ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

AgroParisTech
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

L'Institut des Sciences et Industries du Vivant et de l'Environnement

(AgroParisTech)

Spécialité : Hydrologie

présentée et soutenue publiquement par

Gianluca BOLDETTI

le 17 décembre 2012

Estimation des paramètres des modèles hydrologiques sur des bassins versants non-jaugés: confrontation des approches directes et indirectes.

(Estimation of the parameters of hydrological models on ungauged basins: a comparison of direct and indirect approaches)

Directeur de thèse : V**azken ANDRÉASSIAN**
Co-encadrement de la thèse : **Ludovic OUDIN**

Jury

Mme Anne Catherine FAVRE, Professeure, INPG	Rapporteure
M. Thierry LEVIANDIER, Directeur Scientifique, ENGEES	Rapporteur
M. Cyril KAO, Directeur Scientifique Adjoint, AgroParisTech	Examinateur
M. Pierre RIBSTEIN, Professeur, Université Pierre et Marie Curie	Examinateur
M. Andras BARDOSSY, Professeur, Université de Stuttgart	Examinateur

AgroParisTech
Irstea
Antony

Table of Contents

Acknowledgements **3**

Résumé **4**

Abstract **5**

1 Introduction **7**

Part 1 – Methods, databases and literature **11**

2 Databases used in this thesis **13**

3 Methodological aspects **23**

4 Literature review on the regionalization of rainfall-runoff models **33**

Part 2 – Studies relative to flow statistics and their regionalization **47**

5 Linking flow statistics to physiographic descriptors **49**

6 Using neighbour catchments residuals to improve the efficiency of flow statistics regionalization **63**

Part 3 – Regionalization of rainfall-runoff models – direct methods **89**

7 Physiographic similarity regionalization **91**

8 Joining spatial proximity and physiographic similarity **103**

9 Sensitivity analysis of regionalization methods: how do they react to the lack of similar catchments? **113**

Part 4 – Regionalization of rainfall-runoff models – the indirect path **121**

10 Direct and indirect regionalization **123**

11 How the choice of an efficiency criterion impacts our vision of the 'best' regionalization method **139**

12 Conclusion **149**

13 References **153**

Part 6 – Appendices **157**

14 Sensitivity to the elimination of similar donors: graphic results **157**

15 GR4J **193**

Index **195**

List of Figures **200**

List of Tables **204**

Acknowledgements

This thesis work would have not been possible without the help and support of the following persons:

The thesis director and co-director, Vazken Andréassian and Ludovic Oudin, who provided constant support and useful comments;

The Hydrology unit at Cemagref Antony, and especially Charles Perrin, François Bourgin, Lionel Berthet and Julien Peschard;

The thesis advancement committee (comité de suivi), who provided and additional point of view on the thesis work;

The thesis defense committee, who suggested useful improvements on the initial version of the manuscript;

My family, who sustained my motivation.

A special mention goes to Marine, for being particularly tolerant and understanding during the last months.

Résumé

Les hydrologues ont depuis longtemps l'ambition de produire des modèles qui ne nécessitent pas de calage sur les débits observés. Et pourtant, tous les modèles existants à ce jour nécessitent cette opération, et leur application à des basins non-jaugés dépend donc de la mise en place de procédures de régionalisation, qui identifient des basins versants jaugés dont les données peuvent être utilisés en substitution des chroniques de débit manquantes.

Les méthodes de régionalisation constituent donc un sujet d'intérêt récurrent dans les études hydrologiques, en particulier depuis le lancement de l'initiative PUB de part de l'AISH (Sivapalan et al., 2003). Cependant, l'évaluation des points de force et de faiblesse des différentes approches jusqu'ici proposées est encore difficile, à cause de la rareté d'études comparatives de grande échelle. Le principal objectif de cette thèse est de contribuer à une telle évaluation au travers de la comparaison des performances d'approches classiques et nouvelles sur un grand échantillon composé par 800 bassins versants Français. La thèse poursuit aussi la généralisation de ses résultats par biais d'un test de robustesse étudié ad-hoc, qui reproduit la situation de pays ayant un réseau de jaugeage moins spatialement dense.

L'analyse menée par la thèse se développe en trois parties :

- une première partie dédiée à la régionalisation des statistiques sur le débit, utilisées à la fois comme cas d'étude simple pour tester l'utilisation complémentaire de donnés spatiales et physiographiques en régionalisation, et comme première étape nécessaire pour la mise en place d'une régionalisation "indirecte" des modèles pluie-débit ;
- une deuxième partie dédiée à la traditionnelle régionalisation "directe" des modèles pluie-débit, basée sur les critères de proximité spatiale, similarité physique, ou sur une combinaison des deux ;
- une troisième partie proposant un nouveau schéma de régionalisation dite "indirecte", qui se base sur la régionalisation des statistiques de débit déjà effectuée. Ce type de méthode a été proposé par plusieurs auteurs dans les dernières années, mais à notre connaissance il n'a jamais été comparé directement aux méthodes "directes".

La thèse identifie la sélection de descripteurs physiographiques significatifs comme l'étape la plus importante pour la performance des méthodes "directes" de similarité physique, et montre aussi que ces approches restent plus performantes face à la nouvelle méthode dite "indirecte", même en utilisant un critère d'évaluation a priori favorable à la deuxième. Cette dernière pose à notre avis encore un certain nombre de questions d'ordre méthodologique à résoudre avant d'envisager une utilisation dans un contexte opérationnel.

Abstract

Despite a long-standing ambition to produce a model that does not require calibration against observed runoff data, all current hydrological models require this step: their application on ungauged basins is therefore only possible by means of regionalization procedures, which identify appropriate gauged sites whose data are used in place of the missing runoff record.

Regionalization procedures are therefore a subject of increasing interest in hydrological studies, especially since the start of the IAHS PUB initiative (Sivapalan et al., 2003).

However, assessing the relative merits of the several regionalization approaches developed so far is still difficult, because of the relative lack of large-scale comparative studies.

The main objective of this thesis is to help such assessment by testing classical and novel regionalization approaches on a large dataset of over 800 catchments located in France.

The thesis also aims at generalizing its results by means of a purposely-built robustness test that mimics the situation of more scarcely-gauged nations.

The thesis analysis consists of three main parts:

- A fist part dedicated to the regionalization of flow statistics, used as an exploratory tool to test the complementary use of physiographic and spatial information in the regionalization process, and as required step for an "indirect" regionalization of rainfall-runoff models.
- A second part dedicated to classical "direct" regionalization of rainfall-runoff models, on the basis of spatial proximity, similarity in catchment attributes, or a combination of the two.
- A third part proposing a novel "indirect" regionalization framework, based on the regionalization of flow statistics developed in the first part. This kind of regionalization approach has been advocated by several authors in recent years, but to our knowledge it had not yet been directly compared to "direct" methods.

The thesis identifies the selection of relevant physiographic descriptors as the most important factor affecting the performances of "direct" regionalization methods based on physical similarity, showing that such approaches still seem to outperform the novel "indirect" framework, even when adopting a performance criterion that can be expected to favor the latter. In our opinion, this regionalization approach raises methodological questions that need to be answered before being considered in operational contexts.

1 Introduction

Rainfall-runoff models are a key tool in several contemporary water-management applications, allowing high- and low-flow forecasting, reconstitution of incomplete flow records, correct dimensioning of dams and flood-management structures.

However, despite a long-standing ambition, hydrologists have so far failed at producing a hydrological model that does not require any calibration (Sivapalan et al., 2003): every one of the currently used hydrological models, regardless of its structure, requires the estimation of at least a few parameters in order to realistically reproduce the hydrological behaviour of a particular catchment. Even physically-based models require calibration, due for instance to the poor representativeness of small-scale descriptors when used as parameters in large-scale applications (Bloschl and Sivapalan, 1995).
Unfortunately, while such calibration can only be performed in the presence of simultaneous rainfall and runoff records of sufficient length, flow measurements required are often not available at the site(s) of interest, and in most cases installing a new gauging station is not a realistic option, because of the time and cost required to obtain a meaningful record.

As a consequence, hydrologists often face the challenge of making predictions in an ungauged situation: since the parameters of the current rainfall-runoff models cannot be estimated directly from the catchments' measurable characteristics, parameterising a model for an ungauged basin implies a transfer of information from one or several gauged catchments (often called donors) to the ungauged one (called the receiver). Over the years, a number of techniques have tried to operate such a transfer, all of which are either based on physiographic and climatic catchment descriptors, or on the geographical position of the donor and receiver catchments, or on both. As most of these regionalization approaches use such information to identify donor catchments that are physically and climatically similar and/or spatially close to the receiver, they often go under the names of “physical similarity” (see for isntance: Acreman and Sinclair, 1986; Burn and Boorman, 1993) and “spatial proximity” (Egbuniwe and Todd, 1976).

Regionalization methods relying heavily on spatial proximity are often seen as having less desirable properties than the ones based solely on physical similarity, for essentially two reasons. On one side, spatial proximity does not provide useful information about the link

between model parameters, dominant hydrological processes, and a catchment's physical and climatic properties, while physical similarity allows for a qualitative interpretation. On the other side, an approach driven by spatial proximity clearly requires "close enough" gauging stations, while one would naively hope that physical similarity allows one to identify proper donor catchments even if they are very far from the studied ungauged catchment.
Yet, when considering the relative performance of the two approaches, spatial-proximity methods often perform "disappointingly well", and there seems to be a degree of complementarity between the two approaches (Oudin et al., 2008): if one could know in advance which of the two would work best on each studied ungauged basin, the performance of such an hybrid method would greatly overcome the ones of its components.

This thesis is the consequence of the above considerations and has two main objectives:

(i) To explore the complementary use of physiographic/climatic and geographical information in the context of regionalization, under the assumption that while a method that privileges the former is more desirable, the latter can greatly improve the regionalization's accuracy in some circumstances;

(ii) To address the relative strengths and weaknesses of the tested regionalization approaches on a rich and quite diverse dataset, and evaluating their robustness, especially in regard of the need for spatially-close gauging station (high spatial density of the gauging network).

The first part of this thesis outlines the common background shared by the two main parts of the thesis, consisting in the state-of the art of regionalization studies (literature review), a description of the database used in this study, and a description of a few key methodological points, that essentially regard the techniques used for the evaluation of the regionalization approaches tested throughout the thesis. A special attention is given to the identification of benchmark comparisons and of robustness testing.

The second part of this thesis treats the regionalization of flow statistic, which we chose to deal with before rainfall-runoff models for the following reasons:

(i) Evaluating the relative role of physiographic information and of spatial proximity on hydrological variables that are less affected by the choice of a model structure and by the inevitable issue of parameter identifiability / interdependence. In a way, the regionalization of flow statistics approaches the "ideal case" of a conceptual model whose parameters show perfectly identifiable optimal values, which are in most cases significantly (yet not perfectly) correlated to measurable catchment's attributes.

(ii) As a pre-requisite for the "indirect regionalization" presented in the fourth part of the manuscript

The approach used involved a two-step combination of regressions between statistics and catchment characteristics, followed by a spatial interpolation of the residuals, that can be refined in order to acknowledge nested donor and receiver catchments, and big differences in catchment size.

The third part of this thesis deals with what is, in recent years, probably the most common regionalization approach for conceptual, lumped rainfall-runoff models: the transfer of parameter sets to an ungauged catchment from one or more gauged catchments that are thought to be hydrologically similar to the former. The fundamental element of such approaches is a similarity metric, built on a combination of measurable catchment attributes, so that two catchments showing similar attributes will be considered to have potentially similar hydrological behaviours.

Such attributes can describe physiographic and climatic properties of a catchment, or even its geographical position (but as the objective of a regionalization procedure is to deal with scarcerly-gauged regions, a common objective of regionalization studies is to reduce as much as possible the role of the latter in the similarity metric).

In this thesis work, we proceed in a similar fashion with what is done with flow statistics, by trying to get the best possible results out of the available physiographic and climatic information in the first place, and using spatial proximity in a complementary way in a second one.

The fourth part of this thesis deals with what we call "indirect" regionalization methods, as they are based on the previous regionalization of flow statistics. Once their values have been calculated for a given ungauged catchment, rainfall-runoff parameter sets for the same basin are chosen on the basis of their ability to produce a simulated streamflow record that is consistent with those statistics. These methods can be attractive as the first-step regionalization they are based on has more desirable properties than the direct regionalization of parameter sets, but as we will see, on another side the construction of a statistic-based constraint that produces desirable simulations is not trivial.

The last part of this manuscript features a published article on the issue of outlier identification and treatment in the context of a two-step regionalization of flow statistics, as the one showed in the second part.

Part 1 – Methods, databases and literature

In this part, we provide the context and common background shared by all parts of this thesis, in terms of methodological approach, available data, and placement relative to the existing literature on the topic of regionalization:

- Chapter 2 describes the databases used in this thesis;
- Chapter 3 discusses the methodological aspects relevant to the evaluation of the alternative approaches proposed in the thesis to address the ungaged catchment issue;
- Chapter 4 presents a detailed review of the literature on the regionalization issue.

2 Databases used in this thesis

In this chapter, we present the datasets on which our work is based.

We start with a short justification of our use of a large catchment dataset. Then, we present this dataset; we have a look at the relationship between catchment characterization and physiographic and climatic descriptors, and finish with some synthetic descriptors.

2.1 Why should we use a large set of catchments?

In this study, a data set of 865 French catchments was used. The catchments are spread over France and are subject to a variety of climatic conditions (oceanic, Mediterranean, continental). What is the interest of using such a large dataset in a regionalization study?

The first reason is that a large dataset gives a certain guarantee of diversity. A dataset including catchments of several different hydrological, physiographic and climatic flavours should ensure that all the results of the study can be considered relevant to the general issues that are shared by every regionalization application, and less affected by local specificities. On this point, we notice that our dataset is still far from the "ideal" case, which would be a dataset including many catchments from all over the world, and that our results will be inevitably specific to France.

The second reason is that the impact on regionalization of relatively rare occurrences such as catchments that have some (hydrological) reason to be considered as outliers and hidden data errors is put in better perspective when considering a larger dataset: smaller ones could be "lucky cases" that do not present such imperfection, or on the contrary be greatly affected by a single "bad" catchment.

2.2 How our dataset was made

The selection of catchments was made based on three criteria:

(i) absence of regulation,

(ii) availability of continuous rainfall and streamflow records over a twenty year period (1986-2005),

(iii) amount of missing values in the streamflow record less than 20%.

Table 1 presents the main characteristics of the data set in terms of catchment area, precipitation, potential evapotranspiration (PE) and streamflow.

Table 1: Essential characteristics of the 865 catchment data set

	Min	**0.2 Quantile**	**0.5 Quantile**	**0.8 Quantile**	**Max**
Area (km²)	2	73	208	828	112990
Mean annual precipitation, P (mm/year)	547	818	968	1233	2144
Mean annual runoff, Q (mm/year)	14	202	327	595	6500
Mean annual PE (mm/year)	304	631	670	727	892

2.3 How can we characterize a catchment?

Ideally, physiographic descriptors should be able to represent in a balanced and comprehensive way the main hydrologically relevant traits of a catchment, namely:

- the catchment's climate ;
- its topographic characteristics, such as elevation, area, land use ;
- its lithological and soil characteristics.

Unfortunately this ideal cannot always be fully realized.

Indeed, while topographic and climatic descriptors are widely available, the situation is much more difficult for lithology and soil type/properties: currently available descriptors either seem not to be entirely relevant to the hydrological processes, or to be potentially very useful, but not measurable at the relevant scale (full-catchment, for a lumped model), nor easily re-scaled.

So far, only one attempt of using soil information seems to have given successful results: this is the case of *BFIHOST* (Boorman et al., 1995), a British index of low-flow catchment behaviour that is derived directly from soil maps, with a rather complex procedure. Note, however, that while this descriptor has proven to be quite successful on the catchments for which it has been developed, it does not seem to work nearly as well on a different dataset, as shown by Oudin et al.(2010) and by Schneider et al.(2007), probably due to inconsistencies regarding the type of catchments considered: in this regard Schneider et al.(2007) noticed that *BFIHOST* seems to have a much better predictive value on northern Europe catchment than on southern (and especially Mediterranean) ones.

However, some topographic descriptors seem to be at least indirectly linked to hydrological soil/lithology properties: drainage density (expressed as size of source areas) is one of them (Le Moine, 2008).

We have just mentioned that catchment descriptors should be measured at full-catchment scale for a lumped model, implying that the relevant scale for a distributed one is finer. These trivial observations, as well as the interesting story of *BFIHOST*'s successes and failures, are particular cases of a more general principle, linked with the empirical approach followed in regionalization studies (Sawicz et al., 2011).

This principle says, in our opinion, that the hydrological relevancy of a catchment descriptor is always relative to a specific hydrological application, that is defined by the kind of model used (structure, time step), the scope of the application (as expressed by the efficiency criteria, or by the targeted flow statistics), the kind of catchments considered, and the interactions and correlations with other available descriptors.

Of course, some descriptor choices can give consistent results for several different applications, and identifying such a descriptor list is a reasonable (although ambitious) objective, until the consistency is not expected to be perfect and hydrologists do not mistake relative truths for absolute ones.

Last but not least, it is crucial to remember that measurable descriptors show, in most cases, only the façade of a catchment's structure and functioning, and that two catchments that "look" similar can't be expected to behave similarly in a crude, mechanistic way. Our best efforts should be expected to produce a situation where two apparently similar catchments are likely to have a similar behaviour, but this success should not mislead us: failures will still exist and should not be systematically seen as the consequence of "outliers" or data errors.

In hydrological investigations, well chosen catchment descriptors can provide very strong clues, but cannot be taken as conclusive evidence.

Les us have a look at which descriptors were available for the physiographic and climatic characterization of our catchments (Table 2).

Table 2: List of catchment descriptors available for this study

	Descriptor	**Description**
1	T	Average temperature (°C)
2	W	Average wind speed (m/s)
3	Hum	Average specific humidity (g/kg)
4	A	Area (km²)
5	Z_{min}	Minimum altitude (m)
6	Z_{ave}	Average altitude (m)
7	Z_{max}	Maximum altitude (m)
8	$Z_{0.n}$, $n=1,..9$	Altitude distribution quantiles
9	S_{min}	Minimum slope
10	S_{ave}	Average Slope
11	S_{max}	Maximum Slope
12	$S_{0.n}$	Slope distribution quantiles
13	URBAN	% of surface occupied by Corine land cover classes 111-124
14	AGRIC.	% of surface occupied by Corine land cover classes 211-213
15	FRUIT	% of surface occupied by Corine land cover classes 221-223
16	HYBRID	% of surface occupied by Corine land cover classes 111-124
17	FOREST	% of surface occupied by Corine land cover classes 241-244
18	OTHER	% occupied by remaining Corine land cover classes
19	DD	Drainage Density, expressed in average source area size (km²)

Drainage Density is calculated here as the geometric mean of a catchment's source areas' size. See Le Moine (2008) for more details.

Table 3 shows the cross-correlations between most of our descriptors (the quantiles of altitude and slope have been excluded from this table for sake of brevity). Some descriptors show very strong correlation, such as heights and slopes, forest and agricultural coverage, temperature and specific humidity.

Most of the strongest correlations (more than 0.5), are in our opinion the symptom of the catchment types encountered in our dataset, which seem to belong to a continuum between two extremes: on one side, mountain catchments which tend to be steeper, more forested, colder and rainier, while on the other side lowland catchments tend to be flat, less rainy, with agricultural fields instead of forests. As a result, climate variables, altitude, slope, forest and agricultural land cover appear to be strongly dependent on each other.

Temperature and humidity are also strongly correlated, which isn’t a surprise since the latter is expressed as absolute moisture content, rather than relative.

Finally, a few descriptors do not show strong correlations with other catchment characteristics: it is the case of some land use classes (which are probably less represented in our dataset) minimal slope and catchment area. The later is quite reassuring since it means that for each catchment size several catchment types are represented, and viceversa.

Since in the applications described later in this thesis we choose to use appropriate techniques to select the relevant descriptors, choosing which descriptor should be kept in the strongly correlated pairs or groups is not necessary.

Table 3: Matrix of correlations between descriptors ($p<0.05$ significance in bold characters)

P	**1**																	
T	**-0.54**	**1**																
W	**-0.43**	**0.26**	**1**															
Hum	**-0.52**	**0.95**	**0.31**	**1**														
A	**-0.08**	0.03	-0.02	0.03	**1**													
Zmin	**0.48**	**-0.81**	**-0.33**	**-0.78**	**-0.11**	**1**												
Zmoy	**0.59**	**-0.82**	**-0.44**	**-0.84**	**-0.03**	**0.87**	**1**											
Zmax	**0.56**	**-0.65**	**-0.48**	**-0.69**	**0.15**	**0.63**	**0.89**	**1**										
SMin	**0.32**	**-0.26**	**-0.08**	**-0.26**	**-0.15**	**0.38**	**0.34**	**0.19**	**1**									
SMoy	**0.63**	**-0.55**	**-0.46**	**-0.63**	**-0.06**	**0.52**	**0.8**	**0.83**	**0.36**	**1**								
SMax	**0.55**	**-0.47**	**-0.45**	**-0.52**	**0.16**	**0.32**	**0.65**	**0.82**	0.02	**0.76**	**1**							
URBAN	**-0.2**	**0.16**	**0.1**	**0.11**	**0.02**	**-0.23**	**-0.26**	**-0.2**	**-0.15**	**-0.24**	**-0.13**	**1**						
AGRIC.	**-0.6**	**0.34**	**0.46**	**0.42**	**0.06**	**-0.5**	**-0.65**	**-0.64**	**-0.33**	**-0.7**	**-0.55**	**0.12**	**1**					
FRUIT	**-0.12**	**0.27**	0.04	**0.17**	**-0.02**	**-0.13**	**-0.09**	-0.01	-0.01	0	0.03	**0.09**	**-0.2**	**1**				
HYBRID	**-0.07**	**0.35**	**0.06**	**0.42**	**-0.01**	**-0.18**	**-0.22**	**-0.19**	0	**-0.24**	**-0.16**	-0.01	**-0.08**	0.03	**1**			
FOREST	**0.6**	**-0.43**	**-0.44**	**-0.51**	**-0.05**	**0.52**	**0.65**	**0.61**	**0.33**	**0.7**	**0.5**	**-0.24**	**-0.85**	0.04	**-0.34**	**1**		
OTHER	**0.22**	**-0.47**	**-0.2**	**-0.48**	**-0.01**	**0.35**	**0.48**	**0.47**	0.07	**0.47**	**0.42**	**-0.08**	**-0.28**	-0.05	**-0.18**	**0.08**	**1**	
DD	-0.07	-0.01	**0.17**	-0.01	**-0.02**	**-0.09**	**-0.1**	**-0.13**	**-0.12**	**-0.12**	-0.06	0.05	**0.19**	-0.04	**-0.12**	**-0.13**	-0.01	**1**
	P	*T*	*W*	*Hum*	*A*	*Zmin*	*Zmoy*	*Zmax*	*SMin*	*SMoy*	*SMax*	*URB*	*AGR*	*FRU*	*HYB*	*FOR*	*OTH*	*DD*

2.4 Distribution of a few key descriptors and flow characteristics over our dataset

In this chapter we will look at the shape of the distribution of some of our descriptors. The scope of this analysis is on one side to provide an idea of the diversity of the data set, and on the other to serve as a diagnostic tool prior to the definition of a similarity metric between catchments.

Figure 1 shows the distribution densities of six among the most relevant physiographic and climatic descriptors. It can be seen that the distribution shapes are fairly close to normal – with the exclusion of a few extreme values – apart from Area and Drainage Density, which seem to be log-normally distributed.

Figure 2 shows the distributions of five flow characteristics: average annual runoff (Q), three flow quantiles (Q_{90}, Q_{50}, Q_{10}, where Q_{90} is the daily runoff that is surpassed on 10% of observed days, Q_{50} is the median flow, Q_{10} is the daily runoff surpassed on 90% of the observed days) and Base Flow Index (fraction of "base" runoff over total runoff. The base flow separation has been made by a graphical method linking hydrographs' 5-day minima values). These chart emphasize the diversity of our dataset: from rather dry to very wet catchments, from very unresponsive to very responsive.

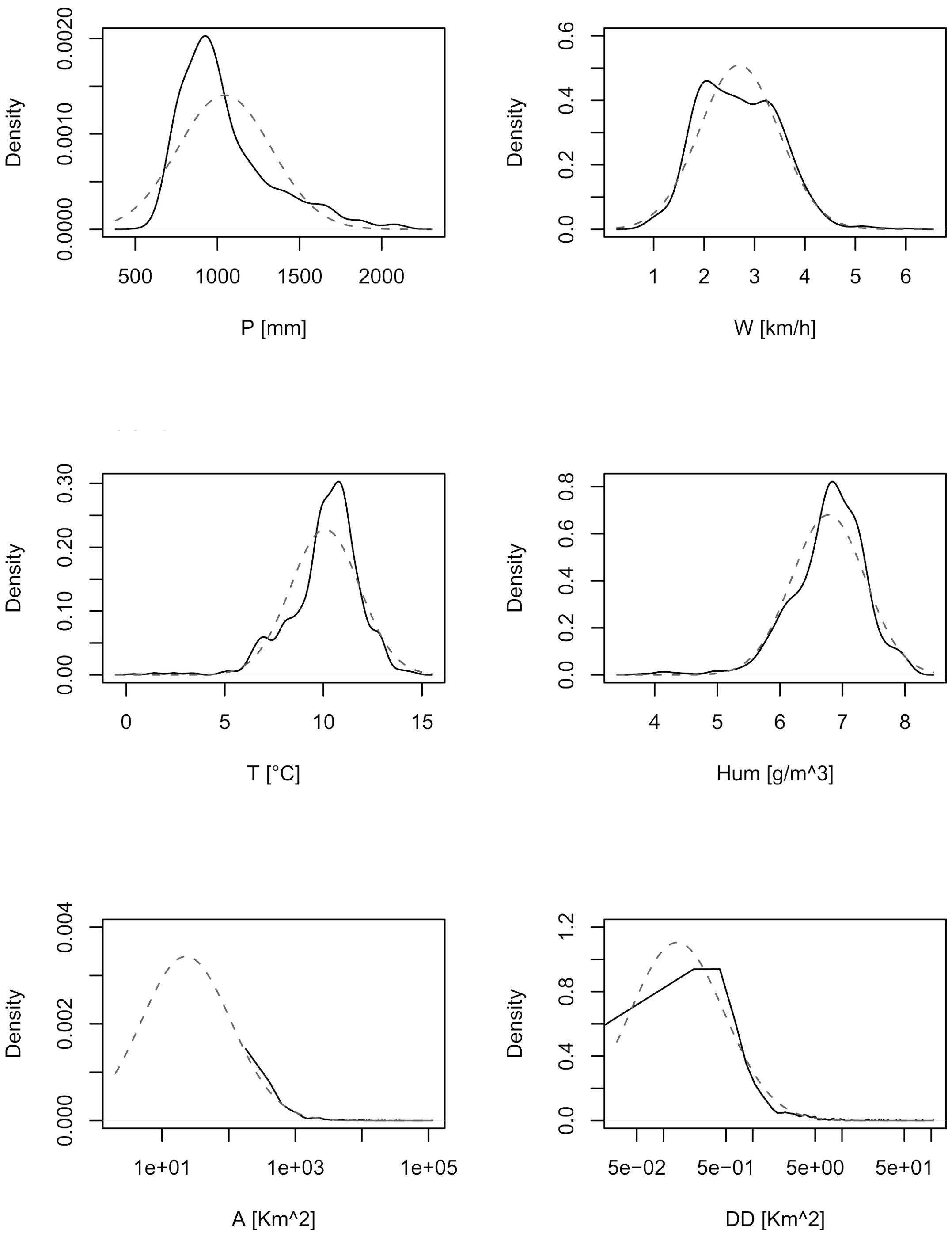


Figure 1: Distribution densities of six physiographic distributors, compared with normal distributions (dashed lines). Note the semi-log scales for Area and Drainage Density (lognormal distributions were employed in these two cases)

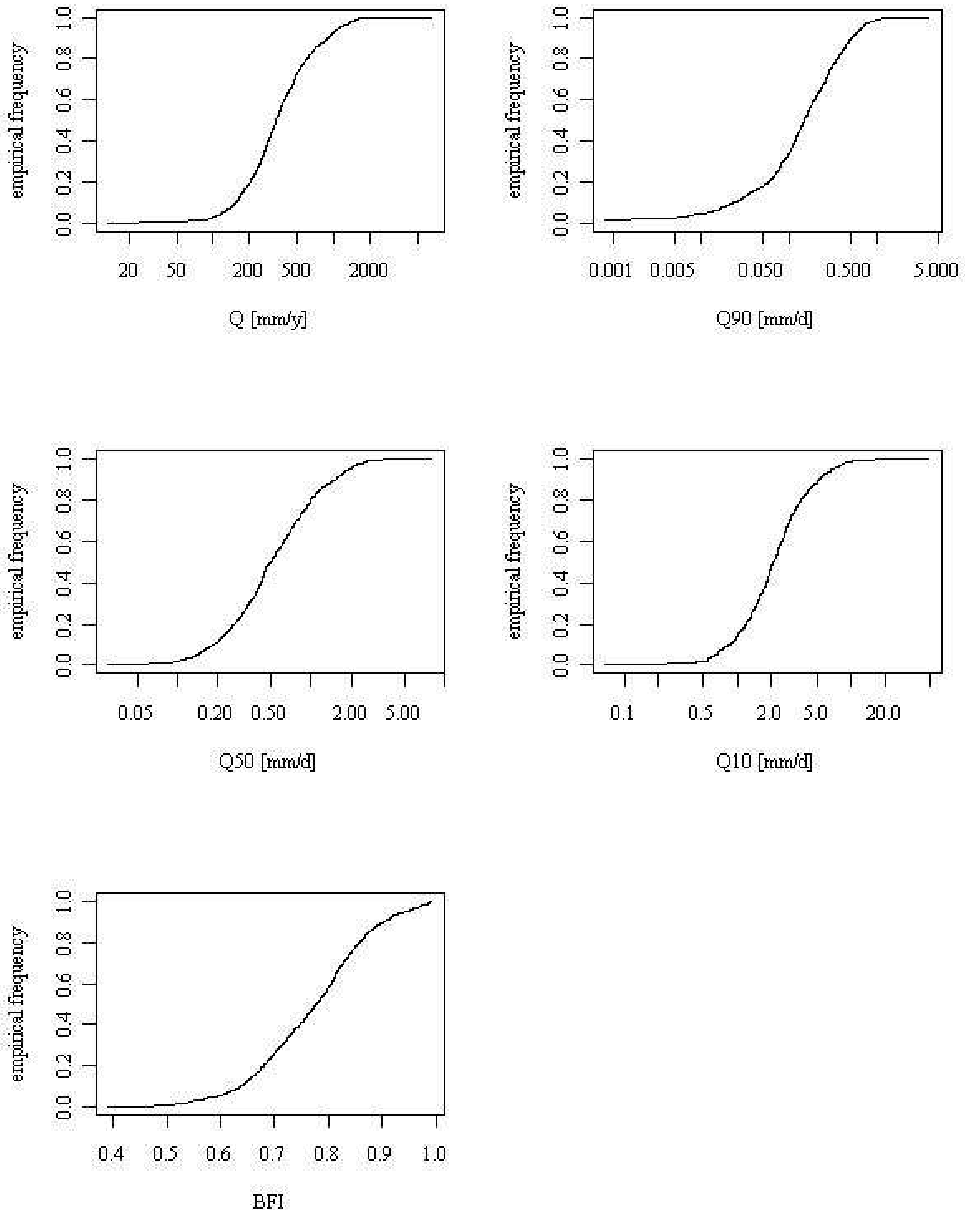


Figure 2: Distributions of five flow characteristics.

3 Methodological aspects

This chapter will clarify a few key points relative to the methodology followed throughout this thesis work, and define the meaning (in the context of this thesis) of often-repeated terms and expressions.

We will discuss successively the jack-knife approach to cross validation, the need for using benchmarks in order to interpret our validation results, and the specific way to address the notion of robustness in a regionalization perspective. Last, we will discuss the differences between *donors* and *receivers*, and argue for a different selection procedure for the donor and the receiver pool.

3.1 General principles of the comparative testing of alternative regionalization methods

3.1.1 Jack-knife approach to cross validation

In this thesis, as in most regionalization studies, a leave-one-out cross validation method has been used to assess the performances of the proposed regionalization procedures. This procedure is commonly referred to as "jack-knife" in hydrological studies, and we will also do so in this thesis work. Note however that this use of the expression jack-knife could sound improper to a statistician, since it usually reserves to the estimation of the variance and bias of a statistic (Efron and Gong, 1983), rather than to the evaluation of a predictive model.

In the context of regionalization studies, the method consists of the following steps:

- Ignoring the gauging data of one catchment in the dataset. We will refer to this catchment as *pseudo-ungauged*, or as *receiver;*
- Estimating model parameters or flow statistics for the pseudo-ungauged catchment using its physiographical and climatic descriptors plus the complete information that we have for the rest of the catchments in the dataset (called *donors*);
- Repeating the procedure so that every catchment in the dataset is treated once as pseudo-ungauged;
- Evaluating the efficiency (or the errors) of the flow simulations (or flow statistics) estimated in such a way, usually taking an average or median of the efficiencies (errors) observed on individual catchments.

3.1.2 There is no absolute truth in this world: we need benchmarks

If looked at in an “absolute” way, the performances of most current regionalization approaches would probably look rather poor, especially in comparison with the performances of a calibrated model[1], and the differences between a method and another would look quite small. Furthermore, such performances are inherently dependent on the model and on the dataset characteristics, among which its spatial density is possibly the most influent: for this reason, we decided to develop a robustness test called “metrological desert” which consist in artificially reducing the dataset’s density (see section 3.3 for a full discussion of this test).

[1] Notice that in this thesis work, for sake of simplicity, we did not use a split-sample calibration/validation approach: therefore, the calibrated parameter set will yield the model’s best possible performances on the time series considered.

Following these two considerations, we think that a “relative” evaluation of regionalization performances is more meaningful than an “absolute” one. We then propose to evaluate the performance of regionalization methods in a relative way, with the help of benchmark comparisons. Such benchmarks should give an example of minimum and maximum, and “acceptable” expected performances, and should be kept equal when switching from a regionalization method to another.

3.1.3 Specificities of different models: how the regionalization exercise differs for a statistical and for a rainfall-runoff model

This thesis covers both the regionalization of flow statistics, and the regionalization of rainfall-runoff models. Though the two exercises share many commonalities (they can even be combined as shown in section 10), they differ on at least one crucial point: while flow statistics generally show significant correlations with physiographic and especially climatic descriptors, this is seldom the case for calibrated model parameters, which should ideally be climate-independent. As a consequence, regressions between physiographic and climatic descriptors and streamflow statistics tend to show at least acceptable results and can be used as the founding element of a regionalization approach (even if we should be aware that this is mostly due to the fact that flow statistics are strongly dependent on the climatic forcings).

On the other hand, as covered in paragraph 4.4.1, regression-like relationships tend to fail when applied to model parameters, even when sophisticated calibration techniques are used in order to obtain more physically-correlated parameter sets: the most successful regionalization approaches for rainfall-runoff models are generally based on data-transfer methods.

Another peculiarity of rainfall-runoff models is that there is a degree of interaction and interdependence between the values of different parameters. This is probably the reason why, when using the information of several donor catchments, linear combination of model parameters does not seem to be the best choice: methods based on model output averaging, as outlined by McIntyre et al. (2005), seem advantageous. In this case, a simulation is run for each of the donor catchments, using the donors's parameter set and the pseudo-ungauged rainfall record. The pseudo-ungauged simulated streamflow record will then be a linear combination of those simulations.

It is worth noting that there might be a secondary reason why output-averaging gives good results: the most common performance criteria are based around the RMSE and are usually more forgiving for conservative, smoother-than-necessary simulations than for simulations which look more "realistic" when not compared with the measured runoff record, at the cost of taking more risks. Averaging several slightly different simulations goes exactly in the direction of a conservative/smooth prediction.

3.2 Catchment selection: differential approach for the donor and the receiver pool

An important methodological issue in regionalization studies is the selection of the catchments on which to test the proposed procedures. Here, we discuss the options, and argue that the best choice is probably a differential approach for the selection of "receivers" (catchments that should be treated as ungauged during a regionalization test) and "donors" (gauged catchments whose data is transferred to the receivers).

- **Receivers:** Ideally, we should use as many "receivers" as possible, setting only reasonable demands about the accepted amount of human influences and the amount of available data (streamflow and precipitation records' length and completeness, available catchment descriptors). The objective of a broad selection of receivers is to be able to test regionalization methods on a large and diverse dataset, in order to (hopefully) ensure that the observed results will be as general as possible, and not specific to a given catchment type.

- **Donors:** On the other hand, there should be no particular limit on donor selection. Most regionalization methods include some kind of donor selection, and in many cases the regionalization exercise is limited to the identification of a few catchments that show some similarities with the receiver, under the assumption that they are correlated to hydrological similarities. Depending on the regionalization application, it is possible to black-list some catchments (never use them as donors) (see paragraphs 6.2 and 6.4).

However, in any case, we insist on the fact that excluding a catchment from the list of potential donors should never lead to its exclusion from the list of receivers. The constitution of the list of receivers should be independent from all regionalization considerations, if the evaluation of the method is to remain sound. Even if a catchment was found so peculiar that it would not look like any other catchment in the dataset, it should not be excluded from the receiver dataset.

3.3 Further methodological requirements to assess the robustness of a regionalization method

3.3.1 Why this question makes sense?

One of the essential characteristics of our dataset is its spatial density. Of course, there is no point in complaining about it, since it offers larger opportunities for testing regionalization methods. But we should still be careful, since we do not want our results to be specific to a high hydrologic density environment. As we show it in Figure 3 below, most of the catchments in our test set have a neighbor catchment closer than 25 km. In half of the cases, this distance is less than 10 km.

Such a strong spatial density of available gauging stations makes pure spatial proximity perform really well for practically any hydrological application on ungauged catchments: indeed it is usually the case that regionalization studies built on a dense gauging network find spatial proximity to be a very good regionalization criterion, and in some cases superior to site-similarity (see e.g.Merz and Blöschl, 2005). But in a real world application, we may deal with catchments for which the closest gauge is further away, and anyway, the spatial density of gauging networks may be different in other countries.

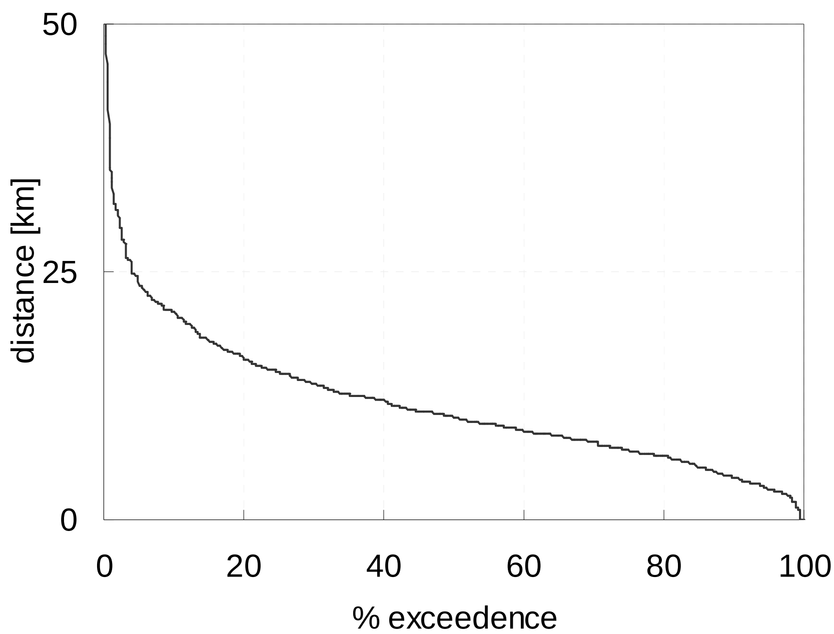


Figure 3: Distribution of the distance of the closest neighbouring catchments over our dataset (distance calculated between catchments' centroids)

An analogous reasoning can be made if we consider catchment's characteristics instead of their positions: having nearly 800 catchments on a relatively small territory means that for each catchment it should be possible to find some very similar donors.

One of the ambitions of this work is to identify regionalization methods that perform well on the majority of our catchments, but also in those few cases where a very close gauged catchment, or a very similar one, are not available. Furthermore, we would like to evaluate how such methods could potentially apply on different datasets, including less-dense ones. For this reason, we need to go beyond the standard testing of regionalization method and imagine a more requiring test, a true "crash test" that will challenge the robustness of the tested regionalization methods to the lack of close and of similar donor catchments.

In the the next paragraph we will introduce a test for the sensitivity of regionalization methods to the lack of close neighbours. A procedure called "metrological desert" has been developed, whose results, and extension to the lack of similar catchments, will be discussed in chapter 9

3.3.2 Assessing the impact of the density of neighbors: metrological desert generation vs random network reduction

In order to build a hydrological "crash test" focusing on the gauging network's density, we must simulate a reduction of the density of our dataset by ignoring some stations. The most intuitive (and also most elegant) method of achieving such an artificially-reduced dataset is to remove from the list of donors a certain number of stations, chosen randomly. As a result, if one looked at the position of these catchments on a map, the reduced dataset would show more or less evenly distributed gauging stations.

However, if we look at our dataset in Figure 4, one can see we face a slightly different challenge: regions where practically every next catchment is gauged and useful for regionalization purposes (not too many human influences) are interrupted by what we called "metrological deserts", i.e. regions where for a number of reasons (mountain regions, flat regions having a mostly artificial river network) the density of the gauging stations is much lower. If we consider an ungauged catchment in one of such metrological deserts, and the distance of the available gauged catchments, we should expect a "threshold" situation: up to a certain distance, there is no gauging station, but just a little further away we might have a very "hydrologically rich" region.

In order to better reproduce such a situation, we developed a different kind of robustness test, which we called "metrological desert", and that will be used extensively in this thesis work. In this case, instead of choosing what percentage of catchments are to be removed from the donors list, we set the desert's radius: when considering a catchment as ungauged, we will ignore all potential donor (gauged) catchments whose centroid is closer than the desert radius.

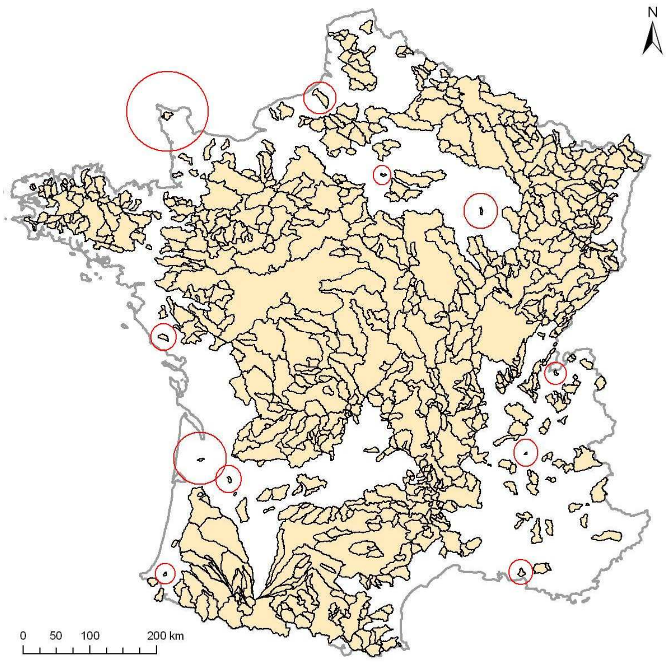


Figure 4: The catchments of our dataset are presented. Red-circled catchments do not have a neighboring basin closer than 20 km

The qualitative difference between the two methods should be clearer by looking at Figure 5: in both examples the same field of ungauged and gauged catchments is considered and the same number of stations is discarded. In the case of the random density reduction we might still have "close" donors, while in the metrological desert case we might still have many donors but none of them will be "close" to the ungauged we are working on.

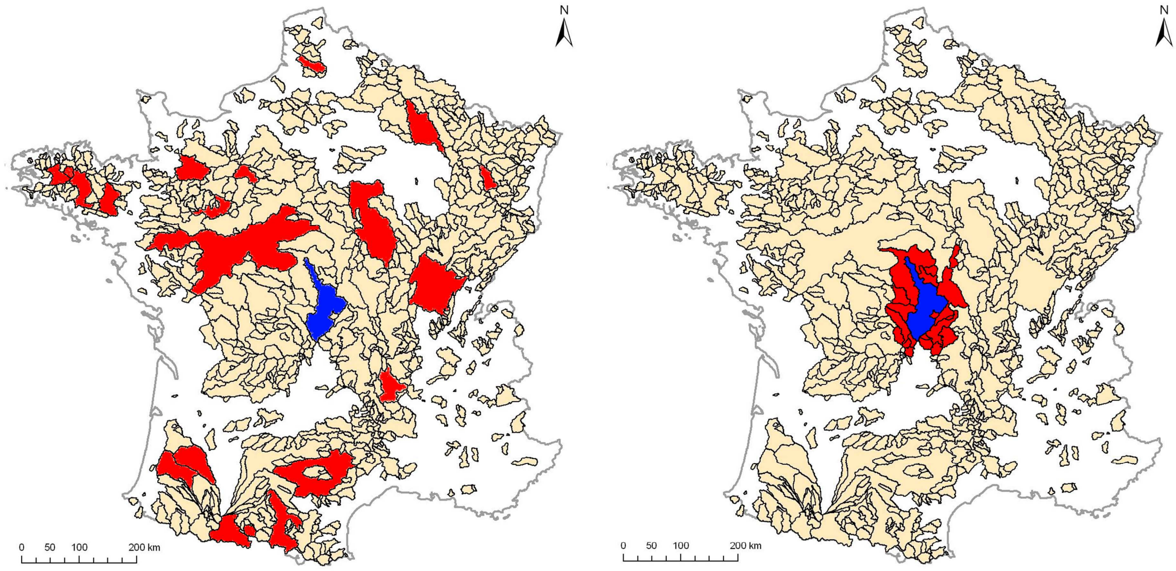


Figure 5: Random density reduction (left) VS metrological desert generation (right). Blue represents the ungauged catchment, beige the authorized donor catchments, red the discarded donor catchments. In both examples, 20 donors have been discarded.

The two methods also differ in how strongly they affect regionalization performance and Figure 6 illustrates this point better. The same model has been regionalized with a similarity-based technique, on a dataset whose density has been reduced randomly in the first case, and with a “desert” approach in the second. It quite clear that the "metrological desert" situation is more challenging than a random network density reduction, even if the number of discarded catchments is much lower (with a 200 km radius, less than 30% of the catchments are discarded). The reason for such behaviour is that, as we discussed in section 4.5.4, similarity metrics and spatial proximity are linked, so that by excluding the closest donors, we are also excluding many of the most similar ones.

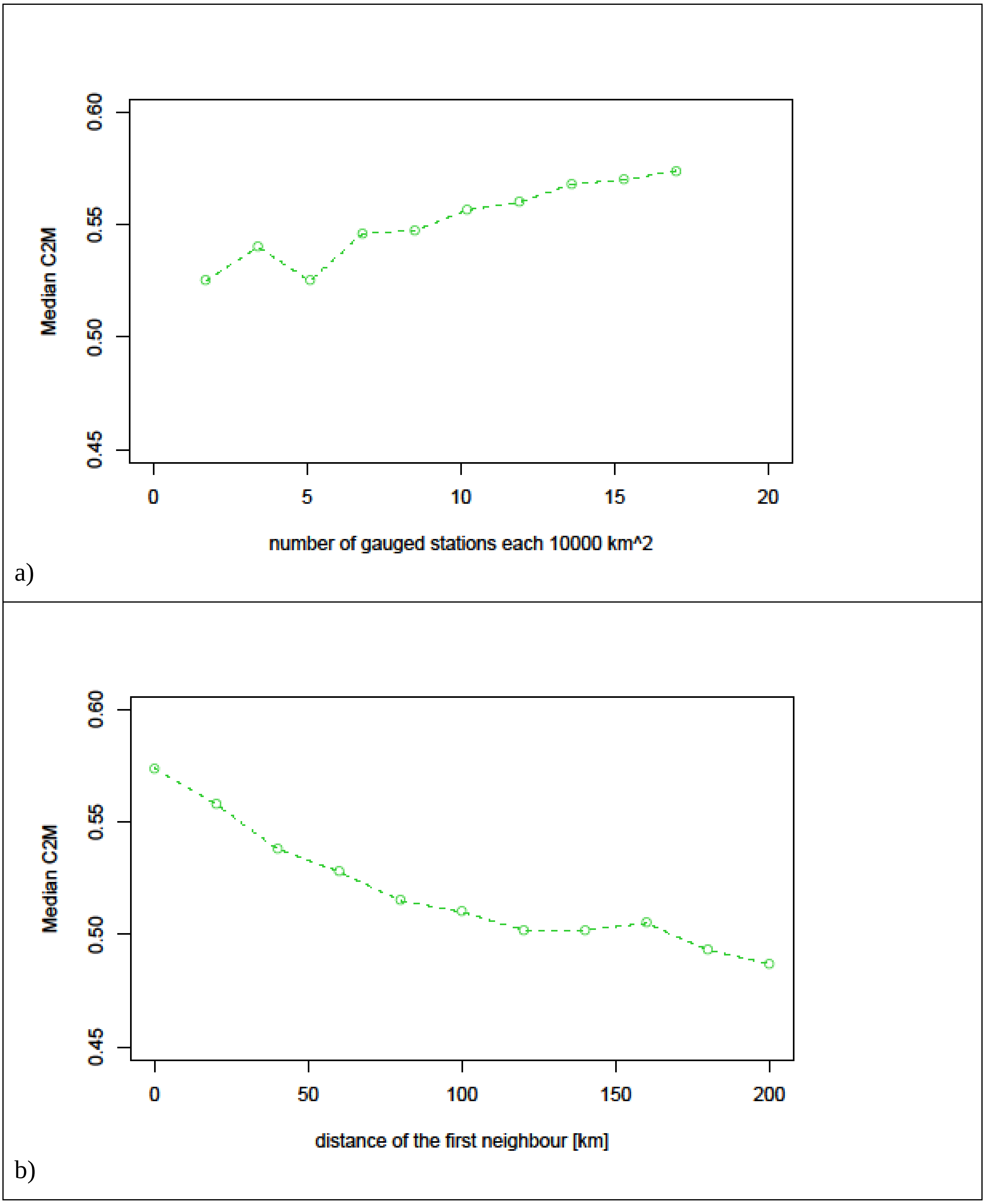


Figure 6: Impact of a progressive reduction of the number of donor catchments on the efficiency of a regionalization exercise (here showed: the median efficiency of a regionalized RR model). a- random reduction of the density of donors, b- creation of a metrological desert.

