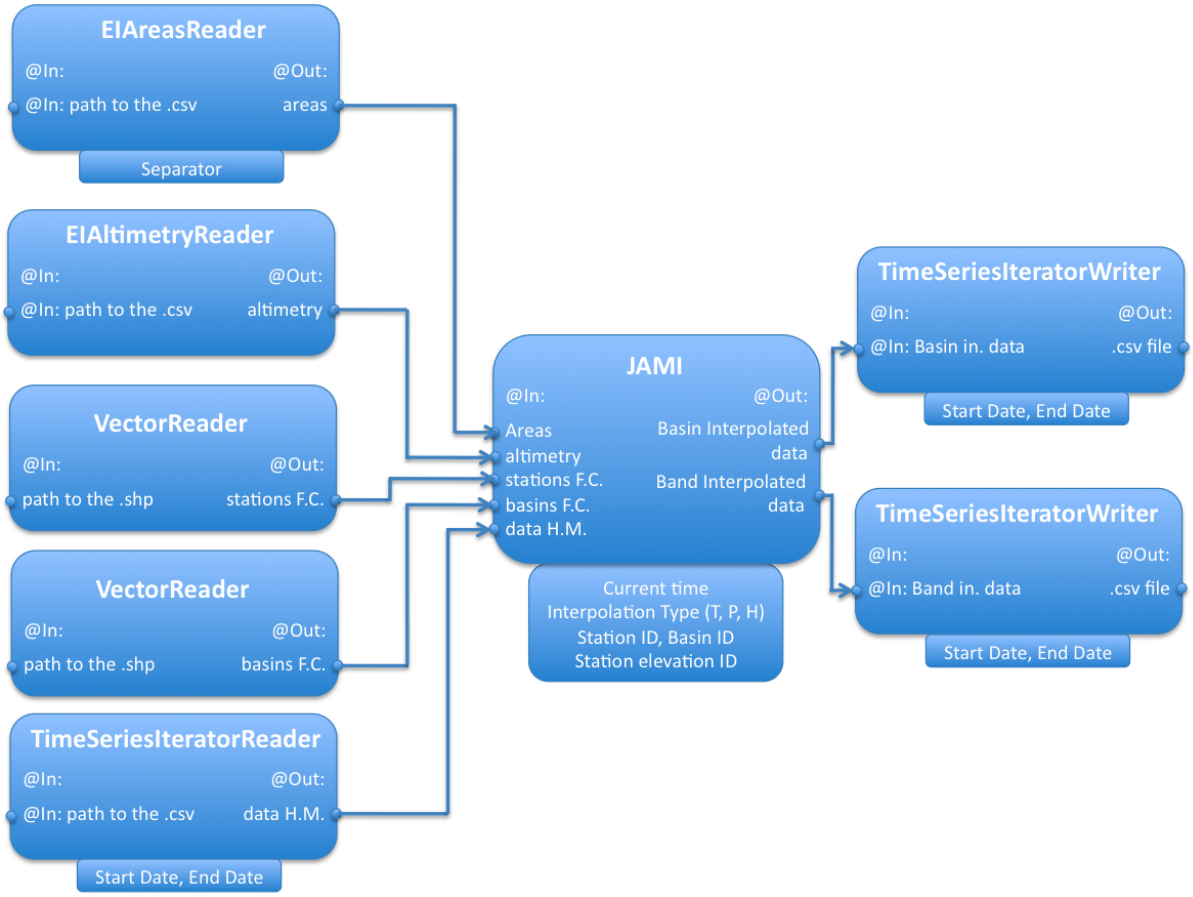

Figure 11.5: OMS3 JAMI wocomponent and data flow.

interpolation variable and the number of the active stations (N_{as}) for the current hillslope and time step. They are explained in the next subsections.

11.5.1 Temperature interpolation

Different interpolation algorithms exist depending on N_{as}:

- N_{as}=0: the program will stop because the temperature is the only needed input data;
- N_{as}=1: the standard atmospheric model it is considered; for each altimetric band b of the current hillslope the temperature in computed using a simple adiabatic transformation:

$$T_b = T_s - \gamma \cdot (z_b - z_s) \tag{11.1}$$

where: z_b[m] and z_s [m] are the centroid altimetric band and station elevation, respectively, T_b [K] and T_s [K] are the bands interpolated and the measured temperature, and γ = 0.006509 [K/m] is the adiabatic laps rate. The interpolated hillslope centroid value is computed as the weigthted mean of the altimetric band temperature with respect to the altimetric band area:

$$T_h = \sum_{i=1}^{n_b} T_i \cdot \frac{A_i}{A_h} \tag{11.2}$$

where n_b is the number of altimetric bands of the hillslope, A_i[km^2] is the area of the i-th altrimetric band, T_i is the interpolated temperature of the i-th altimetric band, and A_h[km^2] is the hillslope area.

