



Topic 1. Civil and Environmental Engineering

Marialaura Bancheri

**A flexible approach to the estimation  
of water budgets and its connection  
to the travel time theory**



Credits of the cover image: Cache La Poudre River, Fort Collins, CO, USA



Except where otherwise noted, contents on this book are licensed under a Creative  
Common Attribution - Non Commercial - No Derivatives  
4.0 International License

ISBN (paper): ..... ; ISBN (online): .....

University of Trento  
Doctoral School in Civil, Environmental and Mechanical Engineering  
<http://web.unitn.it/en/dricam>  
Via Mesiano 77, I-38123 Trento  
Tel. +39 0461 282670 / 2611 - [dicamphd@unitn.it](mailto:dicamphd@unitn.it)

UNIVERSITY OF TRENTO - Italy

Department of Civil, Environmental  
and Mechanical Engineering



Doctoral School in Civil, Environmental and Mechanical Engineering

Topic 1. Civil and Environmental Engineering

Doctoral Thesis - November 2017

MARIALAURA BANCHERI

# **A flexible approach to the estimation of water budgets and its connection to the travel time theory**

**Supervisors:**

Riccardo RIGON (University of Trento, Italy)

Giuseppe FORMETTA (Centre for Ecology and Hydrology, Wallingford,  
United Kingdom)



*Your speed doesn't matter,  
forward is forward.*



## ABSTRACT

The increasing impacts of climate changes on water related sectors are leading the scientists' attentions to the development of comprehensive models, allowing better descriptions of the water and solute transport processes. "Getting the right answers for the right reasons", in terms of hydrological response, is one of the main goals of most of the recent literature. Semi-distributed hydrological models, based on the partition of basins in hydrological response units (HRUs) to be connected, eventually, to describe a whole catchment, proved to be robust in the reproduction of observed catchment dynamics. 'Embedded reservoirs' are often used for each HRU, to allow a consistent representation of the processes. In this work, a new semi-distributed model for runoff and evapotranspiration is presented: five different reservoirs are inter-connected in order to capture the dynamics of snow, canopy, surface flow, root-zone and groundwater compartments.

The knowledge of the mass of water and solute stored and released through different outputs (e.g. discharge, evapotranspiration) allows the analysis of the hydrological travel times and solute transport in catchments. The latter have been studied extensively, with some recent benchmark contributions in the last decade. However, the literature remains obscured by different terminologies and notations, as well as model assumptions are not fully explained. The thesis presents a detailed description of a new theoretical approach that reworks the theory from the point of view of the hydrological storages and fluxes involved. Major aspects of the new theory are the 'age-ranked' definition of the hydrological variables, the explicit treatment of evaporative fluxes and of their influence on the

transport, the analysis of the outflows partitioning coefficients and the explicit formulation of the 'age-ranked' equations for solutes. Moreover, the work presents concepts in a new systematic and clarified way, helping the application of the theory.

To give substance to the theory, a small catchment in the prealpine area was chosen as an example and the results illustrated.

The new semi-distributed model for runoff and evapotranspiration and the travel time theory were implemented and integrated in the semi-distributed hydrological system JGrass-NewAge. Thanks to the environmental modeling framework OMS3, each part of the hydrological cycle is implemented as a component that can be selected, adopted, and connected at run-time to obtain a user-customized hydrological model. The system is flexible, expandable and applicable in a variety of modeling solutions. In this work, the model code underwent to an extensive revision: new components were added (coupled storages water budget, travel times components); old components were enhanced (Kriging, shortwave, longwave, evapotranspiration, rain-snow separation, SWE and melting components); documentation was standardized and deployed.

Since the Thesis regards in wide sense the building of a collaborative system, a discussion of some general purpose tools that were implemented or improved for supporting the present research is also presented. They include the description and the verification of a software component dealing with the long-wave radiation budget and another component dealing with an implementation of some Kriging procedure.



## ACKNOWLEDGEMENTS

This work was a collective effort of persons from all over the world. That's way I decided to write the acknowledgments both in Italian and in English, to express my gratitude in the best way to all my "supporters". I miei ringraziamenti non possono non iniziare dalla persona che mi ha insegnato tutto quello (ancora poco) che so sull'Idrologia, il Professore Riccardo Rigon. La mia fortuna nell'aver trovato un Professore come Riccardo non e' completamente descrivibile a parole. E' stato per me un padre accademico, che mi ha fatto crescere dal punto di vista professionale, insegnandomi tutto quello che ha potuto, trasmettendomi la sua passione e la sua dedizione per il lavoro, spronandomi e incoraggiandomi sempre a dare il meglio e a crederci sempre. Le "invidie" dei miei colleghi dottorandi per un Supervisor cosi' attento e disponibile, sono state per me sempre oggetto di vanto. Non posso non ringraziarlo anche dal punto di vista personale, in quanto si e' dimostrato un esempio di padre, una persona che si e' sempre preoccupata per me e dei miei stati d'animo in quattro anni lontani da affetti e famiglia. Riccardo e la sua famiglia sono stati un punto di riferimento saldo e per questo, la mia gratitudine nei loro confronti e' smisurata.

Vorrei poi ringraziare Giuseppe Formetta, mio secondo Supervisor e fratello accademico. Anche per lui i miei ringraziamenti sono sinceri e profondi, per tutto il supporto che mi ha dato e per avermi permesso di entrare in questo mondo "NewAge", nel migliore dei modi.

Come dicevo, questa tesi e' stato lavoro collettivo di un team di ricerca che si e' consolidato in questi anni e che si e' evoluto in una grande famiglia. I miei ringraziamenti vanno a Francesco Serafin, che mi ha trasmesso tutto il suo sapere informatico, che mi e' stato di grandissimo supporto in momenti chiave del mio

lavoro e con cui ho condiviso pensieri, scopi, arrabbiature e successi. Ringrazio poi Michele Bottazzi, che nonostante sia entrato da poco nella mia vita accademica e personale, lo ha fatto in modo così prepotente, da essersi conquistato un ruolo fondamentale nel mio percorso e nelle mie amicizie. Voglio ringraziare anche Daniele, Stefano e Niccolò, per aver deciso di far parte di questa famiglia e per i momenti condivisi in ufficio in questi anni. Least but not last, I would like to thank Wuletawu Abera, for his precious advises and the help with the Posina catchment modelling.

One year spent in US changed my life and perspectives and it is mandatory to me to thank Dr. Timothy Green, Olaf David and their beloved families. They made my stay in the other part of the world the most beautiful, easiest and best growing experience in my life. I was immediately part of a big family and I will always miss them, from the personal point of view. I hope that I will never miss them from the academical point of view! A sincere thought and thanks go also to Jim, wherever he is right now.

I miei ringraziamenti includono anche il Professore Manfreda e il suo team, Caterina, Leonardo, Silvano e Luciana. Sempre disponibili, mi hanno aiutato e supportato come se fossi stata parte della loro famiglia da molti anni.

Ringrazio il Professore Marco Borga, la Ricercatrice Giulia Zuecco e il Professore Daniele Penna, per avermi gentilmente concesso i loro dati, per i loro preziosi consigli e per la loro infinita disponibilita' nei miei confronti.

I am also thankful to the reviewers of the Thesis, Professor Kirchner and Professor Botter, for their review and their precious advices to improve my work.

Infine ringrazio la scuola di dottorato, soprattutto Marina e Claudia, per il loro supporto in questi anni.

Una nota frase dice: "Quando arrivi al Sud piangi due volte: quando arrivi e quando te ne vai". Io la giro su Trento e dico "Quando arrivi a Trento piangi due volte: quando arrivi e quando te ne vai". Questo e' quello che mi e' successo. La disperazione totale di dovermi trasferire e' stata sostituita da quattro anni indimenticabili, fatti di amicizie sincere. Ringrazio la amica e colonna Elisa, con cui ho condiviso momenti felici e spensierati, che mi ha aiutato in momenti bui e che mi ha

---

supportato in scelte importanti. Ringrazio poi Ines, Karina, Simone, Elena, Stefano, Erica, Fabiano, Oscar, Nadia, Giorgia per essermi stati sempre vicini e per aver condiviso con me tanti momenti speciali.

Mi sono sempre ritenuta una persona fortunata, circondata da tante persone uniche e sempre presenti per me. Ringrazio la mia famiglia che mi ha supportato in questi anni. Mia Mamma, mia sorella, mio cognato e i miei stupendi nipoti, con cui ho condiviso i miei successi accademici, che hanno sopportato i miei malumori, le mie domeniche a lavoro e i miei tempi stretti, senza mai aver smesso di credere in me. Ringrazio la mia nuova famiglia, i miei suoceri e Simona, per avermi concesso tanti momenti spensierati insieme. Ringrazio la papessa Nunzia, sempre presente e sicurezza costante, Francesca la svitata, per tutti i momenti condivisi in giro per il mondo e i miei amici, tutti. Infine ringrazio mio marito Francesco, che, nonostante "la modellazione idrologica non sia proprio il tuo forte", ha sempre creduto in me. Questa tesi e' anche sua, per tutti i consigli tecnici che mi ha dato, per tutte le equazioni differenziali che mi ha aiutato a risolvere, per le decine di volte in cui mi ha convinto a non mollare, per le decine di volte in cui ha preso un treno per donarmi la sua presenza. La mia fortuna e' anche e soprattutto lui.



## TABLE OF CONTENTS

	<b>Page</b>
<b>List of Tables</b>	<b>xiii</b>
<b>List of Figures</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Preliminary on mathematics and informatics</b>	<b>7</b>
2.1 Object Modelling System 3 . . . . .	9
2.2 The JGrass-NewAge hydrological modeling system	13
2.3 JGrass-NewAGE refactoring . . . . .	16
2.4 GEOframe: a system for doing hydrology by com- puter . . . . .	18
2.5 Time continuous Petri Nets . . . . .	24
2.6 The Net3 infrastructure . . . . .	28
2.7 Final remarks . . . . .	31
<b>3 A semi-distributed model for runoff and evapotran- spiration</b>	<b>33</b>
3.1 Embedded reservoir model . . . . .	37
3.1.1 Snow storage . . . . .	41
3.1.2 Canopy storage . . . . .	43
3.1.3 Root zone storage . . . . .	45
3.1.4 Surface runoff . . . . .	47
3.1.5 Groundwater . . . . .	49
3.2 Final remarks . . . . .	49

<b>4</b>	<b>Travel times theory</b>	<b>51</b>
4.1	The geomorphological unit hydrograph from a historical-critical perspective . . . . .	52
4.1.1	Early contributions . . . . .	52
4.1.2	The rise of the width function based geomorphological approach (WFIUH) . . . . .	58
4.2	Age ranked approach . . . . .	65
4.2.1	Definitions of age-ranked quantities . . . . .	69
4.2.2	Backward and forward approaches . . . . .	72
4.2.3	Backward Probabilities . . . . .	73
4.2.4	Forward Probabilities . . . . .	82
4.2.5	The partitioning coefficient $\Theta$ . . . . .	85
4.2.6	Niemi's relation . . . . .	86
4.2.7	Residence times, travel times and life expectancy . . . . .	88
4.2.8	Passive and reactive solutes . . . . .	90
4.2.9	A simple example where probabilities are assigned instead than derived. . . . .	95
4.2.10	Extension to the embedded reservoirs model	99
4.2.11	Final Remarks . . . . .	103
<b>5</b>	<b>Posina application: embedded reservoir model and travel times results</b>	<b>105</b>
5.1	Study area and model setup . . . . .	105
5.1.1	Parameter calibration and verification . . . . .	116
5.2	Results and discussion . . . . .	117
5.2.1	Hourly time step . . . . .	118
5.2.2	Daily time step . . . . .	128
5.3	Less explored and unexplored MS . . . . .	142
5.4	Final remarks . . . . .	145
<b>6</b>	<b>A couple of new ancillary components</b>	<b>149</b>
6.1	LWRB . . . . .	150

6.1.1	Methodology . . . . .	152
6.1.2	The study area: the AmeriFlux Network . . .	161
6.1.3	Results . . . . .	163
6.1.4	Final remarks . . . . .	172
6.2	Kriging . . . . .	174
6.2.1	A little of Kriging theory . . . . .	176
6.2.2	Structure of the SI package . . . . .	179
6.2.3	Study area . . . . .	181
6.2.4	Results . . . . .	183
6.2.5	Final remarks . . . . .	193
<b>7</b>	<b>Conclusions</b>	<b>195</b>
<b>A</b>	<b>Appendix</b>	<b>205</b>
A.1	Calibration algorithms in OMS3 . . . . .	205
A.2	An example of implementation of the Simple Fac- tory Design Pattern . . . . .	208
A.3	Some indices of goodness of fit. . . . .	210
A.4	Symbols, Acronyms, and Notation . . . . .	211
A.5	An example of sim file of the embedded reservoir model for a single HRU . . . . .	214
A.6	List of semivariogram model implemented in SI . .	220
A.7	Kriging dataset: the Isarco River Basin . . . . .	222
A.8	Reproducible research . . . . .	225
	<b>Bibliography</b>	<b>227</b>





## LIST OF TABLES

<b>TABLE</b>	<b>Page</b>
2.1 List of JGrass-NewAge components and associated references. . . . .	14
2.2 Example of table of association between fluxes and their expressions to add to the the Petri nets representation. . . . .	27
2.3 Example of table of association between fluxes and their expressions to add to the the Petri nets representation. . . . .	27
3.1 Table of association between fluxes and expressions related to the snow storage. . . . .	42
3.2 List of symbols, names and units used in the snow storage representation. . . . .	42
3.3 Table of association between fluxes and expressions related to the canopy storage. . . . .	44
3.4 List of symbols, names and units used in the canopy storage representation. . . . .	45
3.5 Table of association between fluxes and expressions related to the root zone storage. . . . .	46
3.6 List of symbols, names and units used in the root zone storage representation. . . . .	46
3.7 Table of association between fluxes and expressions related to the surface runoff storage. . . . .	48

3.8	List of symbols, names and units used in the surface runoff storage representation. . . . .	48
3.9	Table of association between fluxes and expressions related to the groundwater storage. . . . .	49
3.10	List of symbols, names and units used in the groundwater storage representation. . . . .	50
5.1	List of meteorological and discharge stations in the Posina basin. Discharge measurement stations are identified with a *. . . . .	109
5.2	List of parameters of the embedded reservoir model: the components associated to each parameter are specified in the first column, while the calibration ranges are specified in the third column. . . . .	117
5.3	Results in terms of goodness of fit in calibration and validation periods for the two main stations, Bazzoni and Stancari. . . . .	119
5.4	List of optimized values for each model parameters in the hourly time-step case. . . . .	124
5.5	Basic statistics of the residence times, mean travel times and evapotranspiration times in hourly time-step case. . . . .	127
5.6	Results in terms of goodness of fit in calibration and validation periods for the two main stations, Bazzoni and Stancari. No calibration was performed at Bazzoni, therefore only the GOF in validation period are shown. . . . .	130
5.7	List of optimized values for each model parameters in the daily time-step case. . . . .	138
5.8	Basic statistics of the residence times, mean travel times and evapotranspiration times in daily time-step case. . . . .	140

6.1	Clear sky emissivity formulations: $T_a$ is the air temperature [K], $w$ [ $\text{kg}/\text{m}^2$ ] is precipitable water = $4650 [e_0/T_a]$ and $e$ [kPa] is screen-level water-vapor pressure. The models follow the formulations presented in used in Flerchinger (2000). The Angstrom and Brunt model was presented as cited by Niemelä <i>et al.</i> (2001). Konzelmann uses water vapor pressure in [Pa] not [kPa].	155
6.2	Model parameter values as presented in their literature formulation. . . . .	155
6.3	Soil emissivity for surface types (Brutsaert, 2005). . . .	156
6.4	Some general and climatic characteristics of the sites used for calibration: elevation is the site elevation above sea level, $T$ is the annual average temperature, and data period refers to the period of available measurements. . . . .	162
6.5	Results in terms of goodness of fit indices of the fitting between the experimental and the 10 theoretical semi-variograms. All the models shown a good agreement: Bessel model proved to be the best while the Power the worst. . . . .	185
6.6	Results in terms of goodness of fit indices of the fitting between the experimental and the 10 theoretical semi-variograms. All the models shown a good agreement: Bessel model proved to be the best while the Power the worst. . . . .	190
A.1	List of symbols, names and units used in the thesis. . . .	214



## LIST OF FIGURES

FIGURE	Page
2.1 Example of sim file in which are detailed the 3 main fields: <i>component</i> where are specified the component executable to use, <i>parameters</i> where are specified the component parameters, <i>connect</i> where are specified how to connect the components, according to the <i>out-to-in</i> schema . . . . .	12
2.2 Reproducibility VS replicability of reasearch. Figure adapted from <a href="http://rrcns.readthedocs.io/en/latest/reproducible_research.html">http://rrcns.readthedocs.io/en/latest/reproducible_research.html</a> .	19
2.3 GEOframe organization logo. . . . .	21
2.4 Linear reservoir and non linear reservoir representations using the time-varying Petri nets. . . . .	25
2.5 Two coupled storages, upper and the lower, represented using the time-varying Petri Nets. . . . .	26
2.6 ArcHydro framework input data. . . . .	30
2.7 Posina river network schematization according to the River Net3 component: nodes are connected in cascade to the river outlet and simulations are launched according. . . . .	30
3.1 The figure represents the trade-off between the model complexity and the data availability. The figure was adapted from Grayson & Blöschl (2001). . . . .	35

3.2	Representation of the embedded reservoir model using time-varying Petri-Nets. Five components are storage, snow, canopy, root zone, surface flow, and groundwater, which are represented through circles of different colors and specifications. Snow storage is represented using two overlapped circles, since it solved two coupled ODEs. Each storage has been implemented into a different Jgrass-NewAge component. . . . .	40
3.3	Representation of the snow reservoir model using time-varying Petri-Nets. . . . .	41
3.4	Representation of the canopy reservoir model using time-varying Petri-Nets. . . . .	44
3.5	Representation of the root zone reservoir model using time-varying Petri-Nets. . . . .	46
3.6	Representation of the surface runoff reservoir model using time-varying Petri-Nets. . . . .	48
3.7	Representation of the groundwater reservoir model using time-varying Petri-Nets. . . . .	49
4.1	A basin can be continuously subdivided into strips of terrain at the same distance from the outlet. These strips are not necessarily continuous. They are physically connected to the outlet by many channels, but of these channels the physically significant quantity is the length, and therefore they are mathematically equivalent. . . . .	59
4.2	(a) Maps of the distances to outlet. their distribution is the width function. (b) Map of the rescaled distances to outlet. Rescaling factor is $r=100$ . . . . .	61
4.3	A HRU is "vertically" split into a surface and a subsurface domain. The overall response (to channel flow) is obtained just by the sum of the two contributions. . .	64

4.4	A single control volume is considered in which the fluxes are the total precipitation, evapotranspiration and discharge. . . . .	70
4.5	Representation of the evolution of the backward pdf for three injection times ( $t_{in_i}$ , where $i = 1,3$ ) as varying with the injection time $t_{in}$ . The time shift between the three injections was dropped for a direct comparison of the curves. . . . .	76
4.6	Representation of the backward cumulative distribution function for three injection times ( $t_{in_i}$ , where $i = 1,3$ ), as varying with the actual time $t$ . The time shift between the three injections was dropped for a direct comparison of the curves. . . . .	77
4.7	Representation of the evolution of the backward pdf versus the actual time $t$ . The time shift between the three injections was dropped for a direct comparison of the curves. In this case, the area below the curves is not equal to 1. . . . .	78
4.8	Forward residence time probability distribution: in red the relative storage, in green the forward residence time distribution and in blue the relative discharge function. . . . .	83
4.9	Representation of the forward probability of the outputs: in red the relative storage, $s(t, t_{in})$ , in green the output probability, $P_S[t - t_{in} t_{in}]$ and in blue the relative discharge function $\mathcal{F}$ , defined in the text. The difference between $P_S[t - t_{in} t_{in}]$ and $\mathcal{F}$ is the function $\mathcal{G}$ , defined in the text. The orange dashed line represents the generic instant $t$ , after which $P_S[t - t_{in} t_{in}]$ and $\mathcal{F}$ are unknown. . . . .	84

4.10 Variation of the partitioning coefficient in time, for a single injection time in January: after a time scale of 5 months its oscillation became irrelevant and its value tends to its final value of 0.78 . . . . . 86

5.1 Digital elevation model of the Posina River basin and HRU partition. Red dots represent the centroids and numbers represent the ID of the HRUs. Red diamonds represent the discharge measuring points. . . . . 106

5.2 Photo taken on the Posina River at the closure section of Stancari, Vicenza, Italy. . . . . 107

5.3 Power plants restitution immediately before the Stancari section. . . . . 107

5.4 Photo taken on the Posina River at the section of Bazzone, Vicenza, Italy. . . . . 108

5.5 System of buckets for the total throughfall measurements. The buckets were installed and measures collected by the Department of Land, Environment, Agriculture and Forestry of University of Padua. . . . . 108

5.6 Time variation of the LAI in the 1998 for the HRU ID 42. This HRU was chosen since the LAI time series had the less number of no values. . . . . 110

5.7 Schema of the connection of the JGrass-NewAge components, necessities to perform the modeling solution: the blu arrows represents the connection out-to-in made possibile thanks to OMS3. . . . . 111

5.8 Example of 4 maps produced using the Horton-Machine libraries for a single HRU (ID 1): panel *a* shows the digital elevation model, panel *b* shows the distances rescaled by a factor of 10, panel *c* shows the topographic index and panel *d* shows the skyview factor . . . . . 114



- 
- 5.9 Schema of the connection of the JGrass-NewAge components, necessities to perform the travel times analysis: the blu arrows represents the connection out-to-in made possibile thanks to OMS3. . . . . 115
- 5.10 Measured discharge (black line) VS modeled discharge obtained using the embedded reservoir model (blu line) and the Hymod model (red line). Both models were calibrated against one year of discharge measured at Stancari (top plot) and validated using the entire time series of discharge at Stancari and at Bazzoni. . . . . 119
- 5.11 Waterfall charts of the relative contributes of the water balance for the canopy, root zone and groundwater reservoirs. Green bars represent the inputs of the storage, blu bars represents the outputs and red bars represent the change in storage. Five years, from 1994 to 1998, are shown in order to compare the inter-annual variability of each contribute. . . . . 122
- 5.12 Waterfall charts of the relative contributes of the water balance for the canopy, root zone and groundwater reservoirs. Green bars represent the inputs of the storage, blu bars represents the outputs and red bars represent the change in storage. Two selected month, March 1994 and September 1994, are shown in order to compare the annual variability of each contribute. 123
- 5.13 Spatial distribution of AET for six selected time steps. 124

5.14 Temporal evolution of the travel times and residence times: each colored curve represent the travel times computed for each HRU, the black curve represent the residence time for the entire catchment, the blue dashed line represent the trend line. The mean residence time is around 70 days and the trend shows an increasing over time of around 13 minutes/day. . . . . 126

5.15 Temporal evolution of the mean evapotranspiration times: each colored curve represent the evapotranspiration times computed for each HRU, while the blue dashed line represent the trend line. . . . . 126

5.16 Distribution of the mean residence times, travel times and evapotranspiration times. . . . . 127

5.17 Travel time analysis obtained for the entire Posina catchment and the complete mixing case. . . . . 128

5.18 Separation of the baseflow contribute obtained using a cubic spline interpolation of the minimum values of the recorded discharge at Stancari station. . . . . 129

5.19 Comparison between the measured and simulated discharge at Stancari, where the calibration of the parameters was performed. A good agreement of the two series is evident, with a KGE of 0.75. . . . . 130

5.20 Comparison between the measured and simulated baseflow at Stancari, where the calibration of the parameters was performed. . . . . 131

5.21 Comparison between the measured and simulated discharge peaks at Stancari. A very good agreement of the two series is evident, with a KGE of 0.82. . . . . 131

- 5.22 Waterfall charts of the relative contributes of the water balance for the canopy, root zone and groundwater reservoirs. Green bars represent the inputs of the storage, blu bars represents the outputs and red bars represent the change in storage. Five years, from 1994 to 1998, are shown in order to compare the inter-annual variability of each contribute. . . . . 133
- 5.23 Comparison of the time series of LAI, gross precipitation (P) and drainage (D) ratio and AET from the wet canopy. According to the LAI increasing, the ratio decreases till 3 times during summer period. The AET varies according to the storage of the wet canopy and goes to zero when the storage is null. . . . . 135
- 5.24 Time series of PET (red lines) VS the AET from the wet canopy (green lines) and root zone (blue lines). . . . . 136
- 5.25 Scatter plot between throughfall and gross precipitation: the black solid line represents the bisector, while the blu solids lines represents the regression line. As expected throughfall is lower then precipitation, except for really small values (less then 0.1) that could be considered integration errors. . . . . 137
- 5.26 Variation of the ratio between cumulative throughfall and the cumulative gross precipitation with the cumulative gross precipitation. On the small time scales, the ratio presents a high variability while it tends to a final value of 0.88 on the annual scale. . . . . 138
- 5.27 Temporal evolution of the mean travel times and residence times: each colored curve represent the travel times computed for each HRU, the black curve represent the mean residence time, while the blue dashed line represent the trend line. . . . . 139

5.28 Temporal evolution of the mean evapotranspiration times: each colored curve represent the travel times computed for each HRU, while the blue dashed line represent the trend line. . . . . 140

5.29 Distribution of the mean residence times (red curve), travel times (blue curve) and evapotranspiration times (black curve). . . . . 141

5.30 Evolution of the partitioning coefficient in one year of daily simulation: the highest value are achieved in January while the lowest in June. . . . . 141

5.31 Example of a different schema of connections: after the partition, the volume in excess remains in the root zone reservoir and, then, transferred to the surface flow, while the volume infiltrated feeds the groundwater reservoir. . . . . 142

5.32 Root zone storage obtained using the previous schema and one year of data. . . . . 143

5.33 Example of a different schema of connections: a dynamical feedback from the groundwater to the root zone is present. . . . . 144

5.34 Example of a different schema of connections: root zone and groundwater reservoirs exchange dynamical feedbacks through two non-linear fluxes. . . . . 145

6.1 The LWRB component of JGrass-NewAge and the flowchart to model longwave radiation. . . . . 157

6.2 Test site locations in the United State of America. . . . 162

6.3 Results of the clear-sky simulation for four literature models using data from Howland Forest (Maine). . . . 164

6.4	KGE and RMSE values for each clear-sky simulation using literature formulations, grouped by classes of latitude and longitude. Only values of KGE above 0.5 are shown. Only values of RMSE below 100 [ $\text{W m}^{-2}$ ] are shown. . . . .	165
6.5	KGE (best is 1) and RMSE (best is 0) values for each optimized formulation in clear-sky conditions, grouped by classes of latitude and longitude. Only values of KGE above 0.5 are shown. . . . .	166
6.6	KGE and RMSE values for each model in all-sky conditions with the optimized parameters; results are grouped by classes of latitude and longitude. Only values of KGE above 0.5 are shown. . . . .	167
6.7	Results of the model parameters sensitivity analysis. It presents as boxplot the variation of the model performances due to a variation of one of the optimal parameter and assuming constant the others. The procedure is repeated for each model and the blue line represents the smooth line passing through the boxplot medians. . . . .	168
6.8	Comparison between model performances obtained with regression and classic parameters: the KGE values shown are those above 0.3 and results are grouped by latitude classes. . . . .	169
6.9	Comparison between model performances obtained with regression and classic parameters: the KGE values shown are those above 0.3 and results are grouped by longitude classes. . . . .	170

6.10 Boxplots of the KGE values obtained by comparing modeled upwelling longwave radiation, computed with different temperatures (soil surface temperature (SKIN), air temperature (AIR), and soil temperature (SOIL)), against measured data. Results are grouped by seasons. . . . . 172

6.11 The MS optimizes model parameters connecting of the experimental semivariogram component to the theoretical semivariogram to the optimization tool. The output of the optimization are the sill, nugget and range, which, together with the type of model, are the input of the Kriging component. Outputs of the MS can be time-series and maps of interpolated variables. 180

6.12 The MS, respect to the one presented in figure 6.11, connects the particle swarm to the automatic leave-one-out for the assessment of the model performances. 181

6.13 Study area: Isarco River Valley is situated in the North-Est part of Italy and it is one of the main valley in the Alto-Adige region. . . . . 182

6.14 Fitting of the experimental semivariogram for the 15th June 2008 12:00. The 10 theoretical semivariogram models were optimized using the PSO. . . . . 184

6.15 Monthly variation of the NSE index: each dot is the averaged NSE over the entire dataset. The theoretical semivariogram models shown are: Bessel (red dots), Exponential (yellow dots), Gaussian (black dots), Linear (blue dots), Spherical (green dots). . . . . 186

6.16 Maps of spatialized temperature for the 15th February 2008 and for the 15th June 2008 . . . . . 187

6.17	The two bubble plots show the RMSE obtained using the OK (box on the left) and DK (box on the right). The scales of the bubbles are not the same for the two plots for visualization reasons. In fact, being the errors in the DK case is much less than the OK case, a unique bubble scale didn't allow to appreciate the RMSE values in the DK case. . . . .	188
6.18	Boxplots of the semivariograms of the precipitation event of 29th and 30th June 2008: the horizontal line in the middle shows the median, the bottom and top end of the box show the 25th and 75th percentile, respectively, the whiskers (vertical line) shows the range of the data. . . . .	189
6.19	Comparison between the four types of Kriging (OK, red solid line, DK, blue solid line, LOK, brown solid line and LDK, green solid line) and the measured rainfall (black dashed line). . . . .	190
6.20	Spatial interpolation of the precipitation made for each pixel of the DEM (100 m resolution), applying the OK and the linear semivariogram model. . . . .	191
6.21	Comparison between <i>gstat</i> and SI package in the interpolation performances using the DK. . . . .	193
A.1	Schematic representation of LUCA's rounds and step.	207
A.2	Implementation of the Java simple factory for the choice of the clear sky emissivity model . . . . .	209





## INTRODUCTION

The increasing impacts of climate changes and population growth on water quality, supply and distribution are leading the scientists' attention to the development of comprehensive models necessary to a better description of the water and solute transport processes.

The implementation of water management strategies has to rely on robust predictions. 'Getting the right answers for the right reasons' in terms of hydrological response, (Kirchner, 2006), should be the main goal of the modelers, (Hrachowitz *et al.* , 2013, 2014; Soulsby *et al.* , 2015).

In the almost-infinite panorama of hydrological models, this goal is reached in many notable works (e.g., Refsgaard, 1995; Santhi *et al.* , 2001; Manfreda *et al.* , 2005; Therrien *et al.* , 2010; Birkel *et al.* , 2011), and it nearly appears something already overcome. But how many times does a researcher try to disentangle systems of equations presented in a paper without any success? How many times does a researcher try to reproduce a result that doesn't come out? How many times does a

researcher try to use a software that doesn't work? Too many. Therefore 'getting the right answers for the right reasons' should also be combined with the concept of reproducible research. This last concept is usually dismissed by most researchers as unnecessary, and too often science is something that remains closed in a laboratory, in a paper or in a computer.

One of the main goal of this thesis is to present a new way of think, develop and produce research. Some of the concepts presented in this work could appear already debated and solved, but here are treated from a new perspective of the Open Science in the hydrology, both from the hydrological modeling and from the travel time analysis points of view.

The starting point of all was the hydrological modeling system JGrass-NewAge, presented by Formetta *et al.* (2014a) and developed in many successful researches (Formetta *et al.* , 2013a,b, 2014a, 2016; Abera *et al.* , 2017a,b). JGrass-NewAGE was built on top of the Object Modeling System v3 , (OMS3) (David *et al.* , 2013), an environmental modeling framework based on the concept of components, well described in Chapter 2.

In this thesis, to respond to the challenges of the contemporary hydrology, JGrass-NewAGE was enhanced, introducing several structural changes.

The question was: how do optimally develop, test and deploy an hydrological model?

Usually hydrological modelers disregard this aspects, but since much of their work is increasingly based on such models, much more attention has to be devoted to appropriate "software carpentry". Software characteristics that should be included, among others, are:

- ways for algorithm inspections by third parties,

- encapsulation of modeling task that can be separately validated,
- modalities to maintain the traceability of modeling solutions used in papers.

Besides, since the work of researchers is inherently gifted by changes of ideas and paradigms, a further software characteristic would be

- introducing the software mechanisms that allow the flexibility to insert new ideas with minimal programming efforts, when new findings or requests come in.

Therefore cleaning out the informatics of modeling (and connected problematics of documentation) to fulfill the above requirements was an important objective of this Thesis, which presents and implements the seeds of a collaborative system called GEOframe. Moreover, in order to approach systematically the building of reservoirs models, we developed a graphical-mathematical tool to analyze the model structure that illustrates the variables involved and their interactions in terms of graphs. In turn, these graphs are one-to-one mapped to complex systems of ordinary differential equations and with, if the case, partitioning rules among the fluxes. All the previous aspects are presented and deeply discussed in Chapter 2.

A well established strategy in many modern modeling systems is to reduce catchment hydrology to a set of interconnected storages (reservoirs), (Clark *et al.* , 2008; Fenicia *et al.* , 2008; Tague & Dugger, 2010; Hrachowitz *et al.* , 2013). In the past, the passage from the spatially explicit models to more aggregate (lumped) ones has been justified through different strategies and has been usually gathered below the umbrella of "scaling issues", (e.g, Blöschl & Sivapalan, 1995).

Among the others, notable is the work by Gray *et al.* (1993), which tried to offer a mathematical approach to average spatial contributions. His approach brought, for instance, Reggiani *et al.* (1998) and Zehe *et al.* (2006) to formulate a systematic approach for separating spatial environments of catchments. However, their approach remained quite abstract and not widely adopted, so far. Other approaches were recently reviewed in Hrachowitz & Clark (2017), and the interested reader can refer to that paper. In any case, the conceptualization of the domain as a connection of multiple storages to simulate processes interactions and spatial organization, proved to be robust in the reproduction of how the mass of water and solute are stored and released through different outputs (e.g. discharge and evapotranspiration), (Fenicia *et al.* , 2006; Birkel *et al.* , 2011; Bertuzzo *et al.* , 2013; Hrachowitz *et al.* , 2013). Notwithstanding the intentions of the Authors, also the spatially averaged models grew quite complex and challenged their Authors and the community to find the right compromise between the model complexity and the reduction of the number of reservoirs and parameters, (Hrachowitz *et al.* , 2014).

Questions arise like: which is the minimal set of reservoirs that allows to describe correctly the flux dynamics? How is it possible to overcome problems related to the number of parameters and the calibration procedures? How to build a modeling system able to tame the complexity of research needs?

We tried to answer to the previous questions, presenting a new semi-distributed model for runoff and evapotranspiration, which is illustrated and discussed in Chapter 3. The model schematizes each HRU as a connection of storages (reservoirs) and solves the water budget for each one. The work tries to find the correct number of storages to capture the complexity of the hydrological response, making room for the separation between

evaporation and transpiration, in a balance between the data availability and the computational costs.

Travel times analysis is a way to interpret the complexity of such systems, (e.g., Tetzlaff *et al.* , 2008). Travel time models have been studied extensively for many years, (Rigon *et al.* , 2016b, e.g.), with some recent benchmark contributions that can be found in the works of Botter *et al.* (2010, 2011); Botter (2012); van der Velde *et al.* (2012); Cvetkovic *et al.* (2012); Cvetkovic (2013); Ali *et al.* (2014); Benettin *et al.* (2015b); Harman (2015b). Although all of these studies provided valuable advances to the travel time theory, the literature remained obscured by different terminologies and notations, as well as model assumptions, which were not fully explained. One of the scopes of this Thesis is to clarify and extend the concepts of the travel time theory, and show its possible use to dissect reservoirs model physical behavior. We tried to clarify concepts such as the backward and forward probabilities, the role of partitioning coefficients between fluxes and the basic mathematics of travel time distributions. These goals are pursued in Chapter 4 of the Thesis.

To give substance to the theory, as an example a small catchment in the prealpine area, the Posina River, was chosen, already studied in various papers, (Norbiato *et al.* , 2009; Abera *et al.* , 2017b): the semi-distributed model for runoff and evapotranspiration was used to simulate hydrological fluxes and its travel times analysis performed. Results are illustrated in Chapter 5.

Since the Thesis regards in wide sense the building of a collaborative system, Chapter 6 contains the discussion of some general purpose tools that were implemented or improved for supporting the present research. They include the description and the verification of a software component dealing with the

long-wave radiation budget and another component dealing with an implementation of some Kriging spatialization procedure.

## PRELIMINARY ON MATHEMATICS AND INFORMATICS

Contemporary hydrology requires advanced conceptual, mathematical and informatics tools to deal with complex systems of equations, new data collection methods and sources. One of the scopes of this thesis is to line up some of these tools in order to cope with all the previous challenges in the most efficient way. The leading principle of the project was to find the optimal complexity: in a balance between i) the processes to describe, ii) the computational requirements and iii) the necessity to implement software easy to share and maintain within a solid infrastructure, which could become the basis for a community of hydrologists.

The starting point of the work was the hydrological modeling system JGrass-NewAge, presented by Formetta *et al.* (2014a). This system was developed and tested thoroughly (Formetta *et al.* , 2013a,b, 2014a, 2016; Abera *et al.* , 2017a,b). However, the fast evolution of the tools along with those successful researches, posed some issues:

- the procedures to keep track on code modification and

evolution were not assessed;

- documentation procedures were not defined;
- internals of the codes were based on procedural programming, whilst using a object-oriented programming language, and major revision for extensibility and reuse were necessary;
- the lack of a software building system for the growing set of tools;
- the selection and adoption of well supported and well designed libraries for basic numerical operations (like integrating systems of ordinary differential equations, fitting curves, and other tasks);
- the necessity to scale up modeling to arbitrarily large systems, starting from the composition of elementary units, without losing relevant information and keeping tracks of possible feedbacks.

These issues were faced at the beginning of this work, and partially solved conjointly with the ongoing work of Francesco Serafin.

JGrass-NewAGE system is based on the Object Modeling System v3 , (OMS3) (David *et al.* , 2013). The search of a system to implement hydrological models, was necessary to tame the complexity of needs of research and applications, and the will to have a code flexible enough to be improved when new findings or requests come in. The choice of this system came after after the temporary adoption of the OpenMI infrastructure, (Gregersen *et al.* , 2007). The rational of the moving from OpenMI to OMS3 is documented in Rigon (2011) and not fully repeated here.



In order to understand the work of this Thesis, the present chapter covers then:

a short description of OMS3 and its evolution;

a description of the pre-existing JGrass-NewAGE components;

a section is dedicated to some structural changes that were introduced in JGrass-NewAGE to respond to some of the issues listed above;

a mathematical-graphical representation of hydrological model, called Petri Nets, which will be used through the thesis, is presented.

the presentation of a new flexible structure based on graphs for commanding simulations of complex systems, which represents an enhancement recently introduced to OMS3 structure.

## **2.1 Object Modelling System 3**

Object Modeling System v.3, David *et al.* (2013), is a component-based Environmental Modeling Framework (EMF). Components, here, mean self-contained building blocks, modules or units of code. Each component implements a single modeling concept, and the components can be joined together to obtain a Modeling Solution (MS) which can accomplish a complicate task, such as simulating the water budget storages and fluxes. These results are obtained by making leverage a group of features offered by the Java language. For instance, one of these features are the Annotations, a form of syntactic metadata that can be added to the Java source code. The overall capability added by these Annotations, is called self-reflection, (Forman *et al.* , 2004) and is used by OMS3 to provide a hook to connect components among themselves. Annotations are managed, at a higher, external level, by a Domain Specific

Language (DSL), (Mernik *et al.* , 2005), built upon the Groovy language, (Groovy). In practical uses of the DSL, just a subset of the whole language is exposed to the programmer, and, therefore, the researcher (or user) does not have to dive into complex programming issues. A system like OMS , in fact, is built to answer to different users needs, (Rizzoli *et al.* , 2008), and among those, two figures emerge: the researcher/programmer who implements and deploys components, and the researcher/user that concatenates components for obtaining a MS. Our perspective covers, obviously, both the above figures.

OMS3 is "non-invasive", (Lloyd *et al.* , 2011), meaning that, with respect to other similar frameworks, does not change the habits of a good Java programmer. Besides, the OMS user does not need to have an extensive knowledge of libraries, as its Authors say: "there are no interfaces to implement, no classes to extend, no polymorphic methods to override and no framework-specific data types to use", (David *et al.* , 2013).

An important design aspect in the core of OMS3 is the support for implicit multi-threading, a common technique to parallelize internal processing within a software application to fully utilize available processing resources. Multi-threading is a constitutive characteristic of Java language but requires appropriate programming capabilities and algorithm design. These remain encapsulated in the overall OMS3 framework to which is delegated how to run the components, and, when no dependencies are present, to run them in parallel. The arrangement in components, and the implicit parallelism just mentioned, enable hierarchical scaling of environmental models from single desktop computer to large multiprocessor servers, called CSIP, (David *et al.* , 2013).

OMS3 ensures the inter-operability with other frameworks as well as the support of multiple programming language besides

Java. In particular components can be programmed in Fortran, C. R and Python bindings were recently provided.

The advantage of constructing modular software, which helps to minimize couplings, is the production of code that is more flexible, easier to maintain and to be inspected by third parties. Multiple algorithms can be implemented within the same component or in various components, and inserted in MS as alternatives, thus opening the way to compare, inside the same chain of tools, different approaches.

As said, to create and configurate runtime simulations, a DSL is used, which specifies the component executable binaries, the model-specific parameters and the connections between components. In particular, figure 2.1 shows a basic example of a simulation file (from now on sim file). Three main parts can be distinguished:

- *component*, where the component executable to use are specified,
- *parameter*, which specifies the parameters of each component and
- *connect*, where the connection between the components are specified.

Connections are always made according to the *out-to-in* schema, i.e. the output of a component is the input on the following component.

OMS3 comes with a series of features built-in, such as:

- the ability to manage time cycles (each component can treat just a single time step);
- integrated parameters optimizers;
- sensitivity/uncertainty analysis tools;

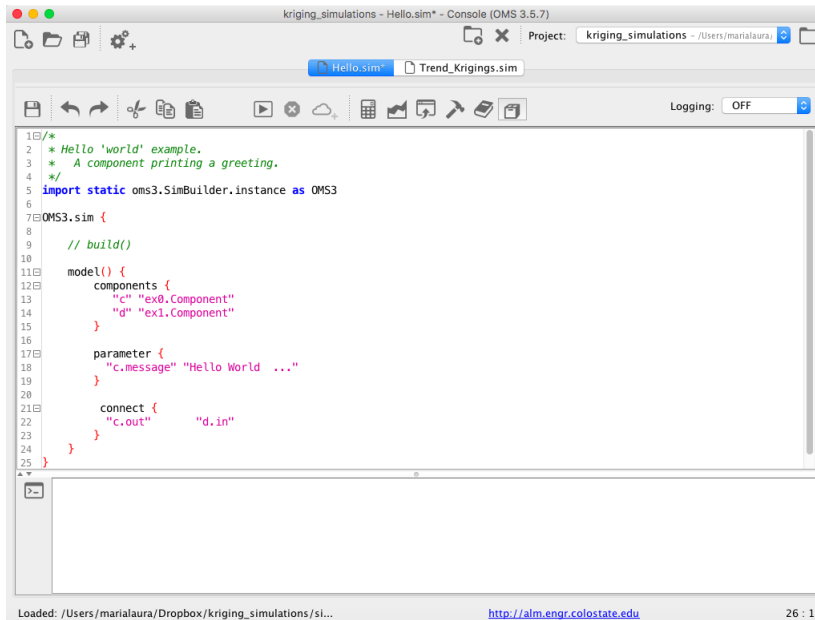


Figure 2.1: Example of sim file in which are detailed the 3 main fields: *component* where are specified the component executable to use, *parameters* where are specified the component parameters, *connect* where are specified how to connect the components, according to the *out-to-in* schema .

- output analysis tools (e.g., statistical evaluation and graphical visualization);
- modeling audit trails (i.e., reproducing model results for legal purposes);
- miscellaneous technical/user documentation.

Two calibration algorithm are present in the core of OMS: Let Us CALibrate (LUCA), (Hay *et al.* , 2006b) and Particle Swarm Optimization (PSO), (Eberhart & Kennedy, 1995). Details of the algorithms are given in Appendix A.1.

For the applications of this thesis, the basic OMS3 structure has been enhanced by Francesco Serafin, which is briefly reported in section 2.6.

## 2.2 The JGrass-NewAge hydrological modeling system

Omitting all JGrass-NewAge history before Giuseppe Formetta research project, which is detailed in Rigon (2014), in the present section the state-of-art of the system, before this Thesis, is presented.

JGrass-NewAge is a semi-distributed, component based hydrological system, (Formetta *et al.* , 2014a). It is written in Java, object oriented and open-source and based on OMS3.

JGrass-NewAge was built upon three main parts:

- the Geographical Information System (GIS) uDig (<http://udig.refractions.net>) for data treatment and visualization;
- the OMS uDig Spatial Toolbox, (Abera *et al.* , 2014);
- OMS3.

Various OMS components to simulate the relevant hydrological processes and pre-process/post-process data were implemented. The complete list of components is specified in table 2.2.

They can be grouped in six categories, according to the task performed:

- geomorphic and DEM analyses;
- spatial extrapolation/interpolation of the meteorological tools;
- estimation of the radiation budget;
- estimation of evapotranspiration;
- estimation of runoff production;

Component	Reference
JGrasstools	Formetta <i>et al.</i> (2014a)
Horton Machine	Rigon <i>et al.</i> (2006)
Kriging	Formetta <i>et al.</i> (2014a)
Shortwave	Formetta <i>et al.</i> (2013b)
Longwave	Formetta <i>et al.</i> (2016)
Priestley Taylor	Formetta <i>et al.</i> (2014a)
Fao-Etp model	Formetta <i>et al.</i> (2014a)
Adige	Formetta <i>et al.</i> (2014a)
Snow	Formetta <i>et al.</i> (2013a)
Cuencas	Formetta <i>et al.</i> (2014a)

Table 2.1: List of JGrass-NewAge components and associated references.

- channel routing;

The DEM analysis is performed using the JGrasstools and the Horton Machine, (Rigon *et al.* , 2006; Abera *et al.* , 2014; Formetta *et al.* , 2014a), which allow, starting from the DEM, to extract the drainage directions, the total contributing areas, the slopes, the river network, the sub-basin partitioning and the topographic characteristics required by computation.

Available tools for the spatial extrapolation/interpolation of the meteorological forcing data are both geostatistic, Kriging techniques and deterministic, an Inverse Distance Weighting (IDW, Cressman (1959)) and Just Another Model Interpolator (JAMI), an undocumented bare bone tools to deal with spatial meteorological data (mainly temperature and precipitation data).

The radiation budget model has been presented and validated in Formetta *et al.* (2013b) and Formetta *et al.* (2016) and includes both shortwave and longwave radiation.

Evapotranspiration can be estimated using two different formulations: the Fao-Evapotraspiration model Allen *et al.* (1998), and the Priestly-Taylor model Priestley & Taylor (1972).

Snow melting and snow water equivalent is treated in a component which includes three models, as described in Formetta *et al.* (2013a).

Two different runoff generation models are implemented, the Duffy's model (Duffy, 1996) and the Hymod model (Moore, 1985), but Duffy model was never really tested.

The discharge, generated at each hillslope, is routed to each associated stream link according to Mantilla & Gupta (2005). In principle, the component de-Saint Venant, that solves the one dimensional de Saint Venant equations, (Casulli & Zanolli, 1998) can also be used for the same scope. However, the latter component, presented at

<http://abouthydrology.blogspot.it/2015/10/a-solver-for-1d-de-saint-venant-equation.html> was never fully documented and tested.

The infrastructure is able to manage complex data structure in input and output such as geographical objects, i.e. *raster* and *shapefile*, commonly used within the GIS and managed with the Geotools library, Turton (2008). Several ancillary components are available for the management of these data structure such as raster readers and writers, the shapefile reader and writer and more.

Although the successful research out-comings, JGrass-NewAGE presented some issues. Mainly, the internals of the codes were based on procedural programming, whilst using a object-oriented programming language, and major revision for extensibility and reuse were necessary. Besides, JGrass-NewAGE was still not exploiting completely the possibilities offered by OMS3 "componentization". All of this suggested a major refactoring of the original codes.

## 2.3 JGrass-NewAGE refactoring

While maintaining the name JGrass-NewAGE through the Thesis, the refactoring of the existing codes, was quite substantial (at design level more than algorithmic level).

A systematic use of Design Patterns (DP), (Gamma *et al.* , 1994; Freeman *et al.* , 2004) was introduced. DP are "a general repeatable solution to a commonly occurring problem in software design. A design pattern isn't a finished design that can be transformed directly into code. It is a description or template for how to solve a problem that can be used in many different situations.", (SourceMaking).

Design patterns implements some rules that allows a cleaner implementation of tasks, and, for instance, separate code parts that are going to vary from those who are thought to remain the same, see appendix A.2 for an implementation example. The adoption of these patterns, once their rational is understood, makes the code more "readable". While fruitful in the large market of programmers, the scientific community remained largely impervious to these techniques, and just a few examples of good practices can be found in scientific literature, (Gardner & Manduchi, 2007; Rouson *et al.* , 2011).

Generically DP suggest answers to the following questions: are there strategies for producing a minimum number of classes without losing functionalities in the work done and promote its extensibility? What goes into a class, and what in other object-oriented features (like methods of a class)? How to create classes with a minimum of generality that can be reused in other problems? How to build classes that can be easily maintained, modified and evolved without disrupting other parts of codes? Answer to these questions became mandatory to further evolve JGrass-NewAGE.

The components revised according to the previous aspects were:



- Kriging: the component was split in 4 different components, experimental semivariogram, theoretical semivariogram, Krigings and leave one out. Simple Factory methods was implemented for the runtime selection of the semivariogram model and of the solver of the linear system of weights;
- Shortwave: code was only cleaned;
- Longwave: Simple Factory methods was implemented for the runtime selection of the clear sky emissivity model;
- Priestley Taylor: code was only cleaned;
- Fao-Etp model: code was only cleaned;
- Snow: the component was split in two components, rain-snow separation and snow modeling (melting and snow water equivalent). Simple Factory methods was implemented for the runtime selection of the melting model.

Some new components were added to the model and some old ones, not used anymore were dropped. Components added to the existing ones are:

- clearness index: computation of the ratio between the incoming and top atmosphere shortwave radiation;
- Net radiation: computation of the net radiation, given the SWRB and LWRB;
- Embedded reservoir model: schematization of the water budget through systems of connected reservoirs. Four different components were implemented, which solve the water budget for the canopy, root zone, surface flow and groundwater;

- Travel times: travel time analysis, backward and forward probabilities;
- Muskingum-Cunge: channel routing.

At algorithmic level, the choice was made to adopt as basis for most mathematical computations the Apache Commons Math Library, <http://commons.apache.org/proper/commons-math/>. Commons Math is a library of lightweight, self-contained mathematics and statistics components addressing the most common problems not available in the Java programming language or Commons Lang. This went to substitute the previously used libraries, such as JAMA.

Contextually to the refactoring, the code was moved to a common community repository, where it is freely available under GPLv3 license,

<https://github.com/geoframecomponents>, as envisioned in Serafin *et al.* (2014). Most of the components were also documented and the documentation made available to the community blog GEOFRAME: <http://geoframe.blogspot.com>.

## **2.4 GEOframe: a system for doing hydrology by computer**

*"I have been frustrated often with statisticians and computer scientists who write papers where they develop new methods and seem to demonstrate that those methods blow away all their competitors. But then no software is available to actually test and see if that is true... In my mind, new methods/analyses without software are just vaporware... If there is no code, there is no paper."*, Leek (2013).

Vaporware and software, in the above citation, well express the the fundamental idea that science must be reproducible and

## Getting more: Replicability

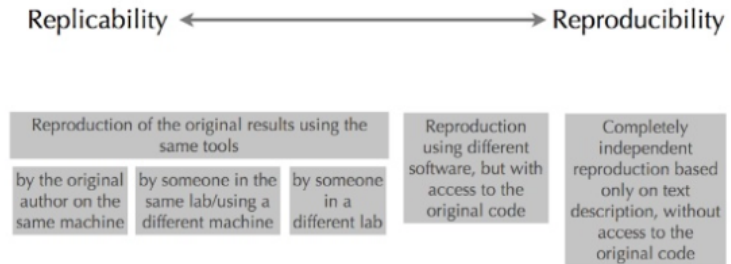


Figure 2.2: Reproducibility VS replicability of reasearch. Figure adapted from [http://rrcns.readthedocs.io/en/latest/reproducible\\_research.html](http://rrcns.readthedocs.io/en/latest/reproducible_research.html)

possibly replicable, Rigon *et al.* (2015); Bancheri *et al.* (2016, 2017). It means that everyone should be able to take what you write, the experiments you did, the mathematics you drew and do it again, obtaining the same results. This statement is as general as difficult to achieve, and usually dismissed by most researchers as unnecessary. The concepts of reproducibility and replicability are different and figure 2.2 well explains their meanings. Replicable science means that the reproduction of the original results using the same tools should be possible by any researcher. Reproducibility infers with the possibility of obtaining results from the scratch, starting only from the text description and using different tools in different contexts. Reproducibility of research results is possible clearly defining concepts, maintaining the consistency of notation between papers on the same topic, using Petri-Nets algebra, together with tables of symbols and associations, see section 2.5. However, even a paper reproducible in theory could be not

reproducible in practice. In fact, since a central core of the science is computation, it adds a further layer of complexity to the science visible in papers, often requiring the use of the right tools, including appropriate programming skills and extra training for the researchers.

What a researcher can do to improve reproducibility and replicability of her/his results?

We delineate here some practices that can help researchers in this task, that we can call the definition of a Reproducible-Research System (RRS), (e.g. Formetta *et al.* , 2014a). The very first step should be share methods and make public any source code, under a copyleft license (for instance GPL v 3.0). In this way everyone can have access to the code, use it and, in case, improve it. Building something 'quick and dirty', assuming an indefinitely delayed cleaning up, is something that could be easily avoided by adopting an open source code approach.

However, working with open-source tools doesn't mean only sharing codes but also doing it in the right way, in order to create interest among the users, which can become also active contributors. Therefore, the second step is providing a good documentation, according to a standard format. Probably, this is the biggest effort to make but also the most important. There is no science without communication and, vice-versa, if we want the widest spreading of our science, documentation is the key. The optimum could be thinking to different actors/users of the software and provide a documentation for each one. For example, for those who develops the components, it should be clarified the overall scope of the component, its design, its classes, the algorithms it uses and some reference to check it all. While users should have information about the IO data , they could not be necessarily interested in component internals.



Figure 2.3: GEOframe organization logo.

Providing example of runs and references, is a further step. This helps the developers and other researchers in all future uses and improvements of the code.

Finally, the last step (at least of our list) is creating a sharing community for ideas, questions, doubts and support.

Trying to follow all the previous steps, GEOframe organization was founded. The organization is thought as a system for doing hydrology by computer, a sharing community for researchers and users. The idea of this community dates back to 2008 but operatively, the organization was born in 2016 and given the logo in Figure 2.3, with this research project. The necessity to share among the scientific community our efforts, led us to think how to do it efficiently, easily and with some basic standard for contributors and users.

GEOframe organization supports the idea of reproducibility and replicability of the research. It fully respond to some of the issues posed in the previous section regarding the procedures to keep track on code evolution, the documentation and the building system. To deploy our ideas, we had to make some crucial decisions:

- GitHub was chosen as the public hosting service for the GEOframe codes and projects. GitHub,

<https://github.com>, is a web-based git , version control repository and Internet hosting service. GitHub takes care of the development history of the source code recording and versioning changes (git), storing those changes in a public repository, (GitHub). In particular, the GEOframe repository created for the development of the source code is

<https://github.com/geoframecomponents>.

- In order to make the assemblage of Java project easier, with all the dependencies from external classes and/or libraries automatically solved and updated to the latest versions, Gradle, (Berglund & McCullough, 2011) was chosen as building system. Other choices could have been Maven, (<http://maven.apache.org>) or IVY (<http://ant.apache.org/ivy/>) but Gradle allows for a more concise representation of tasks. Gradle, in fact uses a Domain Specific Language (DSL) based on Groovy, (Groovy). Gradle script is normally short and clean, it is relatively easy to write and learn and easy to read and maintain. Furthermore it is compatible with Ivy and Maven repositories. Gathering a source code into a project managed by a building system is the key to make it IDE-independent, allowing developers not to change their preferred development tool.
- Sharing the code and do it in the best way also means that the code must be accompanied by test cases and resources (possibly open-data). While writing code, it is pretty usual to commit changes or worse building the entire software without running tests. The continuous integration ensures building and testing of the source code at each commit. Using GitHub as web-based git repository hosting service,

Travis CI, <https://travis-ci.org>, is the best choice for a continuous integration service. The continuous integration service, automatically build the executable codes, checks if the tests are performed correctly and returns a positive answer if all is done properly.