Finally, when considering the computing time and consistency of the two methods, random density reduction has a disadvantage that led us to discard it. When randomly eliminating donor catchments, it is important to be aware of that the results might differ according to how the random number generation algorithm that drives the test is initialized. To eliminate this issue and achieve a representative result, one should perform several realizations of the same procedure and consider an average or median result. Such undesirable behaviour is particularly emphasized when trying to achieve very low network densities: the number of different realizations required to converge to a "stable" result might be quite high. The "metrological desert" being a worst-case scenario that doesn't rely on a random procedure, does not present this problem.

The main drawback of the metrological desert test, as applied in this work, is its “one size fits all” character: our dataset inevitably has regions of higher hydrological diversity across space (where the test should degrade performances more quickly) and more homogeneous ones (where we might expect a more robust regionalization).

3.4 Synthesis of the methodological choices

In this paragraph we would like to quickly summarize our methodological choices for the rest of the thesis.

- The evaluation of regionalization performances will be done with a jackknife (leave-one-out) cross-validation technique, as outlined in paragraph 3.1.1
- Our performance criteria will be R^2 and RMSE for the regionalization of flow statistics, and C2M, a bounded version of the Nash and Sutcliffe efficiency, for rainfall-runoff simulations (see paragraph 7.1.1)
- We will evaluate the robustness in data-sparse situations with the metrological desert test
- The catchment considered as ungauged will be also referred to as *receiver*, while the catchment whose information is used instead of the missing streamflow record will be called *donors*
- The models used will be linear or log-transformed regressions for flow statistics, and GR4J for rainfall-runoff simulations.

4 Literature review on the regionalization of rainfall-runoff models

This chapter deals with a detailed literature review covering the topic of regionalization. We start by defining the ungaged basin, and then propose a structured analysis of available literature:

- Some authors argue that with ungauged basins, the solution lies in "putting more physically measurable" parameters in the model in order to reduce the dependency on calibration;
- Other authors consider that the solution lies in finding a posteriori a relationship between calibrated parameters and relevant physiographic and climatic descriptors (or geographical coordinates);
- Others prefer to look for "similar" catchments, in order to transfer parameters from them;
- Last, some authors in recent years investigated the potential use of previously made regionalization of some flow characteristics to guide them in the choice of model parameters.

4.1 All agree more or less on a definition for an ungaged basin

In this thesis, we will consider as “ungaged” a catchment for which we do not have any measure of runoff at the site we are interested in, although we have rainfall measurements and at least a few physiographic measurements (that can relate to its climatic, topographic, land-cover characteristics). We also suppose that, for a number of neighbouring sites, we do have runoff records long enough to allow the production of basic statistical analyses and the calibration of parsimonious lumped rainfall-runoff hydrological models, in addition to the rainfall and physiographic measurements.

This definition of ungauged catchment is the one shared by most of the existing literature on the subject, although it is not the most general one: one could define as “ungaged” any catchment for which there is no sufficient data to perform the calibration of a rainfall-rinoff model with the usual techniques. Under this general definition, other specific cases can occur: we will cite for instance a study by Winsemius et al. (2009) that cover the case of "scarcerly gauged river basins, where data is uncertain, intermittent or nonconcomitant" under the name of “ungauged”. Rojas Serna (2006) and Seibert and Beven (2009) take a similar point of view, considering the case of "almost ungaged catchments", with results agreeing that it can be in most cases relatively easy, even with limited resources, to turn an "entirely ungauged" catchment into an "almost ungauged", by performing a small number of well-chosen flow measurements.

4.2 For a hydrological fundamentalist, there is no special problem with ungaged basins…

If hydrological modelling was a mechanistic science, the distinction between "gauged" and "ungauged" catchments would not be of any special interest. Indeed, in such a case, the model would so precisely reproduce the relevant processes generating runoff, that all model parameters could be directly estimated from field measurements of some physical property of the catchment. Runoff records would only have the purpose of model validation, but once such a model would have proved to be valid on a large enough set of catchments without requiring any form of additional calibration, one would be sure of its transferability to an ungauged situation. This situation is ironically described by Richard Silberstein in his 2005 paper entitled *Hydrological models are so good, do we still need data?* In this perfect mechanistic world, the definition of “ungauged” based on the absence of runoff data would be useless, a new one could be given, based on the absence of the adequate field measurements.

We argue that in the present state of the art, an entirely mechanistic model is yet to be seen, possibly not because of a wildly inaccurate understanding of the relevant processes, but rather because nearly all catchments, outside experimental ones, are "ungauged" when it comes to measuring all their hydrologically relevant characteristics at the adequate scale(s). In such a situation, all hydrological models need local calibration, and at least in the short term, the runoff-based definition of ungauged basin is the most relevant.

In the following sections, we will review the main approaches that have been proposed to cope with ungauged basins.

4.3 With ungauged basins, the solution lies in "putting more physically measurable" parameters in the model in order to reduce (suppress?) the dependency on calibration

The growing availability of GIS-related information has driven several attempts of building "physically-based", spatially-distributed modelling, with the inherent hope that these data could be the foundation for a truly deterministic model whose parameters could thus be directly extrapolated from maps of physiographic descriptors, without the need for calibration against runoff.

Such a model would ideally be the most desirable for ungauged applications, but unfortunately current distributed, physically-based models still seem to require some form of calibration against streamflow data. As exemplified by Liu and Todini (2002) and Velez et al. (2009) parameter maps are usually re-adjusted after being initially set at the values obtainable from the descriptors maps. This readjustement is often done by means of "correction factors" that still need to be calibrated for individual catchments, or at least for hydrologically homogeneous regions: as a result, the regionalization of such models is not a very different exercise from the regionalization of a conceptual model having a similar number of degrees of freedom.

Very probably, scale effects play a major role in the current inability to derive model parameters directly from descriptor maps: the "lumping" of hydrological behavior from smaller to bigger parts of a catchment is a highly non-linear process, as pointed out by Beven (1989) and more recently by Todini (2011). The first consequence of this non-linearity is that physiographic descriptors measured at a scale different than the model's "pixels" are not necessairly meaningful. The second one is that, even if the modeling at the pixel scale can be described as “physically based”, this is probably not the case for the fluxes between individual pixels, with the exclusion of channel hydraulics.

4.4 With ungauged basins, the solution lies in finding *a posteriori* a relationship between calibrated parameters and relevant physiographic and climatic descriptors (or geographical coordinates)

We make here a distinction between absolute and relative relationships.

4.4.1 Absolute relationships

The estimation of a mathematical formulation that could calculate parameter values as a function of physical descriptors (such as regressions) has probably been the most popular approach used in early attempts to regionalize conceptual RR models. Such an approach is very tempting as on one side it lends itself well to operational purposes, and on the other it might offer an *a posteriori* interpretation of the relationship between apparent catchment properties and hydrological behaviour.

Unfortunately, this method rarely produces satisfying performances, as its underlying assumptions are very rarely respected. As Oudin et al. (2008) noticed, "there are two hypotheses underlying this approach. First, it considers that the link between observable catchment characteristics and model parameters is univocal, whereas unfortunately, most models have been shown to have no unique set of parameters to describe the behavior of a catchment, and the value of these parameters is more or less dependent on the specific conditions of the calibration period and/or possible errors in inputs (see e.g. Yapo et al., 1996). The second underlying hypothesis is that observable catchment descriptors chosen for regressions bring us an information relevant to the behavior of the ungaged catchment. Unfortunately, the spatial variability of the catchment characteristics and the difficulty to observe underground characteristics constitute a major obstacle in identifying hydrologically-relevant catchment descriptors."

McIntyre et al. (2005) also examined model regionalization in the form of regressions between parameter values and catchment descriptors. Although this technique is considered desirable for the same reasons we have stated above, the authors noticed that so far its application has not been overly successful. The study then looks at the specific points that limit such approach. In agreement with Oudin et al. (2008) parameter identifiability is seen as one of major obstacles: the authors would rather like to identify each parameter's optimum value "in a way that considers its intended (functional) role in the model", but notice that this is seldom the case. Interactions between parameters and model structural error are seen as a possible cause, and multi-objective optimization is proposed as a possible solution: for instance some parameters should be calibrated to an objective function that puts emphasis on high flows, others to an objective function that looks at low flows. The authors also advocate the exclusion of outliers without underlining one possible consequence: the regionalized model will have better results on most catchments, but will potentially make even greater errors on badly regionalized ones.

Wagener and Wheater (2006) also tested a method of "sequential regionalization" similar to the one proposed by Lamb et al. (2000). In this case, a regression is calculated for the most identifiable parameter, then its values are fixed at the regionalized value for all catchments and a new optimization is run under this constraint, a regression is found for the second most identifiable parameter, and the procedure is repeated in sequential fashion until a regression is calculated for the least identifiable parameter. Such a method greatly improves the identifiability of some parameters but, on the other side, its performances in terms of simulation efficiency are on par with a non-sequential approach.

Fernandez et al. (2000) proposed, in order to overcome the limits of regression-based regionalizations, to simultaneously run the calibration procedure and the parameter-descriptor regressions, using a compound objective function that valued simulation efficiency as well as good regression fit. Similarly to sequential methods, such an approach led to a big improvement to a relatively secondary issue (in the form of near-perfect regressions) at the expense of efficient parameter sets, when compared to simpler approaches.

4.4.2 Relative relationships

Other authors have proposed relative relationships between model parameters and catchment descriptors. In this case the shift between model parameter values at two sites is expressed as a function of the shift in physical properties.

In this context Buytaert and Beven (2009) proposes a regionalization framework where prior distributions of parameter sets are first borrowed from a donor catchment that is recognized as hydrologically similar, and then modified according to the uncertainty inherent to the regionalization process and to the knowledge of the differences between donor and receiver. The example given is that of otherwise very similar catchments having a different land cover: such a change should influence the parameter(s) that account for evapotranspiration and can be modified on the basis of the existing literature regarding the specific change (for instance, a switch between grassland and pine forest) or of the hydrologist's own experience. The authors hope that repeated tests on several catchments will refine such parameter shifts, in a way that can be interpreted as the establishment of an *a posteriori* relationship between changes in catchment properties and changes in model parameters.

4.5 With ungauged basins, the solution lies in finding one or more similar catchment(s) (in order to transfer parameters from them)

Among the most common regionalization approaches, and probably the most successful up to date, are regionalization methods that look for gauged catchments that are similar to the ungauged target catchment. The method consists in borrowing some hydrological information from them. Such information is usually put together with a simple or weighted average, and is usually exploited in the form of model parameters, simulated streamflow time series, or (more rarely) time series of rainfall and runoff for model calibration (see Goswami and O'Connor, 2006`, for this last option).

Similarity-based approaches seem able to cope with the difficulties of rainfall-runoff modelling regionalization better than regression methods and "physically-based" models, at least from an operational point of view (i.e. when the only objective is to provide the best possible simulations). However, from the perspective of some of the hydrologists interested in gaining a better understanding of hydrological processes and hydrological modeling through the exercise of regionalization, they offer a less "quantitative" interpretation and can thus seem less attractive.

Two main strategies have been used to find appropriate donors: one is the use of geographical distance as a proxy for hydrological similarity, the other the construction of a similarity metric on the basis of physiographic (sometimes also climatic) descriptors. Of course, many cases of "hybrid" approaches exist.

4.5.1 Methods focusing on spatial proximity

Here, we will first look at some examples of studies in which regionalization methods driven by spatial proximity were proposed or judged to be the most successful (including those that use spatial interpolations such as kriging or inverse distance weighting):

Vandewiele and Elias (1995) estimated the parameters of a monthly water-balance model for 75 Belgian catchments, located in a region thought to be quite hydrologically homogeneous. Two spatial-proximity approaches were compared: kriging interpolation or averaging the parameters of neighbouring catchments closer than 30 km. Kriging gave noticeably better results.

Parajka et al. (2005) tested several regionalization schemes for a semi-distributed model, on 320 Austrian catchments. Methods based on spatial proximity, and especially on kriging of model parameters, performed best, followed very closely by a physiographic similarity method considering only one donor. Similarity was identified on the basis of an *a priori* selection of catchment attributes, which performed better than measures focusing on a single characteristic (such as geomorphology, topography, land use, rainfall, soil classes).

Regression-based methods performed worse. Among them, local regressions (i.e. calculated on catchments closer than 50 km to the target), performed better than global methods (calculated once on all the catchments in the dataset). A later study by (Parajka et al., 2007) found that an iterative regional calibration producing parameter sets that are coherent with regional trends improves the results of a kriging-based regionalization, halving the efficiency loss observed when comparing the results obtained with locally calibrated and with regionalized parameters.

Zvolensky et al. (2008) compared several regionalization methods on 23 subcatchments of the Hron River. A nearest-neighbour spatial proximity approach performed better than an approach using the parameter set calibrated on the whole catchment to model each of the sub-catchments. A theoretical case is also presented, where the most similar donor in terms of parameter values is selected. This method is used to evaluate the potential for improvement of the donor selection: the authors conclude that the use of hydrologically relevant physical descriptors is advisable.

Oudin et al.(2008) compared spatial proximity with a simple physical similarity metric based on similarity ranking of physical descriptor values (judged as a "safer" alternative to the normalization of descriptor distributions). The dataset used was very similar to the one used in this thesis work, quite spatially dense, and spatial proximity overperformed physical similarity, even if the two approaches showed a degree of complementarity (this aspect will be covered in greater detail in section 4.5.4).

4.5.2 Methods focusing on physical similarity

In several other studies (Kay et al., 2007; Li et al., 2009; McIntyre et al., 2005; Reichl et al., 2009), on the contrary, a donor selection based on physiographic descriptors is proposed, or shown to perform better than spatial proximity:

McIntyre et al. (2005) tested on 127 UK catchments a similarity measure based on catchment area, annual rainfall, and hydrological soil classification (BFIHOST). The model outputs obtained using the simulations based on the parameters of the 10 most similar donors were averaged, with a weight based on the similarity between the donor and the receiver catchment. This method showed better result than both regression and spatial proximity. The authors commented that the poor results of the spatial proximity approach might be a consequence of the UK geology, which often changes markedly between neighbouring catchments. In this regard, it is also interesting to note that the results of the tested regionalization approaches were noticeably better on the less permeable catchments: the very permeable catchments (in particular the chalk catchments of southern England) are difficult to regionalize and/or to model with rainfall-runoff (RR) models.

Kay et al. (2007) investigated the use of a site-similarity scheme where a specific catchment similarity measure is built for each of the four parameters of a rainfall-runoff model. In each case, the ungauged catchment parameter is calculated as a weighted average of the values calibrated on the 10 most similar donor catchments. Variations on the numbers of donors, catchment descriptors and weighting scheme to be used have been tested. The performances of the method are considered satisfying and worth the additional operational complexity, when compared to a regression approach. On the other side, the ease with which new data can be incorporated in this scheme is identified as an advantage and a way to avoid the case where the ungauged catchment that should be modeled is too "unusual" compared to the available donors, which leads to unsatisfying results.

Li et al. (2009) considered 210 catchments in south-eastern Australia, on which they applied two lumped RR models. For each pseudo-ungauged catchment, either one or eight donors are identified based on spatial proximity, physiographic similarity, or on a mixed approach that integrates the two. In the case where more than one donor is used, model outputs are averaged instead of the parameters. The authors found that the use of eight donors offers a considerable advantage over the use of only one. When comparing the three methods of donor identification, they notice that the differences in performance are mostly found in the poorer modeled catchments, with the integrated spatial-physical approach slightly outperforming spatial proximity and spatial proximity slightly outperforming physical similarity.

Reichl et al. (2009) discusses the identification of a similarity metrics based on physiographic descriptors, for 184 Australian catchments. The most interesting feature of this study probably lies in the relative sparseness of the dataset (many catchments, but of very diverse hydrological behaviour and spread over a very large territory). As a consequence, the authors notice that a high sampling density across the descriptor space would be needed to identify relevant descriptors without an element of experience and intuition, and to optimize a robust similarity metric (one that is relatively independent of the catchments used to develop it).

Despite the difficult case presented by this dataset, however, physiographic similarity is shown to yield better results than spatial proximity and regression approaches.

Overall, there seems to be no specific indication about whether one of the two approaches to identify donor catchments is superior: the dominance of either one (or of a hybrid method) seems to be case-specific, and in most comparison studies such performance difference is not huge.

4.5.3 How to define similarity?

Similarity-driven regionalization studies are based on the implicit assumption that similar physiographic properties imply a similar hydrological behaviour. Spatial-proximity methods share the same foundation, since in such case geographical coordinates are used as a proxy for physiographic properties that either cannot be easily observed, or whose measurements are not available to the modeller.

Is this assumption correct?

An interesting study by Oudin et al. (2010) focused on this specific subject, using a very similar dataset to the one used in this thesis, with the addition of 10 catchments located in southern England:

- In the study, two catchments would be declared hydrologically similar if the model parameters calibrated on the first could produce acceptable simulations on the second (to ensure that this definition is not overly model-specific the authors repeated the test with two different models, which generally agreed on which catchments were similar).
- Then, similarity in physiographic terms was estimated based on several descriptors regarding topography, climate, land cover and soil properties. As the previous steps allowed to find *n* "hydrological cousins" for a given catchment, an equal number of "physical cousins" was selected the same catchment. Finally, the overlap between the two sets of "cousins" was considered: if it was judged to be statistically significant (i.e. not likely to have happened by chance) then physical similarity was considered a good proxy for hydrological similarity, for the catchment originally considered.

For roughly 60% of the dataset, the overlap between physical and hydrological similarity was judged as statistically significant (both models concurred on this). Yet the most interesting considerations regard those catchments for which physical and hydrological similarity did not agree, for both models: these were essentially hydrologically unresponsive catchments, yet often rather small and steep, which indicates that the origin of the unresponsiveness must be of geologic/ lithologic origin. Since the pool of descriptors did not include a geology-related descriptor and the one related to soil properties performed rather poorly compared to others, this seems coherent with the difficulty to match these unresponsive catchments.

What such a study shows is that on one side the assumptions behind physiographic similarity regionalization methods are essentially good, but on the other the absence of a complete description of all hydrologically relevant physical attributes of a catchment limits its application.

4.5.4 Concerning possible complementarities between spatial proximity and physical similarity

In the previous paragraph we have shown that in many cases physiographic similarity is a good proxy for hydrological similarity, yet in practical application such link might be weaker than expected because we often lack all of the relevant physiographic information.
How does spatial proximity relate to these two definitions of similarity?
In several studies, spatial proximity is presented as if it was a completely independent concept from site-similarity. For instance, Kay et al. (2007) interestingly commented a paper by Merz and Blöschl (2004) where it was found that nested catchments tend to be better donors by saying that "this is more likely due to site-similarity than spatial proximity".
We think, on the contrary, that spatial-proximity is simply a clue for site-similarity and "hidden" (either not measured or not measurable) physical properties. In cases where it works better than approaches based exclusively on "strictly physical" descriptors, it does because we are still not successful enough in understanding and directly quantifying the relevant catchment characteristics, which can however be indirectly guessed thanks to their spatial structure, when a dense enough gauging network is available.
In other cases, either the network density is too low, or the spatial variability of hydrological behaviour is too high, or the available physiographic descriptors are enough to characterize the different "hydrological types" found in the dataset, and so physical similarity is a better guess. For these reasons, there seems to be a degree of complementarity between physical similarity and spatial proximity. Oudin et al. (2008), for instance, showed that if we were able to tell in advance what of the two approaches would work better for a given ungauged catchment, such an ideal "combined" method would perform remarkably well.

Another more subtle case that illustrates the nature of the relationship between proximity and similarities comes from the previously mentioned study by Oudin et al. (2010). In one section, the authors specifically focused on the task of finding French catchments that were hydrologically or physically similar to the English catchments in their dataset, so that the possibility to select spatially-close catchments was excluded. Very interestingly, while catchments which were very physically similar could be found, these were not hydrologically similar to the target ones. On the contrary, when hydrologically similar catchments were found, they were quite far from being physically similar to the target sites!

This example shows that in some ways, our unavoidably incomplete similarity metrics are successful indicators of hydrologically similar behaviour as long as at least some of the donor catchments they select are in relative proximity of the target site. Successful applications on dense, or rather homogeneous network should not be mistaken for a success at measuring all the hydrologically relevant physical properties of a catchment.

4.6 With ungauged basins, the solution lies in using a previously made statistical regionalization to guide us in the choice of model parameters

Some authors have recently advocated an *indirect* regionalization method. This method consists in first regionalizing flow statistics that synthetically describe the hydrologic response of the ungauged catchment. In a second time, parameter sets are chosen according to their ability to reproduce the behaviour outlined by the regionalized statistics.

A few existing studies have addressed the issue of indirect regionalization methods.

The study by Yu and Yang (2000) is probably the oldest one on the subject of indirect regionalization, that is presented as an alternative to regression relationships between model parameters and catchment descriptors. The authors regionalized a flow duration curve by means of homogeneous region identification and regression relationships, then they calibrated the parameters of a modified HBV model so that the flow duration curve calculated on the simulated flows would be as close as possible to the regionalized one. The fit of the two curves was evaluated on ten equally spaced flow duration quantiles, each of which was given equal importance in the final weighting. The authors evaluated the results on two catchments, concluding that the method they used resulted in a good fit on low flows and large errors on the peaks (as the objective function used did not put sufficient emphasis on the latter).

Yadav et al. (2007) used regionalized flow-response descriptors to constrain ensemble simulations at ungauged locations. The flow statistics were regionalized by means of linear stepwise regression on 30 watersheds in the UK, and confidence limits for each regression were also calculated. Subsequently, model parameter sets were randomly generated from a uniform distribution and, for each parameter set and catchment, streamflow was simulated and the previously regionalized flow characteristics were derived. Parameter sets were accepted or refused according to whether such statistics fell into the regionalized confidence boundaries or not. The authors state that "the idea of regionalizing such indices stems from the observation that uncertainty involved in regionalizing hydrologic model parameters can be large […] Watershed response characteristics on the other hand are not model-specific. Therefore uncertainties and confounding influences that might arise from the process of model identification are eliminated (or significantly reduced)".

Bardossy (2007) adopted a very similar method, but in this case the model parameters were picked from ensembles of acceptable sets that were previously generated for neighboring catchments. To be considered "acceptable", a parameter set should yield at least 90% of the Nash and Sutcliffe efficiency provided by the optimal parameters for a given catchment. The idea was to generate a larger variety of possible parameter sets, from which to pick for transfer between pairs of catchments in the dataset, using regionalized mean and variance of the streamflow record as acceptability criteria. An interesting result of this study is that for four of the sixteen catchments considered, no satisfying parameter set could be found in ungauged mode, because all candidate sets produced hydrographs whose response characteristics were too far from the regionalized ones. As in the case of Yadav et al. (2007), dependency of the parameters on the model structure, parameter uncertainty and equifinality were given as reasons to develop an indirect regionalization method.

Montanari and Toth (2007) and Castiglioni et al. (2010) both used indirect regionalization methods derived from Whittle's maximum likelihood estimation approach (Whittle, 1953), which is based on matching the mean value and spectral properties of two time series.

- Montanari and Toth (2007) provide extensive details about how Whittle's likelihood can be approximated for the use in hydrological model calibration, particularly in the case of ungauged or scarcely gauged basins. Results are presented only for the second case, i.e. when historical or sparse streamflow data is used to calculate the flow statistics that are needed to calibrate the model.
- Castiglioni et al. (2010) approximate Whittle's likelihood as a similarity of the mean, standard deviation and lag-1 autocorrelation of observed and simulated streamflow records: it is then possible to calibrate a RR model to regional estimates of these statistics. As one might want to emphasize the role of either one of the three statistics (according to the scope of the regionalization), a Pareto ensemble of non-dominated parameter sets (i.e.

those sets that can not be outperformed on one statistic without doing worse on the other two) is identified. Finally, a streamflow time series is generated as mean of all simulations made with non-dominated parameter sets, and its performance is evaluated as NS efficiency. After application of this method on 52 catchments in central Italy, the authors concluded that "regional calibration procedure is potentially able to convey useful information" but at the same time "it is unlikely that regional information is enough to calibrate a RR model with the reliability that is required in real-world applications". Finally, while the use of an indirect regionalization method is advocated as a useful way to constrain the feasible parameter space, the integration of different information is seen as a necessary element to further reduce its size.

Recently, Westerberg et al. (2010) considered calibration of hydrological models using flow duration curves (FDC), but not using regionalized ones. However, they considered the benefits of such a method in the case when rainfall and runoff records are available, but not for sufficiently overlapping periods, which would be treated as ungauged if a traditional time series calibration was used.

4.7 My opinion (before I started this work), how it evolved, and how the solutions I tried to implement relate to the literature

When first approaching the subject of regionalization, and the literature concerning it, it is relatively easy to fall in the trap of considering it to be a “war” of concurring approaches and methods. For instance, as seen in section 4.5, even methods sharing many assumptions, such as spatial proximity and physical similarity, are often presented as opposing choices, sometimes even as radically different ones!

Such was my perspective at the beginning of this work. This attitude was reinforced by the opportunity of working on a large and diverse dataset, seen as a benchmark that can ensure general conclusions about the good (or bad) performances of a regionalization approach, allowing one to eventually propose a “one size fits all”, robust method.

It was exactly the ambition to propose the most general approach that allowed a slight change in perspective, as a consequence of the robustness test that have been performed. When comparing relatively simple, top-down approaches that are likely to be adopted in an engineering context, it is clear that their relative results often depends quite heavily on the characteristics of the dataset they are applied to, an aspect that is usually not given explicit attention in the existing literature.

As a consequence, this thesis work has shaped itself, in its third and fourth parts, as a comparative study that tries to define the conditions under which each of the tested methods should be expected to give acceptable results and those under which it should be expected to fail, at least in terms of spatial density of the dataset.

On the other hand, this work focuses on the complementarity between different regionalization approaches showed by Oudin et al. (2008), and on how “hybrid” methods can take advantage of it.

Part 2 – Studies relative to flow statistics and their regionalization

In this part, we will focus on the regionalization of flow statistics. The main objective of this part is to explore a complementary and two-step use of physiographic/climatic information versus spatial proximity, on an object that can be considered “simpler” than rainfall-runoff models:

- Chapter 5 presents the first part of this work, aiming at relating flow statistics to physical descriptors;
- Chapter 6 presents the use of neighbour catchments residuals to improve the efficiency of flow statistics regionalization.

5 Linking flow statistics to physiographic descriptors

In this chapter, we present exploratory studies aiming at linking simple flow statistics with physiographic descriptors. As relevant statistics, we have chosen the quantiles of each catchment's flow duration curve. We first discuss the specificities of flow statistics regionalization, then we present and discuss our results.

5.1 Brief review of the literature on the regionalization of flow statistics

The estimation of flow statistics at ungauged catchments is often needed for many engineering problems, and performed, as for the regionalization of rainfall runoff models, by means of a transfer of information from gauged catchments to the site of interest.

Several strategies have been used to identify which gauged catchments should be used as donors of information for a particular ungauged site, all implicitly agreeing on the assumption of an hydrological similarity between the donors and receiver sites.

The oldest (and probably most popular) similarity criterion is spatial proximity: e.g. Darlymple (1960) used it to divide a study domain into geographical regions, assuming that within each one the flood frequency response can be considered homogeneous apart from a scaling factor (the index flood). This popular approach has evolved into forms of geostatistical interpolations that in some cases also use information about the organization of catchments along the river network:

- Sauquet et al (2000) developed a method for the interpolation of average annual runoff based on a geostatistical distance between catchments and on a mass-conservation constraint (the total runoff for a given catchment should be equal to the sum of the runoffs of its sub-catchments). The geostatistical distance between catchments a and b is defined as the mean distance between all possible pairs of points in a and b.
- Skøien et al. (2006) proposed a similar method –called top-kriging- for the interpolation of flow statistics.

Other approaches are based on measurable catchment attributes, such as catchment size, land use, geology, soil characteristics, climatic variables: catchments having similar attributes are assumed to be hydrologically similar.

Similarly to what seen in our review of regionalization for rainfall-runoff models, catchment characteristics can be used to form pooling groups of donor sites thought to be similar to the receiver, or to establish regression relationships between flow statistics and catchment attributes: this latter approach is generally more successful than for the parameters of RR models, as can be seen from the following examples:

- Tasker and Stedinger (1989) proposed linear regressions against catchment descriptors for the estimation of flow statistics at ungauged sites
- Smakhtin et al. (1997) regionalized flow duration curves within a hydrologically homogeneous region in South Africa. The procedure involved the normalization of the FDCs of the gauged catchment used in the study by their mean annual runoff; their average was taken as regional normalized FDC. Mean annual runoff (the scaling factor) was then regionalized by means of a regression against mean annual precipitation and catchment area.
- Mazvimavi et al. (2005) compared the use of linear regressions and neural networks in the regionalization of mean annual flow, flow quantiles and base flow index on 52 catchments in Zimbabwe, finding that linear regressions offered better results on the mean annual flow and the base flow index, while flow quantiles presented a non-linear relationship with catchment descriptors and where generally better estimated with neural networks.
- Longobardi and Villani (2008) regionalized the Baseflow Index in region of southern Italy using linear regressions using a catchment permeability index as the only descriptor.

Only in more recent years a limited number of studies focused on the comparison and on the possible integration of spatial proximity and catchment attributes in the regionalization of flow statistics:

- Merz and Blöschl (2005) compared several methods using spatial proximity and catchment descriptors for the regionalization of flood moments in Austria, finding that methods relying only on catchment descriptors performed noticeably worse than those based on spatial proximity alone or on a combination of spatial proximity and catchment descriptors.
- Kjeldsen and Jones (2010) found that applying a nearest-neighbour spatial-proximity based data-transfer procedure to the residuals of a regression model greatly improves the prediction of the index flood for ungauged sites, particularly when the regression is based on fewer catchment descriptors and the gauging network is dense.

In this regard, our aim is regionalize flow statistics using jointly catchment attributes and spatial organization. This is done on flow statistics, as a first trial before following a similar methodology on RR models.

5.2 Regression as a conceptual model of the relationship between physiographic properties, climate and streamflow

As "entry-level" regionalization of flow statistics, we have chosen to fit a stepwise regression between physiographic descriptors and flow statistics, on the whole dataset.

For sake of simplicity, we assumed that some of the hypotheses under which such regression model is acceptable are verified, even though a more rigorous approach would have required the use of appropriate statistical tests. These hypotheses regard:

- The fact that the stepwise procedure requires that explicative variables are be normally distributed
- The degree of correlation between the explicative variables (multicollinearity). If it is too high, the regression's coefficients will be unstable (small changes in the variables' samples will cause big changes in the coefficients)
- That there is indeed a linear relationship between the explicative variables and the regionalized flows: this should be verified a-posteriori by making sure that the regression residuals are not correlated with the explicative variables.

5.2.1 Nation-wide vs local formulations

The literature abounds with methods aiming at identifying homogeneous regions (or homogeneous pooling groups)(see e.g. Viglione et al., 2007). Regionalization studies are often restricted to some previously selected 'homogeneous domain'. In the first case presented here, where we limit ourselves to a simple regression formula, we could try to identify a specific regression (or at least specific parameters) for each of the homogeneous regions.

We purposely chose not to do so, and to fit only one relationship for our entire study domain. We will show later that it provides the most robust results, (even if the regression on the whole dataset yields a poorer performance, compared to the definition of pooling groups or homogeneous sub-regions on which specific regressions are fitted.

This choice also comes from the desire to treat physiographic and climatic information independently from the spatial (geographic) one. This way, the results we obtain when only using the physiographic information represent our ability to observe the dominant hydrological processes and synthesize them with quantitative descriptors, while the performance gain that we will get when accounting for the geographical position of a catchment represents our ignorance of the relevant processes or the inability to observe them at the appropriate scale.

5.2.2 Selecting relevant descriptors

For the purpose of regionalizing each flow statistic, not all of the catchment information available is necessarily relevant. Furthermore, some descriptors might offer redundant information. Consequently, as our objective was to have synthetic, meaningful formulations, some form of preliminary data mining was necessary.

In order to select the relevant physiographic and climatic descriptors for each of the regionalized statistics, we used a stepwise regression method. The key idea of such a procedure is that each explanatory variable must prove significant, i.e. that the performance increase observed when including it in the formulation has a very high probability to be really an effect of the variable's informative value and a very low probability to have happened by chance.

In more detail, the stepwise regression method we used goes through the following stages:

- For each variable, two regression forms are tested: linear and logarithmic;
- For either of the two regression forms, a "forward-entry procedure" is followed in the first place. Starting from a model with zero variables, the descriptor assuring the best increase in correlation coefficient is added to the regression formulation, and its significance is assessed with a Student's *t*-test. This statistical test considers the hypothesis that the improvement of the correlation coefficient is not due to the new variable, but happened by chance. If this hypothesis has a probability lower than 0.05, the variable is considered significant, and kept in the regression formulation.
- Secondly, a "backward removal" procedure is run on the obtained regression model. Each one of the retained explanatory variables is tested again, to see whether its removal causes a performance drop that might also happen by chance. If the *t*-test gives a probability greater than 0.05, the variable is discarded. This removal procedure is used to eliminate redundancies: for instance it can happen that, during the forward-entry phase, a variable identified as significant in the early iterations is "outdated" by a combination of variables added later on.
- The "forward entry – backward removal" cycle is iterated until no variables can be added nor discarded.
- For all flow statistics, the results of the two regression forms (linear and logarithmic) were compared. In all cases, the log-transformed regression gave a better result and resulted in more retained variables. We will illustrate the retained regression form in the following equation:

$$\ln(\hat{Q}) = \alpha + \beta_1 \ln(x_1) + ... + \beta_n \ln(x_n) \qquad \text{Eq. 1}$$

where $\hat{Q}$ is the regionalized flow statistic in mm per time unit, x_i are the physiographic descriptors, *a* is the constant and b_i are the determined coefficients of regression.

5.3 Streamflow statistics considered and results

5.3.1 Streamflow statistics considered

For all of the available catchments, we calculated the following flow statistics, based on records spanning from 1986 to 2005:

– Average annual runoff;
– Percentiles of the Flow Duration Curve (FDC), normalized by the average runoff. We considered eleven quantiles of the FDC, we will refer to these values according to the percentage of non exceedance (Q_5, for instance, is the value that is exceeded 95% of the time). With this nomenclature, we have: Q_5, Q_{10}, Q_{20}, …, Q_{90}, Q_{95};
– Three "slopes" of the FDC were considered. Those allow to describe the responsiveness of the catchment for high, intermediate and low flow values:

$$S_1 = Q_{30} - Q_5 \qquad \text{Eq. 2}$$

$$S_2 = Q_{70} - Q_{30} \qquad \text{Eq. 3}$$

$$S_3 = Q_{95} - Q_{70} \qquad \text{Eq. 4}$$

5.3.2 List of physiographic descriptors

For each of the studied catchments, we had the following physiographic descriptors:

- Climatic descriptors: Average yearly precipitation *P* [mm], average yearly potential evapotranspiration *PE* [mm], average yearly specific humidity [g/kg], average yearly wind speed [m/s]
- Geographic descriptors: Surface *S* [km^2], Elevation [m], Slope. For elevation and slope, maximum, minimum and average values, as well as quantiles of their distributions, were calculated from a DTM. For naming the quantiles, we'll use the

convention that the maximum value would be labelled Q_{100} and the minimum would be labelled Q_0

- Land use descriptors, expressed as % of the total catchment surface classified under specific classes of the Corine Land Cover European land use database (see http://www.eea.europa.eu/publications/COR0-landcover). We chose to aggregate land-cover classes under the following descriptors: urban (Corine land cover classes from 111 to 124), forest (Corine 311-313), agricultural (Corine 211-213, indicating arable land), fruit olives and vineyards (Corine 221-223), hybrid agricultural spaces (Corine 241-244), other (remaining corine classes)

5.3.3 Results

In Table 4 we present an overall review of the regression results. Average yearly runoff is clearly the better reproduced flow statistic, while for the flow duration curve quantiles a trend emerges: peak flows are better reproduced, while the regressions for lower–magnitude flows can be very poor. A possible explanation is that the available catchment descriptors do not contain relevant information about baseflow formation and connection to larger acquifer systems.

These results are also graphically presented in Figure 7 to Figure 10.

Table 4: coefficient of determination and RMSE for the regressions between flow statistics and catchment descriptors (calculated on log-transformed values). Av_Q stands for average annual runoff.

Variable	R^2	RMSE
Av_Q	0.735	0.335
Q_5	0.308	1.022
Q_{10}	0.31	0.915
Q_{20}	0.365	0.726
Q_{30}	0.441	0.613
Q_{60}	0.546	0.518
Q_{50}	0.65	0.446
Q_{60}	0.697	0.399
Q_{70}	0.727	0.367
Q_{80}	0.734	0.355
Q_{90}	0.71	0.368
Q_{95}	0.669	0.402
S_1	0.506	0.508
S_2	0.714	0.397
S_3	0.587	0.522

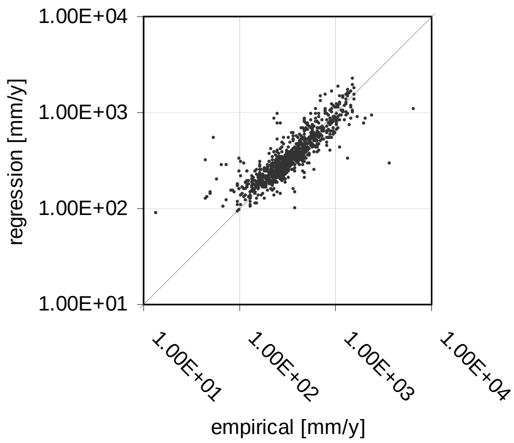


Figure 7: Scatterplot of empirical and regression-calculated values of average yearly runoff

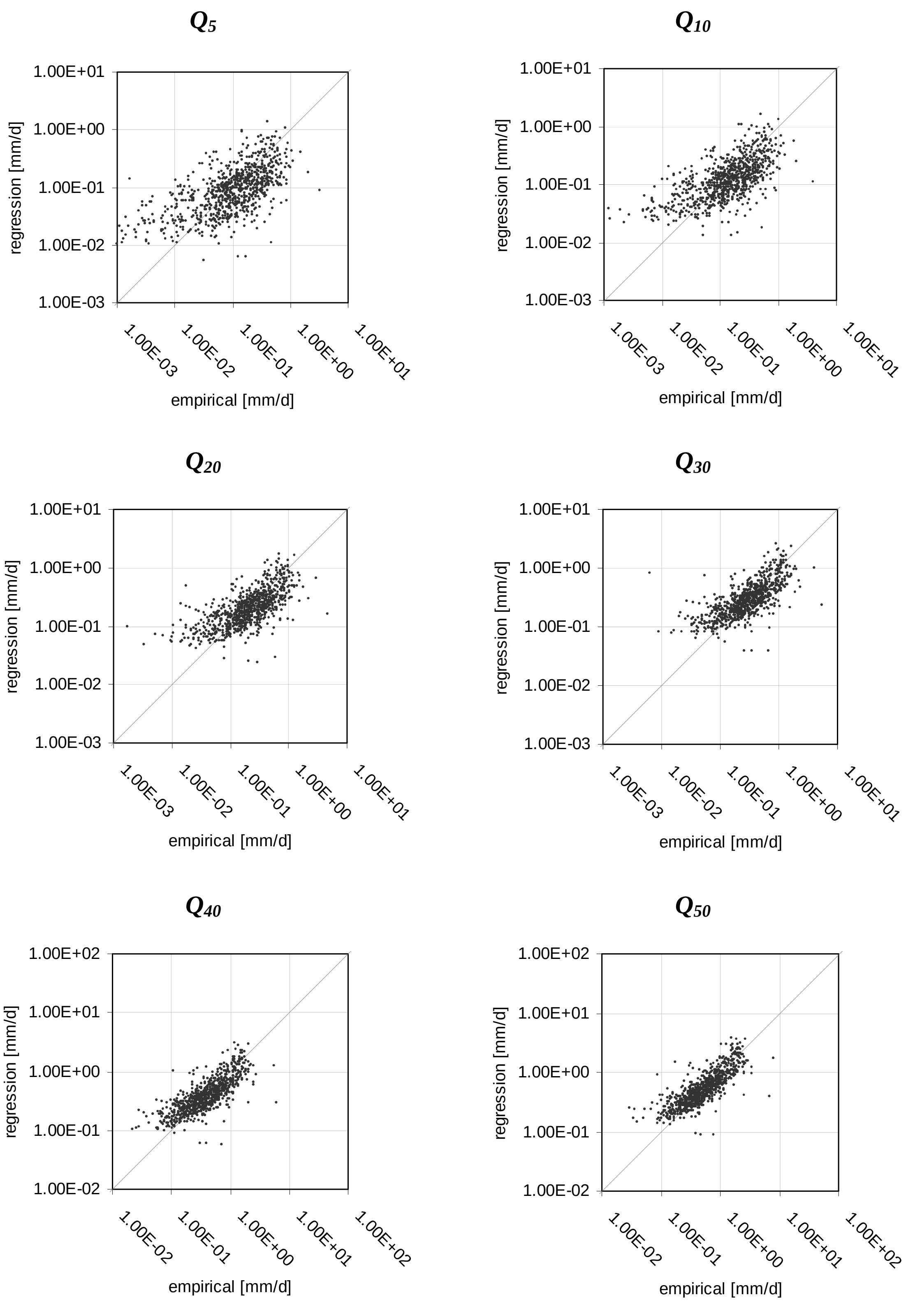


Figure 8: Scatterplots of empirical and regression-calculated values for flow statistics Q_5 to Q_{50}

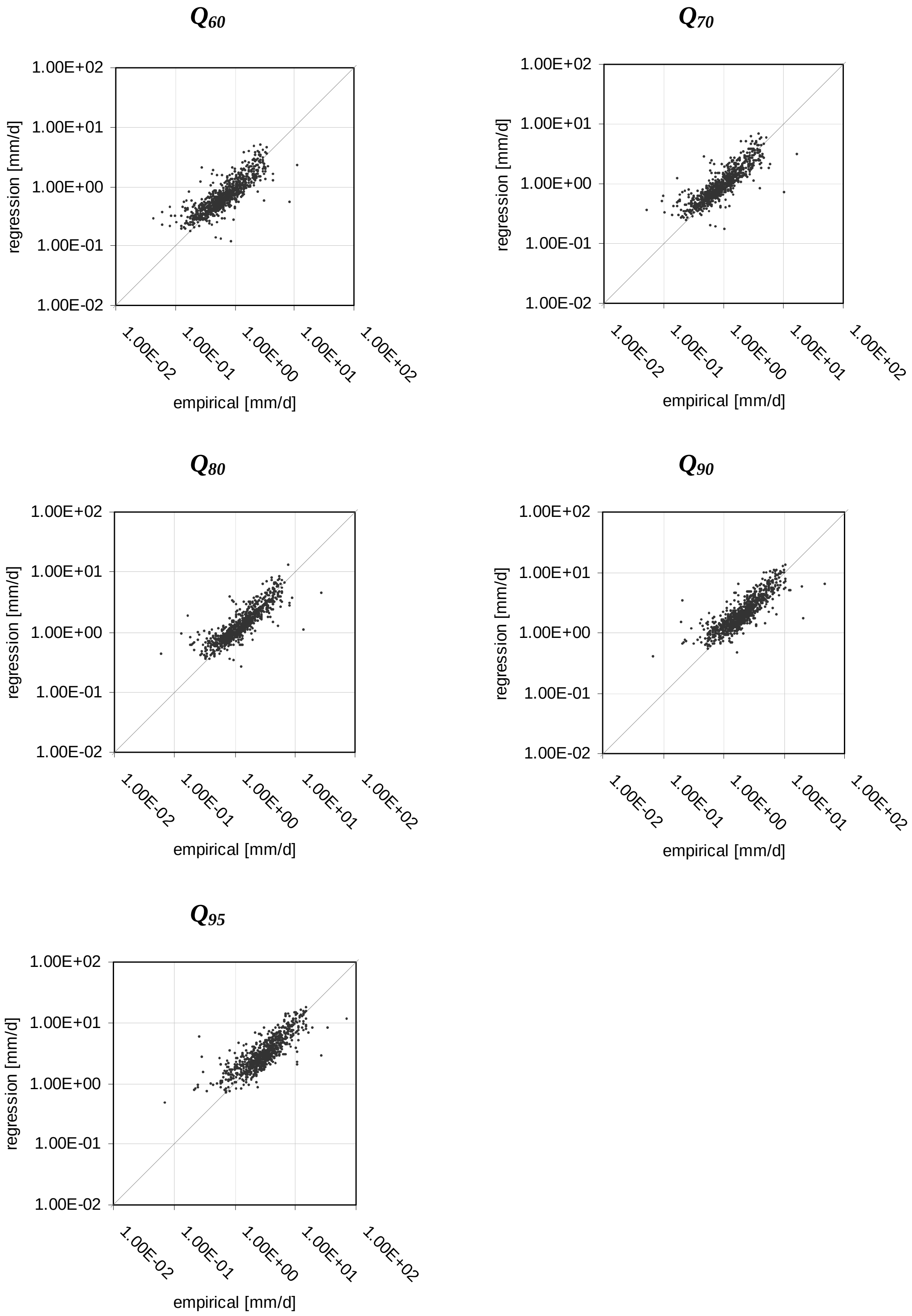


Figure 9: Scatterplots of empirical and regression-calculated values for flow statistics Q_{60} to Q_{95}

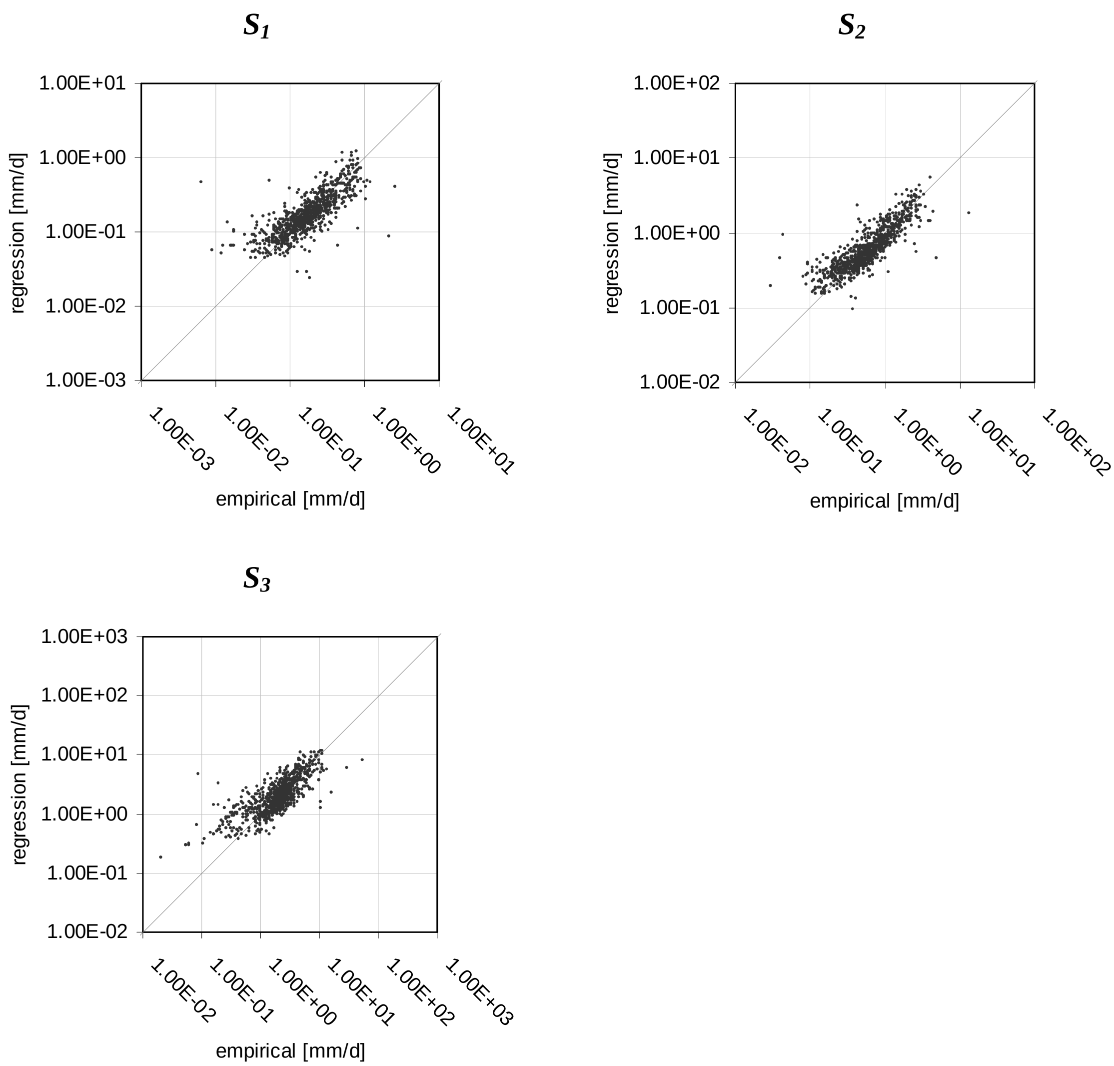


Figure 10: Scatterplots of empirical and regression-calculated values for flow statistics S_1, S_2, S_3

5.3.4 Review of the dependence of the selected statistics on each descriptor

In this section we will have a general look at how the coefficients attributed to each descriptor vary depending on the flow statistic considered, and attempt to explain their hydrological meaning.

Table 5 and Table 6 present the rankings of significance of the selected descriptors for each flow statistic (determined on p-values), and their regression coefficients.

The strongest influence is that of climatic forcings. Mean annual precipitation is the most significant descriptor overall, with a higher significance level for high flows than for low flows, where it is surpassed by evapotranspiration: high PE values mean low values of low flow, and viceversa. On the opposite, low flows are positively correlated to specific humidity and temperature values, and this could be due to the relative simplicity of the formula used here to calculate PE, that only takes latitude and daily temperature into account (Oudin et al., 2005).

Table 5: Rankings of the significance of available descriptors for each flow statistic (threshold at p=0.05).