- N_{as} >1: the standard atmospheric model it is considered for elevations lower than the minimum station elevation and higher than the maximum station elevation; for the points between the maximum and minimum station elevation, the laps rate is computeted:

$$\begin{cases} T_b = T_{s,min} - \gamma \cdot (z_b - z_{s,min}) & z_b < z_{s,min} \\ T_b = T_{s,1} - \frac{T_{s,2} - T_{s,1}}{z_{s,2} - z_{s,1}} \cdot (z_b - z_{s,1}) & z_{s,min} < z_b < z_{s,max} \\ T_b = T_{s,max} - \gamma \cdot (z_b - z_{s,max}) & z_b > z_{s,max} \end{cases} \tag{11.3}$$

where T_b [C] and T_s [C] are the band interpolated and the measured temperature, the subscripts $_{s,min}$ and $_{s,min}$ indicates quantity (elevation (z) and temperature (T) related to the stations with minumum and maximum elevation, and the subscripts $_{s,1}$ and $_{s,2}$ indicates quantity (elevation (z) and temperature (T) related to the stations are located between the centroid altimetric band. The interpolated hillslope centroid value is computed as the weigthted mean of the altimetric band of the temperature with respect to the altimetric band area, as reported in eq.11.2

11.5.2 Relative humidity interpolation

Different interpolation algorithms are presented depending on N_{as}:

- N_{as}=0: default value for relative humidity is assigned to each altimetric band of each hillslope;
- N_{as}=1: a constant value equal to the measured value is assigned to each altimetric band of each hillslope:

The interpolated hillslope centroid value (H_h) is computed as the weigthted mean of the altimetric band's relative humidity with respect to the altimetric band area:

$$H_h = \sum_{i=1}^{n_b} H_i \cdot \frac{A_i}{A_h} \tag{11.4}$$

where n_b is the number of altimetric bands of the hillslope, A_i[km^2] is the area of the i-th altrimetric bands, H_i is the interpolated relative humidity of the i-th altimetric band and A_h[km^2] is the hillslope area.

- N$_{as}$ >1: the relative humidity for altimetric band which centroid elevation is lower than the lowest station elevation $z_{s,min}$, is assigned to the measured value of this station ($H_{s,min}$); the relative humidity for altimetric band whose centroid elevation is higher than the highest station elevation $z_{s,max}$ is assigned equal to the measured value of this station ($H_{s,max}$); the relative humidity for the altimetric band which centroid elevation is between $z_{s,min}$ and $z_{s,max}$ is computeted using a measured lapse rate as presented in eq.11.5

$$\begin{cases} H_b = H_{s,min} & z_b < z_{s,min} \\ H_b = H_{s,1} - \frac{H_{s,2}-H_{s,1}}{z_{s,2}-z_{s,1}} \cdot (z_b - z_{s,1}) & z_{s,min} < z_b < z_{s,max} \\ H_b = H_{s,max} & z_b > z_{s,max} \end{cases} \tag{11.5}$$

where H_b [C] and H_s [C] are the interpolated band and the measured relative humidity, the subscripts $_{s,min}$ and $_{s,min}$ indicates quantity (elevation (z) and relative humidity (H)) related to the stations with minimum and maximum elevation, and the subscripts $_{s,1}$ and $_{s,2}$ indicates quantity (elevation (z) and relative humidity (H)) related to the stations is located between the centroid altimetric band. The interpolated hillslope centroid value is computed as the weigthted mean of the altimetric bands relative humidity with respect to the altimetric band area, as reported in eq.11.4

11.5.3 Amospheric pressure interpolation

Different interpolation algorithms are presented depending on N$_{as}$:

- N$_{as}$=0: a standard adiabatic is considered and for each band of the generic hillslope the interpolated atmospheric pressure is computed as:

$$P_b = 1013.25 \cdot e^{-0.00013 \cdot z_b} \tag{11.6}$$

where z_b [m] is the altimetric band elevation.

The interpolated hillslope centroid value is computed as the weigthted mean of the altimetric bands atmospheric pressure with respect to the altimetric band area:

$$P_h = \sum_{i=1}^{n_b} P_i \cdot \frac{A_i}{A_h} \tag{11.7}$$

where n_b is the number of altimetric bands of the hillslope, A_i[km^2] is the area of the i-th altrimetric band, P_i is the interpolated atmospheric pressure of the i-th altimetric band, and A_h[km^2] is the hillslope area.

- N$_{as}$=1: the standard atmospheric model it is considered; for each altimetric band b of the current hillslope the atmospheric pressure is computed using a simple adiabatic transformation:

$$P_b = P_s \cdot e^{-0.00013 \cdot (z_b - z_s)} \tag{11.8}$$

where: z_b[m] and z_s [m] are the centroid altimetric band and station elevation, P_b [mb] and P_s [mb] are the bands interpolated and the measured atmospheric pressure.