- After having developed some rules for assigning a version to software components, and helped by the GitHub system, we chose to use the Zenodo (<https://zenodo.org>) archival system. "Zenodo is a research data repository. It was created by OpenAIRE and CERN to provide a place for researchers to deposit datasets". So, once a new version of the software is produced, it is uploaded to Zenodo, a Digital Identification Object (DOI) number is given to the version and the code is permanently stored, retrievable, and citable.

There are other aspects that helps the building of a RRS, which are already covered by OMS3, see section 2.1. One of these is the traceability of simulations performed by final users. This is a common problem, documented, for instance, along the Earth System Documentation project (<https://earthsystemcog.org/projects/es-doc-models/>). In OMS the metadating of simulations is included in the construction of .sim files. Their creation is an operation easy and straightforward but has a lot of information inside, see appendix A.5 for an example.

Therefore, sim files, together with all the jars, the resources in input and output are saved in a OMS3 project and should be shared. GEOframe repository, <https://github.com/GEOframeOMSPProjects>, makes our main projects available to any users and for future benchmark tests with other applications.

Finally, GEOframe components are documented.

Documentation for developers is in the "README" files attached to the source code at

<https://github.com/geoframecomponents>, while

documentation for users is available at

<http://geoframe.blogspot.it>.

## 2.5 Time continuous Petri Nets

As told in the Introduction chapter, catchment models can be conceptualized as sequences of reservoirs, Fenicia *et al.* (2008); Birkel *et al.* (2011); Bertuzzo *et al.* (2013); Hrachowitz *et al.* (2013). These exchange water through some derived laws of flux, in a network of connections that reproduce the river network organization. For the purposes of modeling, a basin can be split into hydrologically similar parts, called Hydrologic Response Units (HRUs), for example a hillslope, one of its part or a group of hillslopes, that can be treated as a whole from the mathematical, physical or computational points of view. Within each of these HRUs, parameters and hydrological quantities are considered either as uniform (in space) or described statistically by a distribution function.

When multiple reservoirs are used to depict the hydrology of a single HRU, they are named embedded reservoirs, by meaning that they act in the same spatial position but model different processes.

Is it possible to represent processes in a unique way which corresponds one-to-one to the set of equation to be solved ?

To answer this question, it is required the definition of an algebraic structure of objects to represent (water) budgets, which gives a clear idea of the type of interactions.

Petri Nets is a mathematical modeling languages for the



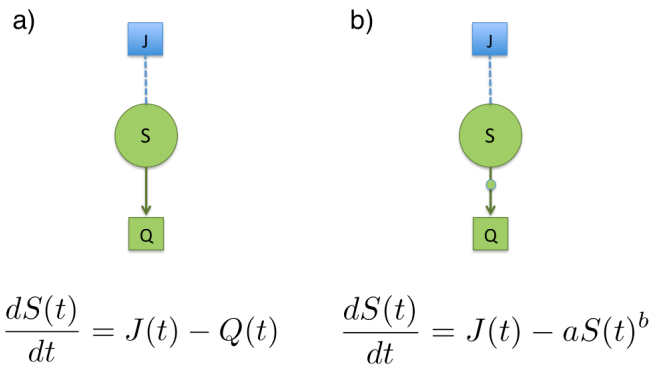


Figure 2.4: Linear reservoir and non linear reservoir representations using the time-varying Petri nets.

description of distributed systems, Murata (1989). Starting from their algebra, we tried to reformulate and adapt them to time-varying hydrological quantities, in order to propose an easy and immediate way to describe complex interactions.

Principle objects in Petri Nets representation are places, draw as circles, transitions, represented as small rectangle and connections represented as arcs. In our version, places represent time-varying storages, for instance the volume of water in groundwater, or the energy content of the same groundwater. A transition (in our case, small square) represents a flux and an arc (positive in the direction of the arrow) connects a place with a transition and viceversa. A place is connected only to transitions and, viceversa, transitions are connected to places. Connections place to place or transition to transition are impossible.

Different storages can be distinguished by different specification and/or different colors. Different arcs can be used for different connections: for example, an external forcing is represented as a dotted arrow, as in panel *a* in fig 2.4, a linear flux is represented as a solid arc, as in panel *a* in fig 2.4, a non linear flux is represented as a solid arc with a circle as in panel *b* in fig 2.4.

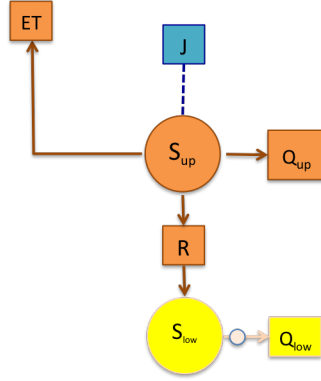


Figure 2.5: Two coupled storages, upper and the lower, represented using the time-varying Petri Nets.

Figure 2.5 shows an example of two coupled storages, upper and the lower, represented using the time-varying Petri Nets. Input of the system is the precipitation  $J$  [ $mm/h$ ], while the output fluxes are the evapotranspiration,  $ET$  [ $mm/h$ ], and the discharge contributes  $Q_{up}$  [ $m^3/s$ ] and  $Q_{low}$  [ $m^3/s$ ]. In particular,  $Q_{up}$  is modeled through a linear reservoir, while  $Q_{low}$ , through a non linear reservoir.  $R$  [ $mm/h$ ] is the recharge term of the lower layer, modeled through a linear reservoir too. Equations 2.1 and 2.2 are in one-to-one correspondence to the net shown in figure 2.5 and helps to clarify the system of ordinary differential equations involved:

$$\frac{dS_{up}(t)}{dt} = J(t) - Q_{up}(t) - R(t) - ET(t) \quad (2.1)$$

$$\frac{dS_{low}(t)}{dt} = R(t) - Q_{low}(t) \quad (2.2)$$

To complete the description of the system, a table of association between the fluxes and their expressions is added:

A table with symbols is necessary to explain the meaning of the symbols and to name the fluxes.

Flux	Expression
$ET$	$f \frac{S_{up}(t)R_n(t)}{S_{max}}$
$Q_{up}(t)$	$a_{up}S_{sup}(t)$
$R(t)$	$b_{up}S_{up}(t)$
$Q_{low}(t)$	$c_{low}S_{low}(t)^{d_{low}}$

Table 2.2: Example of table of association between fluxes and their expressions to add to the the Petri nets representation.

Symbol	Name	Unit
$a_{up}$	linear reservoir coefficient	$[T^{-1}]$
$b_{up}$	linear reservoir coefficient	$[T^{-1}]$
$c_{low}$	non-linear reservoir coefficient	$[T^{-1}]$
$d_{low}$	non-linear reservoir exponent	$[-]$
$ET(t)$	evapotranspiration rate	$[L^3 T^{-1}]$
$J(t)$	precipitation rate	$[L^3 T^{-1}]$
$Q_{low}(t)$	discharge from the lower layer	$[L^3 T^{-1}]$
$Q_{up}(t)$	discharge from the upperer layer	$[L^3 T^{-1}]$
$R(t)$	recharge of groundwater	$[L^3 T^{-1}]$
$R_n(t)$	net radiation	$[W/L^2 T^{-1}]$
$S_{low}(t)$	storage of the lower layer	$[L^3]$
$S_{max}(t)$	upper layer maximum storage	$[L^3]$
$S_{up}(t)$	storage of the upper layer	$[L^3]$
$t$	time	$[T]$

Table 2.3: Example of table of association between fluxes and their expressions to add to the the Petri nets representation.

It can be noticed that any place in a Petri Net correspond to the variation of the conserved quantity, and, therefore, we have as many equations as many circles in the overall graphs.

A system like the one represented in Figure 2.5 covers the description of a HRU. A catchment, as shown in figure 2.7 can be thought as the composition of the HRU and channels links. Each channel is, itself representable as a place with transitions representing its inputs and outputs. Therefore the PN graph of a catchment grows easily very complex but completely explicative, when all information is added (including the tables of symbols

and fluxes), and the burden added to representation helps clarity. PN can be formalized inside the mathematics of the categories, (Mac Lane, 2013), which was evolved to describe in abstract manner the interaction of complex systems, with the goal to revealing symmetries and common characteristics among various complex systems, and understand their emerging properties. However, this topic, goes beyond the scopes of the present thesis. Further information about the time-varying Petri Nets could be found at Murata (1989) and at Rigon *et al.* (2016c).

## 2.6 The Net3 infrastructure

As the above sections show, hydrological systems are non-linear, complex, open dissipative systems. Such systems consist of many coupled processes, depending on the state and on the spatial and temporal scale of analysis. The arrangement of couplings can be represented as a network of information between the variables that measure system processes, (Ruddell & Kumar, 2009), and Petri Nets can be a way to represents such interactions. Therefore, there is the necessity to scale up modeling to arbitrarily large systems starting from the composition of elementary units, without losing relevant information and keeping tracks of possible feedbacks.

The River Net3 component, which comes out from the work of Francesco Serafin, (Serafin *et al.* , 2016; Serafin, 2016), is an attempt to answer to this necessity. It is a new flexible structure that is going to be part of the core of OMS3, based on graphs for commanding simulations of complex systems. Parsing the graphs, Net3 determines and launches processes that can be sent in parallel.

River Net3 component considers a general scheme of catchment model by using catchment natural topology: the watershed

comprehensive of all its features, such as HRUs or sub-basins, channels, monitoring points, lakes, dams and so on, as in figure 2.6, is schematized through a simple list of nodes, identified by their own unique key.

Taking information from the variables of the system, the structure allows the execution of several contemporary processes for independent nodes and is able to spatially connect them to the river outlet.

More information about the implementation can be found at Serafin *et al.* (2016); Serafin (2016).

Figure 2.7 shows the schematization of the Posina River basin, Italy, in 42 HRUs. Each circle represents a node of the network, defined in the confluence of two links and in the measurement points. Each node could have one or more "parent" node and one or more "children" to be connected. According to the tree structure, for example, HRUs number 26, 35 and 37 can be processed in parallel. HRUs 26 and 35 are the "parents" of the HRU 27: only when their simulations are completed, the simulation for HRU 27 is launched and then combined with the results of the HRU 37, and so on, till HRU 9, which represents the closure of the basin.

Besides commanding the parallel execution to processes at HRU level, Net3 also can help to subdivide the preparatory work for a large river basin. In that case, subbasins that do not share areas can be inserted in different graph structures, tested separately and joint eventually.

The River Net3 component is, at the moment, under continuous upgrades and improvements, in order to make it more flexible and reliable. The work is part of Francesco Serafin doctoral project and in the present study was deeply tested with the Posina River case.

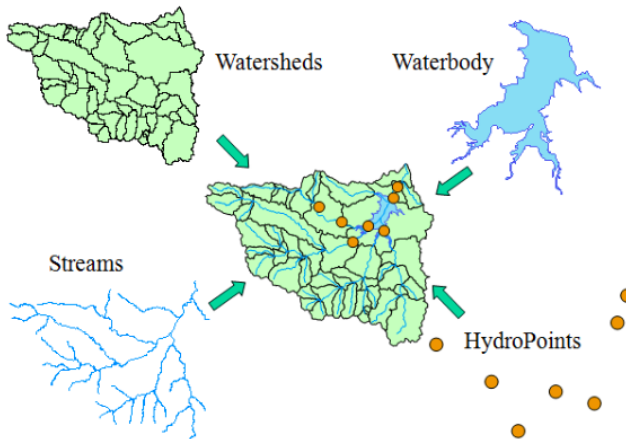


Figure 2.6: ArcHydro framework input data.

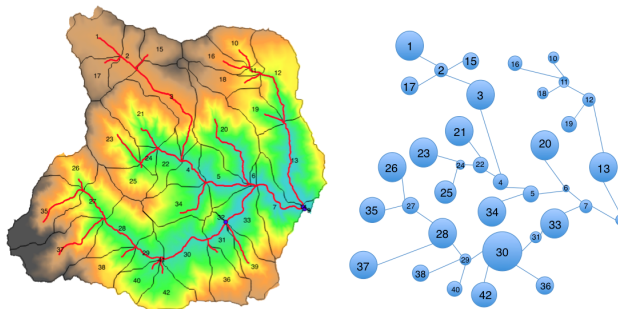


Figure 2.7: Posina river network schematization according to the River Net3 component: nodes are connected in cascade to the river outlet and simulations are launched according.

## 2.7 Final remarks

This chapter introduced the mathematics and informatics used throughout the thesis, which helps the reader to understand the foundation concepts of the present work.

The starting point was the hydrological modeling system JGrass-NewAge, presented by Formetta *et al.* (2014a). While working, in a sense, perfectly, JGrass-NewAGE system presented some issues and major revision for extensibility and reuse were necessary.

The chapter presented:

- the procedures implemented to revise the codes and keep track on code modification and evolution;
- the software improvements to extensibility and reuse, and the refactoring the codes by using design patterns;
- the software building system chosen for the growing set of tools;
- the libraries for basic numerical operations (like integrating systems of ordinary differential equations, fitting curves, and other tasks);
- formalized documentation procedures;

Moreover, the concepts of research reproducibility and replicability are presented, as well as, the suggested steps toward replicable science.

A mathematical-graphical representation of hydrological model, called Petri Nets, which will be used through the thesis, was then introduced. Finally, the last section is dedicated to description of a new flexible structure based on graphs for commanding simulations of complex systems, which represents an enhancement, recently introduced to OMS3 structure.





## A SEMI-DISTRIBUTED MODEL FOR RUNOFF AND EVAPOTRANSPIRATION

Many hydrological models are available in literature, (e.g., Refsgaard, 1995; Santhi *et al.*, 2001; Manfreda *et al.*, 2005; Therrien *et al.*, 2010; Birkel *et al.*, 2011), which vary in complexity and data requirements. The debate on the relative positive and negative aspects of distributed, physically based models versus lumped conceptual models is still ongoing, (Hrachowitz & Clark, 2017). Bottom-up model provide, at small scale, a comprehensive description of hydrological systems. They allow to account for the macropore flows, land-atmosphere exchanges, above and below non-isothermal processes and can be easily extended to model solute transport, (Fatichi *et al.*, 2016; Hrachowitz *et al.*, 2016). However, the use of these models at basin-scale is problematic: the large set-up, the characterization of the parameters and the elevated computational cost, make them still inconvenient for most of the applications at bigger scales, (Sivapalan *et al.*, 2003), and have hampered their widespread in the science community, (Fatichi *et al.*, 2016).

Recent advancement in computing, new data sources and collection methods, i.e. remote sensing, could help to overcome the problems and uncertainties related to such a models.

On the other hand, top-down models, despite of their simplicity, proved a good consistency and predictive ability, are faster in set-up and both computational times and data requirement are reduced. As stated in Levin (1999): "*simple models are a good place to start because their transparent features provide clarity...A simple model can provide a base for elaboration while capturing the essence of a variety of more detailed possible explanations*". Many application of such models proved to be valuable tools to represent the hydrological response, (Fenicia *et al.* , 2006; Hrachowitz *et al.* , 2013), large-scale patterns with self-similarity (Rodríguez-Iturbe & Rinaldo, 2001), as well as reproducing coupled flow and solute transport (Benettin *et al.* , 2013; Rigon *et al.* , 2016b,a). However, the physics of such model is not always well disclosed and they typically fail to reproduce dynamics in stream chemistry, as shown in Fenicia *et al.* (2008).

Grayson & Blöschl (2001) described qualitatively the trade-off between the complexity of the models and their performances: as shown in figure 3.1, for a given data availability there is an "optimum model complexity", beyond which the non-uniqueness reduces the predictive performances of the model. Quite often, in practical applications, too complex model with limited data are used, leading to identifiability problems.

In many modern modeling systems the idea is that the hydrology, at catchment scale, can be reduced to a set of interconnected reservoirs, (Clark *et al.* , 2008; Fenicia *et al.* , 2008; Tague & Dugger, 2010; Hrachowitz *et al.* , 2013; Manfreda *et al.* , 2005, 2012). They can be thought as "well mixed" systems, acting like chemical reactors, where all the inputs are uniformly distributed across the reservoir and perfectly and

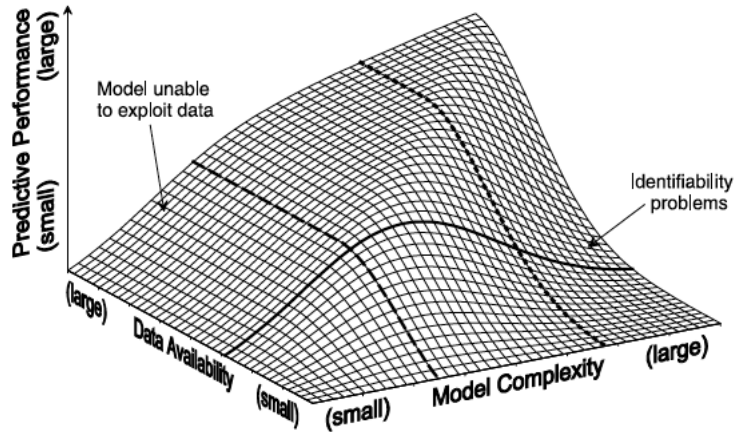


Figure 3.1: The figure represents the trade-off between the model complexity and the data availability. The figure was adapted from Grayson & Blöschl (2001).

instantaneously mixed with everything already in. Some of these reservoirs have a geographical identification, (Rinaldo & Rodríguez-Iturbe, 1996), while others, which we could define as "embedded", have functional reasons, i.e. to attribute the right travel time to water. "Vertical" reservoirs, such as canopy, vadose zone, groundwater, are distinguished (Fenicia *et al.*, 2006; Birkel *et al.*, 2011; Bertuzzo *et al.*, 2013; Hrachowitz *et al.*, 2013) and then spatially aggregated, (Rigon *et al.*, 2016b). The vertical and horizontal exchange of water between the reservoirs is typically expressed by a function of the water storage in the conceptually hierarchically higher reservoir.

Based on these premises, in this chapter an open-source semi-distributed model for runoff and evapotranspiration (from now on embedded reservoir model) is presented. The model schematizes each HRU as a connection of storages (reservoirs) and solves the water budget for each one. The work tries to find the correct number of storages to capture the complexity of the

hydrological response, making room for the separation between evaporation and transpiration, in a balance between the data availability and the computational costs. Five are the reservoirs considered: snow pack, canopy, root zone, surface waters and groundwater. Identified inputs and outputs are:

- snow melting and SWE;
- throughfall;
- evaporation from canopy;
- evapotranspiration from the root zone;
- percolation to the groundwater;
- direct runoff;
- baseflow.

In the almost-infinite panorama of hydrological models, the embedded reservoir doesn't aim to represent the perfect model but rather a modular system. The implementation of each reservoir as a Java component according to OMS3, within the hydrological system JGrass-NewAge, make it possible to connect/disconnect at run-time the reservoirs, according to the data availability, the HRUs characteristics (i.e., whether there is snow or not, whether there is a canopy or not) and, in general, to the modeler requirements. Flexible and extensible, the model allows to take into account a broad range of modeling strategies. Other worth mentioning models that allow such flexibility are FLEX model, Fenicia *et al.* (2011); Kavetski & Fenicia (2011), and SUMMA model Clark *et al.* (2015a,b). These frameworks are based on a general set of conservation equations for mass and energy, with the capability to incorporate multiple choices for

spatial discretization and flux parameterizations, (Clark *et al.* , 2015a).

The rationale behind the embedded reservoir, FLEX and SUMMA modeling systems is the same:

- The possibility to consider several representations of spatial variability and hydrologic connectivity;
- The possibility to simulate a broad range of hydrologic processes, with multiple options for individual processes.

However, the OMS-GEOframe system has the further ambition to offer a system to facilitate the production of tools and models to obtain the goal just presented. Therefore, the embedded reservoir model has a solid infrastructure, which make it well disposed to cover any future aspect which is now missing, lacking or simply to expand.

For all the details about the informatics, the model potentiality offered by OMS3 and mathematic preliminary, see sections 2.1 and 2.2.

### **3.1 Embedded reservoir model**

A semi-distributed model for runoff and evapotranspiration, which describes each HRU with five coupled storages is presented in this section. The number and the configuration of storages chosen was driven by several factors. First of all, the aim was to find the best model structure to explore the water age distributions. The reference starting model was the two layers model presented in Benettin *et al.* (2015a), in which two connected storages, shallow water and groundwater, were used to model the Upper Hafren catchment, mid-Wales (UK). The model well represented the hydrological response of the catchment, resulting in high values of goodness of fit between

the measured and simulated discharge, as well as between the measured and simulated chloride concentration.

However, the analysis of the characteristics of the area investigated in this study, the Posina river basin, led us to the necessity of considering more than two layers. In fact, Posina river basin, see section 5.1 for further details, is a densely vegetated area (around the 74% of the total), with beech, chestnut, maple and hazel as most common species. Therefore, a canopy storage was added to the initial configuration of two storages in order to model the canopy interception, evaporation and throughfall. Moreover, previous study on Posina river, (Norbiato *et al.* , 2009; Abera *et al.* , 2017a) led to the necessity of the introduction of a fourth reservoir accounting for the snow processes. In fact, Norbiato *et al.* (2009) concluded in his work that both the seasonal hydrological balance and the flood regime in the catchment are influenced by snow accumulation and melt. Eventually, a fifth storage was integrated for the surface flow, according to which the volume exceeding the maximum storage capacity of the root zone, goes into direct runoff and it is modeled either with the Width Function IUH (WFIUH), according to D'Odorico & Rigon (2003) or with a non-linear reservoir model.

A similar configuration of 5 reservoirs was proposed by Hrachowitz *et al.* (2013) but nor of the three applications considered the connection of all the five reservoirs connected contemporary. Moreover, different constitutive relationships were used .

DREAM model, (Manfreda *et al.* , 2005), was also a benchmark model taken in consideration in the present study. Four layers, canopy, surface depression, vadose zone and groundwater are considered and, also in this case, the validation gave very high performances. However, no snow process were considered and a

simplification on the modeling of the water stored in the surface depressions was necessary, in order to decrease the number of calibration parameters.

The final configuration of the present model schematizes each HRU through the following five embedded reservoirs:

- Snow pack;
- Canopy;
- Root zone;
- Surface flow;
- Groundwater.

After the detection of the rainfall and the snowfall from the total precipitation, the Hock model, (Hock, 1999), integrated in the snow component of JGrass-NewAge (Formetta *et al.* , 2013a), is used to simulate snow melting and snow water equivalent. The snow melting and/or the rainfall are the inputs of the canopy layer. Its outputs are the throughfall and the evaporation from the wet canopy. If there is no canopy, the melting/rain reaches directly the root zone. The variation of the water storage capacity between different points is described using a parabolic curve for the water storage capacity of the soil, according to (Zhao, 1980). The precipitation exceeding the root zone capacity is sent directly to the volume available for surface runoff, and modeled using the WFIUH or a non-linear reservoir model. The outputs from the root zone storage are the evapotranspiration and the recharge term of the groundwater. Evapotranspiration in the root zone describes both the evaporation from the soils and the transpiration from the canopy. A sixth storage to model the water within the canopy was initially considered but, in this phase, we preferred to not connect it, since we didn't want to

add further calibration parameters and complexities. Baseflow from the groundwater is modeled using a non-linear reservoir. Total runoff is the sum of the direct runoff and of the baseflow.

Figure 3.2 represent the embedded reservoir model using the time-varying Petri-nets. Each place has been implemented in a different component, in order to give the maximum flexibility of connections. Therefore, five different components forming a MS were used for each HRU in which the domain was discretized.

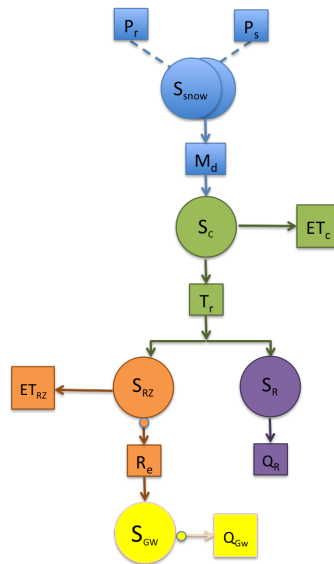


Figure 3.2: Representation of the embedded reservoir model using time-varying Petri-Nets. Five components are storage, snow, canopy, root zone, surface flow, and groundwater, which are represented through circles of different colors and specifications. Snow storage is represented using two overlapped circles, since it solved two coupled ODEs. Each storage has been implemented into a different Jgrass-NewAge component.

The detailed representation of each reservoir using PN, together with the table of associations and of symbols are discussed in the following sections.



### 3.1.1 Snow storage

Snow water equivalent and melting are simulated using the snow component presented in Formetta *et al.* (2013a). Three different models for the simulation of the melting can be chosen: a traditional temperature index method, Cazorzi & Dalla Fontana (1996) model and Hock (1999) model. In this study, we used Hock (1999) model, since it is based both on radiation and on temperature and the required inputs are easier to obtain than in Cazorzi & Dalla Fontana (1996). Figure 3.3 shows the PN representation of the snow, followed by the solved ODE for the two storages, the tables of association and specification of symbols.

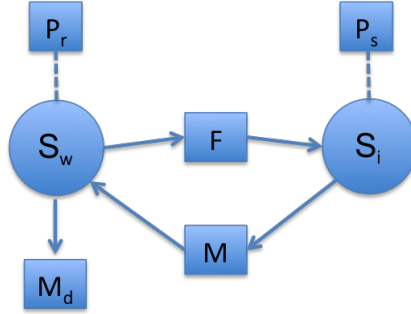


Figure 3.3: Representation of the snow reservoir model using time-varying Petri-Nets.

The snowpack mass balance is computed as follow:

$$\frac{dS_i(t)}{dt} = P_s(t) + F(t) - M(t) \quad (3.1)$$

$$\frac{dS_w(t)}{dt} = P_r(t) - F(t) + M(t) - M_d(t) \quad (3.2)$$

Symbol	Expression
$M_d(t)$	$\leftarrow M(t) \cdot (1 - \alpha_l)$
$F(t)$	$\leftarrow \alpha_f \cdot (T_m(t) - T(t))$
$M(t)$	$\leftarrow \begin{cases} (\alpha_m + \alpha_e \cdot R_s(t)) \cdot (T(t) - T_m(t)) \cdot V_s & T(t) > T_m(t) \\ 0 & T(t) < T_m(t) \end{cases}$

Table 3.1: Table of association between fluxes and expressions related to the snow storage.

Symbol	Name	Unit
$F(t)$	freezing water	$[L^3 T^{-1}]$
$M(t)$	melt	$[L^3 T^{-1}]$
$M_d(t)$	Melting discharge	$[L^3 T^{-1}]$
$P_r(t)$	rainfall	$[L^3 T^{-1}]$
$P_s(t)$	snowfall	$[L^3 T^{-1}]$
$R_s(t)$	shortwave radiation	$[W/L^2 T^{-1}]$
$S_i(t)$	solid water content in the snowpack	$[L^3]$
$S_w(t)$	liquid water in the snowpack	$[L^3]$
$t$	time	$[T]$
$T(t)$	temperature	$[^\circ C]$
$T_m(t)$	melting temperature	$[^\circ C]$
$V_s$	sky view factor	$[-]$
$\alpha_e$	radiation factor	$[L^0 C^{-1} E^{-1} T^{-1}]$
$\alpha_f$	freezing factor	$[L^0 C^{-1} T^{-1}]$
$\alpha_l$	liquid water retention capacity coefficient	$[-]$
$\alpha_m$	melt factor	$[L^0 C^{-1} T^{-1}]$

Table 3.2: List of symbols, names and units used in the snow storage representation.

After the interpolation of the air temperature, using Krigings techniques implemented in the Spatial Interpolation (SI) component, the shortwave radiation was computed with the SWRB component, Formetta *et al.* (2013b). The skyview factor map was obtained from the digital elevation model using the Horton machine, Rigon *et al.* (2006).

Further details about the snow component are in Formetta *et al.* (2013a).

### 3.1.2 Canopy storage

The simplest interception models are those which incorporate empirical, regression-based expressions relating interception loss to precipitation. These linear regressions can be used either to describe individual storm or daily interception loss as a function of daily gross rainfall, if only one rainfall event per day is assumed. Zinke (1967); Jackson (1975); Gash (1979) in their reviews of these simple regression models stressed that, despite their easiness, they don't always give satisfactory quantitative results when the regression coefficients are derived against site specific data.

To overcome the weakness of empirical models, physically based models can be adopted. One of the most rigorous physically based model is the Rutter *et al.* (1971), which solves the water balance equation for the wet vegetation for each time step. This model usually requires frequent (e.g. hourly) data inputs and is computationally demanding, limiting its applicability.

An alternative approach to the water budget solution for each time step is integrating the mass balance for the entire storm. The analytical interception model developed by Gash (1979) is one of the most satisfactory attempts, which retains much of the physical reasoning of the more complex Rutter model, despite a number of simplifying assumptions.

Sparse versions of the Gash (1979) and Rutter *et al.* (1971) adapted by dividing the forest into an open area with no cover and a covered area, were presented by Gash *et al.* (1995) and Valente *et al.* (1997). Both reformulated versions were shown to perform similarly well, Valente *et al.* (1997).

Many of the important models developed after Rutter, were its adaptation and given the new informatics tools available, the solution water-balance equation doesn't represent a computational problem, anymore. Therefore, we decided, in this

study, to model the wet canopy through a slightly modified version of the Rutter model. The original drainage function from the Rutter model was omitted, to avoid further calibration parameters. Aston (1979); Valente *et al.* (1997), suggested the same simplification. Figure 3.4 shows the PN representation of the canopy, followed by the solved ODE for the storage, the tables of association and specification of symbols.

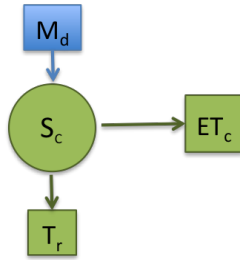


Figure 3.4: Representation of the canopy reservoir model using time-varying Petri-Nets.

The water budget solved for the canopy reservoir is given:

$$\frac{dS_c(t)}{dt} = M_d(t) - T_r(t) - ET_c(t) = (1 - p) M_d(t) - D(t) - ET_c(t) \quad (3.3)$$

Symbol	Expression
$D(t)$	$\leftarrow \max(0, b(S_c(t) - S_{c_{max}}(t)))$
$ET_c(t)$	$\leftarrow \max\left[0, ETp(t) \cdot \min\left(1, \frac{S_c(t)}{S_{c_{max}}(t)}\right)\right]$

Table 3.3: Table of association between fluxes and expressions related to the canopy storage.

The potential evapotranspiration is modeled using Priestley Taylor component, (Priestley & Taylor, 1972; Formetta *et al.* , 2014a), where the parameter  $\alpha$  was set equal to 1.26, as suggested in literature, (Cristea *et al.* , 2012) and according to Abera *et al.* (2017a).

Symbol	Name	Unit
$b_c$	drainage coefficient	$[T^{-1}]$
$D(t)$	drainage	$[L^3 T^{-1}]$
$ET_c(t)$	evaporation	$[L^3 T^{-1}]$
$ET_p(t)$	potential evapotranspiration	$[L^3 T^{-1}]$
$k_c$	LAI coefficient	$[L^3]$
$LAI$	Leaf Area Index	$[L^2 L^{-2}]$
$M_d(t)$	Melting discharge/rain	$[L^3 T^{-1}]$
$p$	free throughfall coefficient	$[-]$
$S_c(t)$	canopy storage	$[L^3]$
$S_{c_{max}}(t)$	canopy maximum retention storage	$[L^3]$
$t$	time	$[T]$
$T_r(t)$	throughfall	$[L^3 T^{-1}]$

Table 3.4: List of symbols, names and units used in the canopy storage representation.

The free throughfall coefficient was derived from measurements.  $S_{c_{max}}(t)$  is modeled as a function of a the time-varying Leaf Area Index (LAI)  $[m^2/m^2]$ , according to Dickinson (1984):

$$S_{c_{max}}(t) = k_c LAI(t) \quad (3.4)$$

Finally the total throughfall is computed as:

$$T_r(t) = D(t) + p \cdot M_d(t) \quad (3.5)$$

### 3.1.3 Root zone storage

Throughfall is partitioned into the infiltration in the root zone and the direct surface flow, according to the saturation conditions of the root zone. The storage accounts for the evaporation from the base soils, the transpiration of the plants and the recharge term of the groundwater.

Figure 3.5 shows the PN representation of the root zone, followed by the solved ODE for the storage, the tables of association and specification of symbols.

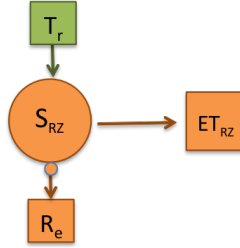


Figure 3.5: Representation of the root zone reservoir model using time-varying Petri-Nets.

The water balance equation solved for the root zone storage is written as:

$$\frac{dS_{rz}(t)}{dt} = (1 - \alpha(t))T_r(t) - Re(t) - ET_{rz}(t) \quad (3.6)$$

Symbol	Expression
$ET_{rz}(t)$	$\leftarrow \min\left(1, \frac{4}{3} \frac{S_{rz}(t)}{S_{rzmax}}\right) \cdot (ET_p(t) - ET_c(t))$
$Re(t)$	$\leftarrow aS_{rz}(t)^b$

Table 3.5: Table of association between fluxes and expressions related to the root zone storage.

Symbol	Name	Unit
$a$	coefficient of the non-linear reservoir model	$[T^{-1}]$
$b$	exponent of the non-linear reservoir model	$[-]$
$ET_c(t)$	evaporation	$[L^3 T^{-1}]$
$ET_{rz}(t)$	evapotranspiration	$[L^3 T^{-1}]$
$ET_p(t)$	potential evapotranspiration	$[L^3 T^{-1}]$
$Re(t)$	recharge term of groundwater	$[L^3 T^{-1}]$
$S_{rz}(t)$	root zone storage	$[L^3]$
$S_{rzmax}(t)$	maximum root zone storage	$[L^3]$
$t$	time	$[T]$
$T_r(t)$	throughfall	$[L^3 T^{-1}]$
$\alpha(t)$	partitioning coefficient between root zone and surface runoff	$[-]$

Table 3.6: List of symbols, names and units used in the root zone storage representation.

The partitioning coefficient  $\alpha(t)$ , is modeled according to Zhao (1980) and also implemented in Hymod model, (Moore, 1985),

while evapotranspiration is modeled using the formulation proposed by Manfreda *et al.* (2005).

### 3.1.4 Surface runoff

The volume exceeding the maximum storage capacity of the root zone, goes into direct runoff and it is modeled either with the Width Function IUH (WFIUH), according to D’Odorico & Rigon (2003) or with a non-linear reservoir model.

The width function is the distribution of the distances of any point in a basin to the outlet, computed following the drainage directions. It can be normalized by the total area to obtain a probability distribution function.

In order to account for the difference between hillslope velocity and channel velocity, which could be from ten to one hundred times, the rescaled width function, (Rinaldo *et al.* , 1995; D’Odorico & Rigon, 2003), is computed, i.e. the path followed by water molecules in hillslopes are magnified by a factor proportional to the ratio of the celerities of water in channels and hillslopes,  $r := u_c / u_h$ .

Moreover, since in humid areas runoff generation usually occurs over already saturated areas, (saturation excess, Dunne (1978)), the topographic index is used to map the runoff-contributing portions of the basin, (D’Odorico & Rigon, 2003). Precipitation falling on saturated areas contributes to the hydrologic response as overland flow, while precipitation falling on unsaturated areas infiltrates completely.

Figure 3.6 shows the PN representation of the surface runoff, followed by the solved ODE for the storage, the tables of association and specification of symbols.

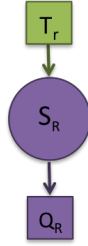


Figure 3.6: Representation of the surface runoff reservoir model using time-varying Petri-Nets.

$$\frac{dS_R(t)}{dt} = \alpha T_r(t) - Q_R(t) \quad (3.7)$$

Model	Symbol	Expression
WFIUH	$Q_R(t)$	$\leftarrow A \cdot \int_0^t uW(t-\tau)\alpha(\tau)T_r(\tau)d\tau$
Non-linear reservoir	$Q_R(t)$	$\leftarrow cS_R(t)^d$

Table 3.7: Table of association between fluxes and expressions related to the surface runoff storage.

Symbol	Name	Unit
$A$	Area of the HRU	$[L^2]$
$c$	coefficient of the non-linear reservoir model	$[T^{-1}]$
$d$	exponent of the non-linear reservoir model	$[-]$
$Q_R(t)$	surface flow	$[L^3T^{-1}]$
$S_R(t)$	runoff storage	$[L^3]$
$T_r(t)$	throughfall	$[L^3T^{-1}]$
$u$	wave celerity	$[LT^{-1}]$
$W(t-\tau)$	WFIUH	$[T^{-1}]$
$\alpha(t)$	partitioning coefficient between root zone and surface runoff	$[-]$

Table 3.8: List of symbols, names and units used in the surface runoff storage representation.

The rescaled width functions were computed for each HRU, starting from the map of the rescaled distances. The latter were obtained from the analysis the digital elevation model using the Horton machine, Rigon *et al.* (2006). Given the dimensions of the study area, the routing of the flow along the channel network



was neglected, according to Rinaldo *et al.* (1991); D’Odorico & Rigon (2003). Further details about the theory are shown in Chapter 4.

### 3.1.5 Groundwater

Finally, in the groundwater storage, recharge rate from the root zone, is transformed into baseflow using a non linear reservoir model. Figure 3.7 shows the PN representation of the groundwater, followed by the solved ODE for the storage, the tables of association and specification of symbols.

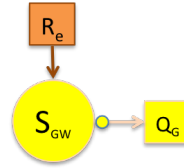


Figure 3.7: Representation of the groundwater reservoir model using time-varying Petri-Nets.

The water budget equation solved for the groundwater storage is written as:

$$\frac{dS_{GW}(t)}{dt} = R_e(t) - Q_{GW}(t) \quad (3.8)$$

Symbol	Expression
$Q_{GW}(t)$	$\leftarrow e^{-\frac{S_{GW}(t)}{S_{GWmax}(t)}} f$

Table 3.9: Table of association between fluxes and expressions related to the groundwater storage.

## 3.2 Final remarks

The chapter presented from the theoretical point of view, the embedded reservoir model. Five storages, snow, canopy, surface flow, root zone and groundwater, with multiple options for the

Symbol	Name	Unit
$e$	coefficient of the non-linear reservoir model	$[L^3 T^{-1}]$
$f$	exponent of the non-linear reservoir model	$[-]$
$Q_{GW}(t)$	groundwater discharge	$[L^3 T^{-1}]$
$Re(t)$	recharge term of groundwater	$[L^3 T^{-1}]$
$S_{GW}(t)$	groundwater storage	$[L^3]$
$S_{GW_{max}}(t)$	maximum groundwater storage	$[L^3]$
$t$	time	$[T]$

Table 3.10: List of symbols, names and units used in the groundwater storage representation.

description of the processes, were implemented and connected in a Modelling Solution. Each of the five storages was presented using the time-varying PN, which clarifies, in a unique way, the connections, the constitutive relations and the mathematical symbols used. The proposed schema of connections, is just one of the almost infinite possible MSs that could be obtained using the proposed infrastructure. As a proof of concepts, an application of the embedded reservoir modelling is presented in Chapter 5, together with examples of 3 other possibile MSs, obtained with different connections of the proposed reservoirs.

## TRAVEL TIMES THEORY

In this chapter we present a brief overview of Geomorphological Instantaneous Unit Hydrograph (GIUH) theories and a new formal approach to estimate the water budget by "travel times". The history of the GIUH is subdivided into two major sections. The first is based on the milestone works of Rodríguez-Iturbe & ValdÈs (1979a), and Gupta & Waymire (1983), which recognized that a treatment of water discharges with "travel times" could give a rich interpretation of the theory of the Instantaneous Unit Hydrograph (IUH). We show how this was possible, what assumptions were made, which of these assumptions can be relaxed, and which have become obsolete and been discarded. The second section focuses on the Width Function Based IUH (WFIUH) approach and its achievements in assessing the interplay of the topology and geometry of the network with water dynamics. The limitations of the WFIUH approach are described, and a way to work around them is suggested. Finally, the theory of travel time and residence time distributions is reworked from the point of view of the

hydrological storages and fluxes involved. The forward and backward travel time distribution functions are defined in terms of conditional probabilities. Previous approaches that used fixed travel time distributions are not consistent with our new derivation. We explain Niemi's formula and show how it can be interpreted as an expression of the Bayes theorem. Some connections between this theory and population theory are identified by introducing an expression which connects life expectancy with travel times. The theory can be applied to conservative solutes, including a method of estimating the storage selection functions. An example, based on the Nash hydrograph, illustrates some key aspects of the theory. Generalization to an arbitrary number of reservoirs is presented.

## **4.1 The geomorphological unit hydrograph from a historical-critical perspective**

Here we discuss the evolution of the geomorphological unit hydrograph in its attempts to assess the interplay of geomorphology and dynamics in estimating stream flows and the overall hydrological budget. In order to do so, we re-read the last thirty five years of evolution of the theory, which we have divided in two parts. The first part covers the classical theory (GIUH), the second looks at the theory based on the width function (WFIUH)

### **4.1.1 Early contributions**

With the 1979a paper "The geomorphologic structure of the hydrologic response", Rodríguez-Iturbe & ValdÈs, marked the beginning of a new era in rainfall-runoff models. The use of geomorphological information to assist in defining the unit hydrograph (or more general hydrological response functions

such as travel time distributions), and the conceptualization of hydrologic response as the convolution of travel time distributions, was represented with a mathematically neat method. It allowed to build up the basin hydrological response, based on spatial information, and synthesized a long history of development in a way that many immediately embraced. The rigorous treatment of the matter made in Gupta *et al.* (1980) consolidated the achievements. Although it may not have been very clear to the early readers of the paper, the treatment was actually quite general and coupled with the method developed to treat groundwater stochastically (e.g. Dagan, 1989), in which a very general formalism was developed to treat any problem per travel times. When the Rodriguez and Valdes paper was written, it was not yet the big data era and digital elevation models were not so pervasive then, nor available for all the earth's surface. River geomorphology was known only from traditional printed maps, and many river courses, especially in the tropics, were hardly known at all. Therefore, the paper also tried to use the information about the shape and form of rivers, given by knowledge of Horton's law of bifurcation ratios, length ratios, area ratios, and Schumm's law of slopes (e.g. Rodríguez-Iturbe & Rinaldo, 2001; Cudennec *et al.*, 2004). According to them, a river's drainage structure could be summarized by only a few numbers, mainly the bifurcation ratio and the length ratio: the first was used to describe the geometrical extension of the river network, and the second to provide the mean travel times in each part of the network. To move from the drainage structure to the hydrograph, a fundamental hypothesis had to be made: during floods the wave celerity could be considered constant along the network, as supported by Leopold & Maddock (1953). In theory, the constancy of celerity was necessary only within each partition of the basin (i.e. in each HRU or state used for its

disaggregation) and not in the overall network (as was actually done in many studies for practical purposes), and actually this assumption can be fully relaxed. Formally, the main equation summarizing all of this reads:

$$Q(t) = A \int_0^t p(t-\tau) J_e(\tau) d\tau$$

$$p(t) = \sum_{\gamma \in \Gamma} p_{\gamma} (p_{\gamma_1} * \dots * p_{\gamma_{\Omega}})(t) \quad (4.1)$$

where  $A$  is the area of the basin,  $J_e$  is the effective precipitation (i.e. the part of precipitation that contributes to the discharge),  $p$  is the instantaneous unit hydrograph, (i.e. the travel times probability distribution function, e.g. Rinaldo & Rodríguez-Iturbe (1996)),  $p_{\gamma}$  is the probability of precipitation hitting each HRU, labeled  $\gamma$  in the set  $\Gamma$  of all the HRUs, and  $(p_{\gamma_1} * \dots * p_{\gamma_{\Omega}})(t)$  denotes the convolution of the distribution of travel times of the water molecules in each state along the path, from the  $\gamma_1$ -th state to the  $\gamma_{\Omega}$ -th one. Technically, the product in (4.1) between  $p$  and  $J_e$  is itself a convolution between the overall travel time distribution with the effective rainfall and, therefore, the notation can be simplified:

$$Q(t) = A \sum_{\gamma \in \Gamma} p_{\gamma} (J_e * p_{\gamma_1} * \dots * p_{\gamma_{\Omega}})(t) \quad (4.2)$$

The summation over the paths implies linearity of the response, while the fact that  $p$  is not conditioned by the injection time implies that the theory is time invariant.

In this way, the theory was very economic: with a few parameters (not explicitly written in equation 4.1)) the entire description of the river basin morphology could be acquired through paths and states and be effectively used to get the hydrologic response. In the original paper the Horton stream hierarchical organization was chosen to identify HRUs and their

characteristics, but this was not, in fact, necessary. The approach for travel time could have been applied, following other heuristics, for different partitions of the river basin. A unique choice of  $J_e$  (i.e. of a uniform precipitation rate per unit area) is explicit in (4.1). However spatially variable rainfall could be easily introduced into GIUH theory by varying the rainfall input in each "state" (Cudennec *et al.*, 2005), assuming

$$p_\gamma := \frac{J_\gamma(t) A_\gamma}{J_e(t) A} \quad (4.3)$$

where  $J_\gamma$  is the amount of effective precipitation pertaining to the  $\gamma$ -th HRU, and  $A_\gamma$  is its area.

In older theories, travel time probability distribution functions were often guessed, and exponential or uniform distributions were used (Dooge, 2003), with the notable exception of the use of time-area functions. However, they can be obtained by solving a (simplified) momentum budget equation under impulsive boundary conditions and null initial conditions. Consider the case of the kinematic wave propagating with constant celerity  $u_s$  along a 1D domain, e.g. Beven (2011). The wave equation is:

$$\frac{\partial p(t, x)}{\partial t} + u_s \frac{\partial p(t, x)}{\partial x} = 0 \quad (4.4)$$

with initial condition  $p_\gamma(x, 0) = 0$ , and boundary condition  $p_\gamma(0, t) = \delta(t) / A_\gamma$ . If  $\hat{p}(x, t)$  is the wave function, solution of equation (4.4), with the equation conservative and the input unitary, it represents for any  $x$  a probability distribution function in  $t$  (since its integral remains unitary for any position). This probability, having as argument the time elapsed from the injection, can be identified with the residence time distribution. Fixing  $x$  at the outlet, denoted as  $x_\Omega$ , we obtain the distribution of water molecule travel times. Because of the linearity of the equation, the general solution for a generic boundary condition,  $J_e(t)$ , can be simply obtained by convolving it with the solution

of (4.4), (e.g. Morse & Feshbach, 1953). So

$$p(t|x_\Omega) = \int_0^t \hat{p}(t - t_i|x_\Omega)J(t_i)dt_i \quad (4.5)$$

which has, unsurprisingly, the form of (4.1). The process is reiterate to downstream states, if we assume as their input, the output of (4.5). Linearity is further used to sum contributions in parallel, according to the old linear theory of IUH (Dooge, 2003), and, as a matter of fact, this completely explains the structure of (4.1), where, in addition, the terms have a geomorphological explanation derived from the HRU tiling of the catchment.

The Mesa & Mifflin (1986) dispersion equation, in turn, can be obtained by using the linear equation:

$$\frac{\partial p(t, x)}{\partial t} + u_s \frac{\partial p(t, x)}{\partial x} = D \frac{\partial^2 p(t, x)}{\partial t^2} \quad (4.6)$$

where  $D$  is a hydrodynamical Dispersion coefficient, for which a complete treatment is given in Rinaldo *et al.* (1991), and can be seen as an approximation of the complete de Saint Venant equation (Henderson, 1966).

#### 4.1.1.1 Limitations of the early theory

As described above, the theory is quite general and can accommodate a lot of spatial heterogeneity, different celerities for different paths, space-variable precipitations, diffusive processes, and any model parameter can be varied for each "state" into which the basin is partitioned. The main limitations come from the richness of the foundational paper itself, which aimed also to synthesize the structure of the network by using the Horton laws, and its claim to solve the problem of predicting hydrological responses in ungauged basins. That is certainly an incorrect assumption, as will be made clear below. Parameters such as celerities and velocities must be considered as space-time averages but they are known to vary consistently



from event to event, according to the maximum stage produced, and this cannot be easily accounted for. Many authors, however, worked to obtain a geomorphoclimatic approach (Rodríguez-Iturbe *et al.* , 1982; Córdova & Rodríguez-Iturbe, 1983; Hall *et al.* , 2001; Grimaldi *et al.* , 2012) in which the parameters were constant for an event, but they could be changed event by event according to the total amount of rainfall. Clearly the theory was an "event-based" one, not aiming to follow the variability of flow velocities during any low stage. Another limitation is the capacity of process description. Because such a theory is only able to describe water transfer and aggregation, it must be associated with a production function in order to provide an estimate of the effective rainfall (Wang *et al.* , 1981; Gupta & Mesa, 1988; Robinson *et al.* , 1995; Woods & Sivapalan, 1999; Sivapalan *et al.* , 2002; Sivapalan, 2003). In recent works some inverse theory has been used to obtain it (de Lavenne *et al.* , 2015) but over the years, researchers and practitioners have mostly used empirical methods, like the Soil Conservation Service's Curve Number, e.g. (Dingman, 1994), to split the rainfall and to obtain the desired quantity. They assumed the hydrograph as the product of a "quick flow" with possible minor contributions from the subsurface, summed to a slower flow, mostly understood to be due to soil flow or groundwater. Normally, this is quite a naive interpretation of the processes, which mostly assumes infiltration excess mechanisms in producing runoff, does not account for mixing of waters from different events, and not at all for evapotranspiration. In fact, it was implicitly assumed that the rainfall measured during an event was net of evapotranspiration.

#### 4.1.2 The rise of the width function based geomorphological approach (WFIUH)

During the late 'eighties digital elevation models (DEMs) started to be available (Band, 1986; Fairfield & Leymarie, 1991; Tarboton *et al.*, 1991) and there was interest to understand how their information could fit into rainfall-runoff models. The old concept of isochrones Ross (1921), or, e.g. Beven (2011), could be derived from the definition of width function (Kirkby, 1976), under the hypothesis of constant celerities throughout the network. The width function, which actually anticipated the GIUH concept, after Kirkby (1976), is the number of sites at the same distance from the outlet, as measured along the streams and the drainage paths. Usually, it is denoted as  $W(x)$  in literature. Assuming that the paths are followed at constant celerity, the travel time probability,  $u_s$ , can be expressed as:

$$p_w(t - t_i) = \int_0^t w(x) f(t - t_i | x) dx \quad (4.7)$$

where  $w(x)$  is the normalized width function (its integral is unitary) and  $f(t - t_i | x)$  is the travel time distribution conditional to the distance to outlet  $x$ , which can be, for instance (e.g. D'Odorico & Rigon, 2003):

$$f(t|x) = \frac{x}{\sqrt{4\pi Dt^3}} \exp \left[ -\frac{(x - u_c t)^2}{4Dt} \right] \quad (4.8)$$

where diffusive processes were accounted for as in Rinaldo *et al.* (1991), and therefore  $D$  is the same diffusion coefficient that appears in (4.6). Equation (4.7) is similar to (4.1) as soon as the integral is replaced by its discrete summation over paths counterpart.

Regarding WFIUH, in the case where the parameters are kept constant, the HRUs are groups of basin pixels, not necessarily connected, at the same distance from the outlet, as shown in

Figure 2 where "strips of pixels" in the same interval of distances are shown.

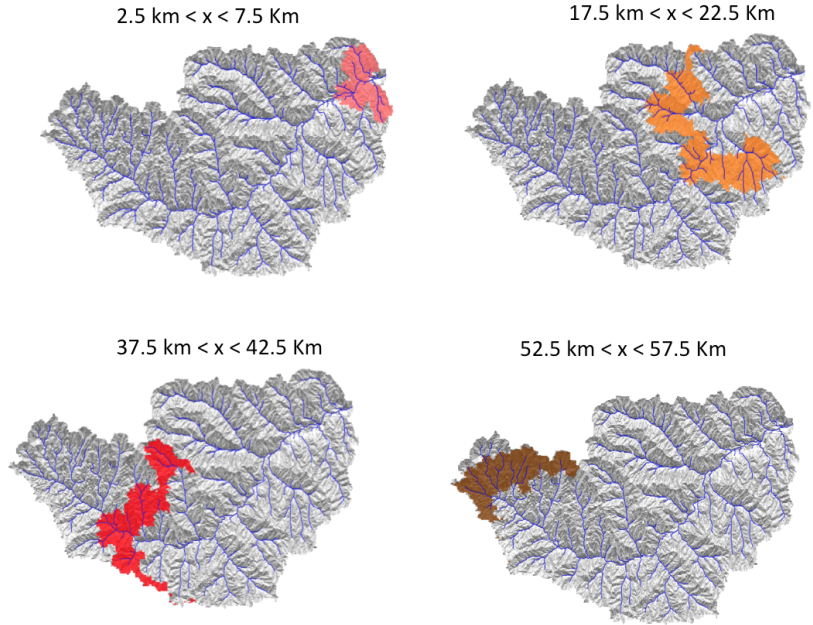


Figure 4.1: A basin can be continuously subdivided into strips of terrain at the same distance from the outlet. These strips are not necessarily continuous. They are physically connected to the outlet by many channels, but of these channels the physically significant quantity is the length, and therefore they are mathematically equivalent.

Therefore, the overall expression of the WFIUH is such that:

$$Q(t) = \int_0^t J(t_i) \int_0^t \frac{x w(x)}{\sqrt{4\pi D(t-t_i)^3}} \exp\left[-\frac{(x-u_c(t-t_i))^2}{4D(t-t_i)}\right] dx dt_i \quad (4.9)$$

The paper by Rinaldo *et al.* (1991) also showed that the detailed description of the hydrodynamic processes within the basin was not necessary, since the main contributions to the moments of the travel time distribution (the first and the second, in that paper) derives not from hydrodynamics but from the arrangement of paths in forming the width function structure. This phenomenon was termed *geomorphological dispersion*,

meaning that the width of the hydrograph is due to the summation of the hydrodynamic dispersion, and the geomorphological dispersion, and the latter dominates the former by an order of magnitude. The analysis in Snell & Sivapalan (1994), where the geomorphic structure of various Australian catchments was analyzed, reinforced the concept that describing the geometry of drainage through the width function was more acceptable than using the Horton laws. Naden (1992) and Naden *et al.* (1999), with very similar mathematics, showed the validity of the approach in reproducing discharges of large basins. Subsequently, two main additions were made to the original width function theory (and by transitive property also to the general theory of GIUH). In the early theories there was an overvaluation of the travel time in channels against travel times in hillslopes. The work by van der Tak & Bras (1990) brought attention to this fact and tried to generalize the theory by envisioning a system where two celerities were used to describe the flow of water, one for hillslope velocity and one for channel velocity, which literature reports to be for ten to one hundred times larger than runoff velocities (eg. Grimaldi *et al.* (2010, 2012)). One simple way to introduce these two velocities into the WFIUH was to consider the path followed by water molecules in hillslopes magnified by a factor proportional to the ratio of the celerities of water in channels and hillslopes,  $r := u_c / u_h$ , such that for any point in a hillslope at distance  $x_h$  from the channel, and  $x = x_h + x_c$  from the catchment outlet (being therefore  $x_c$  the length of the path along the channels), we can consider a *rescaled distance*:

$$x' = x_c + r x_h \quad (4.10)$$

Once this rescaled distance is estimated a unique celerity (i.e. the celerity in channels,  $u_c$ ) is enough to obtain the WFIUH distribution function (Rinaldo *et al.* , 1995; D'Odorico & Rigon,

2003). An example of map of distances to the outlet is reported in Figure 3a, while Figure 3b reports the rescaled isochrones.