	Q_Av	Q_{95}	Q_{10}	Q_{20}	Q_{30}	Q_{40}	Q_{50}	Q_{60}	Q_{70}	Q_{80}	Q_{90}	Q_5	S_1	S_2	S_3
P	1	4	2	2	2	1	1	1	1	1	1	1	1	1	1
PE	6	1	1	1	1	2	2	2	2	4	7		2	4	
Hum		3	3	3	3	4	3	4	4	5	8		4	5	
T		6	6	7	7	5	4	8	10				5		
Wind	7									10	10	7			
A	2						7	5	5	2	2	2		3	2
Slope_0.1	3		10		4	3	8				3	3	3		8
Slope_0.2								3	3	3				2	
Slope_0.7															6
Slope_0.8		9	9												
Slope_0.9		8	8	4			11							6	5
SlopeMin	5									6	6	5		7	7
Z_0.1												6			
Z_0.4											5				4
Z_0.6															3
Z_0.9												4			
Z_av											4				
Zmax	4														
URBAN		7	5	8	8	7	6	7	7	9			7		
FOREST		5	7	5	5	6	5	6	9				6		
FRUIT	8	10				9	10	9	6	7	9			8	
HYBRID		2	4	6	6	8	9	10	8	8					

Table 6: Regression coefficients for each descriptor and flow statistic

	Q_Av	Q_{95}	Q_{10}	Q_{20}	Q_{30}	Q_{40}	Q_{50}	Q_{60}	Q_{70}	Q_{80}	Q_{90}	Q_5	S_1	S_2	S_3
P	1.89	1.18	1.22	1.26	1.52	1.58	1.60	1.72	1.77	1.89	1.97	2.02	1.63	2.07	1.93
PE	-0.54	-12.30	-10.80	-9.44	-7.69	-6.09	-5.50	-4.57	-3.92	-2.44	-1.27		-6.18	-2.50	
Hum		7.52	7.11	6.07	5.20	4.00	3.85	3.44	3.22	2.63	1.36		4.30	2.61	
T		2.22	1.84	1.53	1.20	1.00	0.81	0.53	0.37				0.86		
Wind	0.20									0.13	0.12	0.17			
A	-0.11						-0.04	-0.05	-0.06	-0.10	-0.11	-0.15		-0.09	-0.20
Slope_0.1	0.19		0.20		0.20	0.18	0.11			0.17	0.18	0.25	0.21		0.15
Slope_0.2								0.15	0.14					0.24	
Slope_0.7															0.78
Slope_0.8		-1.51	-1.48												
Slope_0.9		2.02	1.73	0.31			0.10							-0.14	-0.71
SlopeMin	-0.12									-0.10	-0.12	-0.17		-0.08	-0.21
Z_0.1												0.22			
Z_0.4											0.73				1.67
Z_0.6															-1.92
Z_0.9												-0.42			
Z_av											-0.82				
Zmax	-0.13														
URBAN		0.13	0.11	0.07	0.07	0.06	0.04	0.04	0.03	0.02			0.03		
FOREST		0.25	0.19	0.17	0.14	0.10	0.08	0.07	0.04				0.09		
FRUIT	-0.02	-0.06				-0.03	-0.03	-0.02	-0.02	-0.03	-0.02			-0.02	
HYBRID		-0.18	-0.16	-0.11	-0.09	-0.06	-0.04	-0.03	-0.03	-0.03					

All flow quantiles show a moderate positive dependence to the lower quantiles of the slope distribution, but if one looks at the descriptor's significance, it is evident that this phenomenon is more marked for the higher flows and for mean yearly runoff. Our hypothesis is that this dependence could be a byproduct of our criteria for choosing catchments: excluding noticeable human influences means we have fewer stations in the zones of aquifer resurgence, and more upstream catchments that, on average, tend to "leak" some water. Among these, those who have "steeper flatlands" tend to infiltrate less that those who are more markedly flat.

Regarding the dependence on the distribution of heights above the sea level, no big trend is shown, and there seems to be a contradictory dependence for high flows: they seem to be moderately related to the height of the bottom of the catchment, but inversely related to the average height, or the height of the catchment's head.

Although small, the dependence on land cover classes is quite interesting, and is generally more marked for lower flows than for higher ones. Our hypothesis is that land cover is partially a consequence of the climatic and/or hydrological character of a region, and then a proxy for it. See for instance Figure 11: high values of "hybrid" land cover are mostly found in regions having an oceanic or partially oceanic climate.

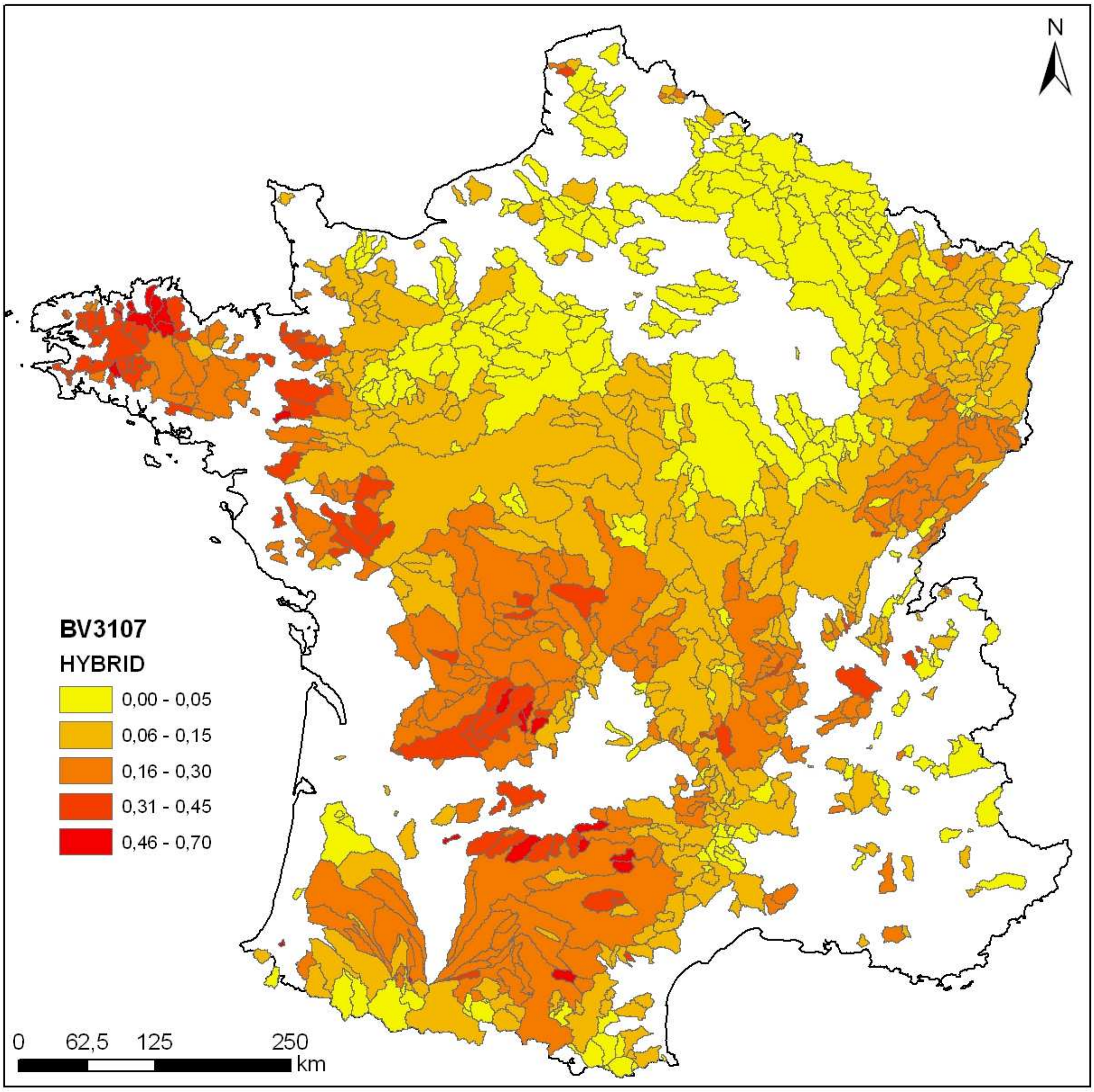


Figure 11: Map of the "hybrid" land cover class, expressed as fraction of the catchment's surface occupied by it. Most of the catchments that are rich with this land cover are climatically influenced by the Atlantic ocean.

6 Using neighbour catchments residuals to improve the efficiency of flow statistics regionalization

In this chapter we will show how geographical distance between catchment centroids can be used as a proxy for those hydrological mechanisms that are poorly related to the available physiographic descriptors, or that cannot be properly modelled with regressions at the national scale.

First, we will present the general Inverse Distance Weighting interpolation technique we used for this purpose, and discuss its results. Then, we will go through two techniques that can further improve the results, based on the surface and on the spatial organization of catchments (accounting for nested donor catchments).

We will also present a specific study (which was published in the Hydrological Sciences Journal) concerning the possible selection of the donor catchments in order to improve the efficiency of residual interpolation.

6.1 Residual's spatial structure as a descriptor of overlooked or not observable properties

Here we call "residual" the following quantity:

$$\vartheta_i = Q_i / \hat{Q}_i \qquad \text{Eq. 5}$$

where $\hat{Q}_i$ is the regression-regionalized flow statistic and Q_i is its empirical value (the one calculated from the 20-year streamflow record).

If one looks at these residuals on a map, a spatial structure is evident, and it is more evident for those statistics which couldn't be reproduced accurately with a regression.

The spatial structure shown by the residuals is likely to be related to information that has been overlooked in the choice of catchment descriptors (such as more detailed descriptors of the climatic forcings) or that is not easily observed, and can be exploited to improve our estimations at ungauged sites.

6.1.1 IDW interpolation

At an ungauged site j, $\hat{\vartheta}_j$ can be estimated as: $\hat{\vartheta}_j = f(\vartheta_i | i = neighbours)$

This can be done with any spatial interpolation technique: here we've chosen to use inverse distance weighting (IDW) for its simplicity, which makes it easy to modify the weights given to each of the interpolated points according to additional criteria (such as whether a gauged catchment is or isn't nested with the ungauged we're interested in, as we'll see in paragraph 6.3).

In IDW, the weight assigned to each "donor" catchment i is calculated as:

$$w_i = \frac{1}{d^{\alpha}}$$

where d is the distance between the centroids of the donor and target catchments.

Then, $\hat{\vartheta}_j$ is obtained with a geometric mean:

$$\hat{\vartheta}_j = \exp\left(\frac{1}{\sum\limits_{\substack{i \\ i \neq j}} w_i} \cdot \sum_{\substack{i \\ i \neq j}} \left(w_i \cdot \ln \vartheta_i \right) \right) \qquad \text{Eq. 6}$$

The inverse distance exponent α has been calculated with an optimization procedure, following a jack-knife technique: each catchment in turn was treated as ungauged and its flow values estimated for a given value of α. A root mean square error on the whole dataset can then be calculated an the inverse distance exponent that minimizes it is chosen.

6.1.2 Results

Table 7 shows the R^2 and RMSE obtained before and after the IDW interpolation of the residuals. A considerable improvement (higher R^2 and lower RMSE) is obtained on all flow statistics, even if for Q_{20} we have a slightly lower R^2 after the interpolation: in this regard, we would like to remember that the inverse distance exponent has been calibrated to minimize RMSE, and it is not assured that this would always lead to a better R^2.

Table 7: Comparison in the results between regression-estimated statistics and regression with IDW interpolation of the residuals.

	R^2		RMSE	
	regression	reg.+IDW	regression	reg.+IDW
Av_Q	0.735	0.770	0.335	0.311
Q_5	0.308	0.407	1.022	0.953
Q_{10}	0.310	0.362	0.915	0.839
Q_{20}	0.365	0.349	0.726	0.680
Q_{30}	0.441	0.437	0.613	0.589
Q_{40}	0.546	0.586	0.518	0.484
Q_{50}	0.650	0.710	0.446	0.404
Q_{60}	0.697	0.742	0.399	0.367
Q_{70}	0.727	0.762	0.367	0.342
Q_{80}	0.734	0.772	0.355	0.327
Q_{90}	0.710	0.758	0.368	0.335
Q_{95}	0.669	0.728	0.402	0.363
S_1	0.506	0.515	0.508	0.503
S_2	0.714	0.762	0.397	0.360
S_3	0.587	0.678	0.522	0.452

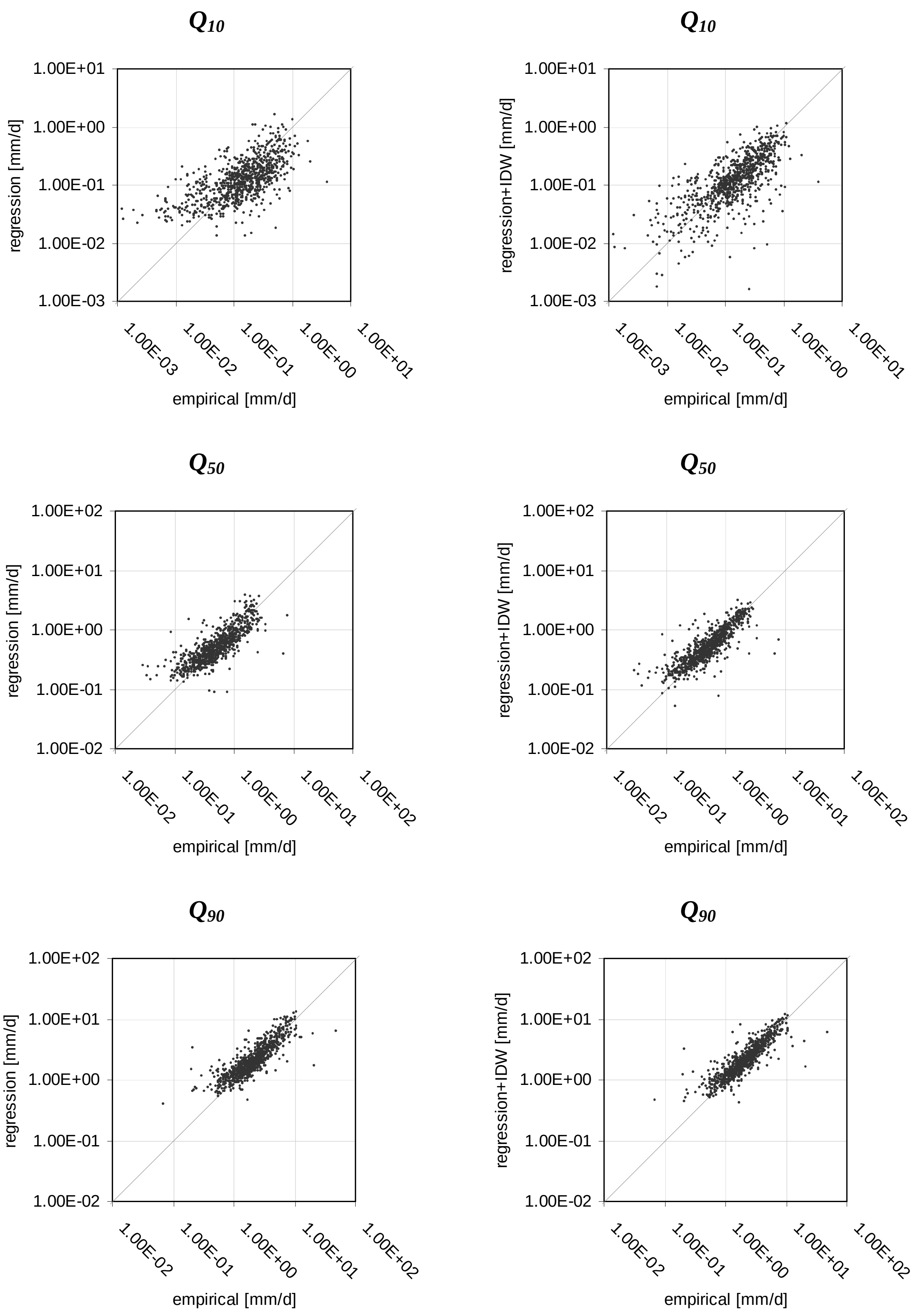


Figure 12: Scatterplots of empirical and estimated flow statistics, with a regression model (left column) or with regression and IDW interpolation of the residuals (right column)

Figure 12 compares the regionalization of some flow statistics with regression only and with IDW interpolation of the residuals. We have chosen to show only three quantiles ($Q_{10,}$ $Q_{50,}$ Q_{90}) since the behaviour is very similar on the other statistics. It can be seen that the additional interpolation narrows the scatterplot cloud and tends to mitigate the biases that tend to especially affect catchments with lower flows.

6.2 Constraints on the surface of donor catchments

With a simple IDW method, it might happen that two catchments (most typically nested ones) with very different areas might have very close centroids, and get very high reciprocal influence. This situation is potentially "dangerous", for several reasons:

- one could logically expect the hydrological behaviour of a large catchment to be different from that of a small nested sub-catchment of a secondary affluent, from a "physical" point of view;
- also, statistical thinking easily leads to expect that if we class all our catchments by size, we should observe much more variability (of significant flow values, or of hydrological behaviour) in the smaller catchments than in the bigger one: big catchments are likely to "average out" extreme behaviours and phenomena that can, on the other hand, be observed at the local scale.

For these reasons, we decided to test limitations on the surface of donor catchments: if the catchment was too small or too large compared to the ungaged receiver, it would not be used as donor.

We optimized the values of the acceptable surface ratios minimizing RMSE, and identified an interesting asymmetrical pattern: no catchment seems to be "too big" to be a donor, suggesting that the smoother, "averaged out" behaviour of the larger catchments provide a safe contribution in the estimation of smaller ones. But there is some improvement when forbidding smaller donors.

Table 8 shows the RMSE on log-values for simple IDW and IDW with area-ratio constraint allows to analyze this phenomenon for different flow quantiles

Table 8: RMSE on log-values for simple IDW and IDW with area-ratio constraint

	RMSE Simple idw	RMSE Area constraint	Area Ratio
Av_Q	0.311	0.141	11.905
Q_5	0.953	0.539	3.226
Q_{10}	0.839	0.467	3.226
Q_{20}	0.680	0.365	5.076
Q_{30}	0.589	0.299	5.076
Q_{40}	0.484	0.232	5.076
Q_{50}	0.404	0.184	5.000
Q_{60}	0.367	0.165	5.000
Q_{70}	0.342	0.153	2.695
Q_{80}	0.327	0.147	11.111
Q_{90}	0.335	0.150	11.905
Q_{95}	0.363	0.162	11.765
S_1	0.503	0.264	5.025
S_2	0.360	0.164	14.706
S_3	0.452	0.214	18.519

In addition to this, we have to say that the surface ratio constraint lost its interest if applied after the donors list was cleaned from outliers with the technique described in chapter 6.4. This is very interesting, because it suggests that the "dangerous" small donors are systematically recognized as "outliers".

These results lead to think that both methods target the extreme behaviours which are very local (specific to small catchments), and can also lead to the following interpretations:

- When estimating the hydrological properties of small, upstream ungaged catchments having lot of downstream data, it is relatively easy to provide reliable, averaged-out, "safe" results, but there is a risk of under-estimating uncertainties and extreme scenarios;
- Conversely, estimating downstream stations with upstream data is likely to produce greater errors, and non-nested catchments of the same size will probably be more helpful than a small nested catchment.

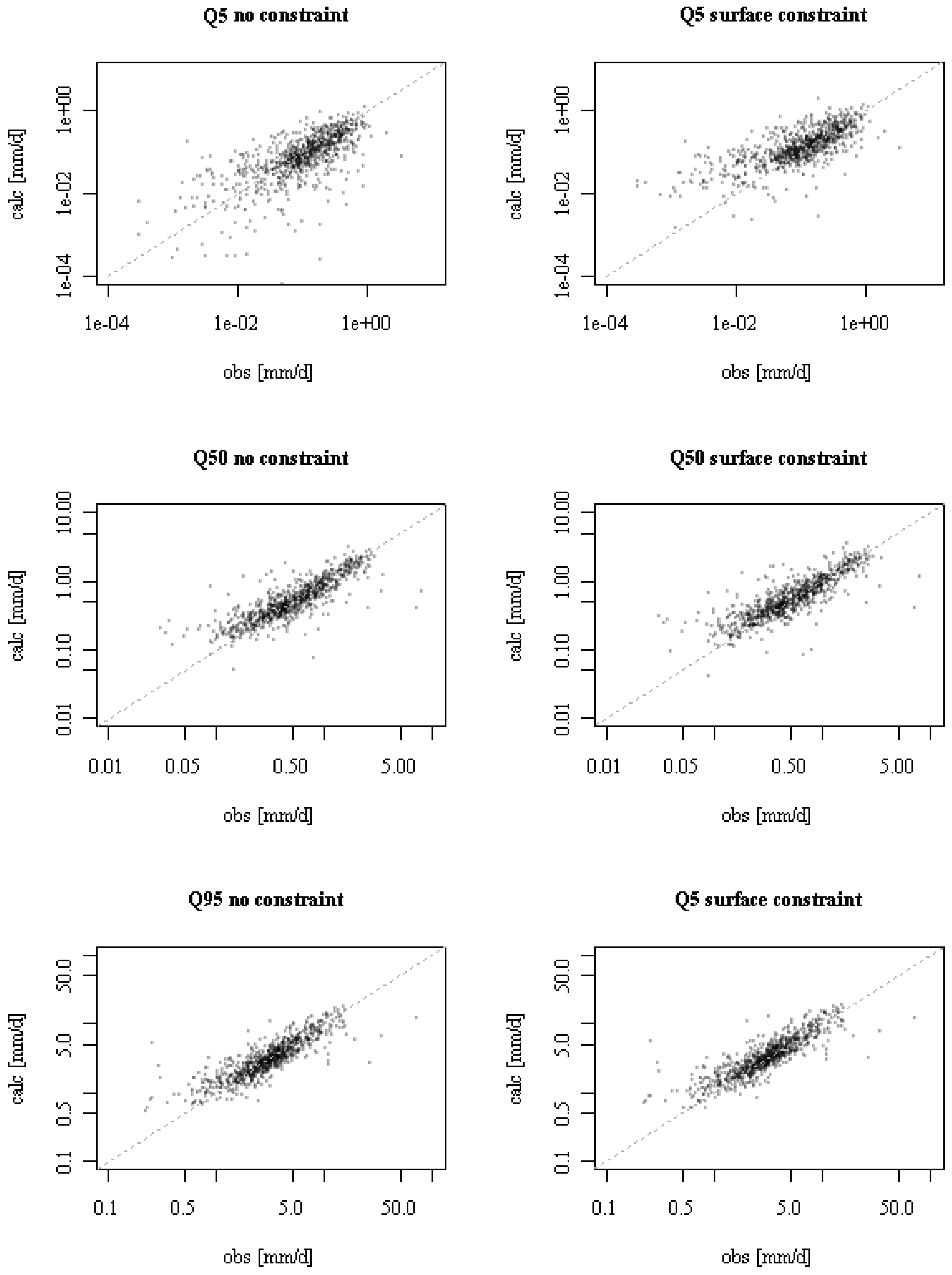


Figure 13: Simple IDW scatterplots (left) confronted with size-constraint method (right). Low to high: Q_{95}, Q_{50}, Q_5

6.3 Accounting for nested donor catchments

It might happen that an ungaged position where we want to estimate some hydrological properties (a reference flow or the parameters of an hydrological model), has one (or more) gauged neighboring station(s) located upstream or downstream on the same river network, so that either the gauged catchment is a part of the ungaged catchment we are interested in, or vice versa. We will say that in this case, the gauged and the ungauged catchments are nested.

Some regionalization studies, especially the one by Merz and Blöschl (2004) found that in a spatially-based scheme, giving more weight to nested donors achieves better results that not discriminating on this basis.

We then decided to test some approaches to account specifically for nested donor catchments. We will present some of the formulations tested, and then discuss which one performs better.

Each donor catchment i is given a weight w_i, that is then used to predict the residual ϑ as a weighted average of the observed residuals ϑ_i. If donor i is nested with the ungauged catchment we are considering , we will modify w_i as follows.

The simplest way to do it is by multiplying w_i by a certain factor *a*, to be calibrated:

a) $$w_i = a \cdot \frac{1}{d^{\alpha}}$$

The exponent α which regulates the weight of the geographic distance can also be modified

b) $$w_i = \frac{1}{d^{\alpha \cdot a}}$$

We can also look at how much area the two catchments share, $f = \frac{Area_{smallercatchment}}{Area_{biggercatchment}}$, and then write the weight w_i as:

c) $$w_i = \left(\frac{1}{1-f}\right)^{a} \cdot \frac{1}{d^{\alpha}}$$

The presented approaches all provide a slight performance increase, but approach c) is clearly offering better performances

Table 9: RMSE on log-values of flow statistics when using simple IDW or IDW giving more weight to nested catchments. The third column shows the esponent "a" presented at point c)

	RMSE Simple idw	RMSE Nested	Exponent
Av_Q	0.311	0.147	22.174
Q_{95}	0.953	0.568	8.967
Q_{10}	0.839	0.497	8.874
Q_{20}	0.680	0.403	10.930
Q_{30}	0.589	0.330	13.719
Q_{40}	0.484	0.258	15.941
Q_{50}	0.404	0.193	18.312
Q_{60}	0.367	0.174	18.144
Q_{70}	0.342	0.162	18.395
Q_{80}	0.327	0.155	17.487
Q_{90}	0.335	0.156	4.931
Q_5	0.363	0.170	3.315
S_1	0.503	0.296	8.177
S_2	0.360	0.170	14.734
S_3	0.452	0.220	1.557

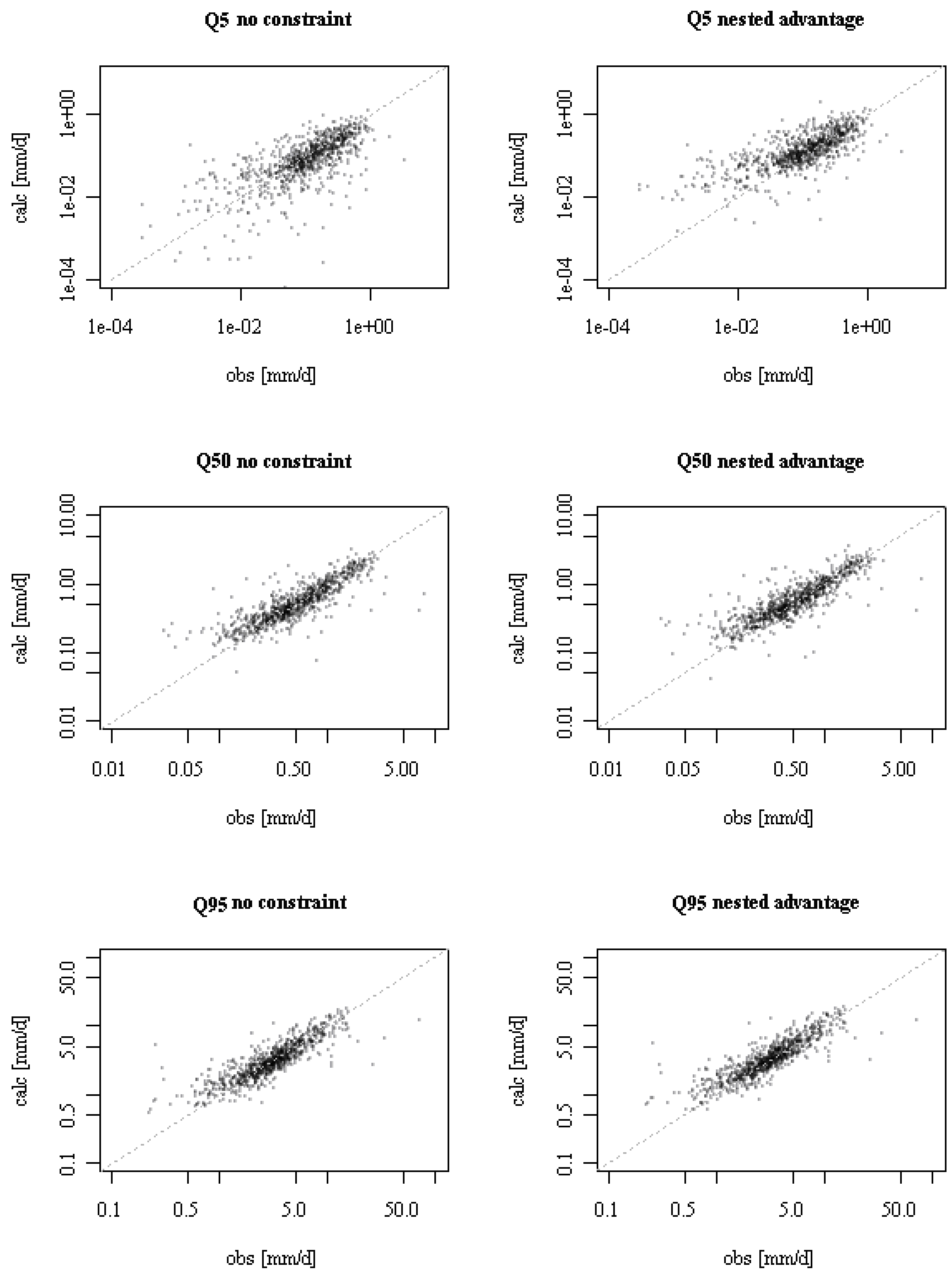


Figure 14: Simple IDW scatterplots (left) confronted with nested donor weighting method (right). Low to high: Q_{95}, Q_{50}, Q_5

6.4 Excluding outliers from the donors' list

One of the issues of every regionalization procedure is the treatment of extreme cases, and where catchments are used as donors, a question arises: should we eliminate the outliers from our donors' list. If yes, which technique should we use to identify them?

The following article, published in the Hydrological Sciences Journal, in the special issue "The Court of Miracles of Hydrology", addresses this point in the context of a two-step regionalization of flow statistics that is in all things analogous to the one presented in this chapter: the core of the method is the use of a regression formulation as first step, followed by an IDW interpolation of the residuals as second step.

A few differences between the context of the article and the rest of this chapter need to be pointed out:

- Instead of FDC quantiles, the article focuses on three flow statistics which are particularly relevant in engineering practice in France, as they are commonly prescribed by the French legislation as project variables;
- The list of available physiographic descriptors used for the article was slightly different from the one used for the regionalization of the FDC quantiles, even if most variables are present in both lists;
- The initial regression between the physiographic descriptors and the regionalized flow statistics was not done using a stepwise regression approach, and the statistical relevancy of the retained descriptors was not evaluated. The regressions were instead obtained empirically by testing all possible combinations of "reasonably few" explanatory variables (no more than five) and then selecting the combination that was subjectively judged to give the best performance for the least number of variables used. The importance given to the use of "as few variables as possible" can be explained as the primary objective of that work was to produce a tool to be used in common engineering practice: there was a fear that a "too complicated" formula would never be used by operational colleagues. The choice of not using a stepwise approach can be defended by saying that while its driving criterion (statistical relevancy) is objective and consistent, such a method still contains a great deal of subjectivity in the choice of the relevancy level to be accepted and of the algorithm to be used (which combination of forward entry and backwards sorting?). In an operational perspective, its advantage lies essentially in the automatization and in the possibility to treat a great number of variables in a less time-consuming way, a

concern that was not existent given the relatively limited number of variables and combinations of variables to be tested.

- The database of precipitation and streamflow records used was, again, slightly different to the one used for the rest of the thesis work. The one used for the main body of this thesis is a more up-to-date version, especially concerning the interpolation of pluviometric records and the inclusion of previously unavailable stations. Out of this renewed database, a new selection of catchments has been made: human influences have been re-evaluated, and it was chosen to work only on records that are reasonably complete over the same 20-year time window, a criterion not used for the previous work.

Outliers are commonly defined as the most extreme values of a sample. When referring to catchments, one possibility is to define the sample as the whole dataset, as done commonly by hydrologists. This means that discarding some outliers would automatically imply a reduction of the dataset's hydrological density, which in our opinion can be counter-productive for regionalization applications.

In the following article we propose an alternative way to identify hydrological outliers: the sample is limited to geographically close catchments, so that outliers will be defined as those catchments whose behaviors differs the most from their neighbors.

The effect of discarding donor catchments which fit this second definition is a smoothing of the hydrological variability the geographical space: local anomalies are ignored. While the overall variability of the dataset is almost unaffected, such procedure produces more conservative estimations of ungauged catchments' flow statistics, and, as a consequence, a more robust regionalization (if the outlier discarding technique is applied to the right degree)

6.5 Final considerations on the results obtained for the regionalization of flow statistics

In the last two chapters, we reviewed a regionalization method for flow statistics based on regression with IDW interpolation of the residuals. This approach is advantageous if compared with a method based on regression alone, and this advantage is greater if one takes in account the relative size of donor catchments compared to the receiver, and their position on the stream network, as explained in sections 6.2 and 6.3.

From an operational point of view, it would be interesting to test whether on a dense gauging network such as the one France has, similar results could be obtained with the use of less catchment descriptors.

From a scientific point of view, we remark that the best regression performances are obtained on average annual flows and on higher-than-median flow quantiles, while low flows tend to get poorer ones. This result and the considerations made in paragraph 5.3.4 lead to the conclusion that we are quite successful at explaining flow statistics that are more directly linked to the climatic input and to the short-term catchment response, while our failure on low flows is probably linked to the lack of adequate descriptors to characterize the long-term hydrologic response of our catchments.

In chapter 10 we will be using regionalized flow statistics to constrain the regionalization of a rainfall-runoff model. In that case we will use a regression + IDW approach (as an example of regionalization on a dense network) and a regression without IDW (as an example of regionalization on sparser networks).

Part 3 – Regionalization of rainfall-runoff models – direct methods

This part focuses on the regionalization of rainfall-runoff models with *direct* methods, i.e. methods that use the available physiographic, climatic and spatial information to identify good donor catchments from which parameter sets are borrowed. This exercise differs from the way in which physiographic and climatic information are used when regionalizing flow statistics, which is essentially regression-based.

Chapter 7 deals with a method based exclusively on physiographic similarity;

Chapter 8 presents two methods (intersection-based and union-based) to combine the benefits of spatial proximity and physiographic similarity.

7 Physiographic similarity regionalization

In this chapter, we explore regionalization methods that are based upon the construction of a similarity metric, which is used to select appropriate donor catchments: as seen in section 4.5, this is probably the most common regionalization approach for rainfall-runoff models.

Such metric can be built in several ways using the available physiographic and climatic descriptors: here we will present two possible methods, that we will call PCA-based (based on a preliminary selection of explanatory variables using PCA) and backwards-sorting (based on backwards sorting of explanatory variables).

7.1 Introduction

In this section we will present two methods to build a site similarity measure out of physiographic descriptors.

In any regionalization exercise, the successful use of physiographic information is attractive for two main reasons:

- A physiographic-based site-similarity measure can easily be confronted with one's understanding of the hydrological processes, at the opposite of spatial proximity.
- One would expect that a similarity metric relying on physical attributes would be more robust when applied to a scarce network of gauged stations, if compared to spatial proximity. This is a reasonable assumption, even though not always verified (Oudin et al., 2008).

The construction of a similarity metric based on physiographic measures faces a few main issues:

- The selection of hydrologically relevant descriptors: if a non-relative descriptor is used during the construction of the similarity metric, it will at best have a neutral effect, and, at worst, a very detrimental one.
- the ranges of variation and the distribution of observed values can differ significantly from one descriptor to another. This poses a problem when one tries to build a metric based on such variables. To overcome this problem, several approaches are possible: here we will, for both methods, normalize the descriptors so that their mean equals 0 and their standard deviation equals 1 (an assumption is made that the distributions of the observed values have similar shapes)
- Physiographic descriptors are often correlated between them to some degree. This implies that some catchment characteristics might be overemphasized in the similarity metric, unless an appropriate weighting/variable selection scheme accounts for this. Correlated descriptors also implies that the similarity metric shouldn't be thought as an Euclidean distance, even when it is built as if it was one, unless a set of uncorrelated explanatory variables is derived from the correlated descriptors (through e.g. Principal Component Analysis).

7.1.1 Common points of the tested regionalization methods

It is important that our readers are aware of the general scheme shared by all the tested regionalization methods.

In all of the following paragraphs, we will test each of our catchments as if they were ungauged, in a jack-knife fashion. At the same time, we will suppose that all the remaining catchments are known, unless when challenging a method's robustness with the "metrological desert" crash-test.

The point of each of the regionalization methods tested is to select a group of donor catchments: these are supposed to be hydrologically similar to the one we treat as ungauged, i.e. a parameter set calibrated on one of the donors should give comparable results on the ungauged.

Once a set of donors is chosen, a simulation is run with each of the donor's calibrated parameter sets, then the obtained time series of streamflow are averaged: the obtained record is the candidate simulation for the pseudo-ungauged catchments and its efficiency will be calculated. This procedure is followed instead of the averaging of parameters since Oudin et al. (2008) showed, using the same model and a very similar database, that flow-averaging gives consistently better performance than parameter-averaging

For each of the proposed methods, the optimal number of donor catchments is set by using the median efficiency obtained on the whole database as a criterion.

The criterion being used is C2M, a bounded version of the Nash-Sutcliffe efficiency whose maximum is 1 and whose minimum is -1, calculated on square-rooted flows.

$$C2M = \frac{NSE}{(2 - NSE)} \qquad \text{Eq. 7}$$

While representing the same concept of the NSE (a comparison between the square errors obtained by a simulation and those obtained by an average of the modeled time series), C2M has two practical advantages:

- it can be averaged (especially useful when the objective is to get "less bad" simulations, rather than improve the peak performances)
- due to the re-scaling effect, higher performances are spread over a broader range of values and can be better evaluated.

See Mathevet et al.(2006) for more details on this criterion.

7.2 Method based on Principal Component Analysis

We first present a method that we call PCA-based because it does not attempt to judge the "hydrological value" of the available physiographic and climatic descriptors when constructing the similarity metric. The only treatment that is applied to explanatory variables is Principal Component Analysis, used to ensure that an Euclidean distance can be "properly" calculated.

7.2.1 Preliminary selection of explanatory variables

This method is characterized by the fact that we didn't try to select the best possible descriptors to be used for the regionalization task, but rather to just discard those that are clearly not useful.

The selection criterion was thus built in comparison with a random choice of donor catchments, with the following procedure:

- For each catchment treated as ungauged, ten different catchments were randomly chosen from our dataset and used as donors.
- The calibrated parameter set of each donor was used to run a simulation on the pseudo-ungauged catchment.
- The time series of streamflows obtained with the ten simulations were averaged, and an efficiency criterion (C2M on square-rooted flows) was calculated.
- The procedure was repeated on the whole dataset several times, with different initializations of the random selection
- The average efficiency of all the simulations was retained as the benchmark to consider a physiographic descriptor acceptable

Each descriptor was used in turn as a similarity measure. Ten catchments having the closest descriptor values to the one treated as ungauged would be used as donors, ten simulations would be run and an average time series was obtained. The average efficiency obtained on the dataset was then compared to the random benchmark, and the descriptor would be discarded if worse than this random benchmark.

As one can see in Table 10, only the land-cover class "fruit" gives results that are worse than a random selection of ten donors, and has not been used in the construction of the similarity metric described in section 7.2.2. It is interesting to note that this land cover class is only found in significant extensions on very few catchments of our dataset, and this alone is a good reason for its exclusion: for the many catchments who have a modest or null coverage of such a class, it won't provide a reliable indicator for site-similarity of any kind, even outside of the context of a hydrological study.

Another interesting point is the relatively high relevance of topographic descriptors, compared to climatic ones. This, put into the perspective of the results obtained for flow statistics, tends to reinforce a belief that model parameters are much less climate-dependent, which is quite reassuring. See section 2.3 for details on each physiographic descriptor.

Table 10: Average efficiencies obtained when using only one physiographic descriptor to define site-similarity

Descriptor	Average efficiency
A	0.385
Zmax	0.371
SMin	0.371
Z_0.6	0.364
Z_0.4	0.362
Z_0.9	0.359
SMax	0.358
Z_0.7	0.358
Z_0.2	0.357
SAv.	0.357
ZAv.	0.356
OTHER	0.355
P	0.354
S_0.4	0.353
FOREST	0.353
Z_0.8	0.353
S_0.3	0.353
Z_0.5	0.353
S_0.1	0.353
Wind	0.352
S_0.9	0.352
S_0.5	0.352
S_0.6	0.351
S_0.2	0.350
Z_0.1	0.350
HYBRID	0.350
DD	0.350
T	0.349
Z_0.3	0.348
S_0.7	0.348
Hum	0.347
PE	0.347
URBAN	0.344
AGRIC.	0.344
S_0.8	0.343
Zmin	0.341
Random Donors	0.322
FRUIT	0.316

7.2.2 Principal Component Analysis as a tool to overcome the issue of correlated descriptors

Once the unnecessary descriptors were discarded, the issue of correlated descriptors was solved using Principal component analysis (PCA). PCA is a well known mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components.

The first seven principal components (explaining 80% of the descriptors' variance) have been derived from the original set of descriptors and used to build a similarity metric. Since the principal components are orthogonal by definition, one can use them to calculate an Euclidean distance between data points (in our case, an ungauged catchment and a potential donor)

$$d = \sqrt{\sum_{i=1}^{n} (pc_{i,u} - pc_{i,c})^2} \qquad \text{Eq. 8}$$

Where the $pc_{i,u}$ is the i-th principal component for the ungauged catchment we are interested in, and $pc_{i,c}$ is the i-th principal component for candidate donor catchment c

7.2.3 Results

Figure 15 shows the general performance of the PCA-based regionalization. The top left chart shows the median C2M efficiency on square-rooted flows obtained when using different numbers of donors. It appears that the optimal number of donors for such a method is six, and in this case the median C2M equals 0.56 (corresponding to NSE=0.72).

The second chart shows the complete distribution of efficiencies obtained on our database catchments, when using six donors and the PCA-based method (black solid line). Only positive efficiencies are shown.

Two grey lines are added to the plot as benchmarks. On the rightmost side we have the performance of a calibrated model (solid line): it represents the performance of an ideal (non existent) regionalization method that would be able to totally substitute the information contained in streamflow time series.

On the other side we have a "minimum demand" benchmark, represented by a random selection of ten donors (dashed line): it represents a method that is truly blind to the ungauged catchment considered.

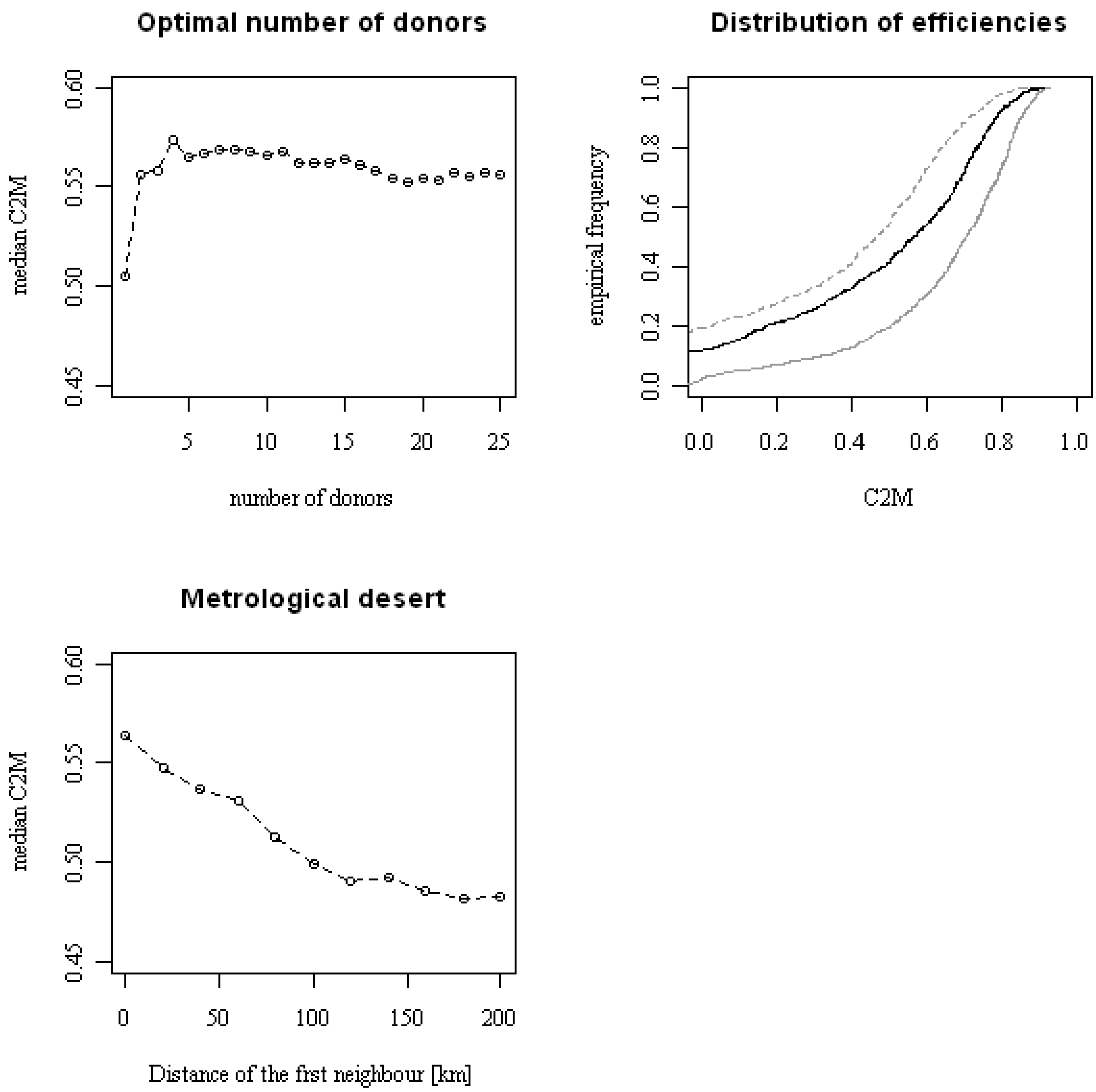


Figure 15: PCA-based Regionalization performances. Top left, median efficiency per number of donor catchments used. Top right, distribution of efficiencies compared to a random selection of donors (dashed line) and calibrated model (solid grey line). Bottom left, performance in a "metrological desert" situation.

These benchmarks are extreme and all regionalization method should fall between them: nevertheless, they should help judging regionalization quality.

Figure 15 shows that there is a large room for progress for the regionalization approach, since its performances are intermediary to the two benchmarks. Besides, the approach does not show a remarkable robustness: excluding donors in a 100 km radius leads to a strong decrease in efficiency (near the efficiency of random donors), meaning by the way that the similarity approach tends to select geographically close donor catchments.

Last, Figure 16 provides an outlook at the shape of the hydrographs obtained in calibration and with the regionalization model.

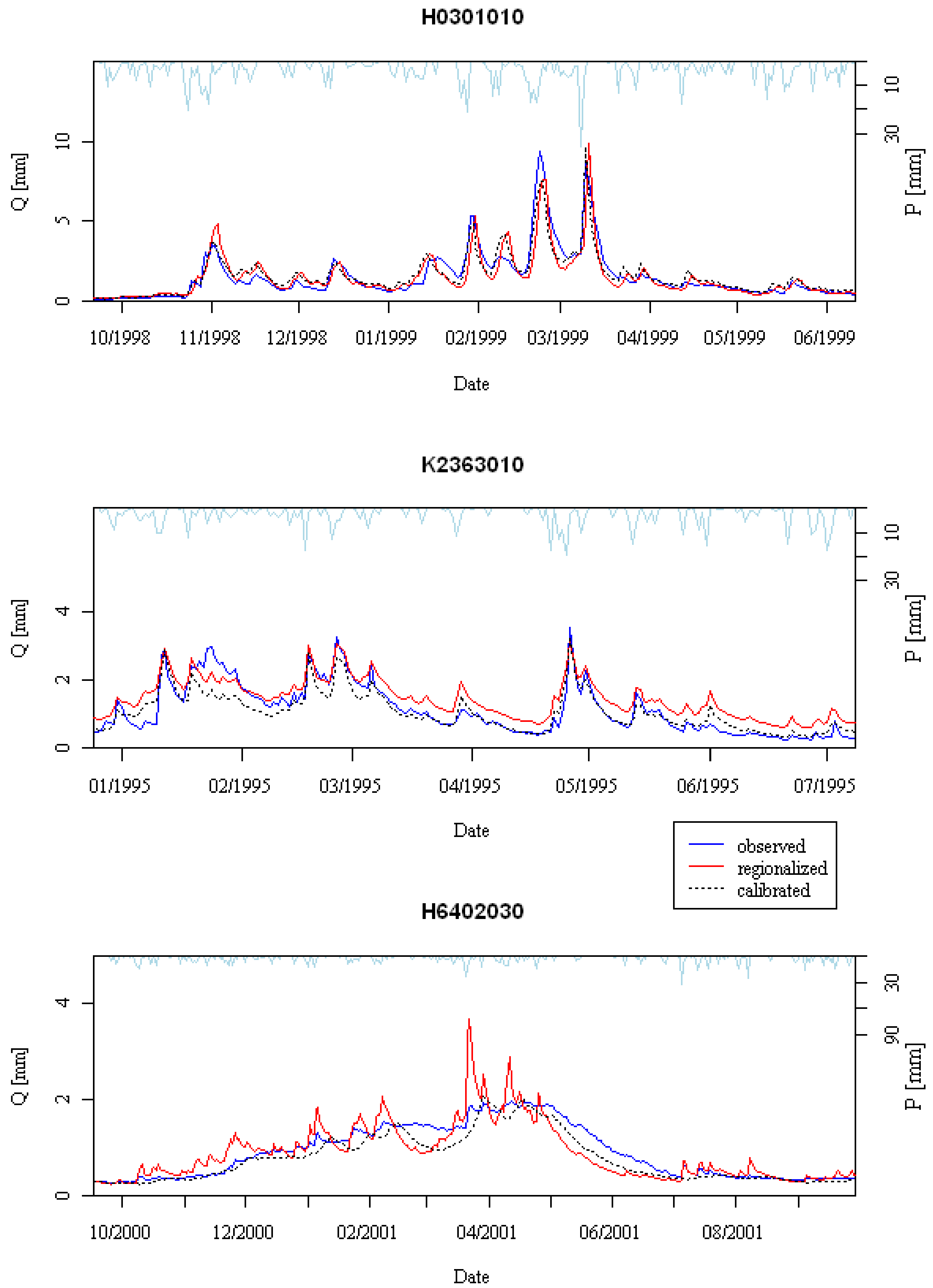


Figure 16: Time-series of observed, regionalized and simulated (with prior calibration) streamflows on three example catchments, of good (H301010) "median" (K2363010) and poor (H6402030) performances

7.3 Backwards sorting method

In this section we present a backwards-sorting method for the construction of a similarity metric. In this case, the available explanatory variables undergo a "backwards sorting" selection, aimed at retaining only the most "hydrologically meaningful" physiographic and climatic descriptors.