- N$_{as}$ >1: the standard atmospheric model it is considered for elevations lower than the minimum station elevation and higher than the maximum station elevation; for the points between the maximum and minimum station elevation, the laps rate is computeted:

$$\begin{cases} P_b = P_{s,min} - P_{s,min} \cdot 0.00013 \cdot (z_b - z_{s,min}) & z_b < z_{s,min} \\ P_b = P_{s,1} - \frac{P_{s,2} - P_{s,1}}{z_{s,2} - z_{s,1}} \cdot (z_b - z_{s,1}) & z_{s,min} < z_b < z_{s,max} \\ P_b = P_{s,max} - P_{s,max} \cdot 0.00013 \cdot (z_b - z_{s,max}) & z_b > z_{s,max} \end{cases} \tag{11.9}$$

where P_b [mb] and P_s [mb] are the interpolated bands and the measured atmospheric pressure, the subscripts $_{s,min}$ and $_{s,min}$ indicates quantity (elevation (z) and atmospheric pressure (P)) related to the stations with minimum and maximum elevation, and the subscripts $_{s,1}$ and $_{s,2}$ indicates quantity (elevation (z) and atmospheric pressure (P)) related to the stations located between the centroid altimetric band. The interpolated hillslope centroid value is computed as the weigthted mean of the altimetric band's atmospheric pressure with respect to the altimetric band area, as reported in eq.11.7

References

[1] JK Aase and SB Idso. A comparison of two formula types for calculating long-wave radiation from the atmosphere. *Water Resources Research*, 14(4):623–625, 1978. 66

[2] R.G. Allen, L.S. Pereira, D. Raes, M. Smith, et al. Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *FAO, Rome*, 300:6541, 1998. 66, 67

[3] R.M. Argent. A case study of environmental modelling and simulation using transplantable components. *Environmental Modelling & Software*, 20(12):1514–1523, 2005. 8

[4] RM Argent, J.M. Perraud, JM Rahman, RB Grayson, and GM Podger. A new approach to water quality modelling and environmental decision support systems. *Environmental Modelling & Software*, 24(7):809–818, 2009. 8

[5] Jeffrey G. Arnold, Raghavan Srinivasan, R. S. Muttiah, and J. R. Williams. Large area hydrologic modeling and assessment, Part I: Model development. *Journal of the American Water Resources Association*, 34(1):73–89, February 1998. ISSN 1093-474X. doi: 10.1111/j.1752-1688.1998.tb05961.x. 86, 87

[6] James C Ascough II, Dana L Hoag, W Marshall Frasier, and Gregory S McMaster. Computer use in agriculture: an analysis of great plains producers. *Computers and electronics in agriculture*, 23(3):189–204, 1999. 24

[7] Albert Baumgartner, Eberhard Reichel, and Richard Lee. *The world water balance: mean annual global, continental and maritime precipitation, evaporation and run-off*. Elsevier Scientific Publishing Company, 1975. 66

[8] Keith J Beven. Dalton medal lecture: How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, 5(1):1–12, 2001. 33

[9] KJ Beven. Calibration, validation and equifinality in hydrological modelling. *Model validation: Perspectives in hydrological science*, 43:55, 2001. 33

[10] K.J. Beven. *Rainfall-runoff modelling: the primer*. John Wiley & Sons Inc, 2001. ISBN 0470866713. 2, 87

[11] KJ Beven and MJ Kirkby. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal*, 24(1):43–69, 1979. 2

[12] R.E. Bird and R.L. Hulstrom. Simplified clear sky model for direct and diffuse insolation on horizontal surfaces. Technical report, Solar Energy Research Inst., Golden, CO (USA), 1981. 51

[13] Roger S Bivand, Edzer J Pebesma, and Virgilio Gómez-Rubio. *Applied spatial data analysis with R*. Springer, 2008. 36

[14] M. Blind, JB Gregersen, et al. Towards an open modelling interface (openmi) the harmonit project. *Advances in Geosciences*, 4:69–74, 2005. 9

[15] J. Boland, L. Scott, and M. Luther. Modelling the diffuse fraction of global solar radiation on a horizontal surface. *Environmetrics*, 12(2):103–116, 2001. 54

[16] A. Bos and A. de Vreng. Parameter optimization of the hymod model using scem-ua and moscem-ua. *Modelling Geo-Ecological Systems Computational Bio-and Physical Geography. University of Amsterdam, 43p*, 2006. 71

[17] G. Botter and A. Rinaldo. Scale effect on geomorphologic and kinematic dispersion. *Water resources research*, 39(10):1286, 2003. ISSN 0043-1397. 85

[18] D. P. Boyle. Multicriteria calibration of hydrological model. *Ph.D. dissertation, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson*, 2001. 70

[19] R.L. Bras and I. Rodríguez-Iturbe. *Random functions and hydrology*. Dover Pubns, 1994. ISBN 0486676269. 73, 83

[20] KAYE Brubaker, ALBERT Rango, and WILLIAM Kustas. Incorporating radiation inputs into the snowmelt runoff model. *Hydrological processes*, 10(10):1329–1343, 1996. 94

[21] E Brun, P David, M Sudul, and G Brunot. A nutnerical tnodel to sildulate snow-cover stratigraphy for operational avalanche forecasting. *Journal of Glaciology*, 38(128), 1992. 93

[22] David Brunt. Notes on radiation in the atmosphere. i. *Quarterly Journal of the Royal Meteorological Society*, 58(247):389–420, 1932. 66

[23] W. Brutsaert. On a derivable formula for long-wave radiation from clear skies. *Water Resources Research*, 11(5):742–744, 1975. 66, 70

[24] W. Brutsaert. *Hydrology: an introduction*. Cambridge Univ Pr, 2005. 66, 70

[25] Mikhail Ivanovich Budyko. *Climate and life*, volume 18. Academic press, 1974. 66