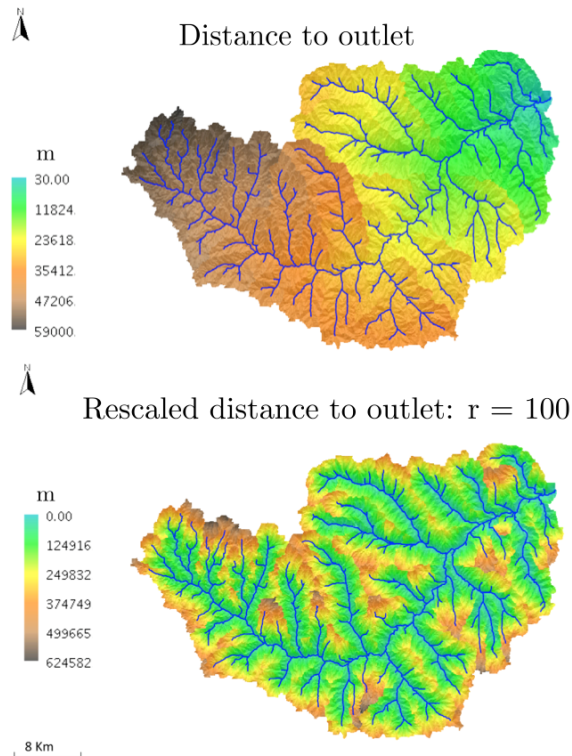


Figure 4.2: (a) Maps of the distances to outlet. their distribution is the width function. (b) Map of the rescaled distances to outlet. Rescaling factor is  $r=100$ .

The merits of the WFIUH are, firstly, to offer a theory which is approachable in a semi-analytical way and, secondly, to make directly apparent the influence of the geomorphic structure of the basin on the hydrograph. Under the simplifying hypothesis of uniform effective rainfall, Rigon *et al.* (2011) was also able to determine the equations for estimating peak flow timing and various related quantities, and to rigorously clarify old concepts like that of concentration time.

One of the main results of the WFIUH regarded the relative role of hillslopes and channels in forming the hydrograph. The WFIUH approach was able to assess that hillslope description of the flow dominates the shape of the hydrograph, since flow celerities in hillslope are much lower than celerities in channels, and cannot be neglected, even in large basins extending for tens of thousand of square kilometers (D’Odorico & Rigon, 2003), unless runoff is generated very close to streams. This is particularly evident when the moments of the discharge distribution are analyzed (Botter & Rinaldo, 2003).

Saco & Kumar (2002a,b) introduced statistical variability in flood celerities, by showing that this produces a third type of dispersion which can be relevant in some cases. The approach of Saco & Kumar was a revisitation of the instantaneous response function (IRF) concept, introduced by ValdÈs *et al.* (1979) in the traditional GIUH approach. Outside the strict field of these theories, Robinson *et al.* (1995), showed that, in the overall response, the time span of precipitation counts and they were able to produce a quantitative assessment through the statistics of the expected values and variances of the travel time distribution.

#### **4.1.2.1 Limitations of the WFIUH theories**

Obviously, what is a true limitation for the general scheme of the GIUH (and not its specific deployments) is a limitation also for the WFIUH approach. One of the main strengths, but also one of the limitations, of the WFIUH theory is that, at least in its more manageable formulations, it freezes its parameters for the whole basin and does not allow the variability that the scheme of the original approach could allow. The rigidity of the scheme is also reflected in its incapacity to account properly (i.e. respect the geometry) for the distribution of rainfall. Even though this fact

has been found to not be extremely relevant in the formation of the hydrograph, for which the total precipitation volumes seems more significant (Smith *et al.* , 2004; Nicotina *et al.* , 2008; Zoccatelli *et al.* , 2011), it has been found that some variability increases the overall reproduction of events and hydrographs (Formetta *et al.* , 2011), at least in a continuous modeling approach. Moreover, recent studies by Pattison *et al.* (2014) showed that contiguous basins can have very different peak flow timing (as driven by strongly differing celerities) due to their soil cover. These cases could not be treated properly with a single width function approach.

This impasse can be overcome by assembling together different WFIUHs to describe a larger basin. Equation (4.1) would then become:

$$Q(t) = A \sum_{\gamma} (J_{\gamma} * w_{\gamma} * f_{\Omega})(t) \quad (4.11)$$

where  $J_{\gamma}$  is the precipitation fallen in the  $\gamma$ -th HRU,  $w_{\gamma}$  the travel time distribution in the HRU according to its (rescaled) width function description,  $f_{\Omega}$  is the convolution of travel time distributions of water molecules in states downstream of the  $\gamma$ -th HRU, and  $*$  denotes the convolution operation. Actually, using the HRU internal variability in this way opens to further disaggregation of its control volume. Should we consider just one (rescaled) width function or more than one, separating, for instance, surface contribution from subsurface flows? In a very recent work, de Lavenne *et al.* (2015), the path of considering variable layers of width functions was investigated and tested by objective methods. It was found that the separation of the whole hydrograph into a surface component (parameterized by two celerities), denoted as  $Q_{sup}$ , and a subsurface one, denoted by  $Q_{sub}$ , was the most acceptable complexity to describe the behavior of the case studies (six French basins of various form

and size). In this way:

$$Q(t) = Q_{sub}(t) + Q_{sup} \quad (4.12)$$

with:

$$Q_{sup} = \alpha A(J_e * p_w)(t) \quad (4.13)$$

and

$$Q_{sub} = (1 - \alpha) A(J_e * p_{sub})(t) \quad (4.14)$$

where the usual symbols and notation are used, except for  $p_{sub}$ , which represents the travel time distribution for subsurface flows, and a partition coefficient  $\alpha$ , which is a matter of calibration, is used to separate the rainfall into the two contributions. This separation is illustrated in Figure 4.

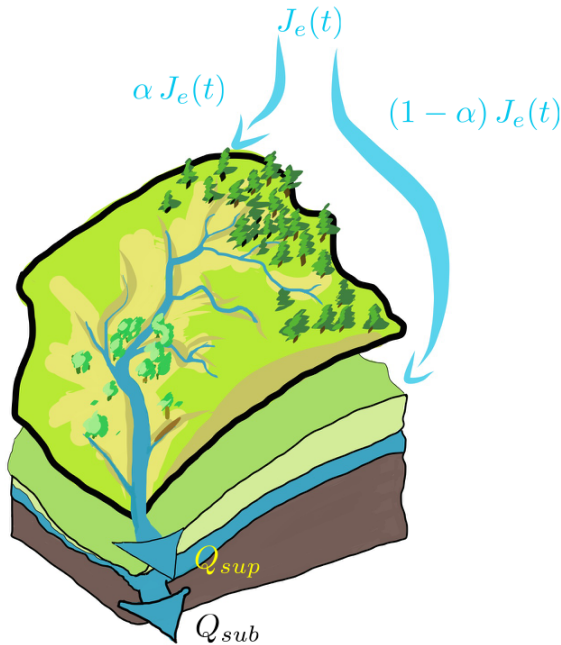


Figure 4.3: A HRU is "vertically" split into a surface and a subsurface domain. The overall response (to channel flow) is obtained just by the sum of the two contributions.

In the same study, notwithstanding the generally optimal response of the WFIUH theory, an accurate analysis of the



results over several events by means of "Cluster Heat Maps" (Wilkinson & Friendly, 2009) clearly showed that there were basins and events within the same basin whose behavior was captured in a relatively worse manner than others by using a unique set of width function parameters (celerities and diffusion coefficients constant in time). This could be interpreted as saying that the process of flood formation is usually slightly different even in close basins and from event to event, thus requiring a more flexible approach than using width functions in order to obtain improved results. A different question, which involves all the aforementioned types of GIUH and WFIUH, is that from this approach, through Niemi's theorem, Niemi (1977), the distribution of water ages,  $p(t - t_i | t)$  (i.e. the distribution of travel time conditional to the exit time, and consequently the mean *age of water*) can be estimated. Unfortunately the age derived from GIUH/WFIUH theories usually does not match the age found in field experiments with tracers. In a remarkable IUH approach Fenicia *et al.* (2008), by gradually increasing the complexity of their rainfall-runoff model by adding reservoirs in sequence or in parallel (convolving or summing travel time distribution functions in our mathematical approach), were able to show that tracer outputs can be properly reproduced, once verifiability is assured by data. It is therefore possible to envision that Fenicia's approach could possibly be adapted to the WFIUH-GIUH case. However, in doing so, adopting the more modern formalism illustrated in the next section could be convenient.

## 4.2 Age ranked approach

As reviewed in the previous section, hydrological travel times have been studied extensively for many years. Some researchers

(Rodríguez-Iturbe & ValdÈs, 1979b; Rinaldo & Rodríguez-Iturbe, 1996), looked at the construction of the hydrologic response using geographical information. Others (e.g., Uhlenbrook & Leibundgut, 2002; Birkel *et al.*, 2014) used travel times to understand catchment processes in relation to tracer experiments, while new experimental techniques were being developed (e.g., Berman *et al.*, 2009; Birkel *et al.*, 2011).

On these premises, (Fenicia *et al.*, 2008; Clark *et al.*, 2011; McMillan *et al.*, 2012; Hrachowitz *et al.*, 2013) aimed to describe both the spatial organization of the catchment and the set of interactions between processes with an assembly of coupled storages (reservoirs) in the number and the organization necessary to give proper hydrological results without adding unwanted parametric complexity (e.g., Klemeš, 1986; Kirchner, 2006). Despite the simplification efforts, the process of adding physical rigor to their models led to quite complex systems. Travel time analysis became a tool to disentangle flux complexities (e.g., Tetzlaff *et al.*, 2008), opening the way for explicit unification of geomorphic theories and storage-based modeling (Rigon *et al.*, 2016b).

A unique framework for understanding all catchment processes was made possible by the recent works of Botter and others (Rinaldo *et al.*, 2011; Botter *et al.*, 2011). This new branch of research is the focus of the present work. In particular, Botter *et al.* (2010) and Botter *et al.* (2011) introduced a newly formulated StorAge Selection function (SAS) related to the probability density function (pdf) of the water age or backward travel-time distribution. With the aid of an a-priori assigned SAS, they were able to write a "master equation" for the travel time probability distribution and solve it, thus systematically connecting the solution of the catchment water budget to travel time aspects of the hydrological flows. Older applications of the

travel time theory mostly assumed the simplest case of complete mixing, within the control volume, which is relaxed by using the SAS concept. Subsequently others van der Velde *et al.* (2012); Benettin *et al.* (2013, 2015b); Harman (2015b) introduced a new form of the SAS and the age-ranked distribution of water and associated compounds. Firstly, van der Velde *et al.* (2012) introduced the transformation from the "injection time" (i.e. when water enters the system), to the dimensionless ranked storage. Subsequently, Harman (2015b) suggested the use of the dimensional ranked storage instead of the dimensionless ranked storage, with the shortcoming of assuming an infinite storage in the parametrization.

These approaches opened the possibility of exploring the nature of storage-discharge relationships, which are usually parameterized within rainfall-runoff models, and can provide fundamental insight into the catchment functions invoked previously (e.g., Seibert & McDonnell, 2002; Kirchner, 2009). Also the traditional work on groundwater flow and catchment scale transport can be associated with the same ideas, but using time invariant travel time distributions (e.g., Dagan, 1984). Instead, Botter *et al.* (2011) used an approach that is inherently non-stationary and has immediately attracted the attention of researchers in that field (e.g., van der Velde *et al.* , 2012; Cvetkovic *et al.* , 2012; Cvetkovic, 2013; Ali *et al.* , 2014). A more detailed history of these concepts can be found in Benettin *et al.* (2013).

All of these studies provided valuable advances to the theory, but the literature remains obscured by different terminologies and notations, as well as model assumptions that are not fully explained.

There remains a need for theoretical developments that are clearly explained and developed using a consistent set of

notations. Questions arise, like: does the theory contain hidden parts that are not consistent or explained well? How does it relate to the instantaneous unit hydrograph theory? How can it be used? What generates time varying backward probabilities? Does the theory fully account for those phenomena which are involved in mobilizing old water (e.g., McDonnell & Beven, 2014; Rinaldo *et al.* , 2015; Kirchner, 2016)?

Questions also remain about how to apply the theory of age-ranked distributions in terms of the model form and parameter estimation. Harman (2015a) noted the importance of selecting an appropriate SAS, but until very recently (Harman, 2015b; Queloz *et al.* , 2015; Benettin *et al.* , 2017), there was no proposed method for selecting the form of an SAS and estimating it from available data. Selection of the SAS for a given watershed remains a topic of importance, since it should not be imposed arbitrarily.

Our work includes a short review of existing concepts that were collected from many (mostly theoretical) papers, which used different conventions and approaches. In the following sections, the theory to date is synthesized into a framework using consistent notation. Besides presenting the concept in a new and organized way, our paper contains some non-trivial answers to the above questions. Clarifications and extensions will be presented and summarized in an integrated manner. These conceptual developments are followed by improved methods for selecting the appropriate form of SAS and estimating its parameters. Guidance for hierarchical approaches to parameter estimation is given, based on available data. Finally, an extension of the age-ranked concepts to the embedded reservoirs model is presented.

### 4.2.1 Definitions of age-ranked quantities

Residence time, travel time and life expectancy of water particles and associated constituents flowing through watersheds are three related quantities whose meaning is well defined by the following equation:

$$T = \underbrace{(t - t_{in})}_{T_r} + \underbrace{(t_{ex} - t)}_{L_e} \quad (4.15)$$

where  $T$  [T] ([T] means time units) is the travel time,  $t$  [T] is the actual time measured by a clock,  $t_{in}$  [T] is the injection time (*i.e.*, the time at which a certain amount of water enters the control volume) and  $t_{ex}$  [T] is the exit time (*i.e.*, the time at which a certain amount of water exits the control volume). Based upon these definitions,  $T_r := t - t_{in}$  [T] is the so called residence time, or the age of water entered at time  $t_{in}$ , and  $L_e := t_{ex} - t$  [T] is the life expectancy of the same water molecules which are inside of the control volume.

Consider, for example, a control volume as the one shown in figure 4.4. Its (bulk) water budget is written as:

$$\frac{dS(t)}{dt} = J(t) - Q(t) - AE_T(t) \quad (4.16)$$

where  $S(t)$  [ $L^3$ ] is the time evolution of the water storage, ([L] denotes length units), but instead of volume, we can measure the storage either as mass, or a depth of water [L] (volume per unit area),  $J(t)$  [ $L^3 T^{-1}$ ] is the precipitation, usually a given (measured) quantity, while the discharge and the actual evapotranspiration,  $Q(t)$  [ $L^3 T^{-1}$ ] and  $AE_T(t)$  [ $L^3 T^{-1}$ ], are modeled. Common simple estimates for the two latter quantities are:

$$Q(t) = \frac{1}{\lambda} S^b(t) \quad (4.17)$$

and

$$AE_T(t) = \frac{S(t)}{S_{max}} E(t) \quad (4.18)$$

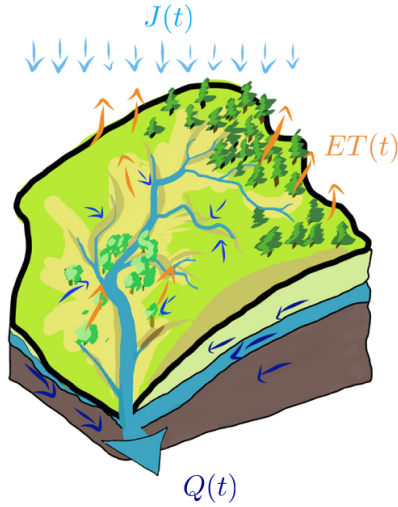


Figure 4.4: A single control volume is considered in which the fluxes are the total precipitation, evapotranspiration and discharge.

where  $\lambda [T]$  and  $b$  are the parameters of the non-linear reservoir model,  $S_{max}$  is the maximum water storage and  $E(t)$  is the potential ET, temporal function of the radiation inputs and atmospheric conditions. Assuming that radiation and various parameters used to model  $Q$  and  $AE_T$  are given, eq.(4.16) can be solved and  $S(t)$  obtained. If  $b = 1$  the budget is a linear ordinary differential equation, and its solution is analytic as in Coddington & Levinson (1955); otherwise, the solution can be obtained through an appropriate numerical solver (e.g., Butcher, 1987). We made the simplification here to use a single storage for illustrative purposes. However, extending the formalism to multiple storages is straightforward, as shown in section 4.2.10.

Being interested in knowing the age of water we need to consider a more general set of equations.

Assume that the water storage  $S(t)$  can be decomposed in its sub-volumes  $s(t, t_{in}) [L^3 T^{-1}]$  which refer to water injected into

the system at time  $t_{in} \in [0, t_p]$ . Thus:

$$S(t) = \int_0^{\min(t, t_p)} s(t, t_{in}) dt_{in} \quad (4.19)$$

where the initial time  $t = 0$  comes before any input into the control volume, and  $t_p$  represents the end of the last precipitation considered in the analysis. The variable  $t$  represents the actual time at which the storage is considered. In the following equations, the reference to  $t_p$  will be dropped for notational simplicity, and any quantity will consider a limited time interval. The functional form of  $s(t, t_{in})$ , as well as the functions we define below, can vary with  $t$  and  $t_{in}$ , so they should be labeled appropriately  $s_{(t, t_{in})}(t, t_{in})$  but this has been avoided for keeping notations simple.

Analogously,  $Q(t)$  [ $L^3 T^{-1}$ ] is the discharge out of the control volume, and  $q(t, t_{in})$  [ $L^3 T^{-2}$ ] is the part of the discharge exiting the control volume at time  $t$  composed of water molecules that entered at time  $t_{in} \in [0, t_p]$ :

$$Q(t) = \int_0^{\min(t, t_p)} q(t, t_{in}) dt_{in} \quad (4.20)$$

Actual evapotranspiration,  $AE_T(t)$  [ $L^3 T^{-1}$ ], is the sum of its parts  $ae_T(t, t_{in})$  [ $L^3 T^{-2}$ ] as:

$$AE_T(t) = \int_0^{\min(t, t_p)} ae_T(t, t_{in}) dt_{in} \quad (4.21)$$

Finally, let  $J(t)$  [ $L^3 T^{-1}$ ] denote the input to the control volume. This input can have an "age", and therefore, it can be defined

$$J(t) = \int_0^{\min(t, t_p)} j(t, t_{in}) dt_{in} \quad (4.22)$$

All these bivariate functions of  $t$  and  $t_{in}$ ,  $s(t, t_{in})$ ,  $q(t, t_{in})$ , and  $ae_t(t, t_{in})$  are null for  $t < t_{in}$  and can present a derivative discontinuity at the origin ( $t = t_{in}$ ). Given the above definitions,

we can rewrite the water budget as a set of age-ranked budget equations:

$$\frac{ds(t, t_{in})}{dt} = j(t, t_{in}) - q(t, t_{in}) - ae_T(t, t_{in}), \quad (4.23)$$

These equations were introduced first by van der Velde *et al.* (2012) and named by Harman (2015a), even though similar ones were present in previous literature, as discussed in Appendix B. In our formulation, however, by using  $t$  and  $t_{in}$  instead of  $t$  and  $T_r$  as independent variables, we do not need to transform the original ordinary differential equations (4.23) into partial differential equations.

#### 4.2.2 Backward and forward approaches

"Backward" and "forward" are well known concepts in the study of travel time distributions. They were first introduced by Niemi (1977), then by Cornaton & Perrochet (2006), for example, and recently refined by Benettin *et al.* (2015b). Benettin *et al.* (2015b), in particular, related the backward concept to the residence time (or age), while the concept of travel time is both a forward or backward. However, according to us, these previous works didn't fully disclose the inner meaning of the two concepts. In fact, in our theory, the probabilities are defined as backward when they "look" in time to the history of water molecules and forward when they "look" in time till their exit from the control volume. According to the previous statements, we can define a backward residence time probability, which is conditioned to  $t$  and "looks" backward to  $t_{in}$ , and a forward residence time probability, which is conditioned to  $t_{in}$  and "looks" forward to  $t$ . In the same way, we can define a backward life expectancy probability, which is conditioned to  $t_{ex}$  and "looks" backward to  $t$ , and a forward life expectancy probability,



which is conditioned to  $t$  and "looks" forward to  $t_{ex}$ . All these concepts will be clarified further in the following sections.

### 4.2.3 Backward Probabilities

Based on the previous definitions, it is easy to define the pdfs of the residence time, travel time and evapotranspiration time. In particular, the pdf of residence time, conditional on the actual time  $t$ , of water particles in storage,  $p_S(T_r|t)$ , can be defined as:

$$p_S(T_r|t) \equiv p_S(t - t_{in}|t) := \frac{s(t, t_{in})}{S(t)} [T^{-1}] \quad (4.24)$$

where " $\equiv$ " means equivalence, and "： $=$ " a definition. Benettin *et al.* (2015b) denoted  $p_S(T_r|t)$  as  $\overleftarrow{p}_S(T_r, t)$  but since this probability density is conditional to the actual time, standard probability notation is clear and unambiguous.

It is evident that this probability is time variant, since the integral and the integrand in equation (4.19) keep a dependence on the clock time  $t$ .

The pdf of travel time is  $p_Q(t - t_{in}|t)$ , where  $t_{ex} = t$ , since we are considering the water exiting the control volume as discharge. It can be defined as:

$$p_Q(t - t_{in}|t) := \frac{q(t, t_{in})}{Q(t)} [T^{-1}], \quad (4.25)$$

This definition for the probability is very restrictive, and can imply inconsistencies in those papers which assume a time invariant backward distribution to obtain tracers concentration, as shown in subsection 4.2.3.2. Eventually, the pdf of travel time for water exiting the control volume as water vapor,

$p_{E_T}(t - t_{in}|t)$ , can be defined as:

$$p_{E_T}(t - t_{in}|t) := \frac{ae_T(t, t_{in})}{AE_T(t)} [T^{-1}], \quad (4.26)$$

It is also possible to define the mean age of water for any of the two outlets, which is given by:

$$\langle T_r(t) \rangle_i = \int_0^{\min(t, t_p)} (t - t_{in}) p_i(t - t_{in} | t) dt_{in} \quad (4.27)$$

for  $i \in \{Q, E_T\}$ , which is a function of the sampling time (and the rainfall input).

After the above definitions, the age-ranked equation (4.23), can be rewritten as:

$$\frac{d}{dt} [S(t) p_S(T_r | t)] = J(t) \delta(t - t_{in}) - Q(t) p_Q(t - t_{in} | t) - AE_t(t) p_{E_T}(t - t_{in} | t) \quad (4.28)$$

when a single "new water" injection of mass is considered, and the bulk quantities  $S(t)$ ,  $Q(t)$ ,  $AE_T(t)$  are known as soon as the bulk water budget, equation (4.16), is solved.  $\delta(t - t_{in})$  is a Delta-dirac function to account for the water particles in precipitation with age zero. The travel time probabilities on the right side of (4.28) are not known. Consequently Botter *et al.* (2011) introduced a storage selection function,  $\omega(t, t_{in})$  [-], for each of the outputs, so that:

$$p_Q(t - t_{in} | t) := \omega_Q(t, t_{in}) p_S(T_r | t) \quad (4.29)$$

and:

$$p_{E_T}(t - t_{in} | t) := \omega_{E_T}(t, t_{in}) p_S(T_r | t) \quad (4.30)$$

Therefore equation (4.28), after the proper substitutions, becomes:

$$\begin{aligned} \frac{d}{dt} [S(t) p_S(T_r | t)] = & J(t) \delta(t - t_{in} | t) + \\ & -Q(t) \omega_Q(t, t_{in}) p_S(T_r | t) - AE_t(t) \omega_{E_T}(t, t_{in}) p_S(T_r | t) \end{aligned} \quad (4.31)$$

Once assigned the  $\omega(t, t_{in})$  values on the basis of some heuristic, as in Botter *et al.* (2011), equation (4.31) represents a linear ordinary differential equation and can be solved exactly as:

$$p_S(T_r | t) = e^{-\int_{t_{in}}^t g(x, t_{in}) dx} \left[ p(0 | t) + \int_{t_{in}}^t \frac{J(y) \delta(y - t_{in})}{S(y)} e^{\int_{t_{in}}^y g(x, t_{in}) dx} dy \right] \quad (4.32)$$

where :

$$g(x, t_{in}) = \frac{1}{S(x)} \left[ \frac{dS(x)}{dt} + Q(x)\omega_Q(x, t_{in}) + AE_t(x)\omega_{E_T}(x, t_{in}) \right] \quad (4.33)$$

and  $p(0|t)$  is the initial condition. This is only valid if equation (4.31) is linear, i.e.  $\omega(t, t_{in})$  is not a function of  $p_S(T_r|t)$ .

Figure 4.5 shows the variation of the  $p_S(T_r|t)$  with the injection time, while the chronological time is kept fixed. The curves were obtained considering three different injections at  $t_{in_1}$ ,  $t_{in_2}$  and  $t_{in_3}$ , and assuming  $\omega_Q(t, t_{in}) = \omega_{E_T}(t, t_{in}) = 1$ .

The conditional probability  $p_S(T_r|t)$  properly integrates to one, as shown in figure 4.6, when it is integrated in  $t_{in}$ . In particular, figure 4.6 shows that  $p_S(T_r|t) = const$ , when  $J(t) = 0$ . In fact, if we consider  $\omega_Q(t, t_{in}) = \omega_{E_T}(t, t_{in}) = 1$ , equation (4.31) is simplified in:

$$\frac{d}{dt}[S(t)p_S(T_r|t)] = -Q(t)p_S(T_r|t) - AE_t(t)p_S(T_r|t) \quad (4.34)$$

and, therefore,

$$\frac{dp_S(T_r|t)}{dt} = -\frac{p_S(T_r|t)}{S(t)} \left[ \frac{dS(t)}{dt} - Q(t) - AE_t(t) \right] = 0 \quad (4.35)$$

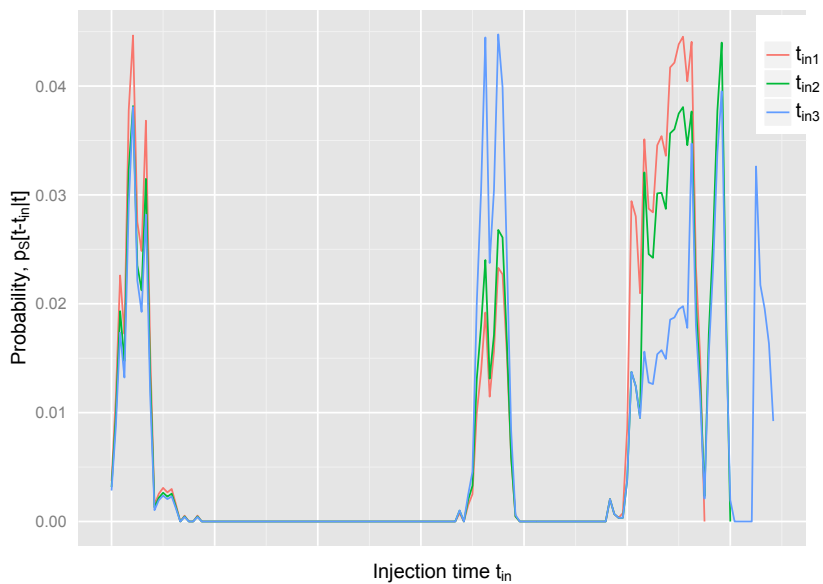


Figure 4.5: Representation of the evolution of the backward pdf for three injection times ( $t_{in_i}$ , where  $i = 1,3$ ) as varying with the injection time  $t_{in}$ . The time shift between the three injections was dropped for a direct comparison of the curves.

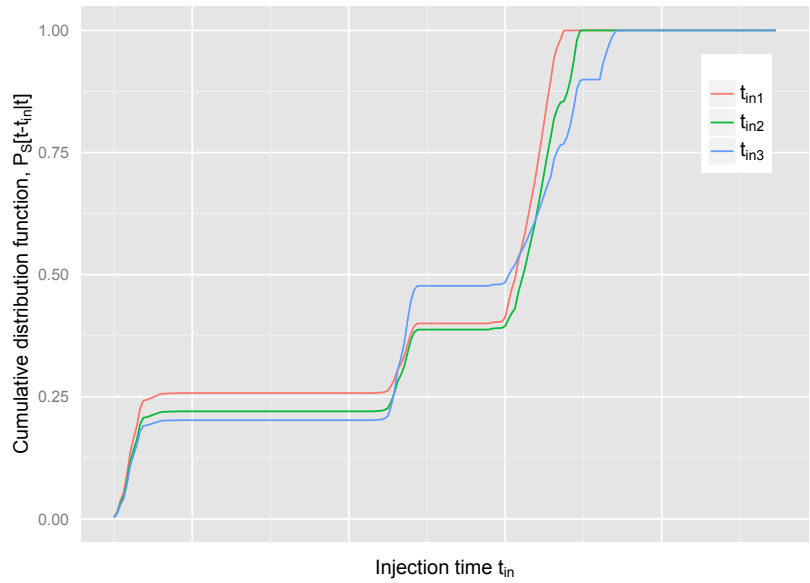


Figure 4.6: Representation of the backward cumulative distribution function for three injection times ( $t_{in_i}$ , where  $i = 1,3$ ), as varying with the actual time  $t$ . The time shift between the three injections was dropped for a direct comparison of the curves.

Figure 4.7 shows the evolution of  $p_S(T_r|t)$  with the actual time  $t$  and the injection time kept fixed. The integral of the area under the three curves, obtained for the same three injections, in this case, is not equal to 1, since the functions are not pdfs in  $t$ .

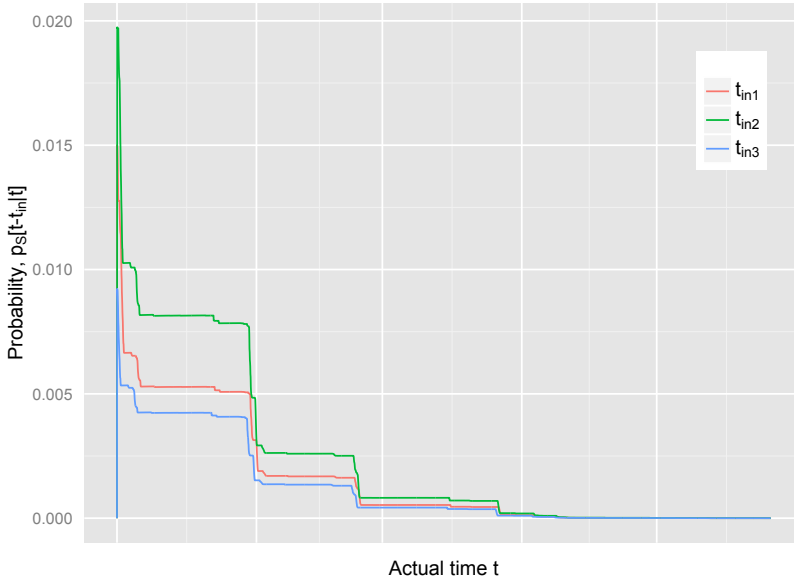


Figure 4.7: Representation of the evolution of the backward pdf versus the actual time  $t$ . The time shift between the three injections was dropped for a direct comparison of the curves. In this case, the area below the curves is not equal to 1.

#### 4.2.3.1 An extension of the theory considering antecedent storage

All the theory shown in the previous sections assumed to start at a time  $t = 0$  that we specified to be sufficiently back in the past. This choice was made for a better comparison with practical cases, where information are available for a limited range of times. Assume, however, that our system is not stationary and we have to add information to it starting from previous times (before  $t = 0$ ). The r.h.s of the equation 4.23 either contains forcings given as function of time or as a function of age-ranked

storages. These forcings represent a limitation to our knowledge, since usually we know them starting from a certain clock time, which we conventionally called  $t = 0$ . However, even if we cannot access to equations for forcings happened before  $t = 0$ , their actions can result in a storage different from zero.

Therefore, if we consider that initial conditions for equation 4.23 are given by  $s(t_{in}, t_{in}) = 0$  and  $s(t, t_{in}) \equiv 0$  for any  $t_{in} < t < 0$ , we have:

$$p_S(t - t_{in}|t) := \begin{cases} \frac{s(t, t_{in})}{\int_{-\infty}^t s(t, t_{in}) dt_{in}} & t > t_{in} > 0 \\ 0 & t_{in} < t < 0 \end{cases} \quad (4.36)$$

where:

$$S(t) = \int_{-\infty}^t s(t, t_{in}) dt_{in} = \underbrace{\int_{-\infty}^0 s(t, \hat{t}_{in}) d\hat{t}_{in}}_{S_0(t)} + \underbrace{\int_0^t s(t, t_{in}) dt_{in}}_{S_t(t)} \quad (4.37)$$

is the water present inside the control volume at time  $t$ , as results of available forcings ( $S_t(t)$ ) and past forcings ( $S_0(t)$ ). We indicated with  $\hat{t}_{in}$  precipitations timing previous than 0.  $S_0(t)$  is the initial storage, necessary also to solve the bulk water budget equation.

On the base of the above probability in 4.36, the mean residence time of the water at  $t = 0$  is:

$$\langle T_r \rangle_0 = \int_0^{\infty} T_r p_S(T_r|0) dT_r \quad (4.38)$$

where in the second integral in the r.h.s. of the equation,  $T_r \equiv 0 - \hat{t}_{in}$  and  $\hat{t}_{in} < 0$ . Next question is how we can assign the initial conditions on storages if we do not have explicit information about the forcings. An educated guess is that we can use a probability distribution function (pdf), for instance a  $\Gamma$  function or others. To stay generic let's say that

$p_S(T_r|0) = g(T_r|0)$  where  $g$  is a known pdf, and  $T_r = 0 - \hat{t}_{in}$ . The knowledge of  $g$  is not enough to extend the system of age-ranked equations to the past. However some inferences can be done. In fact, for any precipitation input at time  $\hat{t}_{in} < 0$ :

$$s(0, \hat{t}_{in}) = p_S(0 - \hat{t}_{in}|0)S(0) = g(T_r|0)S(0) \quad (4.39)$$

The mean residence time can be then estimated, according to its definition, and using  $T_r = t - t_{in}$  as:

$$\langle T_r \rangle_t = \int_t^\infty T_r g(t-t_{in}) dT_r + \int_0^t \underbrace{(t-t_{in})}_{T_r} p_S(\underbrace{(t-t_{in})}_{T_r} | t) d(\underbrace{(t-t_{in})}_{T_r}) \quad (4.40)$$

The first integral accounts for the residence time of water in input before  $t=0$  and the second term is the part simply calculated through the solutions of the age-ranked system but it embeds the initial storage in the definition of the probability. It is apparent that the first integral is decreasing with time and null after an amount of time that is strictly connected with the standard deviation of the pdf  $g$ . Also in the second integral, the effect of the initial storage is vanishing as soon as  $t$  increases. Neglecting the old waters contribution means, on the other side, to decrease the age of water, at least until  $t$  is large enough. This, based on actual knowledge, means that for a correct residence time simulation, models should simulate decades, which could be unfeasible.

At the same time, it seems not completely realistic to use a time-invariant travel time function for  $g$  because, it could be guessed that a time-variant distribution would be more appropriate, since its dependence of time varying forcings. But this is probably the best we can do.

The reader should also observe that we dealt with the residence time. Travel times can be obtained after assuming a SAS or doing appropriate equivalent hypothesis.



#### 4.2.3.2 An observation on fixing the functional form of the backward probability

It can be observed that the backward probability, as defined in (4.24) is quite restrictive, and not very compatible with the assumption of a time invariant backward distribution, often made in literature, (e.g. Kirchner *et al.* , 2000; Kirchner, 2016; Hrachowitz *et al.* , 2010). Most of these papers use a gamma distribution, i.e.

$$g(T_r) = \frac{T_r^{\alpha+1} e^{-\frac{T_r}{\gamma}}}{\gamma^\alpha \Gamma(\alpha)} \quad (4.41)$$

where  $g$  is the incomplete gamma distribution,  $T_r := t - t_{in}$  is the residence time,  $\alpha$  and  $\gamma$  are the two coefficient of the incomplete  $\Gamma$  distribution,  $\Gamma$  is the gamma function.  $g(T_r)$  in (4.41) is certainly a distribution though over the whole domain of  $T_r$ . However, equation (4.24) requires that  $g(T_r)$  would be a probability for any clock time  $t$ , i.e. that:

$$\int_0^{\min(t, t_p)} p_Q(t - t_{in} | t) dt_{in} = 1 \quad (4.42)$$

This is, clearly not obtained with (4.41) (or any other classical distribution), and, in fact,

$$\int_0^{\min(t, t_p)} \frac{(t - t_{in})^{\alpha+1} e^{-\frac{(t-t_{in})}{\gamma}}}{\gamma^\alpha \Gamma(\alpha)} dt_{in} \neq 1 \quad (4.43)$$

where in the formula the travel time  $T_r$  has been explicitly written as function of  $t$  and  $t_{in}$ . It could be argued that the above integral could be approximately equal to unity in real cases, and, seen the success of gamma based approaches to interpret experimental data, this could be true.

However, a better choice for the backward probability should be

a little more complex. For instance:

$$\begin{aligned}
 p_Q(t - t_{in}|t) &= \frac{g(t - t_{in})}{\int_0^{\min(t, t_p)} g(t - t_{in}) dt_{in}} = \\
 &= \frac{\frac{(t - t_{in})^{\alpha+1} e^{-\frac{(t-t_{in})}{\gamma}}}{\gamma^\alpha \Gamma(\alpha)}}{\int_0^{\min(t, t_p)} \frac{(t - t_{in})^{\alpha+1} e^{-\frac{(t-t_{in})}{\gamma}}}{\gamma^\alpha \Gamma(\alpha)} dt_{in}} \quad (4.44)
 \end{aligned}$$

works the right way.

#### 4.2.4 Forward Probabilities

Consider again the age-ranked equation (4.23). Since we want to track the evolution of a water particle while crossing the catchment, we can write its integral form over  $dt$ , as:

$$s(t, t_{in}) = J(t_{in}) - \int_0^t q(t, t_{in}) dt - \int_0^t ae_T(t, t_{in}) dt \quad (4.45)$$

It can be rewritten as a probability conditional to  $t_{in}$ :

$$P_S[t - t_{in}|t_{in}] := 1 - \frac{s(t, t_{in})}{J(t_{in})} = \frac{V_Q(t, t_{in})}{J(t_{in})} + \frac{V_{E_T}(t, t_{in})}{J(t_{in})} \quad (4.46)$$

having defined:

$$V_Q(t, t_{in}) := \int_0^t q(t, t_{in}) dt \quad (4.47)$$

and

$$V_{AE_T}(t, t_{in}) = \int_0^t ae_T(t, t_{in}) dt \quad (4.48)$$

$P_S[t - t_{in}|t_{in}]$ , as shown in figure 4.8, varies (with  $t$ ), as expected, between 0 and 1 and has density:

$$p_S(t - t_{in}|t_{in}) = -\frac{1}{J(t_{in})} \frac{ds(t, t_{in})}{dt} = \frac{q(t, t_{in})}{J(t_{in})} + \frac{ae_T(t, t_{in})}{J(t_{in})} \quad (4.49)$$

It can be observed instead that:

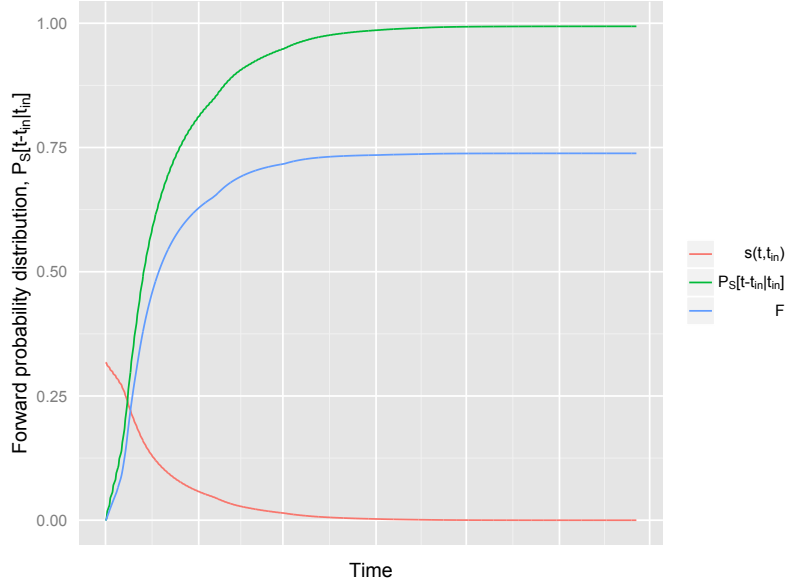


Figure 4.8: Forward residence time probability distribution: in red the relative storage, in green the forward residence time distribution and in blue the relative discharge function.

$$\mathcal{F}(t - t_{in} | t_{in}) := \frac{V_Q(t, t_{in})}{J(t_{in})} \quad (4.50)$$

and

$$\mathcal{G}(t - t_{in} | t_{in}) := \frac{V_{ET}(t, t_{in})}{J(t_{in})} \quad (4.51)$$

are not probability functions, because, their asymptotic value is not 1. Because the forward probabilities are derived, in the case we are describing, on empirical bases from the budgets terms, and not assumed a-priori, their complete shape is known only at  $t \rightarrow \infty$ . For any finite time, the actual knowledge we have, is better represented in Figure 4.9, which shows that the progress of the three curves  $P$ ,  $\mathcal{F}$  and  $\mathcal{G}$  is unknown for future times. In order to normalize  $\mathcal{F}$  and  $\mathcal{G}$ , the asymptotic value of the

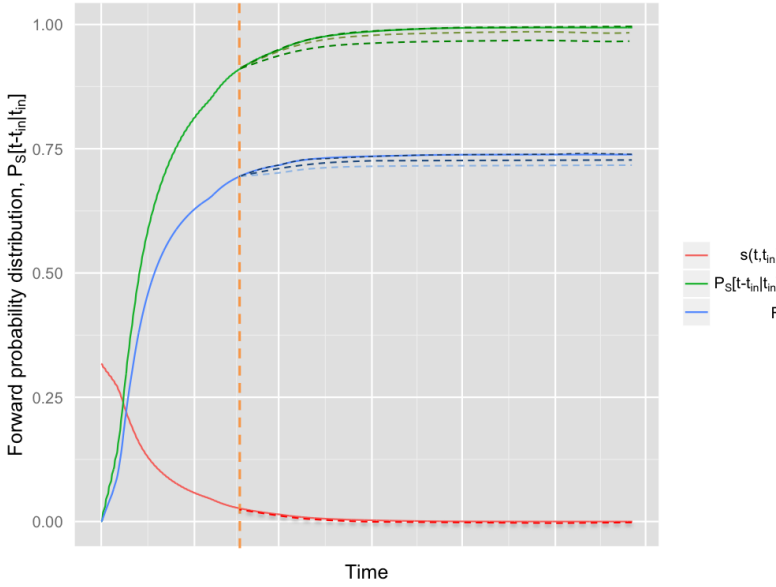


Figure 4.9: Representation of the forward probability of the outputs: in red the relative storage,  $s(t, t_{in})$ , in green the output probability,  $P_S[t - t_{in}|t_{in}]$  and in blue the relative discharge function  $\mathcal{F}$ , defined in the text. The difference between  $P_S[t - t_{in}|t_{in}]$  and  $\mathcal{F}$  is the function  $\mathcal{G}$ , defined in the text. The orange dashed line represents the generic instant  $t$ , after which  $P_S[t - t_{in}|t_{in}]$  and  $\mathcal{F}$  are unknown.

partitioning coefficient is defined among the  $Q$  and  $E_T$ :

$$\Theta(t_{in}) := \lim_{t \rightarrow \infty} \Theta(t, t_{in}) := \lim_{t \rightarrow \infty} \frac{V_Q(t, t_{in})}{V_Q(t, t_{in}) + V_{E_T}(t, t_{in})} \quad (4.52)$$

Then, it is easy to show that:

$$p_Q(t - t_{in}|t_{in}) := \frac{q(t, t_{in})}{\Theta(t_{in})J(t_{in})} \quad (4.53)$$

and

$$p_{E_T}(t - t_{in}|t_{in}) := \frac{ae_T(t, t_{in})}{(1 - \Theta(t_{in}))J(t_{in})} \quad (4.54)$$

are the forward probabilities density function of discharges and evapotranspiration, which properly normalize to 1 when integrated over  $t$ . The two probability density functions  $p_Q$  and

$p_{E_T}$  are related through:

$$p_S(t - t_{in}|t_{in}) = \Theta p_Q(t - t_{in}|t_{in}) + (1 - \Theta) p_{E_T}(t - t_{in}|t_{in}) \quad (4.55)$$

Unlike backward probabilities, the forward probabilities describe how a catchment reacts to precipitation events, but they do not describe the actual time the water takes to move through the catchment. To avoid confusion, the expected value of travel time, weighted by the forward distribution, will be called the mean response time (instead of mean travel time). For discharge, the result is:

$$Q(t) = \int_0^{\min(t, t_p)} p_Q(t - t_{in}|t_{in}) \Theta(t_{in}) J(t_{in}) dt_{in} \quad (4.56)$$

which can be seen as a generalization of the instantaneous unit hydrograph (IUH).

Although  $\Theta$  may be unknown at any finite time, the actual state of the system is obtained by solving the budget equation. More information and details on this partitioning coefficient are provided in the next section.

#### 4.2.5 The partitioning coefficient $\Theta$

$\Theta(t_{in})$  has been introduced to complete the algebra of probabilities, in presence of more than one outflow. However estimation of the coefficient is important by itself, because it summarizes the relevant partitioning of hydrologic fluxes. The first plot in figure 4.10 shows a time-series of  $\Theta(t, t_{in})$  values obtained from a single injection time considering the complete mixing case ( $\omega_Q(t, t_{in}) = \omega_{E_T}(t, t_{in}) = 1$ ). Data used for the simulation are from Posina River, a small catchment in the North Eastern part of pre-alpine mountainous in Veneto region, Italy. At the beginning  $\Theta(t_{in})$  (figure 4.10, top) shows large oscillations due to hourly and daily oscillations, especially in evapotranspiration. Because  $\Theta(t_{in})$  is defined through integrals,

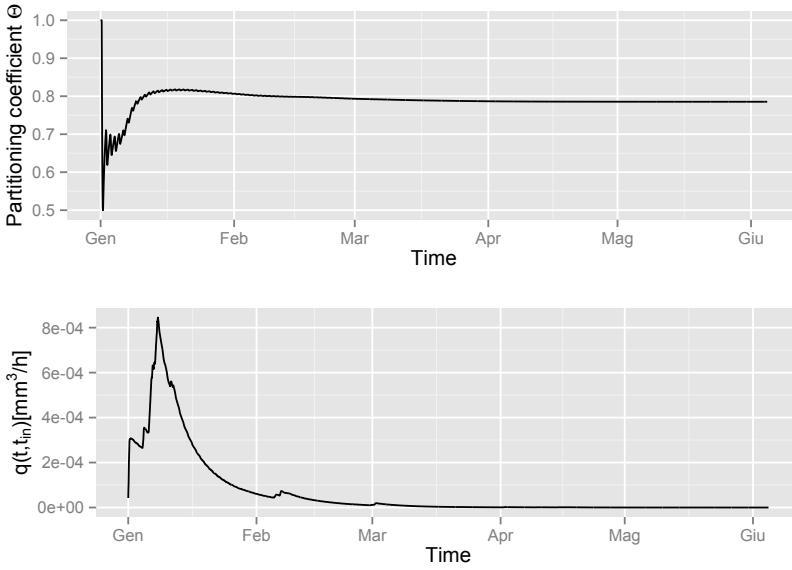


Figure 4.10: Variation of the partitioning coefficient in time, for a single injection time in January: after a time scale of 5 months its oscillation became irrelevant and its value tends to its final value of 0.78

these oscillation are progressively damped and become irrelevant after a couple of weeks (when discharge is still higher than baseflow, as appears from the age-ranked discharge in figure 4.10, bottom).

#### 4.2.6 Niemi's relation

As a result of definitions made in sections (4.2.3) and (4.2.4) two relations exist involving  $q(t, t_{in})$ , i.e. equations (4.25) and (4.53), and  $ae_T(t, t_{in})$ , i.e. equations (4.26) and (4.54). Equating the corresponding two expression, results in:

$$Q(t)p_Q(t - t_{in}|t) = \Theta(t_{in})p_Q(t - t_{in}|t_{in})J(t_{in}) \quad (4.57)$$

and:

$$AE_T(t)p_{E_T}(t - t_{in}|t) = [1 - \Theta(t_{in})]p_{E_T}(t - t_{in}|t_{in})J(t_{in}) \quad (4.58)$$

where  $t = t_{ex}$  since we are considering the particles leaving the control volume as discharge and evapotranspiration. The above relations are known in literature as Niemi's relations or formulas, after Niemi (1977).

Defining the total volume of water injected in the system in  $[0, t_p]$ :

$$V_S(t_p) := \int_0^{\min(t, t_p)} J(t_{in}) dt_{in} = \int_0^{\min(t, t_p)} (Q(t) + AE_T(t)) dt \quad (4.59)$$

it can be observed that:

$$p_J(t_{in}) := \frac{J(t_{in})}{V_S(t_p)} \quad (4.60)$$

can be considered the marginal pdf of the injection times, or the fraction of precipitation at a certain discrete  $t_{in}$  with respect to the total precipitation over a period of  $[0, t_p]$ . Analogously

$$p_Q(t) := \frac{Q(t)}{\Theta(t_{in})V_S(t_{in})} \quad (4.61)$$

is the marginal pdf of the outflow as discharge, or the fraction of discharge at a certain  $t$  generated by precipitation in the same  $[0, t_p]$ . Then, Niemi's relation (4.57) becomes:

$$p_Q(t - t_{in}|t)p_Q(t) = p_Q(t - t_{in}|t_{in})p_J(t_{in}) \quad (4.62)$$

which has the form of the well known Bayes theorem. This shows that the interpretation of the backward and forward probabilities as conditional ones is fully consistent. On the other hands, this reveals that the joint probability of  $T_r$  and  $t$  is:

$$p_S(T_r, t) = p_Q(t - t_{in}|t)p_Q(t) = p_Q(t - t_{in}|t_{in})p_J(t_{in}) \quad (4.63)$$

Because future is unknown, as remarked in section 4.2.4, there should be a working Niemi's relation for any finite time  $t$ , which

does not require the knowledge of the asymptotic value  $\Theta(t_{in})$ .

This can be easily derived after having defined:

$$g(t - t_{in}|t_{in}) := \frac{ae_t(t, t_{in})}{J(t_{in})} \equiv \frac{d\mathcal{G}}{dt} \quad (4.64)$$

and

$$f(t - t_{in}|t_{in}) := \frac{q(t, t_{in})}{J(t_{in})} \equiv \frac{d\mathcal{F}}{dt} \quad (4.65)$$

From these definitions,

$$q(t, t_{in}) = f(t - t_{in}|t_{in})J(t_{in}) \quad (4.66)$$

and

$$ae_t(t, t_{in}) = g(t - t_{in}|t_{in})J(t_{in}) \quad (4.67)$$

and, therefore,

$$Q(t)p_Q(t - t_{in}|t) = f(t - t_{in}|t_{in})J(t_{in}) \quad (4.68)$$

for discharges, and

$$AE_T(t)p_{AE_T}(t - t_{in}|t) = g(t - t_{in}|t_{in})J(t_{in}) \quad (4.69)$$

for evapotranspiration.

As a byproduct, the SAS and the forward functions are shown to be related. For discharge at any time  $t$ , for example,

$$f(t - t_{in}|t_{in}) = \frac{Q(t)\omega_q(t, t_{in})p_S(t - t_{in}|t)}{J(t_{in})} \quad (4.70)$$

#### 4.2.7 Residence times, travel times and life expectancy

The forward probabilities can be related with the life expectancy, i.e. the expected time the water molecules remain in the storage.

In the control volume, we can conceptually denote the subsets of the storage which contains the water molecules expected to exit at time  $t_{ex}$  as:

$$s_{t_{ex}}(t, t_{ex}) \quad (4.71)$$



Analogously to what was done before, we can observe that the quantity

$$p_S(t_{ex} - t|t) := \frac{s_{t_{ex}}(t, t_{ex})}{S(t)} \quad (4.72)$$

has the structure of a probability density function once integrated over all  $t_{ex}$ -s, and it is reasonable to call it the probability density of storage-life expectancy for particles in the control volume at time  $t$ .

However,  $p_S(t_{ex} - t|t)$  can also be related to the forward probabilities discussed in the previous section. In fact, it can be observed that the probability of storage-life expectancy satisfies the following relation with the age-ranked forward quantities:

$$s_{t_{ex}}(t, t_{ex}) = \int_0^{\min(t, t_p)} [q(t_{ex}, t_{in}) + ae_t(t_{ex}, t_{in})] dt_{in} + \int_0^{\min(t, t_p)} [q(t, t_{in}) + ae_t(t, t_{in})] dt_{in} \quad (4.73)$$

where, according to the definitions:

$$\begin{aligned} & \int_0^{\min(t, t_p)} [q(t_k, t_{in}) + ae_t(t_k, t_{in})] dt_{in} = \\ & \int_0^{\min(t, t_p)} [\Theta(t_{in}) p_Q(t_k - t_{in}|t_{in})] J(t_{in}) dt_{in} + \\ & + \int_0^{\min(t, t_p)} [(1 - \Theta(t_{in})) p_{AE_t}(t_k - t_{in}|t_{in})] J(t_{in}) dt_{in} \end{aligned} \quad (4.74)$$

The variable  $t_k$ , used to make the equations above and below more concise, is such that  $t_0 = t_{ex}$  ( $k = 0$ ) and  $t_1 = t$  ( $k = 1$ ). The integral spans the time interval up to  $t_p$  because we are considering the storage derived for precipitation in the finite interval  $[0, t_p]$ . In (4.73) the equality says that the life-storage at time  $t$  is equal to the water injected for any time  $t_{in} \in [0, t_p]$  which is expected to exit as discharge or evapotranspiration at time  $t_{ex}$ . The water still inside the control volume at clock time  $t$

is, however, all the water that entered the volume up to time  $t$ , minus the water that already flowed out.

This integral is not effectively known at time  $t$ , because what is happening between time  $t$  and  $t_{ex}$  is unknown, and so the pdfs (as in Figure 4.8), unless they are specified from some educated guess, as made in the last section of this paper. It follows:

$$\begin{aligned}
 p_S(t_{ex} - t|t) &= \\
 &= \frac{\sum_{k=0}^1 (-1)^k \int_0^{\min(t, t_p)} [\Theta(t_{in}) p_Q(t_k - t_{in}|t_{in})] J(t_{in}) dt_{in}}{S(t)} + \\
 &+ \frac{\sum_{k=0}^1 (-1)^k \int_0^{\min(t, t_p)} [(1 - \Theta(t_{in})) p_{AE_t}(t_k - t_{in}|t_{in})] J(t_{in}) dt_{in}}{S(t)}
 \end{aligned} \quad (4.75)$$

Thus, the relation between the storage-life expectancy and the previously introduced backward and forward probabilities is mediated by an integral equation.

#### 4.2.8 Passive and reactive solutes

The formalism developed in sections 2 to 6 applies in principle to any conservative substance, indicated by a superscript  $i$ . Therefore we have a bulk budget equation for the mass of the substance  $i$ , and age-ranked budget for the same substance:

$$\frac{dS^i(t)}{dt} = J^i(t) - Q^i(t) + R^i(S(t)) \quad (4.76)$$

and

$$\frac{dS^i(t, t_{in})}{dt} = j^i(t, t_{in}) - q^i(t, t_{in}) + r^i(s(t - t_{in})) \quad (4.77)$$

which represent trivial extensions of equations (4.16) and (4.23). To simplify this illustration, we have neglected evapotranspiration, which will be re-introduced eventually, but we have added a sink/source term including any physical or chemical reactions, extending Duffy (2010). However, if the substance is dissolved in water, it is usually treated as

concentration (either in terms of mass, moles or volume per the same quantity of water). Because we have various terms in the equations, concentrations are possibly as many as the terms that appear. In this case, three:

$$C_S^i(t) := \frac{S^i(t)}{S(t)} \quad (4.78)$$

for the concentration in storage;

$$C_J^i(t) := \frac{J^i(t)}{J(t)} \quad (4.79)$$

for concentration in input; and

$$C_Q^i(t) := \frac{Q^i(t)}{Q(t)} \quad (4.80)$$

for discharges. The latter is actually the one which is usually covered in the literature, since it is the one measured at the outlet of a control volume/catchment. For the solute discharge, an integral expression like,

$$Q^i(t) = \int_0^{\min(t, t_p)} \Theta(t_{in}) p_Q(t - t_{in} | t_{in}) J^i(t_{in}) dt_{in} \quad (4.81)$$

is assumed to be valid, where the  $i$  has been dropped from the probability distribution function, assuming that a passive solute moves with the water. Dividing (4.81) by the water discharge, it is obtained:

$$C_Q^i(t) = \int_0^{\min(t, t_p)} \frac{\Theta(t_{in}) p_Q(t - t_{in} | t_{in})}{Q(t)} J^i(t_{in}) dt_{in} \quad (4.82)$$

and, finally, applying the Niemi's formula:

$$\begin{aligned} C_Q^i(t) &= \int_0^{\min(t, t_p)} p_Q(t - t_{in} | t) \frac{J^i(t_{in})}{J(t_{in})} dt_{in} = \\ &= \int_0^{\min(t, t_p)} p_Q(t - t_{in} | t) C_J^i(t_{in}) dt_{in} \end{aligned} \quad (4.83)$$

Therefore the concentration of the passive solute in discharge is known once the concentration of the solute in input is known

together with the backward probability (Rinaldo *et al.*, 2011). The concentration estimated in this way groups substances injected at any time, in agreement with measurement practices. When a sample is taken, the action implies perfect mixing of all the age-ranked waters in the volume where measurements are made.