7.3.1 Variable selection algorithm

In this method, the selection of physical descriptors was made on the basis of the quality of the obtained regionalization.

As a first step, available descriptors are normalized:

$$ds_{i,j} = \frac{d_{i,j} - \bar{d}_i}{\sigma_i} \qquad \text{Eq. 9}$$

Where $ds_{i,j}$ is the normalized value of descriptor *i* for catchment *j*, $d_{i,j}$ is the un-normalized value, $\bar{d}_i$ is the average of d_i over the dataset, and σ_i is its standard deviation.

A "pseudodistance" is then built using all *n* available descriptors, as if they were orthogonal:

$$p = \sqrt{\sum_{i=1}^{n} (ds_{i,u} - ds_{i,c})^2} \qquad \text{Eq. 10}$$

where $ds_{i,u}$ is descriptor i for the ungauged catchment u, and $ds_{i,c}$ is descriptor I for the candidate donor catchment c.

A regionalization procedure using the ten most similar candidate donors is run and its average efficiency is calculated.

The second step involves running the same procedure, this time not using one of the descriptors. This is repeated until all combinations of *n-1* descriptors have been tested, and the one giving the best performances is kept as the new descriptor list.

The whole procedure is iterated until we only have one descriptor left. At this point, if at each iteration the retained combination of *n-m* descriptors was noted along with its average efficiency, the selection of the optimal pool of physiographic descriptor is trivial.

Table 11 shows a list of the descriptors discarded at each iteration, and the average efficiency obtained with the remaining ones. As one can see, the maximum efficiency has been reached at the 11th iteration, so the retained similarity metric uses the 26 remaining descriptors[2] (the descriptors in italic characters have been removed from the list).

Table 11: List of discarded descriptors at each iteration

Iteration	Discarded descriptor	average C2M
1	*Z_0.9*	0.4675
2	*OTHER*	0.4683
3	*SMax*	0.4694
4	*AGRIC.*	0.4701
5	*Z_0.2*	0.4706
6	*S_0.1*	0.4707
7	*S_0.6*	0.4708
8	*HYBRID*	0.4715
9	*DD*	0.4725
10	*FRUIT*	0.4729
11	*Z_0.1*	$\boxed{0.4730}$
12	S_0.7	0.4721
13	S_0.3	0.4724
14	S_0.2	0.4723
15	Z_0.6	0.4720
16	Z_0.3	0.4718
17	S_0.4	0.4722
18	Z_0.4	0.4728
19	Z_0.5	0.4727
20	PE	0.4722
21	Z_0.7	0.4715
22	T	0.4720
23	ZAv.	0.4710
24	Zmax	0.4705
25	S_0.8	0.4692
26	S_0.5	0.4684
27	URBAN	0.4652
28	SMin	0.4652
29	FOREST	0.4638
30	Z_0.8	0.4623
31	Wind	0.4531
32	DD	0.4435
33	P	0.4269
34	S_0.9	0.4118
35	Zmin	0.3935
36	Hum	0.3854
37	A	0.3494

7.3.2 Results

Figure 17 summarizes the performances of the presented regionalization method with the same scheme used in Figure 15.

[2] It should be noted that nothing ensures that the 26 selected descriptors are uncorrelated. If they aren't, the resulting dissimilarity measure can't be considered as an Euclidean distance.

In this case, the optimal number of donor catchments is 7. The distribution of the efficiencies is slightly, yet consistently better than the one obtained with the PCA-based method, with a median C2M of 0.57 (NSE=0.73) and also a slight advantage in the "metrological desert" robustness test.

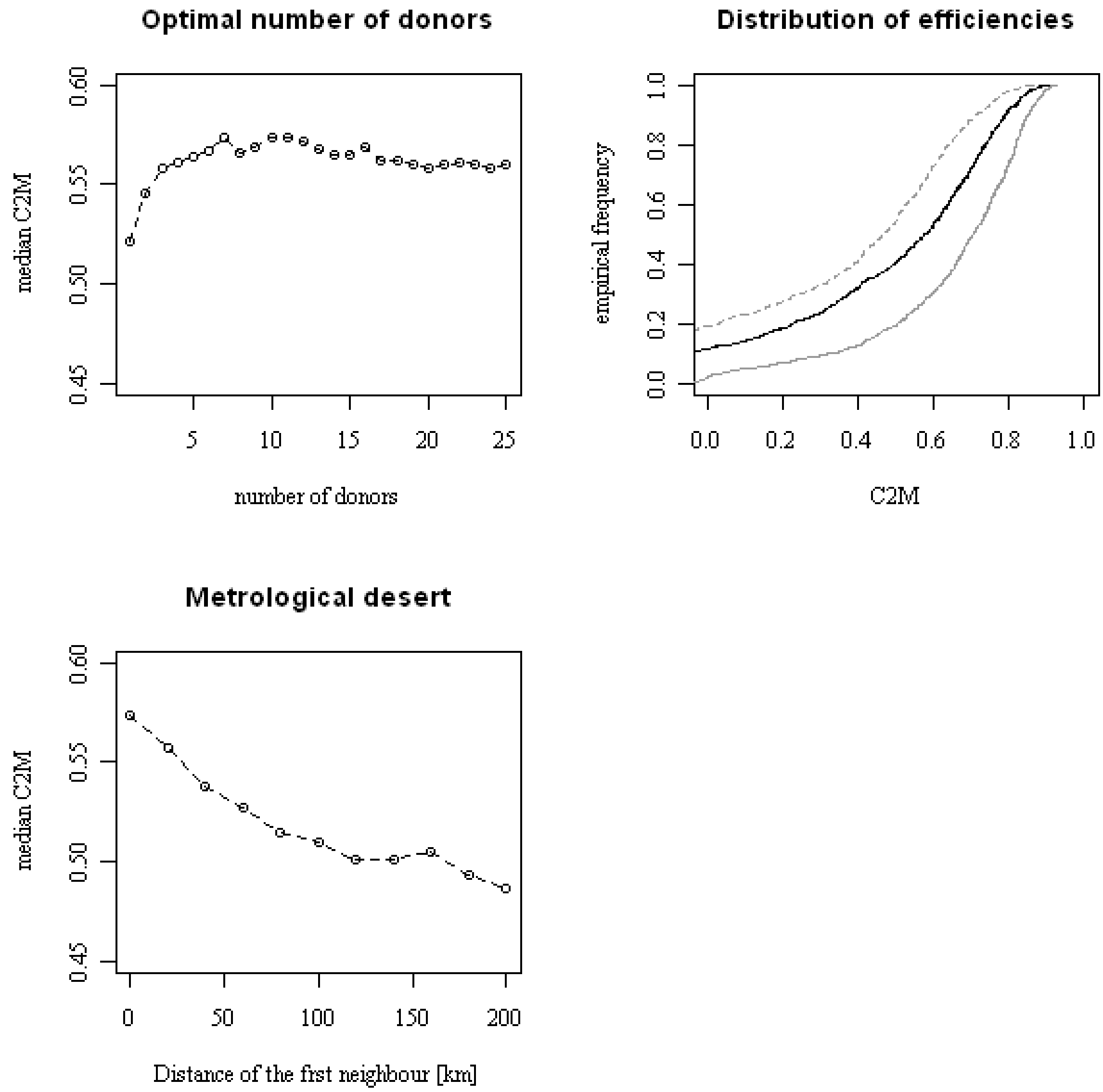


Figure 17: Backwards-sorting Regionalization performances. Top left, median efficiency per number of donor catchments used. Top right, distribution of efficiencies compared to a random selection of donors (dashed line) and calibrated model (solid grey line). Bottom left, performance in a "metrological desert" situation.

8 Joining spatial proximity and physiographic similarity

In this chapter, we present and evaluate two approaches methods allowing a joint use of spatial proximity and physiographic similarity:

- the *intersection-based method*, based on the assumption that good donor catchments are likely to be, at the same time, similar and geographically close to the ungauged catchment of interest. Thus, the best donors will belong to the intersection of the two ensembles ;
- the *union-based method*, based on the assumption that the two approaches may identify good donors independently. Thus donors will be best identified by the union of the two ensembles.

8.1 Introduction

In the previous sections we covered two regionalization methods based on pure physiographic similarity. As already said, our attention has been focused on such methods first because of their expected robustness.

Another reason why an hydrologist should appreciate physical similarity more than spatial proximity is that it is less "black box": it does not really provide any outlook of the hydrological processes that dominate the catchments of a certain region, but at least it gives a possibility for a careful, indirect, rough interpretation.

However, spatial proximity should not totally be dismissed. On one side, there are situations (for instance very dense gauging networks) where its performances might be superior to those of approaches relying on physiographic measures. On the other, as Figure 18 shows, it is to some extent complementary to physiographic similarity.

In Figure 18 we can see a grey dashed line representing the performances of a pure spatial proximity regionalization on our dataset (using four donors), a black dashed line representing the performance of the backwards-sorting physiographic similarity covered in section 7.3, and a black solid line. Such black line represent an ideal (non-existent) method that would, for each ungauged catchment, be able to decide whether in that particular case spatial proximity would give a more accurate guess than physiographic similarity, or vice-versa. Its performances are clearly superior to the other two methods used alone. Of course, the reader should be aware that this example was constructed by "cheating" and is only used to show the complementarity of the two original regionalization approaches.

While we do not expect that a realistic method could come close to the performances of the "ideal" case, we think that Figure 18 clearly shows the interest of investigating methods that combine some degree of physiographic similarity (as a way to ensure robustness and for its "informative" value) with some degree of spatial proximity (whose only value is an eventual increase in performance). The next paragraphs will cover two simple propositions of how such a method could be constructed.

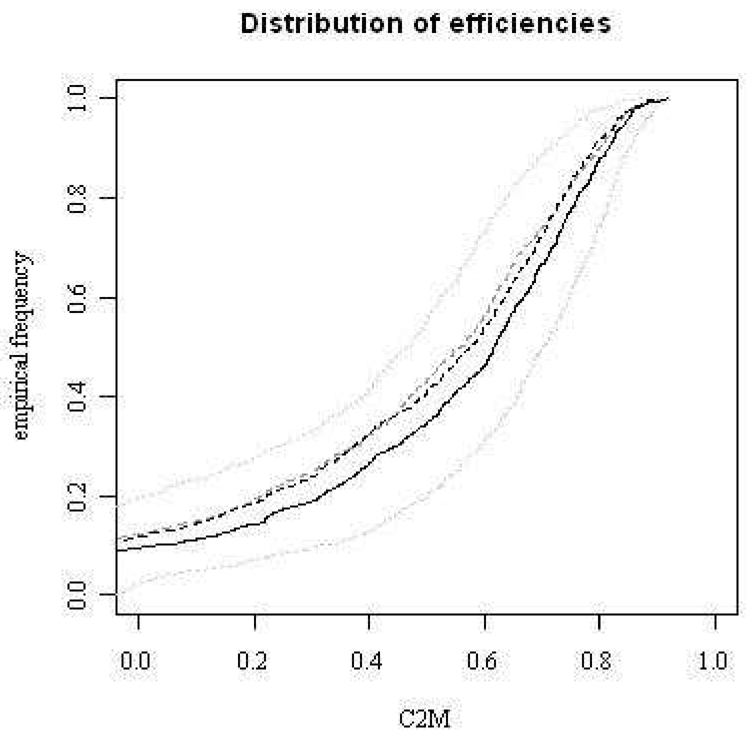


Figure 18: Performances of spatial proximity and physiographic similarity methods (dashed grey and dashed black lines) confronted with the performance of an ideal method perfectly combining the strenghts of the two approaches (solid black line)

8.2 An intersection-based method

8.2.1 description

The idea behind this method is that good donor catchments are likely to be, at the same time, similar and geographically close to the ungaged we are looking at.

To select donors that are close and similar, we proceeded as follows:

- The number of donors to be used was set. Let us say for the sake of this example that we wanted to use 10 catchments.
- Each time we considered a catchment as ungauged, the remaining ones were ranked in two lists of donors. The first was ranked according to geographical distance, the second was ranked for physiographic similarity (as in the backwards-sorting method shown in section 7.3
- We looked at the closest 10 catchments and at the most similar 10 catchments. If these two groups contained the same 10 stations, these would be the retained donors.
- In case we didn't have the same 10 catchments in the two groups, we would progressively increase the size of the two pools of candidates: for instance, the 11 closest one and the 11 most similar.

- We would look at the intersection of the two pooling groups (i.e. catchments appearing both in the group of the closest and in the group of the most similar ones). If 10 catchments were to be found in such an intersection, we would stop and retain these 10. If not, we would keep increasing the sizes of the two candidate groups until 10 candidates could be found.

8.2.2 Results

Figure 19 is a summary of the performance of the intersection regionalization method.

The optimal number of donors is six, with which a median C2M of 0.58 (equivalent to a NSE of 0.73) is obtained.

The performance gain, compared to pure physiographic similarity, is quite small (from 0.574 to 0.578). However, we actually notice a performance decrease in the "metrological desert" robustness test. While this result should be expected as an effect of bringing spatial proximity into the regionalization method, it is quite strong (pure similarity already works better when the closest catchment is more than 20 km away) and makes the proposed "intersection" method a poor choice.

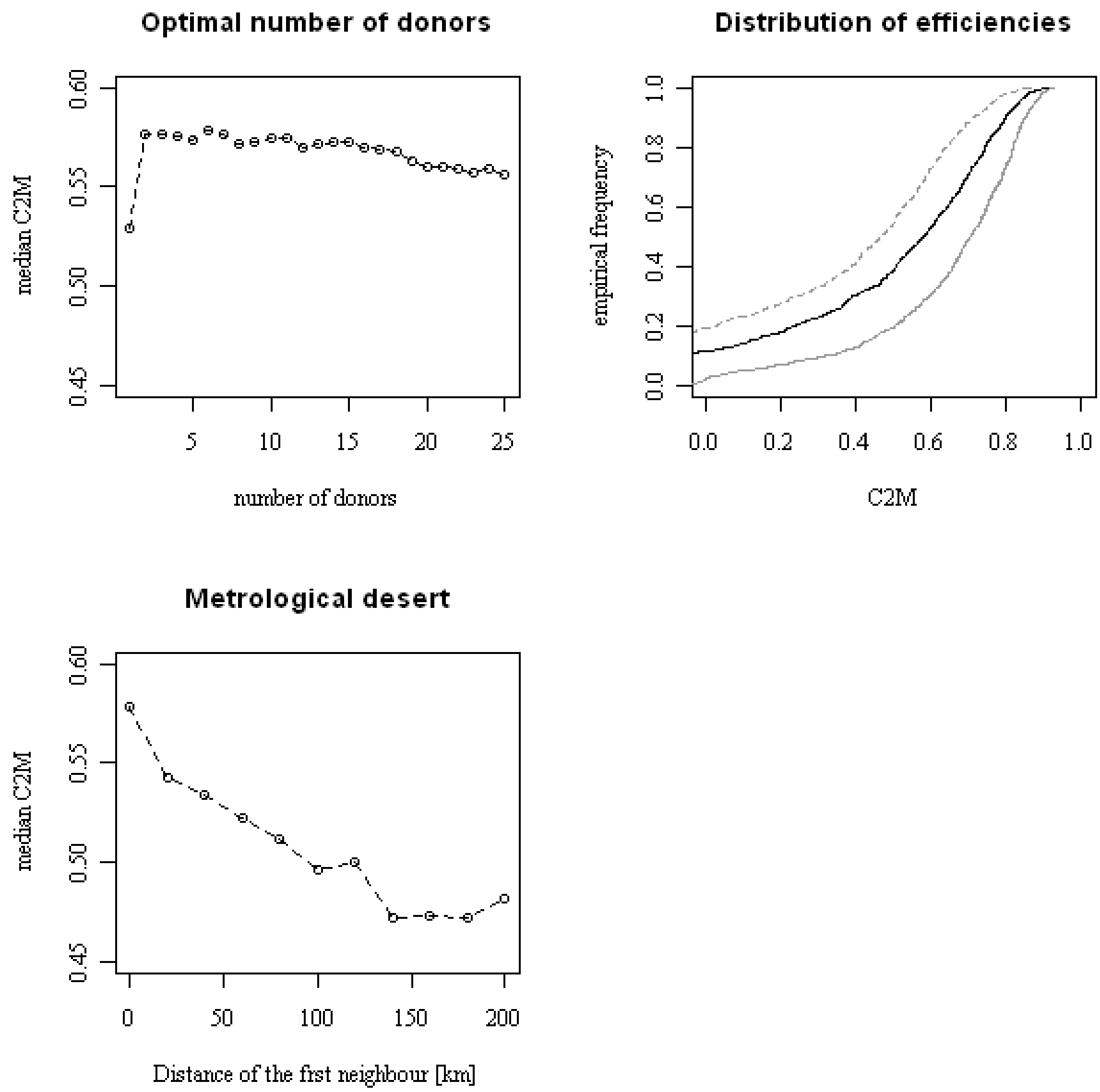


Figure 19: Combining spatial proximity and physical similarity, results of the intersection regionalization method. Top left, median efficiency per number of donor catchments used. Top right, distribution of efficiencies compared to a random selection of donors (dashed line) and calibrated model (solid grey line). Bottom left, performance in a "metrological desert" situation.

8.3 A union-based method

8.3.1 Description

This approach is based on the idea that –for our dataset- pure spatial proximity and pure physiographic similarity will identify a certain number of "good" donor catchments when used alone.

Pure spatial proximity works best (on our database) with four donor catchments, backwards-sorting physiographic similarity works best with 11 donors. We then simply pasted the two donor lists, obtaining a group of 15 donors. Notice that, when a catchment is both in the 4 closest and in the 11 most similars, it is counted twice.

8.3.2 Results

Figure 20 shows two charts about the performance of the union-based regionalization method: the distribution of the efficiencies obtained on the catchments we tested as ungauged, and the median performance in the "metrological desert" robustness test. In comparison to the previously treated proposals, we did not test different numbers of donors.

The median performance obtained is 0.58 in C2M, or an NSE of 0.74. This result is only marginally better than the intersection-based regionalization: however, the robustness of this approach seems to be much more satisfying. Pure physiographic similarity would only have a clear advantage on catchments who don't have any donor closer than 180 km, while when at least one donor closer than 100 km is available, the union-based method is superior.

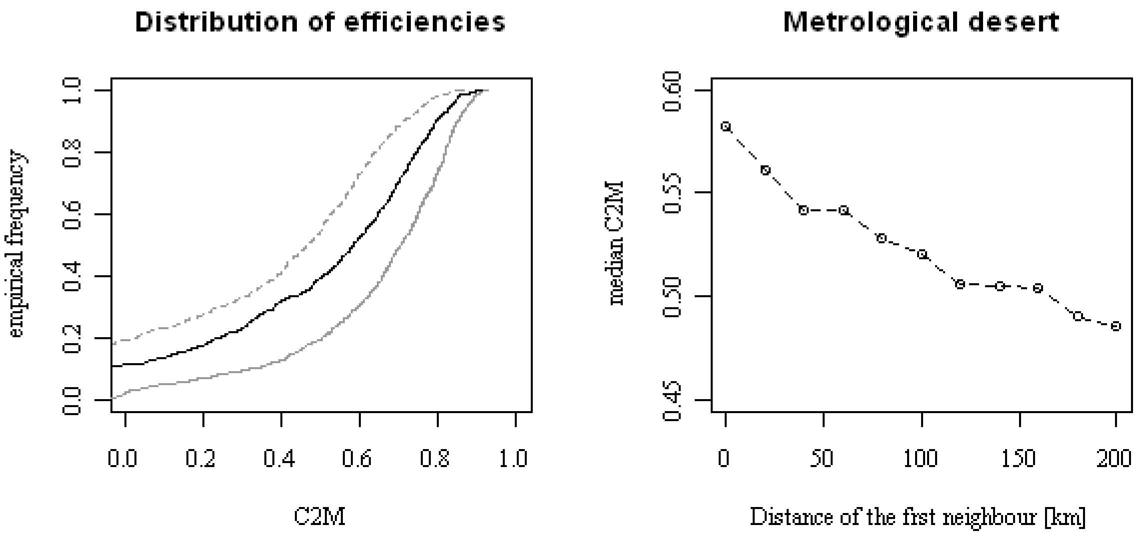


Figure 20: Combining spatial proximity and physical similarity, results of the union regionalization method. Left, distribution of efficiencies compared to a random selection of donors (dashed line) and calibrated model (solid grey line). Right, performance in a "metrological desert" situation.

8.4 Comparison of the tested regionalization approaches

Figure 21 and Figure 22 allow a comparison of the performances of the tested direct regionalization approaches. All approaches perform very similarly for the better modelled catchments, with the most noticeable differences being concentrated between empirical frequencies 0.1 and 0.4.

Overall the "union" combination of spatial proximity and physical similarity performs best, even if it is not very far from the other three tested methods, and constitutes a marginal improvement over a backwards-sorting based similarity approach, despite a theoretically much bigger margin for improvement.

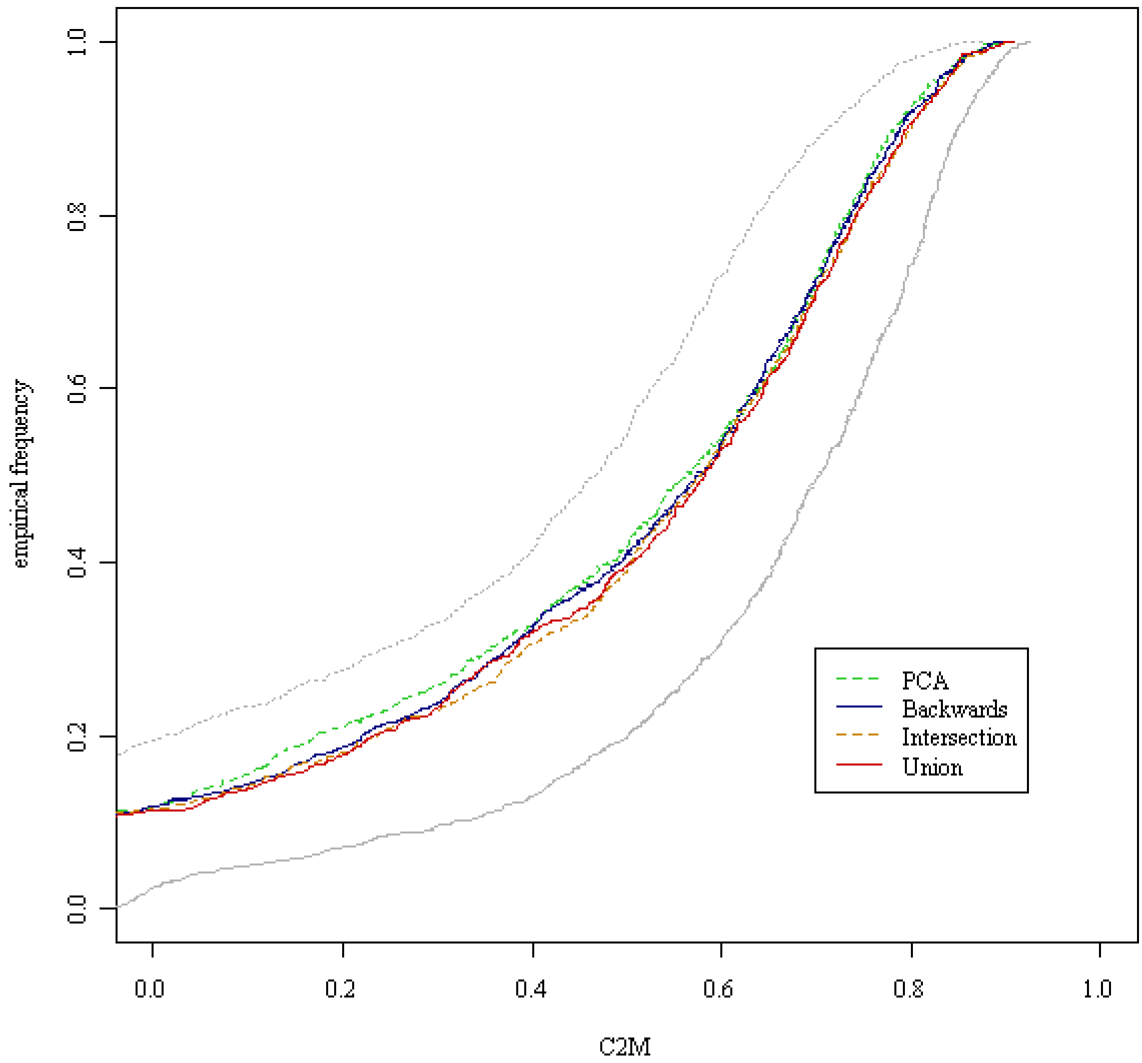


Figure 21: Distribution of the performances of the tested direct regionalizations, compared to two benchmarks: random donor selection (dotted grey line), calibrated model (solid grey line)

All methods appear to have similar robustness, with the possible exception of the "intersection" one (which probably relies too much on spatial proximity). In all cases, a noticeable improvement over spatial proximity can be noticed.

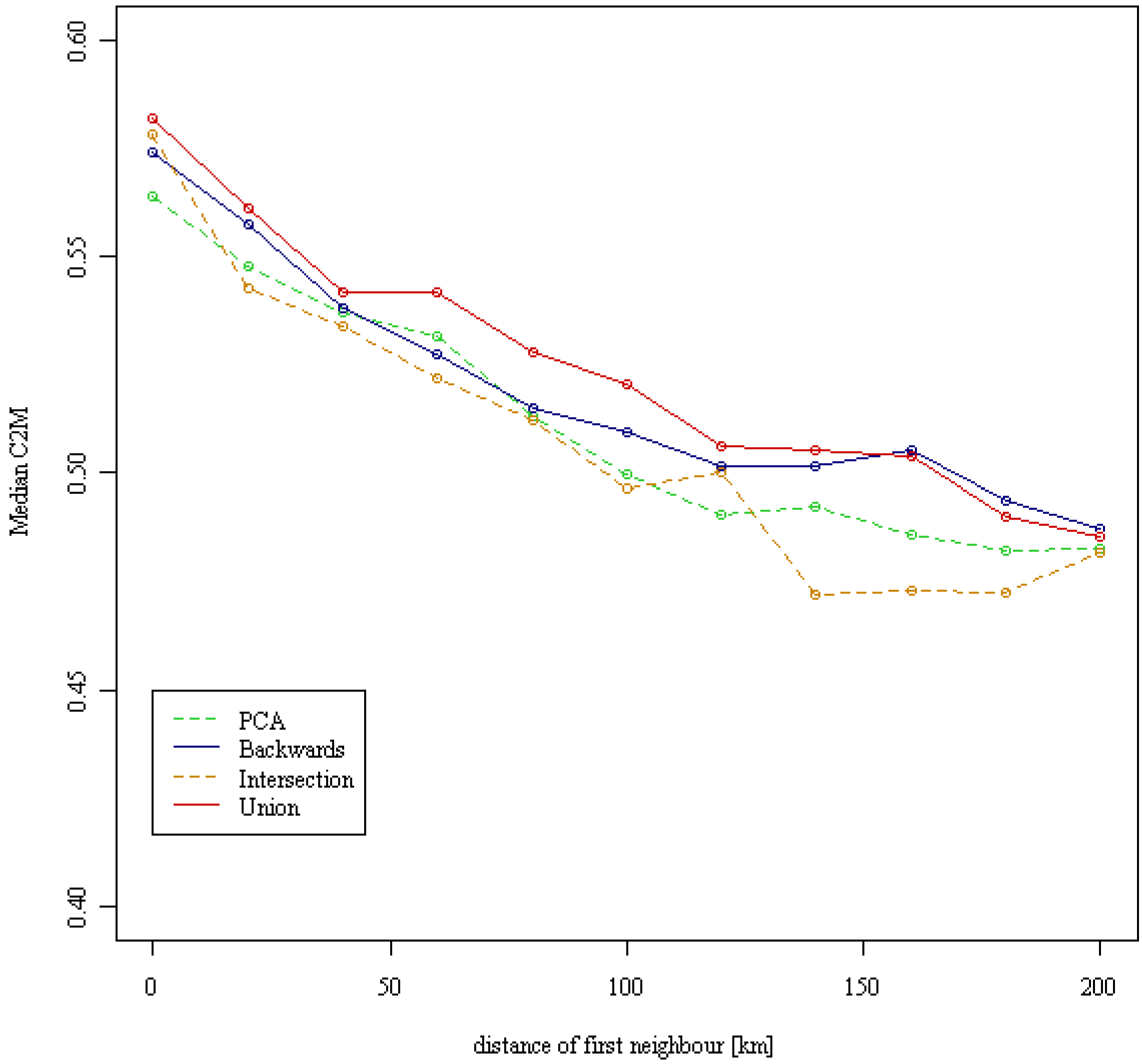


Figure 22: Comparison of the performances of the tested direct regionalizations under the "metrological desert" robustness test

9 Sensitivity analysis of regionalization methods: how do they react to the lack of similar catchments?

In this chapter, we will present three assessments of the robustness of the proposed regionalization methods, based on a simple but requiring test called the “metrological desert”. This test is based on the elimination from the donor list of those catchments which are geographically closest or most similar to the receiver catchment.

9.1 Introduction

In this chapter we will resume the results of the “metrological desert” test introduced in section 3.3, and propose its extension to the elimination of physiographically similar catchments.

9.1.1 Results of the elimination neighboring donors

Figure 23 shows the sensitivity of the four regionalization approaches presented in chapters 7 and 8 to the elimination of the closest donors. For each approach, three lines show how the values of the 0.9 quantile, median, and 0.1 quantile of the performance distribution decrease.

Interestingly, the differences in regionalization robustness seem to be much greater on the worse-modeled catchments than on the rest of the distribution: on the 0.1 quantile, it is clear that the “Intersection” approach (which has a strongest element of spatial proximity among the tested alternatives) is by far the less robust, while the remaining three methods have similar performances (with the backwards-sorting method performing less badly). Looking at the median performances, the relative lack of robustness of the “Intersection” approach is confirmed, even if the differences between regionalization methods are much smaller, to the point that the remaining three approaches can be considered to be equivalent. Finally, the 0.9 quantile shows a slight disadvantage of the PCA-based similarity approach.

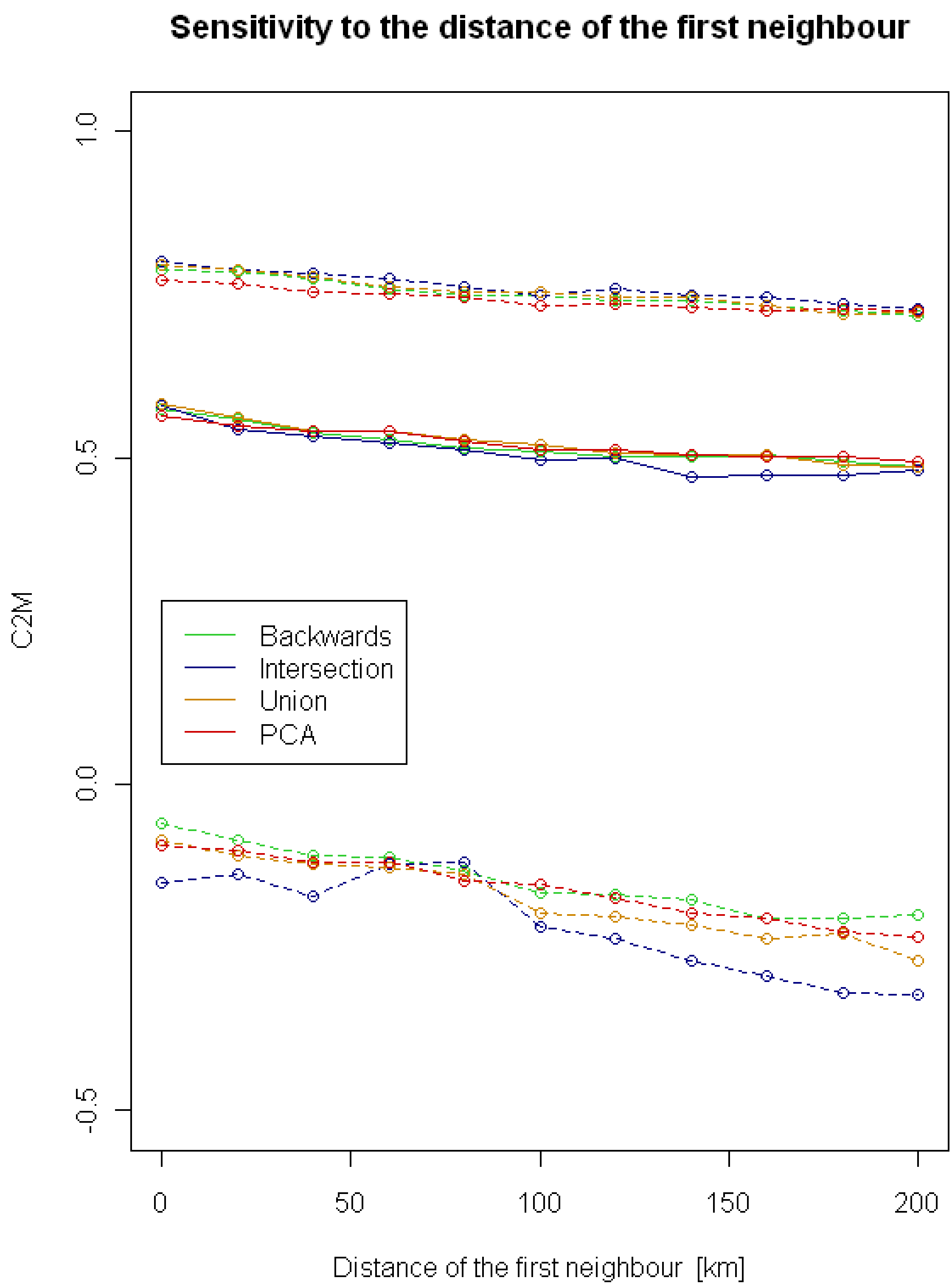


Figure 23: Performances of several regionalization approaches in a "metrological desert" situation. Upper dashed line: 0.9 quantile of the performance distribution. Continuous line: median. Lower dashed line: 0.1 quantile

Overall, we argue that the most robust approaches when facing a lack of close donors are the one based on backwards-sorting similarity and the “union” similarity/proximity hybrid, but it is not possible to detect a significant difference between the two.

9.2 Sensitivity of regionalization methods to the lack of similar catchments

In analogy with the "metrological desert" approach, we tested the sensitivity of regionalization methods to the removal from the dataset of donors whose characteristics were closest to the ones of the receiver catchment. This test was repeated individually for each physiographic descriptor, calculating the difference between the receiver's and the potential donors' values and eliminating the donors who fell below a certain threshold. Although this procedure is quite redundant, we think that excluding donors on the basis of a chosen similarity metric wouldn't yield "neutral" results (i.e. we expect that methods which use the closest similarity metrics would be the most affected).

The thresholds to be tested have been determined considering that most physiographic descriptors are roughly normally distributed. Consequently, we decided to set the maximum threshold to be tested for each physiographic characteristic to half of the descriptor's standard deviation, as this would eliminate nearly 40% of the donors for an average case, and we thought that it was not necessary to test an even harder constraint.

9.2.1 Results

For ease of reading, results will be presented in graphic form in appendix 14, while this paragraph will provide an overall review.

The overall sensitivity to the lack of similar donors seems to be comparable to what is observed when we excluded geographical neighbours (or even lower if one considers that in that case we were excluding about 20% of the donors instead of 40%). Other similarities can be found if one notices that the sensitivity of the median and lower quantiles of the performance distribution seems to be greater than for the upper quantiles, both in terms of average performance decrease and in terms of difference between one regionalization strategy and another.

Interesting trends can be observed if one considers the performance of the "intersection" method, which is the one containing the strongest compromise with spatial proximity, in comparison with the other approaches. This method seems to have comparable, or even slightly better results when donors are eliminated on the basis of climatic descriptors and of drainage density; on the other hand its performances are generally inferior when the sensitivity to altitude, slope, and some land cover classes quantiles is considered. In our opinion, this might indicate that altitude and slope and some kinds of land cover are more spatially correlated than other descriptors, and as a consequence those donors who are similar when considering these properties are more likely to be geographically close to the receiver catchment.

Catchment area constitute an exception to the above described scenario, as the regionalization's sensitivity to the lack of donors sharing similar characteristics seemed to be higher for these descriptors than for the rest. Catchment area appeared to be the most significant among the available descriptors when constructing hydrological similarity measures, so the high sensitivity to the removal of donors having a similar size to the receiver is no surprise. Furthermore, its distribution over our dataset is log-normal, which means that on average, more than 40% of the available donors are eliminated when using the 0.5 σ threshold. A similar behaviour is observed when excluding donors which share a similar drainage density with the receiver: in this case, one could invoke again the strong spatial organization of drainage density, since DD did not appear as a very significant descriptor when constructing hydrological similarity measures.

9.3 Sensitivity of regionalization methods to thresholds of model efficiency

In this section we wish to examine the reaction of regionalization methods to the lack of well-modelled donors: how important are they to obtain good regionalization results? Similarly to the procedures applied in the rest of this chapter, we will exclude from the donor list catchments whose calibration efficiency exceeds a certain threshold, which is moved lower and lower between C2M=1 (perfect simulation) and 0.5 (poor, but not catastrophic simulation).

9.3.1 Results

As it can be seen in Figure 24, the negative impact of the lack of well-modelled donors is equally extreme for all regionalization methods tested, to the point that in our opinion discussing the relative merits of each of them in such a context does not make sense.

For all methods, the decrease in performance seem to be concentrated on the "medianly" and worse regionalized catchments, especially when donors having calibrated a performance between C2M=0.9 and C2M=0.65 are excluded. These boundaries contain roughly 60% of the available donors, as only 1.5% of the catchments in our dataset yield a calibrated efficiency higher than 0.9, and 38% have a calibrated efficiency lower than 0.65. But in comparison, randomly removing 60% of our donors would have a much milder impact, as suggested in Figure 26, which leads us to conclude that the presence of a majority of well-modeled catchments in the donor list is required to obtain satisfying regionalization results for most ungauged catchments. The impact on the upper quantiles of the regionalized performance distribution is still very strong but less extreme: however, our hypothesis is that these few “lucky” cases can’t be expected to yield acceptable results for the right reasons.

The observations made in this test can be corroborated with the results of a similar experience, illustrated in Figure 25. In this case, the worse modeled catchments, up to an efficiency threshold, are excluded from the regionalization of the GR4J. This operation leads to a mild decrease of the median performances for all regionalization methods tested, until a threshold corresponding to C2M=0.5 is reached: then, the regionalization performances decrease steeply as some of the better-modeled donors are excluded. These results are consistent with a similar test exposed by Oudin et al. (2008), with the difference that in that case the exclusion of badly modeled donors initially led to a mild increase of the median regionalization performances, and that the steep performance decrease occurred when a threshold of C2M=0.67 is reached.

The upper and lower quantiles of the regionalization’s performance distribution also confirm the general trend of performance decrease when badly modeled donors are excluded, although with some differences: the upper quantiles do not seem affected until donors having an efficiency greater than C2M=0.8 are excluded, while the lower quantiles are equally affected throughout the whole test.

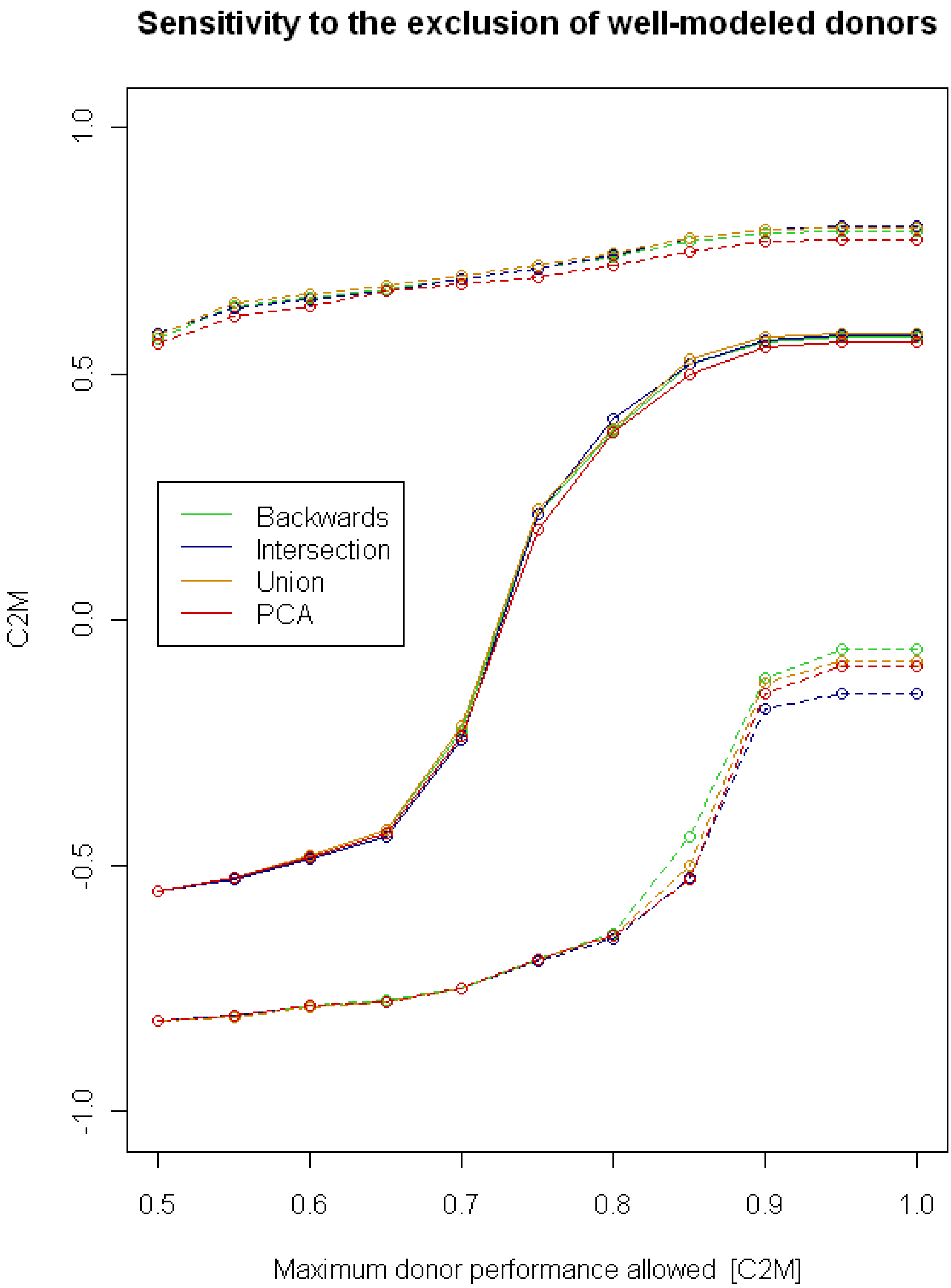


Figure 24: Sensitivity of several regionalization approaches to the lack of well-modeled donors. Upper dashed line: 0.9 quantile of the performance distribution. Continuous line: median. Lower dashed line: 0.1 quantile

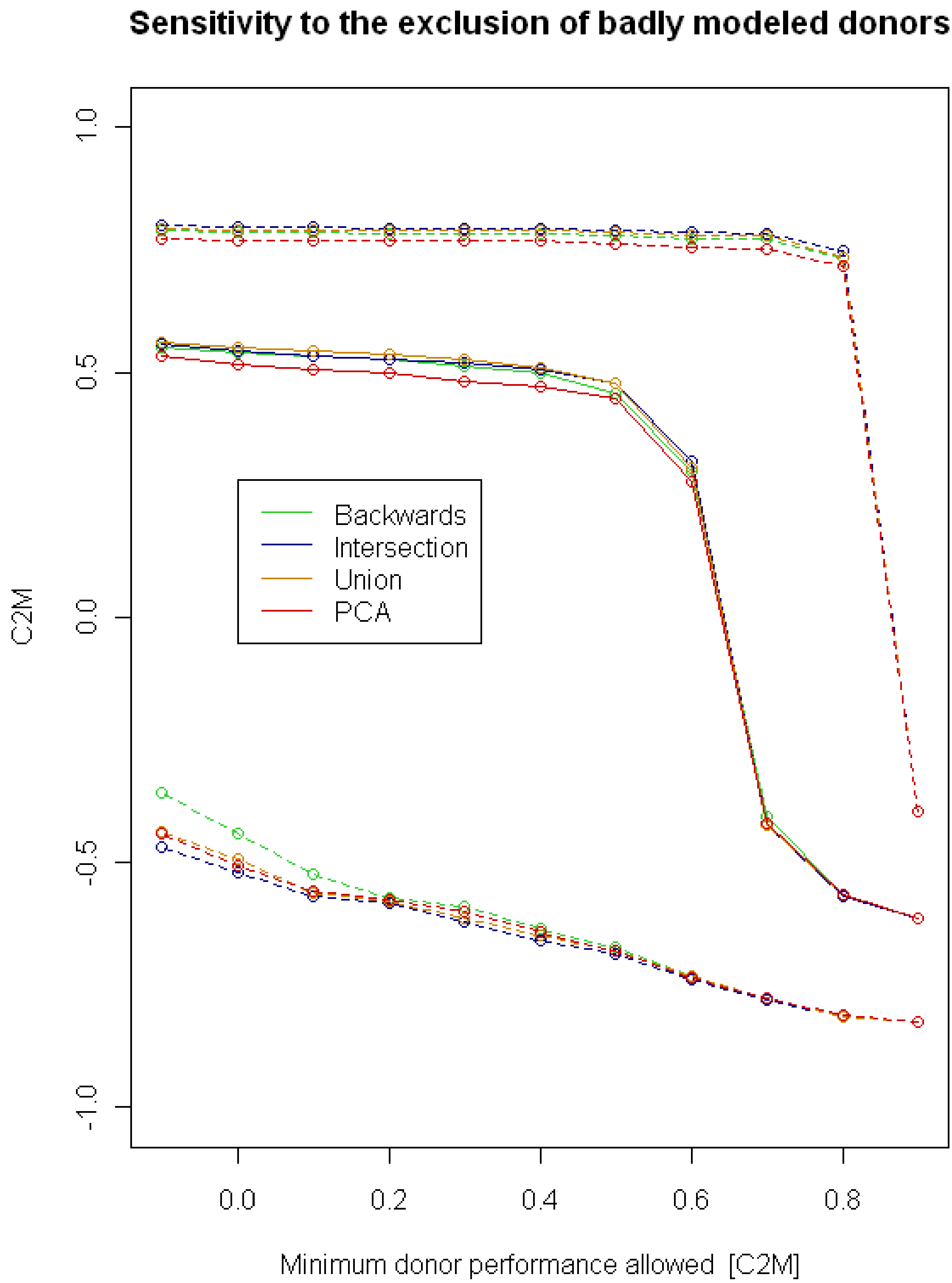


Figure 25: Sensitivity of several regionalization approaches to the exclusion of badly modeled donors. Upper dashed line: 0.9 quantile of the performance distribution. Continuous line: median. Lower dashed line: 0.1 quantile

Part 4 – Regionalization of rainfall-runoff models – the indirect path

In this part, we present the results of what we have chosen to call the indirect path to regionalizing a hydrological model. We see the problem of model parameterization on an ungaged basin as a problem of choosing one or several parameter sets in a library. The originality of the method presented here lies in the fact that we rely on a previously implemented regionalization of statistical flow values (i.e. flow quantiles) to constrain the choice of the parameters from the library: only those parameter sets allowing to best reproduce the regionalized flow quantiles will be retained for rainfall-runoff simulation.

10 Direct and indirect regionalization

In this chapter, we first justify the reasons why we thought an indirect approach could be advantageous, in relation with the existing literature on the subject. We then identify the principal issues that we'd like to explore concerning the subject of indirect regionalization, give the details of the regionalization procedure we used, and comment the obtained results.

10.1 Introduction

When estimating model parameters at an ungauged location, several approaches are possible. The two most common approaches are:

- Fitting a regression between model parameters (calibrated at gauged catchments) and catchment physiographic descriptors;
- Transferring (or interpolating) model parameters from *proxy* catchments, i.e. catchments that can be close in the geographic space (spatial proximity) or in a geographic-physiographic space (physical similarity).

These two methods have one point in common: they try to regionalize the model parameters in a single step. We could call them *direct* regionalization methods.

On the other hand, some authors recently advocated an *indirect* regionalization method. This method consists in first regionalizing flow statistics that synthetically describe the hydrologic response of the ungauged catchment. In a second time, parameter sets are chosen according to their ability to reproduce the behaviour outlined by the regionalized statistics. Throughout this report, we will refer to such a method as an *indirect* regionalization.

10.1.1 Why could an indirect regionalization be advantageous?

Direct regionalization methods face two kinds of difficulties:

- calibrated model parameters usually show little or no correlation with physiographic descriptors: as a consequence, regressions usually yield poor results;
- approaches based on parameter transfer from proxy catchments perform better than regressions, but are less robust than them in data-sparse situations. Their performance is more affected by the presence (or absence) of "good donors" in the dataset, for the ungauged catchment considered.

In comparison, the regionalization of flow statistics seems to be an easier task. As a consequence, the idea lying behind indirect regionalization is that if one could successfully identify efficient parameter sets with the use of flow statistics, such a method could be used later to regionalize model parameters.

The overall scope of this chapter is to test this assumption, i.e. to outline if and under which conditions an indirect regionalization approach could be a better choice than direct ones.

10.2 Review of the relevant scientific literature

10.2.1 How does the work presented in this chapter relate to the existing literature?

In section 4.6, we have mentioned a few studies proposing indirect regionalization methods. All provide similar frameworks for selecting model parameters based on their ability to reproduce some flow response statistics, and generally agree on the reasons for which such a procedure could be preferable over a direct transfer of model parameters.

Another common point is that all of the presented methods select a range of feasible parameter sets rather than one set: however, this choice somehow limits the possibility of a comparison with traditional direct regionalization methods. Also concerning the assessment of indirect regionalization's performance, it is not clear if such methods can perform well when evaluated with traditional criteria (as the Nash and Sutcliffe efficiency) or if their interest is the consequence of a change of paradigm in model evaluation. In this regard, we feel that the assessment of indirect regionalization's performances in a "traditional" framework (one streamflow simulation, a single-objective evaluation criterion based on the comparison of simulated and observed hydrographs) is required as a preliminary.