[26] RJC Burnash, RL Ferral, and RA McGuire. A generalized streamflow simulation system-Conceptual modeling for digital computers, report, Joint Fed. and State River Forecast Cent. *US Natl. Weather Serv./Calif. State Dept. of Water Resour., Sacramento, Calif*, 1973. 2

[27] P.A. Burrough, R.A. McDonnell, P.A. Burrough, and R. McDonnell. *Principles of geographical information systems*, volume 333. Oxford university press Oxford, 1998. 8

[28] F. Cazorzi and G. Dalla Fontana. Snowmelt modelling by combining air temperature and a distributed radiation index. *Journal of Hydrology*, 181(1-4):169–187, 1996. 94, 95, 96

[29] JP Chiles. P. delfiner. 1999. geostatistics: modeling spatial uncertainty. 37

[30] J.G. Corripio. Modelling the energy balance of high altitude glacierised basins in the central andes. *PhD dissertation, University of Edinburgh; 2002*, 2002. 50, 51, 52, 53, 55, 66, 95

[31] J.G. Corripio. Vectorial algebra algorithms for calculating terrain parameters from dems and solar radiation modelling in mountainous terrain. *International Journal of Geographical Information Science*, 17(1):1–24, 2003. 50, 51, 70, 95

[32] N.H. Crawford and R.K. Linsley. Digital simulation in hydrology Standford watershed model 4. 1966. 2

[33] N. Cressie. Statistics for spatial data. *Terra Nova*, 4(5):613–617, 1992. 34, 40

[34] G.P. Cressman. An operational objective analysis system. *Monthly Weather Review*, 87 (10):367–374, 1959. 34

[35] M DallAmico, S Endrizzi, S Gruber, and R Rigon. A robust and energy-conserving model of freezing variably-saturated soil. *The Cryosphere*, 5(2):469–484, 2011. 93

[36] O. David, S.L. Markstrom, K.W. Rojas, L.R. Ahuja, and I.W. Schneider. The object modeling system. *Agricultural system models in field research and technology transfer*, pages 317–331, 2002. 8, 11

[37] D.R. Dawdy and T. O'Donnell. Mathematical models of catchment behavior. *J. Hydraul. Div. ASCE*, 91(HY4):123–127, 1965. 104

[38] P. D'Odorico and R. Rigon. Hillslope and channel contributions to the hydrologic response. *Water resources research*, 39(5):1113, 2003. 2, 71, 85

[39] J.C.I. Dooge. A general theory of the unit hydrograph. *J. Geophys. Res*, 64(2):241–256, 1959. 2

[40] Q. Duan, S. Sorooshian, and V. Gupta. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4):1015–1031, 1992. 105

[41] Q. Duan, S. Sorooshian, and V.K. Gupta. Optimal use of the sce-ua global optimization method for calibrating watershed models. *Journal of Hydrology*, 158(3):265–284, 1994. 109

[42] Christopher J. Duffy. A two-state integral-balance model for soil moisture and groundwater dynamics in complex terrain. *Water Resour. Res.*, 32(8):2421–2434, 1996. URL `http://dx.doi.org/10.1029/96WR01049`. 70

[43] R.C. Eberhart and Y. Shi. Particle swarm optimization: developments, applications and resources. In *Proceedings of the 2001 congress on evolutionary computation*, volume 1, pages 81–86. Piscataway, NJ, USA: IEEE, 2001. 74, 97

[44] Russ Eberhart, Pat Simpson, and Roy Dobbins. *Computational intelligence PC tools*. Academic Press Professional, Inc., San Diego, CA, USA, 1996. ISBN 0-12-228630-8. 105

[45] Stefano Endrizzi. *Snow cover modelling at a local and distributed scale over complex terrain*. PhD thesis, Ph. D. dissertation. Dept. of Civil and Environmental Engineering, University of Trento, Italy, 2007. 93

[46] DG Erbs, SA Klein, and JA Duffie. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. *Solar Energy*, 28(4):293–302, 1982. 54

[47] Steven K Esbensen and Yochanan Kushnir. *The heat budget of the global ocean: An atlas based on estimates from surface marine observations*. Number 29. Oregon State University, 1981. 66

[48] W.A. Flügel. Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Br
"ol, Germany. *Hydrological Processes*, 9(3-4):423–436, 1995. ISSN 1099-1085. 24, 69

[49] EJ Førland, N.M. Institutt, and Nordic Working Group on Precipitation (NWGP). *Manual for operational correction of Nordic precipitation data*. Norwegian Meteorological Institute, 1996. 95

[50] G. Formetta, A. Antonello, S. Franceschi, O. David, and R. Rigon. Parameter optimization of the hymod model using scem-ua and moscem-ua. *Submitted to Environmental modelling and software*, 2006. 94

[51] G. Formetta, R. Mantilla, S. Franceschi, A. Antonello, and R. Rigon. The jgrass-newage system for forecasting and managing the hydrological budgets at the basin scale: models

of flow generation and propagation/routing. *Geoscientific Model Development*, 4(4):943–955, 2011. doi: 10.5194/gmd-4-943-2011. URL `http://www.geosci-model-dev.net/4/943/2011/`. 33, 55, 83, 84, 94