The bulk substance budget can instead be written as:

$$\begin{aligned} \frac{dS^i(t)}{dt} &= \frac{dC_S^i(t)S(t)}{dt} = J^i(t) - Q^i(t) + R^i(S(t)) = \\ &= J^i(t) - C_Q^i(t)Q(t) + R^i(S(t)) \end{aligned} \quad (4.84)$$

and the missing concentration  $C_S^i(t)$  can be easily estimated with the help of (4.78) since  $S(t)$  is also known.

The above is essentially the same of equation (12) in Duffy (2010), but the age-ranked formalism can be used to understand a little more about the processes in action. Starting from the quantities that appear in equation (4.77), the backward probability can be defined as:

$$p^i(t - t_{in}|t) := \frac{s^i(t, t_{in})}{S^i(t)} \quad (4.85)$$

and analogous definitions (e.g. equation 4.25) can be given for the discharge and the inputs, such as to obtain, after the appropriate substitutions:

$$\begin{aligned} &\frac{d}{dt} C_S^i(t)S(t)p_S(t - t_{in}|t) = \\ &= J^i(t)\delta(t - t_{in}) - C_Q^i(t)Q(t) \underbrace{\omega_Q(t, t_{in})p_S(t - t_{in}|t)}_{p_Q(t-t_i|t)} + r^i(t, t_{in}) \end{aligned} \quad (4.86)$$

which is the master equation (equation 4.31) for the substance  $i$ . Many of the superscripts  $i$  were dropped, because the

$i$ -substance does not modify the velocity (i.e., it behaves like water).

The braces were added to emphasize that  $p_Q(t - t_{in}|t)$  should have been left, and we could solve the system of equations directly for  $p_S(t - t_{in}|t)$  and  $p_Q(t - t_{in}|t)$ , obtaining eventually the age-ranked quantities, using (4.76).

In fact, in (4.86) all the quantities are known, either because solution of the solute budget (4.76) or the water master equation (equation 4.31), or a known input ( $J(t)$ ). The only quantity that is unknown (and usually guessed) is  $\omega_Q(t, t_{in})$ . However, (4.86) and (4.31) can be seen as two coupled equations in  $p_S(t - t_{in}|t)$  and  $\omega_Q(t, t_{in})$ , and we can conclude that the SAS can be derived rather than imposed.

From a practical point of view there could be some obstacles in the correct determination of the SAS, because the distribution of the input of the substance can be unknown. In this case (4.86) can be used to back-trace the the passive solute injection, after having made educated guesses on the SAS. In the presence of more than one solute, the flow of every solute obeys the same probabilities  $p_S$  and  $p_Q$ . This redundancy can then be used for improving their estimation by applying the appropriate statistical techniques.

For the sake of simplicity we neglected evapotranspiration.

However, now that the concepts are established, we can observe that incorporating  $AE_T$  involves a second SAS, which remains undetermined. Various approaches can be chosen to overcome this fact. For instance, it can be assumed that

$\omega_Q(t, t_{in}) = \omega_{E_T}(t, t_{in})$ . Nevertheless the main experimental approach would be to find a second passive tracer transported through vegetation. In this case, if a third equation similar to (4.86), but containing evapotranspiration, would hold, it would permit the determination of the missing SAS coefficient.

An interesting approach is shown in Queloz *et al.* (2015) and in Benettin *et al.* (2017), where the "direct SAS" method was introduced. This approach considers a single catchment-scale SAS, coupled to the measured input and output fluxes. In the residence time domain, the SAS function can be parametrized through a power law defining the preference for old/young water, of the kind:

$$\omega(p_S) = kp_S^{k-1} \quad (4.87)$$

where the only parameters involved is  $k$ , which indicates the "preference" for releasing water from younger/older storage components. The SAS function as defined by equation 4.87 does not vary in time, but it can easily be made time-variant, as in Benettin *et al.* (2017). The use "direct-SAS" approach proved feasible, computationally efficient and it was able to well reproduce the observed deuterium signature.

Duffy (2010), as in Carrera & Medina (1999), added an equation for water age similar to ours (4.76) and (4.86). This is necessary when dealing with spatially distributed properties (see Appendix B) but not at our spatially integrated scales. In fact, in our case, water age can be estimated directly from its definition (4.27), since the probability distribution of residence time is known.

Finally, in order to clarify this theory, an example of  $r^i$  could be:

$$r^i(t, t_{in}) := k_1(s^i(t, t_{in}) - k_2s_{eq}^i) \quad (4.88)$$

where  $k_1$  and  $k_2$  are suitable reaction constants and  $s_{eq}^i$  represents an equilibrium storage. Whilst more complex reactions can be envisioned, this type of reaction (or sink term), being linear, does not alter the essential traits of the theory described above.

### 4.2.9 A simple example where probabilities are assigned instead than derived.

With the scope to further clarify the formalism, we assume in this section that the forward pdfs introduced in the previous sections are known. We use the concept of linear reservoir, which has a long history in surface hydrology, e.g. Dooge (1973). First consider only one outflow, the bulk equation for the water budget of a single linear reservoir is:

$$\frac{dS(t)}{dt} = \sum_{t_{in}=1}^n R_{t_{in}} - \frac{1}{\lambda} S(t) \quad (4.89)$$

where it has been assumed, for simplicity, that  $J(t) = \sum_{t_{in}=1}^n R_{t_{in}}$ , i.e. that the precipitation is accounted as a sequence of instantaneous impulses at different times  $t_{in}$ s. By definition of the linear reservoir:

$$Q(t) = \frac{1}{\lambda} S(t) \quad (4.90)$$

where  $\lambda$  [T] is the mean response time (not to be confused with the mean “travel time” derived from the backward distributions) in the reservoir. If this is the case, assuming that the age-ranked storages behave linearly, the age-ranked water budgets can be written as:

$$\frac{ds(t, t_{in})}{dt} = R_{t_{in}} \delta(t - t_{in}) - \frac{1}{\lambda} s(t, t_{in}) \quad (4.91)$$

where it is

$$q(t, t_{in}) = \frac{1}{\lambda} s(t, t_{in}) \quad (4.92)$$

Equation (4.91), after integration over  $t_{in}$  reduces to equation (4.89). By definition, it is  $s(t, t_{in}) = 0$  for  $t < t_{in}$  and the solution, for  $t > t_{in}$  is well known as:

$$s(t, t_{in}) = R_{t_{in}} e^{\frac{t_{in}-t}{\lambda}} \quad (4.93)$$

The equivalent solution, for  $S(t)$  gives:

$$S(t) = \int_{t_{in}}^t R_{t_{in}} e^{-(t-t_{in})/\lambda} dt_{in} \quad (4.94)$$

and the backward probability can be written, then as:

$$p_S(t - t_{in}|t) = \frac{R_{t_{in}} e^{-\frac{t-t_{in}}{\lambda}}}{\int_{t_{in}}^t R_{t_{in}} e^{-(t-t_{in})/\lambda} dt_{in}} \quad (4.95)$$

If, and only if,  $R_{t_{in}} = const$  the probability simplifies, and it is time invariant, i.e. dependent only on the residence time  $T_r = t - t_{in}$ . Please notice that, in this case, we did not appeal to equation (4.31) to estimate the backward probability. Instead we used the definitions in equation (4.95).

Because discharges are just linearly proportional to the storage, it is easy to show that  $p_q(t - t_{in}|t) = p_S(t - t_{in}|t)$  and, therefore, in this case,  $\omega(t, t_{in}) = 1$ . This shows that the linear reservoir case, where for all injection times the mean residence time is equal (to  $\lambda$ ), the SAS function is necessarily unitary. However, a more general case, can be set if the mean residence time is a function of  $t_{in}$ , meaning that equation (4.91) can be modified into:

$$\frac{ds(t, t_{in})}{dt} = R_{t_{in}} \delta(t - t_{in}) - \frac{1}{\lambda_{t_{in}}} s(t, t_{in}) \quad (4.96)$$

and its solution for  $t > t_{in}$  is the same as (4.93), but with  $\lambda$  muted into  $\lambda_{t_{in}}$ . However, due to the dependence of  $\lambda_{t_{in}}$  on the injection time, the SAS is not anymore a constant, being equal to:

$$\begin{aligned} \omega_Q(t, t_{in}) &:= \frac{p_q(t - t_{in}|t)}{p_S(t - t_{in}|t)} = \lambda_{t_{in}}^{-1} \frac{\int_{t_{in}}^t R_{t_{in}} e^{-(t-t_{in})/\lambda_{t_{in}}} dt_{in}}{\int_{t_{in}}^t \lambda_{t_{in}}^{-1} R_{t_{in}} e^{-(t-t_{in})/\lambda_{t_{in}}} dt_{in}} = \\ &= \lambda_{t_{in}}^{-1} \frac{\int_{t_{in}}^t R_{t_{in}} e^{+t_{in}/\lambda_{t_{in}}} dt_{in}}{\int_{t_{in}}^t \lambda_{t_{in}}^{-1} R_{t_{in}} e^{t_{in}/\lambda_{t_{in}}} dt_{in}} \end{aligned} \quad (4.97)$$

This seems to suggest that imposing the characteristics of the pdf could completely determine the  $\omega_Q(t, t_{in})$ . Vice versa, as already known, assigning  $\omega_Q(t, t_{in})$  from some heuristic,



obviously, would determine a mean residence time dependence on the injection time.

Non trivial  $\omega(t, t_{in})$  can also be derived from assuming a sequence of linear reservoirs, as in the so called Nash model, Dooge (1973). Without entering in details, a sequence of linear reservoirs implies that just the last reservoir maintains a linear relation between storage and outflow. Instead a nonlinear relationship exists between the whole storage and the same outflow, implying also a nonlinear SAS.

Even if semi-analytical results are not feasible using non-linear reservoirs, suitably tuning the parameters of each age-ranked equation cannot change the form of the SAS, as is also suggested by arguments below.

Other aspects come into play when there are multiple outputs. Expanding the previous linear case to include evapotranspiration, the bulk equation, becomes:

$$\frac{dS(t)}{dt} = \sum_{t_{in}=1}^n R_{t_{in}} - \left( \frac{1}{\lambda} - aet(t) \right) S(t) \quad (4.98)$$

where the actual evapotranspiration is assumed to equal:

$$AE_T(t) = S(t)aet(t) \quad (4.99)$$

with a linear dependence on the soil water content, as for instance in Rodriguez-Iturbe *et al.* (1999). The equations of water budget for the generations becomes:

$$\frac{ds(t, t_{in})}{dt} = R_{t_{in}} \delta(t - t_{in}) - \left( \frac{1}{\lambda_{t_{in}}} + ae(t, t_{in}) \right) s(t, t_{in}) \quad (4.100)$$

where the bivariate dependence of  $ae(t, t_{in})$  on the actual time and the injection time can be justified by arguing that, water of different ages is not perfectly mixed in the control volume and plant roots sample water of different ages in different modes, according to their spatial distributions. Since equation (4.100)

remains a linear ordinary differential equation, it can be solved analytically, and:

$$s(t, t_{in}) = R_{t_{in}} e^{-\Lambda(t, t_{in})} \quad (4.101)$$

where:

$$\Lambda(t, t_{in}) := \int_{t_{in}}^t \left( \frac{1}{\lambda_{t_{in}}} + ae(t', t_{in}) \right) dt' \quad (4.102)$$

and:

$$S(t) = \int_0^t R_{t_{in}} e^{-\Lambda(t, t_{in})} dt_{in} \quad (4.103)$$

Notably, the outflows terms can be expressed as a function of the storage:

$$q(t, t_{in}) + aet(t, t_{in}) = \mu(t, t_{in}) s(t, t_{in}) \quad (4.104)$$

the problem remains linear and analytically solvable. The quantity  $\mu(t, t_{in})$  is usually called age and mass-specific output rate, Calabrese & Porporato (2015). Solving equation (4.100) it is not even necessary to show that:

$$\omega_{E_T}(t, t_{in}) \neq 1 \quad (4.105)$$

The latter condition is regained if and only if  $aet(t, t_{in}) = aet(t)$ , i.e. it depends only on the current time (which is a condition that requires the perfect mixing of aged waters). In fact, in case a dependence on  $t_{in}$  remains, then, trivial algebra says that:

$$p_{E_T}(t - t_{in} | t) = \frac{ae(t, t_{in}) s(t, t_{in})}{\int_{t_{in}}^t ae(t, t_{in}) s(t, t_{in}) dt_{in}} \quad (4.106)$$

which implies:

$$\omega_{E_T}(t, t_{in}) := \frac{p_{E_T}(t - t_{in} | t)}{p_S(t - t_{in} | t)} = \frac{ae(t, t_{in}) \int_{t_{in}}^t R_{t_{in}} e^{-\Lambda(t, t_{in})}}{\int_{t_{in}}^t ae(t, t_{in}) S(t, t_{in}) dt_{in}} \quad (4.107)$$

Obviously these results, obtained by imposing a travel time probability, can be inconsistent with tracers results, because both approaches require estimates of the  $\omega$  functions, which are not known well.

#### 4.2.10 Extension to the embedded reservoirs model

The age-ranked theory can be easily extended to the embedded reservoir model presented in the Chapter 3. To summarize the main connections of the embedded model: the snow melting and/or the rainfall are the inputs of the canopy layer. Its outputs are the throughfall and the evaporation from the wet canopy. If there is no canopy, the melting/rain reaches directly the root zone.

The outputs from the root zone storage are the evapotranspiration and the recharge term of the groundwater. The throughfall exceeding the root zone capacity is sent directly to the volume available for surface runoff. Baseflow from the groundwater is modeled using a non-linear reservoir. Total runoff is the sum of the direct runoff and of the baseflow. The water budget solved for the canopy reservoir is given by:

$$\frac{dS_c(t)}{dt} = M_d(t) - T_r(t) - ET_c(t) = (1 - p) M_d(t) - D(t) - ET_c(t) \quad (4.108)$$

Applying the theory developed in the main text, the age-ranked equations for this storage is given by:

$$\begin{aligned} \frac{dS_c(t) p_c(t - t_{in}|t)}{dt} = & (1 - p) M_d(t_{in}) \delta(t - t_{in}) + \\ & - \omega_D(t, t_{in}) D(t) p_c(t - t_{in}|t) - \omega_{ET_c}(t, t_{in}) ET_c(t) p_c(t - t_{in}|t) \end{aligned} \quad (4.109)$$

Once the two SASs in Eq. (4.109), i.e.  $\omega_D(t, t_{in})$  and  $\omega_{ET_c}(t, t_{in})$ , are assigned, also the probability  $p_c(t - t_{in}|t)$ , and the age-ranked storage  $s_c(t, t_{in})$  can be determined. Given the characteristics of the storage, which doesn't present particular flow paths, it is reasonable to consider

$$\omega_D(t, t_{in}) = \omega_{ET_c}(t, t_{in}) = 1.$$

The root zone reservoir obeys the following budget equation:

$$\frac{dS_{rz}(t)}{dt} = (1 - \alpha(t))T_r(t) - Re(t) - ET_{rz}(t) \quad (4.110)$$

The age-ranked equations for this storage is given by:

$$\begin{aligned} \frac{dS_{rz}(t)p_{rz}(t - t_{in}|t)}{dt} = & (1 - \alpha(t))Tr(t_{in})\delta(t - t_{in}) + \\ & -\omega_{Re}(t, t_{in})Re(t)p_{rz}(t - t_{in}|t) - \omega_{ET_{rz}}(t, t_{in})ET_{rz}(t)p_{rz}(t - t_{in}|t) \end{aligned} \quad (4.111)$$

which is solvable, once the two SASs in Eq. (4.111), i.e.  $\omega_{Re}(t, t_{in})$  and  $\omega_{ET_{rz}}(t, t_{in})$ , are assigned.

The equation for the runoff reservoir is:

$$\frac{dS_R(t)}{dt} = \alpha(t)T_R(t) - Q_R(t) \quad (4.112)$$

$\alpha$  is estimated with a complex scheme (derived from Hymod, (Zhao, 1980; Moore, 1985)) but is assumed to produce homogeneous mixing. Therefore, the age-ranked equation are:

$$\frac{dS_R(t)p_R(t - t_i|t)}{dt} = \alpha(t)T_r(t, t_i) - Q_R(t)\omega_R(t, t_i)p(t - t_i|t) \quad (4.113)$$

It must be observed, however, that, if  $Q_R$  is given by the width function theory, then

$$Q_R(t, t_i) = A \int_0^t u, W(t - \tau)T_R(\tau, t_i)d\tau \quad (4.114)$$

which is a function of both  $t$  and  $t_i$  even without assuming mixing.

Finally, the groundwater reservoir obeys the following budget equation:

$$\frac{dS_{GW}(t)}{dt} = R_e(t) - Q_{GW}(t) \quad (4.115)$$

Eq. (4.115) can then be associated with the age-ranked master equation:

$$\frac{dS_{GW}(t)p_{GW}(t-t_{in}|t)}{dt} = Re(t, t_{in}) - Q_{GW}(t)\omega_{GW}(t, t_{in})p_{GW}(t-t_{in}|t) \quad (4.116)$$

where  $Re(t, t_{in})$  is the input to the groundwater reservoir which comes with aged waters, and is given by solving Eq. (4.111)

because it is  $Re(t, t_{in}) = Re(t)p_{rz}(t-t_{in}|t)$ .

In turn Eq. (4.116) is solvable and can be used to obtain all the age-ranked functions relative to the groundwater storage.

The overall system is the sum of the three reservoirs where:

$$S(t) = S_c(t) + S_{rz}(t) + S_{GW}(t) \quad (4.117)$$

and

$$s(t, t_{in}) = s_c(t, t_{in}) + s_{rz}(t, t_{in}) + s_{GW}(t, t_{in}) \quad (4.118)$$

Therefore

$$p_S(t-t_{in}|t) := \frac{s(t, t_{in})}{S(t)} \quad (4.119)$$

is the backward residence time distribution for the compound system.

The total discharge is given by the contribute from the surface runoff and from the groundwater:

$$Q(t) = Q_R(t) + Q_{GW}(t), \quad (4.120)$$

and

$$q(t, t_{in}) = q_R(t, t_{in}) + q_{GW}(t, t_{in}), \quad (4.121)$$

the global travel time distribution is:

$$p_Q(t-t_{in}|t) := \frac{q(t, t_{in})}{Q(t)} \quad (4.122)$$

Besides, the total evapotranspiration is given by the contribute from the canopy and from the root zone:

$$ET(t) = ET_c(t) + ET_{rz}(t), \quad (4.123)$$

and

$$et(t, t_{in}) = et_c(t, t_{in}) + et_{rz}(t, t_{in}), \quad (4.124)$$

the global travel time distribution is:

$$p_{ET}(t - t_{in}|t) := \frac{et(t, t_{in})}{ET(t)} \quad (4.125)$$

It follows that the compound systems behaves like having SAS given by:

$$\omega_Q(t, t_{in}) = \frac{p_Q(t - t_{in}|t)}{p_S(t - t_{in}|t)} \quad (4.126)$$

$$\omega_{ET}(t, t_{in}) = \frac{p_{ET}(t - t_{in}|t)}{p_S(t - t_{in}|t)} \quad (4.127)$$

On the basis of the global probability distribution functions, the behavior of a tracer  $i$  can be obtained from Niemi's relations as:

$$C_Q^i(t) = \int_0^{\min(t, t_p)} p_Q(t - t_{in}|t) C_J^i(t_{in}) dt_{in} \quad (4.128)$$

This concentration does not distinguish between waters coming from the surface and the groundwater reservoirs. However, the theory can do it by substituting Eq. (4.128) in place of  $p_Q(t - t_{in}|t)$ ,  $p_{QR}(t - t_{in}|t)$  or  $p_{QGW}(t - t_{in}|t)$ .

Because it must be:

$$p_Q(t - t_{in}|t) = (1 - \Theta_Q(t)) p_{QR}(t - t_{in}|t) + \Theta_Q(t) p_{QGW}(t - t_{in}|t) \quad (4.129)$$

where:

$$\Theta_Q(t) = \frac{Q_{sat}(t)}{Q_R(t) + Q_{GW}(t)} \quad (4.130)$$

is the appropriate partitioning coefficient. To obtain the last equations, it is sufficient to apply the definitions for the probabilities. The case treated show that any set of coupled reservoirs can be analyzed from the travel time point of view, no matter how complex the system is.

#### 4.2.11 Final Remarks

In this chapter, we reviewed existing concepts that were collected from many different papers, and presented them in a new systematic way. We established a consistent framework that offers a unified view of the travel time theories across surface water and groundwater.

It contains several clarifications and extensions.

Clarifications include:

- the concepts of forward and backward conditional probabilities and a small but important change in notation;
- their one-to-one relation with the water budget (and the age-ranked functions) from which the probabilities were derived (after the choice of SASs);
- the proper way to choose backward probabilities. Specifically, it was shown that the usual way to assign time invariant backward probabilities is inappropriate. We also show how to do it correctly, and introduced a minimal time variability.
- the fact that time-invariant forward probabilities usually imply time-varying backward probabilities, i.e. travel time distributions.
- the rewriting of the master equation by Botter, Bertuzzo and Rinaldo as an ordinary differential equation (instead of a partial differential equation);
- the role and nature of the partitioning coefficient between discharge and evapotranspiration (which is unknown at any time except asymptotically);

- the significance of the SAS functions with examples;
- the relationship of the present theory with the well known theory of the instantaneous unit hydrograph; and
- we added information and clarified some links of the present theory with [Delhez *et al.* (1999) and [Duffy (2010)].

Extensions include:

- new relations among the probabilities (including the relation between expectancy of life and forward residence time probabilities);
- an analysis of the partitioning coefficients (which are shown to vary seasonally);
- an explicit formulation of the equations for solutes which would permit direct determination of the SAS on the basis of experimental data;
- tests of the effects of various hypotheses, e.g., assuming a linear model of forward probability and gamma model for the backward probabilities;
- an extension of Niemi's relation (and a new normalization);
- the presentation of Niemi's relation as a special case of the Bayes Theorem;
- a system of equations from which to obtain the SAS experimentally.

The extension of the theory to any passive substance diluted in water clearly opens the way to new developments of the theory and applications of tracers.



## POSINA APPLICATION: EMBEDDED RESERVOIR MODEL AND TRAVEL TIMES RESULTS

As a proof of concepts presented in the Chapters 3 and 4, this chapter presents an example derived from a real case (Posina River Basin), which comes with open source code available to any researcher.

### 5.1 Study area and model setup

Posina River basin is a small catchment (around  $116 \text{ km}^2$ ), located in the the Alpine part of Veneto Region, Italy. Altitude spans from 418 a.m.s.l. to 2211 a.m.s.l., see figure 5.1. The river is a tributary of the Astico river, which flows into the Bacchiglione-Brenta river system and then into the Adriatic sea. Around the 74% of the Posina is densely vegetated, Norbiato *et al.* (2009); Penna *et al.* (2015). Beech, chestnut, maple and hazel are the main species in the catchment. The climate is characterized as wet, with an mean annual precipitation of 1740 millimeters and annual runoff of 1000 millimeters, Abera *et al.*

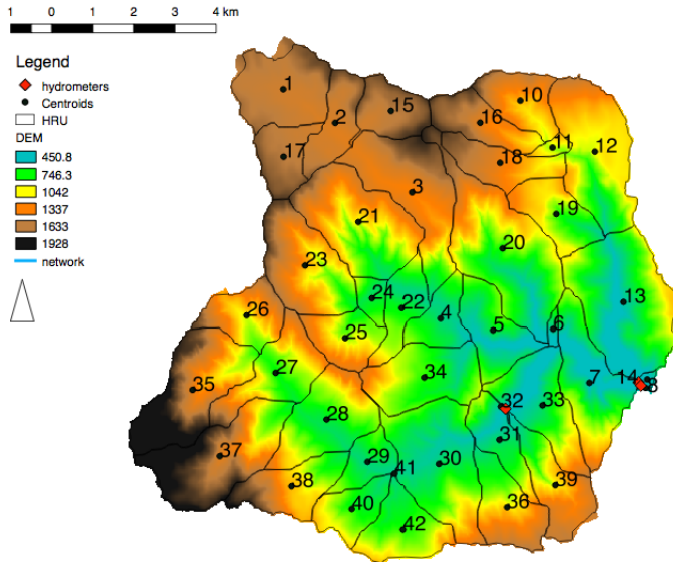


Figure 5.1: Digital elevation model of the Posina River basin and HRU partition. Red dots represent the centroids and numbers represent the ID of the HRUs. Red diamonds represent the discharge measuring points.

(2017a). The average annual temperature is  $9.7^{\circ}\text{C}$ ; average monthly temperatures range from  $1.2^{\circ}\text{C}$  in January and  $18.7^{\circ}\text{C}$  in July. Data are retrieved from 12 meteorological stations and three discharge gauges, whose coordinates and elevations are reported in table 5.1. An ultrasound hydrometer and a hydrometer are installed. Immediately before Stancari station there is the restitution of a power plant, figure 5.3. Bazzoni upstream station and the hydrometer are shown in figure 5.4, while the system of buckets installed for measurements of the volume of throughfall is shown in figure 5.5.

Both meteorological stations and hydrometers provided hourly data for the period 1994 to 1998. The basin was partitioned in 42 HRUs and rainfall and temperature data were interpolated for each HRU centroids using Kriging interpolation algorithm,



Figure 5.2: Photo taken on the Posina River at the closure section of Stancari, Vicenza, Italy.



Figure 5.3: Power plants restitution immediately before the Stancari section.



Figure 5.4: Photo taken on the Posina River at the section of Bazzoni, Vicenza, Italy.



Figure 5.5: System of buckets for the total throughfall measurements. The buckets were installed and measures collected by the Department of Land, Environment, Agriculture and Forestry of University of Padua.

City	Elevation	Longitude	Latitude
Folgaria UPO	1168	1668428	5086815
Lavarone UPO	1171	1674754	5089860
Brustol Velo d'Astico	328	1682121	5074661
Contra' Doppio Posina	725	1672938	5075022
Molini Laghi	597	1675208	5078024
Monte Summano	619	1687964	5069297
Passo Xomo Posina	1056	1674012	5071777
Pedescala	308	1683840	5079537
Valli del Pasubio	600	1672265	5069542
Castana Arsiero	430	1679369	5076164
*Rio Freddo at Valoje (22.24km <sup>2</sup> )	390	1681507	5075248
*Posina at Stancari (116.2km <sup>2</sup> )	388	1681524	5075140
*Posina at Bazzoni (38.82km <sup>2</sup> )	453	1678208	5074606

Table 5.1: List of meteorological and discharge stations in the Posina basin. Discharge measurement stations are identified with a \*.

according to Abera *et al.* (2017a).

LAI data used in the study are satellite-based with 1 km of spatial resolution, retrieved from Copernicus Global Land Service, <http://land.copernicus.eu/global/products/lai>. LAI data are available at monthly time-step from 1998. Therefore a simple linear interpolation was used to reconstruct LAI daily and hourly time-series for 1998, as shown in figure 5.6, and then replicated for the entire period of analysis.

The parameter  $p$ , which is the proportion of rain falling through canopy without interception, was set equal to 0.65, according to field measurements, kindly shared by Marco Borga and Giulia Zuecco of the Dept of Land, Environment, Agriculture and Forestry of University of Padua, (Zuecco *et al.*, 2014).

Figure 5.7 shows the modeling solution used in the present study for the testing of the embedded reservoir model.

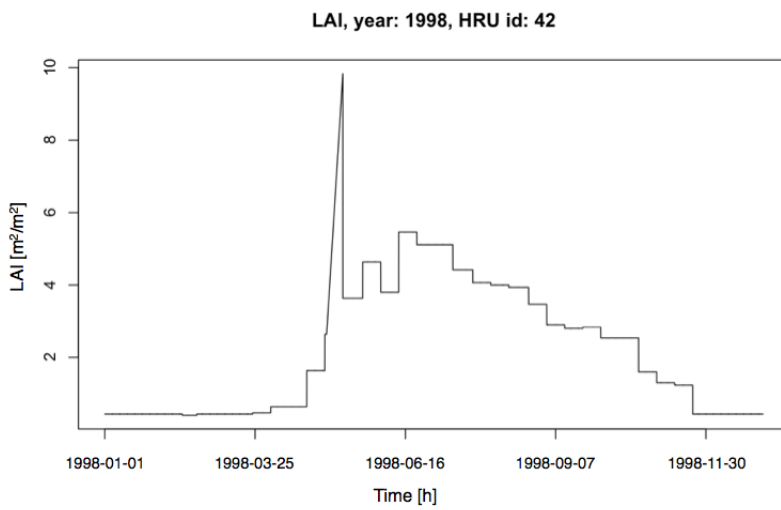


Figure 5.6: Time variation of the LAI in the 1998 for the HRU ID 42. This HRU was chosen since the LAI time series had the less number of no values.

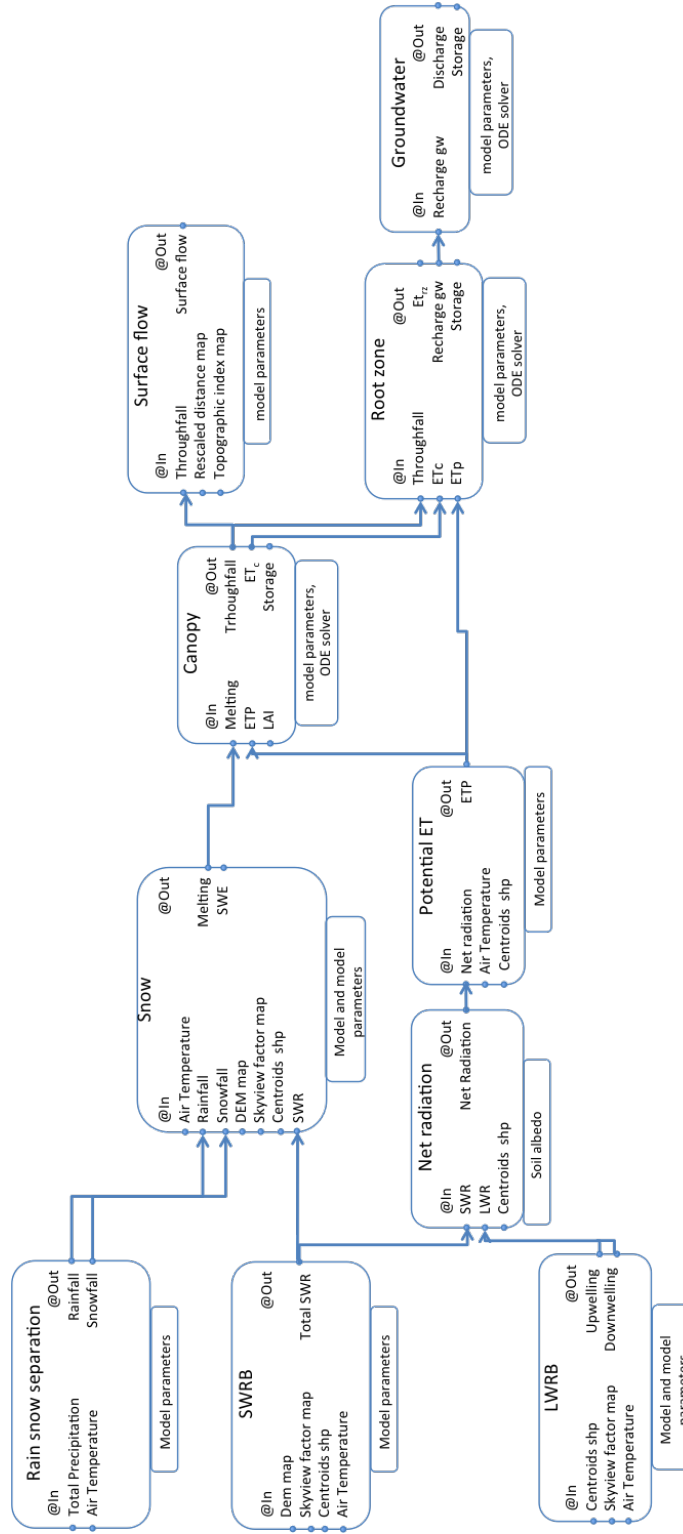


Figure 5.7: Schema of the connection of the JGrass-NewAge components, necessary to perform the modeling solution: the blu arrows represents the connection out-to-in made possible thanks to OMS3.

- ShortWave Radiation Balance component (SWRB), (Formetta *et al.* , 2013b), was used to estimate the total shortwave radiation. Inputs of SWRB are the digital elevation model, the skyview factor map, the shapefile of the centroids of the HRUs, the air temperature and the model parameters, according to (Corripio, 2003). Output of the component is the total shortwave, which is both input of the net radiation and of the snow components.
- LongWave Radiation Balance component (LWRB), (Formetta *et al.* , 2016), was used to estimate the total longwave radiation. LWRB takes in input the skyview factor map, the shapefile of the centroids of the HRUs, the air temperature and the model parameters and gives in output both downwelling and upwelling longwave radiations.
- Total longwave and shortwave are the inputs of Net radiation component, which, according to the soil albedo, computes the net radiation. The value of the soil albedo was set according to Henderson-Sellers & Wilson (1983).
- Rain-snow separation detects, from total precipitation, rainfall and snowfall, according to the method presented in Formetta *et al.* (2013a).
- Snow component, (Formetta *et al.* , 2013a), allows to simulate the snow water equivalent (SWE) and the snow melting. Besides the rainfall and snowfall, snow components takes in inputs the digital elevation model, the skyview factor map, the shapefile of the centroids of the HRUs, the air temperature, the shortwave and the parameters of the chosen model.
- Priestley-Taylor model, (Priestley & Taylor, 1972), was used to simulate potential evapotranspiration using the ETp



component, which takes in input, besides the net radiation, the air temperature and the shapefile of the centroids of the HRUs.

- Canopy storage component takes in input the LAI, potential evapotranspiration and melting/rain, while the outputs are the throughfall, evaporation from the wet canopy and the storage. The model parameters and the name of the solver of the ODE, must also be specified.
- Surface flow component takes in input the throughfall, and, in the case of the WFIUH approach, the rescaled distance and the topographic index and gives in output the surface flow. The model parameters must also be specified.
- Root zone storage component was used to simulate the drainage toward the groundwater, the evaporation from the bare soil and transpiration from the canopy. The component takes in input the potential evapotranspiration, the evapotranspiration from the wet canopy and the throughfall. As well as in other similar cases, the model parameters and the name of the solver of the ODE, must be specified.
- Finally, the groundwater component was used to simulate the baseflow through a non-linear reservoir and takes input the recharge term from groundwater. The model parameters and the name of the solver of the ODE, must also be specified.

The scheme is repeated for each HRU, which differs for the values of the meteorological forcing and geographical effects. The Net3 system, eventually, takes care of each HRU and joins them to produce the total values (of discharge, ET, etc) for the whole basin.

The maps of skyview factor, of rescaled distances and of topographic index, shown in figure 5.8, were obtained for each HRU using the Horton-Machine libraries, Rigon *et al.* (2006).

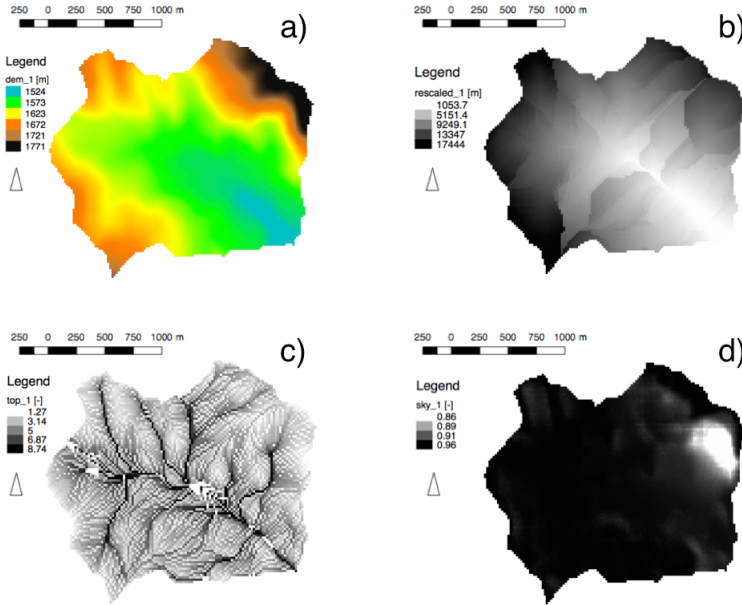


Figure 5.8: Example of 4 maps produced using the Horton-Machine libraries for a single HRU (ID 1): panel *a* shows the digital elevation model, panel *b* shows the distances rescaled by a factor of 10, panel *c* shows the topographic index and panel *d* shows the skyview factor .

As soon as the water budget quantities are obtained, for each storage, the age-ranked equations, from 4.109 to 4.130 can be solved. Figure 5.9 shows the modeling solution used for the testing of the travel times theory, in the case of the canopy storage.

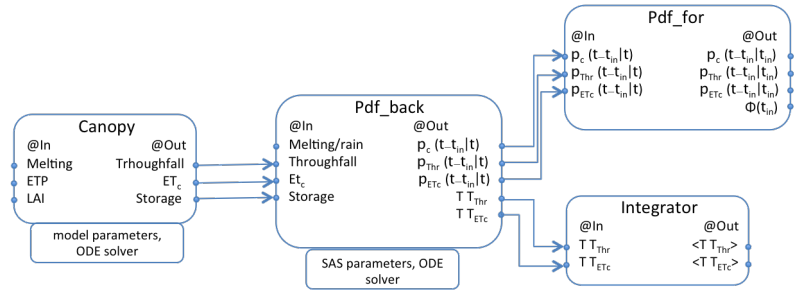


Figure 5.9: Schema of the connection of the JGrass-NewAge components, necessary to perform the travel times analysis: the blue arrows represent the connection out-to-in made possible thanks to OMS3.

- the component backward pdfs, solves the age-ranked equations and computes the backward residence times and travel times pdfs. Input of the components are the terms of the water balance of the investigated reservoir. In this case, the input is the melting/rain, while the outputs are the evaporation from the wet canopy and the throughfall. The parameter of the SAS should be defined, together with the numerical solver of the ODE. Outputs of the component are bi-dimensional matrices in which each row represents the actual time and each column represents the injection time;
- the mean travel times are computed using the integrator component, which integrates the age-ranked quantities in the injection-times domain;
- the component forward pdfs, computes the forward residence and travel times pdfs and partitioning coefficient,  $\Theta(t_{in})$ . Inputs are the backward pdfs from the previous component.

The MS can be repeated in the same way for each analyzed reservoir or for the compound system, as shown in chapter 4,

just changing the inputs of the backward pdfs component.

As regards the integrators for the solution of the ODEs for each storage two choices are implemented, the Dormand-Prince integrator, Dormand & Prince (1980) and the Euler integrator, Butcher (1987). Dormand-Prince integrator is an embedded Runge-Kutta integrator of order 5 used in local extrapolation mode (i.e. the solution is computed using the high order formula) with step-size control (and automatic step initialization) and continuous output. The Euler algorithm, on the other hand, is the simplest solver that can be used to integrate ordinary differential equations and it is based on forward differences.

In this case, the Dormand-Prince integrator was chosen for all the components.

### **5.1.1 Parameter calibration and verification**

Although each HRU has different inputs (precipitation, temperature, radiation,  $ET_p$ , LAI), embedded model parameters are calibrated at catchment scale versus rainfall-runoff data. The parameters, shown in table 5.1.1, resulted the optimal number to describe the hydrological response, with the minimal computational costs. The table shows the involved reservoir, the name of the parameter and their range of calibration.

Separation of the rainfall and snowfall from the total precipitation was made using the set of parameters calibrated in Abera *et al.* (2017a) and the rain-snow separation method presented in Formetta *et al.* (2013a). Since there were no SWE measurements for the reference dataset also the snow parameters were calibrated using the measured discharge at Stancari.

The model was calibrated against the measured discharge at Stancari station and then validated using the entire dataset, both

Reservoir	Parameter	Range
Snow	$\alpha_m$	[0.01 – 1.0]
Snow	$\alpha_f$	[0.0001 – 0.1]
Snow	$\alpha_e$	[1.0E-5 – 1.0E-4]
Snow	$\alpha_l$	[0.3 – 0.9]
Canopy	$k_c$	[0.1 – 0.3]
Root zone	$S_{rz_{max}}$	[100-400]
Root zone	$a$	[0.0001-0.01]
Root zone	$b$	[1.0 - 2.0]
Root zone	$B$	[0.1-0.8]
Direct runoff	$pSat$	[20.0-80.0]
Direct runoff	$c$	[0.0-0.1]
Direct runoff	$d$	[1.0-3.0]
Groundwater	$S_{gw_{max}}$	[500-1000]
Groundwater	$e$	[100-600]
Groundwater	$f$	[1-10]

Table 5.2: List of parameters of the embedded reservoir model: the components associated to each parameter are specified in the first column, while the calibration ranges are specified in the third column.

at Stancari and at Bazzoni stations. Moreover, a further validation was performed by applying the Hymod model, (Moore, 1985), which was calibrated and validated against the same dataset, both at Stancari and at Bazzoni stations. The objective functions used to evaluate model performances are the Kling-Gupta Efficiency (KGE), (Gupta *et al.* , 2009) and the Nash-Satcliffe Efficiency (NSE), (Nash & Sutcliffe, 1970), better described in Appendix A.3.

## 5.2 Results and discussion

The following sections present the results obtained from the application of the embedded reservoir model and a preliminary travel times analysis using the Posina river dataset, previously described. Two applications were performed using different time-steps, hourly and daily.

### **5.2.1 Hourly time step**

This section presents the results obtained using the original Posina dataset at hourly time-step. The calibration was made versus one year of measured discharge at Stancari station, while the validation was made on the entire dataset, both at Stancari and at Bazzoni. Surface runoff was modeled using the non-linear reservoir model. Figure 5.10 shows the comparison between the measured and the simulated discharges obtained using both the embedded reservoir and Hymod models at the two closure stations (at Stancari and at Bazzoni). The indices of goodness (KGE and NS) for both models are reported in table 5.2.1. Both models show good performances in reproducing the discharge at Stancari station. These performances are, however, biased from the hourly fluctuations due to the hydropower plants and the bottling plants located upstream the closure sections.

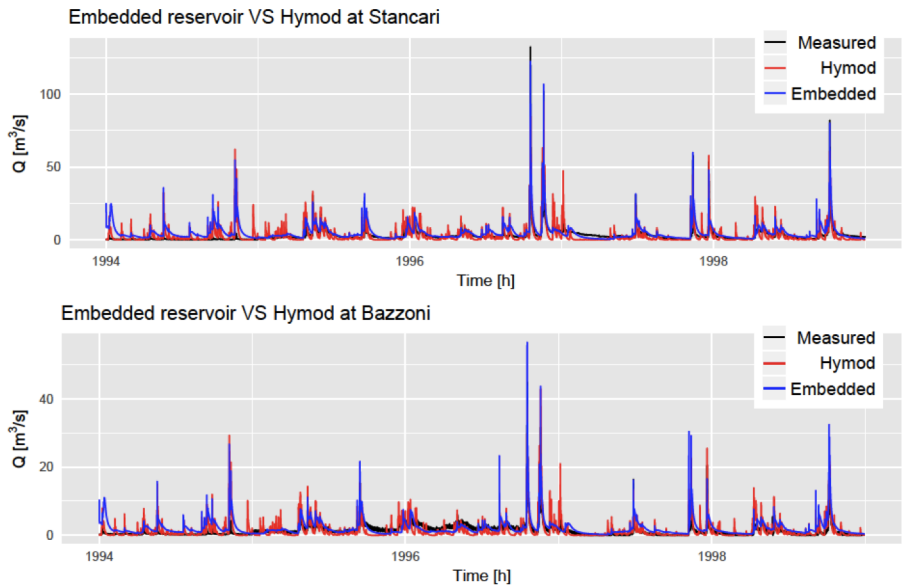


Figure 5.10: Measured discharge (black line) VS modeled discharge obtained using the embedded reservoir model (blu line) and the Hymod model (red line). Both models were calibrated against one year of discharge measured at Stancari (top plot) and validated using the entire time series of discharge at Stancari and at Bazzoni.

Gauge station	Embedded		Hymod	
	KGE	NS	KGE	NS
*Posina at Bazzoni (38.82km <sup>2</sup> )	0.55	0.46	0.57	0.37
*Posina at Stancari (116.2km <sup>2</sup> )	0.70	0.65	0.63	0.57

Table 5.3: Results in terms of goodness of fit in calibration and validation periods for the two main stations, Bazzoni and Stancari.

The analysis of the relative amount of the input and output contributes to the overall water balance, was made using the waterfall charts, (Mitchell *et al.* , 2003), and are shown in figures 5.11, at annual scale and 5.12, at monthly scale.

In particular, figure 5.11 shows the waterfall charts for the canopy, root zone and groundwater reservoirs for years from 1994 to 1998. Green bars represent the mean annual volume of

the input of the reservoirs, blue bars represent the mean annual volume of the output fluxes, and the red bars represent the annual variation of the storage. At annual scale, it is clear that increases in precipitation tend to contribute directly to  $Q$  with minor effects on  $S$  and  $ET$ , confirming results in (Abera *et al.* , 2017a). The melting/rain represent the main source of variability in the budget, and (therefore) the  $ET$  shows a smoother behavior since the radiation, which is the main driver, is consistent across the years.

Analyzing the canopy reservoir, it is clear that the greatest part of the input (melting/rain) contributes to the throughfall ( around the 95% of the total input), while the the evaporation from the wet canopy represents a minor contribute.

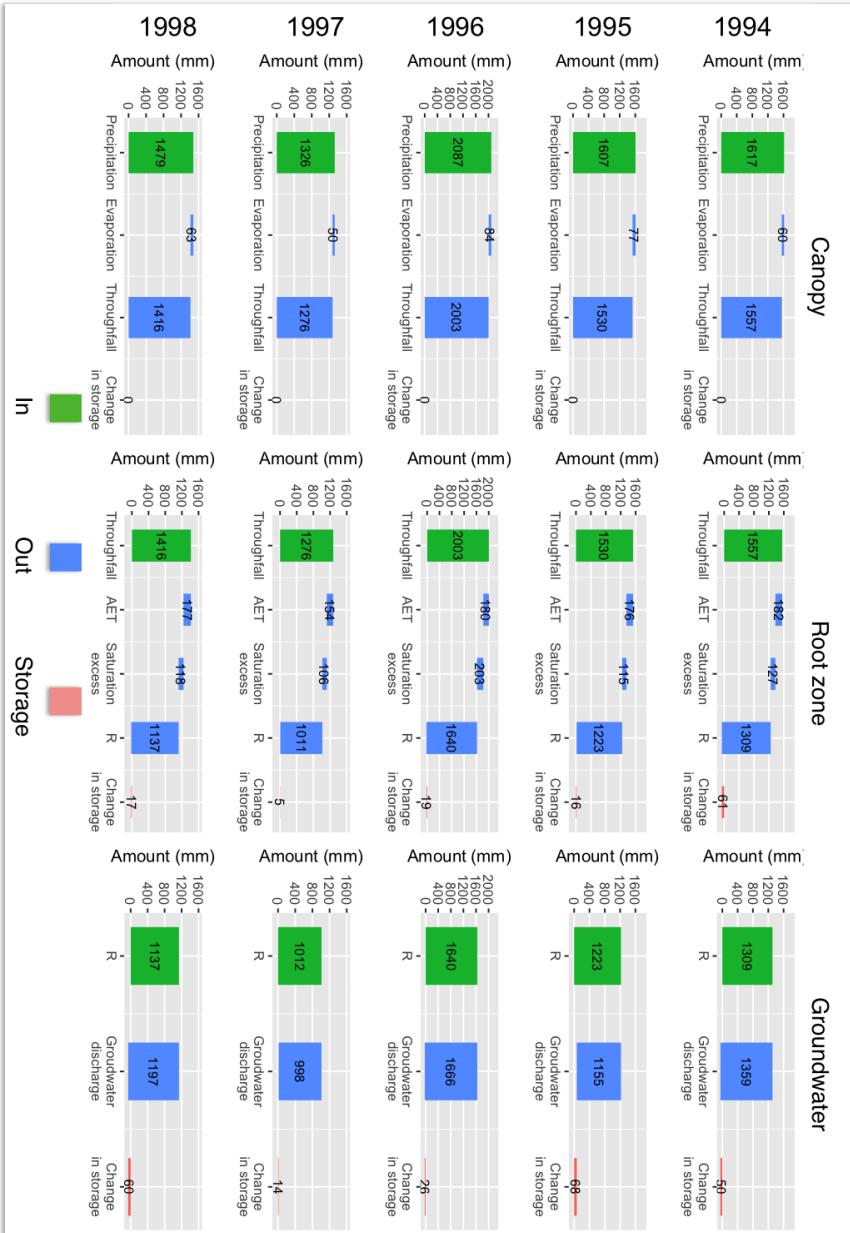
As regards the root zone, the evapotranspiration contribute is greater than in the canopy and it represents around the 14% of the actual input of the storage, which is given by the throughfall minus the saturation excess. The total actual evapotranspiration (evaporation from the wet canopy plus evapotranspiration from the root zone) represents around the 20% of annual input. This value, although is in the range reported in (Abera *et al.* , 2017a), is lower then expected (around the 30%). This is probably due to calibration problems at hourly time step caused by the hourly fluctuations of the measured discharge ( see also the results obtained at daily time-step shown in next section). The highest contribute in the root zone reservoir is represented by the recharge term of the groundwaters, where almost the total input goes into the baseflow discharge.

The waterfall charts of two selected month (March and September 1994), shows a greater variability and that, at smaller temporal aggregation, the role of storage buffers is great. The % of the relative contributes changes, and the total actual evapotranspiration varies from around 25 % in March to around



10% in September. Also the changes in storages appear marked, especially in March for the root zone reservoir.

Figure 5.11: Waterfall charts of the relative contributes of the water balance for the canopy, root zone and groundwater reservoirs. Green bars represent the inputs of the storage, blu bars represents the outputs and red bars represent the change in storage. Five years, from 1994 to 1998, are shown in order to compare the inter-annual variability of each contribute.



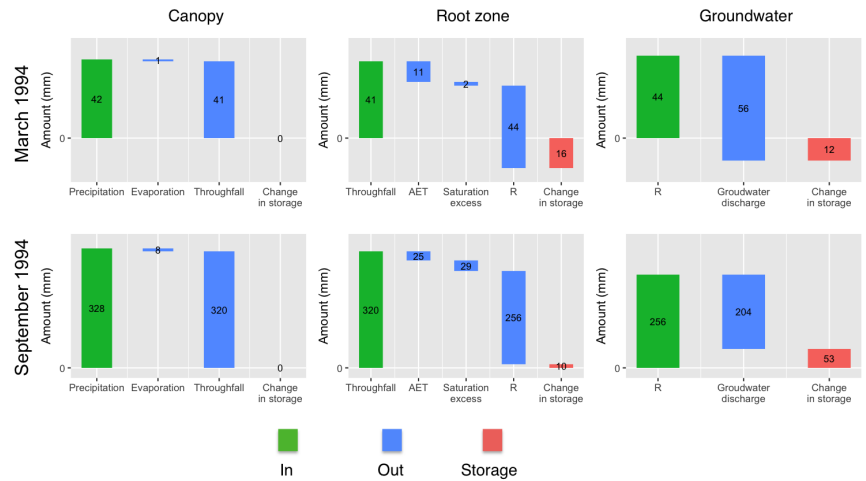


Figure 5.12: Waterfall charts of the relative contributes of the water balance for the canopy, root zone and groundwater reservoirs. Green bars represent the inputs of the storage, blue bars represents the outputs and red bars represent the change in storage. Two selected month, March 1994 and September 1994, are shown in order to compare the annual variability of each contribute.

Plot like like Figure 5.12 and 5.13 can be produced for any of the HRU, for any hour or more aggregated temporal scale.

Figure 5.13 shows the daily total actual evapotranspiration (evaporation from the wet canopy plus evapotranspiration from the root zone) in selected days of the years. As it is clear from the maps, the AET does not vary much over the year. This situation is common to many other places in humid areas, (Lewis *et al.*, 2000; Oishi *et al.*, 2010). Some HRUs presents values of AET always smaller than the rest of the basins. This is mainly given to a lower net radiation input and, thus, a lower ETP. The lower net radiation is due to a different exposition of the HRU and a lower sky-view factor (0.6). Since the evaporation from the wet canopy doesn't represent an important contribute to the total AET, the spatial variability of the canopy cannot be appreciated.

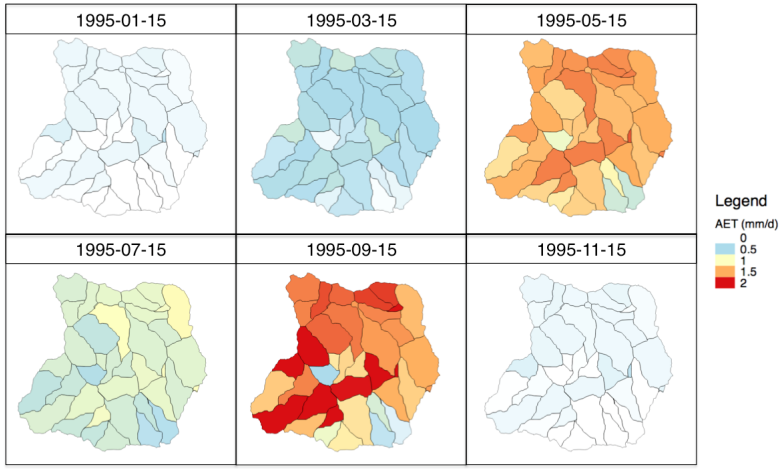


Figure 5.13: Spatial distribution of AET for six selected time steps.

The optimized values of the parameters of the hourly time-step modeling solution are shown in table 5.4.

Reservoir	Parameter	Value
Snow	$\alpha_m$	0.2531
Snow	$\alpha_f$	0.008
Snow	$\alpha_e$	8.03E-5
Snow	$\alpha_l$	0.69
Canopy	$k_c$	0.03
Root zone	$S_{r^{z_{max}}}$	207.91
Root zone	$a$	$6E^{-6}$
Root zone	$b$	2.54
Root zone	$B$	0.13
Surface runoff	$c$	0.016
Surface runoff	$d$	2.49
Groundwater	$S_{GW_{max}}$	652.97
Groundwater	$e$	141.00
Groundwater	$f$	9.58

Table 5.4: List of optimized values for each model parameters in the hourly time-step case.

A preliminary analysis of the travel times was performed as soon as the water budget terms for each HRU were obtained using the

embedded reservoir model. The complete mixing case ( $\omega_Q(t, t_{in}) = \omega_{E_T}(t, t_{in}) = 1$ ) was considered. Figure 5.14, shows the temporal evolution of mean travel times and mean residence times. Each colored curve represent the temporal evolution of the travel times computed for each HRU, the black curve represent the temporal evolution of the residence time for the entire catchment, the blue dashed line represent the trend line. The mean residence time for the overall period is around 77 days but it is clear from the trend line that it increases along the investigate period. In particular, the trend line has an intercept of 70 days and a slope of 0.01(day/day), which means that there is an increasing of the mean residence time of around 13 minutes each day. The lowest value of the the travel times is reached in correspondence of the peaks of discharge of the October 1996 and November 1998. The temporal evolution is characterized by seasonal and inter-annual fluctuations due to long term variability of climate conditions. The mean travel times shows a rather irregular behavior, reflecting the continuous interplay between reservoirs.

Figure 5.15, shows the results the temporal evolution of the mean evapotranspiration times. Each colored curve represent the temporal evolution of the evapotranspiration times computed for each HRU and the blue dashed line represent the trend line. In this case, the mean value is around 11 days and there trend line is horizontal, which means that there is no increasing of the mean evapotranspiration time over the time. The mean evapotranspiration times, are, as expected, lower then the mean travel times and present a more dynamical behavior. This means that they are easily affected by single events, producing high frequency fluctuations.