Another question is raised by a slight disagreement in the use of indirect regionalization in the four papers: while Yadav et al. (2007) and Westerberg et al.(2010) approach it as a form of calibration, and do not couple it with other information, Castiglioni et al. (2010) and Bardossy (2007) explicitly or implicitly advocate its use in combination with other methods to constrain the feasible parameter space. **Do indirect regionalization methods perform acceptably on their own, or should they be rather considered as an additional criterion in a regionalization method combining different approaches?**

10.3 Issues of concern for implementing an indirect regionalization scheme

Here, we list the main issues which need to be addressed when planning an indirect regionalization scheme. The questions listed here are further dealt with in section 10.5.

10.3.1 How does the first level of regionalization affect the second?

Obviously, since the indirect regionalization comprises two steps, the accuracy of the first step will have an impact on the efficiency of the overall scheme.

We believe that, when evaluating the performances of an indirect regionalization scheme, this point should not be overlooked: we should not do as if we were always capable of estimating flow statistics with small errors, even if this may indeed be the case for some catchments. In this chapter, we will evaluate the impact of two different flow statistics' regionalization approaches on the overall performance of a model at ungauged locations. A comparison with the ideal case in which flow statistics could be estimated without errors will also be provided.

10.3.2 How to constrain the initial choice of possible parameter sets?

Looking at the literature, one can see a remarkable difference in how an indirect regionalization is used to identify candidate parameter sets:

- In the case of Yadav et al. (2007), candidate parameter sets are picked from a broad range of possible values, whose limits are set a priori and should reflect the expected range of variation of each model parameter over the whole study area.
- On the contrary, Bardossy (2007) starts from a much narrower choice of possible parameter sets: the criterion used is that candidate sets should perform acceptably on a selected "donor" catchment, which is supposed to be hydrologically similar to the ungauged catchment of interest.

In this chapter we will, as a first step, test the proposed indirect regionalization using all the optimal parameter sets of the catchments considered as gauged. We think this is the most challenging situation for such a method, practically equivalent (given the number of catchments in our database) to a case where parameter sets would be generated from an a priori distribution of "likely" values. At the same time, it is the only test we can think of that would address the performances of such indirect regionalization independently of the criterion used to further constraint the parameter choice.

In a second step, we will try to constrain the choice of candidate parameter sets with an additional criterion based on spatial proximity, similarly to what was done in sections 6 and 8, as an example of how the combination of different regionalization approaches can improve performance.

10.3.3 Can such a method be robust?

As a final point we would like to address the robustness of the indirect regionalization method, i.e. how its performance is affected by the quality and quantity of available data.

More specifically, since we will work on data from a very spatially-dense gauging network, our attention will be focused on what happens when the density of donors is reduced

10.4 Method

10.4.1 General choices

- **Context**

We evaluate the performance of an indirect regionalization method when applied in the following context:

– A lumped four-parameter model (GR4J) is used;

– We assume that an ungauged catchment is one for which we do not have streamflow measurements However, physiographic descriptors as well as precipitation input time series are available;

– The objective of the regionalization is to produce one streamflow time series;

– The efficiency of each simulation is evaluated as C2M on square-rooted flows;

– Flow statistics are regionalized using two simple methods: a regression between statistics and catchment descriptors, fitted on all available catchments, and a regression whose residuals are interpolated with inverse distance weighting (IDW), i.e. the approaches developed in chapters 5 and 6 are used.

- **Flow statistics considered**

For all of the available catchments, we calculated the following flow statistics, on records concerning years between 1986 and 2005:

– Average yearly runoff;

– Percentiles of the flow duration curve:

we considered eleven quantiles of the FDC, and will refer to these values according to the percentage of exceedance (Q_{10}, for instance, is the value that is exceeded 10% of the time). With this nomenclature, we have: $Q_5, Q_{10}, Q_{20}, \ldots Q_{90}, Q_{95}$;

– Lag:

The lag time of the catchment, estimated as the time shift for which rainfall and runoff records show the highest correlation. For instance, if the runoff record appears to be mostly correlated with the rainfall of two days before, we will have a two days lag.

- **Parameter sets: initial choice and evaluation**

In this chapter, we initially considered a broad library of possible parameter sets for each catchment treated as ungauged. This library contains the parameter sets that where calibrated on all the remaining stations in the dataset.

Each parameter set has been evaluated according to the following scheme:

- A simulation is run with the parameter set under evaluation and the rainfall record of the ungauged catchment considered;
- The flow statistics mentioned above are calculated on the obtained simulation: we will refer to them as $\tilde{S}$;
- $\tilde{S}$ are confronted with the regionalized estimation $\hat{S}$: for each statistic, the following error measure is calculated

$$err_i = \left| \frac{\tilde{S}_i - \hat{S}_i}{\sigma_i} \right| \qquad \text{Eq. 11}$$

  where σ_i is the standard deviation of the observed values (this is equivalent to working on normalized variables).

- A penalty score is calculated, as sum of all errors:

$$p = \sum err_i \qquad \text{Eq. 12}$$

- The available parameter sets are ranked according to their penalty score

At this point, one can pick the best n parameter sets, run a simulation for each of them with the ungauged's rainfall record, and average such simulations into a single time series. The number n is specific to the dataset and to the regionalization method uses: for this reason, we will determine its optimal value following a jack-knife procedure

A calibrated penalty score, where the penalty score would be a weighted average of the errors on each statistic, was also tested. The weights have been determined by a jack-knife calibration. Although this technique did slightly improve the regionalization performance, the increase is modest and apparently specific to the number of parameter sets one wants to use (e.g. weights calibrated to select one parameter set do not offer an advantage when one wants to select ten parameter sets). As we want to focus on the generalities of the indirect method, and particularly on the influence of the first-step regionalization on its outcome, we will not cover the calibration of a weighted penalty score in greater detail.

10.4.2 Criterion used to further constrain the choice of parameter sets

As outlined in section 10.3.2, after evaluating the ability of regionalized flow statistics to identify efficient parameter sets out of a broad range of possible values, we will test it on a narrower "library" of possible values, built using an additional constraint, based on spatial proximity.

This choice has been based on the fact that the gauging network we work with is particularly spatially-dense, and reflects a scheme already used in other parts of this work: using spatial proximity as a "last resource" to improve the performances of a regionalization method.

10.4.3 Three benchmark comparisons.

In order to help evaluating our results, we provide the comparison with two benchmark approaches:

- random selection: ten parameter sets are randomly selected from the library of candidate parameter sets. The resulting flow time-series are averaged. This benchmark indicates the lower limit of acceptable performance: any regionalization method should perform better than this (hopefully much better).
- "spatial proximity": we considered the optimal sets of the first three neighbors, and averaged the time series. This method is quite unsophisticated, lacks robustness, but performs very well when the distance between the ungauged catchment of interest and the next gauged basins is short enough, as in our case. We would like other regionalization methods to have similar (or better) performances on a full-density network and to show better robustness when the density is reduced.
- "ideal case": we will use the calibrated model's performance. For a large dataset such as the one we are working with, it is practically equivalent to a method that could choose the best possible parameter set among those calibrated on the gauged catchments. On a smaller dataset, there would be a more noticeable difference.

10.5 Discussion of results

In this section, we will have a look at how the proposed indirect regionalization method performed. The results will be presented in the following order: we will address successively the number of donors to be retained, the impact of the initial regression's accuracy, the options for further constraining the choice of the parameter sets, and the dependency of the results on donor stations density.

10.5.1 Number of parameter sets to be retained

Figure 26 shows how the performance of the proposed indirect regionalization varies when selecting different numbers of donors. Three cases are shown, corresponding to two different first-step regionalizations of the flow statistics, and a "cheat" case where we supposed that one could regionalize flow statistics with no errors.

The most remarkable behavior showed by the three cases is the lack of a significant performance decrease when selecting many donors, although in the "cheat" case we do have a slight decrease if more than 30 donors are selected. This result is rather surprising, because we would expect that only those parameter sets which allow reproducing best the flow statistics should contribute to a good simulation, while after a while, the contribution of the most different parameter sets should degrade the performance of the regionalization approach.

During the rest of our discussion, we will show results obtained with 50 donors, for all methods (this is an arbitrary choice).

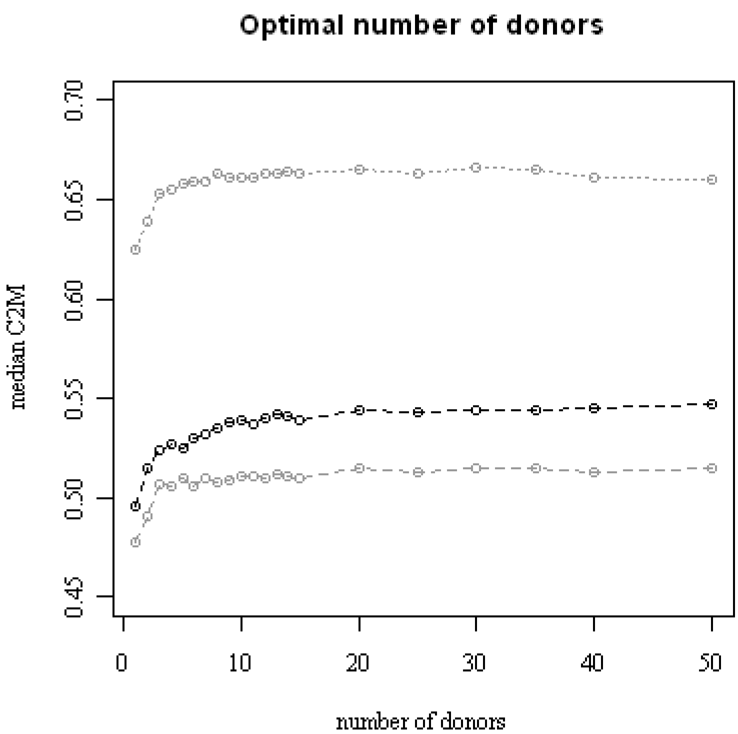


Figure 26: Optimal number of donors for an indirect regionalization scheme. Black line: statistics regionalized with a regression and an IDW interpolation of the residuals. Grey dashed line: statistics regionalized using a regression. Grey dotted line: "true" statistics (cheat)

10.5.2 *Impact of statistics' regionalization quality on the following regionalization of RR model parameter sets*

In order to assess the impact of the initial regionalization's accuracy on the overall results, we followed the indirect regionalization procedure described in section 10.4 with two kinds of flow statistics' regionalizations: a nation-wide regression and a regression coupled with IDW interpolation of the residuals.

Figure 27 shows the results obtained with flow statistics estimated with a plain regression approach (i.e. excluding regionalization of residuals). In this case, the indirect regionalization cannot reach the efficiency of the spatial proximity benchmark comparison. The obtained median efficiency is 0.51 if expressed in C2M, equivalent to a NSE of 0.68

This result is quite disappointing since it seems that the regionalization approach does not benefit from the information on flow statistics.

To assess whether this result is due to a lack of predictive efficiency of the approach used to regionalize flow statistics, we also tested the more refined approach, i.e. using IDW interpolation of the residuals.

The results are presented in Figure 28: although there is a visible improvement, the spatial proximity benchmark approach still performs better. The two methods have roughly the same performance for the better-modeled catchments, but the spatial proximity approach is still superior for the worse modeled ones.

The median efficiency obtained in this case is 0.55 expressed as C2M, corresponding to a NSE of 0.71

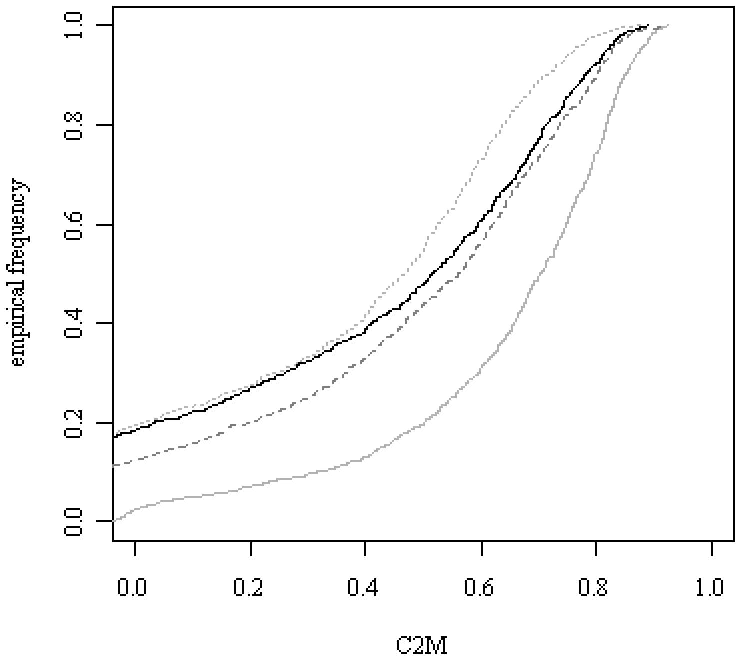


Figure 27: Performance of the proposed "plain regression approach". The black line presents the results obtained with flow statistics obtained through a "plain regression approach", confronted with three benchmark comparisons: random donor (dotted grey), spatial proximity (dashed grey) and calibrated model (solid grey)

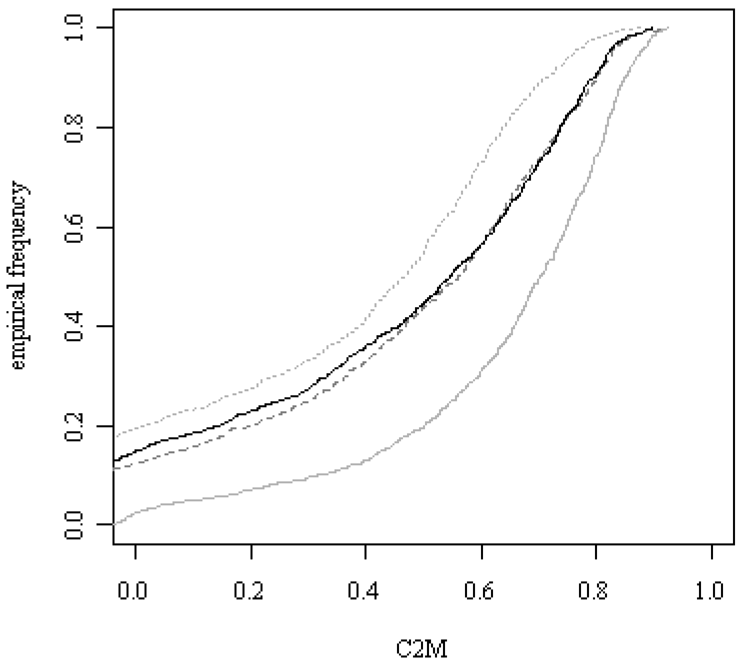


Figure 28: Performance of the proposed "regression + residuals interpolation". The black line presents the results obtained with flow statistics obtained through a regression approach combined with an IDW-based interpolation of residuals. It is confronted with three benchmark

comparisons: random donor (dotted grey), spatial proximity (dashed grey) and calibrated model (solid grey)

At this stage, the reasons of the failure of the indirect regionalization approaches are not clear since it may be attributable to either the fact that the “hydrological similarity” based on flow statistics does not match the “parametric similarity”, or the fact that the performance of the regionalization technique for flow statistics is too low, which does not allow to identify truly “hydrologically” similar catchments.

With the experiment presented in Figure 29, our aim is to estimate the maximum possible margin of improvement, if only we could estimate perfectly the flow statistics at the ungauged location. Of course, this is impossible, this is why we consider this method as a "cheat". The purpose of such a comparison is to show what margin of improvement can be filled just by improving the first-step regionalization of flow statistics over the two techniques we used.

Obviously, it is clear that one could eventually do much better than pure spatial proximity, getting very close to the efficiency of the calibrated model (median C2M=0.66, NSE=0.80).
These results show that the level of predictability of flow statistics, though apparently satisfying, is too low to help the regionalization procedures. Consequently, the indirect regionalization approach might be improved if regionalization of flow statistics is.

However, one can also wonder at this stage if the efforts should rather be put on direct regionalization approaches: note that the performances of an indirect approach with “perfectly regionalized” statistics are still far from those of a calibration performed directly against the performance criterion, and also from those of a “cheating” approach that chooses (a posteriori) the most efficient parameter set among those calibrated for the database catchments, excluding the one that was calibrated for the one considered as receiver (see Figure 30). This last result confirms that similarity based on the reproduction of flow statistics and similarity based on model parameters do not match very well, as both hypothetical methods use information from the receiver’s flow record and choose among the same parameter sets.

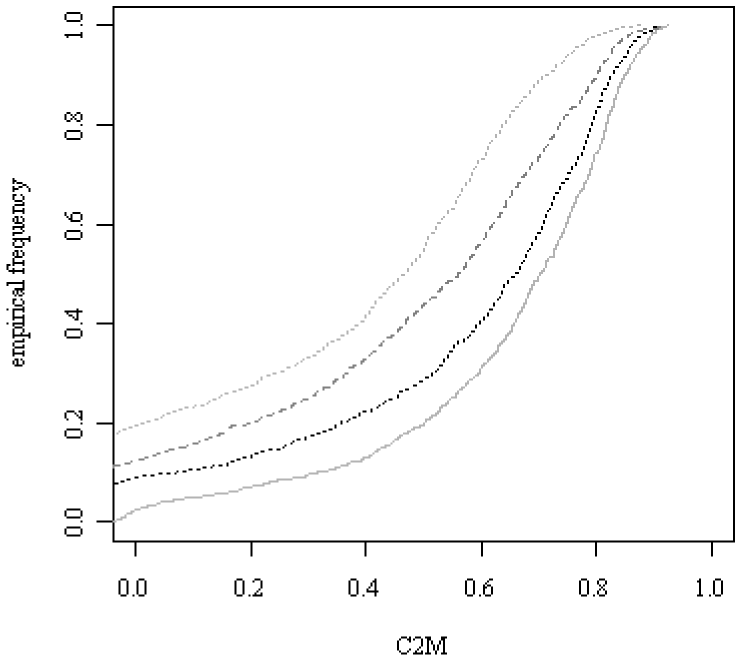


Figure 29: Performance of an ideal case where the flow statistics could be regionalized with no errors (black dashed line). It is confronted with three benchmark comparisons: random donor (dotted grey), spatial proximity (dashed grey) and calibrated model (solid grey)

Incidentally, one can also consider the reproduction of statistic flow signatures as an alternative calibration criterion, as done by Westerberg et al. (2010): in this case, further analysis would be needed to assess for which applications it produces desirable simulations, and for which others it does not.

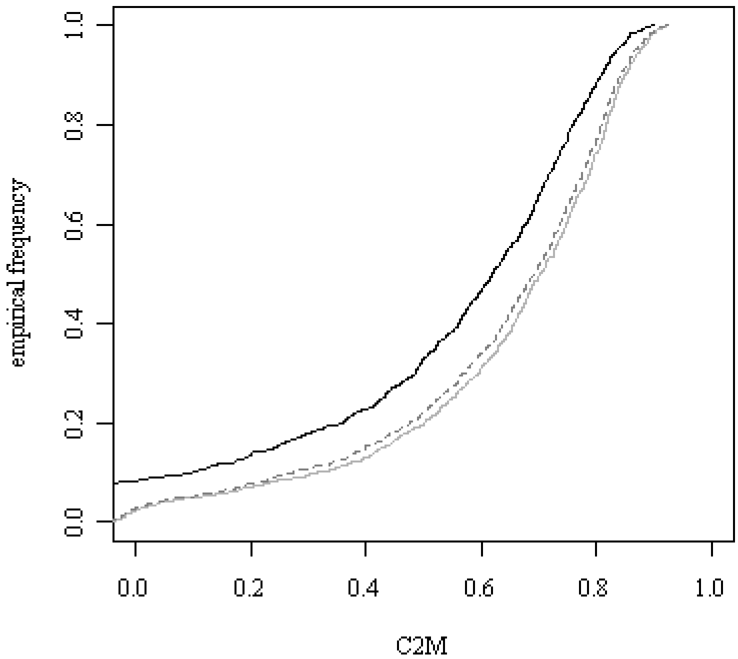


Figure 30: Performance of an ideal case where the flow statistics could be regionalized with no errors (black solid line). It is confronted with two benchmark comparisons: calibrated model (grey solid line) and a "cheating" method that selects a-posteriori the best possible donor among the catchments available in the database (grey dashed line).

10.5.3 Could it be advantageous to constrain the choice of parameter sets with an additional criterion?

One of the issues that this section aims to address is whether an indirect regionalization approach can give an acceptable performance when used alone, i.e. when making a selection out of a broad range of possible parameter sets, or whether it is necessary to combine it with other ways of constraining their choice. For this purpose, as explained in paragraph 10.4.2, we employed a constraint based on spatial proximity: we initially select a pool of n closest neighbours as donors, we evaluate them with the described indirect regionalization procedure, and finally consider only half of them (as an arbitrary choice). In the end, the best result seems to be obtained when considering the best 5 donors out of the closest 10.

Figure 5 shows the results obtained in this case: finally, the selection based on flow statistics offers a slight but quite consistent improvement over the closest neighbor (median C2M=0.59, NSE=0.74). It is interesting to notice that the improvement seems to be concentrated on "better-modeled" catchments. Note that similar results were observed when selecting 10 donors out of the 20 most physiographically similar catchments.

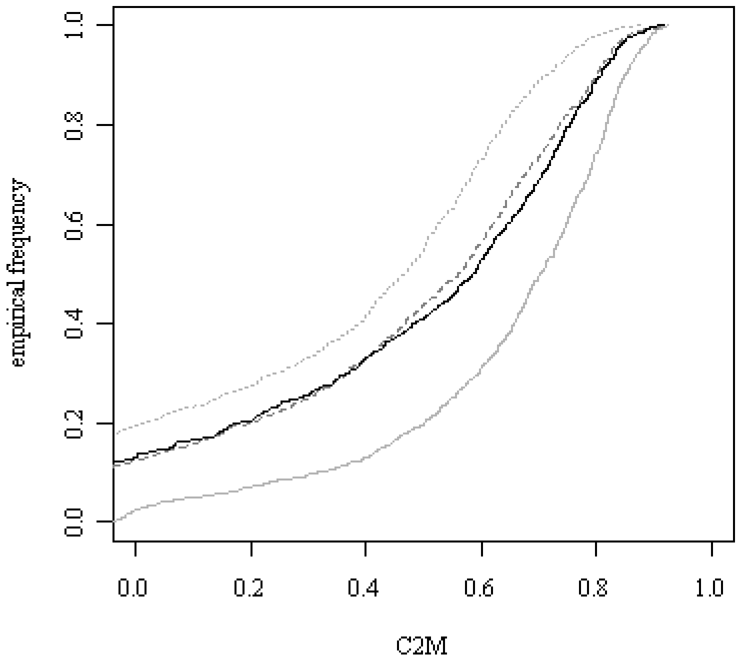


Figure 31 Performance of the proposed approach when selecting 5 parameter sets out of the the first 10 neighboring catchments (solid black line). It is confronted with three benchmark comparisons: random donor (dotted grey), spatial proximity (dashed grey) and calibrated model (solid grey)

10.5.4 Robustness of the method: application of the metrological desert test

As explained in paragraph 10.3.3, one of our objectives is addressing the performance of the presented indirect regionalization method in data-sparse situations. We will address this point using the "metrological desert" test presented in section 3.3 and in chapter 9

Figure 32 shows, with a black dashed line, the median efficiency obtained by an indirect regionalization using 50 neighbours and streamflow statistics estimated with a simple regression (for which we assume that the performance decrease in a "metrological desert" situation could be negligible). This is the same case already presented in Figure 27 and has been chosen for the metrological desert test since in a sparse-network situation we consider the IDW interpolation of the residuals to be unreliable.

As a comparison, the chart also shows the results of the backwards-sorting physiographic similarity approach discussed in section 7.3: this method has been chosen as an example of good robustness.

Indirect regionalization (or, at least, the selection of parameter sets) shows a remarkable robustness: the obtained performance is only dependent on the performance of the first-step

estimation of streamflow, and on spatially-sparse networks it is likely to perform as well as similarity-based methods or even better.
Of course, in cases where some stations could provide useful data for flow statistic estimation but not for model calibration, we could expect even better performances.

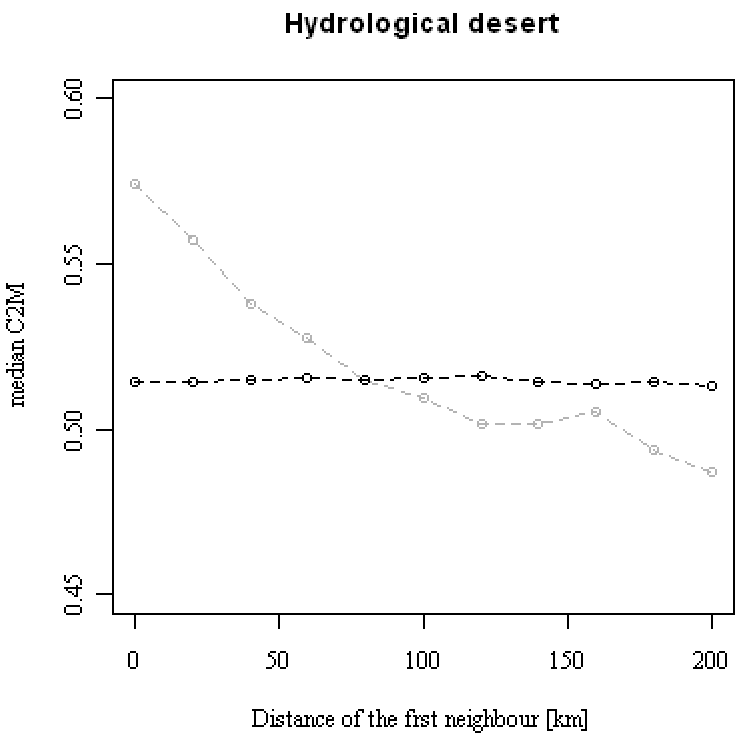


Figure 32: "Metrological desert" test. The median efficiency of an indirect regionalization method using regression-estimated flow statistics (black line) is confronted with the optimal physiographic-similarity method identified in Part 3 (grey line).

11 How the choice of an efficiency criterion impacts our vision of the 'best' regionalization method

In this chapter, we will test the regionalization methods seen in the previous chapters from the point of view of different efficiency criteria, in order to investigate whether their order of preference should be considered as specific to the chosen criterion. Specific “sub-criteria” will be used to better address the relative strengths and weaknesses of the two families of regionalization methods (traditional and indirect).

11.1 What is the best regionalization method when we adopt an FDC-based performance criterion?

In chapter 10 we have seen that indirect regionalization based on the reproduction regionalized FDC quantiles has worse performances than traditional methods based on site-similarity, at least when the performance criterion of choice is C2M (bounded NSE).

It is not clear whether the poor performance of such methods is mostly due to a poor regionalization of the FDC quantiles, or to the fact that the constraint imposed by the reproduction of the FDC is quite different from the one imposed by RMSE-based criteria. The purpose of this chapter is to clarify this point by means of a change in the performance criterion: an FDC-based constraint will be used to evaluate the regionalized models, instead of C2M. For each catchment in our database, a new parameter set will be calibrated, according to such constraint. These parameter sets will then be regionalized according to the same procedures outlined in chapters 7 and 8.

11.1.1 Performance criterion used for calibration

The first step of our evaluation is the calibration of new parameter sets under an FDC-based constraint.

The performance criterion used is almost identical to the "penalty score" presented in section 10.4.2, ie a sum of normalized errors on empirical FDC quantiles. The only difference one is that the "lag" statistic (time shift that maximizes the correlation between rainfall and runoff records) is not used in this case, as its discrete nature (its empirical values can only be integer: 0, 1, 2, etc.) leads to discontinuities in the optimization surfaces, which make calibration procedures either very time-consuming or inefficient.

11.1.2 Regionalization results

Figure 33 shows the results of direct and indirect regionalization methods when evaluated with an FDC-based criterion. For ease of reading we have chosen to only show two direct methods, as the performances of the remaining two were very similar. We have also included two benchmarks: the calibration performance (solid grey line) and the performance of a "worst case" regionalization using 10 randomly chosen parameter sets. Readers will note that the best performances correspond to the lowest values of the chosen criterion.

The performances of the best direct and the best indirect methods are comparable, with only slight differences at the two extremes of the distribution (an indirect method based on FDC quantiles regionalized with regression+IDW tends to be better than direct methods on worse-modeled catchments, while the opposite is true for better-modelled ones). An indirect method based on FDC quantiles regionalized using nation-wide regressions only has the worse performances in the group. This behaviour suggests that the indirect regionalization's inadequacy outlined in section 10.5 is mostly due to the difference between RMSE-based and FDC-based constraints.

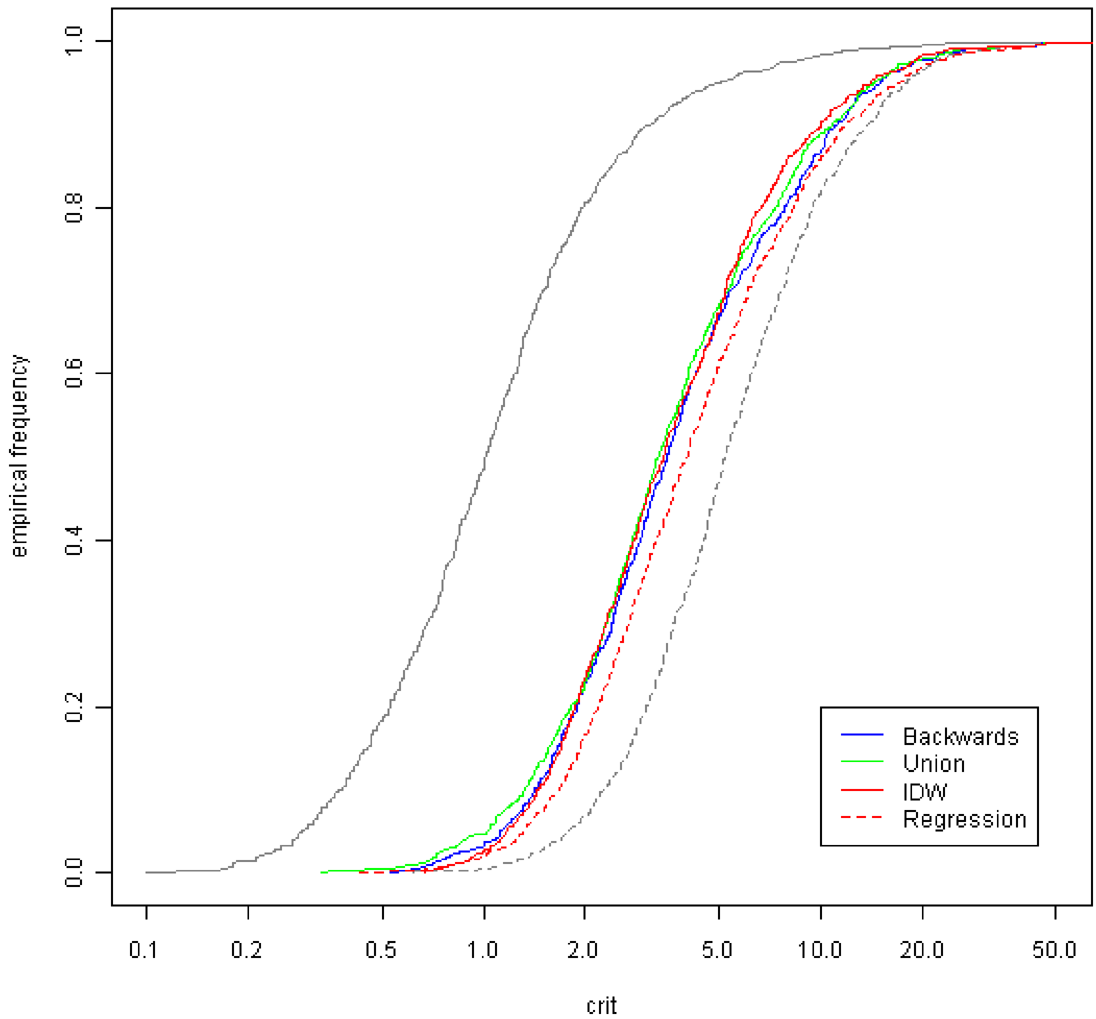


Figure 33: Performance distribution of regionalization results according to an FDC-based criterion. Continuous grey line: calibration performance. Dashed grey line: random regionalization.

11.2 Use of the Gupta et al. decomposition of NSE as diagnostic tool : where do lie the differences between C2M and the FDC-based criterion used in this chapter?

Gupta et al. (2009) proposed a decomposition of the NSE criterion based measures of correlation coefficient, normalized bias and relative variability (alpha) between the observed and the simulated runoff. This decomposition can be useful to understand in which ways the constraints imposed by C2M and by the proposed FDC-based criteria differ, and consequently to explain why the indirect regionalization method discussed in chapter 10 had disappointing results. We will first look at calibrated parameter sets in order to only focus on the two criteria, and then move our attention to regionalized ones.

11.2.1 Detail of the NSE decomposition used

Below we will briefly detail the three sub-components of the NSE used as diagnostic criteria:

- Correlation coefficient

$$\rho = \frac{E\left[(O_i - \mu_o)(S_i - \mu_s)\right]}{\sigma_o \sigma_s} \qquad \text{Eq. 13}$$

- Bias

$$\beta_n = (\mu_s - \mu_o)/\sigma_o \qquad \text{Eq. 14}$$

- Relative variability (alpha)

$$\alpha = \sigma_s / \sigma_o \qquad \text{Eq. 15}$$

Where *s* stands for simulated, *o* for observed.

11.2.2 Difference in calibrated parameter sets

Figure 34 compares the correlation coefficients resulting from parameter sets that have been calibrated with either C2M on square-rooted flows, or an FDC-based constraint. While in both cases most of the simulations have a correlation of at least 0.9, there is a significant difference between the two calibration criteria, with the FDC-based constraint yielding lower correlations.

Figure 35 shows the distribution of bias for the two criteria. The results obtained with C2M are good, as most of the simulations show very little bias, and overall, a very slight tendancy for overestimation. The “indirect” constraint shows a worse behaviour, even if th average of observed biases is closer to 0.

Figure 36 shows the relative variability observed in the two cases. While we think that both criteria are unsatisfying in this regard, the FDC-based constraint produces again the worst results of the two, with a more pronounced tendancy for over-variability.

Overall we can say that when evaluated with the proposed sub-criteria, an FDC-based constraint seems to have worse performances than C2M, in particular regarding the correlation and bias of the simulations vs the observations.

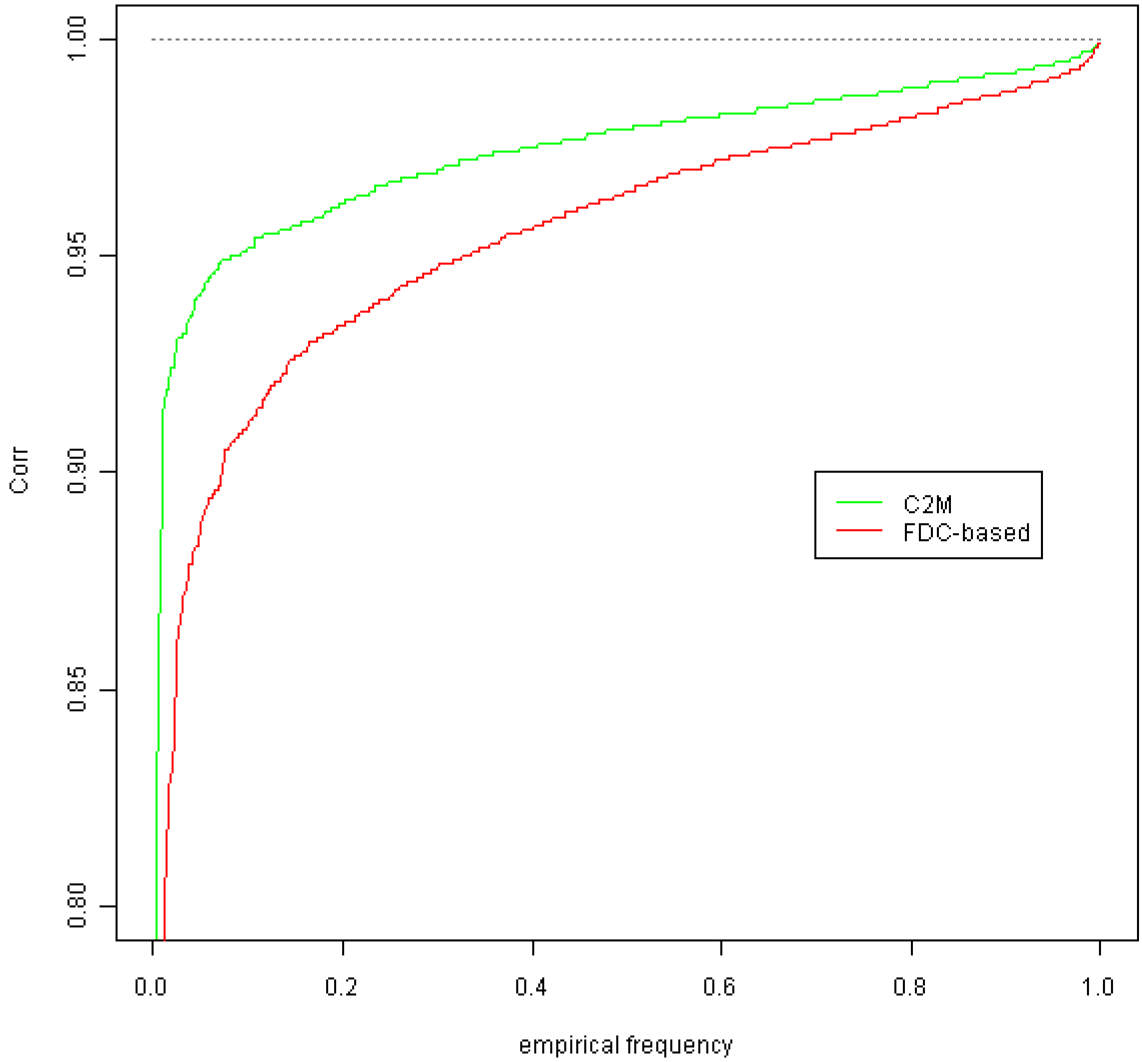


Figure 34: distribution of correlations for C2M-calibrated and FDC-calibrated parameters

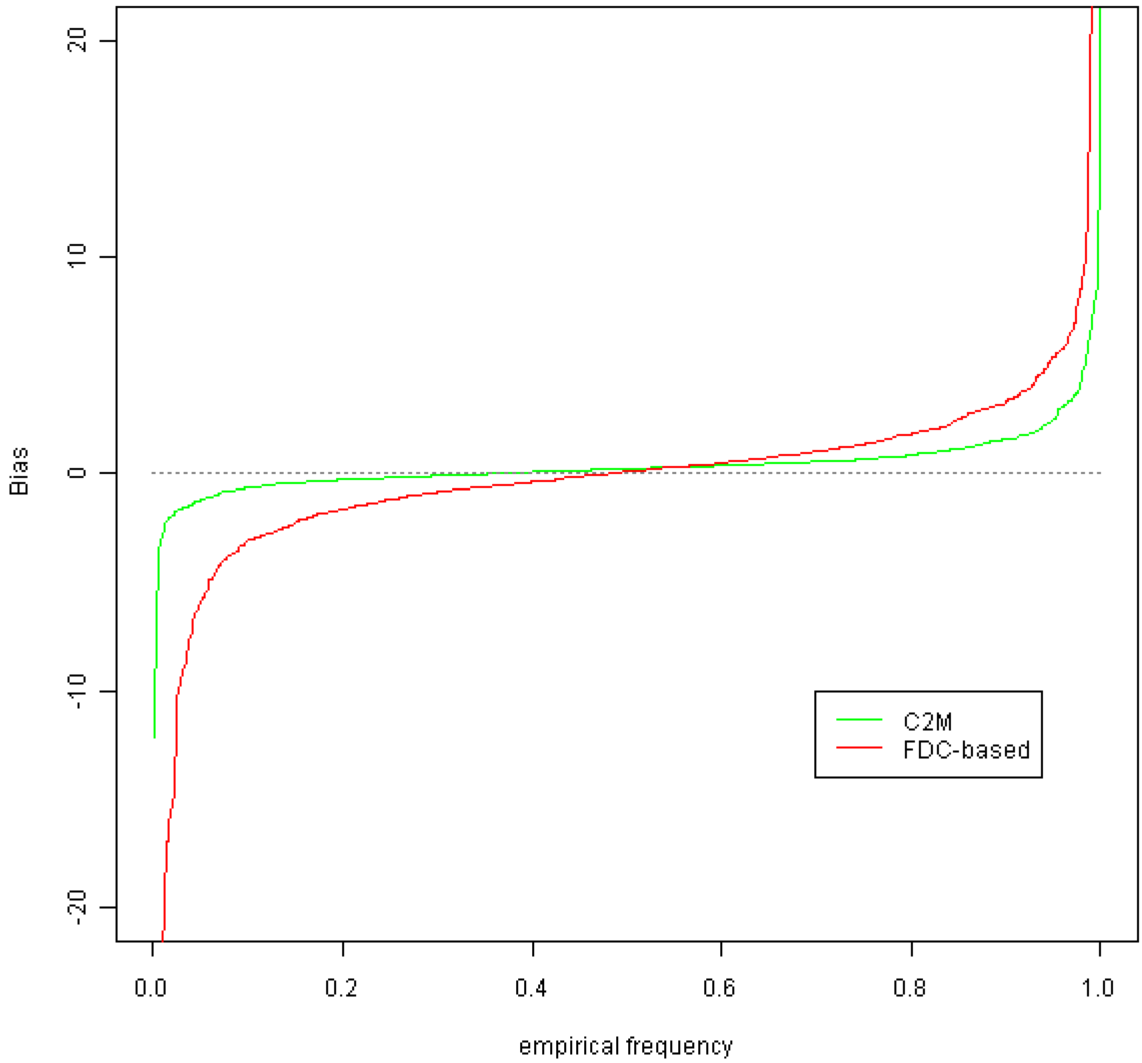


Figure 35: distribution of bias for C2M-calibrated and FDC-calibrated parameters

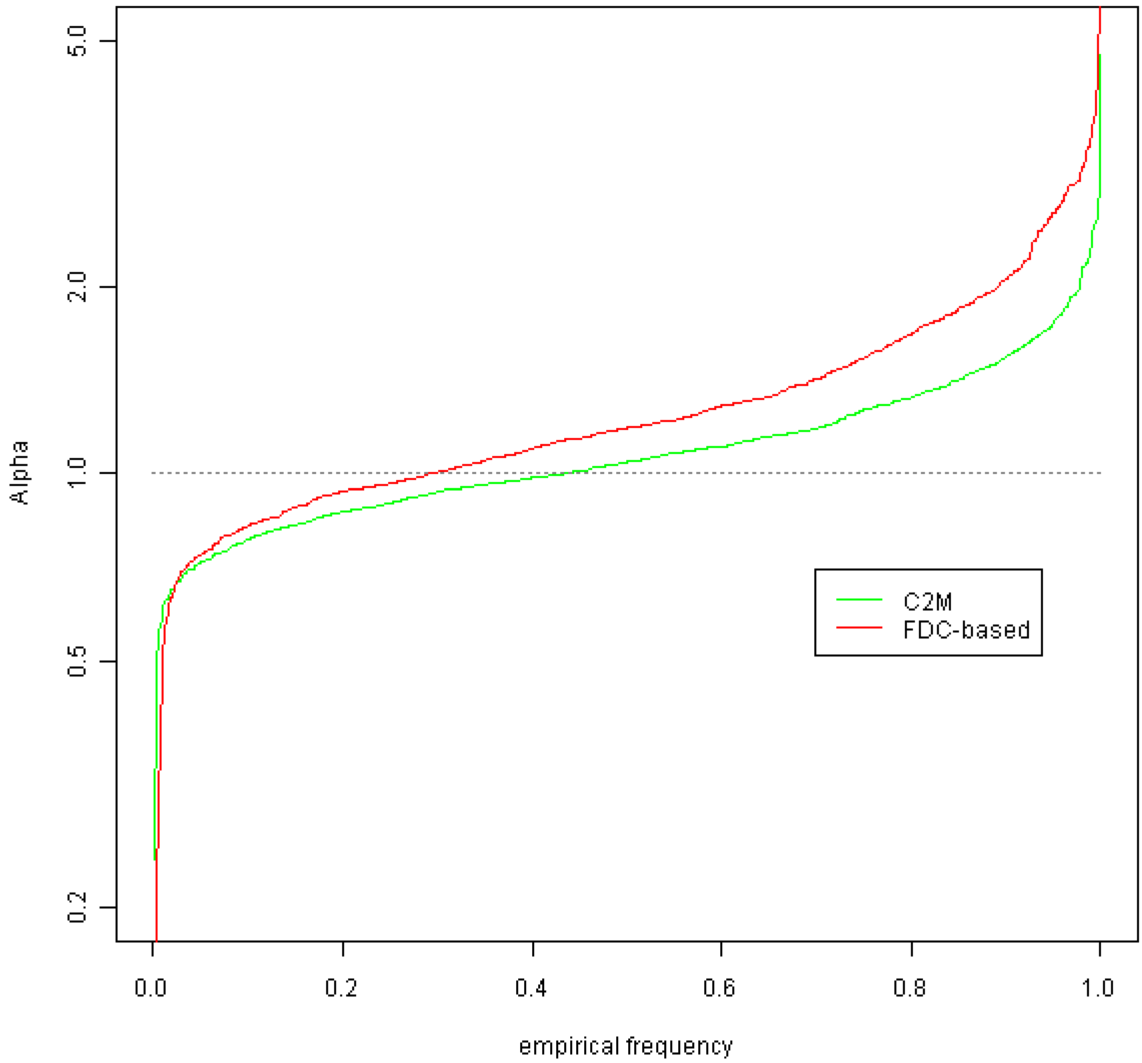


Figure 36: distribution of relative variability (alpha) for C2M-calibrated and FDC-calibrated parameters

11.2.3 Difference in regionalized parameter sets

After discussing the difference between the calibrated parameter sets (that the regionalized ones try to mimic), it is time to evaluate the differences between parameter sets regionalized with traditional site-similarity methods and with "indirect" ones.

Figure 37, Figure 38 and Figure 39 show an interesting trend: while indirect methods perform worse than the similarity based ones, the difference is smaller than when considering the two calibration criteria that these regionalization approaches try to mimic. This is particularly true for bias, where the two regionalization approaches can be considered to be almost equivalent (they are equally far from a neutral bias, even if direct methods tend to overestimate and indirect ones to underestimate) and for relative variability, where we can only observe a small difference.

The largest weakness of “indirect” approaches in comparison to direct ones is shared with the FDC-based constraint they try to mimic: poor correlation between simulated and observed runoff. This can be the consequence of lack of statistics that efficiently summarize a catchment’s dynamical response, and/or of the fact that calibrating against a small number of statistical properties is, in a way, reminiscent of calibrating against a small number of flow records, and results in a loss of information in comparison to criteria that use all the points of a time series.