[52] G. Formetta, R. Rigon, J. L. Chávez, and O. David. Modeling short wave solar radiation using the jgrass-newage system. *Geoscientific Model Development Discussions*, 5(4):4355–4393, 2012. doi: 10.5194/gmdd-5-4355-2012. URL `http://www.geosci-model-dev-discuss.net/5/4355/2012/`. 94, 95, 96

[53] M. Franchini, J. Wendling, C. Obled, and E. Todini. Physical interpretation and sensitivity analysis of the TOPMODEL. *Journal of Hydrology*, 175(1-4):293–338, 1996. 2

[54] Josef Fürst and Thomas Hörhan. Coding of watershed and river hierarchy to support gis-based hydrological analyses at different scales. *Computers & Geosciences*, 35(3):688–696, 2009. 30

[55] H.J. Gardner, G. Manduchi, and Ltd MyiLibrary. *Design Patterns for E-science*. Springer, 2007. 8

[56] D.C. Garen and D. Marks. Spatially distributed energy balance snowmelt modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results. *Journal of Hydrology*, 315(1-4):126–153, 2005. 34, 35, 70

[57] D.C. Garen, G.L. Johnson, and C.L. Hanson. Mean areal precipitation for daily hydrologic modeling in mountainous regions1. *JAWRA Journal of the American Water Resources Association*, 30(3):481–491, 1994. 35

[58] Jonathan L Goodall and David R Maidment. A spatiotemporal data model for river basin-scale hydrologic systems. *International Journal of Geographical Information Science*, 23 (2):233–247, 2009. 23

[59] P. Goovaerts. *Geostatistics for natural resources evaluation*. Oxford University Press, USA, 1997. 34, 55, 70, 113

[60] P. Goovaerts. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of hydrology*, 228(1):113–129, 2000. 34

[61] R.B. Grayson, I.D. Moore, and T.A. McMahon. Physically based hydrologic modeling: 1. a terrain-based model for investigative purposes. *Water Resources Research*, 28(10): 2639–2658, 1992. 8

[62] W.H. Green and G.A. Ampt. Studies on soil physics, 1. The flow of air and water through soils. *Journal of Agricultural Sciences*, 4:11–24, 1911. 87

[63] H.V. Gupta, H. Kling, K.K. Yilmaz, and G.F. Martinez. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2):80–91, 2009. 57, 82, 97

[64] V.K. Gupta and CT Waymire. A representation of an instantaneous unit hydrograph from geomorphology. *Water Resources Research*, 16(5):855–862, 1980. 2

[65] G.H. Hargreaves and Z.A. Samani. Estimation of potential evapotranspiration. *Journal of Irrigation and Drainage Division, Proceedings of the American Society of Civil Engineers*, 108:223–230, 1982. 87

[66] L.E. Hay, G.H. Leavesley, M.P. Clark, S.L. Markstrom, R.J. Viger, and M. Umemoto. Step wise, multiple objective calibration of a hydrologic model for a snowmelt dominated basin1. *JAWRA Journal of the American Water Resources Association*, 42(4):877–890, 2006. 82, 103, 109

[67] N. Helbig, H. Löwe, B. Mayer, and M. Lehning. Explicit validation of a surface shortwave radiation balance model over snow-covered complex terrain. *Journal of Geophysical ResearchAtmospheres*, 115:D18113, 2010. 53

[68] D.V. Hoyt. A model for the calculation of solar global insolation. *Solar Energy*, 21(1): 27–35, 1978. 53

[69] G. Hudson and H. Wackernagel. Mapping temperature using kriging with external drift: theory and an example from scotland. *International journal of Climatology*, 14(1):77–91, 1994. 34

[70] R.P. Ibbitt. *Systematic parameter fitting for conceptual models of catchment hydrology.* Hydrology Program, Colorado State University, Engineering Research Center, 1970. 105

[71] R.P. Ibbitt and T. ODonnell. Fitting methods for conceptual catchment models. *Journal of the Hydraulics Division*, 97(9):1331–1342, 1971. 105

[72] M. Iqbal. An introduction to solar radiation. 1983. 49, 51, 52, 53, 55

[73] C.H. Jarvis and N. Stuart. A comparison among strategies for interpolating maximum and minimum daily air temperatures. part i: The selection of guiding topographic and land cover variables. *Journal of Applied Meteorology*, 40(6):1060–1074, 2001. 70

[74] S.K. Kampf and S.J. Burges. A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water resources research*, 43(5):W05423, 2007. ISSN 0043-1397. 2

[75] D. Kavetski, G. Kuczera, and S.W. Franks. Calibration of conceptual hydrological models revisited: 1. overcoming numerical artefacts. *Journal of Hydrology*, 320(1):173–186, 2006. 95

[76] J. Kennedy. Swarm intelligence. *Handbook of nature-inspired and innovative computing*, pages 187–219, 2006. 105

[77] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 4, pages 1942–1948. IEEE, 1995. 94, 97, 103, 105

[78] T. Kokkonen, H. Koivusalo, T. Jakeman, and J. Norton. Construction of a degree–day snow model in the light of the ten iterative steps in model development. In *Proceedings of the iEMSs Third Biennial Meeting: Summit on Environmental Modelling and Software. Environmental Modelling and Software Society, Burlington, USA*, 2006. 94, 95