The comparison of the distribution of the mean residence times (red curve), travel times (blue curve) and evapotranspiration

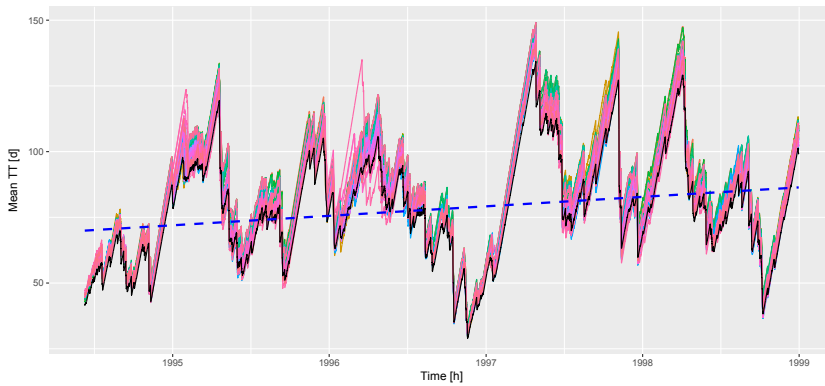


Figure 5.14: Temporal evolution of the travel times and residence times: each colored curve represent the travel times computed for each HRU, the black curve represent the residence time for the entire catchment, the blue dashed line represent the trend line. The mean residence time is around 70 days and the trend shows an increasing over time of around 13 minutes/day.

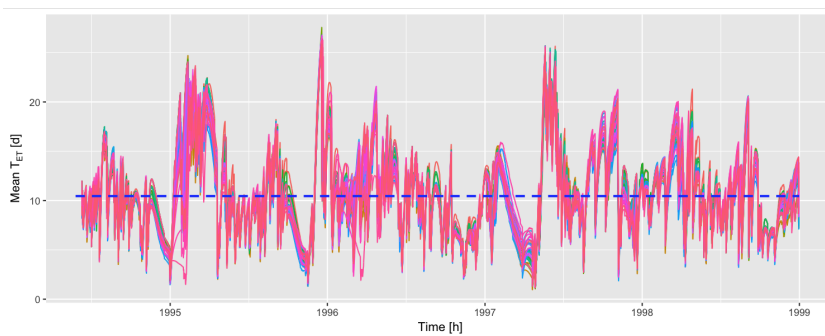


Figure 5.15: Temporal evolution of the mean evapotranspiration times: each colored curve represent the evapotranspiration times computed for each HRU, while the blue dashed line represent the trend line.

times (black curve) are shown in figure 5.16. The basic statistics of each distribution are presented in table 5.5. All the distributions are Gaussian, which is something expected, given the nature of the processes involved.

Term	Mean	Median	Mode	CV
Residence times	78.16	77.41	77	0.25
Travel times	84.24	83.12	82	0.26
Evapotranspiration times	10.63	10.12	8.2	0.39

Table 5.5: Basic statistics of the residence times, mean travel times and evapotranspiration times in hourly time-step case.

Figure 5.17 shows different time-series of the partitioning coefficient: each curve represents the time evolution of  $\Theta(t, t_{in})$  obtained considering twelve precipitation events, one for each month of a year of rainfall data and the considering the complete mixing case ( $\omega_Q(t, t_{in}) = \omega_{ET}(t, t_{in}) = 1$ ). The highest values of the coefficient ( $\Theta(t_{in}) = 0.95$ , in this case, are achieved during the coldest months of the year, in which the evapotranspiration flux is lower. On the contrary, smaller  $\Theta(t_{in})$  values were obtained in the summer months, with a minimum in June of 0.72.

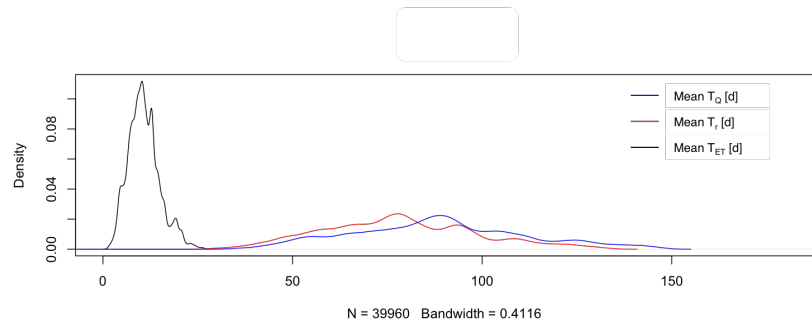


Figure 5.16: Distribution of the mean residence times, travel times and evapotranspiration times.

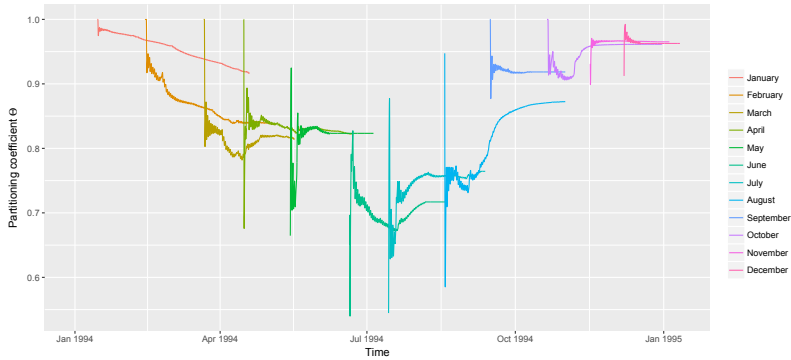


Figure 5.17: Travel time analysis obtained for the entire Posina catchment and the complete mixing case.

### 5.2.2 Daily time step

A second application of the embedded reservoirs was implemented calibrating the MS at a daily scale, instead of the hourly one. This was done to filter out the fluctuations induced by the hydropower and bottling plants, and see what this causes on the overall water budget (annual). Moreover, from the available discharge dataset, the contribution of runoff and baseflow have been estimated by interpolation of a cubic spline on the minimum values of the time series, as in Manfreda *et al.* (in review), see figure 5.18. Surface runoff was modeled using the non-linear reservoir model.



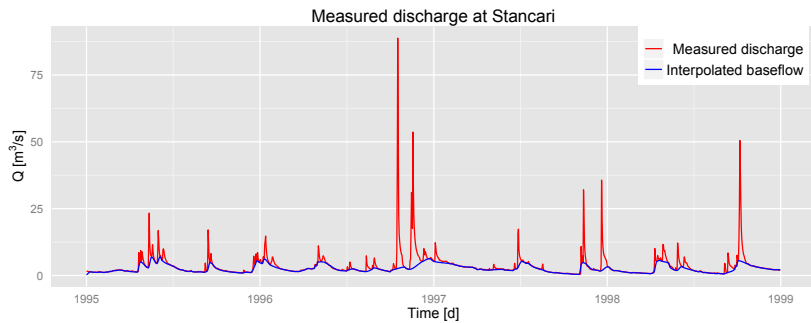


Figure 5.18: Separation of the baseflow contribute obtained using a cubic spline interpolation of the minimum values of the recorded discharge at Stancari station.

The calibration of the parameter was made against the discharge recorded at the closure of the basin (Posina at Stancari) in the period 1995-1996, while the validation was made using the entire dataset of measured discharge at Stancari, including also the measurements at Bazzoni.

Results of the application of the model are shown in figure 5.19, 5.20 and 5.21. In particular, figure 5.19 shows the comparison between the observed discharge against the simulated discharge at Stancari, where the model parameters were calibrated and at Bazzoni. The performances of the model are good and better than the hourly case, as shown in table 5.2.2, with a KGE of 0.75 and a NSE of 0.7, both evaluated on the entire period of observation (1995-1998). The indices of goodness are, as expected, better in the calibration period (1995-1996) with a KGE of 0.78, then in the validation period (1995-1998), with a KGE of 0.75. Further verification of the model performances were made comparing the measured and the simulated discharge at Bazzoni. In this case, the parameters were not calibrated but the overall results are still good, with a KGE of 0.65.

Figure 5.20, shows the comparison between the baseflow

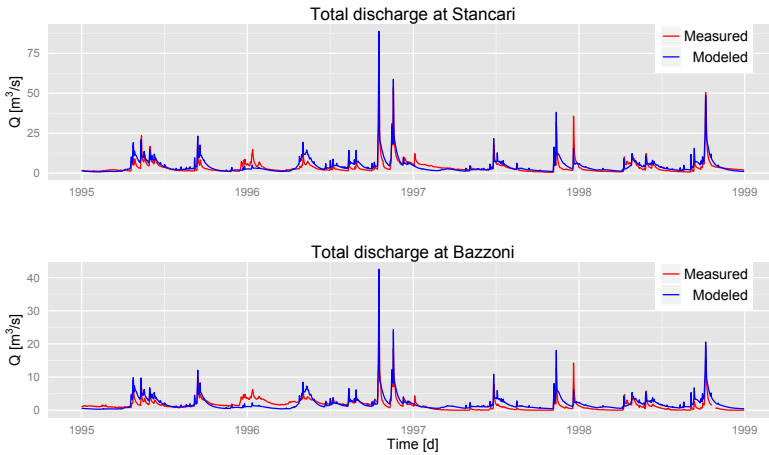


Figure 5.19: Comparison between the measured and simulated discharge at Stancari, where the calibration of the parameters was performed. A good agreement of the two series is evident, with a KGE of 0.75.

Gauge station	Calibration		Validation	
	KGE	NS	KGE	NS
*Posina at Bazzoni (38.82km <sup>2</sup> )	–	–	0.65	0.46
*Posina at Stancari (116.2km <sup>2</sup> )	0.78	0.72	0.75	0.70

Table 5.6: Results in terms of goodness of fit in calibration and validation periods for the two main stations, Bazzoni and Stancari. No calibration was performed at Bazzoni, therefore only the GOF in validation period are shown.

extracted with the mathematical filter and the baseflow simulated at Stancari, where the model parameters were calibrated. Some fitting problems are evident in correspondence of the peaks, probably due to calibration procedure.

In fact, figure 5.21 shows the results of the calibration of the model in correspondence of the peak events of October and November 1996. A almost perfect agreement is obtained between the two series, with a KGE of 0.82, NSE 0.91, MAE 2.95  $m^3/s$ , RMSE 4.30  $m^3/s$ , PBIAS (%)17.90. The results confirmed

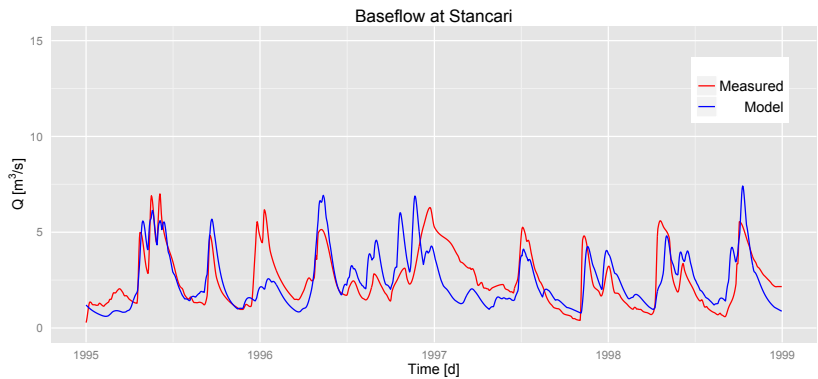


Figure 5.20: Comparison between the measured and simulated baseflow at Stancari, where the calibration of the parameters was performed.

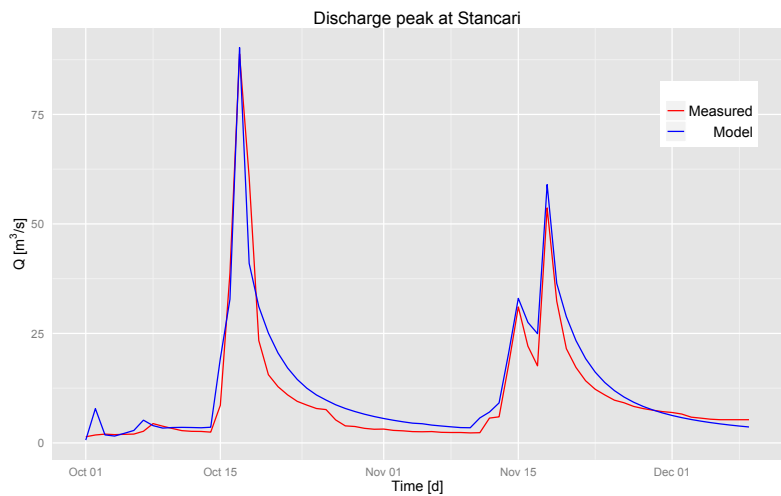


Figure 5.21: Comparison between the measured and simulated discharge peaks at Stancari. A very good agreement of the two series is evident, with a KGE of 0.82.

that calibration problems could be hidden in the previous results.

The waterfall charts for the analysis of the relative contributes in the case of daily time-step are shown in figure 5.22. It has to be compared with figure 5.11 that shows the estimation of the same

quantity but obtained with the hourly simulation. What is immediately clear is that the evaporation from the wet canopy is bigger (almost ten times) than in the hourly case. The reservoir, in fact, are exchanging volumes of smaller water than in the hourly case and the storage available for the evaporation of the wet canopy is bigger. As regards the root zone, the volumes of actual evapotranspiration are comparable to the hourly case and around  $180 \text{ mm/year}$ . The total actual ET, in this case, represents around the 30 % of the total inputs, which is coherent with the results of Abera *et al.* (2017b). Also in this case, an increasing of the input of the reservoirs is reflected in an increasing of the discharge, with smooth variation of the AET and small changes in the storage.

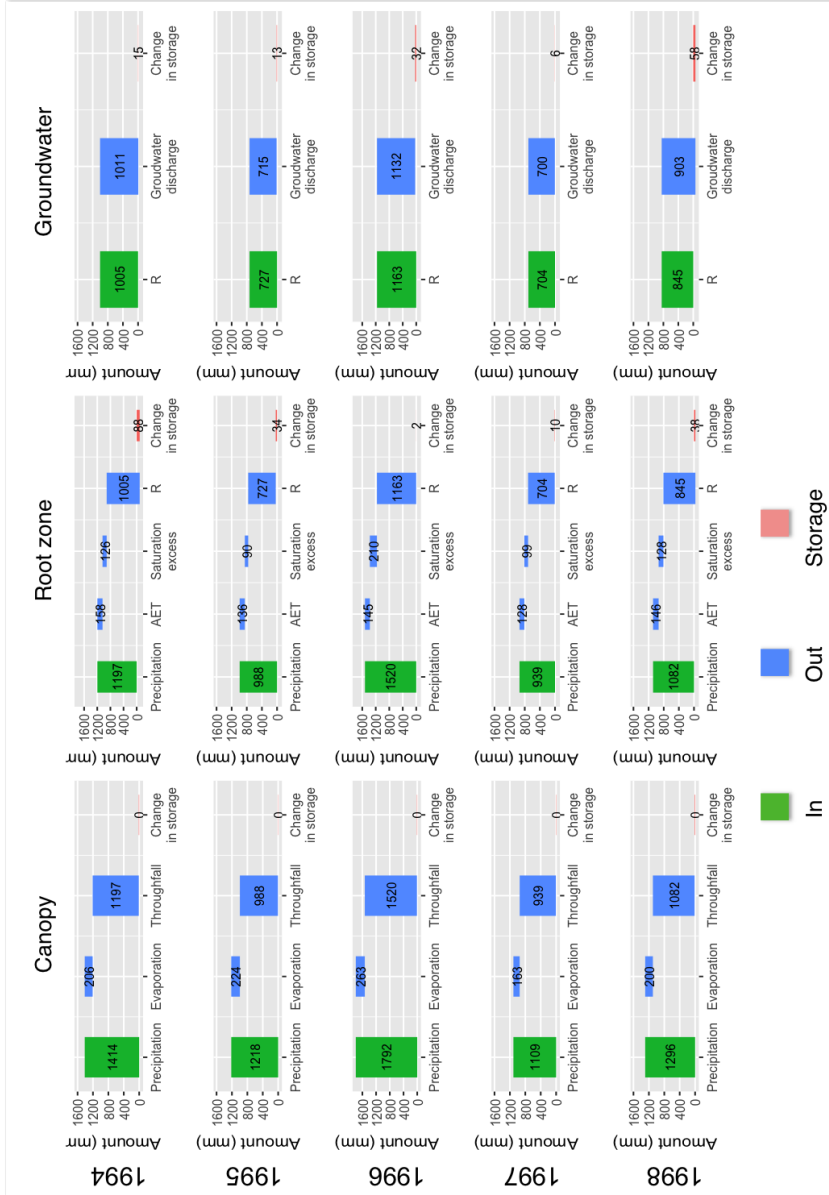


Figure 5.2.2: Waterfall charts of the relative contributes of the water balance for the canopy, root zone and groundwater reservoirs. Green bars represent the inputs of the storage, blu bars represents the outputs and red bars represent the change in storage. Five years, from 1994 to 1998, are shown in order to compare the inter-annual variability of each contribute.

Thanks to the model structure, it is possible to investigate each input and output of the storages. An interesting analysis of the fluxes could be done on the canopy layer. Figure 5.23 shows the comparison of the LAI, in the first graph, the ratio between the gross precipitation (P) and the drainage from the canopy (D), in the second graph, and the evaporation from the wet canopy, in the third graph. These results were obtained for the HRU ID 42, which is one of the HRU with the less no values in the input dataset. It is evident that the interception is greater when LAI increases, with a decrease of the ratio between P and D of almost 3 times. The ratio in fact, passes from around 35 in April 1998 to around 10 during summer in September 1998. AET increases and decreases according to the normal seasonality of PET. Zero AET is due to no precipitation and, thus, no water on the wet canopy.

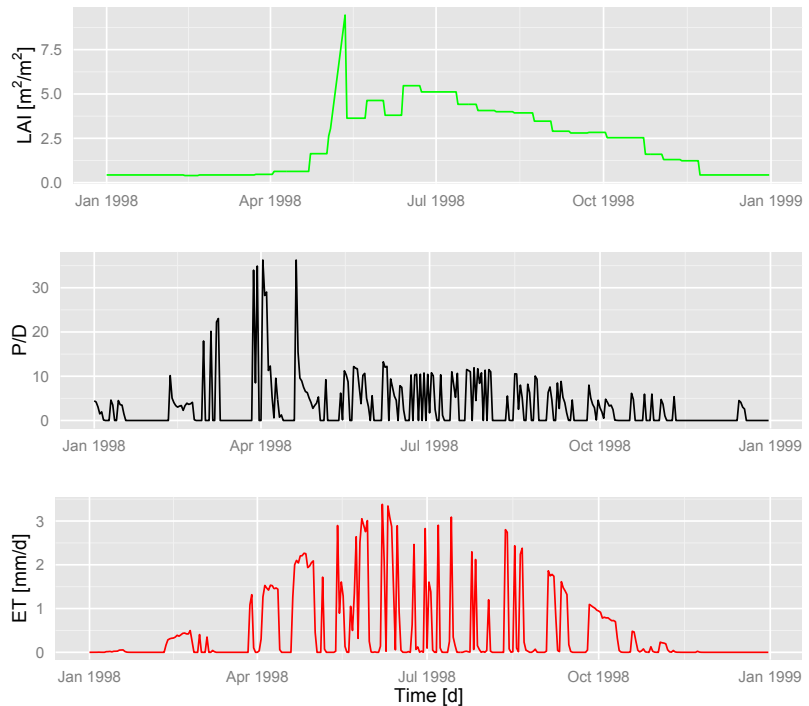


Figure 5.23: Comparison of the time series of LAI, gross precipitation (P) and drainage (D) ratio and AET from the wet canopy. According to the LAI increasing, the ratio decreases till 3 times during summer period. The AET varies according to the storage of the wet canopy and goes to zero when the storage is null.

Figure 5.24, shows the relative amount of actual evapotranspiration from the wet canopy and from the root zone, respect to the potential evapotranspiration obtained for a single HRU (ID 42). The evaporation from the wet canopy represents the 58 % of the total actual evapotranspiration. The total actual evapotranspiration (wet canopy plus root zone) is 1730 mm in five years, which represents the 25 % of the whole budget, confirming the results obtained in Abera *et al.* (2017a).

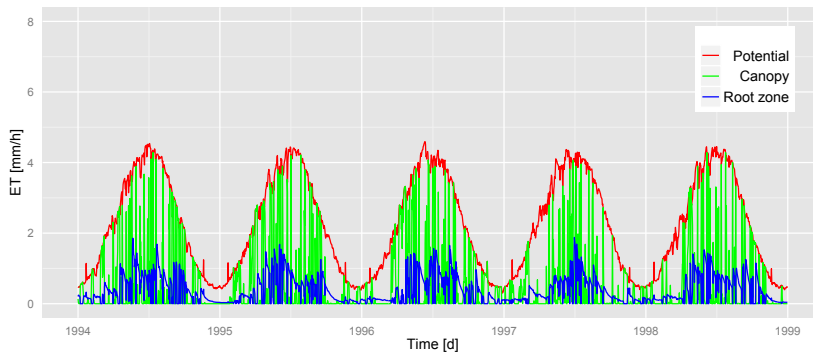


Figure 5.24: Time series of PET (red lines) VS the AET from the wet canopy (green lines) and root zone (blue lines).

Figure 5.25 shows the scatterplot of the modeled throughfall compared to the gross precipitation: black solid line represents the bisector while the blue solid line represents the regression line. As expected, the throughfall is lower than the gross precipitation, especially for high precipitation rates. For small values (precipitation  $< 1$  mm), some negligible errors are present, probably due to integration procedures.



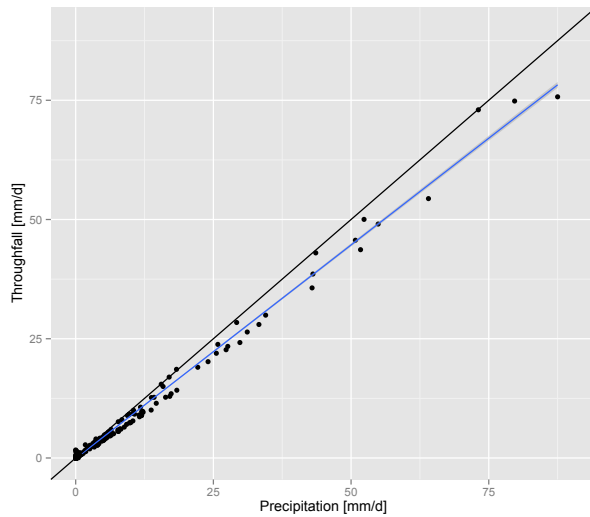


Figure 5.25: Scatter plot between throughfall and gross precipitation: the black solid line represents the bisector, while the blue solid line represents the regression line. As expected throughfall is lower than precipitation, except for really small values (less than 0.1) that could be considered integration errors.

Also interesting is the variation of ratio between the cumulative throughfall and the cumulative gross precipitation, with the cumulative gross precipitation, which is representative of the relative importance of interception losses, figure 5.26. On the small time scales, the ratio presents a high variability while it tends to a constant value after annual integration. The behavior of the curve presents seasonality with a larger throughfall in winter, as expected. At annual scale, however, a variation from 0.85 to 0.89 is still present, probably due to a seasonal variation of the LAI.

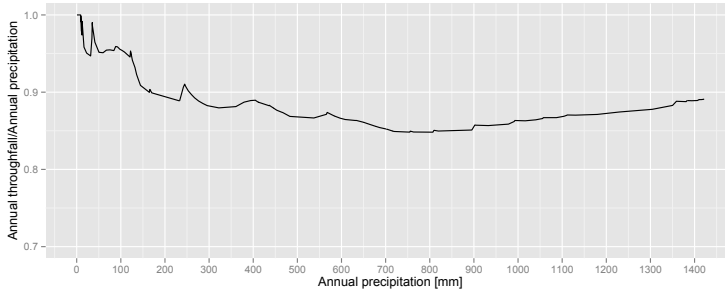


Figure 5.26: Variation of the ratio between cumulative throughfall and the cumulative gross precipitation with the cumulative gross precipitation. On the small time scales, the ratio presents a high variability while it tends to a final value of 0.88 on the annual scale.

The optimized values of the parameters for the daily time-step modeling solution are shown in table 5.7.

Reservoir	Parameter	Value
Snow	$\alpha_m$	0.95
Snow	$\alpha_f$	0.08
Snow	$\alpha_e$	6.09E-5
Snow	$\alpha_l$	0.32
Canopy	$k_c$	0.25
Root zone	$S_{rz_{max}}$	220.20
Root zone	$a$	0.003
Root zone	$b$	1.74
Root zone	$B$	0.42
Surface runoff	$c$	0.26
Surface runoff	$d$	2.49
Groundwater	$S_{gw_{max}}$	532.20
Groundwater	$e$	255.68
Groundwater	$f$	7.58

Table 5.7: List of optimized values for each model parameters in the daily time-step case.

The same preliminary analysis of the travel times was performed for the daily case. The complete mixing case

( $\omega_Q(t, t_{in}) = \omega_{E_T}(t, t_{in}) = 1$ ) was considered. Figure 5.27, show the temporal evolution of the travel times and mean residence

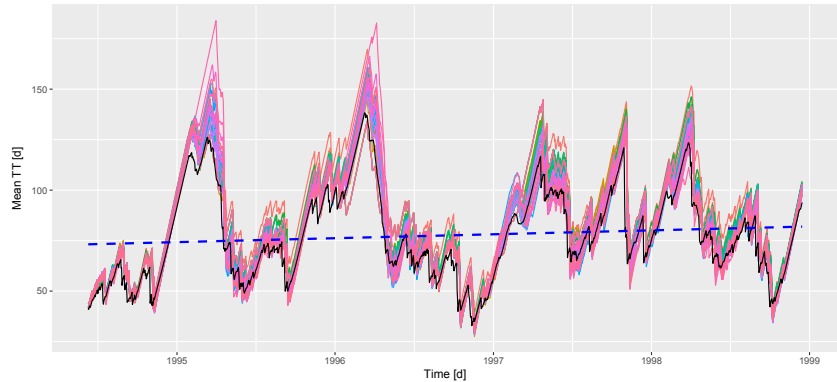


Figure 5.27: Temporal evolution of the mean travel times and residence times: each colored curve represent the travel times computed for each HRU, the black curve represent the mean residence time, while the blue dashed line represent the trend line.

times. Each colored curve represent the temporal evolution of the travel times computed for each HRU, the black curve represent the temporal evolution of the residence time for the entire catchment, the blue dashed line represent the trend line. In this case, the mean residence time for the overall period is around 77 days, which is slightly bigger than the hourly case. Also in this case, the trend line show an increases of the mean residence time, which is lower than in the hourly case. In fact, the trend line has an intercept of 73 days and a slope of 0.0055 (day/day), which means that there is an increasing of the mean residence time of around 7 mins each day. The lowest value of the the travel times is reached in correspondence of the peaks of discharge of the October 1996 and November 1998. The temporal evolution is characterized by seasonal and inter-annual fluctuations, which are slightly dumped respect to the hourly case.

Figure 5.28, shows the temporal evolution of the mean evapotranspiration times for the daily time-step. Each colored curve represent the temporal evolution of the

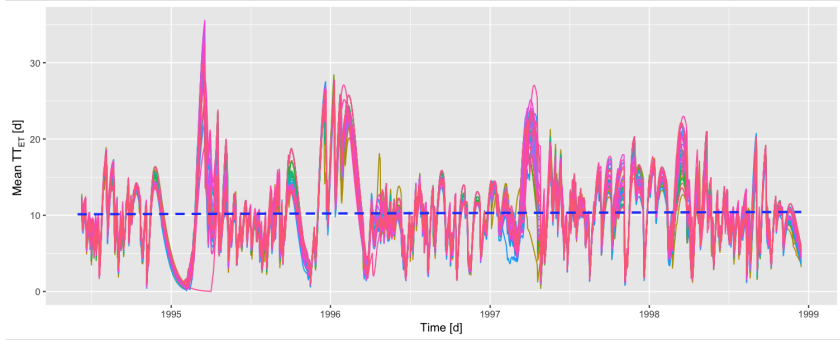


Figure 5.28: Temporal evolution of the mean evapotranspiration times: each colored curve represent the travel times computed for each HRU, while the blue dashed line represent the trend line.

evapotranspiration times computed for each HRU and the blue dashed line represent the trend line. In this case, the trend is almost horizontal (slope of 0.0003 day/day) and also in this case their temporal evolution shows a more dynamical behavior compared to the travel times. The mean value is around 12 days, while the minimum values are of 1-2 days, reached in driest periods.

Finally, the same comparison of the distributions of the mean residence times (red curve), travel times (blue curve) and evapotranspiration times (black curve) is performed and shown in figure 5.29. The basic statistics of each distribution are presented in table 5.8.

Term	Mean	Median	Mode	CV
Residence times	77.54	73.32	68	0.28
Travel times	84	79.60	74	0.30
Evapotranspiration times	10.3	9.8	9	0.47

Table 5.8: Basic statistics of the residence times, mean travel times and evapotranspiration times in daily time-step case.

Figure 5.30 shows different time-series of the partitioning coefficient: each curve represents the time evolution of  $\Theta(t, t_{in})$

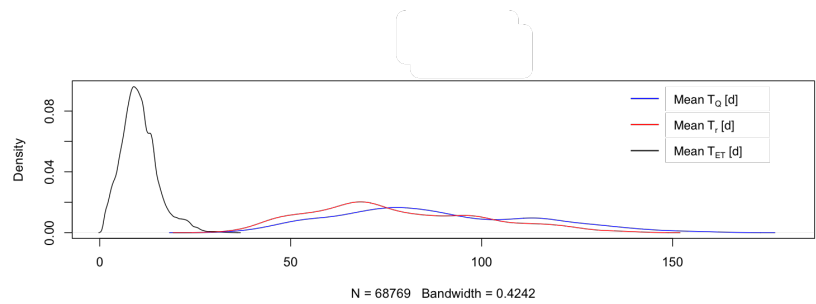


Figure 5.29: Distribution of the mean residence times (red curve), travel times (blue curve) and evapotranspiration times (black curve).

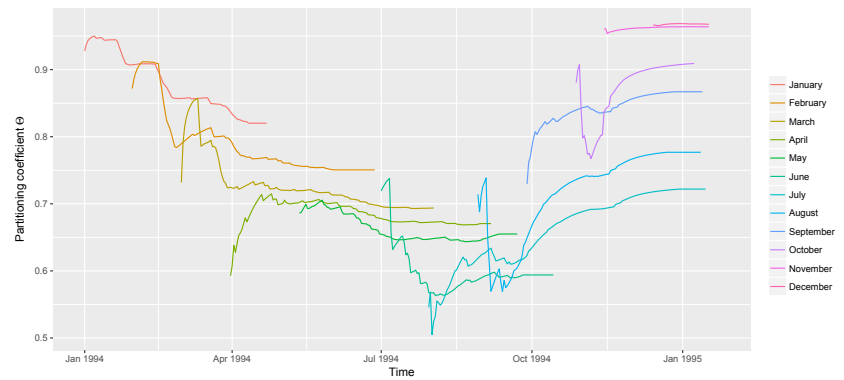


Figure 5.30: Evolution of the partitioning coefficient in one year of daily simulation: the highest value are achieved in January while the lowest in June.

obtained considering twelve precipitation events at daily time-scale, considering the complete mixing case ( $\omega_Q(t, t_{in}) = \omega_{ET}(t, t_{in}) = 1$ ). The curves present a smoother behavior respect to the hourly case due to the different time step. The bigger volumes of total AET resulted in the daily time-scale determined smaller values of  $\Theta$ , especially during the summer months. The minimum value is now around 0.6, while in the hourly time-scale was around 0.75.

### 5.3 Less explored and unexplored MS

The flexible structure allowed by the OMS3 components permits to experiment several MSs. For instance the one depicted in figure 5.31, in which the role of the groundwater and the surface flow are exchanged. After this exchange, the amount of water assigned to the root zone remains fixed by the partitioning rule, having consequences on the overall behavior of the system.

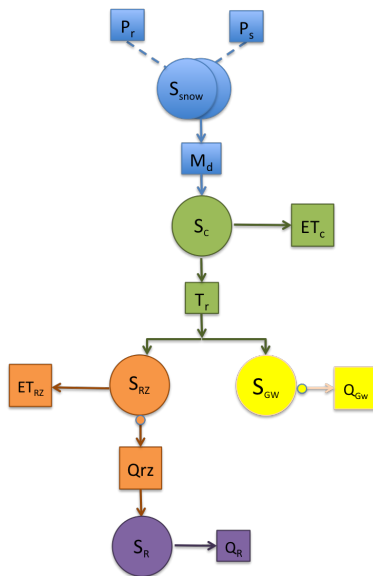


Figure 5.31: Example of a different schema of connections: after the partition, the volume in excess remains in the root zone reservoir and, then, transferred to the surface flow, while the volume infiltrated feeds the groundwater reservoir.

The results of the previous connection, shown in figure 5.32, proved to be not correct, since the root zone reservoir is unable to sustain all the output fluxes, leading to an excessive emptying of the storage. In fact, the mean storage is around  $15 \text{ mm}^3 / \text{mm}^2$ , proving to be not physical. The simulation shown in the figure 5.32 was made using only on year of data, since the numerical solver was not able to solve the ODE given the low values of the

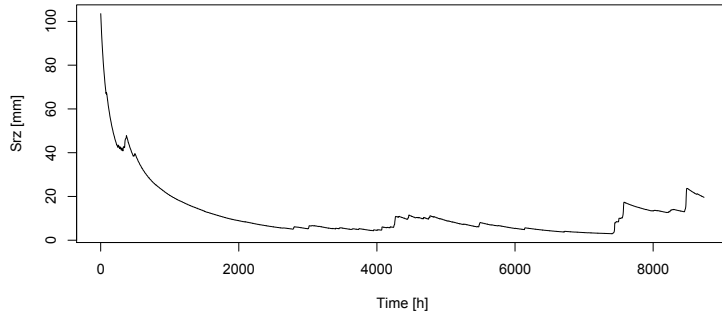


Figure 5.32: Root zone storage obtained using the previous schema and one year of data.

storage.

Although the negative results obtained using the previous test, it shows the great flexibility of the presented infrastructure, which easily allows to experiment different hydrological connectivities, while including multiple options for individual process.

Making treasure of the experience made, other types of connections should be (in principle) tried, as the one depicted by the figure 5.33. In this case, respect to the schema proposed in figure 5.31, a dynamical feedback is present from the groundwater to the root zone. This could help to overcome the previous problems of the root zone storage, preventing its emptying.

Other feedbacks can be introduced by considering, for instance, a riparian zone, where channel water can interact with groundwater. In the figure below  $Q_{R_i}$  and  $Q_{R_j}$  are upstream runoff fluxes that, instead of being simply added to the global stream, is partially partitioned to the groundwater of the downstream system. In this case the interaction between the root zone and the groundwater is not ruled, anymore, by the partition described in section 3.1.3, but the two reservoirs

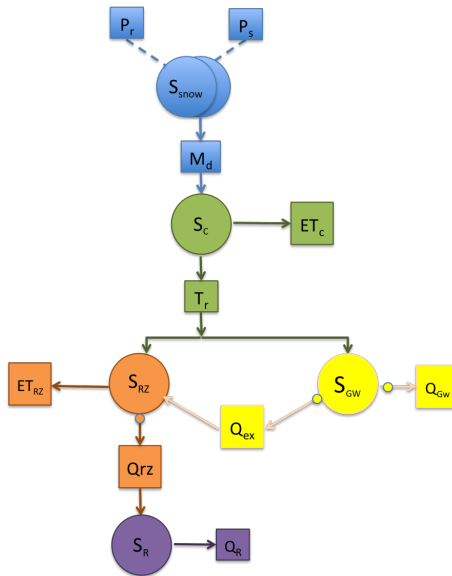


Figure 5.33: Example of a different schema of connections: a dynamical feedback from the groundwater to the root zone is present.

exchange dynamical feedbacks through two non-linear fluxes,  $R_e$  and  $Q_{ex}$ , which allows exfiltration from groundwater to the root zone, when the conditions allows it.

These PN were shown to enlarge the range of possible MSs, which open the research to multiple analysis. For example, a comparison between embedded reservoirs model and Richards equation based soil-water-atmosphere-plant model can be performed, similar to the one presented in Romano *et al.* (2011). Certainly, they would require an improvement of the actual Net3 infrastructure, due to the complication that feedbacks introduce in the resulting system of ODEs. These complication (which requires the simultaneous resolution of some equations, or, from the the numerical point of view, iterative loops) is particularly impacting since the mechanism that, in this moment, parallelize computation in Net3 does not include them. However, this is material for future work.



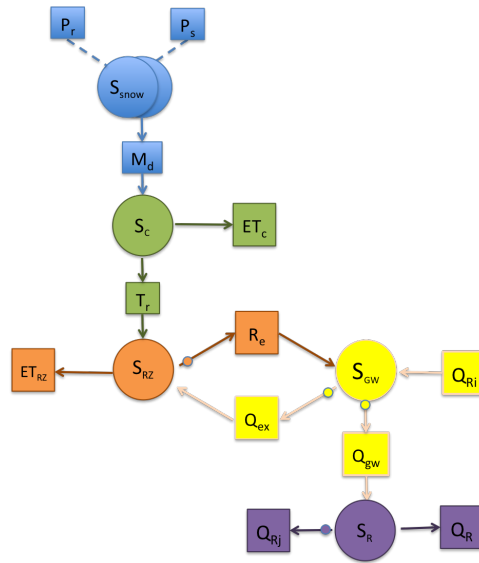


Figure 5.34: Example of a different schema of connections: root zone and groundwater reservoirs exchange dynamical feedbacks through two non-linear fluxes.

## 5.4 Final remarks

The chapter presented the results of the semi-distributed model for runoff and evapotranspiration, described in Chapter 3. The model considers five storages, snow, canopy, root zone, groundwater, runoff for each HRU in which the basin is subdivided. Each storage is implemented as a OMS3 component. In this case, the modeling solution simulates the discharge and evapotranspiration, trying to distinguish all the different contributes of the storages connected. The MS is integrated in JGrass-NewAge, taking advantage of all the ancillary existing components, including those for radiation and Kriging described in the next chapter. The partition of the fluxes in the five storages is not particularly original per se. However, the use of components, allows inspection and producing the water budget of each one of them, which is not usually easily

obtainable with different informatics.

The discharge prediction is in line with other modeling performances, either with hourly or daily calibrations.

From the analysis of the waterfall charts obtained for both the hourly and the daily time-step cases, it is clear that increases in precipitation tend to contribute directly to Q with minor effects on S and ET. The melting/rain represents the main source of variability in the budget, and (therefore) the ET shows a smoother behavior, since the radiation, which is the main driver, is consistent across the years. In the case of hourly simulations, the total actual evapotranspiration (evaporation from the wet canopy plus evapotranspiration from the root zone) represents around the 20% of annual input and is lower than expected (around the 30%). This is probably due to calibration problems caused by the hourly fluctuations of the measured discharge.

In the daily application, the waterfall charts shown that the reservoir are exchanging smaller volumes of water and the storage available for the evaporation of the wet canopy is bigger. Therefore, the total actual ET, in this case, represents around the 30 % of the total inputs, which is coherent with the results of Abera *et al.* (2017b).

A preliminary travel time analysis was performed on the results obtained with the embedded reservoirs model, both at hourly and daily time-steps. Both cases show an increasing trend in the mean travel times and mean residence times which means that the final values will be reached with longer simulations, not possible in this moment, due to the available dataset. However, the increasing trend shows that the model proved to be able in the reproduction of the travel times, since the water is getting older with time. The temporal evolution of the mean travel times and residence times are characterized by seasonal and inter-annual fluctuations, while the temporal evolution of the

mean evapotranspiration times is highly fluctuating, which means that the evapotranspiration times are easily affected by single events. The mean value of the residence times is around 77 days, mean value of the travel times is around 84 days, mean value of the evapotranspiration times is around 10 days,. The analysis of the partitioning coefficient, shown. In both cases, the computation of the partitioning coefficients between discharge and evapotranspiration shown a coherence respect to the relative amounts of the volume computed, with the lower values achieved during the summer months.

Further analysis on the travel times and on the mixing are upcoming.

At the end of the chapter, other MSs are shown and (partially) tested, taking advantage of the great flexibility of the presented infrastructure.



## A COUPLE OF NEW ANCILLARY COMPONENTS

This chapter presents two ancillary components, the LongWave Radiation Balance component (LWRB) and the Spatial Interpolation (SI) components, which two examples of the refactoring of JGrass-NewAge and of GEOframe best practices.

Ten algorithms for estimating downwelling longwave atmospheric radiation ( $L_{\downarrow}$ ) and one for upwelling longwave radiation ( $L_{\uparrow}$ ) are integrated into the JGrass-NewAge modeling system. The algorithms are tested against energy flux measurements available for 24 sites in North America to assess their reliability. These new JGrass-NewAge model components are used i) to evaluate the performances of simplified models (SMs) of  $L_{\downarrow}$ , as presented in literature formulations, and ii) to determine by automatic calibration the site-specific parameter sets for  $L_{\downarrow}$  in SMs. Moreover, the  $L_{\uparrow}$  SM is tested by using three different temperatures (surface soil temperature, air temperature at 2 m elevation, and soil temperature at 4 cm depth) and model performances are then assessed.

A package for the interpolation of climatological variables, such

as temperature and precipitation, using Kriging techniques is then presented. The purposes are (1) to present a geostatistical software easy to use, easy to plug-in in a hydrological model, fast and flexible (2) to show a practical example of an accurately designed software in the perspective of reproducible research, (3) to show the goodness of the software applications, in order to have a reliable alternative to other traditionally used tools. Ten types of theoretical semivariograms and four types of Krigings were implemented and then gathered into Object Modeling System compliant components. The package provides real time optimization for semivariogram and kriging parameters, improving the estimation of spatially distributed data. The software was tested on Isarco River basin, Italy, using temperature and rainfall data retrieved from 97 meteorological stations. For both variables, good interpolation results were obtained and then compared to the results from the notable R package, *gstat*.

## 6.1 LWRB

Longwave radiation is an important component of the radiation balance on earth and it affects many phenomena, such as evapotranspiration, snow melt (Plüss & Ohmura, 1997), glacier evolution (MacDonell *et al.*, 2013), vegetation dynamics (Rotenberg *et al.*, 1998), plant respiration, and primary productivity (Leigh Jr, 1999). Longwave radiation is usually measured with pyrgeometers, but these are not normally available in basic meteorological stations, even though an increasing number of projects has been developed to fill the gap (Augustine *et al.*, 2000, 2005; Baldocchi *et al.*, 2001). The use of satellite products to estimate longwave solar radiation is increasing (GEWEX, Global Energy and Water cycle Experiment,

ISCCP the International Satellite Cloud Climatology Project) but they have too coarse a spatial resolution for many hydrological uses. Therefore, models have been developed to solve energy transfer equations and compute radiation at the surface (e.g., Key & Schweiger, 1998; Kneizys *et al.*, 1988). These physically based and fully distributed models provide accurate estimates of the radiation components. However, they require input data and model parameters that are not easily available. To overcome this issue, simplified models (SM), which are based on empirical or physical conceptualizations, have been developed to relate longwave radiation to atmospheric proxy data such as air temperature, water vapor deficit, and shortwave radiation. They are widely used and provide clear sky (e.g., Ångström, 1915; Brunt, 1932; Idso & Jackson, 1969) and all-sky estimations of downwelling ( $L_{\downarrow}$ ) and upwelling ( $L_{\uparrow}$ ) longwave radiation (e.g., Brutsaert, 1975; Iziomon *et al.*, 2003).

SM performances have been assessed in many studies by comparing measured and modeled  $L_{\downarrow}$  at hourly and daily time-steps (e.g., Sugita & Brutsaert, 1993b; Iziomon *et al.*, 2003; Juszak & Pellicciotti, 2013; MacDonell *et al.*, 2013; Schmucki *et al.*, 2014). Hatfield *et al.* (1983) was among the first to present a comparison of the most used SMs in an evaluation of their accuracy. They tested seven clear-sky algorithms using atmospheric data from different stations in the United States. In order to validate the SMs under different climatic conditions, they performed linear regression analyses on the relationship between simulated and measured  $L_{\downarrow}$  for each algorithm. The results of the study show that the best models were Brunt (1932), Brutsaert (1975) and Idso (1981). Flerchinger *et al.* (2009) made a similar comparison using more formulations (13) and a wider data-set from North America and China, considering all possible sky conditions. Finally, Carmona *et al.* (2014) evaluated the

performance of six SMs, with both literature and site-specific formulations, under clear-sky conditions for the sub-humid Pampean region of Argentina.

However, none of the above studies have developed a method to systematically estimate site-specific model parameters for location where measurements are not available using basic site characteristics.

This work presents the LongWave Radiation Balance package (LWRB) of the JGrass-NewAGE modeling system, (Formetta *et al.*, 2014a). LWRB implements 10 formulations for  $L_{\downarrow}$  and one for  $L_{\uparrow}$  longwave radiation. The package was systematically tested against measured  $L_{\downarrow}$  and  $L_{\uparrow}$  longwave radiation data from 24 stations across the contiguous USA, chosen from the 65 stations of the AmeriFlux Network. Unlike all previous works, the LWRB component follows the specifications of OMS3 framework, (David *et al.*, 2013). Therefore, it can use all of the JGrass-NewAge tools for the automatic calibration algorithms, data management and GIS visualization, and it can be seamlessly integrated into various modeling solutions for the estimation of water budget fluxes (Formetta *et al.*, 2014a). Moreover, differently from other studies, all the tools used in this paper are open-source, well documented, and ready for practical use by other researchers and practitioners.

### 6.1.1 Methodology

The SMs for  $L_{\uparrow}$  [ $\text{W m}^{-2}$ ] and  $L_{\downarrow}$  [ $\text{W m}^{-2}$ ] longwave radiation are based on the Stefan-Boltzmann equation:

$$L_{\downarrow} = \epsilon_{all-sky} \cdot \sigma \cdot T_a^4 \quad (6.1)$$

$$L_{\uparrow} = \epsilon_s \cdot \sigma \cdot T_s^4 \quad (6.2)$$



where  $\sigma = 5.670 \cdot 10^{-8}$  [ $\text{W m}^{-2} \text{K}^{-4}$ ] is the Stefan-Boltzmann constant,  $T_a$  [K] is the air temperature,  $\epsilon_{all-sky}$  [-] is the effective atmospheric emissivity,  $\epsilon_s$  [-] is the soil emissivity and  $T_s$  [K] is the surface soil temperature. To account for the increase of  $L_1$  in cloudy conditions,  $\epsilon_{all-sky}$  [-] is formulated according to eq. (6.3):

$$\epsilon_{all-sky} = \epsilon_{clear} \cdot (1 + a \cdot c^b) \quad (6.3)$$

where  $c$  [-] is the cloud cover fraction and  $a$  [-] and  $b$  [-] are two calibration coefficients. Site specific values of  $a$  and  $b$  are presented in Brutsaert (1975), ( $a=0.22$  and  $b=1$ ), Iziomon *et al.* (2003) ( $a$  ranges between 0.25 and 0.4 and  $b=2$ ) and Keding (1989) ( $a=0.183$  and  $b=2.18$ ). In our modeling system  $a$  and  $b$  are calibrated to fit measurement data under all-sky conditions. The cloud cover fraction,  $c$ , can be estimated from solar radiation measurements (Crawford & Duchon, 1999), from visual observations (Alados-Arboledas *et al.*, 1995, Niemelä *et al.*, 2001), and from satellite data (Sugita & Brutsaert, 1993a) or it can be modeled as well. In this study we use the formulation presented in Campbell (1985) and Flerchinger (2000), where  $c$  is related to the clearness index  $s$  [-], i.e. the ratio between the measured incoming solar radiation,  $I_m$  [ $\text{W m}^{-2}$ ], and the theoretical solar radiation computed at the top of the atmosphere,  $I_{top}$  [ $\text{W m}^{-2}$ ], according to  $c = 1 - s$  (Crawford & Duchon, 1999). This type of formulation needs a shortwave radiation balance model to estimate  $I_{top}$  and meteorological stations to measure  $I_m$ ; also, it cannot estimate  $c$  at night. In our application, the fact that the SMs are fully integrated into the JGrass-NewAge system allows us to use the shortwave radiation balance model (Formetta *et al.*, 2013b) to compute  $I_{top}$ . Night-time values of  $c$  are computed with a linear interpolation between its values at the last hour of daylight and the first hour of daylight on consecutive days. The computation of the first

and last hour of the day is based on the model proposed in Formetta *et al.*, 2013b that follows the approach proposed in Corripio (2002), equations (4.23)-(4.25). The sunrise occurs at  $t = 12 \cdot (1 - \omega/\pi)$  and the sunset will be at  $t = 12 \cdot (1 + \omega/\pi)$  where  $\omega$  is the hour angle, i.e. the angle between the observer meridian and the solar meridian. It is zero at noon and positive before noon. Those equations are based on the assumption that sunrise and sunset occur at the time when the z coordinate of the sun vector equals zero.

The formulation presented in equation 6.3 was proposed by Bolz (1949) applied in other studies (Carmona *et al.*, 2014; Maykut & Church, 1973; Jacobs, 1978; Niemelä *et al.*, 2001). Evaluating the effectiveness of different formulations respect to equation 6.3 is still an open question which is not object of the current paper. It has been investigated in several studies (i.e., Flerchinger *et al.*, 2009; Juszak & Pellicciotti, 2013, and references therein) and some of them recommended the one proposed by Unsworth & Monteith (1975).

Ten SMs from literature have been implemented for the computation of  $\epsilon_{clear}$ . Table 6.1 specifies assigned component number, component name, defining equation, and reference to the paper from which it is derived. X, Y and Z are the parameters provided in literature for each model, listed in Table 6.2.

ID	Name	Formulation	Reference
1	Angstrom	$\epsilon_{clear} = X - Y \cdot 10^{2e}$	Ångström (1915)
2	Brunt's	$\epsilon_{clear} = X + Y \cdot e^{0.5}$	Brunt (1932)
3	Swinbank	$\epsilon_{clear} = (X \cdot 10^{-13} \cdot T_a^5) / (\sigma \cdot T_a^4)$	Swinbank (1963)
4	Idso and Jackson	$\epsilon_{clear} = 1 - X \cdot \exp(-Y \cdot 10^{-4} \cdot (273 - T_a)^2)$	Idso & Jackson (1969)
5	Brutsaert	$\epsilon_{clear} = X \cdot (e/T_a)^{1/Z}$	Brutsaert (1975)
6	Idso	$\epsilon_{clear} = X + Y \cdot 10^{-4} \cdot e \cdot \exp(1500/T_a)$	Idso (1981)
7	Monteith and Unsworth	$\epsilon_{clear} = X + Y \cdot \sigma \cdot T_a^4$	Monteith & Unsworth (1990)
8	Konzelmann	$\epsilon_{clear} = X + Y \cdot (e/T_a)^{1/8}$	Konzelmann <i>et al.</i> (1994)
9	Prata	$\epsilon_{clear} = [1 - (X + w) \cdot \exp(-(Y + Z \cdot w)^{1/2})]$	Prata (1996)
10	Dilley and O'Brien	$\epsilon_{clear} = (X + Y \cdot (T_a/273.16)^6 + Z \cdot (w/25)^{1/2}) / (\sigma \cdot T_a^4)$	Dilley & O'Brien (1998)

Table 6.1: Clear sky emissivity formulations:  $T_a$  is the air temperature [K],  $w$  [ $\text{kg}/\text{m}^2$ ] is precipitable water =  $4650 [e_0/T_a]$  and  $e$  [kPa] is screen-level water-vapor pressure. The models follow the formulations presented in used in Flerchinger (2000). The Angstrom and Brunt model was presented as cited by Niemelä *et al.* (2001). Konzelmann uses water vapor pressure in [Pa] not [kPa].

The models presented in Table 6.1 were proposed with coefficient values ( $X, Y, Z$ ) strictly related to the location in which the authors applied the model and where measurements of  $L_{\downarrow}$  radiation were collected. Coefficients reflect climatic, atmospheric and hydrological conditions of the sites, and are reported in Table 6.2.

ID	Name	X	Y	Z
1	Angstrom	0.83	0.18	-0.07
2	Brunt	0.52	0.21	[-]
3	Swinbank	5.31	[-]	[-]
4	Idso and Jackson	0.26	-7.77	[-]
5	Brutsaert	1.72	7	[-]
6	Idso	0.70	5.95	[-]
7	Monteith and Unsworth	-119.00	1.06	[-]
8	Konzelmann et al	0.23	0.48	[-]
9	Prata	1.00	1.20	3.00
10	Dilley and O'brien	59.38	113.70	96.96

Table 6.2: Model parameter values as presented in their literature formulation.

The formulation of the  $L_{\uparrow}$  requires the soil emissivity, which usually is a property of the nature of a surface, and the surface soil temperature. Table 6.3 shows the literature values (Brutsaert,

2005) of the soil emissivity for different surface types:  $\epsilon_s$  varies from a minimum of 0.95 for bare soils to a maximum of 0.99 for fresh snow.

Nature of surface	Emissivity
Bare soil (mineral)	0.95 – 0.97
Bare soil (organic)	0.97 – 0.98
Grassy vegetation	0.97 – 0.98
Tree vegetation	0.96 – 0.97
Snow (old)	0.97
Snow (fresh)	0.99

Table 6.3: Soil emissivity for surface types (Brutsaert, 2005).

It is well known that surface soil temperature measurements are only available at a few measurement sites, therefore, under the hypothesis that difference between soil and air temperatures is not too big, it is possible to simulate  $L_{\uparrow}$  using the air temperature (Park *et al.*, 2008). In our approach three different types of temperature were used to simulate  $L_{\uparrow}$ , specifically: surface soil temperature (where available), air temperature at 2 m height, and soil temperature at 4 cm depth.

The LWRB package (see flowchart in Figure 6.1) is part of the JGrass-NewAge system and was preliminary tested in Formetta *et al.* (2014b). Model inputs depend on the specific SM being implemented and the purpose of the run being performed (calibration, verification, simulation). The inputs are meteorological observations such as air temperature, relative humidity, incoming solar radiation, and sky clearness index. The LWRB is also fed by other JGrass-NewAGE components, such as the shortwave radiation balance (SWRB) (Formetta *et al.*, 2013b). To test model performances (i.e. verification), the LWRB can be connected to the system's Verification component; to execute the parameter calibration algorithm (Formetta *et al.*, 2014a), it can be connected to the LUCA (Let Us CALibrate) component. In

turn, all these components can and/or need to be connected to other ones, as the problem under examination may require. Model outputs are  $L_{\downarrow}$  and  $L_{\uparrow}$ . These can be provided in single points of specified coordinates or over a whole geographic area, represented as a raster map. For the latter case a digital elevation model (DEM) of the study area is necessary in input.

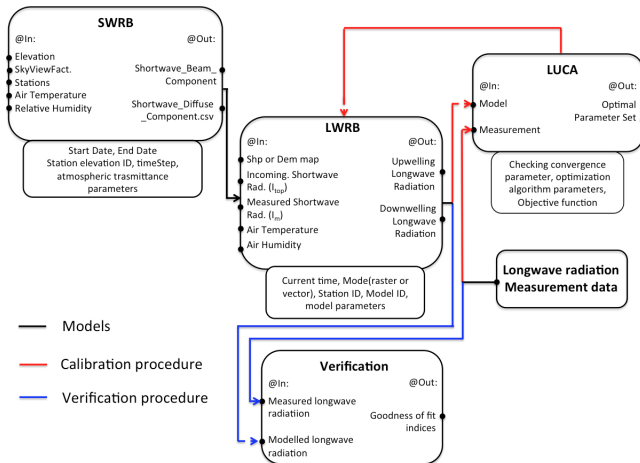


Figure 6.1: The LWRB component of JGrass-NewAge and the flowchart to model longwave radiation.

The subsection 6.1.1.1 present the calibration and the verification procedure. Moreover a model sensitivity analysis procedure is presented in subsection 6.1.1.2 and a multi-regression model to relate optimal parameter set and easy available meteorological data is proposed in subsection 6.1.1.3.

### 6.1.1.1 Calibration of $L_{\downarrow}$ longwave radiation models

Model calibration estimates the site-specific parameters of  $L_{\downarrow}$  models by tweaking them with a specific algorithm in order to best fit measured data. To this end, we use the LUCA calibration algorithm proposed in Hay *et al.* (2006a), which is a part of the OMS core and is able to optimize parameters of any OMS component. See Appendix A.1 for further information.