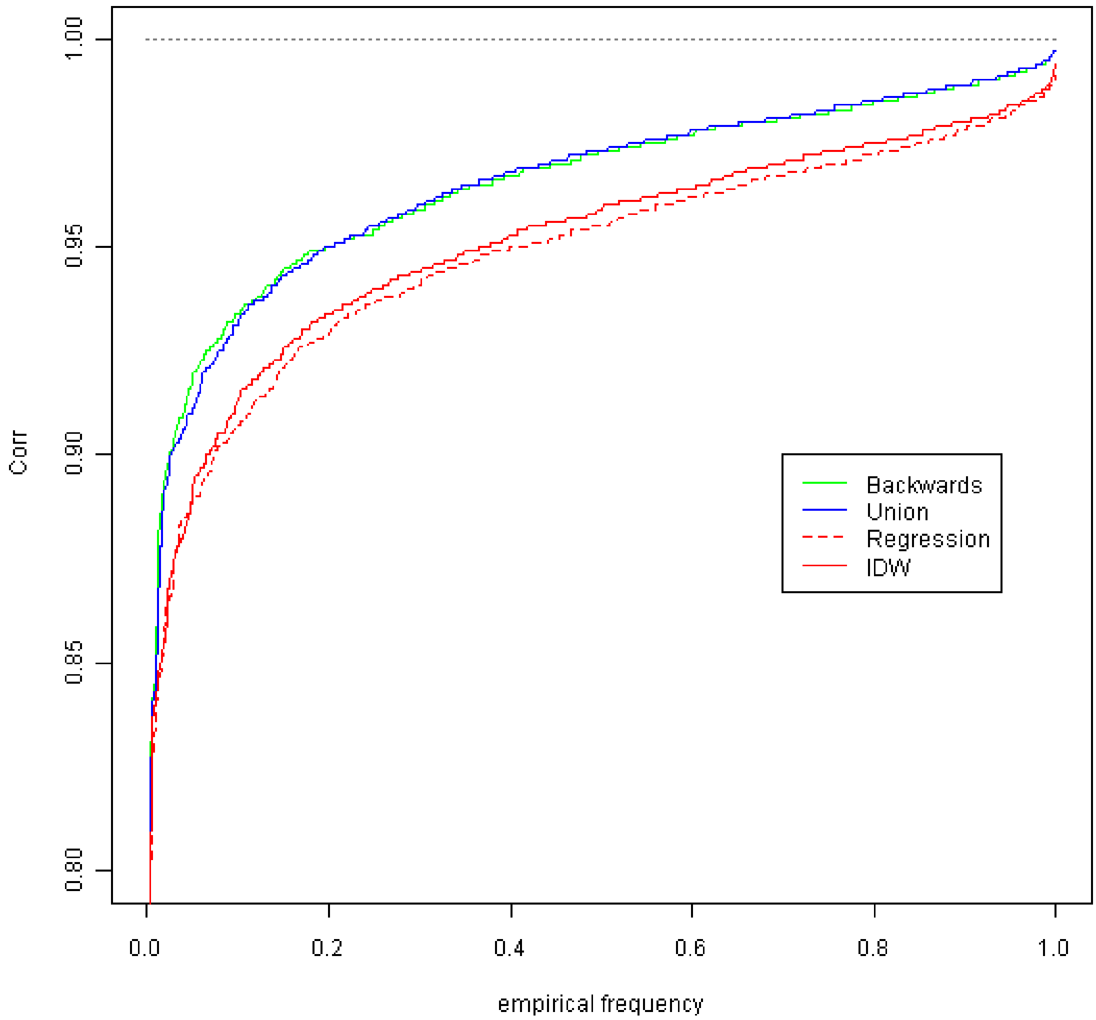


Figure 37: distribution of correlations for direct and indirect regionalization methods

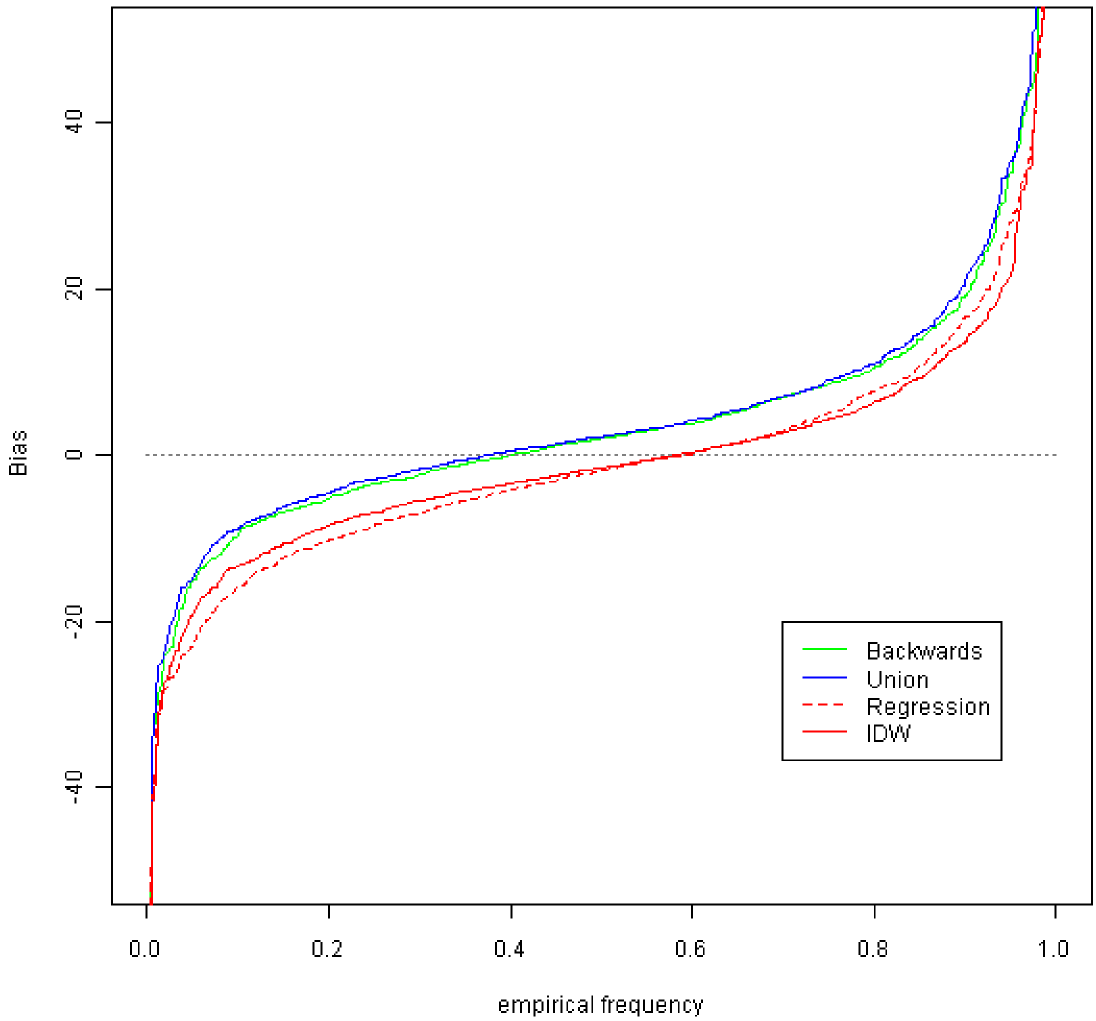


Figure 38: distribution of bias for direct and indirect regionalization methods

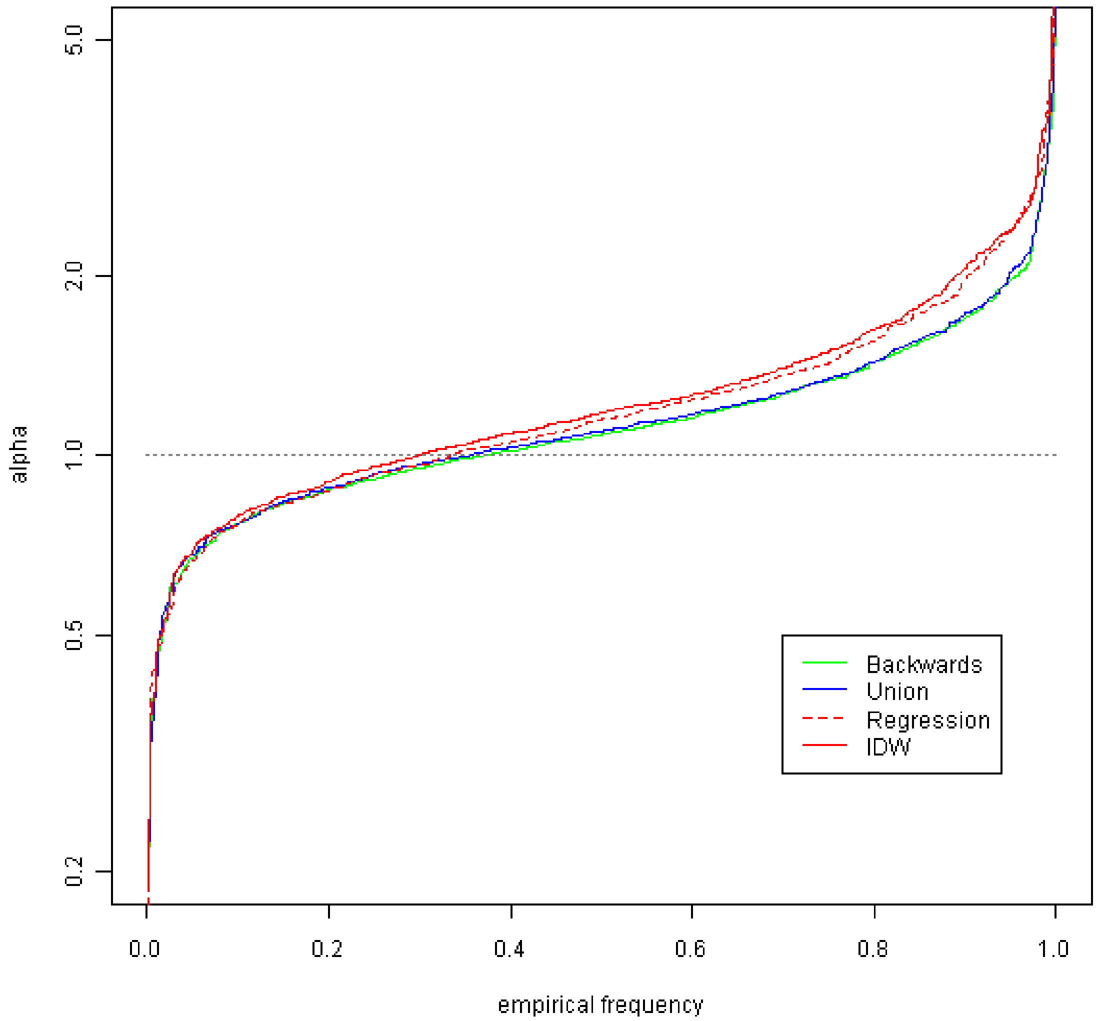


Figure 39: distribution of relative variability for direct and indirect regionalization methods

12 Conclusion

Since the kickoff of the PUB decade in 2003 (Sivapalan et al., 2003), there has been an increasing interest in regionalization issues.

Despite a number of case studies and some comparative studies, it is still difficult to assess the relative merits of the several regionalization approaches developed so far. The objective of this thesis was to give an outlook of those relative merits on the French territory.

To this aim, we developed a 3-step analysis:

- The first step (Part 2) was to develop an efficient approach to regionalize flow statistics.
- The second step (Part 3) was to assess the performance of the classical "direct" regionalization procedure.
- The third step (Part 4) was to use the insights gleaned from the 2 first steps and propose a novel framework for regionalizing models: the so-called "indirect" regionalization procedure, which has not been compared with direct approaches in previous studies.

To reach more general conclusions, we developed a methodology to assess both the performance of the tested regionalization approaches and their robustness in a context of sparse hydrometric network. Indeed, the French territory has a quite dense hydrometric network compared to other countries and the robustness test developed here may partly explain the disparate results found in the literature on the relative merits of regionalization approaches.

The second part of the thesis brought some interesting insights on our ability to predict flow statistics. Our approach was two-fold. First, we wish to explain flow statistics with the only knowledge of catchment characteristics, since this approach might be both more robust and more conceptually satisfying compared to approaches based solely on interpolation. Second, we wished to explain the residuals of the regression-based approach using information on their spatial organization. This allowed better performance, at the expense of a lack of robustness if considering a poorly gauged network.

The third part of the thesis aimed at testing several options for the "direct" regionalization approach on the basis of the GR4J model simulations:

- The most important step for the construction of a successful physical similarity metric appeared to be the selection of the most "hydrologically relevant" catchment descriptors, among the many available ones. A metric based on Principal Component Analysis, with rather loose criteria for the selection of relevant descriptors, was compared with one based on a strictly performance-oriented selection of relevant descriptors, at the expense of loose assumptions in the variable treatment (explanatory variables were treated as uncorrelated, even if this was not the case): the latter approach gave the best results. In this regard, it is important to emphasize that most, if not all, of the available catchment descriptors could *a priori* be considered to be hydrologically relevant from a subjective point of view: a good variable selection process should not be driven by a descriptor's relevance when considered alone, but rather by its role in a compound metric.
- Both of the tested similarity methods performed slightly better than the spatial-proximity alternative, with median C2M efficiency criteria of 0.56 and 0.57 for the two similarity-based methods and 0.54 for proximity. This result is in slight contradiction with previous large-scale regionalization studies (see e.g. Oudin et al., 2008; Parajka et al., 2005). This could be due to the refined approach proposed here, aimed at selecting the most hydrologically relevant catchment descriptors. Thus, one could consider that there still exists a room for progress in regionalization approaches if some other relevant catchment descriptors are proposed (particularly in regard to sub-surface characteristics) and/or new similarity metrics are tested.
- Last, two simple methods of combining similarity metrics and spatial proximity were tested, with a marginal performance increase over physical similarity alone, and, in one case, a significant decrease in robustness, despite the relatively high potential for improvement showed by an *a posteriori* combination of the two approaches.

Considering perspectives of further work on similarity and proximity approaches, we suggest that:

- This thesis work may have suffered from the relative lack of data concerning pedology (the nature of soil). The lack is “relative” in the sense that although this information is available in form of soil classifications, it should be first rearranged to obtain a limited number numerical soil descriptors that refer to the hydrological behavior of soils.
- The similarity approaches might benefit from a weighting of the donors (giving more importance to the donors who are classified as most similar to the ungauged)
- The complementary use of the similarity and proximity criteria could be further investigated, with the objective of predicting *a priori* if a certain ungauged catchment should be treated with either of the two approaches.

In the fourth part of this thesis work we focused on the performances of "indirect" regionalization methods, which are based on a previous regionalization of flow statistics. This investigation was particularly exploratory. Indeed, whereas this kind of approaches has been advocated by several authors in recent years (see e.g. Bardossy, 2007; Castiglioni et al., 2010; Westerberg et al., 2010; Yadav et al., 2007), they had been tested as operational approaches by only a few authors, and to our knowledge their performance had not been compared to "direct" regionalization schemes.

Our results suggest that the performance of these approaches strongly depends on the accuracy on the statistics' estimation, and that very precise estimates would be required to outperform the "direct" approaches.

In this regard it is important to note that even with a "perfect" estimation of flow statistics, one would get results that are still relatively far from those of the calibrated model. We have shown that this is largely due to the fact that asking the model to reproduce certain statistical properties of the observed time series is a different constraint than the one imposed by RMSE-based performance criteria used to calibrate the RR model, such as C2M (or NSE).

We believe that future work on the subject of indirect regionalization schemes should at first focus on this issue, trying to address two questions:

- If we calibrated a rainfall-runoff model so that the simulated record matches some of the statistical properties of the observed one, would we get useful simulation? In other words, can calibration (and consequently regionalization) against flow statistics be regarded as a useful performance criterion for some applications?
- Can we tweak the target flow statistics and the way we calculate the error in their reproduction so that the resulting simulations are close to optimal if evaluated with our traditional criteria?

On the positive side, indirect methods seem to benefit of the robustness of regression-based regionalization of flow statistics, and thus they have interesting performances (when compared to direct approaches) in the case of spatially-sparse gauging networks.

Finally, although their performances when used alone seem to be less satisfying than those of more traditional approaches, indirect regionalization methods seem to integrate very well in a multi-criterion approach, as showed by the example of an "hybrid" approach using spatial proximity as well: the development of multi-criteria regionalization could be a subject of further work on its own.

Overall, among the relatively simple regionalization methods tested, site-similarity based on an accurate selection of physiographic and climatic catchment descriptors seems to be the most reasonable choice, especially if one considers that our selection of catchment descriptors did not include soil or geological properties, and that this method performed better than spatial proximity despite the high density of our gauging network.

However, our results also underline that although similarity metrics show a desirable "informative" content, one should be aware of their approximate nature and of their robustness limits: regionalization based on site-similarity still requires a relatively dense gauging network to perform at its best. Pure spatial proximity can be considered an acceptable surrogate if the gauging network is locally dense and if only few catchment descriptors are available.

Indirect methods, finally, need further investigation; but if evaluated in terms of a traditional RMSE-based performance criterion, they are only interesting in semi-ungauged situations, or in the case of very sparse gauging stations.

The thesis does not give definitive answers on regionalization approaches combining different criteria, although a combination of direct and indirect approaches has been shown to give promising results. We believe that this point in particular merits a more systematic attention in future works on the subject of model regionalization.

13 References

Acreman, M.C. and Sinclair, C.D., 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. Journal of Hydrology, 84(3-4): 365-380.

Bardossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. Hydrology and Earth System Sciences, 11(2): 703-710.

Beven, K., 1989. Changing ideas in hydrology : The case of physically-based models. Journal of Hydrology, 105: 157-172.

Bloschl, G. and Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. Hydrological Processes, 9(3-4): 251-290.

Boorman, D.B., Hollis, J.M. and Lilly, A., 1995. Hydrology of soil types: a hydrologically-based classification of the soils of the United Kingdom. Report - UK Institute of Hydrology, 126.

Burn, D.H. and Boorman, D.B., 1993. Estimation of hydrological parameters at ungauged catchments. Journal of Hydrology, 143(3-4): 429-454.

Buytaert, W. and Beven, K., 2009. Regionalization as a learning process. Water Resources Research, 45(11).

Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A. and Montanari, A., 2010. Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach. Advances in Water Resources, 33(10): 1235-1242.

Darlymple, T., 1960. Flood Frequency Analysis. Water Supply Paper 1543_a. US Geological Survey, Reston, VA.

Efron, B. and Gong, G., 1983. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. The American Statistician, 37(1): 36-48.

Egbuniwe, N. and Todd, D.K., 1976. APPLICATION OF THE STANFORD WATERSHED MODEL TO NIGERIAN WATERSHEDS. Water Resources Bulletin, 12(3): 449-460.

Fernandez, W., Vogel, R.M. and Sankarasubramanian, A., 2000. Regional calibration of a watershed model. Hydrological Sciences Journal, 45(5): 689-707.

Goswami, M. and O'Connor, K.M., 2006. Flow simulation in an ungauged basin: an alternative approach to parameterization of a conceptual model using regional data, Large sample basin experiments for hydrological model parameterization: results of the model parameter experiment, pp. 149-158.

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

Kay, A.L., Jones, D.A., Crooks, S.M., Kjeldsen, T.R. and Fung, C.F., 2007. An investigation of site-similarity approaches to generalisation of a rainfall-runoff model. Hydrology and Earth System Sciences, 11(1): 500-515.

Kjeldsen, T.R. and Jones, D.A., 2010. Predicting the index flood in ungauged UK catchments: On the link between data-transfer and spatial model error structure. Journal of Hydrology, 387(1-2): 1-9.

Lamb, R., Crewett, J. and Calver, A., 2000. Relating hydrological model parameters and catchment properties to estimate flood frequencies from simulated river flows. Proceedings of the BHS 7th National Hydrology Symposium, Newcastle: 357–364.

Le Moine, N., 2008. Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modèles pluie-débit? PhD Thesis, Université Pierre et Marie Curie, Paris, 322 pp.

Li, H., Zhang, Y., Chiew, F.H.S. and Xu, S., 2009. Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. Journal of Hydrology, 370(1-4): 155-162.

Liu, Z. and Todini, E., 2002. Towards a comprehensive physically-based rainfall-runoff model. Hydrology and Earth System Sciences, 6(5): 859-881.

Longobardi, A. and Villani, P., 2008. Baseflow index regionalization analysis in a mediterranean area and data scarcity context: Role of the catchment permeability index. Journal of Hydrology, 355(1-4): 63-75.

Mathevet, T., Michel, C., Andréassian, V. and Perrin, C., 2006. A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins., IAHS Red Books Series n°307, pp. 211-219.

Mazvimavi, D., Meijerink, A.M.J., Savenije, H.H.G. and Stein, A., 2005. Prediction of flow characteristics using multiple regression and neural networks: A case study in Zimbabwe. Physics and Chemistry of the Earth, 30(11-16 SPEC. ISS.): 639-647.

McIntyre, N., Lee, H., Wheater, H., Young, A. and Wagener, T., 2005. Ensemble predictions of runoff in ungauged catchments. Water Resources Research, 41.

Merz, R. and Blöschl, G., 2004. Regionalisation of catchment model parameters. Journal of Hydrology, 287(1-4): 95-123.

Merz, R. and Blöschl, G., 2005. Flood frequency regionalisation - Spatial proximity vs. catchment attributes. Journal of Hydrology, 302(1-4): 283-306.

Montanari, A. and Toth, E., 2007. Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins? Water Resources Research, 43(5).

Oudin, L., Andréassian, V., Perrin, C., Michel, C. and Le Moine, N., 2008. Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. Water Resources Research, 44(3).

Oudin, L. et al., 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. Journal of Hydrology, 303(1-4): 290-306.

Oudin, L., Kay, A., Andréassian, V. and Perrin, C., 2010. Are seemingly physically similar catchments truly hydrologically similar? Water Resources Research, 46(11).

Parajka, J., Blöschl, G. and Merz, R., 2007. Regional calibration of catchment models: Potential for ungauged catchments. Water Resources Research, 43(6).

Parajka, J., Merz, R. and Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. Hydrology and Earth System Sciences, 9(3): 157-171.

Perrin, C., Michel, C. and Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology, 279: 275-289.

Reichl, J.P.C., Western, A.W., McIntyre, N.R. and Chiew, F.H.S., 2009. Optimization of a similarity measure for estimating ungauged streamflow. Water Resources Research, 45(W10423).

Rojas Serna, C., Michel, C., Perrin, C. and Andréassian, V., 2006. Ungaged catchments: How to make the most of a few streamflow measurements? In: V. Andréassian, A. Hall, N. Chahinian, C. Perrin and J. Schaake (Editors), Large sample basin experiments for hydrological model parameterisation. Results of the MOdel Parameter Experiment (MOPEX). IAHS, Wallingford.

Sauquet, E., Gottschalk, L. and Leblois, E., 2000. Mapping average annual runoff: A hierarchical approach applying a stochastic interpolation scheme. Hydrological Sciences Journal, 45(6): 799-815.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P.A. and Carrillo, G., 2011. Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. Hydrology and Earth System Sciences, 15(9): 2895-2911.

Schneider, M.K., Brunner, F., Hollis, J.M. and Stamm, C., 2007. Towards a hydrological classification of European soils: Preliminary test of its predictive power for the base flow index using river discharge data. Hydrology and Earth System Sciences, 11(4): 1501-1513.

Seibert, J. and Beven, K.J., 2009. Gauging the ungauged basin: How many discharge measurements are needed? Hydrology and Earth System Sciences, 13(6): 883-892.

Sivapalan, M. et al., 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal, 48(6): 857-880.

Skøien, J.O., Merz, R. and Blöschl, G., 2006. Top-kriging - Geostatistics on stream networks. Hydrology and Earth System Sciences, 10(2): 277-287.

Smakhtin, V.Y., Hughes, D.A. and Creuse-Naudin, E., 1997. Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa. Hydrological Sciences Journal, 42(6): 919-936.

Tasker, G.D. and Stedinger, J.R., 1989. An operational GLS model for hydrologic regression. Journal of Hydrology, 111(1-4): 361-375.

Todini, E., 2011. History and perspectives of hydrological catchment modelling. Hydrology Research, 42(2-3): 73-85.

Tulu, T., 1991. Simulation of streamflows for ungauged catchments. Journal of Hydrology, 129(1-4): 3-17.

Vandewiele, G.L. and Elias, A., 1995. Monthly water balance of ungauged catchments obtained by geographical regionalization. Journal of Hydrology, 170(1-4): 277-291.

Vandewiele, G.L., Xu, C.Y. and Huybrecht, W., 1991. Regionalisation of physically-based water balance models in Belgium. Application to ungauged catchments. Water Resources Management, 5: 199-208.

Velez, J.J., Puricelli, M., Unzu, F.L. and Frances, F., 2009. Parameter extrapolation to ungauged basins with a hydrological distributed model in a regional framework. Hydrology and Earth System Sciences, 13(2): 229-246.

Viglione, A., Laio, F. and Claps, P., 2007. A comparison of homogeneity tests for regional frequency analysis. Water Resources Research, 43(3).

Wagener, T. and Wheater, H.S., 2006. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. Journal of Hydrology(320): 132-154.

Westerberg, I.K. et al., 2010. Calibration of hydrological models using flow-duration curves. Hydrol. Earth Syst. Sci. Discuss., 7(6): 9467-9522.

Whittle, P., 1953. Estimation and information in stationary time series. Arkiv för Matematik, 2(5): 423-434.

Winsemius, H.C., Schaefli, B., Montanari, A. and Savenije, H.H.G., 2009. On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information. Water Resour. Res., 45(12): W12422.

Yadav, M., Wagener, T. and Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources, 30(8): 1756-1774.

Yapo, P.O., Gupta, H.V. and Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology, 181(1-4): 23-48.

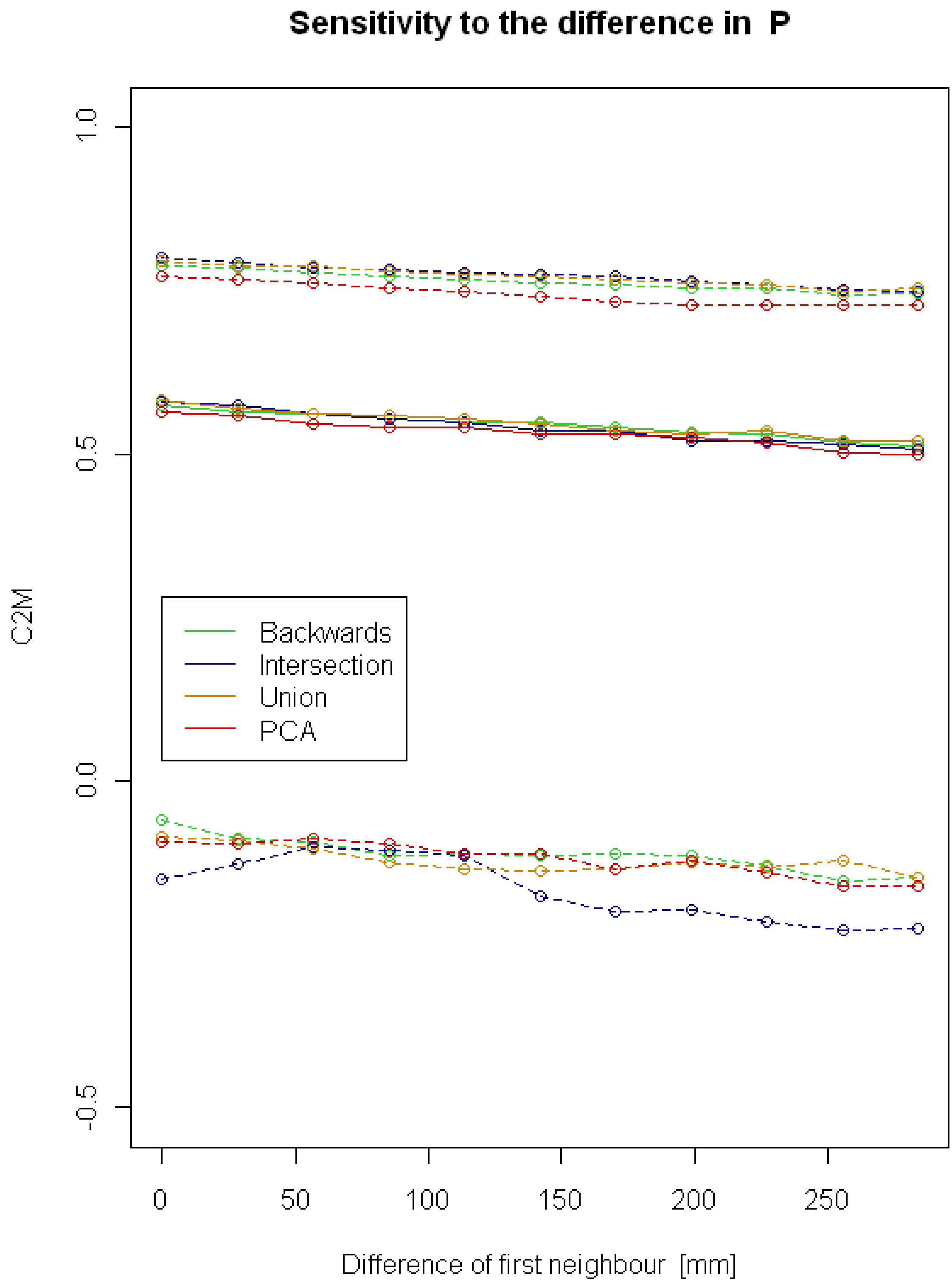
Yu, P.S. and Yang, T.C., 2000. Using synthetic flow duration curves for rainfall-runoff model calibration at ungauged sites. Hydrological Processes, 14(1): 117-133.

Zvolensky, M., Kohnova, S., Hlavcova, K., Szolgay, J. and Parajka, J., 2008. Regionalisation of rainfall-runoff model parameters based on geographical location of gauged catchments. Journal of Hydrology and Hydromechanics, 56(3): 176-189.

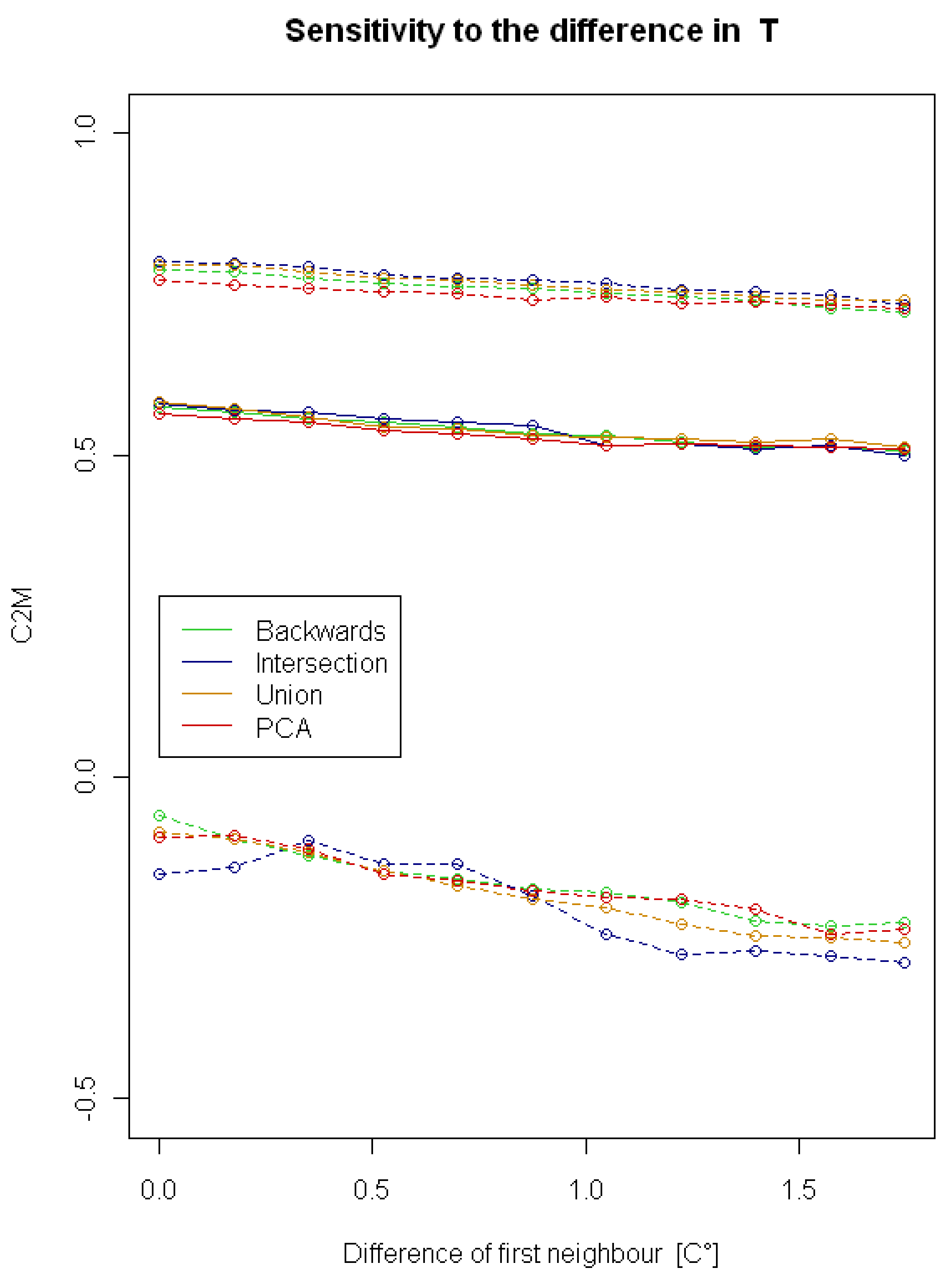
Part 6 – Appendices

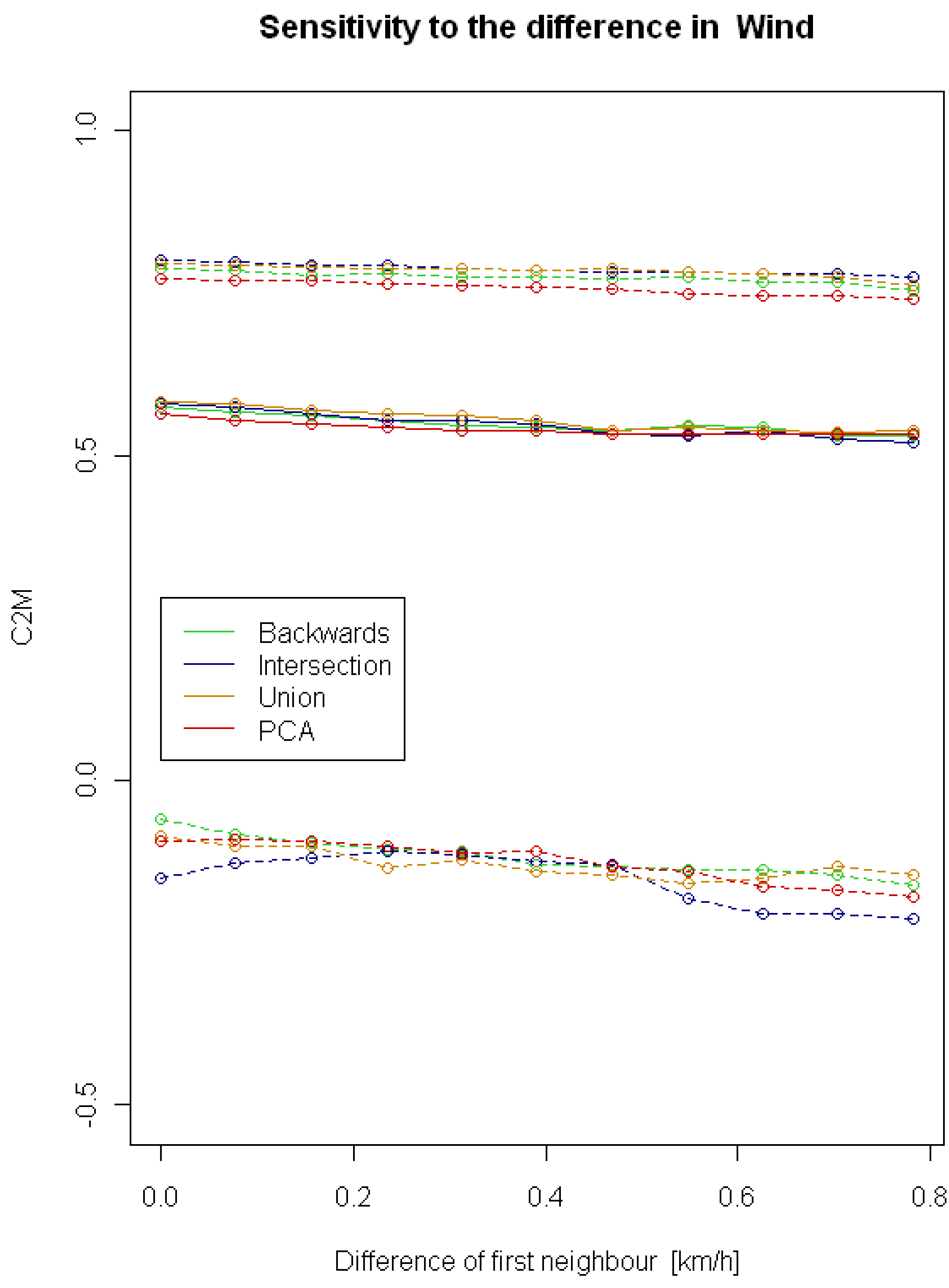
14 Sensitivity to the elimination of similar donors: graphic results

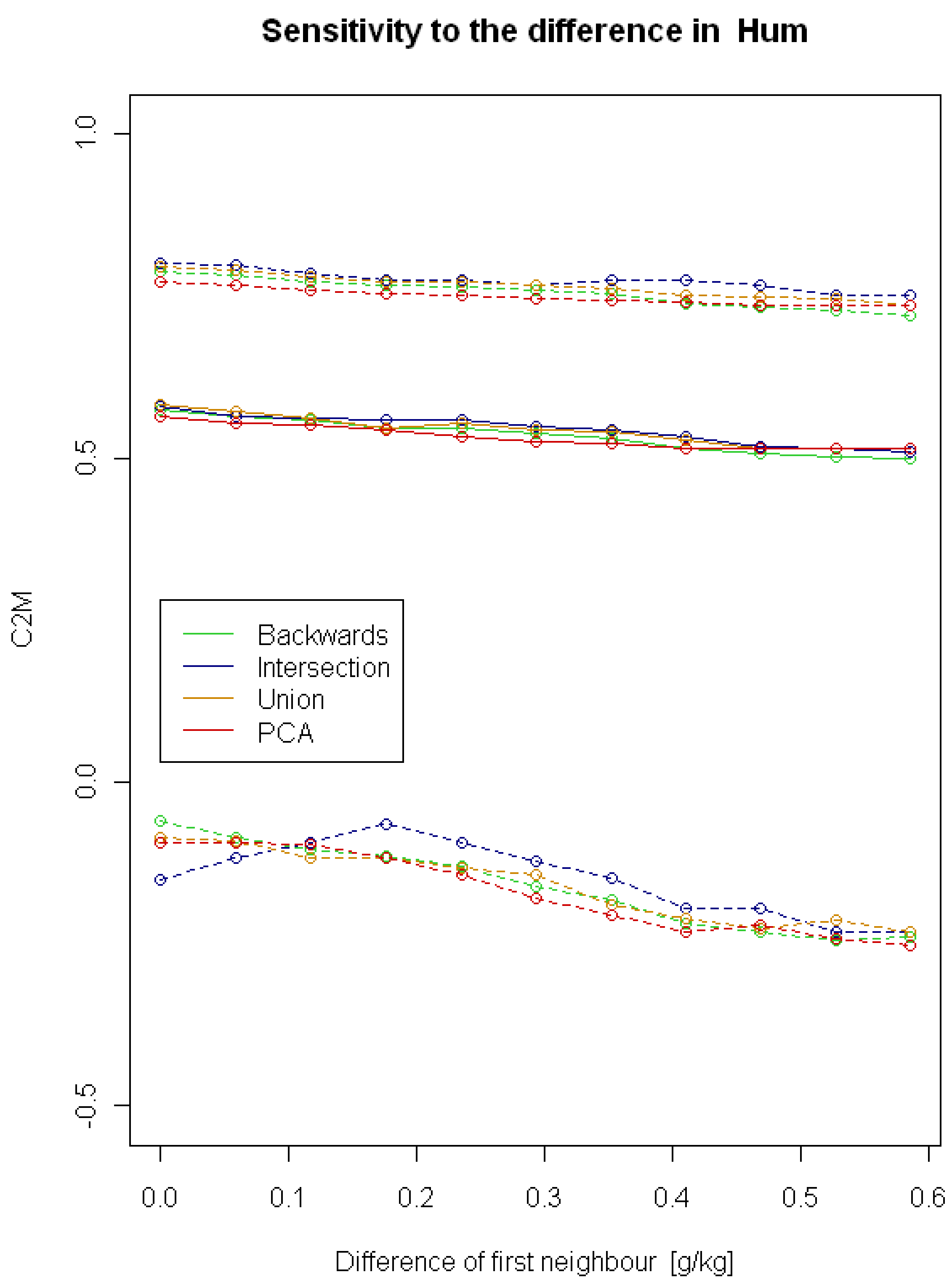
In the following pages we show the results of the elimination of donors which are similar to the receiver catchment, according to each descriptor in our list. The graphics follow the same conventions of those in chapter 9: the upper dashed line represents the 0.9 quantile of the performance distribution, the continuous line shows the median, and the lower dashed line the 0.1 quantile.

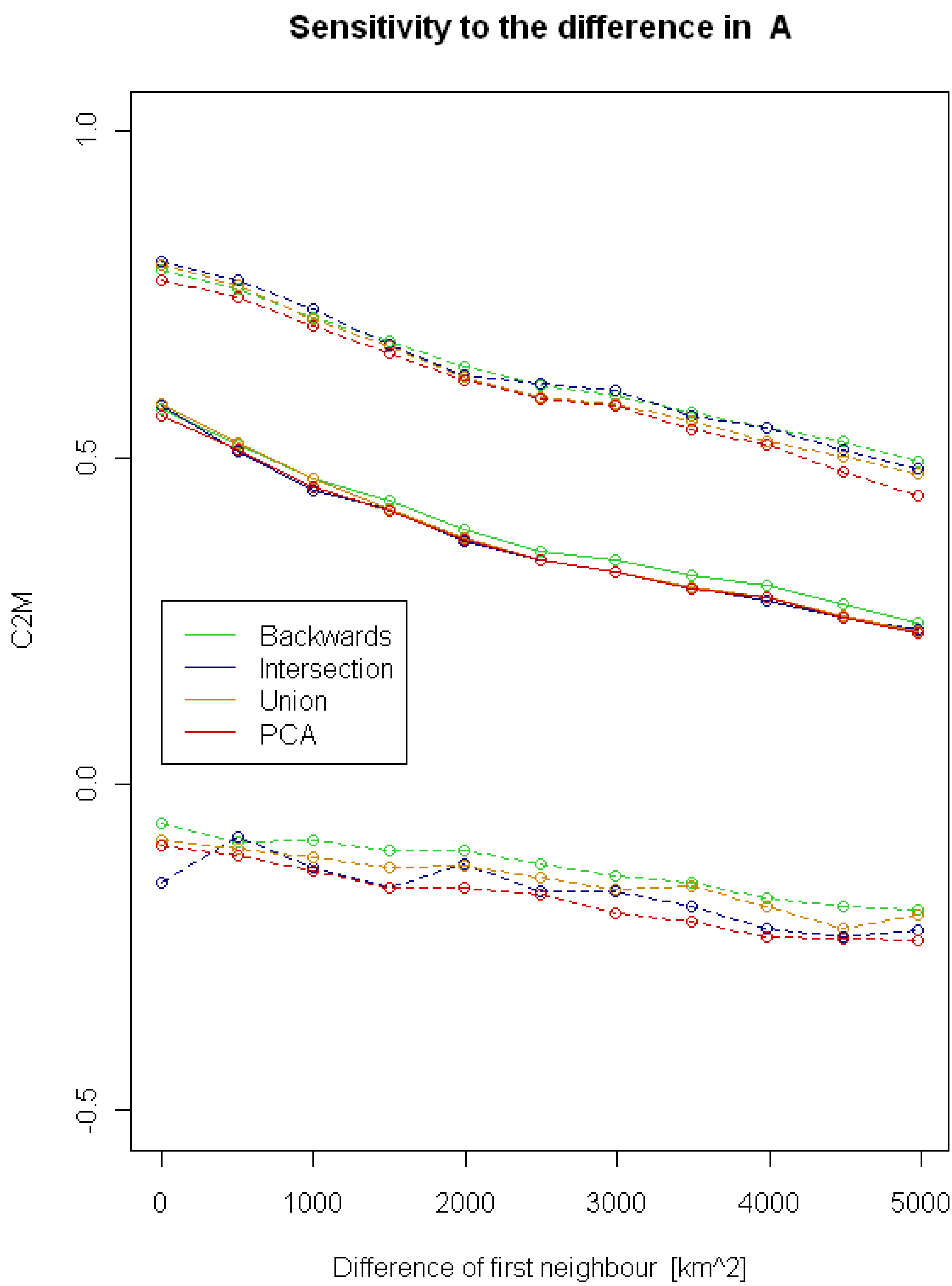


Sensitivity to the difference in P
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0
50
100
150
200
250
Difference of first neighbour [mm]

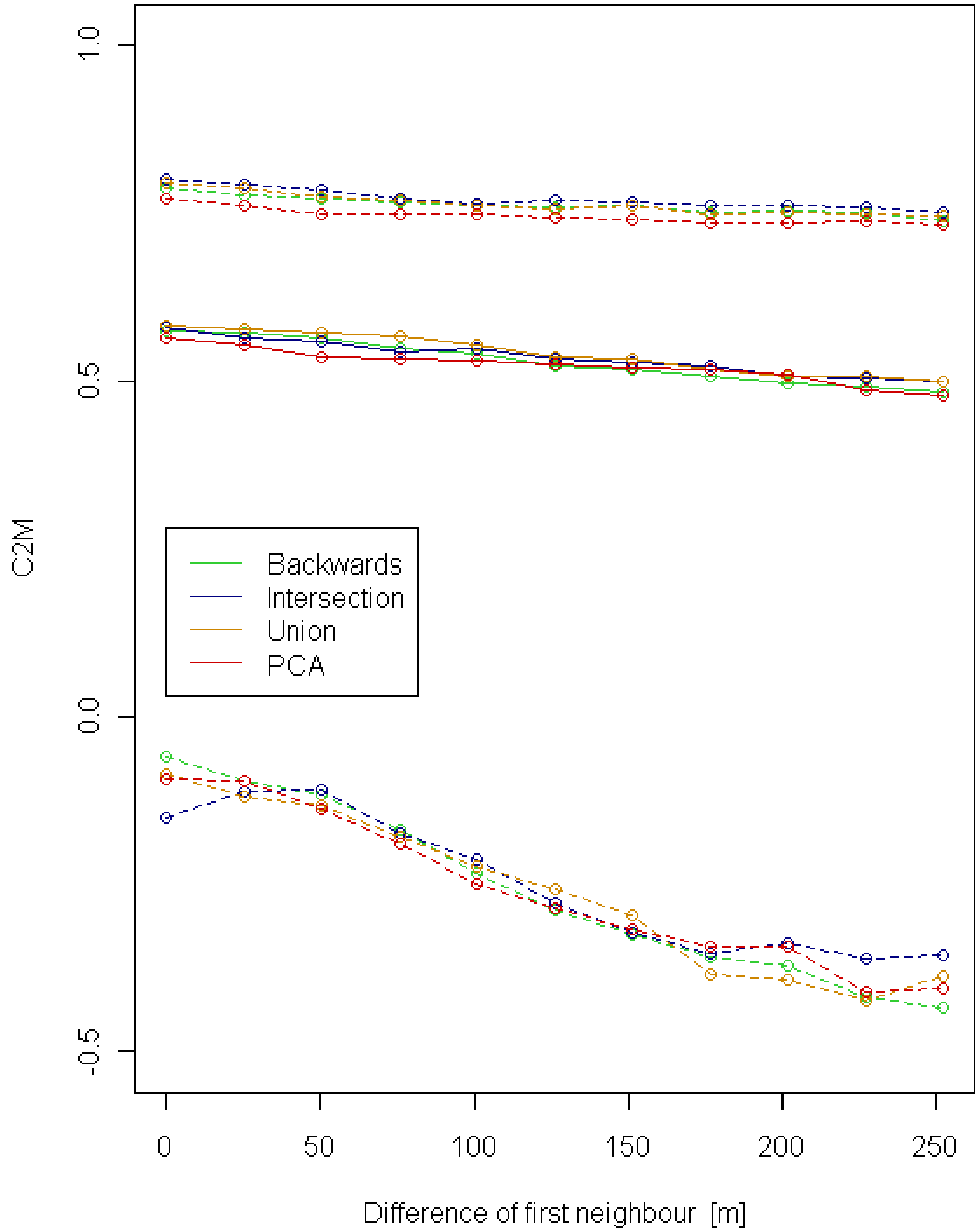

Sensitivity to the difference in T
C2M
1.0
0.5
0.0
-0.5
0.0
0.5
1.0
1.5
Difference of first neighbour [C°]
Backwards
Intersection
Union
PCA

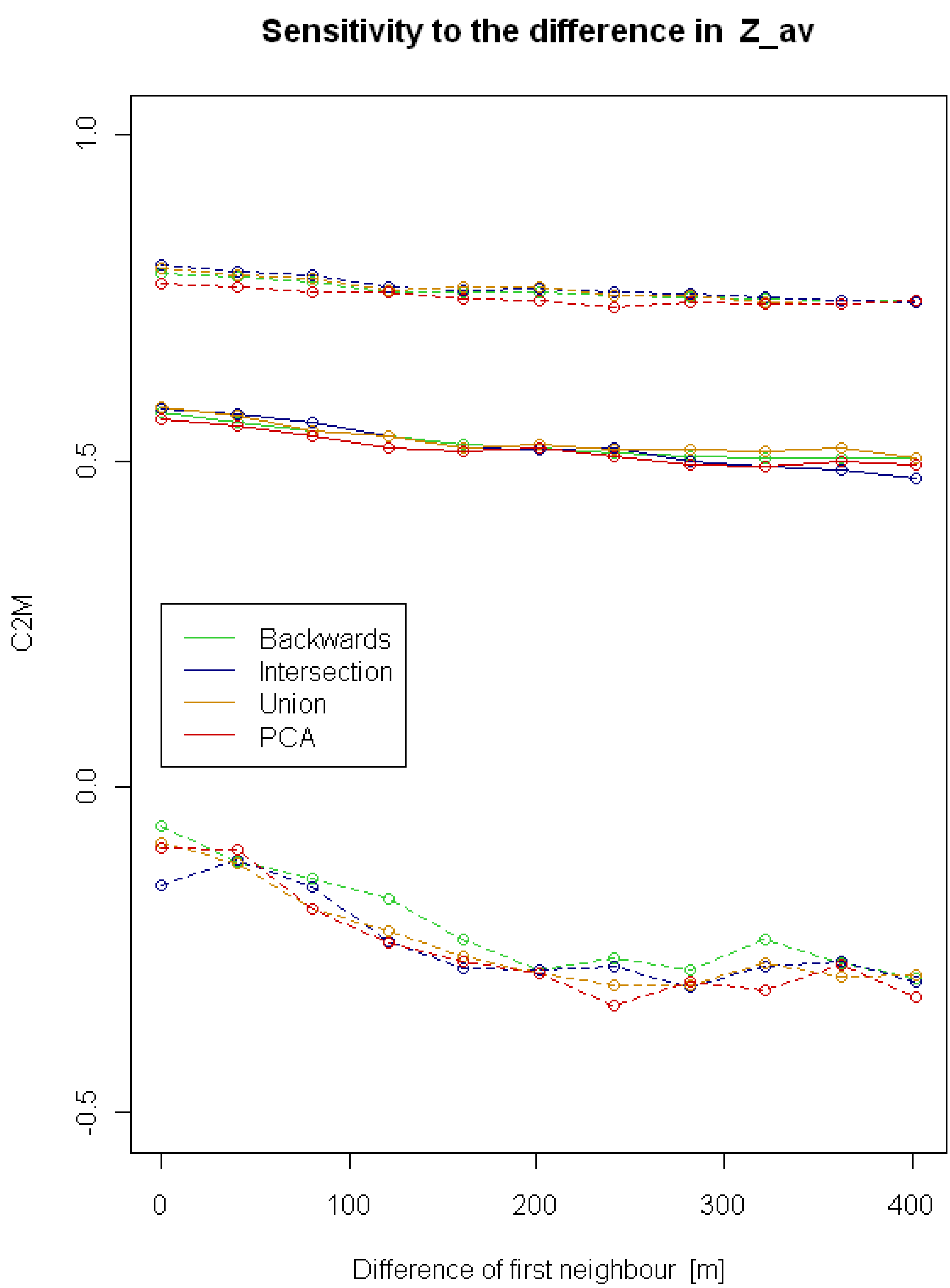

Sensitivity to the difference in Wind
1.0
0.5
0.0
-0.5
C2M
Backwards
Intersection
Union
PCA
0.0
0.2
0.4
0.6
0.8
Difference of first neighbour [km/h]