[79] P. Krause. *Das hydrologische Modellsystem J2000: Beschreibung und Anwendung in großen Flußgebieten.* Forschungszentrum, Zentralbibliothek, 2001. 91

[80] P. Krause. Quantifying the impact of land use changes on the water balance of large catchments using the J2000 model. *Physics and Chemistry of the Earth, Parts A/B/C*, 27(9-10):663–673, 2002. ISSN 1474-7065. 24, 69

[81] William P Kustas, Albert Rango, and Remko Uijlenhoet. A simple energy budget algorithm for the snowmelt runoff model. *Water Resources Research*, 30(5):1515–1527, 1994. 94

[82] GH Leavesley, RW Lichty, BM Troutman, and LG Saindon. Precipitation-runoff modeling system: User's manual. *Available from Books and Open File Report Section, USGS Box 25425, Denver, Co 80225. USGS Water Resources Investigations Report 83-4238, 1983. 207 p, 54 fig, 15 tab, 51 ref, 8 attach.*, 1983. 2, 91, 103

[83] D.R. Legates and G.J. McCabe Jr. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1):233–241, 1999. 109

[84] Michael Lehning, Ingo Völksch, David Gustafsson, Tuan Anh Nguyen, Manfred Stähli, and Massimiliano Zappa. Alpine3d: a detailed model of mountain surface processes and its application to snow hydrology. *Hydrological Processes*, 20(10):2111–2128, 2006. 93

[85] L.B. Leopold and T. Maddock. The hydraulic geometry of stream channels and some physiographic implications, Professional Paper 252. *US Geological Survey, Washington, DC*, 1953. 2

[86] K.N. Liou. *An introduction to atmospheric radiation*, volume 84. Academic press, 2002. 49

[87] Jiajia Liu, Zuhao Zhou, Yangwen Jia, Hao Wang, and Xiangdong Chen. A stem-branch-topological codification for watershed subdivision and identification to support distributed hydrological modeling at large river basins. *Hydrological Processes*, 2013. 24

[88] CD Lloyd. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in great britain. *Journal of Hydrology*, 308(1-4):128–150, 2005. 8

[89] D.R. Maidment. Gis and hydrologic modeling-an assessment of progress. In *Proceedings of the Third International Conference on Integrating GIS and Environmental Modelling*, 1996. 8

[90] D.R. Maidment. *Arc Hydro: GIS for water resources*, volume 1. ESRI press, 2002. 23

[91] P.V. Mandapaka, W.F. Krajewski, R. Mantilla, and V.K. Gupta. Dissecting the effect of rainfall variability on the statistical structure of peak flows. *Advances in Water Resources*, 32(10):1508–1525, 2009. 83

[92] R. Mantilla. Physical basis of statistical self-similarity in peak ows on random self-similar networks. *PhD dissertation, University of Colorado, Boulder; 2007*, 2001. 83

[93] R. Mantilla and V.K. Gupta. A GIS numerical framework to study the process basis of scaling statistics in river networks. *IEEE Geoscience and Remote Sensing Letters*, 2(4): 404–408, 2005. 73, 74

[94] R. Mantilla, V.K. Gupta, et al. Role of coupled flow dynamics and real network structures on hortonian scaling of peak flows. *Journal of Hydrology*, 322(1-4):155–167, 2006. 73

[95] D. Marechal. A soil-based approach to rainfall-runoff modelling in ungauged catchments for England and Wales. 2004. 108

[96] Danny Marks, James Domingo, Dave Susong, Tim Link, and David Garen. A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological Processes*, 13(1213):1935–1959, 1999. 93

[97] J Martinec. Snowmelt-runoff model for stream flow forecasts. *Nordic hydrology*, 6(3): 145–154, 1975. 93

[98] Jaroslav Martinec, Albert Rango, and E Major. The snowmelt-runoff model(srm) user's manual. 1983. 93

[99] Jaroslav Martinec, Albert Rango, Ralph Roberts, Michael F Baumgartner, and Gabriela M Apfl. *Snowmelt runoff model (SRM) user's manual*. Geographisches Institut der Universität, 1994. 93

[100] T. Maxwell and R. Costanza. A language for modular spatio-temporal simulation. *Ecological modelling*, 103(2-3):105–113, 1997. 8

[101] M. Menabde and M. Sivapalan. Linking space-time variability of river runoff and rainfall fields: a dynamic approach. *Advances in water resources*, 24(9-10):1001–1014, 2001. 74

[102] D.B. Michelson. Systematic correction of precipitation gauge observations using analyzed meteorological variables. *Journal of hydrology*, 290(3):161–177, 2004. 95

[103] J.L. Monteith. Evaporation and environment. In G.E. Fogg, editor, *Symposium of the Society for Experimental Biology, The State and Movement of Water in Living Organisms, Vol. 19*, pages 205–234, New York, 1965. Academic Press, Inc. 87

[104] RJ Moore. The probability-distributed principle and runoff production at point and basin scales. *Hydrol. Sci. J.*, 30(1):165, 1985. 70, 71, 86

[105] R.V. Moore and C.I. Tindall. An overview of the open modelling interface and environment (the openmi). *Environmental Science & Policy*, 8(3):279–286, 2005. 8