The calibration procedure for  $L_{\downarrow}$  follows these steps:

- The theoretical solar radiation at the top of the atmosphere ( $I_{top}$ ) is computed using the SWRB (see Figure 1);
- The clearness index,  $c$ , is calculated as the ratio between the measured incoming solar radiation ( $I_m$ ) and  $I_{top}$ ;
- Clear-sky and cloud-cover hours are detected by a threshold on the clearness index (equal to 0.6), providing two subsets of measured  $L_{\downarrow}$ , which are  $L_{\downarrow clear}$  and  $L_{\downarrow cloud}$ . On one side, a threshold of 0.6 to define the clear-sky conditions helps in the sense that allow to define time-series of measured clear-sky  $L_{\downarrow}$  with comparable length in all the stations, and this is useful for a reliable calibration process. On the other side, it introduces a small error in computing the emissivity in all-sky condition using equation (3). Although the effects of this small error would need further investigations, they could be compensated by the optimization of the parameters  $a$  and  $b$ , that are non-linearly related to the emissivity in all-sky conditions;
- The parameters  $X$ ,  $Y$ , and  $Z$  for the models in Table 6.1 are optimized using the subset  $L_{\downarrow clear}$  and setting  $a=0$  in eq. 6.3;
- The parameters  $a$  and  $b$  for eq. 6.3 are optimized using the subset  $L_{\downarrow cloud}$  and using the  $X$ ,  $Y$ , and  $Z$  values computed in the previous step.

The calibration procedure provides the optimal set of parameters at a given location for each of the ten models. As well as parameter calibration, we carry out a model parameter sensitivity analysis and we provide a linear regression model relating a set of site-specific optimal parameters with mean air temperature, relative humidity, precipitation, and altitude.

As presented in previous applications (e.g., Hatfield *et al.*, 1983; Flerchinger *et al.*, 2009), we use the SMs with the original coefficients from literature (i.e. the parameters of Table 6.2) and compare the performances of the models against available measurements of  $L_{\downarrow}$  and  $L_{\uparrow}$  for each site. The goodness of fit is evaluated by using two goodness-of-fit estimators: the Kling-Gupta Efficiency (KGE) and the root mean square error (RMSE). See Appendix A.3 for further details.

#### **6.1.1.2 Sensitivity analysis of $L_{\downarrow}$ models**

For each  $L_{\downarrow}$  model we carry out a model parameters sensitivity analysis to investigate the effects and significance of parameters on performance for different model structures (i.e. models with one, two, and three parameters). The analyses are structured according to the following steps:

- we start with the optimal parameter set, computed by the optimization process for the selected model;
- all parameters are kept constant and equal to the optimal parameter set, except for the parameter under analysis;
- 1000 random values of the analyzed parameter are picked from a uniform distribution centered on the optimal value with width equal to  $\pm 30\%$  of the optimal value; in this way 1000 model parameter sets were defined and 1000 model runs were performed;
- 1000 values of KGE are computed by comparing the model outputs with measured time series.

The procedure was repeated for each parameter of each model and for each station of the analyzed dataset.

### 6.1.1.3 Regression model for parameters of $L_{\downarrow}$ models

The calibration procedure previously presented to estimate the site specific parameters for  $L_{\downarrow}$  models requires measured downwelling longwave data. Because these measurements are rarely available, we implement a straightforward multivariate linear regression (Chambers & Hastie, n.d.; Wilkinson & Rogers, 1973) to relate the site-specific parameters X, Y and Z to a set of easily available site specific climatic variables, used as regressors  $r_i$ . To perform the regression we use the open-source R software (<https://cran.r-project.org>) and to select the best regressors we use algorithms known as "best subsets regression", which are available in all common statistical software packages. The regressors we have selected are: mean annual air temperature, relative humidity, precipitation, and altitude. The models that we use for the three parameters are presented in equations (6.4), (6.5), and (6.6):

$$X = i_X + \sum_{k=1}^N \alpha_k \cdot r_k + \epsilon_X \quad (6.4)$$

$$Y = i_Y + \sum_{k=1}^N \beta_k \cdot r_k + \epsilon_Y \quad (6.5)$$

$$Z = i_Z + \sum_{k=1}^N \gamma_k \cdot r_k + \epsilon_Z \quad (6.6)$$

where  $N=4$  is the number of regressors (annual mean air temperature, relative humidity, precipitation, and altitude);  $r_k$  with  $k=1, \dots, 4$  are the regressors;  $i_X$ ,  $i_Y$ , and  $i_Z$  are the intercepts;  $\alpha_k$ ,  $\beta_k$ , and  $\gamma_k$  are the coefficients; and  $\epsilon_X$ ,  $\epsilon_Y$ , and  $\epsilon_Z$  are the normally distributed errors. Once the regression parameters are determined, the end-user can estimate site specific X, Y and Z parameter values for any location by simply substituting the values of the regressors in the model formulations.



### 6.1.2 The study area: the AmeriFlux Network

To test and calibrate the LWRB SMs we use 24 meteorological stations of the AmeriFlux Network (<http://ameriflux.ornl.gov>). AmeriFlux is a network of sites that measure water, energy, and CO<sub>2</sub> ecosystem fluxes in North and South America. The dataset is well-known and used in several applications such as Xiao *et al.* (2010), Barr *et al.* (2012), and Kelliher *et al.* (2004). Data used in this study are the Level 2, 30-minute average data. Complete descriptions and downloads are available at the Web interface located at <http://public.ornl.gov/ameriflux/>.

We have chosen 24 sites that are representative of most of the contiguous USA and span a wide climatic range: going from the arid climate of Arizona, where the average air temperature is 16 °C and the annual precipitation is 350 mm, to the equatorial climate of Florida, where the average air temperature is 24 °C and the annual precipitation is 950 mm. Some general and climatic characteristics for each site are summarized in Table 6.4, while Figure 6.2 shows their locations. The 30-minute average data have been cumulated to obtain continuous time series of averaged, hourly data for longwave radiation, air and soil temperature, relative humidity, precipitation, and soil water content. Longwave radiation was measured with Eppley Pyrgeometers with uncertainty of  $\pm 3$  [W m<sup>-2</sup>].

SiteID	State	Latitude	Longitude	Elevation (m)	Climate	T (°C)	Data period
1	AZ	31.908	-110.840	991	semiarid	19	2008 – 2013
2	AZ	31.591	-110.509	1469	temperate, arid	16	2002 – 2011
3	AZ	31.744	-110.052	1372	temperate, semi-arid	17	2007 – 2013
4	AZ	31.737	-109.942	1531	temperate, semi-arid	17	2004 – 2013
5	AZ	31.821	-110.866	116	subtropical	19	2004 – 2014
6	AZ	35.445	-111.772	2270	warm temperate	9	2005 – 2010
7	AZ	35.143	-111.727	2160	warm temperate	9	2005 – 2010
8	AZ	35.089	-111.762	2180	warm temperate	8	2005 – 2010
9	CA	37.677	-121.530	323	mild	16	2010 – 2012
10	CA	38.407	-120.951	129	mediterranean	15	2000 – 2012
11	FL	25.365	-81.078	0	equatorial savannah	24	2004 – 2011
12	ME	45.207	-68.725	61	temperate continental	5	1996 – 2008
13	ME	45.204	-68.740	60	temperate continental	6	1996 – 2009
14	MN	44.995	-93.186	301	continental	6	2005 – 2009
15	MN	44.714	-93.090	260	snowy, humid summer	8	2003 – 2012
16	MO	38.744	-92.200	219	temperate continental	13	2004 – 2013
17	MT	48.308	-105.102	634	continental	5	2000 – 2008
18	NJ	39.914	-74.596	30	temperate	12	2005 – 2012
19	OK	36.427	-99.420	611	cool temperate	15	2009 – 2012
20	TN	35.931	-84.332	286	temperate continental	15	2005 – 2011
21	TN	35.959	-84.287	343	temperate	14	1994 – 2007
22	TX	29.940	-97.990	232	warm temperate	20	2004 – 2012
23	WA	45.821	-121.952	371	strongly seasonal	9	1998 – 2013
24	WV	39.063	-79.421	994	temperate	7	2004 – 2010

Table 6.4: Some general and climatic characteristics of the sites used for calibration: elevation is the site elevation above sea level, T is the annual average temperature, and data period refers to the period of available measurements.

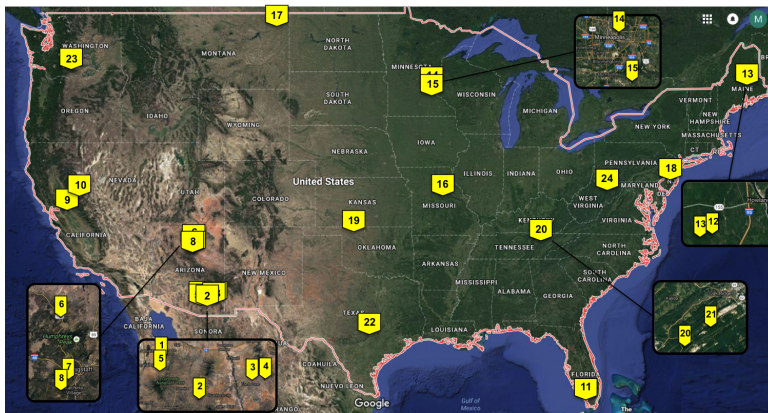


Figure 6.2: Test site locations in the United State of America.

### 6.1.3 Results

#### 6.1.3.1 Verification of $L_{\downarrow}$ models with literature parameters

When implementing the ten  $L_{\downarrow}$  SMs using the literature parameters, in many cases, they show a strong bias in reproducing measured data. A selection of representative cases is presented in Figure 6.3 which shows scatterplots for four SMs in relation to one measurement station. The black points represent the hourly estimates of  $L_{\downarrow}$  provided by literature formulations, while the solid red line represents the line of optimal predictions. Model 1 (Ångström, 1915) shows a tendency to lie below the 1:1 line, indicating a negative bias (percent bias of -9.8) and, therefore, an underestimation of  $L_{\downarrow}$ . In contrast, model 9 (Prata, 1996) shows an overestimation of  $L_{\downarrow}$  with a percent bias value of 26.3.

Figure 6.4 presents the boxplot of KGE (first column) and RMSE (second column) obtained for each model under clear-sky conditions, grouped by classes of latitude and longitude. In general all the models except the Model 8 (Konzelmann *et al.*, 1994) provided values of KGE higher than 0.5 and RMSE lower than 100 [ $\text{W m}^{-2}$ ] for all the latitude and longitude classes. Model 8 is the less performing model for many of the stations likely because the model parameters were estimated for the Greenland where snow and ice play a fundamental role on the energy balance. Its KGE values range between 0.33 and 0.62 on average, while its RMSE values are higher than 100 [ $\text{W m}^{-2}$ ] except for latitude classes  $>40^{\circ}\text{N}$  and longitude classes  $>-70^{\circ}\text{W}$ . Model 6 (Idso, 1981) and Model 2 (Brunt, 1932) provide the best results and the lower variability, independently of the latitude and longitude ranges where they are applied. Their average KGE values are between 0.75 and 0.92, while the RMSE has a maximum value of 39 [ $\text{W m}^{-2}$ ]. Moreover, all the models except

2 and 6 show a high variability of the goodness of fit through the latitude and longitude classes.

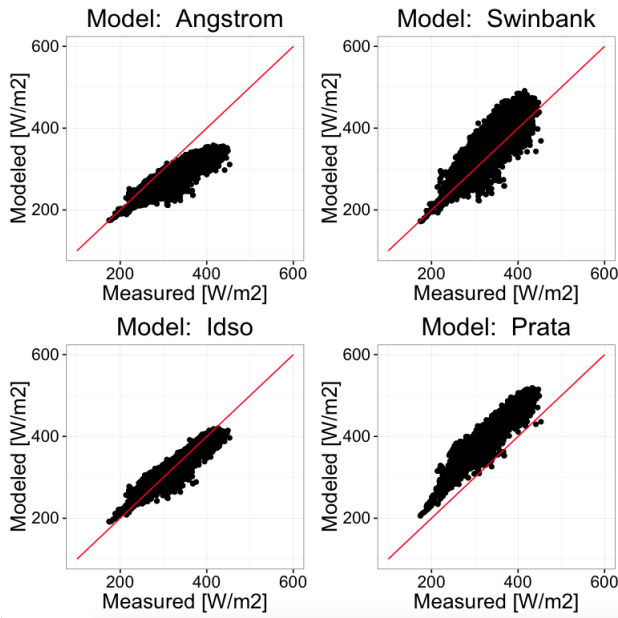


Figure 6.3: Results of the clear-sky simulation for four literature models using data from Howland Forest (Maine).

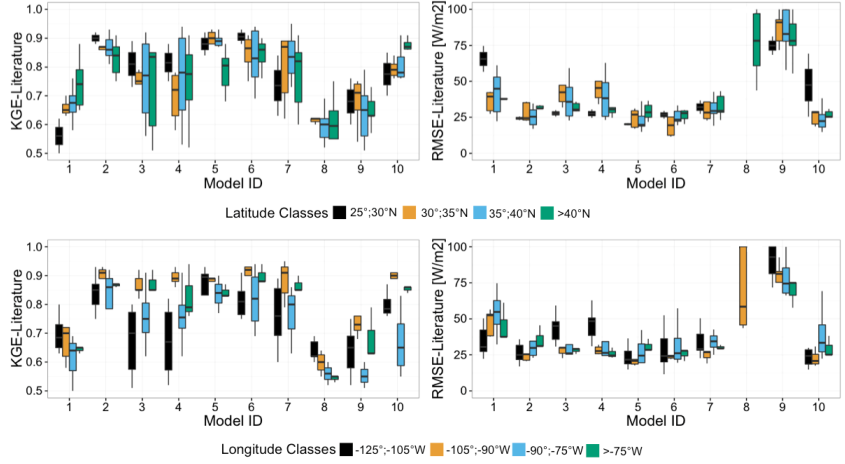


Figure 6.4: KGE and RMSE values for each clear-sky simulation using literature formulations, grouped by classes of latitude and longitude. Only values of KGE above 0.5 are shown. Only values of RMSE below 100 [W m<sup>-2</sup>] are shown.

### 6.1.3.2 $L_1$ models with site-specific parameters

The calibration procedure greatly improves the performances of all ten SMs. Optimized model parameters for each model are reported in the supplementary material (Table S1). Figure 6.5 presents the boxplots of KGE and RMSE values for clear-sky conditions grouped by classes of latitude and longitude. The percentage of KGE improvement ranges from its maximum value of 70% for Model 8 (which is not, however, representative of the mean behavior of the SMs) to less than 10% for Model 6, with an average improvement of around 35%. Even though variations in model performances with longitude and latitude classes still exist when using optimized model parameters, the magnitude of these variations is reduced with respect to the use of literature formulations. The calibration procedure reduces the RMSE values for all the models to below 45 [W m<sup>-2</sup>], even for Model 8, which also in this case had the maximum improvement. Model 6 (Idso, 1981) and Model 2 (Brunt, 1932) provide the best results

on average for all the analyzed latitude and longitude classes.

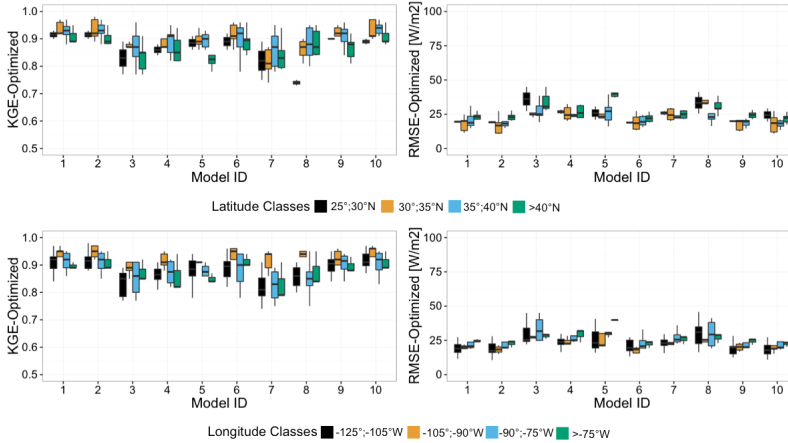


Figure 6.5: KGE (best is 1) and RMSE (best is 0) values for each optimized formulation in clear-sky conditions, grouped by classes of latitude and longitude. Only values of KGE above 0.5 are shown.

Figure 6.6 presents the boxplots of KGE and RMSE values for each model under all-sky conditions, grouped by latitude and longitude classes. In general, for all-sky conditions we observe a deterioration of KGE and RMSE values with respect to the clear-sky optimized case, with a decrease in KGE values up to a maximum of 25% on average for Model 10. This may be due to uncertainty incorporated in the formulation of the cloudy-sky correction model (eq. 6.3): it seems that sometimes the cloud effects are not accounted for appropriately. This, however, is in line with the findings of Carmona *et al.* (2014).

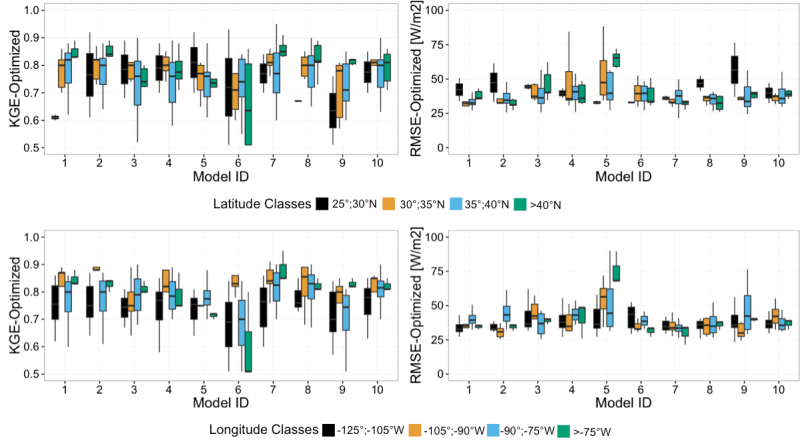


Figure 6.6: KGE and RMSE values for each model in all-sky conditions with the optimized parameters; results are grouped by classes of latitude and longitude. Only values of KGE above 0.5 are shown.

### 6.1.3.3 Sensitivity analysis of $L_1$ models

The results of the models sensitivity analysis are summarized in Figures 6.7-a and 6.7-b for models 1 to 5 and models 6 to 10, respectively. Each figure presents three columns, one for each parameter. Considering model 1 and parameter X: the range of X is subdivided into ten equal-sized classes and for each class the corresponding KGE values are presented as a boxplot. A smooth blue line passing through the boxplot medians is added to highlight any possible pattern to parameter sensitivity. A flat line indicates that the model is not sensitive to parameter variation around optimal value. Results suggest that models with one and two parameters are all sensitive to parameter variation, presenting a peak in KGE in correspondence with their optimal values; this is more evident in models with two parameters. Models with three parameters tend to have at least one insensitive parameter, except for Model 1, that could reveal a possible over-parameterization of the modeling process.

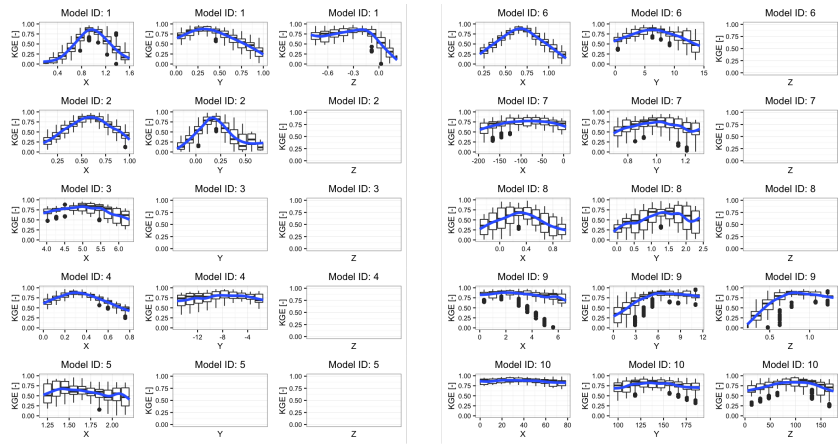


Figure 6.7: Results of the model parameters sensitivity analysis. It presents as boxplot the variation of the model performances due to a variation of one of the optimal parameter and assuming constant the others. The procedure is repeated for each model and the blue line represents the smooth line passing through the boxplot medians.

### 6.1.3.4 Regression model for parameters of $L_1$ models

A multivariate linear regression model was estimated to relate the site-specific parameters  $X$ ,  $Y$  and  $Z$  to mean annual air temperature, relative humidity, precipitation, and altitude. The script containing the regression model is available, as specified in Reproducible Research section below.

The performances of the  $L_1$  models using parameters assessed by linear regression are evaluated through the leave-one-out cross validation (Efron & Efron, 1982). We use 23 stations as training-sets for equations (6.4), (6.5), and (6.6) and we perform the model verification on the remaining station. The procedure is repeated for each of the 24 stations.

The cross validation results for all  $L_1$  models and for all stations are presented in Figures (6.8) and (6.9), grouped by classes of latitude and longitude, respectively. They report the KGE comparison between the  $L_1$  models with their original



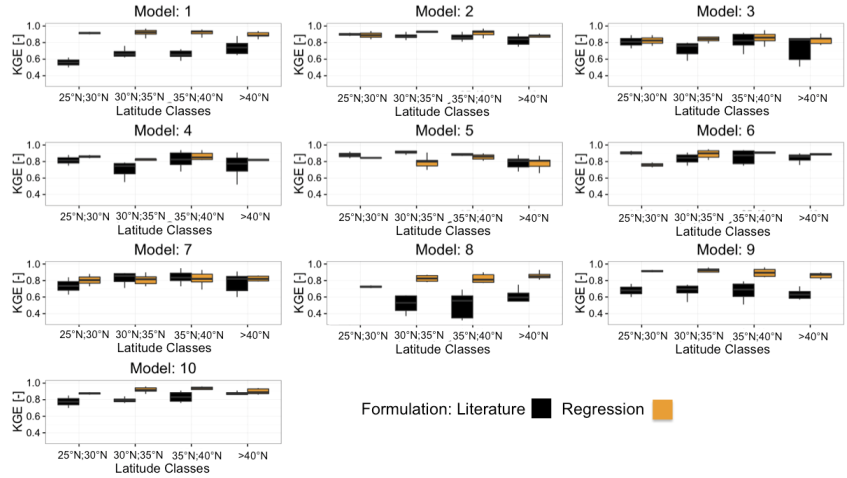


Figure 6.8: Comparison between model performances obtained with regression and classic parameters: the KGE values shown are those above 0.3 and results are grouped by latitude classes.

parameters (in black) and with the regression model parameters (in black).

In general, the use of parameters estimated with regression model gives a good estimation of  $L_1$ , with KGE values of up to 0.92. With respect to the classic formulation, model performance with regression parameters improved for all the models independently of the latitude and longitude classes. In particular for Model 8 the KGE improved from 0.26 for the classic formulation to of 0.92, on average. Finally, the use of the parameters estimated by the regression model provides a reduction of the model performances variability for all the models except Model 5 and 8, for longitude class  $-125;-105^\circ\text{W}$  and  $-105;-90^\circ\text{W}$  respectively.

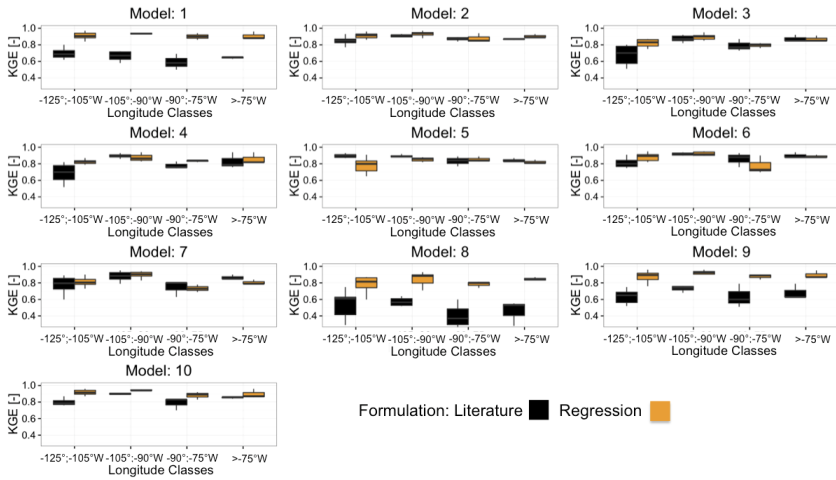


Figure 6.9: Comparison between model performances obtained with regression and classic parameters: the KGE values shown are those above 0.3 and results are grouped by longitude classes.

### 6.1.3.5 Verification of the $L_1$ model

Figure 6.10 presents the results of the  $L_1$  simulations obtained using the three different temperatures available at experimental sites: soil surface temperature (skin temperature), air temperature, and soil temperature (measured at 4 cm below the surface). The figure shows the performances of the  $L_1$  model for the three different temperatures used in terms of KGE, grouping all the stations for the whole simulation period according to season. This highlights the different behaviors of the model for periods where the differences in the three temperatures are larger (winter) or negligible (summer). The values of soil emissivity are assigned according the soil surface type, according to Table 6.4 (Brutsaert, 2005). Although many studies investigated the influence of snow covered area on longwave energy balance (e.g., Plüss & Ohmura, 1997; Sicart *et al.*, 2006), the SMs do not explicitly take into account of it. As presented in König-Langlo & Augstein (1994), the effect of snow could be

implicitly taken into account by tuning the emissivity parameter. The best fit between measured and simulated  $L_{\uparrow}$  is obtained with the surface soil temperature, with an all-season average KGE of 0.80. Unfortunately, the soil surface temperature is not an easily available measurement. In fact, it is available only for 8 sites of the 24 in the study area. Very good results are also obtained using the air temperature, where the all-season average KGE is around 0.76. The results using air temperature present much more variance compared to those obtained with the soil surface temperature. However, air temperature (at 2 m height) is readily available measure, in fact it is available for all 24 sites. The use soil temperature at 4 cm depth provides the least accurate results for our simulations, with an all-season average KGE of 0.46. In particular, the use of soil temperature at 4 cm depth during the winter is not able to capture the dynamics of  $L_{\uparrow}$ . It does, however, show a better fit during the other seasons. This could be because during the winter there is a substantial difference between the soil and skin temperatures, as also suggested in Park *et al.* (2008).

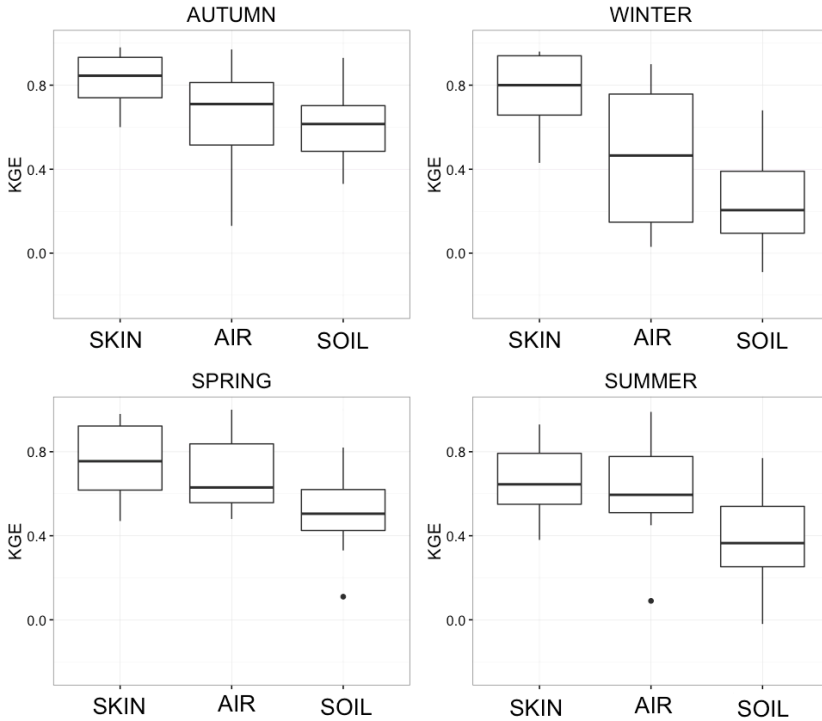


Figure 6.10: Boxplots of the KGE values obtained by comparing modeled upwelling longwave radiation, computed with different temperatures (soil surface temperature (SKIN), air temperature (AIR), and soil temperature (SOIL)), against measured data. Results are grouped by seasons.

#### 6.1.4 Final remarks

The main achievements of this work include: i) a broad assessment of the classic  $L_{\downarrow}$  longwave radiation parameterizations, which clearly shows that the Idso (1981) and Brunt (1932) models are the more robust and reliable for all the test sites, confirming previous results (Carmona *et al.*, 2014); ii) a site specific assessment of the  $L_{\downarrow}$  longwave radiation model parameters for 24 AmeriFlux sites that improved the performances of all the models; iii) the set up of a regression model that provides an estimate of optimal parameter sets on the basis climatic data; iv) an assessment of  $L_{\downarrow}$  model

performances for different temperatures (skin temperature, air temperature, and soil temperature at 4 cm below surface), which shows that the skin and the air temperature are better proxy for the  $L_{\uparrow}$  longwave radiation. Regarding longwave downwelling radiation the Brunt (1932) model is able to provide on average the best performances with the regression model parameters independently of the latitude and longitude classes. For the Idso (1981) model the formulation with regression parameter provided lower performances with respect to the literature formulation for latitude between 25°N and 30°N.

The integration of the package into JGrass-NewAge will allow users to build complex modeling solutions for various hydrological scopes. In fact, future work will include the link of the LWRB package to the existing components of JGrass-NewAge to investigate  $L_{\downarrow}$  and  $L_{\uparrow}$  effects on evapotranspiration, snow melting, and glacier evolution. Finally, the methodology proposed provides the basis for further developments such as the possibility to: i) investigate the effect of different all-sky emissivity formulation and quantify the influence of the clearness index threshold ii) verify the usefulness of the regression models for climates outside the contiguous USA; iii) analyze in a systematic way the uncertainty due to the quality of meteorological input data on the longwave radiation balance in scarce instrumented areas.

For the present paper we used code version v.0.9. Versions till 0.94 are also available on the repository. Researchers interested in replicating or extending our results are invited to download our codes at:

<https://github.com/geoframecomponents/LongWaveRadiationBalance>.

Instructions for using the code can be found at:

<http://geoframe.blogspot.co.uk/2016/04/>

`lwrp-component-latest-documentation.html`.

Regression of parameters were performed in R and are available at

[https://github.com/GEOframeOMSProjects/OMS\\_Project\\_LWRB/blob/master/docs/Regression.R](https://github.com/GEOframeOMSProjects/OMS_Project_LWRB/blob/master/docs/Regression.R).

## 6.2 Kriging

Meteorological forcing data such as rainfall, temperature, solar radiation and others are the dominant controlling factors of the hydrological cycle, energy balance and ecosystem processes (Ly *et al.*, 2013). They are the natural input to distributed and semi-distributed hydrological models and their quality and precision affect the accuracy of results.

Several algorithms for the spatial interpolation of meteorological data are available in literature: Thiessen polygons (e.g., Thiessen, 1911; WMO, 1994), inverse distance methods (Ly *et al.*, 2013), interpolation with splines (e.g., Hutchinson, 1995; Mitášová & Mitáš, 1993), Kriging (e.g., Matheron, 1981; Goovaerts, 1997) or other types of interpolation (e.g., Robeson, 1992; Li & Heap, 2011, and references therein).

The evaluation of these algorithms was assessed by several authors, (Tabios & Salas, 1985; Jarvis & Stuart, 2001), concluding that Krigings are one of the best techniques for the interpolation of the spatial behavior of climatological variables. For monthly rainfall and storm totals, Tabios & Salas (1985), Creutin & Obled (1982), shown that it is preferable to other rainfall interpolation methods. Goovaerts (2000), Lloyd (2005), Basistha *et al.* (2008), Ly *et al.* (2011), confirmed these results.

Krigings can be applied to a wide range of datasets (e.g., Stahl *et al.*, 2006; Phillips *et al.*, 1992), allowing the estimation of the variance of interpolated quantities (e.g., Li & Heap, 2011).

Auxiliary variables can be used to improve the interpolation, such as terrain-related parameters (e.g., relief, slope and aspect) as investigated in Attorre *et al.* (2007). Not surprisingly, Carrera-Hernández & Gaskin (2007) found that the use of elevation as a secondary variable improves temperature prediction.

However, the interpolation could be computationally more demanding than other techniques, since it requires the fitting of the theoretical semivariogram against the experimental for each or group of time-steps.

To overcome this problem, most of applications that implement Kriging interpolators either use long time series and long time-step, such as daily, (Verfaillie *et al.* , 2006; Buytaert *et al.* , 2006), monthly or yearly (Hevesi *et al.* , 1992; Goovaerts, 2000; Boer *et al.* , 2001; Todini, 2001), or short time series and shorter time steps (such as rainfall events) [e.g., Haberlandt (2007)] and keeps the semivariogram as static (fixed during the whole simulations). Furthermore, between the geostatistical tools available to the scientific community, few are open-source (i.e., SAGA GIS kriging, R gstat, High Performance Geostatistics Library HPGL), but none implements a quick way to plug-in to hydrological model, and are easy connectible to automatic calibration algorithms.

The work presents a Spatial Interpolation (SI) package which makes hourly (or sub-hourly, when it is reasonable) estimate of any spatially distributed environmental data. SI implements 10 theoretical semivariogram models and 4 types of Kriging algorithms, for a total of forty interpolation options.

It is designed according to the Object Modeling System (OMS) framework (David *et al.* , 2013) to be compatible with the JGrass-NewAGE system, (Formetta *et al.* , 2014a). In particular, in this work, four components are used to obtain the optimization

of the parameters of the theoretical semivariogram and perform automatically and easily a jackknife to assess the error of estimates.

The work is organized as follows: first the theory of the Kriging is introduced in section 6.2.1. Section 6.2.3 describes the study area and the experimental setup. The results of the application of the SI package on temperature and rainfall datasets and a comparison of results with the R *gstat* are discussed in section 6.2.4.

### 6.2.1 A little of Kriging theory

Kriging is a group of geostatistical techniques used to interpolate the value of random fields based on spatial autocorrelation of measured data. The measurements value  $z(\vec{x}_\alpha)$  and the unknown value  $z(\vec{x})$ , where  $\vec{x}$  is the location, given according to a certain cartographic projection, are considered as particular realizations of random variables  $Z(\vec{x}_\alpha)$  and  $Z(\vec{x})$  (Goovaerts, 1997; Isaaks & Srivastava, 1989). The estimation of the unknown value  $z^\lambda(\vec{x})$ , where the true unknown value is  $Z^\lambda(\vec{x})$ , is obtained as a linear combination of the  $N$  values at surrounding points, Goovaerts (1999):

$$Z^\lambda(\vec{x}) - m(\vec{x}) = \sum_{\alpha=1}^N \lambda(\vec{x}_\alpha) [Z(u_\alpha) - m(\vec{x}_\alpha)] \quad (6.7)$$

where  $m(\vec{x})$  and  $m(\vec{x}_\alpha)$  are the expected values of the random variables  $Z(\vec{x})$  and  $Z(\vec{x}_\alpha)$ .  $\lambda(\vec{x}_\alpha)$  is the weight assigned to datum  $z(\vec{x}_\alpha)$ . Weights are chosen to satisfy the conditions of minimizing the error of variance of the estimator  $\sigma_\lambda^2$ , that is:

$$\underset{\lambda}{\operatorname{argmin}} \sigma_\lambda^2 \equiv \underset{\lambda}{\operatorname{argmin}} \operatorname{Var}\{Z^\lambda(\vec{x}) - Z(\vec{x})\} \quad (6.8)$$

under the constraint that the estimate is unbiased, i.e.

$$E\{Z^\lambda(\vec{x}) - Z(\vec{x})\} = 0 \quad (6.9)$$



The latter condition, implies that:

$$\sum_{\alpha=1}^N \lambda_{\alpha}(\vec{x}_{\alpha}) = 1 \quad (6.10)$$

As shown in various textbooks, e.g. Kitanidis (1997), the above conditions bring to a linear system whose unknown is the tuple of weights, and the system matrix depends on the semivariograms among the couples of the known sites.

When it is made the assumption of isotropy of the spatial statistics of the quantity analyzed, the semivariogram is given by Cressie & Cassie (1993):

$$\gamma(h) := \frac{1}{2N_h} \sum_{i=1}^{N_h} (Z(\vec{x}) - Z(\vec{x}_i))^2 \quad (6.11)$$

where the distance  $(\vec{x}, \vec{x}_i) \equiv h$ ,  $N_h$  denotes the set of pairs of observations at location  $x$  and at locations  $\vec{x}_i$  at distance  $h$  apart from  $\vec{x}$ . In order to be extended to any distance, experimental semivariogram need to be fitted to a theoretical semivariogram model. Examples of theoretical semivariogram models are detailed in Appendix A.6.

These theoretical semivariograms contain parameters (called nugget, sill and range) to be fitted against the existing data, before being introduced in the Kriging linear system, whose solution returns the weights in eq (6.7).

Three main variants of Kriging can be distinguished, (Goovaerts, 1997):

- Simple Kriging (SK), which considers the mean,  $m(\vec{x})$ , to be known and constant throughout the study area;
- Ordinary Kriging (OK) , which account for local fluctuations of the mean, limiting the stationarity to the local neighborhood. In this case, the mean is unknown.

- Kriging with a trend model (here Detrended Kriging, DK), which considers that the known local mean  $m(\vec{x}_\alpha)$  varies within the local neighborhood.

The trend can be, for example, a linear regression model between the investigated variables and a auxiliary variable, such as elevation or slope. According to Goovaerts (1997), the trend should be subtracted from the original data and the OK of the residuals performed. The final interpolated values will be the sum of the interpolated values and the previously estimated trend.

Variants of OK and DK are the local ordinary kriging (LOK) and local detrended Kriging (LDK). In this case the estimate is only influenced by the measurements belonging to a neighbor.

The SI package implements the OK and the DK, since local mean may vary significantly over the study area and the SK assumption of the known stationary mean could be too strict, (Goovaerts, 1997). Moreover, Goovaerts (2000) found that, in the case of trend, detrended kriging provides better results than coKriging and it is not as computationally demanding. In SI package, the neighbor stations in the local case are defined either in a maximum searching radius or as a number of stations closer to the interpolation point.

To estimate the errors produced by the interpolation using Kriging techniques, we chose the leave-one-out cross validation technique (Efron & Efron, 1982). In fact, Goovaerts (1997) states that the standard deviation cannot be used as a direct measures of estimation precision. Moreover the procedure allows to evaluate the impact of the different models on interpolation results, (Isaaks & Srivastava, 1989; Goovaerts, 1997; Prudhomme & Reed, 1999; Martin & Simpson, 2003; Aidoo *et al.* , 2015).

Leave-one-out cross validation consists of removing one data point at a time and performing the interpolation for the location

of the removed point, using the remaining stations. The approach is repeated until every sample has been, in turn, removed and estimates are calculated for each point. Interpolated and measured values for each station were compared and the goodness of fit indexes, such as Root mean square error (RMSE) and Nash-Sutcliffe Efficiency (NSE), were calculated to assess model performances. See Appendix A.3 for further details about the goodness of fit indexes.

### 6.2.2 Structure of the SI package

Key aspects of the Krigings techniques, from the computationally point of view, are the estimation of the theoretical semivariogram parameters, which requires a best fitting techniques, and, subsequently their introduction in a theoretical model and the linear system which serves to obtain the weights  $\lambda_\alpha$ . The SI package has been structured in 4 components (ExperimentalVariogram, TheoreticalVariogram, Krigings and LeaveOneOut), shown in diagrams 6.11 and 6.12. In the figures, each block represents a component and the blue arrows represent the connections between them (the data they share). The components connected as in diagram 6.11 or in diagram 6.12 represent a Modeling Solution (MS), which performs a modeling task (i.e., the optimization of the model parameters).

Inputs and outputs of SI package are time series of measured values and geographic objects, such as *shapefiles* and *raster*, containing geographical information on the measurements stations and on interpolation points. These data are managed with the Geotools library, (Turton, 2008), as in Formetta *et al.* (2014a) and are easily managed and visualized by most of the GIS systems.

Figure 6.11 shows the MS obtained linking the experimental

variogram component to the theoretical variogram component and to the optimization algorithm. This MS optimizes the semivariogram parameters, sill, nugget and range and feeds the Kriging component. The inputs of the experimental variogram are the time series of the measured variables and the shapefile with the spatial coordinates of the stations. The outputs are the experimental semivariogram values and the distance vector, which feed the theoretical semivariogram component together with the name of the model chosen and the sill, nugget and range. Particle swarm is the one that actually optimizes the theoretical model and gives in output the sill, nugget and range for the Kriging component. Further inputs of the calibrator are the objective function to be optimized (in this case the RMSE) and other internal parameters, such as the number of iteration and the tolerance.

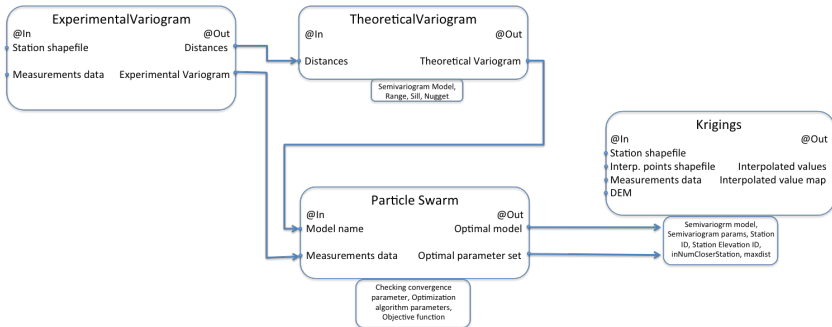


Figure 6.11: The MS optimizes model parameters connecting of the experimental semivariogram component to the theoretical semivariogram to the optimization tool. The output of the optimization are the sill, nugget and range, which, together with the type of model, are the input of the Kriging component. Outputs of the MS can be time-series and maps of interpolated variables.

The optimized parameters, together with the dataset of measurements, the shapefile of the stations and the shapefile with the interpolation points are the inputs of the Krigings component. Data in input to be spatialized can vary and being,

for instance, temperature, rainfall (as in Abera *et al.* (2017a)), parameters ancillary of other models, for instance those for radiation (as in Formetta *et al.* (2013b)), where previous version of these components were used. The outputs are either time series or map of interpolated values.

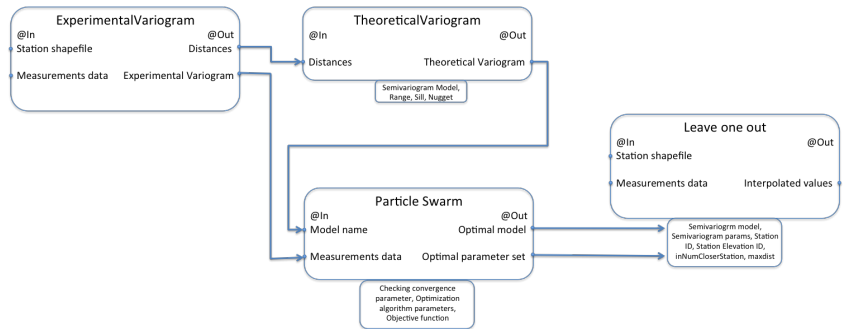


Figure 6.12: The MS, respect to the one presented in figure 6.11, connects the particle swarm to the automatic leave-one-out for the assessment of the model performances.

Figure 6.12 show the implementation of the iterative procedure necessary to estimate errors of interpolation, called leave-one-out (or jackknife). Given  $n$  spatially distributed measures,  $n - 1$  measures are used for the interpolation, while the remaining one is used for comparison. The operation is repeated  $n$  times excluding each time a different measure location, so to obtain a set of estimated error as large as the number of sample sites. Because our package is dealing with time-varying fields, the operation is repeated for each time step where the measure are available and the site error is actually a temporal mean.

### 6.2.3 Study area

To test the performances of the modeling solutions, we used the SI package to interpolate temperature and rainfall data from 97 stations, in the Isarco River valley, Italy, shown in Figure 6.13 and

detailed in Appendix A.7. Isarco River is a left tributary of the Adige River, in Trentino-Alto Adige Region, northern Italy.

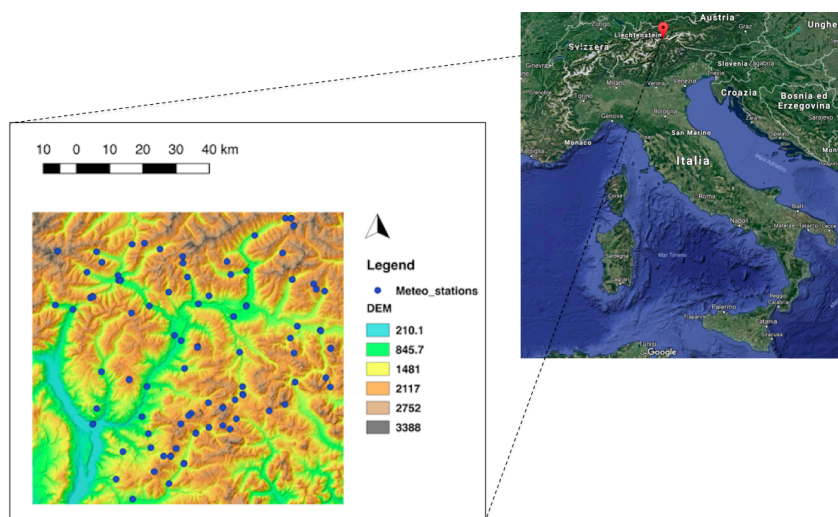


Figure 6.13: Study area: Isarco River Valley is situated in the North-East part of Italy and it is one of the main valley in the Alto-Adige region.

The catchment is around  $4200 \text{ km}^2$ , the river length is around 95 km and the altitude spans from 210 m a.s.l. to 3400 m a.s.l. Climate is typically alpine and characterized by dry winters, snow and glacier-melt in spring, and humid summers and autumns. Data used in the testing were provided by Provincia Autonoma di Bolzano, and collected into the Adige database (<http://abouthydrology.blogspot.it/2016/09/the-adige-database-or-database-newage.html>) during the CLIMAWARE and GLOBAQUA projects.

To test the SI package we used the data from the year 2008, since, in the available dataset (2003-2013) was the year with the smallest number of missing data.

A quality check was also made, in order to eliminate the out-layers from the dataset. Moreover, the spatial distribution of the no-value was analyzed, in order to asses the number of bins

of distances in which compute the semivariance. For each time step, we found that around the 10% of stations were not recording data. Therefore, since the mean number of active stations for each time step was 70-80, we decided to consider 8 bins. This choice was also supported by a visual inspection of the experimental semivariance shape, which confirmed that using 8 bins, the number of stations involved were not too low or too high.

In order to assess the goodness of the package performances, two applications were performed:

- an interpolation of one year of hourly temperature data;
- an interpolation of a hourly rainfall event.

Firstly, the analysis of the semivariance was performed and experimental semivariograms were fitted using the 10 theoretical models. The models that gave the best fitting were then used for the interpolation of the two variables using the 4 types of Krigings. Thus, Krigings performances were assessed using the leave-one-out cross validation. Finally, results obtained from the interpolation of the temperature dataset were compared to the results obtained with R *gstat*, in order to assess the differences between the two packages, their easiness of use and their performances.

## **6.2.4 Results**

### **6.2.4.1 Application of SI on temperature dataset**

The first application of SI package was made using the temperature dataset. The hourly semivariograms were computed and then fitted using the 10 available theoretical models.

Figure 6.14 shows the results of the fitting of the experimental semivariogram for a single time step, 15th June 2008. The black

dots represent the experimental semivariance, while each colored curve represents a different optimized theoretical model.

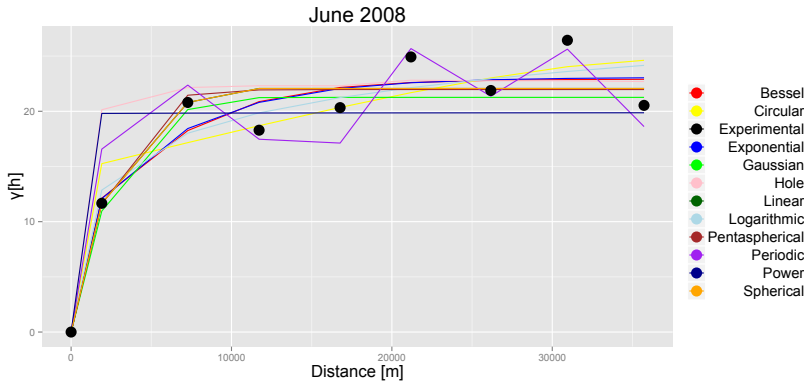


Figure 6.14: Fitting of the experimental semivariogram for the 15th June 2008 12:00. The 10 theoretical semivariogram models were optimized using the PSO.

Table 6.5 reports the main indices of goodness, NSE, RMSE,  $R^2$  and PBIAS, computed between the experimental semivariogram and the 10 theoretical semivariogram models. All the models gave satisfactory results and, therefore, we chose to use the best 5, Bessel, Exponential, Gaussian, Linear and Spherical for the interpolation of the temperature dataset.

In order to assess the goodness of the 4 types of Krigings, OK, LOK, DK, LDK, we performed the leave one out cross validation using the optimized hourly values of sill, nugget and range.

Figures 6.15 show the results in terms of NSE between the measured and interpolated values of temperature using the four types of Krigings, OK, LOK, DK, LDK and the five semivariogram models. Each point represents the averaged monthly NSE over the 97 meteorological stations. The two local cases were performed using the ten closest stations to the interpolation point.



Semivariogram model	NSE	RMSE	$R^2$	PBIAS
Bessel	0.92	2.14	0.92	-0.20
Circular	0.88	2.59	0.88	0.0
Exponential	0.92	2.10	0.92	-3.80
Gaussian	0.90	2.39	0.91	0.35
Hole	0.77	3.61	0.81	7.90
Linear	0.91	2.28	0.91	0.0
Logarithmic	0.92	2.17	0.92	0.0
Pentaspherical	0.91	2.29	0.91	0.0
Periodic	0.90	2.18	0.92	0.0
Power	0.72	3.99	0.73	-3.70
Spherical	0.91	2.28	0.91	0.0

Table 6.5: Results in terms of goodness of fit indices of the fitting between the experimental and the 10 theoretical semivariograms. All the models shown a good agreement: Bessel model proved to be the best while the Power the worst.

Since temperatures in this dataset present a strong trend with the elevation ( $R^2 \sim 0.9$ ), results obtained using the OK and LOK are worse than the ones obtained with DK, LDK, which present higher values of NSE, till 0.98.

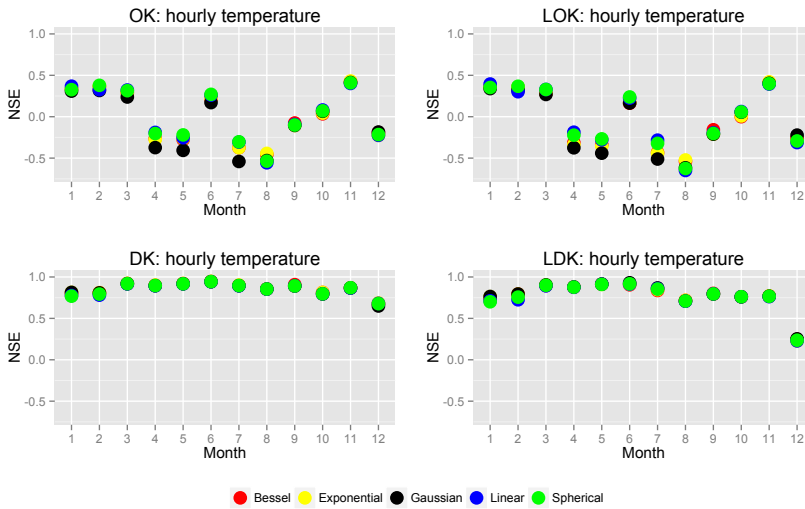


Figure 6.15: Monthly variation of the NSE index: each dot is the averaged NSE over the entire dataset. The theoretical semivariogram models shown are: Bessel (red dots), Exponential (yellow dots), Gaussian (black dots), Linear (blue dots), Spherical (green dots).

The SI package works both using single points and raster maps, as shown in figure 6.16. The spatialization of the temperature was made for each pixel of the DEM (100 m resolution), applying the LDK and the exponential semivariogram model. The two plots in the figure show two different dates in the 2008, one during winter (15th February 2008 12:00) and one in summer (15th June 2008 12:00).

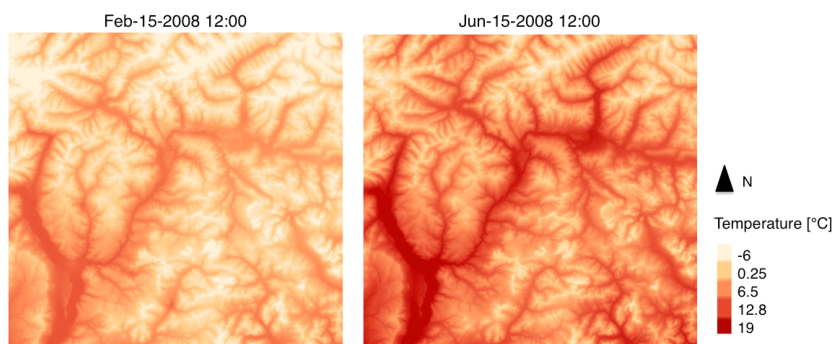


Figure 6.16: Maps of spatialized temperature for the 15th February 2008 and for the 15th June 2008

Figure 6.17 shows the bubble plots of the RMSE obtained between the measured hourly temperature in June 2008 and the interpolated with the OK and DK, overlapped to the DEM. The size of the bubble is representative of the magnitude of the error: bigger errors are obtained in the case of OK for the stations at higher elevation, which are corrected in the case of DK. The biggest error in the OK interpolation ( $RMSE = 11.95^{\circ}C$ ) is obtained for the station ID 90145 ( $Z=3399$  amasl) which is then reduced to  $1.83^{\circ}C$ .

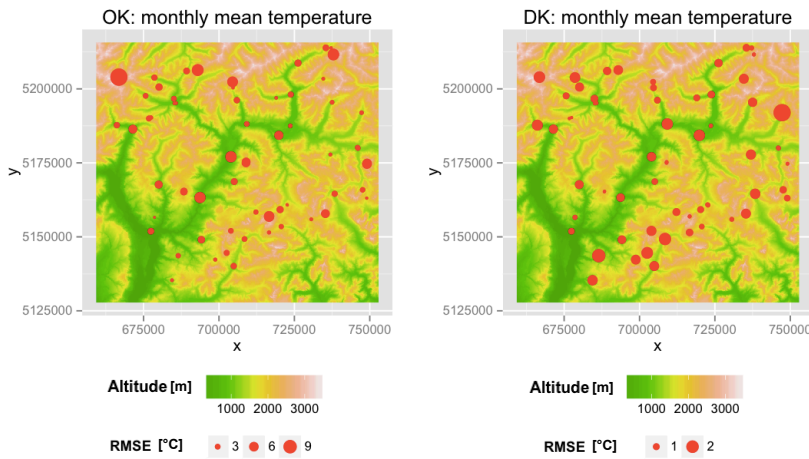


Figure 6.17: The two bubble plots show the RMSE obtained using the OK (box on the left) and DK (box on the right). The scales of the bubbles are not the same for the two plots for visualization reasons. In fact, being the errors in the DK case is much less than the OK case, a unique bubble scale didn't allow to appreciate the RMSE values in the DK case.

#### 6.2.4.2 Application of SI on rainfall dataset

The application on the rainfall dataset was made at event scale. We chose a rainfall event of 11 h between the 29th and 30th June 2008. The event was chosen since it is the longest and the most intense recorded from the highest number of stations.

Figure 6.18 shows the boxplots of the 11 hourly semivariograms with 8 bins of lag distance, while red line represents the best theoretical semivariogram optimized, which in this case was obtained using the linear model. It is clear a nugget effect (optimized nugget value of 6), probably due to error measurements. Moreover, given the shape of the optimized semivariogram, the variables are correlated at any distance.