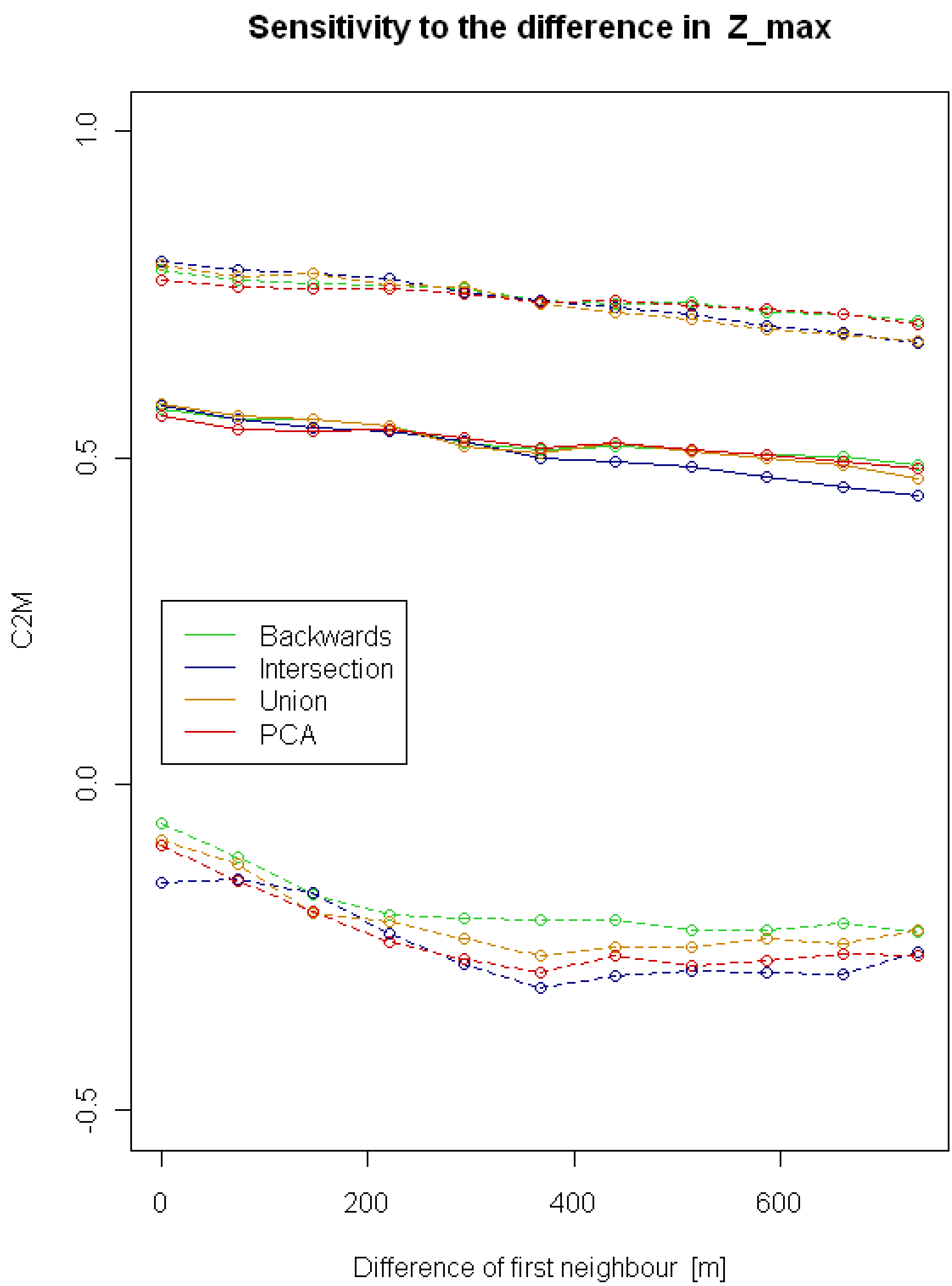

Sensitivity to the difference in Hum
C2M
1.0
0.5
0.0
-0.5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Difference of first neighbour [g/kg]
Backwards
Intersection
Union
PCA


Sensitivity to the difference in A
C2M
1.0
0.5
0.0
-0.5
0
1000
2000
3000
4000
5000
Difference of first neighbour [km^2]
Backwards
Intersection
Union
PCA

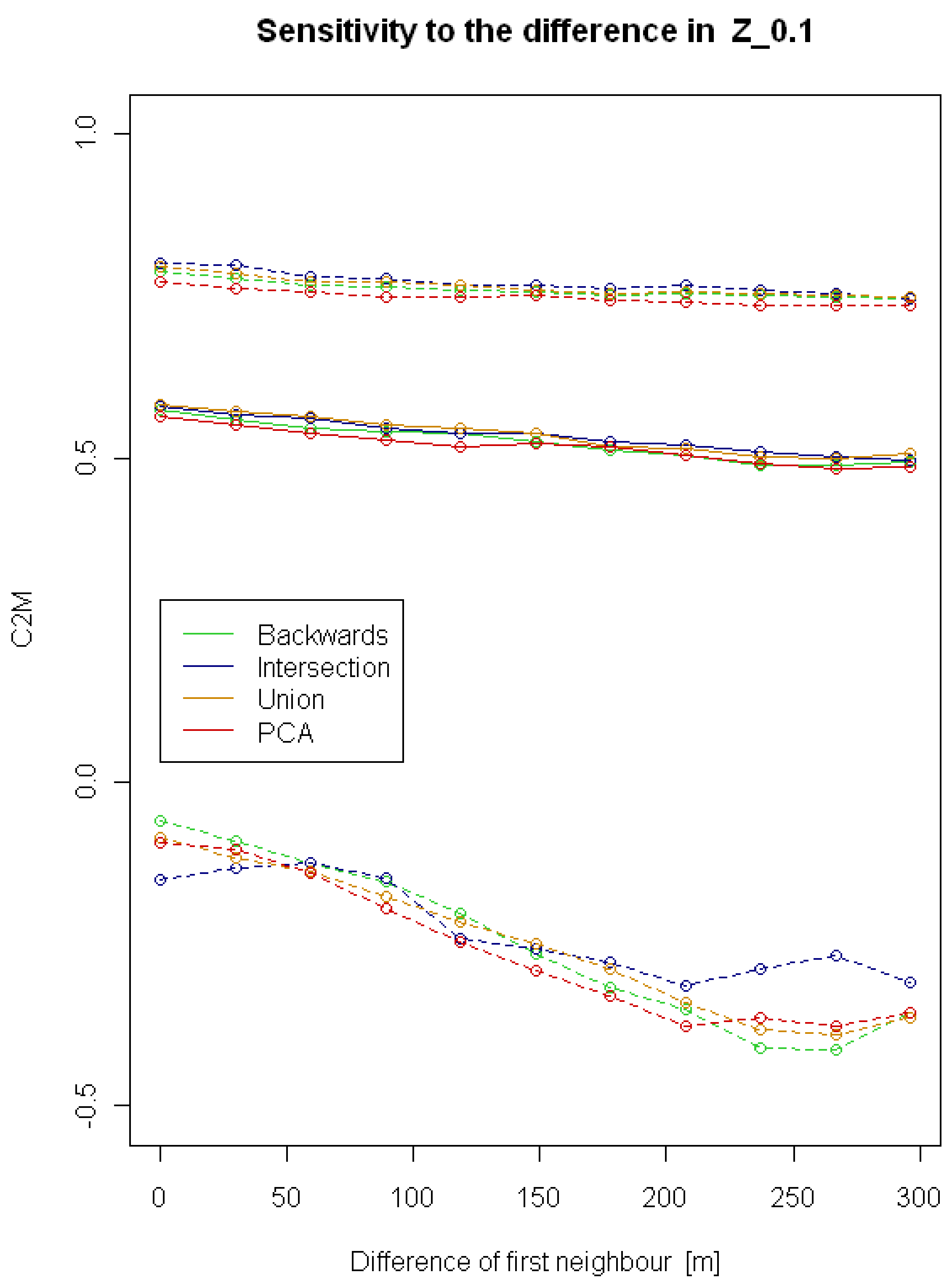
Sensitivity to the difference in Z_min



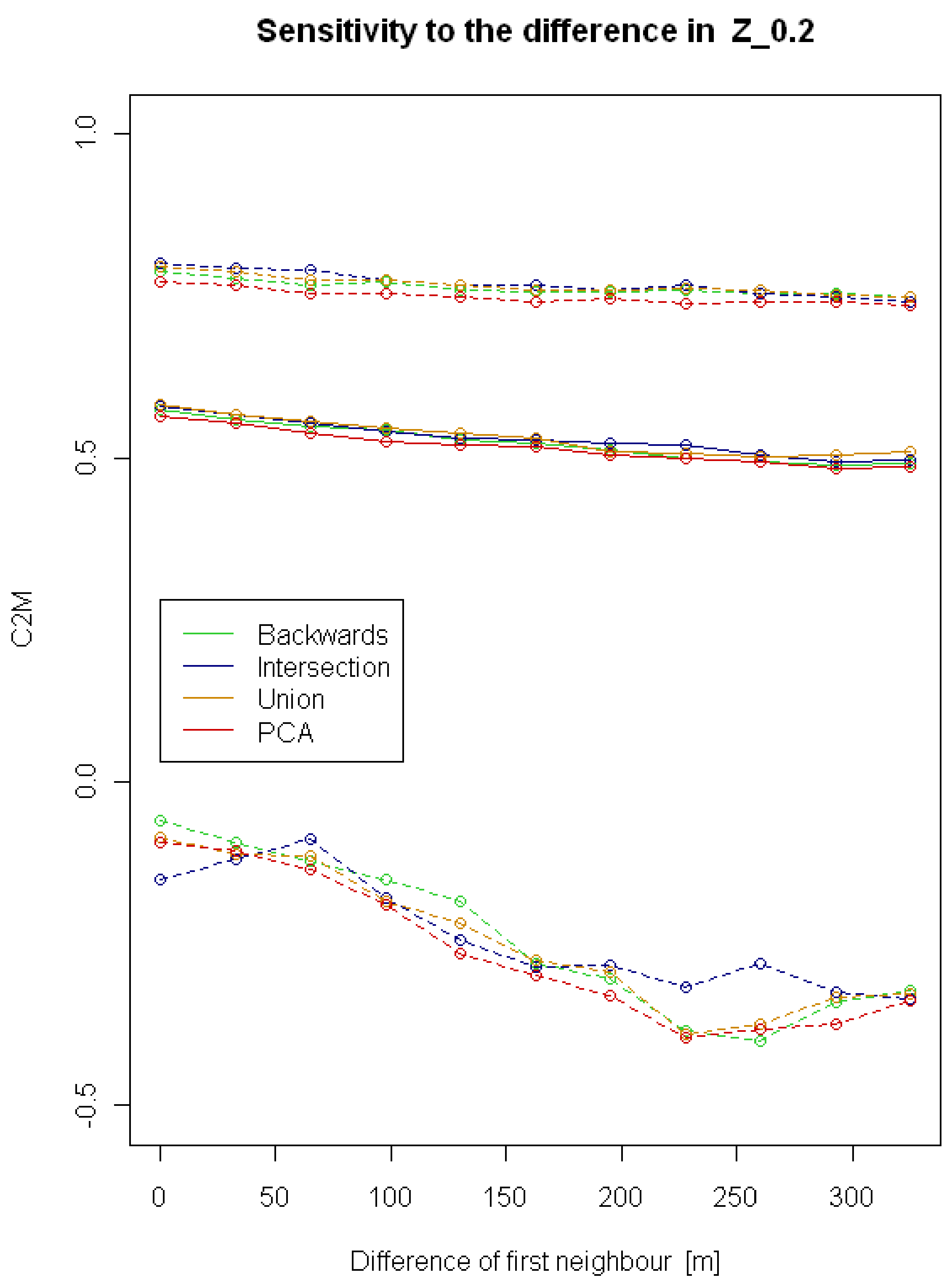

Sensitivity to the difference in Z_av
C2M
1.0
0.5
0.0
-0.5
0
100
200
300
400
Difference of first neighbour [m]
Backwards
Intersection
Union
PCA

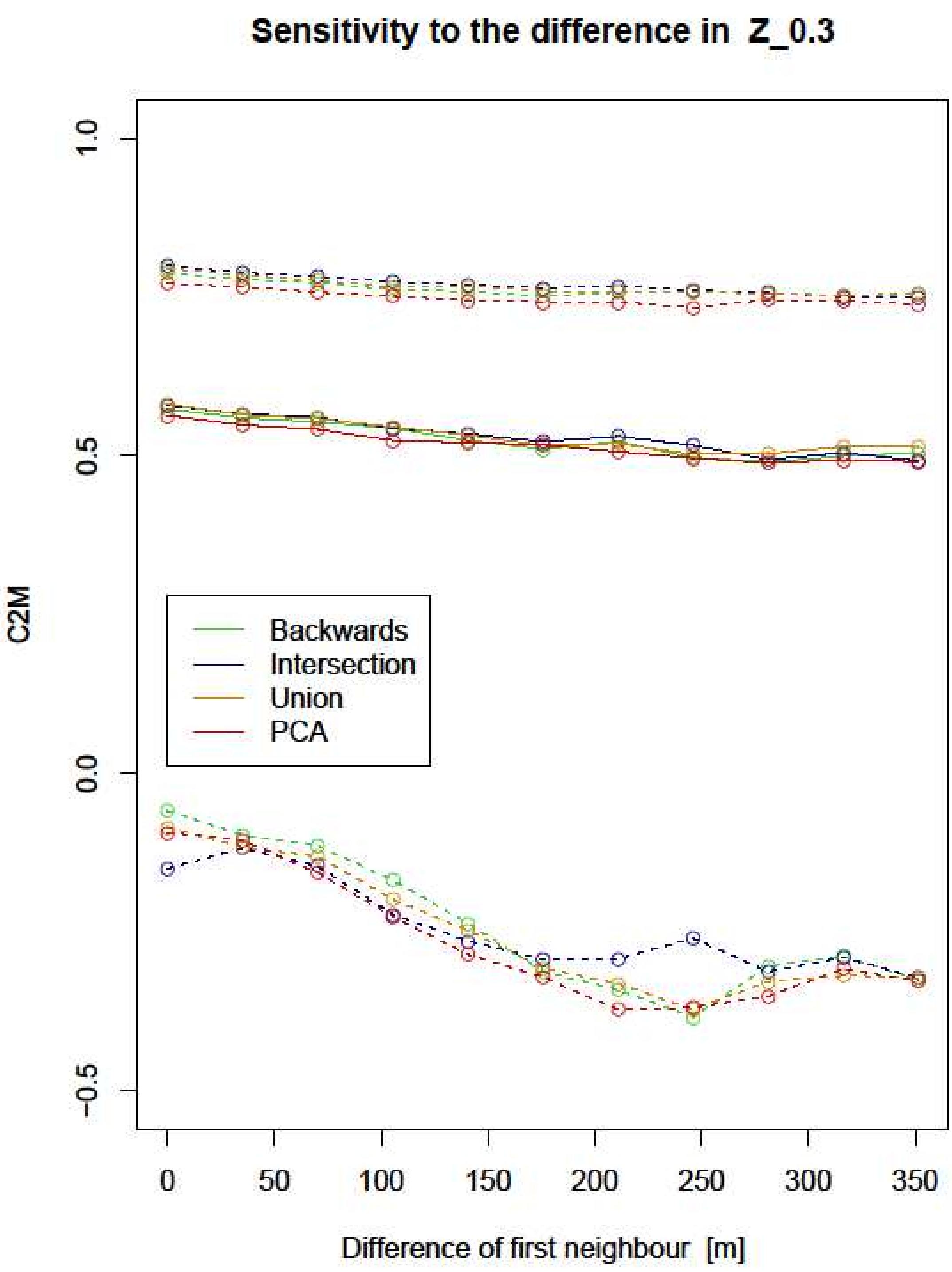


Sensitivity to the difference in Z_max
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0
200
400
600
Difference of first neighbour [m]



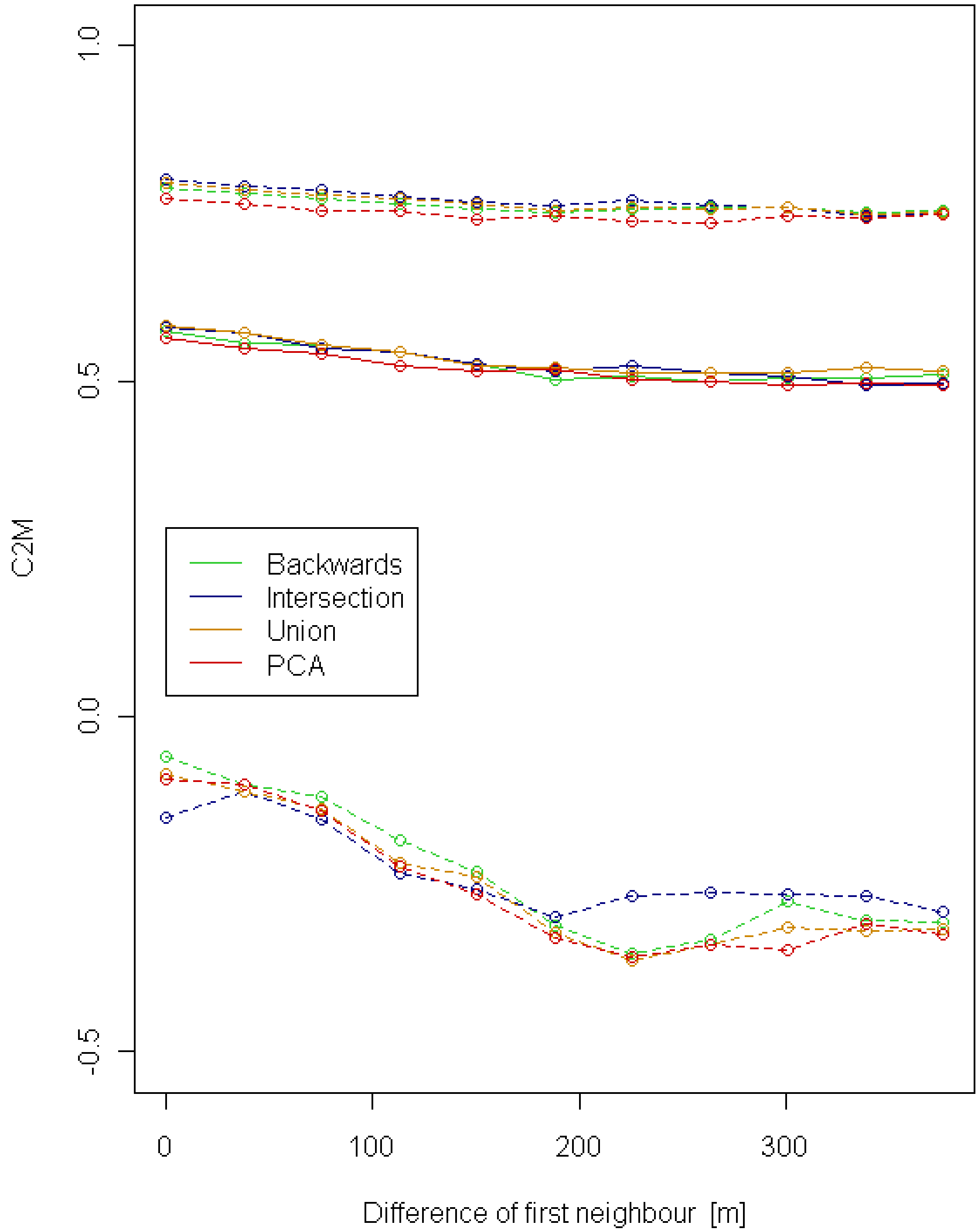
Sensitivity to the difference in Z_0.1
C2M
1.0
0.5
0.0
-0.5
0
50
100
150
200
250
300
Difference of first neighbour [m]
Backwards
Intersection
Union
PCA

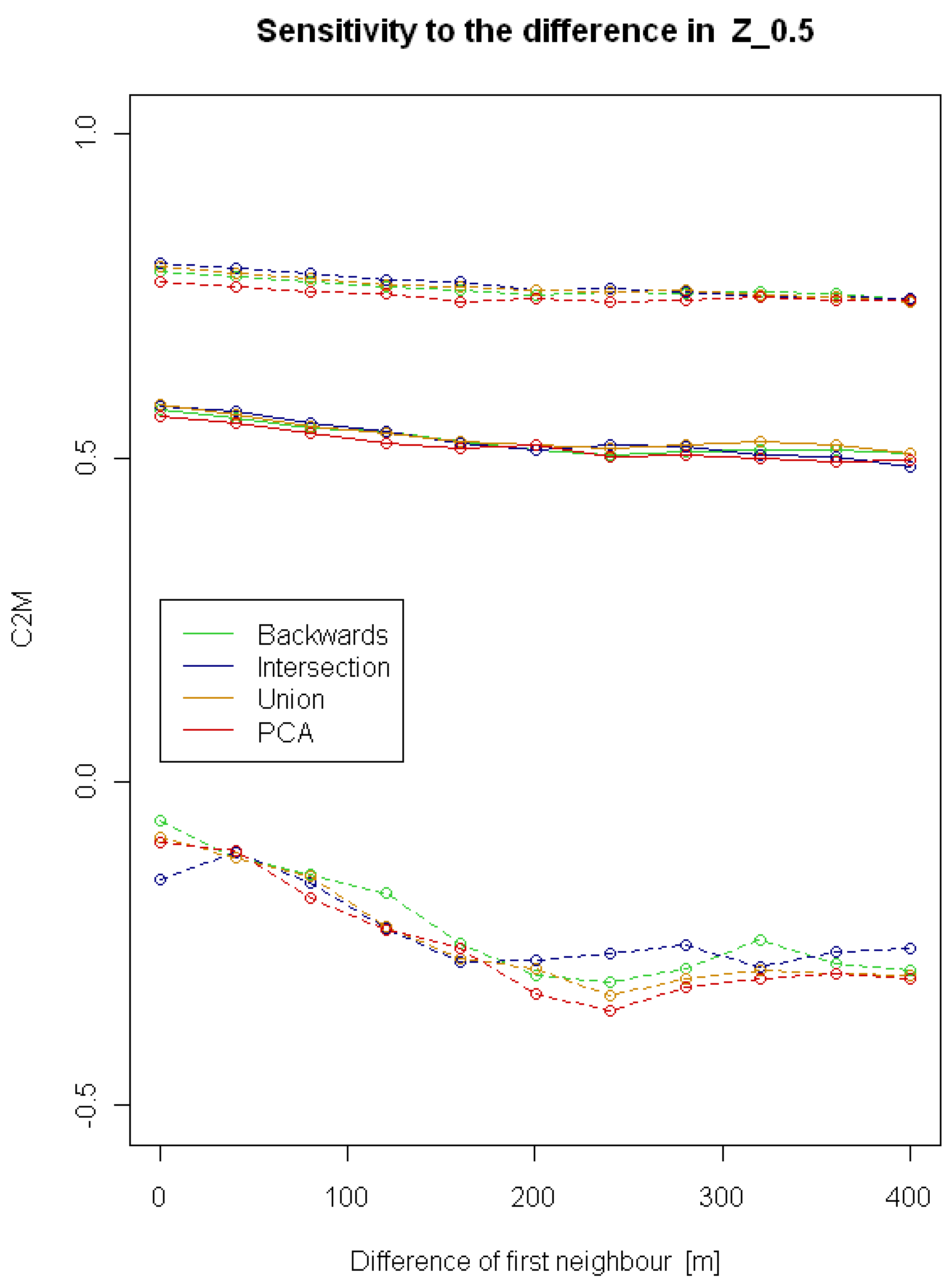

Sensitivity to the difference in Z_0.2
C2M
1.0
0.5
0.0
-0.5
0
50
100
150
200
250
300
Difference of first neighbour [m]
Backwards
Intersection
Union
PCA

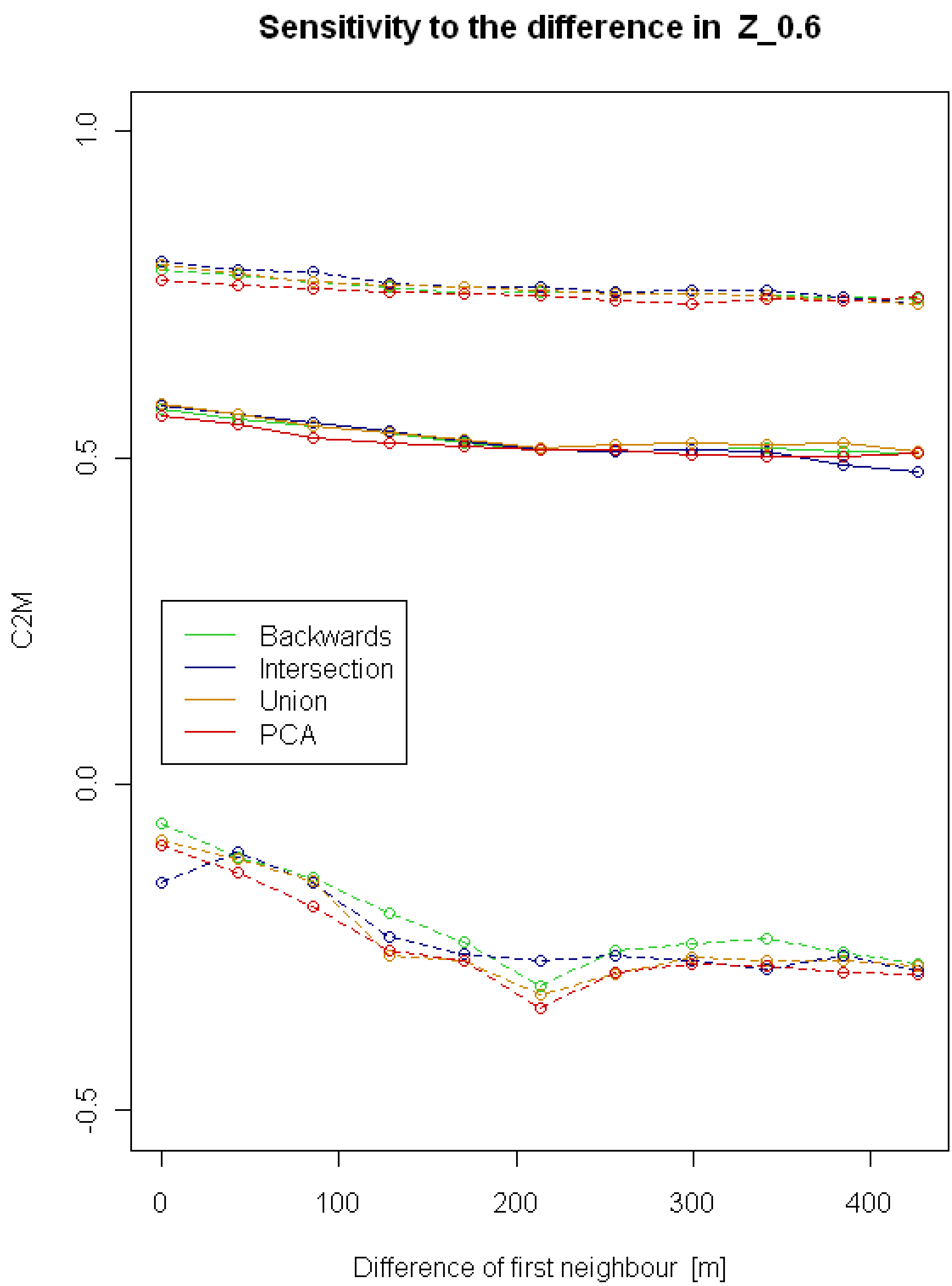


Sensitivity to the difference in Z_0.3
C2M
1.0
0.5
0.0
−0.5
0
50
100
150
200
250
300
350
Difference of first neighbour [m]
Backwards
Intersection
Union
PCA

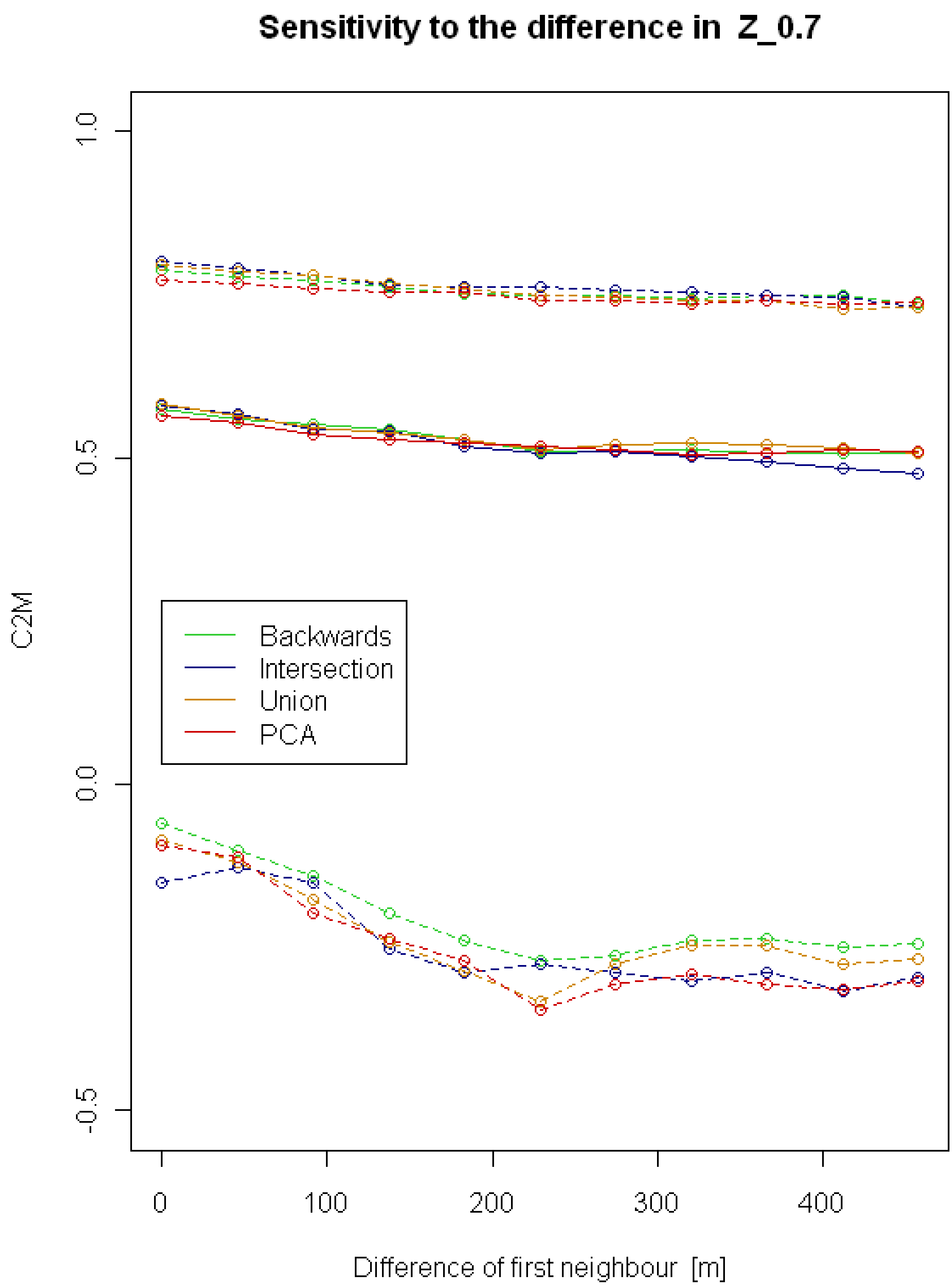
Sensitivity to the difference in Z_0.4



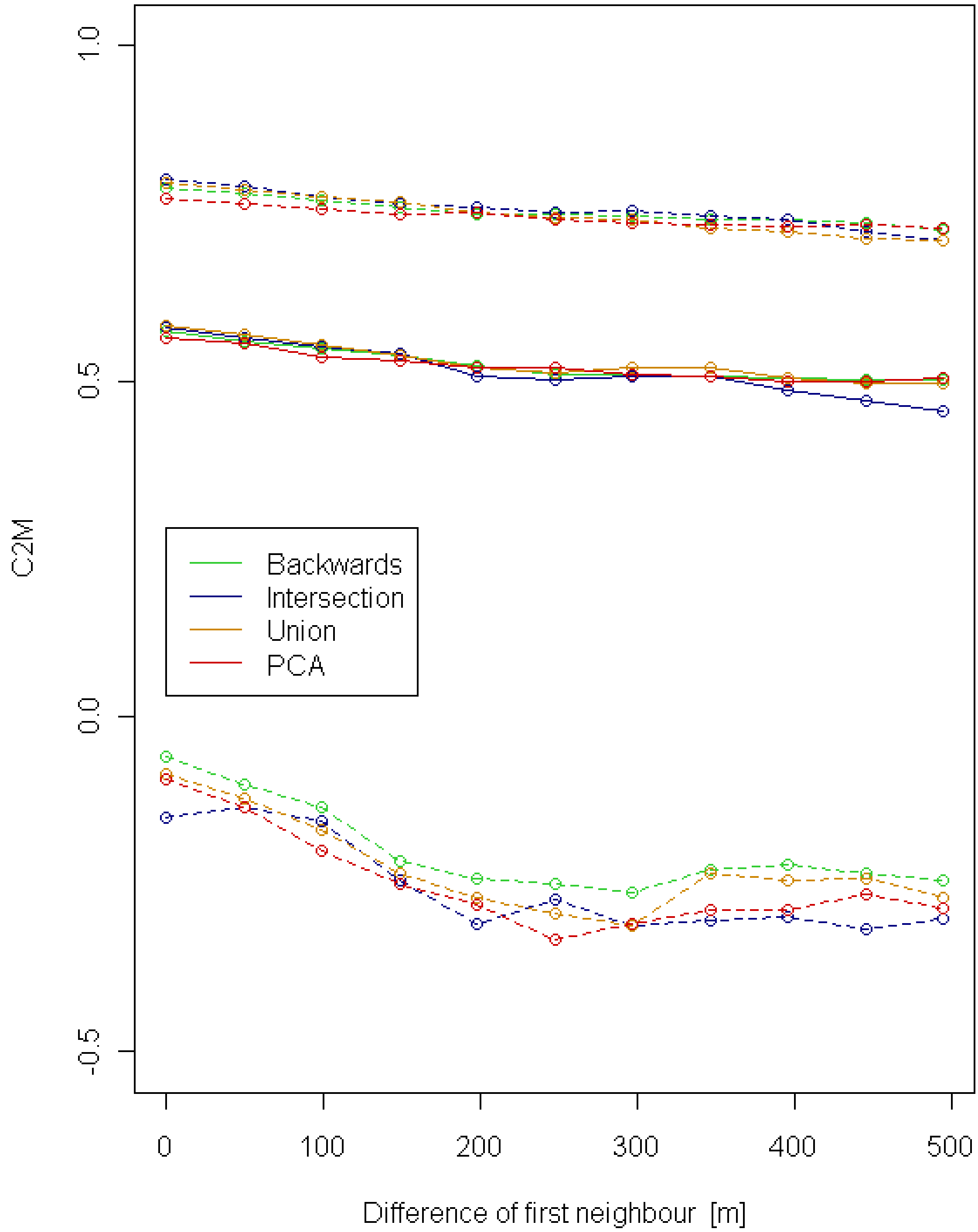

Sensitivity to the difference in Z_0.5
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0
100
200
300
400
Difference of first neighbour [m]

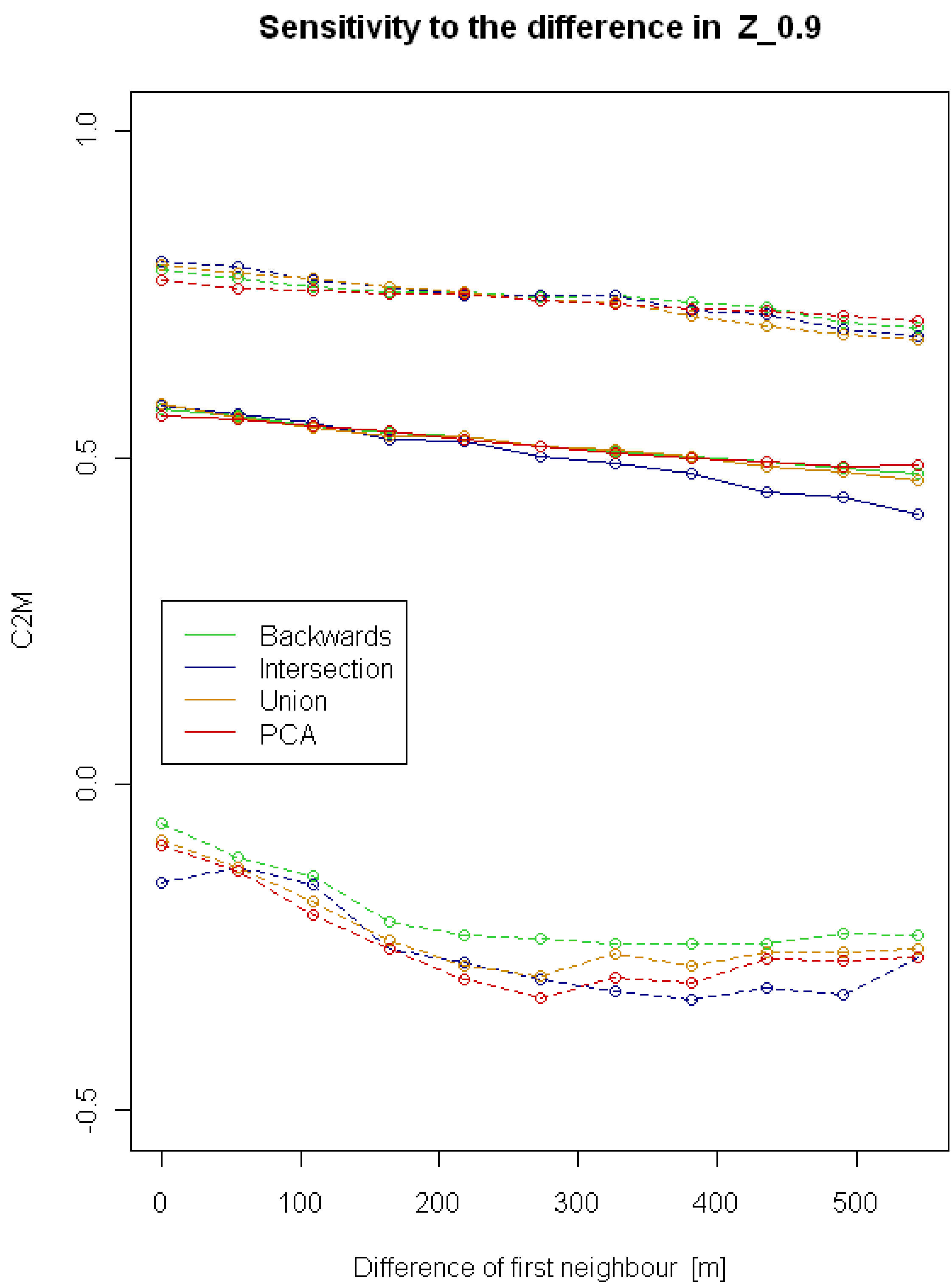


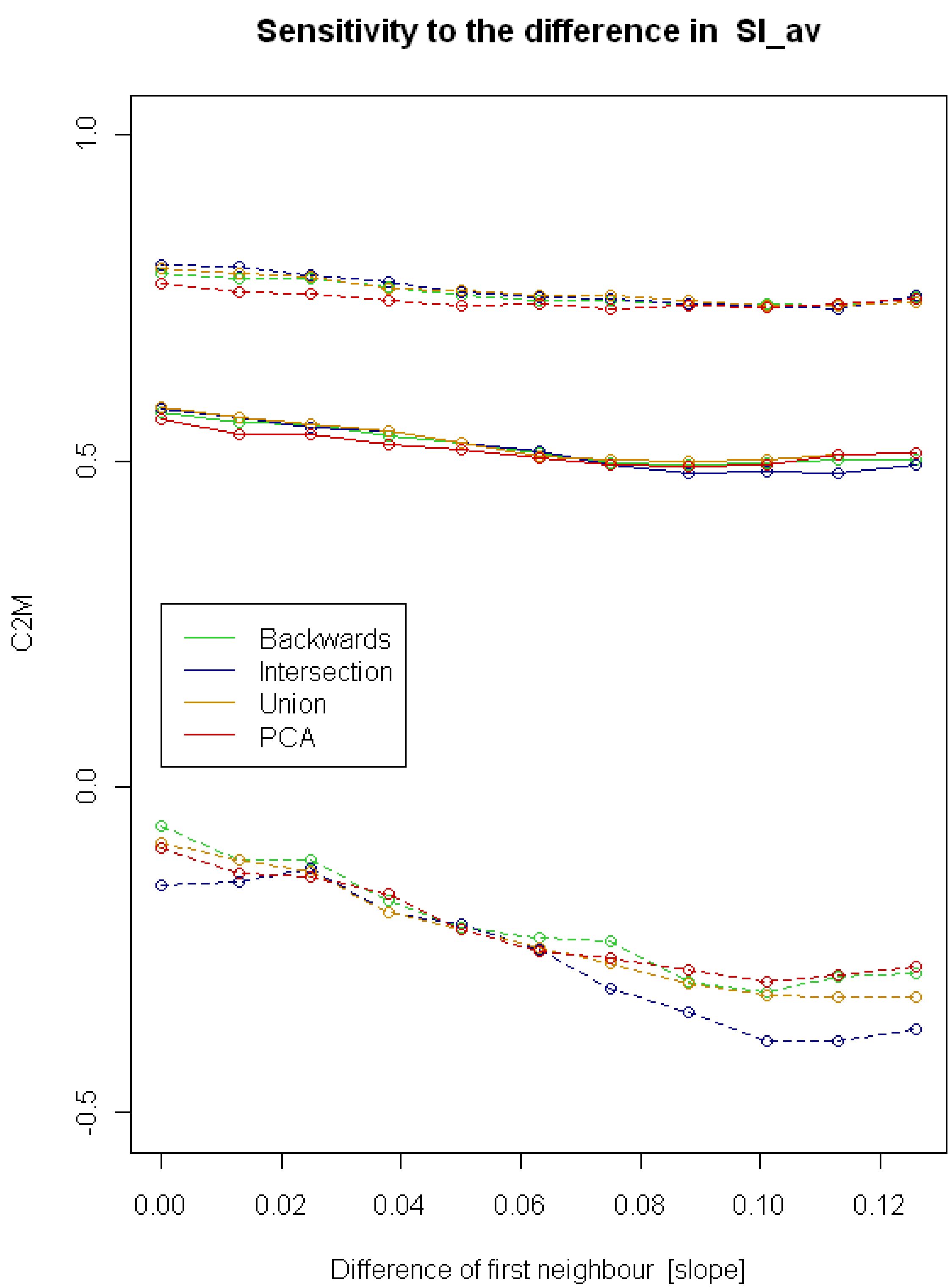
Sensitivity to the difference in Z_0.6
C2M
1.0
0.5
0.0
-0.5
0
100
200
300
400
Difference of first neighbour [m]
Backwards
Intersection
Union
PCA

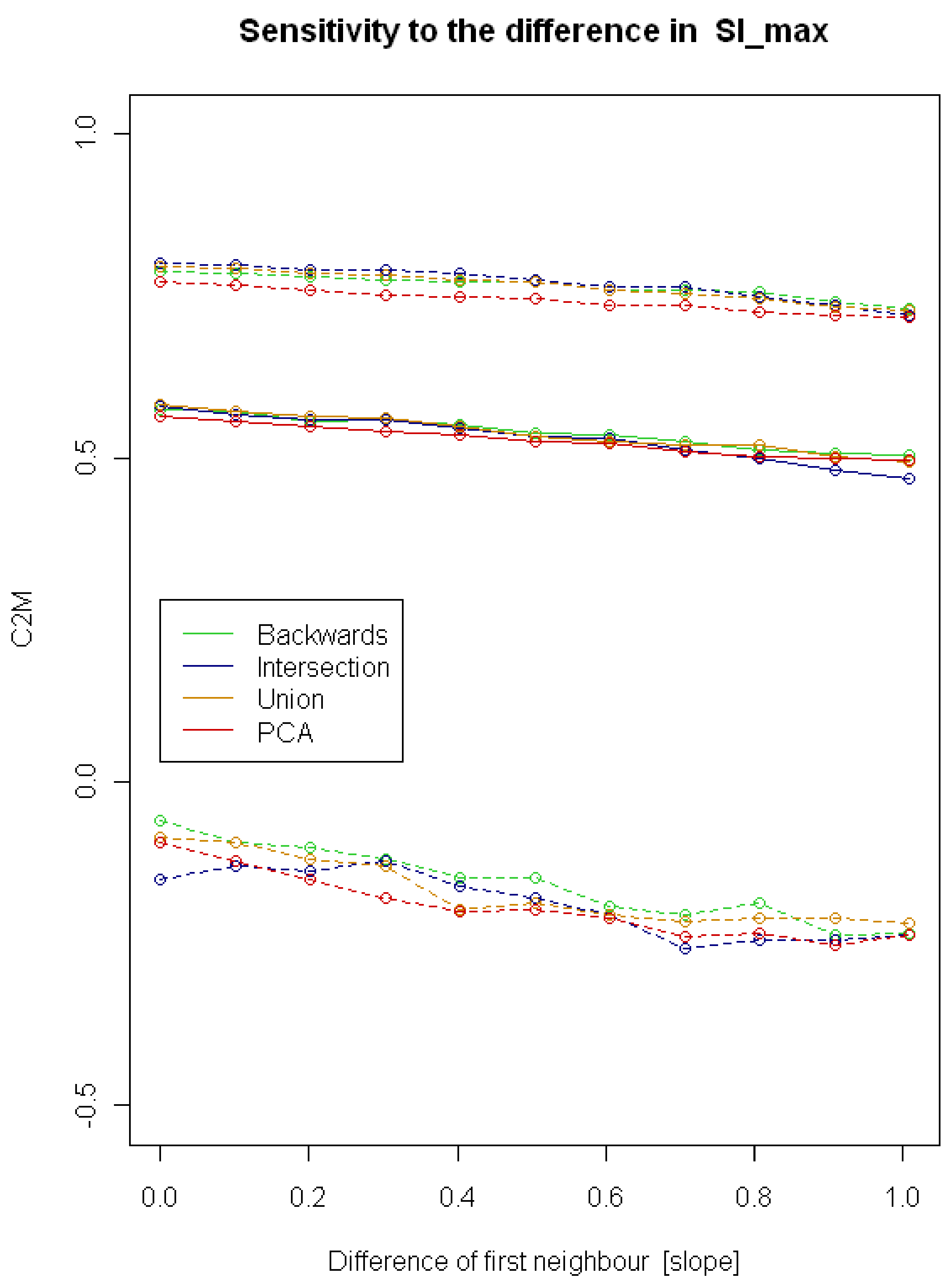

Sensitivity to the difference in Z_0.7
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0
100
200
300
400
Difference of first neighbour [m]

Sensitivity to the difference in Z_0.8

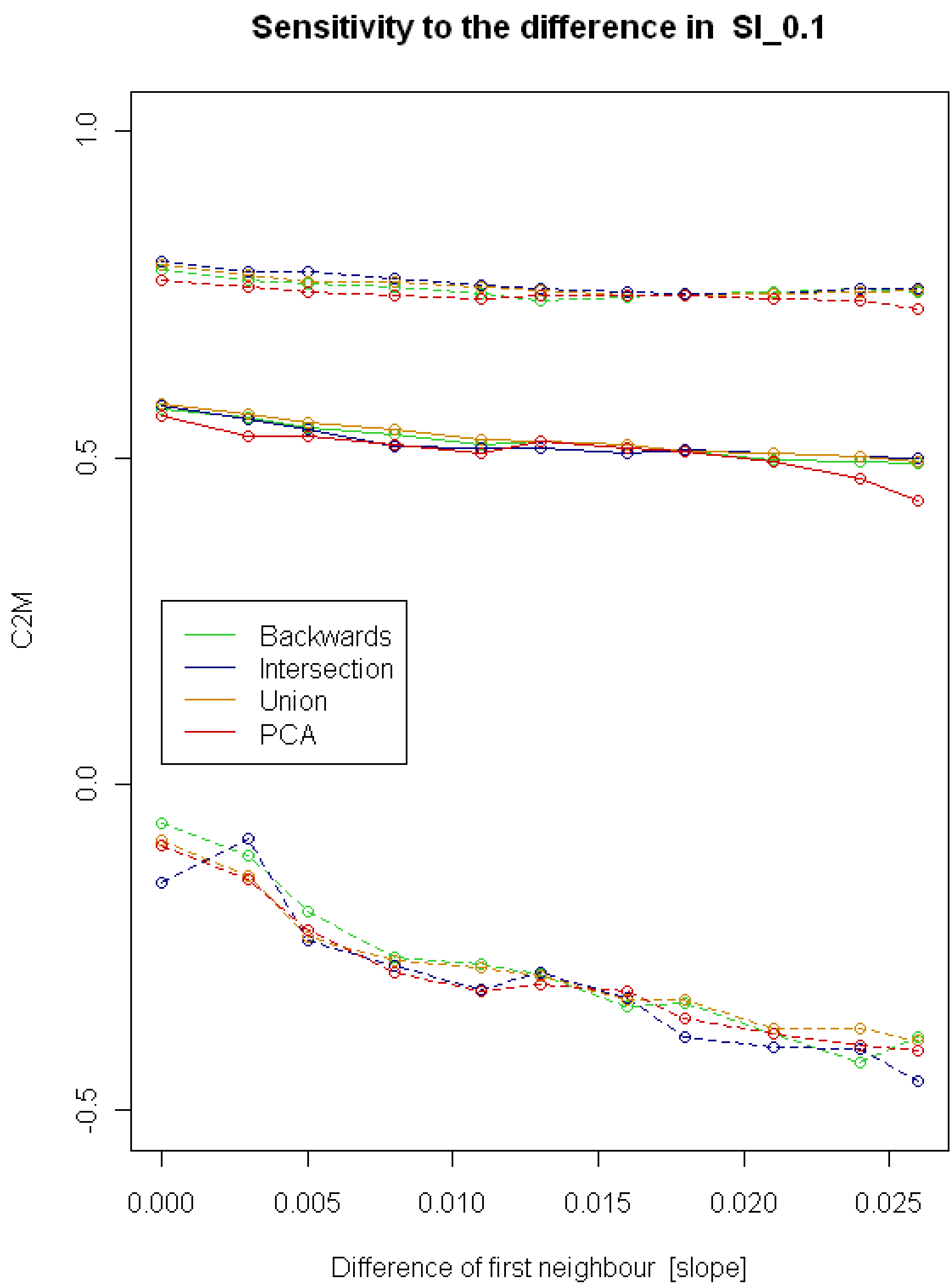



Sensitivity to the difference in Z_0.9
1.0
0.5
0.0
-0.5
C2M
Backwards
Intersection
Union
PCA
0
100
200
300
400
500
Difference of first neighbour [m]