[106] JC Morrill, RE Dickinson, and AN Hahmann. Sensitivity of a land surface model to the diurnal distribution of downward longwave radiation. *Journal of the Meteorological Society of Japan*, 77(1B):265–279, 1999. 66

[107] R Moussa, F Colin, C Dagès, JC Fabre, P Lagacherie, X Louchart, M Rabotin, D Raclot, and M Voltz. Distributed hydrological modelling of farmed catchments (mhydas): assessing the impact of man-made structures on hydrological processes. 24

[108] T.J Mulvaney. On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and of flood discharges in a given catchment. *Transactions and Minutes of the Proceeding of the Institute of Civil Engineers of Ireland.*, 1851. 2

[109] S. L. Neitsch, Jeffrey G. Arnold, J. R. Kiniry, and J. R. Williams. Soil and Water Assessment Tool (SWAT) Theoretical Documentation, Version 2005, (BRC Report 02-05), 2005. 86

[110] M. Neteler and H. Mitasova. Open source gis: A grass gis approach. the international series in engineering and computer science, 2008. 10

[111] S.V. Ollinger, J.D. Aber, G.M. Lovett, S.E. Millham, R.G. Lathrop, and J.M. Ellis. A spatial model of atmospheric deposition for the northeastern us. *Ecological Applications*, 3(3):459–472, 1993. 34

[112] JF Orgill and KGT Hollands. Correlation equation for hourly diffuse radiation on a horizontal surface. *Solar energy*, 19(4):357–359, 1977. 54

[113] Stefano Orlandini, Giovanni Moretti, Marco Franchini, Barbara Aldighieri, and Bruno Testa. Path-based methods for the determination of nondispersive drainage directions in grid-based digital elevation models. *Water resources research*, 39(6):1144, 2003. 26, 30

[114] Independent Panel. Report to us department of interior and state of idaho on failure of teton dam. *US Government Printing Office, Washington, DC*, 1976. 19

[115] Edzer J. Pebesma. Multivariable geostatistics in s: the gstat package. *Computers and Geosciences*, 30:683–691, 2004. 36, 39, 113

[116] D.L. Phillips, J. Dolph, and D. Marks. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology*, 58(1):119–141, 1992. 34

[117] AJ Prata. A new long-wave formula for estimating downward clear-sky radiation at the surface. *Quarterly Journal of the Royal Meteorological Society*, 122(533):1127–1151, 1996. 52

[118] C. H. B. Priestley and R. J. Taylor. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, 100(2):81–92, February 1972. ISSN 0027-0644. doi: 10.1175/1520-0493(1972)100⟨0081:OTAOSH⟩2.3.CO;2. URL `http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2`. 87

[119] C.H.B. Priestley. *Turbulent transfer in the lower atmosphere*. University of Chicago Press Chicago, Ill., 1959. 67

[120] CHB Priestley and RJ Taylor. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly weather review*, 100(2):81–92, 1972. 67

[121] Gauthier Quesnel, Raphaël Duboz, and Éric Ramat. The Virtual Laboratory Environment – An operational framework for multi-modelling, simulation and analysis of complex dynamical systems. *Simulation Modelling Practice and Theory*, 17:641–653, April 2009. 7

[122] Timothy J Randle. *Geomorphology and river hydraulics of the Teton River upstream of Teton Dam, Teton River, Idaho*. US Department of the Interior, Bureau of Reclamation, 2000. 19

[123] S. Reed, V. Koren, M. Smith, Z. Zhang, F. Moreda, D.J. Seo, and P. DMIP. Overall distributed model intercomparison project results. *Journal of Hydrology*, 298(1-4):27–60, 2004. 2

[124] DT Reindl, WA Beckman, and JA Duffie. Diffuse fraction correlations. *Solar Energy*, 45 (1):1–7, 1990. 54, 61

[125] R. Rigon. The horton machine: a system for dem analysis : the reference manual. 2006. 11

[126] R. Rigon, G. Bertoldi, and T.M. Over. GEOtop: A distributed hydrological model with coupled water and energy budgets. *Journal of Hydrometeorology*, 7(3):371–388, 2006. ISSN 1525-7541. 2, 70

[127] R. Rigon, E. Ghesla, C. Tiso, and Cozzini A. The horton machine: a system for dem analysis : the reference manual. 2006. 25

[128] R. et al. Rigon. The horton machine: a system for dem analysis. 2013. 25

[129] Riccardo Rigon, Ignacio Rodriguez-Iturbe, Amos Maritan, Achille Giacometti, David G Tarboton, and Andrea Rinaldo. On hack's law. *Water Resources Research*, 32(11):3367–3374, 1996. 27

[130] A. Rinaldo and I. Rodriguez-Iturbe. Geomorphological theory of the hydrological response. *Hydrological processes*, 10(6):803–829, 1996. 2

[131] A. Rinaldo, A. Marani, R. Rigon, et al. Geomorphological dispersion. *Water Resour. Res*, 27(4):513–525, 1991. 2

[132] AE Rizzoli, MGE Svensson, E. Rowe, M. Donatelli, RM Muetzelfeldt, T. Wal, F.K. Evert, and F. Villa. *Modelling framework (SeamFrame) requirements*. SEAMLESS, 2005. 7, 8