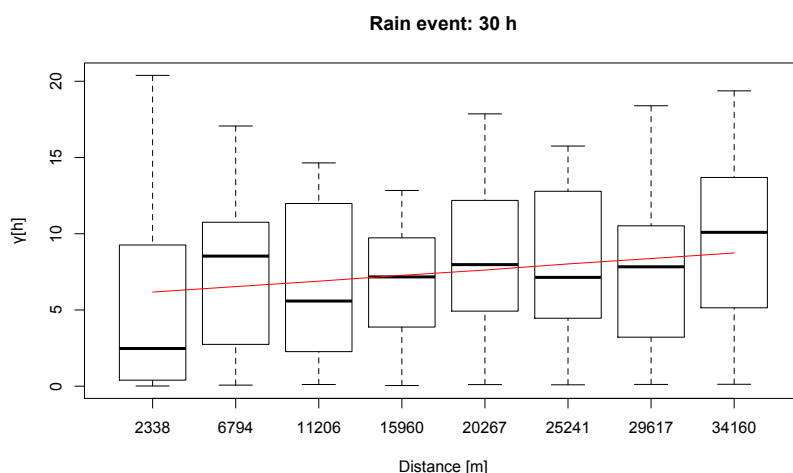


Figure 6.18: Boxplots of the semivariograms of the precipitation event of 29th and 30th June 2008: the horizontal line in the middle shows the median, the bottom and top end of the box show the 25th and 75th percentile, respectively, the whiskers (vertical line) shows the range of the data.

The optimized value of range, nugget and sill were used for the 4 types of Kriging interpolations. Figure 6.19 shows the comparison of the results obtained for two stations (ID 1152 and ID 1270), chosen at different elevation (953 m a.s.l. and 2100 m a.s.l., respectively).

Table 6.6 shows the indexes of goodness between the measured and the interpolated rainfall for the 4 types of Kriging and the two stations. The performance are overall good in the case of station ID 1152 and the best interpolator is the local ordinary kriging computed using the 5 closer stations. Results of the station ID 2170 are slightly worse in the case of OK, and worst in cases of LOK, DK, LDK, probably due to the highest elevation of the station. In this case the best interpolator is the LOK computed using the 5 closer stations.

Kriging type	ID 1152			ID 1270		
	NSE	RMSE	R <sup>2</sup>	NSE	RMSE	R <sup>2</sup>
OK	0.69	1.64	0.77	0.60	1.48	0.64
LOK	0.73	1.53	0.74	0.31	1.88	0.45
DK	0.61	1.58	0.77	0.46	1.66	0.54
LDK	0.66	1.71	0.72	0.35	1.82	0.37

Table 6.6: Results in terms of goodness of fit indices of the fitting between the experimental and the 10 theoretical semivariograms. All the models shown a good agreement: Bessel model proved to be the best while the Power the worst.

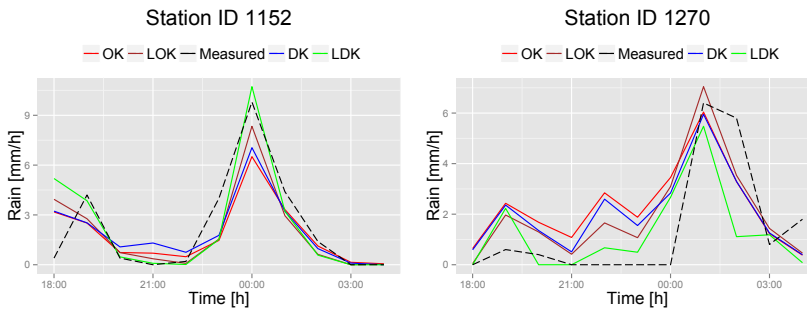


Figure 6.19: Comparison between the four types of Kriging (OK, red solid line, DK, blue solid line, LOK, brown solid line and LDK, green solid line) and the measured rainfall (black dashed line).

The spatial interpolation of the precipitation was also made for each pixel of the DEM (100 m resolution), applying the OK and the linear semivariogram model. Figure 6.20 show the results of the interpolation for the June 30th 2008 at 00:00. As it appears from the map, the rainfall intensities are higher in the river valley, with a value of 9.8 [mm/h] measured at the station ID 1152.

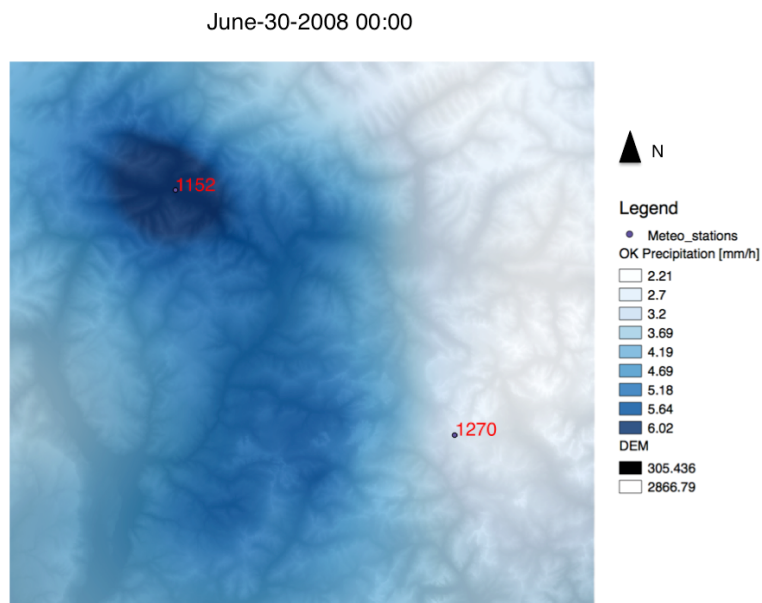


Figure 6.20: Spatial interpolation of the precipitation made for each pixel of the DEM (100 m resolution), applying the OK and the linear semivariogram model.

### 6.2.4.3 Comparison with R package *gstat*

A comparison with the R package *gstat* was made in order to highlight the differences and similarities of both packages, and to justify the introduction of an alternative software.

Regarding the features, both softwares are open-source, released under GPL v3 license and both working on all major operating systems (MS Windows, Mac OSX, Linux).

*gstat* is developed in C with a part of the code in R language and must be executed using the R various environments. SI is developed in Java (using Java 7 and older features) as a group of OMS components and can be executed inside the OMS console. Moreover, because the non-invasiveness of OMS, the SI components can be used as a stand-alone Java or embedded

with mixed-language codes.

Moving to functional differences, *gstat* computes both omnidirectional and directional semivariogram, while SI, so far, does not implement directional semivariograms. Moreover, *gstat* makes available five more theoretical semivariogram models respect to SI: Matern, Matern with Stein's parameterizations, Wave, Spline and Legendre. However, adding the desired model to the SI package is easy and straightforward, thanks to the design pattern implemented. An example of how to do it operatively, is shown in Appendix A.2

Despite C language is usually considered faster than Java, because *gstat* uses R language procedure, is generally slower than Java (few hours compared to one day for the interpolation of a year of hourly data).

Regarding the estimate that the two packages offers, they are usually different. Comparison were made either with the OK and the DK by utilizing the same temperature data used in section 6.2.3. Semivariograms were computed using the same number of bins and the same cutoff distance.

Figure 6.21 shows the results of the comparison in terms of NSE: the overall performances of both tools are good. However, SI performs always better, and sometimes significantly better.



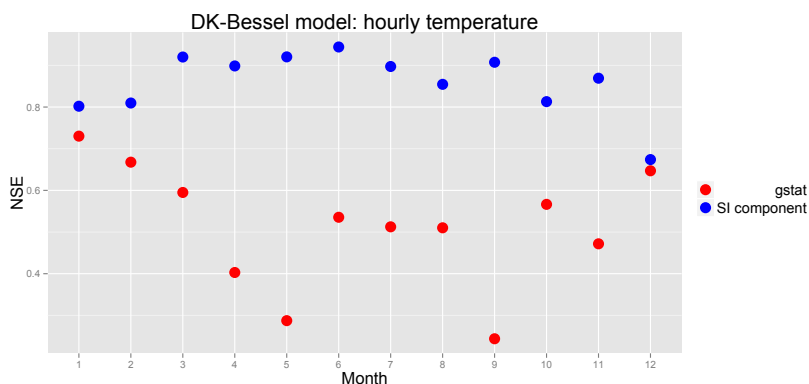


Figure 6.21: Comparison between *gstat* and SI package in the interpolation performances using the DK.

### 6.2.5 Final remarks

The goals of the work are multiple: (1) to present a geostatistical software easy to use, easy to plug-in in a hydrological model, fast and flexible (2) to show a practical example of an accurately designed software in the perspective of reproducible research, (3) to show the goodness of the software applications, in order to have a reliable alternative to other traditionally used tools.. Several characteristics make the SI package a good competitor tool among the available in literature:

- The integration in JGrass-NewAge system, which allows the building of a variety of modeling solutions, thanks to a simple connection to other components;
- the package is flexible, easily maintainable and suitable for future improvements, thanks to a design pattern for the choice of the type of the theoretical semivariogram and for the definition of the neighbor in the LOK and LDK;
- The use of git, GitHub, Gradle, Travis and Zenodo, to pursue the scope of reproducible and replicable research.

Further developments of the package are easy and straightforward, such as: the possibility of integrating new types of Krigings, a different selection of the stations, the integration of non-linear relationship between the interpolated variable and an auxiliary variable.

## CONCLUSIONS

This thesis presents a flexible approach to the estimation of water budgets and travel times, based on a solid informatics infrastructure.

The starting point of the work was the hydrological modeling system JGrass-NewAge, presented by Formetta *et al.* (2014a). While working, in a sense, perfectly, JGrass-NewAGE system presented some issues and major revision for extensibility and reuse were necessary. The thesis' work:

- implemented procedures to revise the codes and keep track on code modification and evolution;
- improved software extensibility and reuse, refactoring the codes by using of design patterns;
- chose a suitable software building system, for the growing set of tools;
- selected well supported and well designed libraries for basic numerical operations (like integrating systems of

ordinary differential equations, fitting curves, and other tasks);

- experimented a system to scale up modeling to arbitrarily large systems, starting from the composition of elementary units;
- formalized documentation procedures.

These topics, mentioned in the introductory chapter, were faced at the beginning of the work, and partially solved conjointly with Francesco Serafin.

In Chapter 3, an open-source semi-distributed model for runoff and evapotranspiration (from now on named Embedded Reservoir Model, ERM) was implemented and presented. The model schematizes each HRU as a connection of storages (the embedded reservoirs) and solves the water budget for each one. Eventually, all the contributions of the HRU's are aggregated to obtain the overall catchment hydrological fluxes.

Specifically, we tried to find the correct number of storages to capture the complexity of the hydrological response, making room for a better treatment of evapotranspiration, in a balance between the data availability and the computational costs. Five are the reservoirs considered for each HRU: snow pack, canopy, root zone, surface waters and groundwater. Identified outputs of the reservoirs are:

- snow melting and SWE;
- throughfall;
- evaporation from canopy;
- evapotranspiration from the root zone;
- percolation to the groundwater;

- direct runoff;
- baseflow.

In the almost-infinite panorama of hydrological models, the EMR doesn't aim to represent the perfect model but rather wants to represent a possible exploitation of our modular system. The implementation of each reservoir as a Java component according to OMS3, within the hydrological system JGrass-NewAge, makes it possible to connect/disconnect at run-time the reservoirs, according to the data availability, the HRUs characteristics (i.e., whether there is snow or not, whether there is a canopy or not) and, in general, to the modeler requirements.

Flexible and extensible, EMR allows to take into account a broad range of modeling strategies. The use of components allows the inspection and the estimation of the water budget terms for each HRU in which the domain was discretized. This feature is not usually easily obtainable with different informatics.

In a sense, we can say that we embraced the modeling manifesto presented in Clark *et al.* (2015a), called SUMMA (Structure for Unifying Multiple Modeling Alternative), whose main requirements for systematic analysis of models are (listed from Clark *et al.* (2015a), where further details can be obtained):

- (1) Capabilities to experiment with different representations of spatial variability and hydrologic connectivity;
- (2) Inclusion of a broad range of dominant biophysical and hydrologic processes, with multiple options for individual processes;
- (3) Clean separation of the model equations from their numerical solution;

(4) Flexibility to adjust model parameters.

As regards point (1), in our system the catchment can be subdivided according to the modeler requirements. This opens the possibilities of testing different connections and strategies that are theoretically unlimited. The use of the components, built on top of a solid informatics, supported by OMS3, is in line with point (2). Point (3) is also allowed by OMS3 and further enhanced with the contributions coming from this Thesis, using particular techniques of object oriented software design. Finally, as regards the modeling parameter issue, listed at point (4), the infrastructure design allows this type of investigations too.

Seen as a deployment of the SUMMA concept, our implementation differs from the original one for its informatics, and, furthermore, for the focus, which we gave to sets of connected reservoirs (i.e. ODEs) instead to interconnected partial differential equations.

The schematization of catchments and processes as networks of connected reservoirs is not new. However, the work presented in this Thesis, while proposing a flexible approach:

- put together the storage approach with the theory of travel times;
- designed a graphical algebraic notation, helping in model analysis and understanding.

The Thesis and the related papers present a new consistent theoretical framework which offers a unified view of the travel time theories, across surface water and groundwater.

Clarifications include:

- the concepts of forward and backward conditional probabilities and a small but important change in notation;

- their one-to-one relation with the water budget (and the age-ranked functions) from which the probabilities were derived (after the choice of SASs);
- the proper way to choose backward probabilities. Specifically, it was shown that the usual way to assign time invariant backward probabilities is inappropriate. We also show how to do it correctly, and introduced a minimal time variability.
- the fact that time-invariant forward probabilities usually imply time-varying backward probabilities, i.e. travel time distributions.
- the rewriting of the master equation by Botter, Bertuzzo and Rinaldo as an ordinary differential equation (instead of a partial differential equation);
- the role and nature of the partitioning coefficient between discharge and evapotranspiration (which is unknown at any time except asymptotically);
- the significance of the SAS functions with examples;
- the relationship of the present theory with the well known theory of the instantaneous unit hydrograph; and
- we added information and clarified some links of the present theory with [Delhez *et al.* (1999) and [Duffy (2010)].

Some extensions of the theory of travel times were also presented:

- new relations among the probabilities (including the relation between expectancy of life and forward residence time probabilities);

- an analysis of the partitioning coefficients (which are shown to vary seasonally);
- an explicit formulation of the equations for solutes which would permit direct determination of the SAS on the basis of experimental data;
- tests of the effects of various hypotheses, e.g., assuming a linear model of forward probability and gamma model for the backward probabilities;
- an extension of Niemi's relation (and a new normalization);
- the presentation of Niemi's relation as a special case of the Bayes Theorem; and
- a system of equations from which to obtain the SAS experimentally.

The extension of the theory to any passive substance diluted in water clearly opens the way to new developments of the theory and applications of tracers. The theory of travel times, based on the approach described in Chapter 4, was implemented in software components inside JGrass-NewAge system, described in Chapters 2.

The EMR modeling solution and the travel times theory were tested on a prealpine basin, the Posina River catchment. The focus was not just to reproduce discharges but all the components of the hydrological budget and observe how the closure of the budget happens at monthly and annual temporal scales of aggregation. This was performed for any of the budget terms mentioned above.

The practical analysis of the Posina catchment shown in this Thesis, can be considered "just" an example, given the many possibilities of investigation offered by the infrastructure, which were not yet fully explored.



Two applications, at hourly time-step and daily time-step, were performed and some results can be drawn. The discharge prediction is in line with other modeling performances, either with hourly or daily calibrations. From the analysis of the waterfall charts obtained for both the hourly and the daily time-step cases, it is clear that increases in precipitation tend to contribute directly to Q with minor effects on S and ET. The snow melting/rain represents the main source of variability in the budget, and (therefore) the ET shows a smoother behavior, since the radiation, which is the main driver, is consistent across the years.

In the case of hourly simulations, the total actual evapotranspiration (evaporation from the wet canopy plus evapotranspiration from the root zone) represents around the 20% of annual input and is lower than expected from other studies Abera *et al.* (2017b) (around the 30%), and has to be further checked in future. Discrepancies could be due to calibration problems caused by the hourly fluctuations of the measured discharge, or to the hypothesis made for obtaining the global budget closure.

In the daily timescale application, the reservoir are exchanging smaller volumes of water than at hourly time scale, and the storage available for the evaporation of the wet canopy is bigger. Therefore, the total actual ET, in this case, represents around the 30 % of the total inputs.

A preliminary travel time analysis was performed on the results obtained with the embedded reservoirs model, both at hourly and daily time-steps. The estimated mean values are for the two cases very similar ( around 77 days for the residence times, around 84 days for the travel times and around 10 days for the evapotranspiration times), which demonstrates that it is possible to perform the travel times analysis, when it is

convenient, also at daily time-scale, reducing the computational requirements of the calculations.

Both cases show an increasing trend in the temporal evolution of the mean travel times and mean residence times, which means that their values are getting bigger (older water) with time.

Longer simulations were not possible, due to the available dataset. However, the trend shows that the model is able to capture the temporal behavior of mean travel times and mean residence times.

Seasonal and inter-annual fluctuations characterized the temporal evolution of the mean travel times and residence times, while the temporal evolution of the mean evapotranspiration times is highly fluctuating, which means that the evapotranspiration times are easily affected by single events. The analysis of the partitioning coefficient was also performed. In both cases, the computation of the partitioning coefficients between discharge and evapotranspiration shown a coherence respect to the relative amounts of the volume computed, with the lower values achieved during the summer months.

Further analysis on the travel times and on the mixing are upcoming.

A mention has also to be made to the two main ancillary components that were refactored: the component for the estimation of the longwave radiation (LWRB) and the component for the spatialization of hydrological variables, based on Kriging techniques.

The work on LWRB covers:

- a broad assessment of the classic  $L_{\downarrow}$  longwave radiation parameterizations, which clearly shows that the Idso (1981) and Brunt (1932) models are the more robust and reliable for all the test sites, confirming previous results (Carmona *et al.*, 2014);

- a site specific assessment of the  $L_{\downarrow}$  longwave radiation model parameters for 24 AmeriFlux sites that improved the performances of all the models;
- the set up of a regression model that provides an estimate of optimal parameter sets on the basis climatic data;
- an assessment of  $L_{\uparrow}$  model performances for different temperatures (skin temperature, air temperature, and soil temperature at 4 cm below surface), which shows that the skin and the air temperature are better proxy for the  $L_{\uparrow}$  longwave radiation.

Regarding longwave downwelling radiation the Brunt (1932) model was found to provide on average the best performances with the regression model parameters independently of the latitude and longitude classes. For the Idso (1981) model the formulation with regression parameter provided lower performances with respect to the literature formulation for latitude between  $25^{\circ}\text{N}$  and  $30^{\circ}\text{N}$ .

As regards the Kriging spatialization, the work presents a Spatial Interpolation (SI) package, which:

- is a geostatistical software easy to use, easy to plug-in in a hydrological model, fast and flexible;
- proved to be able to accurately spatialize temperature and rainfall in a complex topography;
- has a faster option that traditional tools (i.e. gstat);
- has a full control on the use of variograms, and the possibility to exchange them at run-time.

The two packages, LWRB and SI, can be considered the state-of-art of the GEOframe system, a system for doing

hydrology by computer, whose the major scope is producing  
reproducible science.



## A.1 Calibration algorithms in OMS3

Two principle model calibration algorithms are part of the core of OMS3: Let Us CALibrate (LUCA) and Particle Swarm Optimization (PSO).

LUCA, Hay *et al.* (2006b), is a stepwise multiple-objective and automated procedure. The algorithm is based on a global search algorithm, which is the Shuffled Complex Evolution (SCE) Duan *et al.* (1993) and the objective functions that determines the goodness of fit between measured and simulated values to evaluate the performances. SCE consists of the following steps:

1. The set of parameters to be calibrated is considered as a point in  $n$  dimension space, where  $n$  is the number of parameters. SCE generates random points, within the lower and upper bound of each parameter.
2. The model is run with every set of parameters generated and an objective function is used to calculate a criterion

value for each point.

3. The points are divided into smaller groups, called complexes, such that points of good and bad criterion values are equally distributed.
4. Several points are selected from the complex to construct a sub-complex. In the sub-complex, a new point is generated, and a point that has a bad criterion value is replaced with this new point. This evolution step is repeated several times with different random points in a sub-complex.
5. All points in the complexes are combined together to be one group.
6. Steps (3) – (5), which represent a shuffling loop, are repeated until:
  - The number of model executions reaches the maximum.
  - The percent change in the best criterion value of the current shuffling loop respect to previous ones is less than a specified percentage.

The output is the parameter file containing the point (a parameter set) that has the best criterion value. Figure A.1 shows two important concept of LUCA calibration algorithm: steps and rounds. A step is associated with a parameter set, while a round consists of the execution of one or more steps. A classic SCE is a calibration with 1 step and 1 round.

PSO algorithms (Eberhart & Kennedy, 1995) are stochastic, population-based algorithms inspired by social behavior and movement dynamics of insects, birds, and fish. The goal is to find the global optimum in the search space, which is defined as

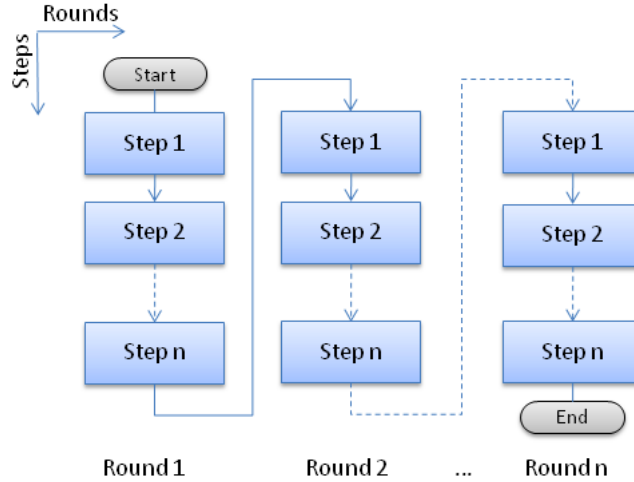


Figure A.1: Schematic representation of LUCA's rounds and step.

the point representing the best fitness. 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, to adjust its "flying" to take into account the information the other particles provide in order to determine the global optimum point. The flying experience is determined by the flying velocities, i.e. the rate of change of their position in parameters space. In a  $N$ -dimensional search space, at the step  $t$ , the position of the  $i$ -th particle of the swarm and its velocity are represented by  $N$ -dimensional vectors respectively:

$$X_i^t = x_{i,1}^t, x_{i,2}^t, \dots, x_{i,N}^t \text{ and } V_i^t = v_{i,1}^t, v_{i,2}^t, \dots, v_{i,N}^t.$$

At each time step, the velocity and position of each particles (i.e. of the parameter set) are updated. As in the case of LUCA, the system evolves until one of these two conditions is reached:

- The number of model executions reaches the maximum.
- The relative or absolute tolerance between the last two global optima fitness is reached.

## A.2 An example of implementation of the Simple Factory Design Pattern

As described 2.1, each OMS3 component implements a single modeling concept. For example, the Longwave Radiation Balance (LWRB) component allows the estimate of the downwelling and upwelling longwave radiation. According to the theory detailed in Formetta *et al.* (2016), there are several formulations of the downwelling clear sky emissivity. In LWRB component, 10 formulations were implemented for the runtime choice of the model to use.

From the informatics design point of view, this means that there are 10 (one for each model implemented) pieces of code which vary within the whole block (component). More/less models could be added/deleted. Therefore the code is probably going to be modified in future developments. In order to improve the component flexibility, it is necessary to identify the aspects of the application that vary and separate from what stays the same. "Encapsulating" what varies won't effect the rest of the code, resulting in fewer unwanted consequences from the changes of code and more flexibility. This is the very first of many Java design principles.

Kriging, LWRB, snow components were re-implemented according to the design principle of encapsulating behaviors.

The Simple Factory class was used to encapsulate what varies. It isn't actually a design pattern but more a programming idiom. It handles the object creation following another basic design principle: program to an interface, not an implementation. By coding to an interface, the code will work with any new classes implementing that interface through polymorphism. Through polymorphism, indeed, the concrete implementation of the proper model object is assigned at runtime. The choice of this



pattern was also due to the enforcing of a third important object oriented programming principle: the dependency inversion principle. The dependencies of the client are not demanded to the concrete subclasses but only to the abstract classes and interfaces (Ellis *et al.*, 2007).

Figure A.2 shows the implementation of the simple factory for the choice of the clear sky emissivity model. The concrete classes, Konzelmann and Brutsaert, implement the same interface, Model. The SimpleModelFactory generates object of concrete class from a given information (a string containing the name of the chosen model). The component LWRB uses the pattern to get the object of the concrete class.

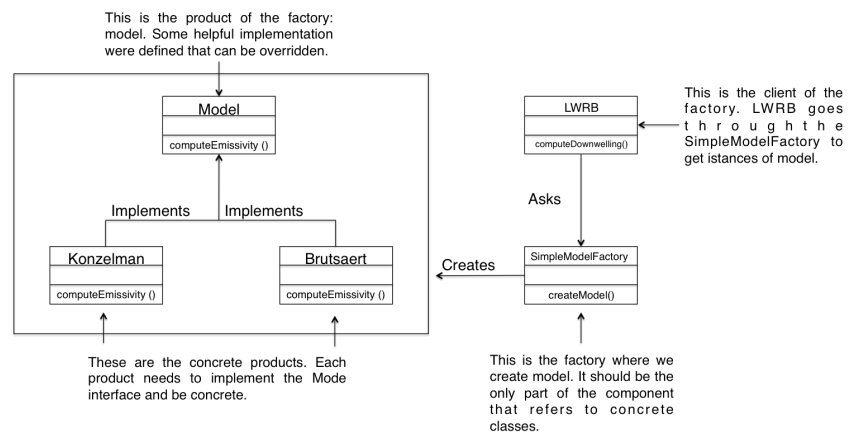


Figure A.2: Implementation of the Java simple factory for the choice of the clear sky emissivity model

Adding a new model to the options of the SimpleModelFactory is easy and straightforward:

1. create a new concrete class that implements the Model interface;
2. generate a new object of the concrete class based on the string 'name'.

### A.3 Some indices of goodness of fit.

The KGE incorporates into one objective function three different statistical measures of the relation between measured and simulated data: the correlation coefficient,  $r$ ; the variability error,  $a = \sigma_S / \sigma_M$ ; and the bias error,  $b = \mu_S / \mu_M$ .  $\mu_S$  and  $\mu_M$  are the mean values, while  $\sigma_S$  and  $\sigma_M$  are the standard deviations of measured and simulated data.

$$KGE = 1 - \sqrt{(r - 1)^2 + (a - 1)^2 + (b - 1)^2} \quad (A.1)$$

The maximum agreement between predicted and observed values is when  $KGE=1$ .

The Nash-Sutcliffe Efficiency (NSE) is a normalized model efficiency coefficient. It determines the relative magnitude of the residual variance compared to the measured data variance (Nash & Sutcliffe, 1970)

$$NSE = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \overline{O_i})^2} \quad (A.2)$$

where  $P_i$  is the predicted value and  $O_i$  observed value at a given time step. It varies between  $-\infty$  to 1, where 1 corresponds to the maximum agreement between predicted and observed values.

The RMSE, on the other hand, is presented in eq. A.3:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (M_i - S_i)^2} \quad (A.3)$$

where M and S represents the measured and simulated time-series respectively and N is their length.

## A.4 Symbols, Acronyms, and Notation

Symbol	Name	Unit
$a$	coefficient of the non-linear reservoir model	$[T^{-1}]$
$ae_T(t, t_{in})$	age-ranked evapotranspiration	$L^3 T^{-2}$
$ae_T(t, t_{ex})$	age-ranked $et$ conditioned to the exit time	$L^3 T^{-2}$
$a_{up}$	linear reservoir coefficient	$[T^{-1}]$
$b$	exponent of the non-linear reservoir model	$[-]$
$b_c$	canopy coefficient	$[T^{-1}]$
$b_{up}$	linear reservoir coefficient	$[T^{-1}]$
$c$	coefficient of the non-linear reservoir model	$[T^{-1}]$
$c_{low}$	non-linear reservoir coefficient	$[T^{-1}]$
$d$	exponent of the non-linear reservoir model	$[-]$
$d_{low}$	non-linear reservoir exponent	$[-]$
$e$	coefficient of the non-linear reservoir model	$[L^3 T^{-1}]$
$e$	screen level vapor pressure	$[kPa]$
$f$	exponent of the non-linear reservoir model	$[-]$
$f(t - t_{in} t_{in})$	time derivative of the relative $q$ function	$T^{-1}$
$f_{up}$	partitioning coefficient	$-$
$g(t - t_{in} t_{in})$	time derivative of the relative $et$ function	$T^{-1}$
$g(T_r)$	incomplete Gamma distribution	$T^{-1}$
$j(t, t_{in})$	age-ranked rainfall rate	$L^3 T^{-2}$
$j^i(t, t_{in})$	age-ranked input of the substance $i$	$L^3 T^{-2}$
$k_{1,2}$	reaction's constants	$-$
$k_c$	LAI coefficient	$[L^3]$
$p$	free throughfall coefficient	$[-]$
$p^i(t - t_{in} t)$	travel time backward pdf of the substance $i$	$T^{-1}$
$p_{E_r}(t - t_{in} t)$	evapotranspiration time backward pdf	$T^{-1}$
$p_{E_r}(t - t_{in} t_{in})$	evapotranspiration time forward pdf	$T^{-1}$
$p_J(t_{in})$	marginal pdf of the outflow as discharge	$-$
$p_{low}(t - t_{in} t)$	travel time backward pdf of the lower storage	$T^{-1}$
$p_Q(t - t_{in} t)$	travel time backward pdf	$T^{-1}$
$p_Q(t - t_{in} t_{in})$	travel time forward pdf	$T^{-1}$
$p_Q(t_{in})$	marginal pdf of the injection times	$-$
$p_S(T_r t)$	residence time backward pdf	$T^{-1}$
$p_S(t - t_{in} t_{in})$	residence time forward pdf	$T^{-1}$
$p_S(t_{ex} - t t)$	life expectancy backward pdf	$T^{-1}$
$p_{sat}(t - t_{in} t)$	travel time backward pdf of the saturated storage	$T^{-1}$
$p_{Sat}$	saturation percentage	$\%$
$p_{sup}(t - t_{in} t)$	travel time backward pdf of the upper storage	$T^{-1}$
$q(t, t_{in})$	age-ranked discharge	$L^3 T^{-2}$
$q(t, t_{ex})$	age-ranked discharge conditioned to the exit time	$L^3 T^{-2}$
$q^i(t, t_{in})$	age-ranked output of the substance $i$	$L^3 T^{-2}$

Symbol	Name	Unit
$q_{low}(t, t_{in})$	age-ranked discharge for the lower reservoir	$L^3 T^{-2}$
$q_{sat}(t, t_{in})$	age-ranked discharge for the saturated reservoir	$L^3 T^{-2}$
$r^i(t, t_{in})$	age-ranked sink/source term	$L^3 T^{-2}$
$s(t, t_{in})$	age-ranked water storage	$L^3 T^{-1}$
$s^i(t, t_{in})$	age-ranked water storage of the substance $i$	$L^3 T^{-2}$
$s_{eq}^i$	equilibrium storage	$L^3 T^{-1}$
$s_{ex}(t, t_{ex})$	age-ranked water storage conditioned to the exit t	$L^3 T^{-1}$
$s_{low}(t, t_{in})$	age-ranked water storage lower reservoir	$L^3 T^{-1}$
$s_{up}(t, t_{in})$	age-ranked water storage upper reservoir	$L^3 T^{-1}$
$s_{sat}(t, t_{in})$	age-ranked water storage saturated reservoir	$L^3 T^{-1}$
$t$	actual time	T
$t_{ex}$	exit time	T
$t_{in}$	injection time	T
$t_p$	end time of the last precipitation in the analysis	T
$w$	precipitable water	$[ML^{-2}]$
$A$	Area of the HRU	$[L^2]$
$AE_T(t)$	actual evapotranspiration	$L^3 T^{-1}$
$C_J^i(t)$	concentration in input	–
$C_S^i(t)$	concentration in storage	–
$C_Q^i(t)$	concentration in discharge	–
$D(t)$	drainage	$[L^3 T^{-1}]$
$ET_c(t)$	evaporation	$[L^3 T^{-1}]$
$ET_{rz}(t)$	evapotranspiration	$[L^3 T^{-1}]$
$ET_p(t)$	potential evapotranspiration	$[L^3 T^{-1}]$
$F(t)$	freezing water	$[L^3 T^{-1}]$
$\mathcal{F}(t - t_{in} t_{in})$	relative discharge function	–
$\mathcal{G}(t - t_{in} t_{in})$	relative evapotranspiration function	–
$I_{top}$	incoming radiation	$[W/L^2 T^{-1}]$
$I_{top}$	top atmosphere shortwave radiation	$[W/L^2 T^{-1}]$
$R_s(t)$	shortwave radiation	$[W/L^2 T^{-1}]$
$J(t)$	rainfall rates	$L^3 T^{-1}$
$J^i(t)$	input rates of the substance $i$	$L^3 T^{-1}$
$J(t_{in})$	precipitation at a certain $t_{in}$	$L^3 T^{-1}$
$LAI$	Leaf Area Index	$[L^2 L^{-2}]$
$L_e$	life expectancy	T
$L \uparrow$	downwelling longwave radiation	$[W/L^2 T^{-1}]$
$L \downarrow$	upwelling longwave radiation	$[W/L^2 T^{-1}]$
$M(t)$	melt	$[L^3 T^{-1}]$
$M_d(t)$	Melting discharge/rain	$[L^3 T^{-1}]$
$P_r$	rainfall	$[L^3 T^{-1}]$
$P_s$	snowfall	$[L^3 T^{-1}]$
$P_S(t - t_{in} t_{in})$	residence time forward probability function	–
$Q(t)$	discharge	$L^3 T^{-1}$
$Q^i(t)$	output rates of the substance $i$	$L^3 T^{-1}$
$Q_1$	recharge to the saturated reservoir	$L^3 T^{-1}$

Symbol	Name	Unit
$Q_{GW}(t)$	groundwater discharge	$[L^3 T^{-1}]$
$Q_l$	runoff produced by the lower reservoir	$L^3 T^{-1}$
$Q_{low}(t)$	discharge from the lower layer	$[L^3 T^{-1}]$
$Q_{sat}$	outflow from the saturated storage	$L^3 T^{-1}$
$Q_{up}(t)$	discharge from the upper layer	$[L^3 T^{-1}]$
$Re(t)$	recharge term of groundwater	$[L^3 T^{-1}]$
$R(t)$	recharge of groundwater	$[L^3 T^{-1}]$
$R^i(S(t))$	sink/source term	$L^3 T^{-1}$
$R_n$	ret radiation	$[W/L^2 T^{-1}]$
$R_s(t)$	shortwave radiation	$[W/L^2 T^{-1}]$
$R(t, t_{in})$	input to the lower reservoir	$L^3 T^{-1}$
$R_{t_{in}}$	instantaneous impulses at different $t_{in}s$	$L^3$
$S(t)$	volume of water stored in a control volume	$L^3$
$S_c(t)$	canopy storage	$[L^3]$
$S_{c_{max}}(t)$	canopy maximum retention storage	$[L^3]$
$S_{GW}(t)$	groundwater storage	$[L^3]$
$S_{GW_{max}}(t)$	maximum groundwater storage	$[L^3]$
$S_i(t)$	solid water content in the snowpack	$[L^3]$
$S^i(t)$	stored mass of the substance $i$ stored	$L^3$
$S_{low}(t)$	storage of the lower layer	$[L^3]$
$S_{max}(t)$	maximum storage	$[L^3]$
$S_R(t)$	runoff storage	$[L^3]$
$S_{rz}(t)$	root zone storage	$[L^3]$
$S_{rz_{max}}(t)$	maximum root zone storage	$[L^3]$
$S_{up}(t)$	storage of the upper layer	$[L^3]$
$S_w(t)$	liquid water in the snowpack	$[L^3]$
$S_{sat}$	water stored in the saturated storage	$L^3$
$T(t)$	temperature	$[^{\circ}C]$
$T$	travel time	T
$T_a(t)$	air temperature	$[^{\circ}C]$
$T_s(t)$	soil temperature	$[^{\circ}C]$
$T_r$	residence time	T
$T_r(t)$	throughfall	$[L^3 T^{-1}]$
$T_m$	melting temperature	$[^{\circ}C]$
$u$	wave celerity	$[LT^{-1}]$
$V_{AET}(t, t_{in})$	time integral age-ranked $et$	$L^3 T^{-1}$
$V_s$	sky view factor	[-]
$V_S(t_p)$	total volume injected in the volume in $[0, t_p]$	$L^3 T^{-1}$
$V_Q(t, t_{in})$	time integral of the age-ranked discharge	$L^3 T^{-1}$
$W(t - \tau)$	WFIUH	$[T^{-1}]$
$\alpha$	partitioning coeff root zone-surface runoff	[-]
$\alpha_G$	coefficient of the gamma distribution	-
$\alpha_e$	radiation factor	$[L^{\circ}C^{-1}E^{-1}T^{-1}]$
$\alpha_f$	freezing factor	$[L^{\circ}C^{-1}T^{-1}]$
$\alpha_l$	liquid water retention capacity coefficient	[-]
$\alpha_m$	melt factor	$[L^{\circ}C^{-1}T^{-1}]$
$\alpha_{\tau}$	partitioning coefficient	[-]

Symbol	Name	Unit
$\delta(t - t_{in})$	Delta-dirac distribution	$T^{-1}$
$\gamma$	coefficient of the gamma distribution	–
$\gamma(h)$	semivariance	–
$\epsilon_{clear}$	Clear sky emissivity	–
$\lambda$	coefficient of the non-linear reservoir model	T
$\mu(t, t_{in})$	age and mass-specific output rate	–
$\omega_{ET}(t, t_{in})$	SAS for evapotranspiration	–
$\omega_{low}(t, t_{in})$	SAS for runoff produced by the lower reservoir	–
$\omega_Q(t, t_{in})$	SAS for discharge	–
$\omega_{Q_1}(t, t_{in})$	SAS for the recharge to the saturated reservoir	–
$\omega_R(t, t_{in})$	SAS for the recharge to the lower reservoir	–
$\omega_{Q_{sat}}(t, t_{in})$	SAS for runoff produced by the saturated storage	–
$\sigma$	Stefan-Boltzmann constant	$[WL^{-2}K^{-4}]$
$\Theta(t_{in})$	partitioning coefficient	–
$\Gamma$	Gamma function	–

Table A.1: List of symbols, names and units used in the thesis.

## A.5 An example of sim file of the embedded reservoir model for a single HRU

```
import static oms3.SimBuilder.instance as OMS3
def home = oms_prj
def startDate= "1994-01-01 00:00"
def endDate= "1998-12-31 00:00"
def ID=1

OMS3.sim (name: "1") {

resource "$oms_prj/lib"

model(while: "reader_data_rain.doProcess") {
  components {
    // list of the component to use

    "reader_data_ET" "org.jgrasstools.gears.io.
      timedependent.OmsTimeSeriesIteratorReader"
    "reader_data_LAI" "org.jgrasstools.gears.io.
      timedependent.OmsTimeSeriesIteratorReader"
```

```

"reader_data_rain" "org.jgrasstools.gears.io.
    timedependent.OmsTimeSeriesIteratorReader"
"reader_data_snow" "org.jgrasstools.gears.io.
    timedependent.OmsTimeSeriesIteratorReader"
"reader_data_SWRB" "org.jgrasstools.gears.io.
    timedependent.OmsTimeSeriesIteratorReader"
"reader_data_temp" "org.jgrasstools.gears.io.
    timedependent.OmsTimeSeriesIteratorReader"
"reader_dem" "org.jgrasstools.gears.io.
    rasterreader.OmsRasterReader"
"reader_sky" "org.jgrasstools.gears.io.
    rasterreader.OmsRasterReader"
"snow" "snowMeltingPointCase.SnowMeltingPointCase"
"vreader_station" "org.jgrasstools.gears.io.
    shapefile.OmsShapefileFeatureReader"

"waterBudgetRZ" "rootZone.WaterBudgetRootZone"
"waterBudgetCanopyOUT" "canopyOUT.
    WaterBudgetCanopyOUT"
"waterBudgetRunoff" "runoff.WaterBudgetRunoff
    "

"reader_width" "org.jgrasstools.gears.io.
    rasterreader.OmsRasterReader"
"reader_topIndex" "org.jgrasstools.gears.io.
    rasterreader.OmsRasterReader"

"waterBudgetGW" "groundWater.
    WaterBudgetGroundWater"

"sum" "sumSeries.SumSeries"
"outNode1" "ex0.Out1"

"writer_Qgw" "org.jgrasstools.gears.io.
    timedependent.OmsTimeSeriesIteratorWriter"
"writer_Qro" "org.jgrasstools.gears.io.
    timedependent.OmsTimeSeriesIteratorWriter"

}

// parameter of the components
parameter{

```

```
// reader input data

"reader_data_rain.file" "${home}/data/${ID}/rain_${ID}.csv"
"reader_data_rain.idfield" "ID"
"reader_data_rain.tStart" "${startDate}"
"reader_data_rain.tEnd" "${endDate}"
"reader_data_rain.tTimestep" 60
"reader_data_rain.fileNovalue" "-9999"

"reader_data_snow.file" "${home}/data/${ID}/snow_${ID}.csv"
"reader_data_snow.idfield" "ID"
"reader_data_snow.tStart" "${startDate}"
"reader_data_snow.tEnd" "${endDate}"
"reader_data_snow.tTimestep" 60
"reader_data_snow.fileNovalue" "-9999"

"reader_data_SWRB.file" "${home}/data/${ID}/total_${ID}.csv"
"reader_data_SWRB.idfield" "ID"
"reader_data_SWRB.tStart" "${startDate}"
"reader_data_SWRB.tEnd" "${endDate}"
"reader_data_SWRB.tTimestep" 60
"reader_data_SWRB.fileNovalue" "-9999"

"reader_data_temp.file" "${home}/data/${ID}/airT_${ID}.csv"
"reader_data_temp.idfield" "ID"
"reader_data_temp.tStart" "${startDate}"
"reader_data_temp.tEnd" "${endDate}"
"reader_data_temp.tTimestep" 60
"reader_data_temp.fileNovalue" "-9999"

"reader_dem.file" "${home}/data/${ID}/dem_${ID}.asc"
"reader_sky.file" "${home}/data/${ID}/sky_${ID}.asc"
"vreader_station.file" "${home}/data/${ID}/centroids_netnum_${ID}.shp"
```



```

// snow component
"snow.fStationsid" "netnum"
"snow.model" "Hock"
"snow.tStartDate" "${startDate}"
"snow.combinedMeltingFactor" 0.955102
"snow.freezingFactor" 0.089217
"snow.alfa_l" 0.304315
"snow.radiationFactor" 0.000061
"snow.meltingTemperature" 1.94

"reader_data_ET.file" "${home}/data/${ID}/etp_${ID}
}.csv"
"reader_data_ET.idfield" "ID"
"reader_data_ET.tStart" "${startDate}"
"reader_data_ET.tEnd" "${endDate}"
"reader_data_ET.tTimestep" 60
"reader_data_ET.fileNovalue" "-9999"

"reader_data_LAI.file" "${home}/data/${ID}/LAI_${
ID}.csv"
"reader_data_LAI.idfield" "ID"
"reader_data_LAI.tStart" "${startDate}"
"reader_data_LAI.tEnd" "${endDate}"
"reader_data_LAI.tTimestep" 60
"reader_data_LAI.fileNovalue" "-9999"

// canopy component
"waterBudgetCanopyOUT.solver_model" "dp853"
"waterBudgetCanopyOUT.kc_canopy_out" 0.25
"waterBudgetCanopyOUT.p" 0.65

// root zone component
"waterBudgetRZ.solver_model" "dp853"
"waterBudgetRZ.s_RootZoneMax" 250
"waterBudgetRZ.pB" 1
"waterBudgetRZ.Pmax" 0.2
"waterBudgetRZ.b_rz" 1.744739
"waterBudgetRZ.A" 3.79
"waterBudgetRZ.inTimestep" 60

// surface runoff component
"reader_width.file" "${home}/data/${ID}/rescaled_${

```

```
    {ID}.asc"
"reader_topIndex.file" "${home}/data/${ID}/top_${
    ID}.asc"
"waterBudgetRunoff.pCelerity" 1
"waterBudgetRunoff.pSat" 20
"waterBudgetRunoff.inTimestep" 60
"waterBudgetRunoff.ID" "${ID}"

// groundwater component
"waterBudgetGW.solver_model" "dp853"
"waterBudgetGW.a" 255
"waterBudgetGW.b" 7.58
"waterBudgetGW.A" 3.79
"waterBudgetGW.timeStep" 60
"waterBudgetGW.Smax" 532

// writer output data
"writer_Qgw.tStart" "${startDate}"
"writer_Qgw.tTimestep" 60
"writer_Qgw.fileNovalue" "-9999"

"writer_Qro.file" "${home}/output/Qro_${ID}.csv"
"writer_Qro.tStart" "${startDate}"
"writer_Qro.tTimestep" 60
"writer_Qro.fileNovalue" "-9999"

}

\textcolor{red}{//connections}
connect {

    "reader_data_rain.outData" "snow.
        inRainfallValues"
    "reader_data_snow.outData" "snow.
        inSnowfallValues"
    "reader_data_temp.outData" "snow.
        inTemperatureValues"
    "reader_data_SWRB.outData" "snow.
        inShortwaveRadiationValues"
    "reader_dem.outRaster" "snow.inDem"
    "reader_sky.outRaster" "snow.inSkyview"
```

```

"vreader_station.geodata" "snow.inStations"

"snow.outMeltingDischargeHM" "
  waterBudgetCanopyOUT.inHMRain"
"reader_data_ET.outData" "waterBudgetCanopyOUT
.inHMETp"
"reader_data_LAI.outData" "
  waterBudgetCanopyOUT.inHMLAI"

"waterBudgetCanopyOUT.outHMThroughfall" "
  waterBudgetRunoff.inRainValues"
"reader_width.outRaster" "waterBudgetRunoff.
.inRescaledDistance"
"reader_topIndex.outRaster" "waterBudgetRunoff
.inTopindex"
"waterBudgetRZ.alpha" "waterBudgetRunoff.alpha
"

"waterBudgetCanopyOUT.outHMThroughfall" "
  waterBudgetRZ.inHMRain"
"reader_data_ET.outData" "waterBudgetRZ.
.inHMETp"
"waterBudgetCanopyOUT.outHMAET" "
  waterBudgetRZ.inHMEwc"

"waterBudgetRZ.outhMR" "waterBudgetGW.
.inHMRechargeValues"

"waterBudgetRunoff.outhMDischarge" "sum.
.inHMDischarge"
"waterBudgetGW.outhMDischarge" "sum.
.inHMDischarge2"

"waterBudgetGW.outhMDischarge" "writer_Qgw.
.inData"

```

```

        "waterBudgetRunoff.outHMDischarge" "
            writer_Qro.inData"

        "sum.outHMQtot" "outNode1.inval"

    }

// Net3 spatial connections
outFluxes {
    "outNode1.outvall1" ""
}
}
}

```

## A.6 List of semivariogram model implemented in SI

Using  $n$  to represent the nugget,  $h$  to represent lag distance,  $r$  to represent range, and  $s$  to represent sill, the ten theoretical semivariogram models most frequently used in literature are:

- Bessel semivariogram

$$\gamma(h) = s \cdot \left(1 - \text{frachr} \cdot k1\left(\frac{h}{r}\right)\right) \quad (\text{A.4})$$

- Circular semivariogram

$$\begin{cases} \gamma(h) = n + s \cdot \left\{ \frac{2}{\pi} \cdot \left[ \frac{h}{r} \cdot \sqrt{1 - \left(\frac{h}{r}\right)^2} \right] + \arcsin\left(\frac{h}{r}\right) \right\} & h < r \\ \gamma(h) = n + s & h \geq r \end{cases} \quad (\text{A.5})$$

- Exponential semivariogram

$$\gamma(h) = n + s \cdot \left(1 - e^{-\frac{h}{r}}\right) \quad (\text{A.6})$$

- Gaussian semivariogram

$$\gamma(h) = n + s \cdot \left[1 - e^{-\left(\frac{h}{r}\right)^2}\right] \quad (\text{A.7})$$

- Hole semivariogram

$$\gamma(h) = n + s \cdot \left[1 - \frac{\sin\left(\frac{h}{r}\right)}{\frac{h}{r}}\right] \quad (\text{A.8})$$

- Linear semivariogram

$$\begin{cases} \gamma(h) = n + s \cdot \frac{h}{r} & h < r \\ \gamma(h) = n + s & h \geq r \end{cases} \quad (\text{A.9})$$

- Logarithmic semivariogram

$$\gamma(h) = n + s \cdot \log\left(\frac{h}{r}\right) \quad (\text{A.10})$$

- Pentaspherical semivariogram

$$\begin{cases} \gamma(h) = n + s \cdot \left\{ \frac{15}{8} \frac{h}{r} + \left(\frac{h}{r}\right)^3 \cdot \left[ -\frac{5}{4} + \frac{3}{8} \left(\frac{h}{r}\right)^5 \right] \right\} & h < r \\ \gamma(h) = n + s & h \geq r \end{cases} \quad (\text{A.11})$$

- Periodic semivariogram

$$\gamma(h) = n + s \cdot \left[ 1 - \cos\left(2\pi \frac{h}{r}\right) \right] \quad (\text{A.12})$$

- Power semivariogram

$$\gamma(h) = n + s \cdot h^r \quad (\text{A.13})$$

- Spherical semivariogram

$$\begin{cases} \gamma(h) = n + s \cdot \left[ 1.5 \cdot \frac{h}{r} - 0.5 \cdot \left(\frac{h}{r}\right)^3 \right] & h < r \\ \gamma(h) = n + s & h \geq r \end{cases} \quad (\text{A.14})$$

## A.7 Kriging dataset: the Isarco River Basin

Station ID	Elevation	X	Y
1008	254	677379	5151854
1010	560	703978	5177054
1025	1250	746077	5179955
1123	1205	704851	5139977
1131	821	723632	5187440
1137	2906	749059	5174631
1138	1990	677518	5190276
1139	2145	676739	5189931
1140	3105	737918	5211545
1142	2006	737151	5213768
1145	2985	716594	5156827
1146	2050	722698	5160726
1147	2260	688344	5165237
1152	943	685746	5195128
1153	2473	708956	5175007
1260	2777	692951	5206370
1262	1645	720648	5153469
1270	2100	730594	5155931
1274	1314	738357	5164560
1284	2142	716596	5151487
1311	1736	748986	5163080
1324	2265	747634	5165909
1326	2615	735292	5157843
1332	1750	700847	5142153

---

Station ID	Elevation	X	Y
1343	1385	708425	5149125
90072	1147	666135	5187742
90074	644	671444	5186398
90130	1330	689271	5206060
90133	1246	678592	5203835
90135	1960	680117	5200634
90138	948	685044	5196675
90140	1440	697647	5204620
90145	3399	666809	5203984
90147	1364	675668	5197644
90148	2145	676779	5190080
90149	1990	677414	5190356
90155	943	685228	5195148
90156	943	685786	5195277
90159	850	694445	5187499
90162	590	702029	5178611
90166	1219	745961	5180192
90168	1285	736938	5177815
90170	2340	737994	5173329
90172	1131	739010	5181375
90175	1412	747278	5191964
90176	2747	743952	5194145
90177	2152	744722	5192575
90182	1320	737507	5195474

---

Station ID	Elevation	X	Y
90186	2006	737193	5213916
90187	3105	737960	5211693
90189	1450	735444	5213892
90192	1080	726201	5208726
90193	2155	717838	5200867
90196	1562	734591	5203425
90202	1141	718972	5196967
90203	870	723822	5198107
90211	828	723674	5187588
90216	1558	720262	5159216
90218	1428	722518	5163202
90220	2050	722543	5160892
90222	2985	716635	5156974
90225	1150	721627	5173593
90230	820	719843	5184315
90232	750	709213	5188155
90233	1349	712423	5190536
90234	2808	704501	5202375
90235	2050	704653	5200382
90236	1159	705920	5196174
90239	1410	700296	5191661
90266	490	693721	5163281
90267	840	692236	5154144
90269	1022	694161	5149040



Station ID	Elevation	X	Y
90273	1616	698761	5142305
90275	1128	694888	5144827
90276	2125	695895	5137598
90287	2100	689061	5185368
90293	966	680012	5167662
90296	2260	688384	5165385
90298	1140	678615	5156585
90312	254	677473	5151945
90337	1470	686486	5143630
90354	1562	684427	5135344
90387	2906	749101	5174778
90458	1750	700886	5142300
90459	1205	704891	5140124
90467	1000	689420	5129334
90532	2040	702547	5144578
90533	2050	703949	5152028
90534	1385	708465	5149272
90631	1465	712386	5150914
90639	2376	718717	5150433
90651	1350	707040	5155154
90652	1050	700371	5133994

## A.8 Reproducible research

Our effort is always to share our findings, our methods, our codes and our results among the communities, toward the reproducible research. As clear stated in Chapter 2, this effort requires a backward major effort to make our research as much as flexible, easy to maintain and easy to be inspected by third parties. In this way, we expect active feedbacks from the community, which are, for sure, a great contribution to our success. Our source codes and projects are always shared on the GEOframe organization on Github, as well as the dataset used and results are shared on Zenodo. In particular, the source code used in this thesis are available at:

- Embedded reservoir model v 1.0: [https://github.com/geoframecomponents/WATER\\_BUDGET/tree/v1.0](https://github.com/geoframecomponents/WATER_BUDGET/tree/v1.0)
- Travel times theory v 1.0: <https://github.com/geoframecomponents/TravelTimes/tree/v1.0>

- **Krigings interpolation v 0.9.7:** <https://github.com/geoframecomponents/Krigings/tree/v0.9.7>
- **Longewave radiation balance v 0.94:**  
<https://github.com/geoframecomponents/LongWaveRadiationBalance/tree/v0.94>
- **All the other ancillary components are available at:**  
<https://github.com/geoframecomponents>

As regards the OMS projects, they are available at:

- **Posina results:**  
<https://doi.org/10.5281/zenodo.1043942>
- **Krigings interpolation:**  
<https://doi.org/10.5281/zenodo.1044052>
- **All the other projects for the ancillary components:**  
<https://github.com/GEOframeOMSPProjects>

Documentations of each component are available at:

<http://geoframe.blogspot.it/>.

All the other findings, thoughts and open debates are available at

<http://abouthydrology.blogspot.it>.

## BIBLIOGRAPHY

- Abera, Wuletawu, Antonello, Andrea, Franceschi, Silvia, Formetta, Giuseppe, & Rigon, Riccardo. 2014.  
The uDig Spatial Toolbox for hydro-geomorphic analysis.  
*Geomorphological Techniques*, **2**(4.1), 1–19.
- Abera, Wuletawu, Formetta, Giuseppe, Borga, Marco, & Rigon, Riccardo. 2017a.  
Estimating the water budget components and their variability in a pre-alpine basin with JGrass-NewAGE.  
*Advances in Water Resources*, **104**, 37–54.
- Abera, Wuletawu, Formetta, Giuseppe, Brocca, Luca, & Rigon, Riccardo. 2017b.  
Modeling the water budget of the Upper Blue Nile basin using the JGrass-NewAge model system and satellite data.  
*Hydrology and Earth System Sciences*, **21**(6), 3145.
- Aidoo, Eric N, Mueller, Ute, Goovaerts, Pierre, & Hyndes, Glenn A. 2015.  
Evaluation of geostatistical estimators and their applicability to characterise the spatial patterns of recreational fishing catch rates.  
*Fisheries research*, **168**, 20–32.
- Alados-Arboledas, L, Vida, J, & Olmo, FJ. 1995.  
The estimation of thermal atmospheric radiation under cloudy conditions.  
*International journal of climatology*, **15**(1), 107–116.
- Ali, Melkamu, Fiori, Aldo, & Russo, David. 2014.  
A comparison of travel-time based catchment transport models, with application to numerical experiments.  
*Journal of Hydrology*, **511**, 605–618.
- Allen, Richard G, Pereira, Luis S, Raes, Dirk, Smith, Martin, *et al.* . 1998.

Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56.  
FAO, Rome, **300**(9), D05109.

Ångström, Anders Knutsson. 1915.

*A study of the radiation of the atmosphere: based upon observations of the nocturnal radiation during expeditions to Algeria and to California.*  
Vol. 65.  
Smithsonian Institution.

Aston, AR. 1979.

Rainfall interception by eight small trees.  
*Journal of hydrology*, **42**(3), 383–396.

Attorre, Fabio, Alfo, Marco, De Sanctis, Michele, Francesconi, Fabio, & Bruno, Franco. 2007.

Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale.  
*International Journal of Climatology*, **27**(13), 1825–1843.

Augustine, John A, DeLuisi, John J, & Long, Charles N. 2000.

SURFRAD-A national surface radiation budget network for atmospheric research.  
*Bulletin of the American Meteorological Society*, **81**(10), 2341–2357.

Augustine, John A, Hodges, Gary B, Cornwall, Christopher R, Michalsky, Joseph J, & Medina, Carlos I. 2005.

An update on SURFRAD-The GCOS Surface Radiation budget network for the continental United States.  
*Journal of Atmospheric and Oceanic Technology*, **22**(10), 1460–1472.

Baldocchi, Dennis, Falge, Eva, Gu, Lianhong, Olson, Richard, Hollinger, David, Running, Steve, Anthoni, Peter, Bernhofer, Ch, Davis, Kenneth, Evans, Robert, *et al.* . 2001.

FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities.  
*Bulletin of the American Meteorological Society*, **82**(11), 2415–2434.

Bancheri, Marialaura, Formetta, Giuseppe, Serafin, Francesco, Rigon, Riccardo, Green, Timothy R, & David, Olaf. 2016.

Replicability of a modelling solution using NewAGE-JGrass.

- Bancheri, Marialaura, Serafin, Francesco, Formetta, Giuseppe, Rigon, Riccardo, & David, Olaf. 2017.  
JGrass-NewAge hydrological system: an open-source platform for the replicability of science.  
*Page 17109 of: EGU General Assembly Conference Abstracts*, vol. 19.
- Band, Lawrence E. 1986.  
Topographic Partition of Watersheds with Digital Elevation Models.  
*Water Resources Research*, **22**(1), 15–24.
- Barr, Jordan G, Engel, Vic, Smith, Thomas J, & Fuentes, José D. 2012.  
Hurricane disturbance and recovery of energy balance, CO<sub>2</sub> fluxes and canopy structure in a mangrove forest of the Florida Everglades.  
*Agricultural and Forest Meteorology*, **153**, 54–66.
- Basistha, Ashoke, Arya, DS, & Goel, NK. 2008.  
Spatial distribution of rainfall in Indian himalayas—a case study of Uttarakhand region.  
*Water Resources Management*, **22**(10), 1325–1346.
- Benettin, Paolo, Rinaldo, Andrea, & Botter, Gianluca. 2013.  
Kinematics of age mixing in advection-dispersion models.  
*Water Resources Research*, **49**(12), 8539–8551.
- Benettin, Paolo, Kirchner, James W, Rinaldo, Andrea, & Botter, Gianluca. 2015a.  
Modeling chloride transport using travel time distributions at Plynlimon, Wales.  
*Water Resources Research*, **51**(5), 3259–3276.
- Benettin, Paolo, Rinaldo, Andrea, & Botter, Gianluca. 2015b.  
Tracking residence times in hydrological systems: forward and backward formulations.  
*Hydrological Processes*, **29**(25), 5203–5213.
- Benettin, Paolo, Soulsby, Chris, Birkel, Christian, Tetzlaff, Doerthe, Botter, Gianluca, & Rinaldo, Andrea. 2017.  
Using SAS functions and high-resolution isotope data to unravel travel time distributions in headwater catchments.  
*Water Resources Research*, **53**(3), 1864–1878.
- Berglund, Tim, & McCullough, Matthew. 2011.

*Building and Testing with Gradle.*

" O'Reilly Media, Inc. " .

Berman, Elena SF Gupta, Manish, Gabrielli, Chris, Garland, Tina, &  
McDonnell, Jeffrey J. 2009.

High-frequency field-deployable isotope analyzer for hydrological  
applications.

*Water Resources Research*, **45**(10).

Bertuzzo, E, Thomet, M, Botter, G, & Rinaldo, A. 2013.

Catchment-scale herbicides transport: Theory and application.

*Advances in water resources*, **52**, 232–242.

Beven, Keith J. 2011.

*Rainfall-runoff modelling: the primer.*

John Wiley & Sons.

Birkel, Christian, Soulsby, Chris, & Tetzlaff, Doerthe. 2011.

Modelling catchment-scale water storage dynamics: reconciling dynamic  
storage with tracer-inferred passive storage.

*Hydrological Processes*, **25**(25), 3924–3936.

Birkel, Christian, Soulsby, Chris, & Tetzlaff, Dörthe. 2014.

Developing a consistent process-based conceptualization of catchment  
functioning using measurements of internal state variables.

*Water Resources Research*, **50**(4), 3481–3501.

Blöschl, Günter, & Sivapalan, Murugesu. 1995.

Scale issues in hydrological modelling: a review.

*Hydrological processes*, **9**(3-4), 251–290.

Boer, Eric PJ, de Beurs, Kirsten M, & Hartkamp, A Dewi. 2001.

Kriging and thin plate splines for mapping climate variables.

*International Journal of Applied Earth Observation and Geoinformation*,  
**3**(2), 146–154.

Bolz, HM. 1949.

Die Abhängigkeit der infraroten Gegenstrahlung von der Bewölkung.

*Z Meteorol*, **3**, 201–203.

Botter, Gianluca. 2012.

Catchment mixing processes and travel time distributions.

*Water Resources Research*, **48**(5), n/a–n/a.

- W05545.
- Botter, Gianluca, & Rinaldo, Andrea. 2003.  
Scale effect on geomorphologic and kinematic dispersion.  
*Water Resour. Res.*, **39**(Oct.), 10 PP.
- Botter, Gianluca, Bertuzzo, Enrico, & Rinaldo, Andrea. 2010.  
Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox.  
*Water Resources Research*, **46**(3).
- Botter, Gianluca, Bertuzzo, Enrico, & Rinaldo, Andrea. 2011.  
Catchment residence and travel time distributions: The master equation.  
*Geophysical Research Letters*, **38**(11).
- Brunt, David. 1932.  
Notes on radiation in the atmosphere. I.  
*Quarterly Journal of the Royal Meteorological Society*, **58**(247), 389–420.
- Brutsaert, Wilfried. 1975.  
On a derivable formula for long-wave radiation from clear skies.  
*Water Resources Research*, **11**(5), 742–744.
- Brutsaert, Wilfried. 2005.  
*Hydrology: an introduction*.  
Vol. 61.  
Wiley Online Library.
- Butcher, John Charles. 1987.  
*The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*.  
Wiley-Interscience.
- Buytaert, Wouter, Celleri, Rolando, Willems, Patrick, Bièvre, Bert De, & Wyseure, Guido. 2006.  
Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes.  
*Journal of Hydrology*, **329**(3), 413–421.
- Calabrese, Salvatore, & Porporato, Amilcare. 2015.  
Linking age, survival, and transit time distributions.  
*Water Resources Research*, **51**(10), 8316–8330.

Campbell, Gaylon S. 1985.

*Soil physics with BASIC: transport models for soil-plant systems.*

Vol. 14.

Elsevier.

Carmona, Facundo, Rivas, Raúl, & Caselles, Vicente. 2014.

Estimation of daytime downward longwave radiation under clear and cloudy skies conditions over a sub-humid region.

*Theoretical and applied climatology*, **115**(1-2), 281–295.

Carrera, J, & Medina, A. 1999.

A discussion on the calibration of regional groundwater models.

*In: International Workshop of EurAgEng's Field of Interest on Soil and Water, Leuven (Belgium), 24-26 Nov 1999.*

Wageningen Pers.

Carrera-Hernández, JJ, & Gaskin, SJ. 2007.

Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico.

*Journal of Hydrology*, **336**(3), 231–249.

Casulli, V, & Zanolli, P. 1998.

A conservative semi-implicit scheme for open channel flows.

*International Journal of Applied Science & Computations*, **5**(1-10), 98.

Cazorzi, F, & Dalla Fontana, G. 1996.

Snowmelt modelling by combining air temperature and a distributed radiation index.

*Journal of Hydrology*, **181**(1-4), 169–187.

Chambers, John M, & Hastie, TJ (eds.).

*"Linear Models." in Statistical Models in S.*

Chapman Hall.

Chap. 4, pages 95–144.

Clark, Martyn P, Slater, Andrew G, Rupp, David E, Woods, Ross A, Vrugt,

Jasper A, Gupta, Hoshin V, Wagener, Thorsten, & Hay, Lauren E. 2008.

Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models.

*Water Resources Research*, **44**(12).



- Clark, Martyn P, McMillan, Hilary K, Collins, Daniel BG, Kavetski, Dmitri, & Woods, Ross A. 2011.  
Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses.  
*Hydrological Processes*, **25**(4), 523–543.
- Clark, Martyn P, Nijssen, Bart, Lundquist, Jessica D, Kavetski, Dmitri, Rupp, David E, Woods, Ross A, Freer, Jim E, Gutmann, Ethan D, Wood, Andrew W, Brekke, Levi D, *et al.* . 2015a.  
A unified approach for process-based hydrologic modeling: 1. Modeling concept.  
*Water Resources Research*, **51**(4), 2498–2514.
- Clark, Martyn P, Nijssen, Bart, Lundquist, Jessica D, Kavetski, Dmitri, Rupp, David E, Woods, Ross A, Freer, Jim E, Gutmann, Ethan D, Wood, Andrew W, Gochis, David J, *et al.* . 2015b.  
A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies.  
*Water Resources Research*, **51**(4), 2515–2542.
- Coddington, Earl A, & Levinson, Norman. 1955.  
*Theory of ordinary differential equations*.  
Tata McGraw-Hill Education.
- Córdova, JoséR, & Rodríguez-Iturbe, Ignacio. 1983.  
Geomorphoclimatic estimation of extreme flow probabilities.  
*Journal of Hydrology*, **65**(1), 159–173.
- Cornaton, Fabien, & Perrochet, Pierre. 2006.  
Groundwater age, life expectancy and transit time distributions in advective–dispersive systems: 1. Generalized reservoir theory.  
*Advances in Water Resources*, **29**(9), 1267–1291.
- Corripio, Javier G. 2003.  
Vectorial algebra algorithms for calculating terrain parameters from DEMs and solar radiation modelling in mountainous terrain.  
*International Journal of Geographical Information Science*, **17**(1), 1–23.
- Corripio, Javier Gonzalez. 2002.  
*Modelling the energy balance of high altitude glacierised basins in the Central Andes*.  
Ph.D. thesis, University of Edinburgh.

Crawford, Todd M, & Duchon, Claude E. 1999.

An improved parameterization for estimating effective atmospheric emissivity for use in calculating daytime downwelling longwave radiation. *Journal of Applied Meteorology*, **38**(4), 474–480.

Cressie, Noel AC, & Cassie, Noel A. 1993.

*Statistics for spatial data*.  
Vol. 900.  
Wiley New York.

Cressman, George P. 1959.

An operational objective analysis system.  
*Mon. Wea. Rev.*, **87**(10), 367–374.

Creutin, JD, & Obled, Ch. 1982.

Objective analyses and mapping techniques for rainfall fields: an objective comparison.  
*Water Resources Research*, **18**(2), 413–431.

Cristea, Nicoleta C, Kampf, Stephanie K, & Burges, Stephen J. 2012.

Revised coefficients for Priestley-Taylor and Makkink-Hansen equations for estimating daily reference evapotranspiration.  
*Journal of Hydrologic Engineering*, **18**(10), 1289–1300.

Cudennec, C., Fouad, Y., Gatot, I. S., & Duchesne, J. 2004.

A geomorphological explanation of the unit hydrograph concept.  
*Hydrol. Process.*, **18**(4), 603–621.

Cudennec, C., Slimani, M., & Le Goulven, P. 2005.

Accounting for sparsely observed rainfall space-time variability in a rainfall-runoff model of a semiarid Tunisian basin.  
*Hydrolog. Sci. J.*, **50**(4), 617–630.

Cvetkovic, Vladimir. 2013.

How accurate is predictive modeling of groundwater transport? A case study of advection, macrodispersion, and diffusive mass transfer at the Forsmark site (Sweden).  
*Water Resources Research*, **49**(9), 5317–5327.

Cvetkovic, Vladimir, Carstens, Christoffer, Selroos, Jan-Olof, & Destouni, Georgia. 2012.

Water and solute transport along hydrological pathways.  
*Water resources research*, **48**(6).

- Dagan, Gedeon. 1984.  
Solute transport in heterogeneous porous formations.  
*Journal of fluid mechanics*, **145**, 151–177.
- Dagan, Gedeon. 1989.  
*Flow and Transport in Porous Formations*.  
Springer-Verlag Heidelberg Berlin New York.  
465 p.
- David, Olaf, Ascough, JC, Lloyd, Wes, Green, Tim R, Rojas, KW, Leavesley, George H, & Ahuja, Lajpat R. 2013.  
A software engineering perspective on environmental modeling framework design: The Object Modeling System.  
*Environmental Modelling & Software*, **39**, 201–213.
- de Lavenne, A., Boudhra, H., & Cudennec, C. 2015.  
Streamflow prediction in ungauged basins through geomorphology-based hydrograph transposition.  
*Hydrology Research*, **46**(2), 291–302.
- Delhez, Eric JM, Campin, Jean-Michel, Hirst, Anthony C, & Deleersnijder, Eric. 1999.  
Toward a general theory of the age in ocean modelling.  
*Ocean Modelling*, **1**(1), 17–27.
- Dickinson, Robert E. 1984.  
Modeling evapotranspiration for three-dimensional global climate models.  
*Climate processes and climate sensitivity*, 58–72.
- Dilley, AC, & O'brien, DM. 1998.  
Estimating downward clear sky long-wave irradiance at the surface from screen temperature and precipitable water.  
*Quarterly Journal of the Royal Meteorological Society*, **124**(549), 1391–1401.
- Dingman, S. L. 1994.  
*Physical Hydrology*.  
New York: Macmillan.
- D'Odorico, Paolo, & Rigon, Riccardo. 2003.  
Hillslope and channel contributions to the hydrologic response.  
*Water resources research*, **39**(5).

Dooge, J. C. I. 2003.

*Linear Theory of Hydrologic Systems.*

Vol. 1.

Katlenburg-Lindau, Germany: EGU Reprint Series.

Dooge, James. 1973.

*Linear theory of hydrologic systems.*

Agricultural Research Service, US Department of Agriculture.

Dormand, John R, & Prince, Peter J. 1980.

A family of embedded Runge-Kutta formulae.

*Journal of computational and applied mathematics*, **6**(1), 19–26.

Duan, QY, Gupta, Vijai K, & Sorooshian, Soroosh. 1993.

Shuffled complex evolution approach for effective and efficient global minimization.

*Journal of optimization theory and applications*, **76**(3), 501–521.

Duffy, Christopher J. 1996.

A Two-State Integral-Balance Model for Soil Moisture and Groundwater Dynamics in Complex Terrain.

*Water resources research*, **32**(8), 2421–2434.

Duffy, Christopher J. 2010.

Dynamical modelling of concentration–age–discharge in watersheds.

*Hydrological processes*, **24**(12), 1711–1718.

Dunne, Thomas. 1978.

Field studies of hillslope flow processes.

*Hillslope hydrology*, 389–227.

Eberhart, Russell, & Kennedy, James. 1995.

A new optimizer using particle swarm theory.

Pages 39–43 of: *Micro Machine and Human Science, 1995. MHS'95.*,

*Proceedings of the Sixth International Symposium on.*

IEEE.

Efron, Bradley, & Efron, B. 1982.

*The jackknife, the bootstrap and other resampling plans.*

Vol. 38.

SIAM.

- Ellis, Brian, Stylos, Jeffrey, & Myers, Brad. 2007.  
The factory pattern in API design: A usability evaluation.  
*Pages 302–312 of: Proceedings of the 29th international conference on Software Engineering.*  
IEEE Computer Society.
- Fairfield, John, & Leymarie, Pierre. 1991.  
Drainage networks from grid digital elevation models.  
*Water Resources Research*, **27**(5), 709–717.
- Fatichi, Simone, Vivoni, Enrique R, Ogden, Fred L, Ivanov, Valeriy Y, Mirus, Benjamin, Gochis, David, Downer, Charles W, Camporese, Matteo, Davison, Jason H, Ebel, Brian, *et al.* . 2016.  
An overview of current applications, challenges, and future trends in distributed process-based models in hydrology.  
*Journal of Hydrology*, **537**, 45–60.
- Fenicia, F, Savenije, HHG, Matgen, P, & Pfister, L. 2006.  
Is the groundwater reservoir linear? Learning from data in hydrological modelling.  
*Hydrology and Earth System Sciences Discussions*, **10**(1), 139–150.
- Fenicia, Fabrizio, Savenije, Hubert H. G., Matgen, Patrick, & Pfister, Laurent. 2008.  
Understanding catchment behavior through stepwise model concept improvement.  
*Water Resour. Res.*, **44**(1), W01402.
- Fenicia, Fabrizio, Kavetski, Dmitri, & Savenije, Hubert HG. 2011.  
Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development.  
*Water Resources Research*, **47**(11).
- Flerchinger, GN. 2000.  
*The Simultaneous Heat and Water (SHAW) Model: Technical Documentation, Northwest Watershed Research Center, USDA Agricultural Research Service, Boise.*  
Tech. rept. Idaho, Technical Report NWRC 2000-09, 37 pp.
- Flerchinger, GN, Xaio, Wei, Marks, Danny, Sauer, TJ, & Yu, Qiang. 2009.  
Comparison of algorithms for incoming atmospheric long-wave radiation.  
*Water resources research*, **45**(3).

- Forman, Ira R, Forman, Nate, & Ibm, John Vlissides. 2004.  
Java reflection in action.
- Formetta, G., Mantilla, R., Franceschi, S., Antonello, A., & Rigon, R. 2011.  
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.
- Formetta, G, Kampf, SK, David, O, & Rigon, R. 2013a.  
The Cache la Poudre river basin snow water equivalent modeling with NewAge-JGrass.  
*Geoscientific Model Development Discussions*, 6(3).
- Formetta, G, Rigon, R, Chávez, JL, & David, O. 2013b.  
Modeling shortwave solar radiation using the JGrass-NewAge system.  
*Geoscientific Model Development*, 6(4), 915–928.
- Formetta, G, Antonello, A, Franceschi, S, David, O, & Rigon, R. 2014a.  
Hydrological modelling with components: A GIS-based open-source framework.  
*Environmental Modelling & Software*, 55, 190–200.
- Formetta, G, David, O, & Rigon, R. 2014b.  
Testing site-specific parameterizations of longwave radiation integrated in a GIS-based hydrological model.  
*International Environmental Modelling and Software Society (iEMSs) 7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA, Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)*.
- Formetta, G., Bancheri, M., David, O., & Rigon, R. 2016.  
Site specific parameterizations of longwave radiation.  
*Hydrology and Earth System Sciences Discussions*, 2016, 1–22.
- Freeman, Eric, Robson, Elisabeth, Bates, Bert, & Sierra, Kathy. 2004.  
*Head First Design Patterns: A Brain-Friendly Guide*.  
" O'Reilly Media, Inc." .
- Gamma, Erich, Helm, Richard, Johnson, Ralph, & Vlissides, John. 1994.  
*Design patterns: elements of*.
- Gardner, Henry, & Manduchi, Gabriele. 2007.  
State.

- Design Patterns for e-Science*, 185–190.
- Gash, JHC. 1979.  
An analytical model of rainfall interception by forests.  
*Quarterly Journal of the Royal Meteorological Society*, **105**(443), 43–55.
- Gash, John HC, Lloyd, CR, & Lachaud, G. 1995.  
Estimating sparse forest rainfall interception with an analytical model.  
*Journal of Hydrology*, **170**(1-4), 79–86.
- Goovaerts, P. 2000.  
Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall.  
*Journal of hydrology*, **228**(1), 113–129.
- Goovaerts, Pierre. 1997.  
*Geostatistics for natural resources evaluation*.  
Oxford university press.
- Goovaerts, Pierre. 1999.  
Geostatistics in soil science: state-of-the-art and perspectives.  
*Geoderma*, **89**(1), 1–45.
- Gray, William G, Leijnse, Anton, Kolar, Randall L, & Blain, Cheryl A. 1993.  
*Mathematical tools for changing scale in the analysis of physical systems*.  
CRC Press.
- Grayson, Rodger, & Blöschl, Günter. 2001.  
*Spatial patterns in catchment hydrology: observations and modelling*.  
CUP Archive.
- Gregersen, JB, Gijsbers, PJA, & Westen, SJP. 2007.  
OpenMI: Open modelling interface.  
*Journal of hydroinformatics*, **9**(3), 175–191.
- Grimaldi, S., Petroselli, A., & Nardi, F. 2012.  
A parsimonious geomorphological unit hydrograph for rainfall-runoff modelling in small ungauged basins.  
*Hydrolog. Sci. J.*, **57**(1), 73–83.
- Grimaldi, Salvatore, Petroselli, Andrea, Alonso, Gustavo, & Nardi, Fernando. 2010.

Flow time estimation with spatially variable hillslope velocity in ungauged basins.

*Adv. Water. Resour.*, **33**(10, Sp. Iss. SI), 1216–1223.

Groovy, Apache. 2014.

*Groovy. A multi-faceted language for the Java platform.*

<http://groovy-lang.org>.

Gupta, Hoshin V, Kling, Harald, Yilmaz, Koray K, & Martinez, Guillermo F. 2009.

Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling.

*Journal of Hydrology*, **377**(1), 80–91.

Gupta, Vijay K., & Mesa, Oscar J. 1988.

Runoff generation and hydrologic response via channel network geomorphology – Recent progress and open problems.

*J. Hydrol.*, **102**(1), 3–28.

Gupta, Vijay K., & Waymire, Ed. 1983.

On the formulation of an analytical approach to hydrologic response and similarity at the basin scale.

*J. Hydrol.*, **65**(1), 95–123.

Gupta, Vijay K., Waymire, Ed, & Wang, C. T. 1980.

A representation of an instantaneous unit hydrograph from geomorphology.

*Water Resour. Res.*, **16**(5), 855–862.

Haberlandt, Uwe. 2007.

Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event.

*Journal of Hydrology*, **332**(1), 144–157.

Hall, M. J., Zaki, A. F., & Shahin, M. M. A. 2001.

Regional analysis using the Geomorphoclimatic Instantaneous Unit Hydrograph.

*Hydrol. Earth Syst. Sci.*, **5**(1), 93–102.

Harman, Ciaran. 2015a.

Internal versus external controls on age variability: Definitions, origins and implications in a changing climate.



- In: 2015 AGU Fall Meeting.*  
Agu.
- Harman, Ciaran J. 2015b.  
Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed.  
*Water Resources Research*, **51**(1), 1–30.
- Hatfield, JL, Reginato, R JI, & Idso, SB. 1983.  
Comparison of long-wave radiation calculation methods over the United States.  
*Water Resources Research*, **19**(1), 285–288.
- Hay, Lauren E, Leavesley, George H, Clark, Martyn P, Markstrom, Steve L, Viger, Roland J, & Umemoto, Makiko. 2006a.  
*Step wise, multiple objective calibration of a hydrologic model for a snowmelt dominated basin1.*
- Hay, Lauren E, Leavesley, George H, & Clark, MP. 2006b.  
Use of remotely-sensed snow covered area in watershed model calibration for the Sprague River, Oregon.  
*In: Joint 8th Federal Interagency Sedimentation Conference and 3rd Federal Interagency Hydrologic Modeling Conference, Reno, Nevada.*
- Henderson, Francis Martin. 1966.  
*Open channel flow.*  
Macmillan, New York.  
522 pp.
- Henderson-Sellers, A, & Wilson, MF. 1983.  
Surface albedo data for climatic modeling.  
*Reviews of Geophysics*, **21**(8), 1743–1778.
- Hevesi, Joseph A, Istok, Jonathan D, & Flint, Alan L. 1992.  
Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: structural analysis.  
*Journal of applied meteorology*, **31**(7), 661–676.
- Hock, Regine. 1999.  
A distributed temperature-index ice-and snowmelt model including potential direct solar radiation.  
*Journal of Glaciology*, **45**(149), 101–111.

Hrachowitz, M., & Clark, M. P. 2017.

HESS Opinions: The complementary merits of competing modelling philosophies in hydrology.  
*Hydrology and Earth System Sciences*, **21**(8), 3953–3973.

Hrachowitz, M, Soulsby, C, Tetzlaff, D, Malcolm, IA, & Schoups, G. 2010.

Gamma distribution models for transit time estimation in catchments: Physical interpretation of parameters and implications for time-variant transit time assessment.  
*Water Resources Research*, **46**(10).

Hrachowitz, Markus, Savenije, H, Bogaard, TA, Tetzlaff, D, & Soulsby, C. 2013.

What can flux tracking teach us about water age distribution patterns and their temporal dynamics?  
*Hydrology and Earth System Sciences*, *17* (2), 2013.

Hrachowitz, Markus, Fovet, Ophelie, Ruiz, Laurent, Euser, T, Gharari, S,

Nijzink, R, Freer, J, Savenije, HHG, & Gascuel-Oudou, C. 2014.  
Process consistency in models: The importance of system signatures, expert knowledge, and process complexity.  
*Water Resources Research*, **50**(9), 7445–7469.

Hrachowitz, Markus, Benettin, Paolo, Breukelen, Boris M, Fovet, Ophelie,

Howden, Nicholas JK, Ruiz, Laurent, Velde, Ype, & Wade, Andrew J. 2016.  
Transit times–The link between hydrology and water quality at the catchment scale.  
*Wiley Interdisciplinary Reviews: Water*.

Hutchinson, MF. 1995.

Interpolating mean rainfall using thin plate smoothing splines.  
*International journal of geographical information systems*, **9**(4), 385–403.

Idso, Sherwood B. 1981.

A set of equations for full spectrum and 8-to 14- $\mu\text{m}$  and 10.5-to 12.5- $\mu\text{m}$  thermal radiation from cloudless skies.  
*Water resources research*, **17**(2), 295–304.

Idso, Sherwood B, & Jackson, Ray D. 1969.

Thermal radiation from the atmosphere.  
*Journal of Geophysical Research*, **74**(23), 5397–5403.

Isaaks, Edward H, & Srivastava, R Mohan. 1989.

- An introduction to applied geostatistics.*  
Vol. 561.  
Oxford university press New York.
- Iziomon, MOSES G, Mayer, HELMUT, & Matzarakis, ANDREAS. 2003.  
Downward atmospheric longwave irradiance under clear and cloudy skies:  
Measurement and parameterization.  
*Journal of Atmospheric and Solar-Terrestrial Physics*, **65**(10), 1107–1116.
- Jackson, JJ. 1975.  
Relationships between rainfall parameters and interception by tropical  
forest.  
*Journal of Hydrology*, **24**(3-4), 215–238.
- Jacobs, JD. 1978.  
Radiation climate of Broughton Island.  
*Energy budget studies in relation to fast-ice breakup processes in Davis  
Strait*, **26**, 105–120.
- Jarvis, Claire H, & Stuart, Neil. 2001.  
A comparison among strategies for interpolating maximum and minimum  
daily air temperatures. Part I: The selection of „Äguiding,Ä topographic  
and land cover variables.  
*Journal of Applied Meteorology*, **40**(6), 1060–1074.
- Juszak, I, & Pellicciotti, F. 2013.  
A comparison of parameterizations of incoming longwave radiation over  
melting glaciers: model robustness and seasonal variability.  
*Journal of Geophysical Research: Atmospheres*, **118**(8), 3066–3084.
- Kavetski, Dmitri, & Fencia, Fabrizio. 2011.  
Elements of a flexible approach for conceptual hydrological modeling: 2.  
Application and experimental insights.  
*Water Resources Research*, **47**(11).
- Keding, I. 1989.  
*Klimatologische Untersuchung ueber die atmosphaerische Gegenstrahlung  
und Vergleich vom Berechnungsverfahren anhand langjaehriger Messungen  
im Oberrheintal.*  
Offenbach am Main: Selbstverlag des Deutschen Wetterdienstes.
- Kelliher, FM, Ross, DJ, Law, BE, Baldocchi, DD, & Rodda, NJ. 2004.

Limitations to carbon mineralization in litter and mineral soil of young and old ponderosa pine forests.

*Forest Ecology and Management*, **191**(1), 201–213.

Key, Jeffrey R, & Schweiger, Axel J. 1998.

Tools for atmospheric radiative transfer: Streamer and FluxNet.

*Computers & Geosciences*, **24**(5), 443–451.

Kirchner, James W. 2006.

Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology.

*Water Resources Research*, **42**(3).

Kirchner, James W. 2009.

Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward.

*Water Resources Research*, **45**(2).

Kirchner, James W. 2016.

Aggregation in environmental systems–Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments.

*Hydrology and Earth System Sciences*, **20**(1), 279–297.

Kirchner, James W, Feng, Xiahong, & Neal, Colin. 2000.

Fractal stream chemistry and its implications for contaminant transport in catchments.

*Nature*, **403**(6769), 524–527.

Kirkby, M. J. 1976.

Tests of the random network model, and its application to basin hydrology.

*Earth Surface Processes*, **1**(3), 197–212.

Kitanidis, Peter K. 1997.

*Introduction to geostatistics: applications in hydrogeology*.

Cambridge University Press.

Klemeš, Vit. 1986.

Operational testing of hydrological simulation models.

*Hydrological Sciences Journal*, **31**(1), 13–24.

Kneizys, Francis X, Shettle, EP, Abreu, LW, Chetwynd, JH, & Anderson, GP. 1988.

- Users guide to LOWTRAN 7.*  
Tech. rept. DTIC Document.
- König-Langlo, Gert, & Augstein, Ernst. 1994.  
Parameterization of the downward long-wave radiation at the Earth's surface in polar regions.  
*Meteorologische zeitschrift, NF 3, Jg. 1994, H. 6, 343–347.*
- Konzelmann, Thomas, van de Wal, Roderik SW, Greuell, Wouter, Bintanja, Richard, Henneken, Edwin AC, & Abe-Ouchi, Ayako. 1994.  
Parameterization of global and longwave incoming radiation for the Greenland Ice Sheet.  
*Global and Planetary change, 9(1), 143–164.*
- Leek, Jeff. 2013.  
*Statisticians and computer scientists - if there is no code, there is no paper.*  
[http://simplystatistics.org/2013/01/23/statisticians-and-computer-scientists-if-there-is-no-code-](http://simplystatistics.org/2013/01/23/statisticians-and-computer-scientists-if-there-is-no-code/)
- Leigh Jr, Egbert Giles. 1999.  
*Tropical Forest Ecology: A View from Barro Colorado Island: A View from Barro Colorado Island.*  
Oxford University Press.
- Leopold, Luna B., & Maddock, Thomas. 1953.  
The Hydraulic Geometry of Stream Channels and Some Physiographic Implications.  
*US Geol. Surv. Prof. Pap., 252, 56.*
- Levin, Simon A Simon A. 1999.  
*Fragile dominion complexity and the commons.*
- Lewis, D, Singer, MJ, Dahlgren, RA, & Tate, KW. 2000.  
Hydrology in a California oak woodland watershed: a 17-year study.  
*Journal of Hydrology, 240(1), 106–117.*
- Li, Jin, & Heap, Andrew D. 2011.  
A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors.  
*Ecological Informatics, 6(3), 228–241.*
- Lloyd, CD. 2005.

Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain.

*Journal of Hydrology*, **308**(1), 128–150.

Lloyd, Wes, David, Olaf, Ascough, JC, Rojas, Ken W, Carlson, Jack R, Leavesley, George H, Krause, Peter, Green, Tim R, & Ahuja, Lajpat R. 2011.

Environmental modeling framework invasiveness: Analysis and implications.

*Environmental modelling & software*, **26**(10), 1240–1250.

Ly, Sarann, Charles, Catherine, & Degre, Aurore. 2011.

Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium.

*Hydrology & Earth System Sciences*, **15**(7).

Ly, Sarann, Charles, Catherine, & Degré, Aurore. 2013.

Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review.

Mac Lane, Saunders. 2013.

*Categories for the working mathematician*.

Vol. 5.

Springer Science & Business Media.

MacDonell, Shelley, Nicholson, Lindsey, & Kinnard, Christophe. 2013.

Parameterisation of incoming longwave radiation over glacier surfaces in the semiarid Andes of Chile.

*Theoretical and applied climatology*, **111**(3-4), 513–528.

Manfreda, S, Fiorentino, M, & Iacobellis, V. 2005.

DREAM: a distributed model for runoff, evapotranspiration, and antecedent soil moisture simulation.

*Advances in Geosciences*, **2**, 31–39.

Manfreda, S, Funicelli, L, & Mancusi, L. 2012.

Previsione idrologica per la gestione degli impianti idroelettrici.

*Atti del Convegno di idraulica e Costruzioni Idrauliche*.

Manfreda, S, Mita, L, Dal Sasso, S, Samela, C, Mancusi, L, & Fiorentino, M. in review.

- Exploiting Physical Information for the Calibration of the Lumped 2 Model AD2.  
*Hydrological Processes*.
- Mantilla, Ricardo, & Gupta, Vijay K. 2005.  
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.
- Martin, Jay D, & Simpson, Timothy W. 2003.  
A study on the use of kriging models to approximate deterministic computer models.  
*Pages 2–6 of: Proceedings of DETC*, vol. 3.
- Matheron, Georges. 1981.  
Splines and kriging: their formal equivalence.  
*Down-to-earth statistics: solutions looking for geological problems*, **8**, 77–95.
- Maykut, Gary A, & Church, Phil E. 1973.  
Radiation climate of Barrow Alaska, 1962–66.  
*Journal of Applied Meteorology*, **12**(4), 620–628.
- McDonnell, Jeffrey J, & Beven, Keith. 2014.  
Debates–The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph.  
*Water Resources Research*, **50**(6), 5342–5350.
- McMillan, Hilary, Tetzlaff, Doerthe, Clark, Martyn, & Soulsby, Chris. 2012.  
Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach.  
*Water Resources Research*, **48**(5).
- Mernik, Marjan, Heering, Jan, & Sloane, Anthony M. 2005.  
When and how to develop domain-specific languages.  
*ACM computing surveys (CSUR)*, **37**(4), 316–344.
- Mesa, OscarJ., & Mifflin, EdwardR. 1986.  
On the Relative Role of Hillslope and Network Geometry in Hydrologic Response.

Pages 1–17 of: Gupta, V.K., Rodríguez-Iturbe, I., & Wood, E.F. (eds), *Scale Problems in Hydrology*.

Water Science and Technology Library, vol. 6.

Springer Netherlands.

Mitášová, Helena, & Mitáš, Lubos. 1993.

Interpolation by regularized spline with tension: I. Theory and implementation.

*Mathematical geology*, **25**(6), 641–655.

Mitchell, V. Grace, McMahon, Thomas A., & Mein, Russell G. 2003.

Components of the Total Water Balance of an Urban Catchment.

*Environmental Management*, **32**(6), 735–746.

Monteith, John Lennox, & Unsworth, MH. 1990.

*Principles of Environmental Physics*.

Butterworth-Heinemann.

Moore, RJ. 1985.

The probability-distributed principle and runoff production at point and basin scales.

*Hydrological Sciences Journal*, **30**(2), 273–297.

Morse, PM, & Feshbach, H. 1953.

*Methods of theoretical physics*.

Cambridge University Press.

Murata, Tadao. 1989.

Petri nets: Properties, analysis and applications.

*Proceedings of the IEEE*, **77**(4), 541–580.

Naden, P, Broadhurst, P, Tauveron, N., & Walker, A. 1999.

River routing at the continental scale: use of globally-available data and an a priori method of parameter estimation.

*Hydrology and Earth System Sciences*, **3**(1), 109–123.

Naden, P. S. 1992.

Spatial variability in flood estimation for large catchments: the exploitation of channel network structure.

*Hydrolog. Sci. J.*, **37**(1), 53–71.

Nash, J Eamonn, & Sutcliffe, Jonh V. 1970.



- River flow forecasting through conceptual models part I—A discussion of principles.  
*Journal of hydrology*, **10**(3), 282–290.
- Nicotina, L., Alessi Celegon, E., Rinaldo, A., & Marani, M. 2008.  
On the impact of rainfall patterns on the hydrologic response.  
*Water Resources Research*, **44**(12), W12401.
- Niemelä, Sami, Räisänen, Petri, & Savijärvi, Hannu. 2001.  
Comparison of surface radiative flux parameterizations: Part I: Longwave radiation.  
*Atmospheric Research*, **58**(1), 1–18.
- Niemi, Antti J. 1977.  
Residence time distributions of variable flow processes.  
*The International Journal of Applied Radiation and Isotopes*, **28**(10), 855–860.
- Norbiato, Daniele, Borga, Marco, Merz, Ralf, Blöschl, Günther, & Carton, Alberto. 2009.  
Controls on event runoff coefficients in the eastern Italian Alps.  
*Journal of Hydrology*, **375**(3), 312–325.
- Oishi, A Christopher, Oren, Ram, Novick, Kimberly A, Palmroth, Sari, & Katul, Gabriel G. 2010.  
Interannual invariability of forest evapotranspiration and its consequence to water flow downstream.  
*Ecosystems*, **13**(3), 421–436.
- Park, Gi-Hyeon, Gao, Xiaogang, & Sorooshian, Soroosh. 2008.  
Estimation of surface longwave radiation components from ground-based historical net radiation and weather data.  
*Journal of Geophysical Research: Atmospheres (1984–2012)*, **113**(D4).
- Pattison, Ian, Lane, Stuart N., Hardy, Richard J., & Reaney, Sim M. 2014.  
The role of tributary relative timing and sequencing in controlling large floods.  
*Water Resources Research*, **50**(7), 5444–5458.
- Penna, D, Meerveld, HJ, Oliviero, O, Zuecco, G, Assendelft, RS, Dalla Fontana, G, & Borga, M. 2015.

Seasonal changes in runoff generation in a small forested mountain catchment.

*Hydrological processes*, **29**(8), 2027–2042.

Phillips, Donald L, Dolph, Jayne, & Marks, Danny. 1992.

A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain.

*Agricultural and Forest Meteorology*, **58**(1-2), 119–141.

Plüss, Christian, & Ohmura, Atsumu. 1997.

Longwave radiation on snow-covered mountainous surfaces.

*Journal of Applied Meteorology*, **36**(6), 818–824.

Prata, AJ. 1996.

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.

Priestley, CHB, & Taylor, RJ. 1972.

On the assessment of surface heat flux and evaporation using large-scale parameters.

*Monthly weather review*, **100**(2), 81–92.

Prudhomme, Christel, & Reed, Duncan W. 1999.

Mapping extreme rainfall in a mountainous region using geostatistical techniques: a case study in Scotland.

*International Journal of Climatology*, **19**(12), 1337–1356.

Queloz, Pierre, Carraro, Luca, Benettin, Paolo, Botter, Gianluca, Rinaldo, Andrea, & Bertuzzo, Enrico. 2015.

Transport of fluorobenzoate tracers in a vegetated hydrologic control volume: 2. Theoretical inferences and modeling.

*Water Resources Research*, **51**(4), 2793–2806.

Refsgaard, J. C. 1995.

MIKE SHE.

*Computer Models of Watershed Hydrology*, **1113**.

Reggiani, Paolo, Sivapalan, Murugesu, & Hassanizadeh, S Majid. 1998.

A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics.

*Advances in Water Resources*, **22**(4), 367–398.

- Rigon, R, Ghesla, E, Tiso, C, & Cozzini, A. 2006.  
The HORTON machine: a system for DEM analysis The reference manual.  
*Università degli Studi di Trento*.
- Rigon, R., D'Odorico, P., & Bertoldi, G. 2011.  
The geomorphic structure of the runoff peak.  
*Hydrol. Earth Syst. Sci.*, **15**(6), 1853–1863.
- Rigon, R., Bancheri, M., & Green, T. R. 2016a.  
Age-ranked hydrological budgets and a travel time description of  
catchment hydrology.  
*Hydrology and Earth System Sciences Discussions*, **2016**, 1–22.
- Rigon, Riccardo. 2011.  
*Going Beyond the Present Stato-of-Art in hydrological Modeling. My point of  
view*.  
[http://abouthydrology.blogspot.it/2011/03/  
going-beyond-present-stato-of-art-in.html](http://abouthydrology.blogspot.it/2011/03/going-beyond-present-stato-of-art-in.html).
- Rigon, Riccardo. 2014.  
*JGrass-NewAGE history - Version zero and version one*.  
[http://abouthydrology.blogspot.it/2014/11/  
jgrass-newage-history-version-zero-and.html](http://abouthydrology.blogspot.it/2014/11/jgrass-newage-history-version-zero-and.html).
- Rigon, Riccardo, Serafin, Francesco, & Bancheri, Marialaura. 2015.  
*Theory and Practice of Reproducible Research*.  
[http://abouthydrology.blogspot.it/2015/07/  
theory-and-practice-of-reproducible.html](http://abouthydrology.blogspot.it/2015/07/theory-and-practice-of-reproducible.html).
- Rigon, Riccardo, Bancheri, Marialaura, Formetta, Giuseppe, & de Lavenne,  
Alban. 2016b.  
The geomorphological unit hydrograph from a historical-critical  
perspective.  
*Earth Surface Processes and Landforms*, **41**(1), 27–37.
- Rigon, Riccardo, Bancheri, Marialaura, & Serafin, Francesco. 2016c.  
*Reservoirology3*.  
[http://abouthydrology.blogspot.it/2016/11/  
reservoirology-3.html](http://abouthydrology.blogspot.it/2016/11/reservoirology-3.html).
- Rinaldo, A., Vogel, G. K., Rigon, R., & Rodriguez-Iturbe, I. 1995.  
Can one gauge the shape of a basin?  
*Water Resour. Res.*, **31**(4), 1119–1127.

- Rinaldo, A, Beven, Keith J, Bertuzzo, E, Nicotina, L, Davies, Jessica, Fiori, A, Russo, D, & Botter, G. 2011.  
Catchment travel time distributions and water flow in soils.  
*Water resources research*, **47**(7).
- Rinaldo, Andrea, & Rodríguez-Iturbe, Ignacio. 1996.  
Geomorphological theory of the hydrological response.  
*Hydrol. Process.*, **10**(6), 803–829.
- Rinaldo, Andrea, Marani, Alessandro, & Rigon, Riccardo. 1991.  
Geomorphological dispersion.  
*Water Resour. Res.*, **27**(4), 513–525.
- Rinaldo, Andrea, Benettin, Paolo, Harman, Ciaran J, Hrachowitz, Markus, McGuire, Kevin J, Van Der Velde, Ype, Bertuzzo, Enrico, & Botter, Gianluca. 2015.  
Storage selection functions: A coherent framework for quantifying how catchments store and release water and solutes.  
*Water Resources Research*, **51**(6), 4840–4847.
- Rizzoli, AE, Leavesley, G, Ascough, JC, Argent, RM, Athanasiadis, IN, Brillhante, V, Claeys, FHA, David, O, Donatelli, M, Gijbbers, P, *et al.* . 2008.  
Chapter seven integrated modelling frameworks for environmental assessment and decision support.  
*Developments in Integrated Environmental Assessment*, **3**, 101–118.
- Robeson, Scott Michael. 1992.  
Spatial interpolation, network bias, and terrestrial air temperature variability.
- Robinson, Justin S., Sivapalan, Murugesu, & Snell, John D. 1995.  
On the relative roles of hillslope processes, channel routing, and network geomorphology in the hydrologic response of natural catchments.  
*Water Resour. Res.*, **31**(12), 3089–3101.
- Rodríguez-Iturbe, Ignacio, & Rinaldo, Andrea. 2001.  
*Fractal river basins: chance and self-organization*.  
Cambridge University Press.
- Rodríguez-Iturbe, Ignacio, González-Sanabria, Marcelo, & Bras, Rafael L. 1982.  
A geomorphoclimatic theory of the instantaneous unit hydrograph.  
*Water Resour. Res.*, **18**(4), 877–886.

- Rodríguez-Iturbe, Ignacio, Porporato, A, Ridolfi, Luca, Isham, V, & Coxi, DR. 1999.  
Probabilistic modelling of water balance at a point: the role of climate, soil and vegetation.  
*Pages 3789–3805 of: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 455.  
The Royal Society.
- Rodríguez-Iturbe, Ignacio, & ValdÈs, Juan B. 1979a.  
The geomorphologic structure of hydrologic response.  
*Water Resour. Res.*, **15**(6), 1409–1420.
- Rodríguez-Iturbe, Ignacio, & ValdÈs, Juan B. 1979b.  
The geomorphologic structure of hydrologic response.  
*Water Resources Research*, **15**(6), 1409–1420.
- Romano, N, Palladino, M, & Chirico, GB. 2011.  
Parameterization of a bucket model for soil-vegetation-atmosphere modeling under seasonal climatic regimes.  
*Hydrology and Earth System Sciences*, **15**(12), 3877–3893.
- Ross, Cecil Napier. 1921.  
The calculation of flood discharges by the use of a time contour plan.  
*Transactions of the Institution of Engineers, Australia*, **2**, 85–92.
- Rotenberg, E, Mamane, Y, & Joseph, JH. 1998.  
Long wave radiation regime in vegetation-parameterisations for climate research.  
*Environmental modelling & software*, **13**(3), 361–371.
- Rouson, Damian, Xia, Jim, & Xu, Xiaofeng. 2011.  
*Scientific software design: the object-oriented way*.  
Cambridge University Press.
- Ruddell, Benjamin L, & Kumar, Praveen. 2009.  
Ecohydrologic process networks: 1. Identification.  
*Water Resources Research*, **45**(3).
- Rutter, AJ, Kershaw, KA, Robins, PC, & Morton, AJ. 1971.  
A predictive model of rainfall interception in forests, 1. Derivation of the model from observations in a plantation of Corsican pine.  
*Agricultural Meteorology*, **9**, 367–384.

Saco, Patricia M., & Kumar, Praveen. 2002a.

Kinematic dispersion in stream networks 1. Coupling hydraulic and network geometry.

*Water Resour. Res.*, **38**(Nov.), 14 PP.

Saco, Patricia M., & Kumar, Praveen. 2002b.

Kinematic dispersion in stream networks 2. Scale issues and self-similar network organization.

*Water Resour. Res.*, **38**(11), 1245.

Santhi, C, Arnold, Jeffrey G, Williams, Jimmy R, Dugas, William A, Srinivasan, Raghavan, & Hauck, Larry M. 2001.

Validation of the swat model on a large river basin with point and nonpoint sources.

*JAWRA Journal of the American Water Resources Association*, **37**(5), 1169–1188.

Schmucki, Edgar, Marty, Christoph, Fierz, Charles, & Lehning, Michael. 2014.

Evaluation of modelled snow depth and snow water equivalent at three contrasting sites in Switzerland using SNOWPACK simulations driven by different meteorological data input.

*Cold Regions Science and Technology*, **99**, 27–37.

Seibert, Jan, & McDonnell, Jeffrey J. 2002.

On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration.

*Water Resources Research*, **38**(11).

Serafin, Francesco. 2016.

*Creating a new infrastructure for GEOtop 3.0.*

Ph.D. thesis, University of Trento.

Serafin, Francesco, Rigon, Riccardo, Cordano, Emanuele, & Formetta, Giuseppe. 2014.

*Master Thesis in Environmental and Land Engineering: Patterns for the application of modern informatics to the integration of PDEs: the case of the Boussinesq Equation.*

Serafin, Francesco, Bancheri, Marialaura, Rigon, Riccardo, & David, Olaf. 2016.

A Java binary tree data structure for environmental modelling.

*In: 8th International congress on environmental modelling and software, Toulouse, France.*

- Sicart, Jean-Emmanuel, Pomeroy, JW, Essery, RLH, & Bewley, D. 2006.  
Incoming longwave radiation to melting snow: observations, sensitivity and estimation in northern environments.  
*Hydrological processes*, **20**(17), 3697–3708.
- Sivapalan, M. 2003.  
Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection?  
*Hydrol. Process.*, **17**(5), 1037–1041.
- Sivapalan, M., Jothityangkoon, C., & Menabde, M. 2002.  
Linearity and nonlinearity of basin response as a function of scale: Discussion of alternative definitions.  
*Water Resour. Res.*, **38**, 5.
- Sivapalan, Murugesu, Blöschl, Günter, Zhang, Lu, & Vertessy, Rob. 2003.  
Downward approach to hydrological prediction.  
*Hydrological processes*, **17**(11), 2101–2111.
- Smith, Michael B., Koren, Victor I., Zhang, Ziya, Reed, Seann M., Pan, Jeng-J., & Moreda, Fekadu. 2004.  
Runoff response to spatial variability in precipitation: an analysis of observed data.  
*Journal of Hydrology*, **298**(1), 267 – 286.  
The Distributed Model Intercomparison Project (DMIP).
- Snell, John D., & Sivapalan, Murugesu. 1994.  
On geomorphological dispersion in natural catchments and the geomorphological unit hydrograph.  
*Water Resour. Res.*, **30**(7), 2311–2323.
- Soulsby, C, Birkel, C, Geris, J, Dick, J, Tunaley, C, & Tetzlaff, D. 2015.  
Stream water age distributions controlled by storage dynamics and nonlinear hydrologic connectivity: Modeling with high-resolution isotope data.  
*Water Resources Research*, **51**(9), 7759–7776.
- SourceMaking. 2014.  
*Design Patterns*.  
[https://sourcemaking.com/design\\_patterns](https://sourcemaking.com/design_patterns).

- Stahl, K, Moore, RD, Floyer, JA, Asplin, MG, & McKendry, IG. 2006.  
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.
- Sugita, M, & Brutsaert, W. 1993a.  
Comparison of land surface temperatures derived from satellite observations with ground truth during FIFE.  
*International Journal of Remote Sensing*, **14**(9), 1659–1676.
- Sugita, Michiaki, & Brutsaert, Wilfried. 1993b.  
Cloud effect in the estimation of instantaneous downward longwave radiation.  
*Water Resources Research*, **29**(3), 599–605.
- Swinbank, W CQJR. 1963.  
Long-wave radiation from clear skies.  
*Quarterly Journal of the Royal Meteorological Society*, **89**(381), 339–348.
- Tabios, Guillermo Q, & Salas, Jose D. 1985.  
*A comparative analysis of techniques for spatial interpolation of precipitation1*.
- Tague, Christina, & Dugger, Aubrey L. 2010.  
Ecohydrology and climate change in the mountains of the Western USA—A review of research and opportunities.  
*Geography Compass*, **4**(11), 1648–1663.
- Tarboton, David G., Bras, Rafael L., & Rodriguez-Iturbe, Ignacio. 1991.  
On the extraction of channel networks from digital elevation data.  
*Hydrological Processes*, **5**(1), 81–100.
- Tetzlaff, D, McDonnell, JJ, Uhlenbrook, S, McGuire, KJ, Bogaart, PW, Naef, F, Baird, AJ, Dunn, SM, & Soulsby, C. 2008.  
Conceptualizing catchment processes: simply too complex?  
*Hydrological Processes*, **22**(11), 1727–1730.
- Therrien, R, McLaren, RG, Sudicky, EA, & Panday, SM. 2010.  
HydroGeoSphere: a three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport.  
*Groundwater Simulations Group, University of Waterloo, Waterloo, ON*.



- Thiessen, Alfred H. 1911.  
Precipitation averages for large areas.  
*Monthly weather review*, **39**(7), 1082–1089.
- Todini, Ezio. 2001.  
Influence of parameter estimation uncertainty in Kriging: Part  
1-Theoretical Development.  
*Hydrology and Earth System Sciences Discussions*, **5**(2), 215–223.
- Turton, Ian. 2008.  
Geo tools.  
*Pages 153–169 of: Open source approaches in spatial data handling*.  
Springer.
- Uhlenbrook, S, & Leibundgut, Ch. 2002.  
Process-oriented catchment modelling and multiple-response validation.  
*Hydrological Processes*, **16**(2), 423–440.
- Unsworth, Michael H, & Monteith, JL. 1975.  
Long-wave radiation at the ground I. Angular distribution of incoming  
radiation.  
*Quarterly Journal of the Royal Meteorological Society*, **101**(427), 13–24.
- ValdÈs, Juan B., Fiallo, Yolanda, & Rodrìguez-Iturbe, Ignacio. 1979.  
A rainfall-runoff analysis of the geomorphologic IUH.  
*Water Resources Research*, **15**(6), 1421–1434.
- Valente, Fernanda, David, JS, & Gash, JHC. 1997.  
Modelling interception loss for two sparse eucalypt and pine forests in  
central Portugal using reformulated Rutter and Gash analytical models.  
*Journal of Hydrology*, **190**(1-2), 141–162.
- van der Tak, Laurens D., & Bras, Rafael L. 1990.  
Incorporating hillslope effects into the geomorphologic instantaneous unit  
hydrograph.  
*Water Resources Research*, **26**(10), 2393–2400.
- van der Velde, Y, Torfs, PJJE, Zee, SEATM, & Uijlenhoet, R. 2012.  
Quantifying catchment-scale mixing and its effect on time-varying travel  
time distributions.  
*Water Resources Research*, **48**(6).

- Verfaillie, Els, Van Lancker, Vera, & Van Meirvenne, Marc. 2006.  
Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas.  
*Continental Shelf Research*, **26**(19), 2454–2468.
- Wang, C. T., Gupta, Vijay K., & Waymire, Ed. 1981.  
A geomorphologic synthesis of nonlinearity in surface runoff.  
*Water Resour. Res.*, **17**(3), 545–554.
- Wilkinson, GN, & Rogers, CE. 1973.  
Symbolic description of factorial models for analysis of variance.  
*Applied Statistics*, 392–399.
- Wilkinson, Leland, & Friendly, Michael. 2009.  
The History of the Cluster Heat Map.  
*The American Statistician*, **63**(2), 179–184.
- WMO. 1994.  
*Guide to hydrological practices*.
- Woods, Ross, & Sivapalan, Murugesu. 1999.  
A synthesis of space-time variability in storm response: Rainfall, runoff generation, and routing.  
*Water Resour. Res.*, **35**(8), 2469–2485.
- Xiao, Jingfeng, Zhuang, Qianlai, Law, Beverly E, Chen, Jiquan, Baldocchi, Dennis D, Cook, David R, Oren, Ram, Richardson, Andrew D, Wharton, Sonia, Ma, Siyan, *et al.* . 2010.  
A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data.  
*Remote sensing of environment*, **114**(3), 576–591.
- Zehe, E, Lee, H, & Sivapalan, M. 2006.  
Dynamical process upscaling for deriving catchment scale state variables and constitutive relations for meso-scale process models.  
*Hydrology and Earth System Sciences Discussions*, **10**(6), 981–996.
- Zhao, R-J. 1980.  
The xinanjiang model.  
*In: Proceedings of the Oxford Symposium*.
- Zinke, Paul J. 1967.  
*Forest interception studies in the United States*.

Forest Hydrology. Oxford, UK: Pergamon Press.

Zoccatelli, D., Borga, M., Viglione, A., Chirico, G. B., & Blöschl, G. 2011.  
Spatial moments of catchment rainfall: rainfall spatial organisation, basin morphology, and flood response.  
*Hydrology and Earth System Sciences*, **15**(12), 3767–3783.

Zuecco, Giulia, Oliviero, Omar, Penna, Daniele, van Meerveld, Ilja, Hopp, Luisa, Dalla Fontana, Giancarlo, & Borga, Marco. 2014.  
Spatial and temporal variability of throughfall at the plot scale in the Italian pre-Alps.  
*In: EGU General Assembly Conference Abstracts*, vol. 16.



The increasing impacts of climate changes on water related sectors are leading the scientists' attentions to the development of comprehensive models, allowing better descriptions of the water and solute transport processes. "Getting the right answers for the right reasons", in terms of hydrological response, is one of the main goals of most of the recent literature. 'Embedded reservoirs' are often used for each HRU, to allow a consistent representation of the processes. In this work, a new semi-distributed model for runoff and evapotranspiration is presented: five different reservoirs are inter-connected in order to capture the dynamics of snow, canopy, surface flow, root-zone and groundwater compartments. The knowledge of the mass of water and solute stored and released through different outputs allows the analysis of the hydrological travel times and solute transport in catchments. The latter have been studied extensively, with some recent benchmark contributions in the last decade. The thesis presents a detailed description of a new theoretical approach that reworks the travel time theory from the point of view of the hydrological storages and fluxes involved. To give substance to the theory, a small catchment in the prealpine area was chosen as an example and the results illustrated. The new semi-distributed model for runoff and evapotranspiration and the travel time theory were implemented and integrated in the semi-distributed hydrological system JGrass-NewAge. The system is flexible, expandable and applicable in a variety of modeling solutions. In this work, the model code underwent to an extensive revision: new components were added, old components were enhanced, documentation was standardized and deployed.

**Marialaura Bancheri** received a Master's Degree in in Hydraulic Engineering at University of Salerno, Italy in 2013. Her main research focus is the hydrological modeling of complex catchments and the development of open-source tools, using innovative software techniques. The experiences gained during the PhD increased her background as a civil engineer, giving her a wider view on the application of the engineering in different fields. During the PhD career, She learned to develop project proposals, to identify and the collect data, to implement models and tools with correct software design, and to disseminate the results in effective ways. She was also for one year in US at the USDA as a research scientist, working on the interperability between agricultural and hydrological models. During the PhD, She was author of four publications on international journals and attendend sevaral national and international conferences.