[133] I. Rodríguez-Iturbe and J.B. Valdés. The geomorphologic structure of hydrologic response. *Water Resources Research*, 15(6):1409–1420, 1979. 2

[134] Ignacio Rodríguez-Iturbe and Amilcare Porporato. *Ecohydrology of water-controlled ecosystems: soil moisture and plant dynamics*. Cambridge University Press, 2005. 31

[135] BB Ross, DN Contractor, and VO Shanholtz. A finite-element model of overland and channel flow for assessing the hydrologic impact of land-use change. *Journal of Hydrology*, 41(1-2):11–30, 1979. ISSN 0022-1694. 24, 69

[136] R. Rosso. Nash model relation to Horton order ratios. *Water Resources Research*, 20(7): 914–920, 1984. 2

[137] F. Rubel and M. Hantel. Correction of daily rain gauge measurements in the baltic sea drainage basin. *Nordic Hydrology*, 30(3):191–208, 1999. 95

[138] L.K. Sherman. Streamflow from rainfall by the unit-graph method. *Engineering News Record*, 108(14):501–505, 1932. 2

[139] RO Slatyer and IC McIlroy. Practical microclimatology. *Practical Microclimatology.*, 1961. 67

[140] S. Sorooshian, Q. Duan, and V.K. Gupta. Calibration of rainfall-runoff models: application of global optimization to the sacramento soil moisture accounting model. *Water Resources Research*, 29(4):1185–1194, 1993. 105

[141] JW Spencer. Fourier series representation of the position of the sun. *Search*, 2(5):172, 1971. 50

[142] K. Stahl, RD Moore, JA Floyer, MG Asplin, and IG McKendry. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agricultural and Forest Meteorology*, 139(3):224–236, 2006. 34

[143] A. Stehr, P. Debels, F. Romero, and H. Alcayaga. Hydrological modelling with swat under conditions of limited data availability: evaluation of results from a chilean case study. *Hydrological sciences journal*, 53(3):588–601, 2008. 63, 98

[144] Gil Strassberg, Norman L Jones, and David R Maidment. *Arc Hydro Groundwater: GIS for Hydrogeology*. Esri Press, 2011. 24

[145] Z Su. The surface energy balance system (sebs) for estimation of turbulent heat fluxes. *Hydrology and Earth System Sciences Discussions*, 6(1):85–100, 2002. 66

[146] David G Tarboton, Charles H Luce, et al. *Utah energy balance snow accumulation and melt model (UEB)*. Citeseer, 1996. 93

[147] A.H. Thiessen. Precipitation averages for large areas. *Monthly weather review*, 39(7): 1082–1089, 1911. 34

[148] E. Todini. The arno rainfall–runoff model. *Journal of Hydrology*, 175(1-4):339–382, 1996. 71

[149] Soil Conservation Service USDA. National Engineering Handbook. chapter Section 4. U.S. Dept. of Agriculture, Soil Conservation Service, Washington, DC, 1972. 87

[150] M.W. Van Liew, JG Arnold, and DD Bosch. Problems and potential of autocalibrating a hydrologic model. 2005. 63, 98

[151] KL Verdin and JP Verdin. A topological system for delineation and codification of the Earth's river basins. *Journal of Hydrology*, 218(1-2):1–12, 1999. ISSN 0022-1694. 30

[152] D. Viviroli, M. Zappa, J. Gurtz, and R. Weingartner. An introduction to the hydrological modelling system PREVAH and its pre-and post-processing-tools. *Environmental Modelling & Software*, 24(10):1209–1222, 2009. ISSN 1364-8152. 24, 69

[153] J.A. Vrugt and B.A. Robinson. Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences*, 104(3): 708–711, 2007. 105

[154] J.A. Vrugt, C.J.F. ter Braak, M.P. Clark, J.M. Hyman, and B.A. Robinson. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation. *Water Resources Research*, 44(12):W00B09, 2008. 80

[155] H. Wackernagel. *Multivariate geostatistics: an introduction with applications*. Springer Verlag, 2003. 34

[156] C.G. WESSELUNG, D.J.A.N. KARSSENBERG, P.A. BURROUGH, and W. DEURSEN. Integrating dynamic environmental models in gis: the development of a dynamic modelling language. *Transactions in GIS*, 1(1):40–48, 2007. 8

[157] J. R. Williams. Flood routing with variable travel time or variable storage coefficients. *Trans. ASAE*, 12(1):100–103, 1969. 87

[158] C.J. Willmott, S.G. Ackleson, R.E. Davis, J.J. Feddema, K.M. Klink, D.R. Legates, J. ODonnell, and C.M. Rowe. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research*, 90(C5):8995–9005, 1985. ISSN 0148-0227. 84, 109

[159] John P Wilson and John C Gallant. *Terrain analysis: principles and applications*. Wiley, 2000. 25

[160] Fabrizio Zanotti, Stefano Endrizzi, Giacomo Bertoldi, and Riccardo Rigon. The geotop snow module. *Hydrological Processes*, 18(18):3667–3679, 2004. 93, 94

Declaration

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such.

Trento (Italy), 20-02-2013