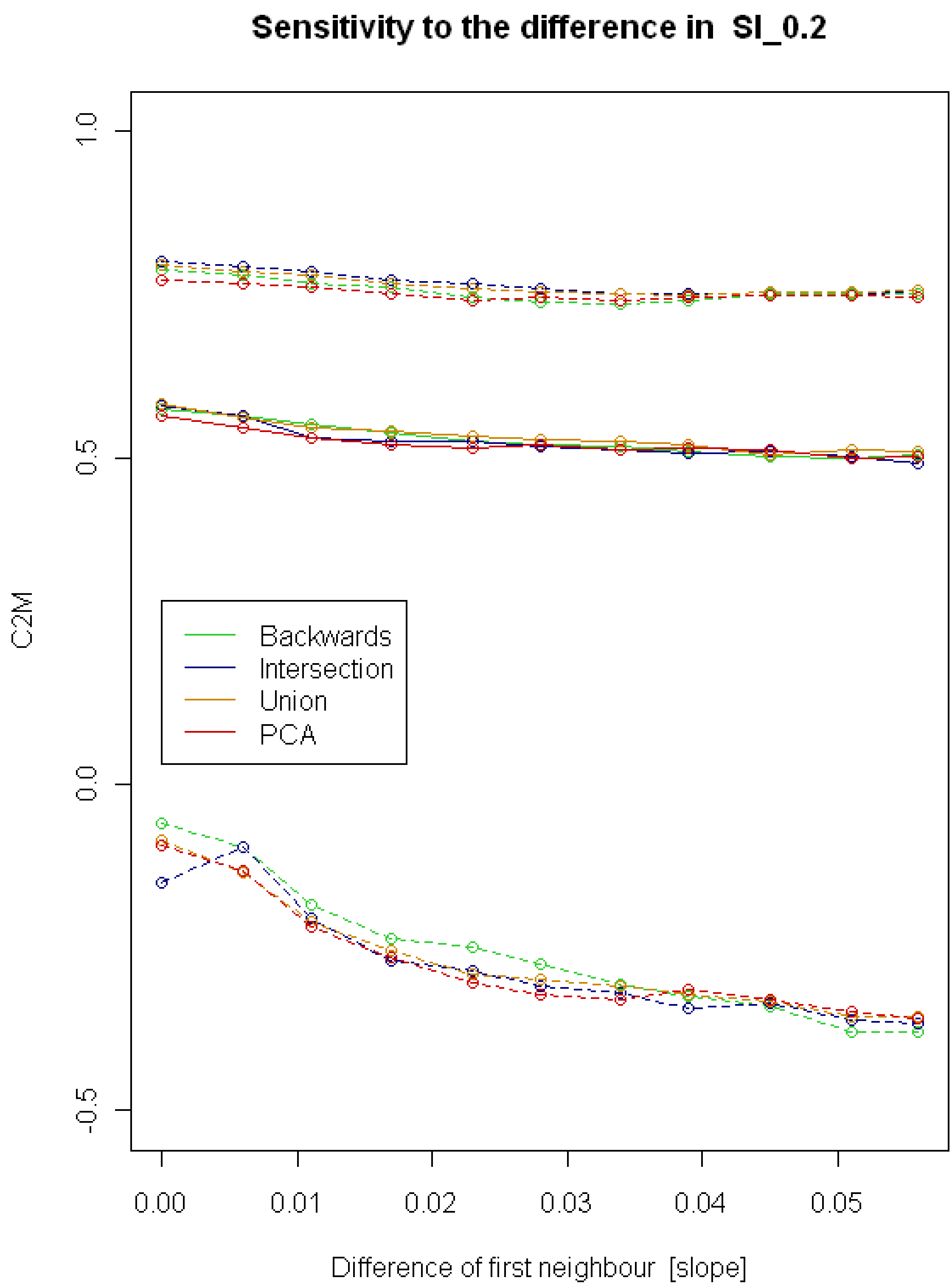

Sensitivity to the difference in SI_av
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Difference of first neighbour [slope]

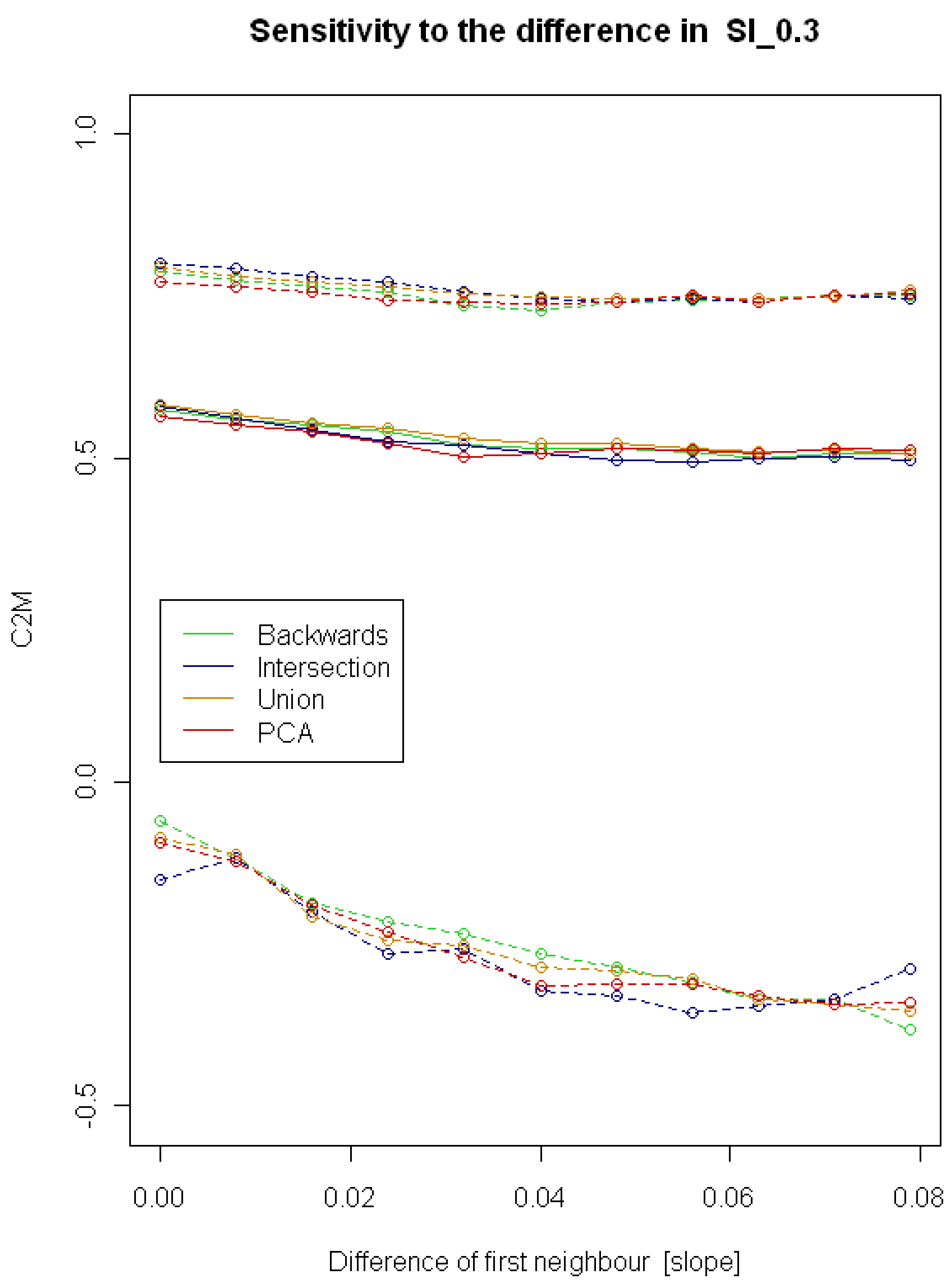


Sensitivity to the difference in SI_max
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.0
0.2
0.4
0.6
0.8
1.0
Difference of first neighbour [slope]

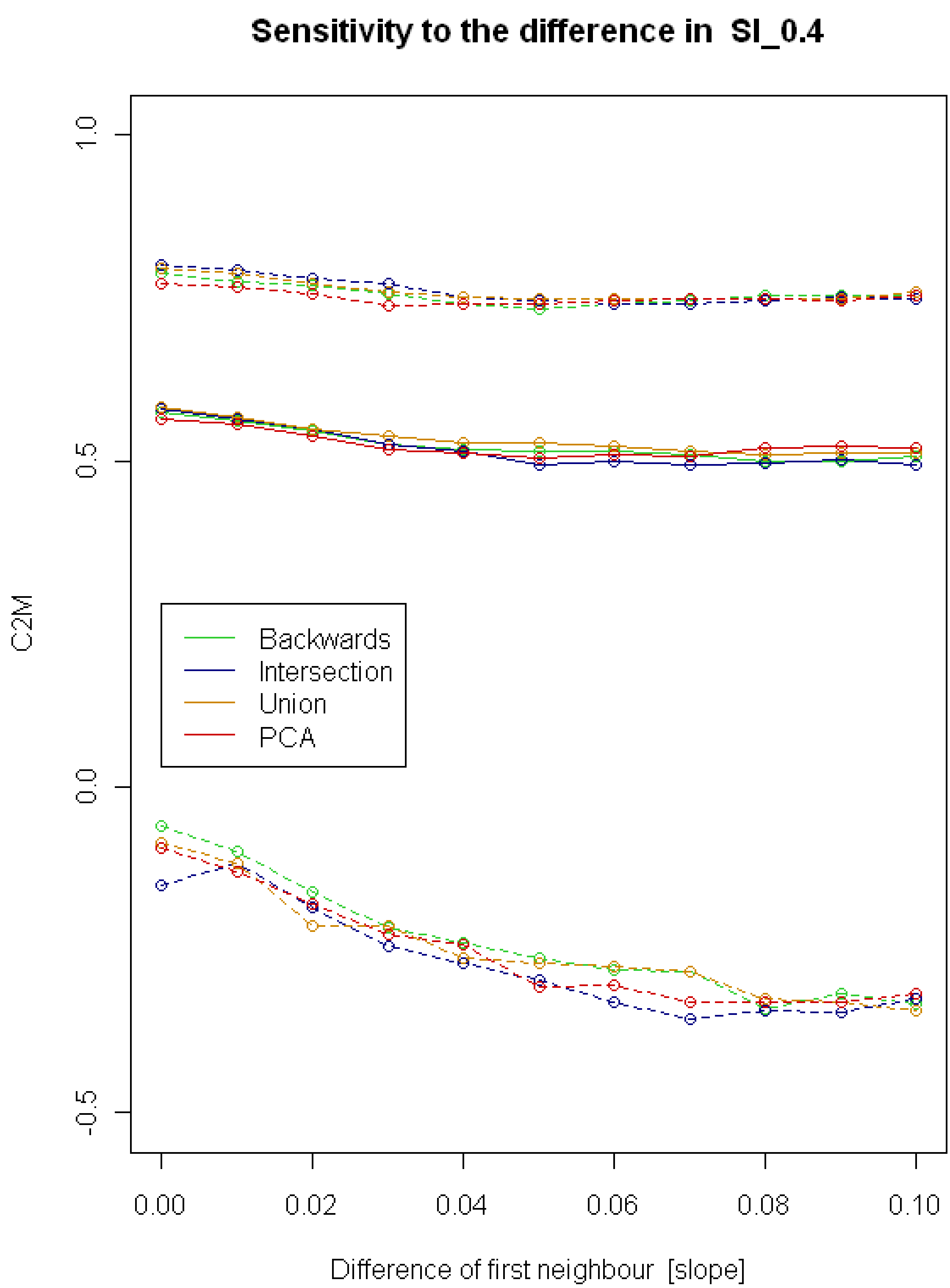


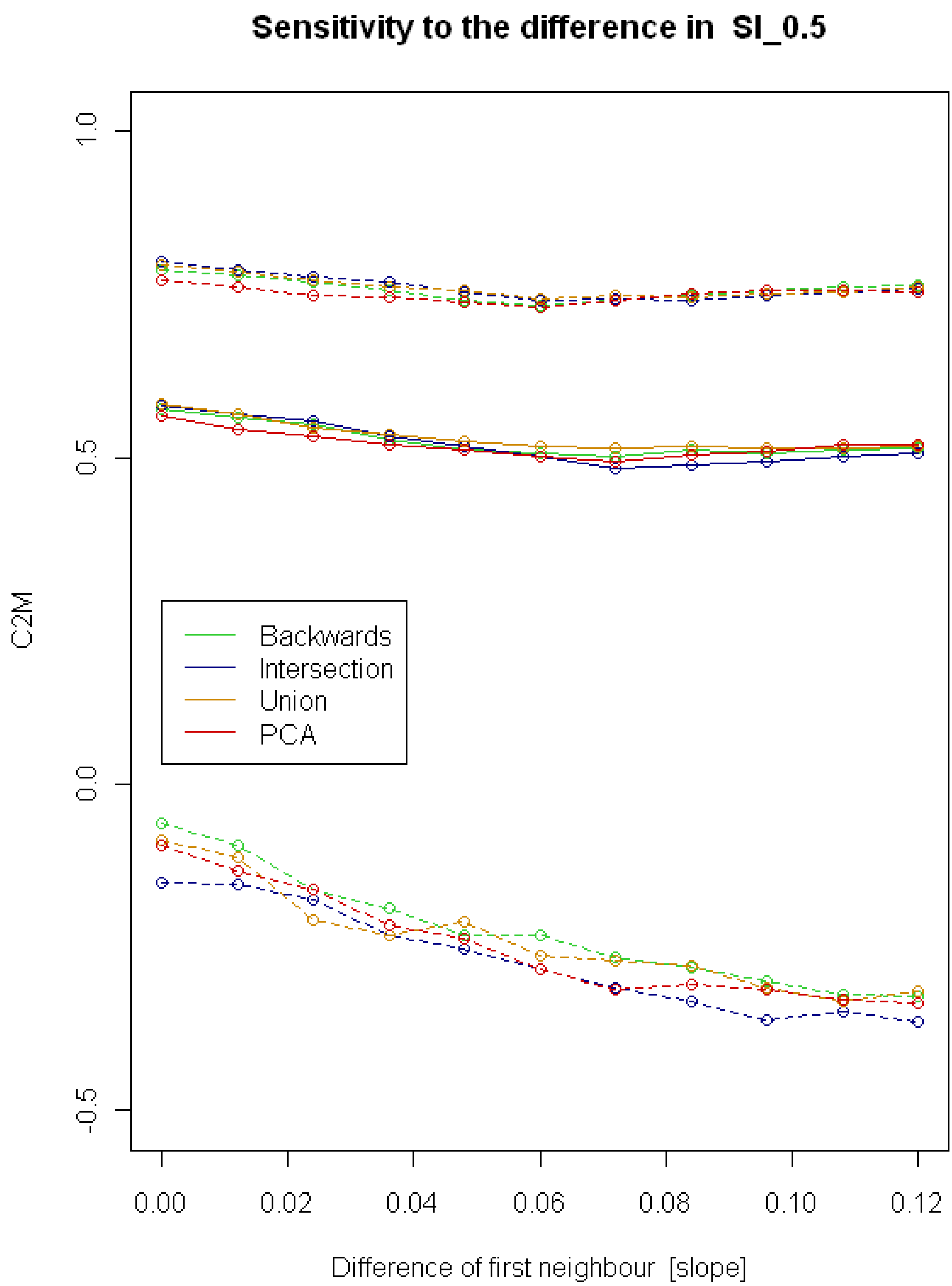
Sensitivity to the difference in SI_0.1
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.000
0.005
0.010
0.015
0.020
0.025
Difference of first neighbour [slope]


Sensitivity to the difference in SI_0.2
1.0
0.5
0.0
-0.5
C2M
Backwards
Intersection
Union
PCA
0.00
0.01
0.02
0.03
0.04
0.05
Difference of first neighbour [slope]

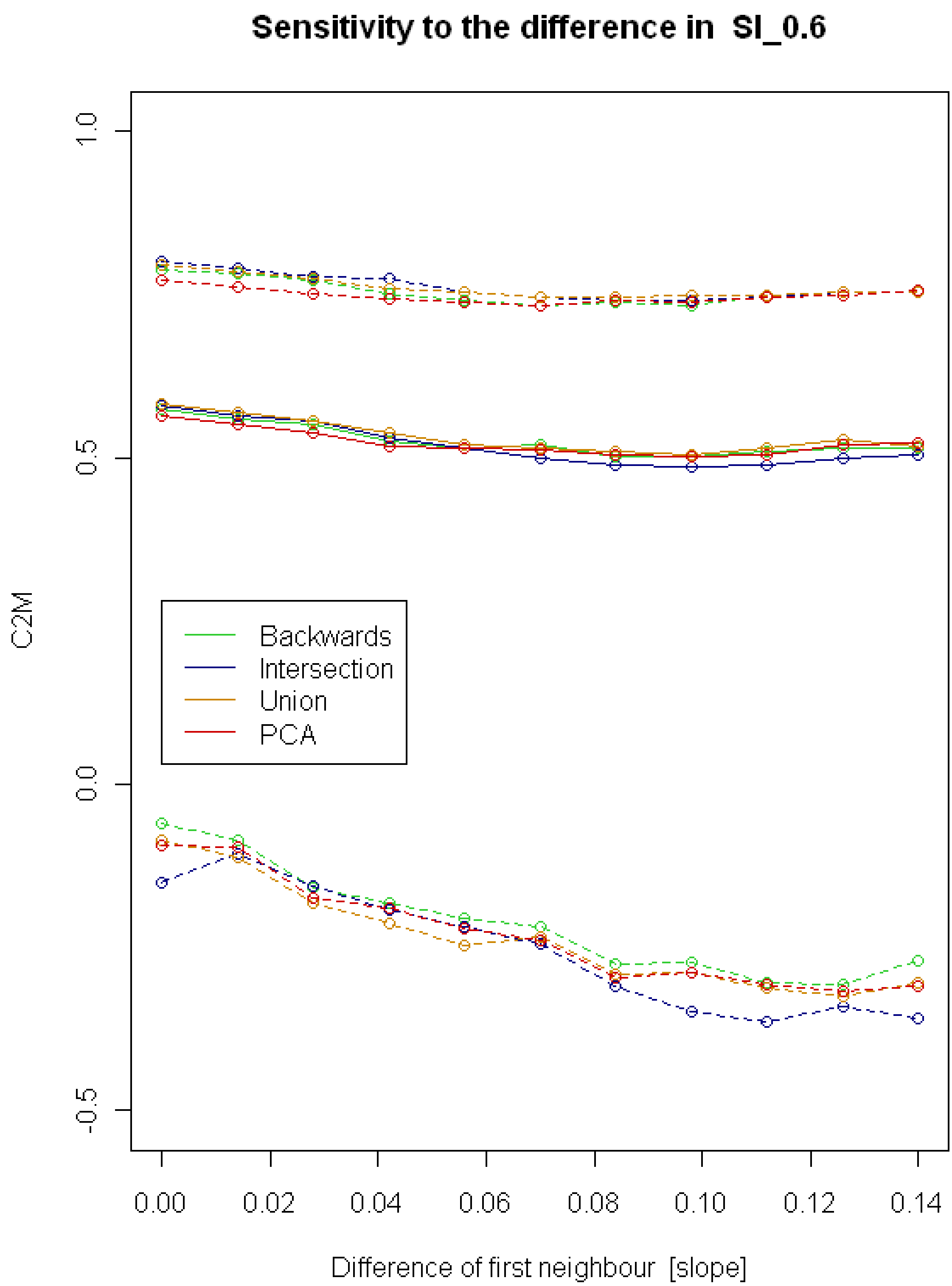


Sensitivity to the difference in SI_0.3
1.0
0.5
0.0
-0.5
C2M
Backwards
Intersection
Union
PCA
0.00
0.02
0.04
0.06
0.08
Difference of first neighbour [slope]

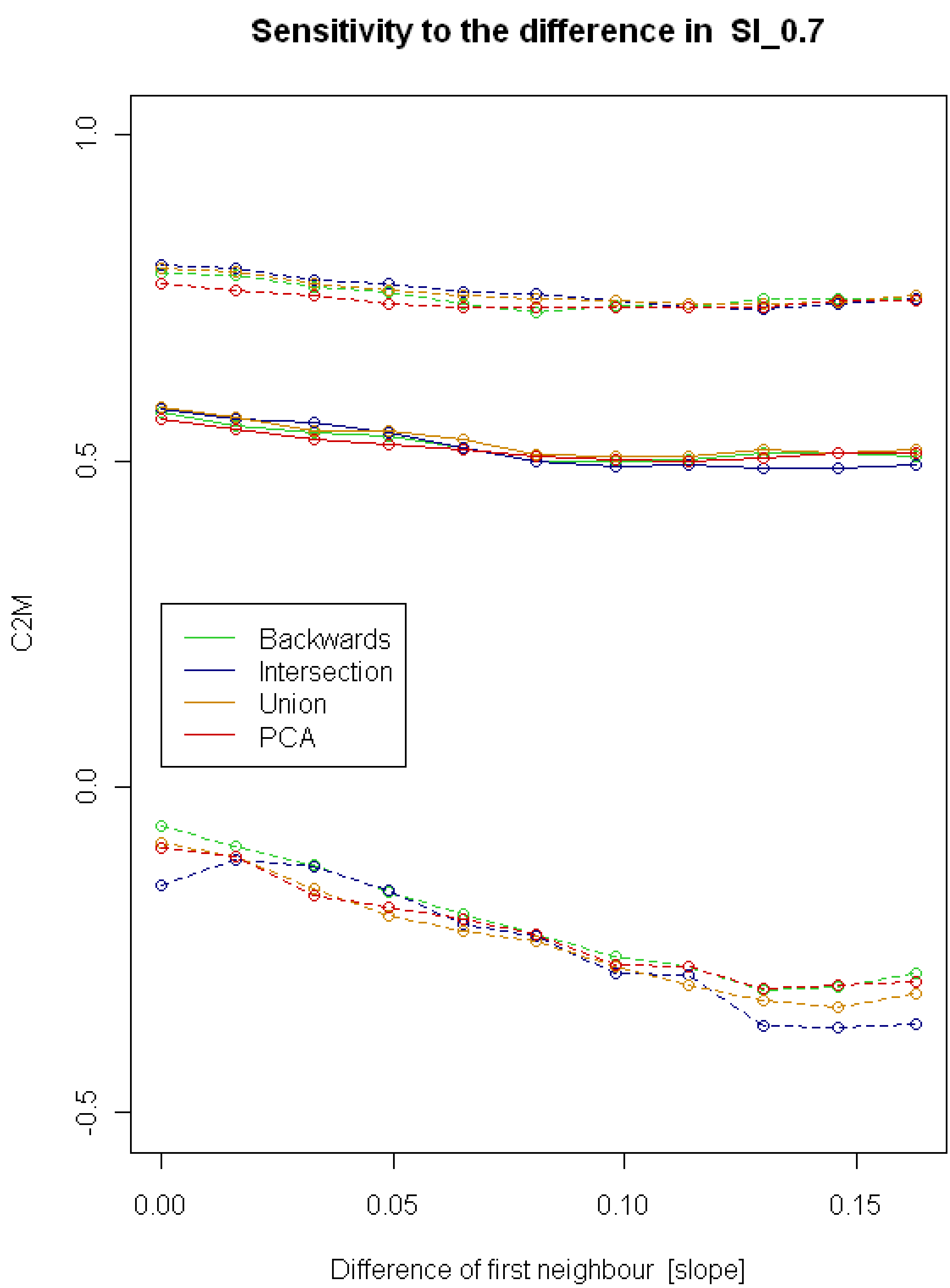

Sensitivity to the difference in SI_0.4
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.02
0.04
0.06
0.08
0.10
Difference of first neighbour [slope]

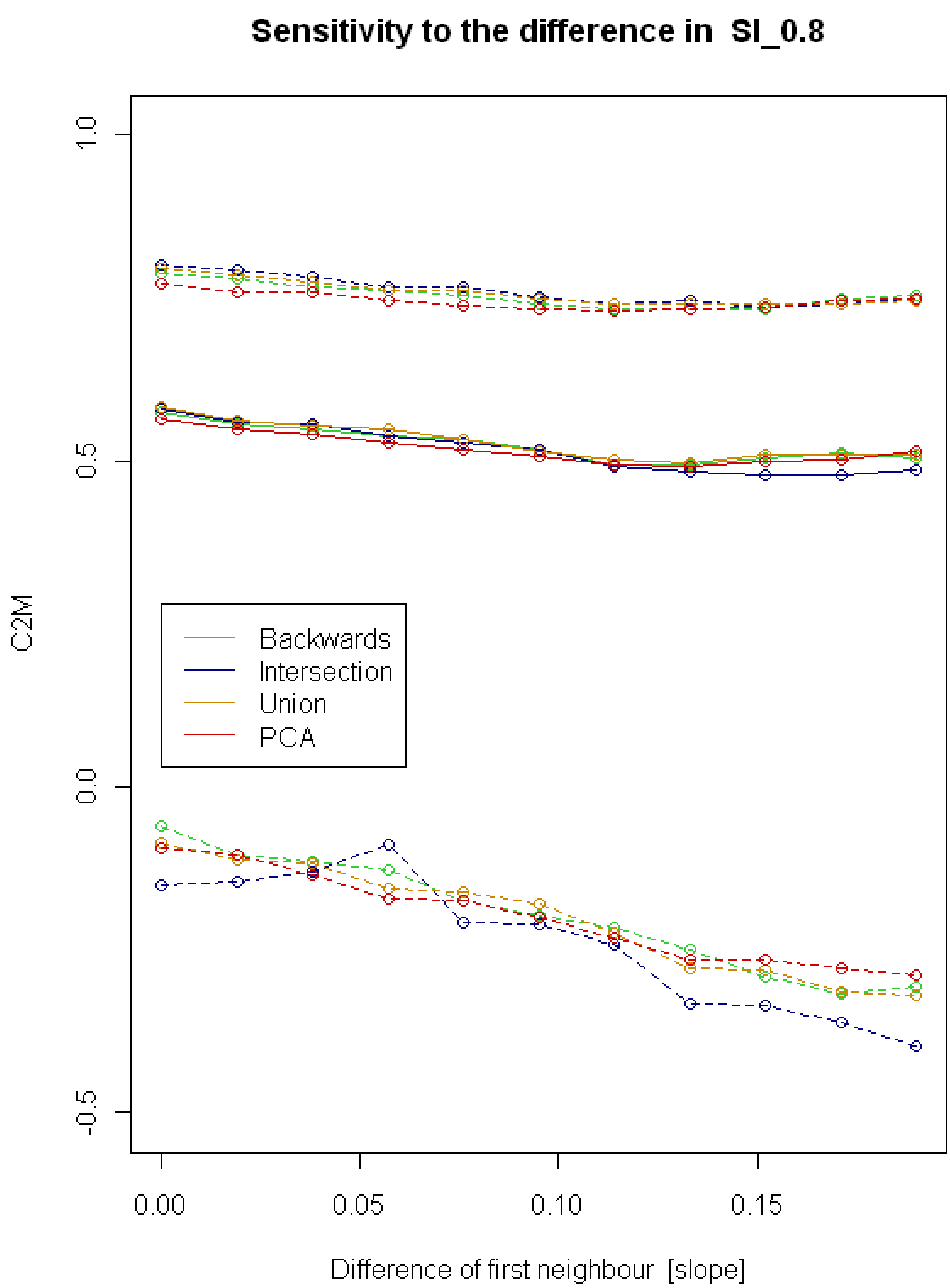


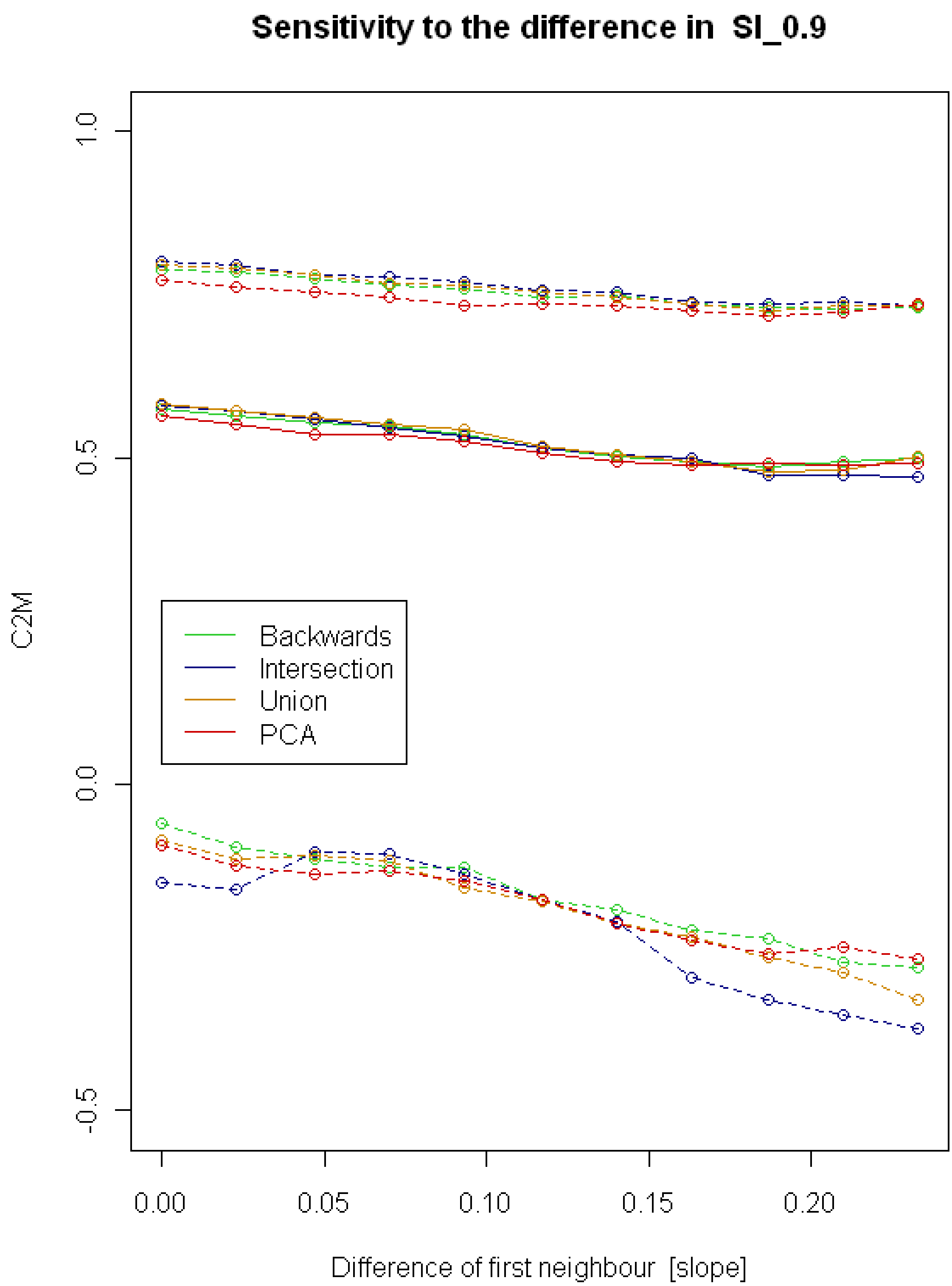
Sensitivity to the difference in SI_0.5
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Difference of first neighbour [slope]



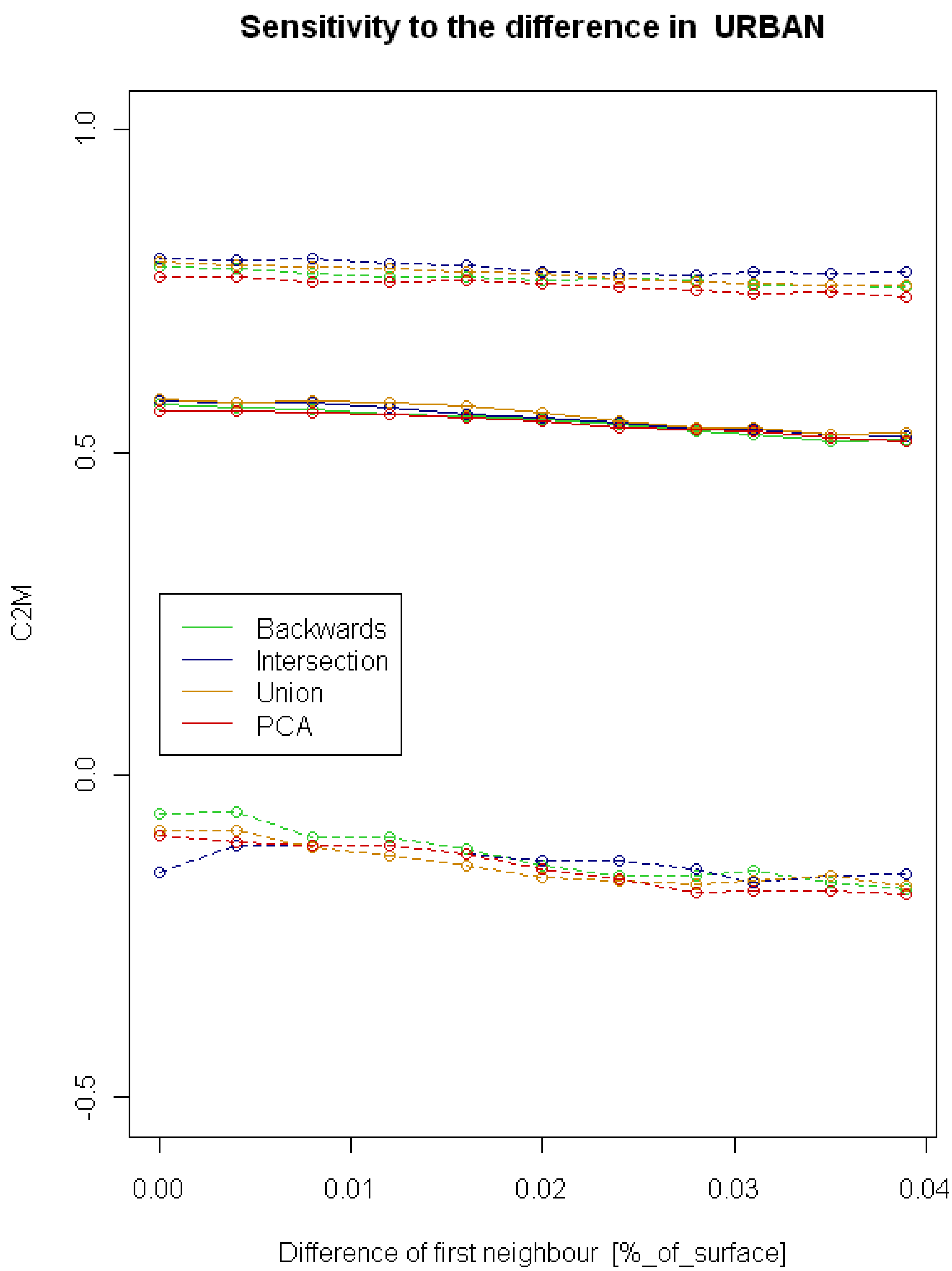
Sensitivity to the difference in SI_0.6
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Difference of first neighbour [slope]


Sensitivity to the difference in SI_0.7
1.0
0.5
0.0
-0.5
C2M
Backwards
Intersection
Union
PCA
0.00
0.05
0.10
0.15
Difference of first neighbour [slope]

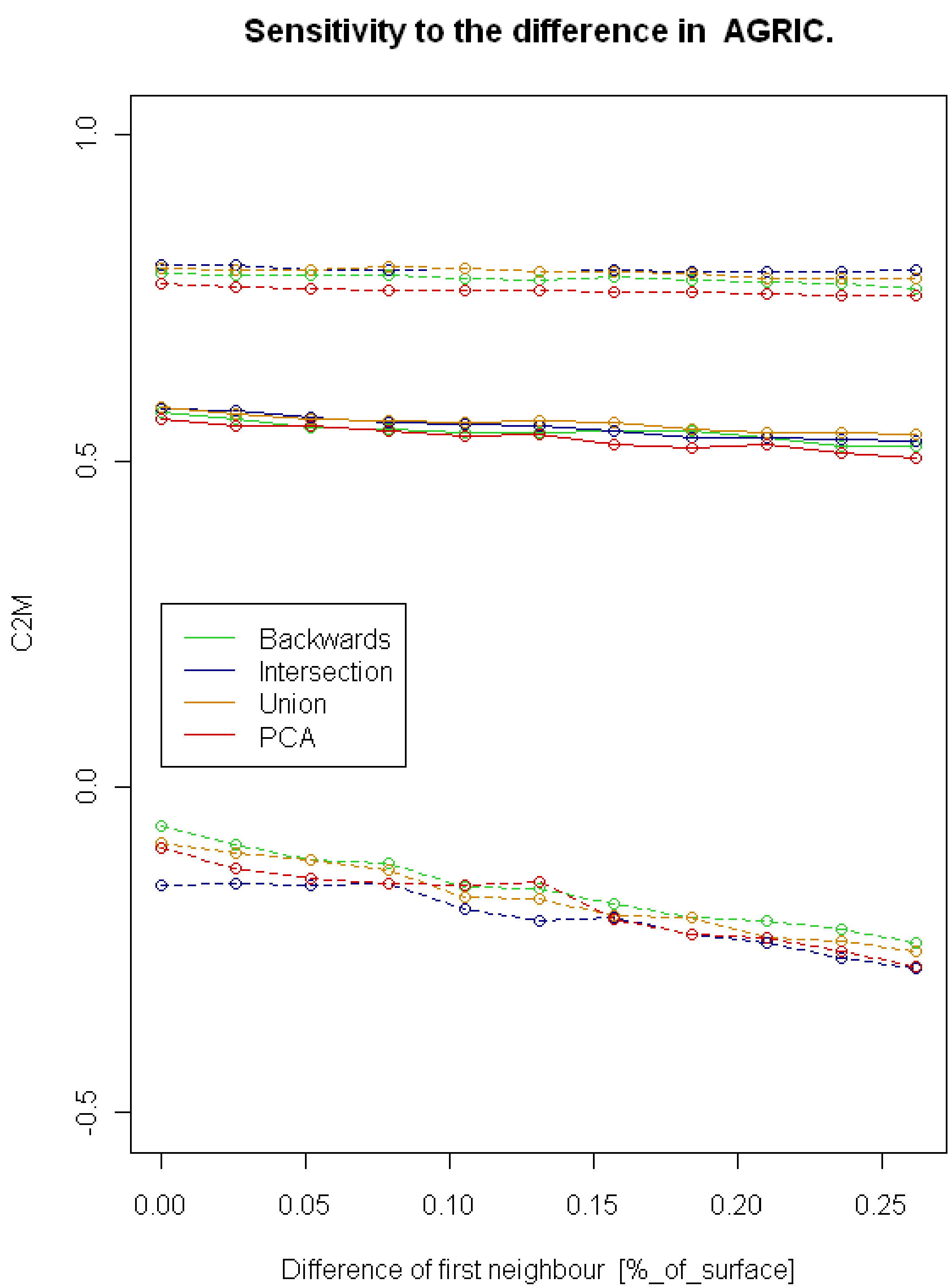

Sensitivity to the difference in SI_0.8
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.05
0.10
0.15
Difference of first neighbour [slope]



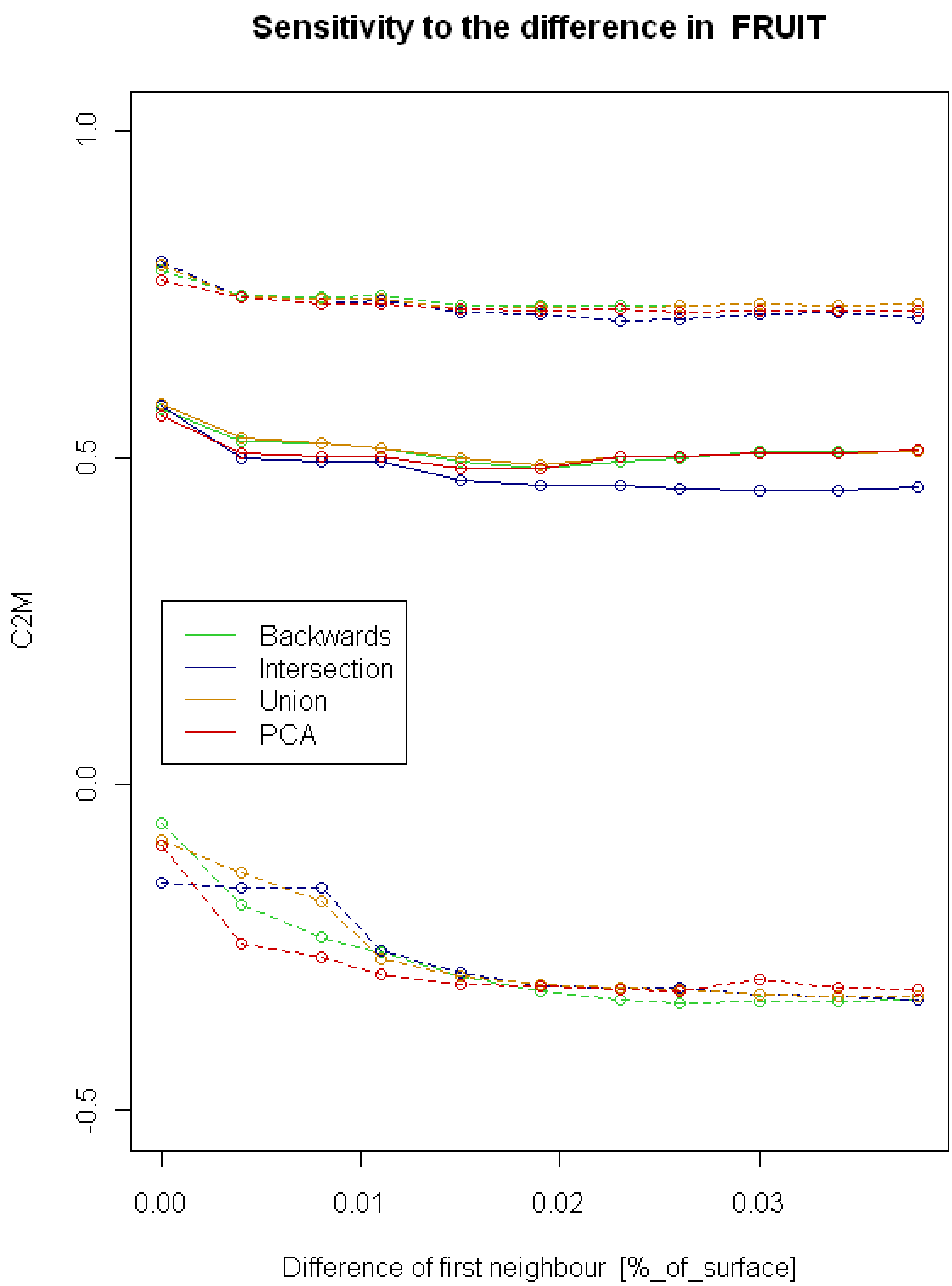
Sensitivity to the difference in SI_0.9
C2M
1.0
0.5
0.0
-0.5
0.00
0.05
0.10
0.15
0.20
Difference of first neighbour [slope]
Backwards
Intersection
Union
PCA



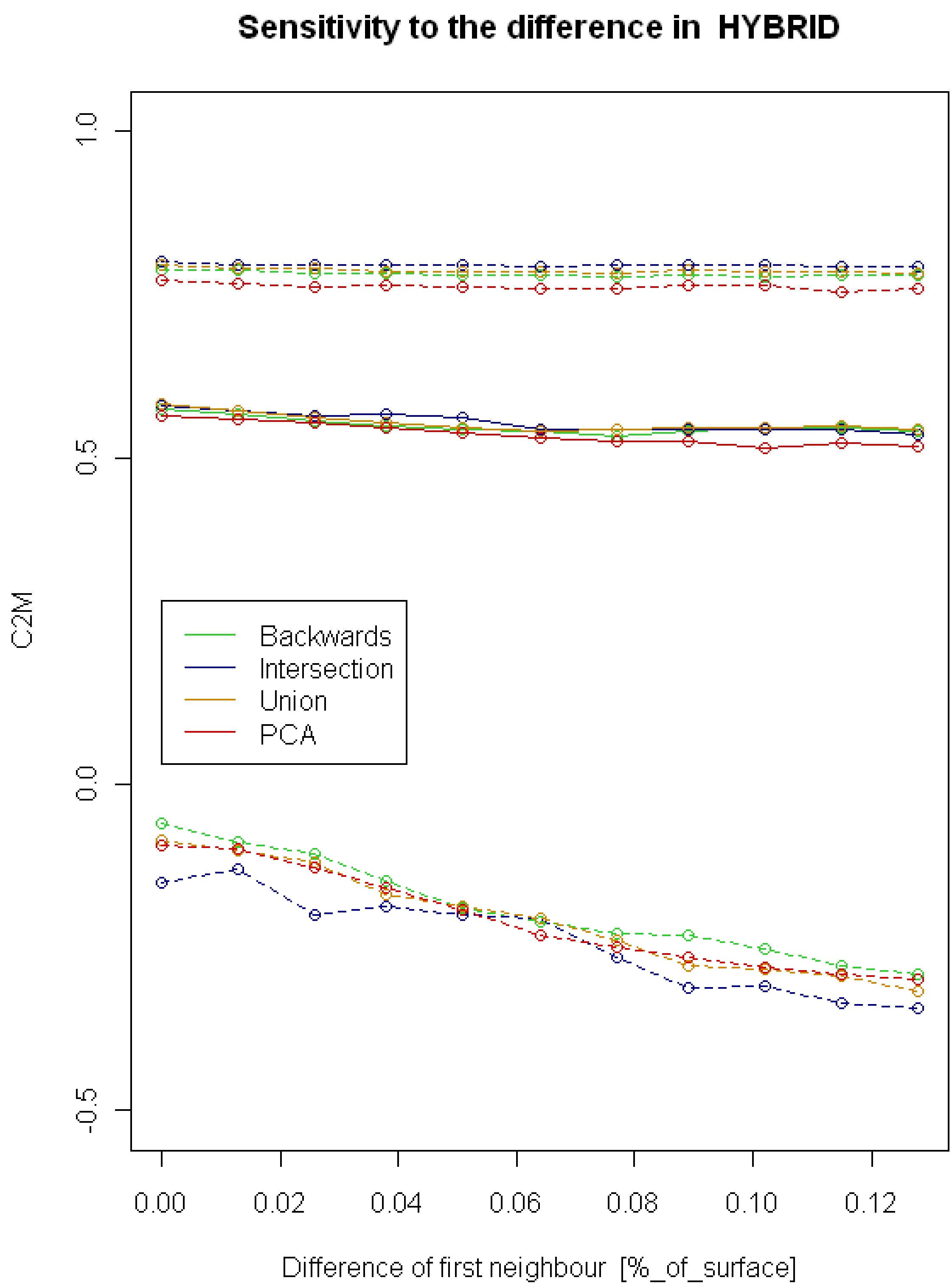
Sensitivity to the difference in URBAN
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.01
0.02
0.03
0.04
Difference of first neighbour [%_of_surface]

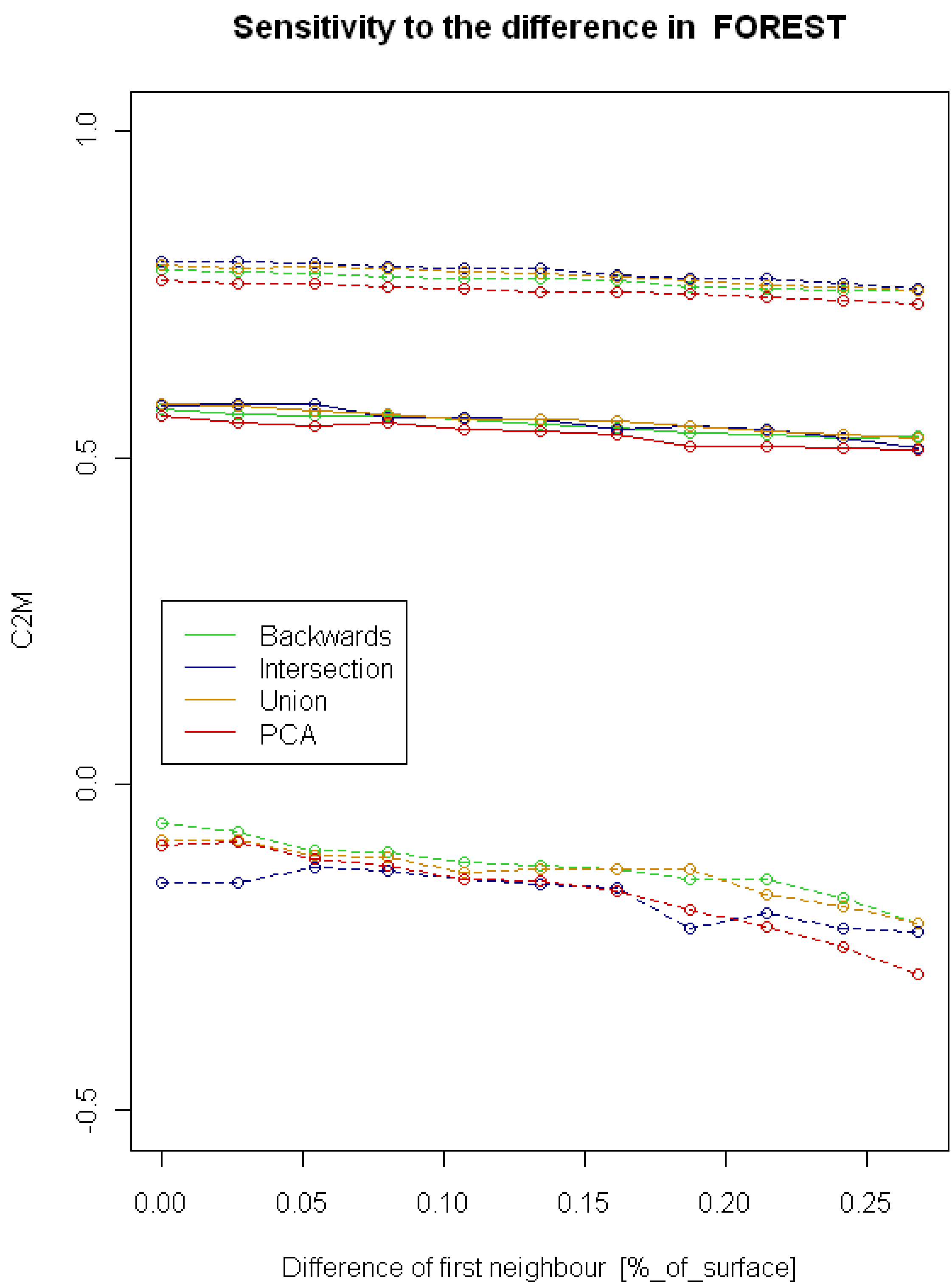


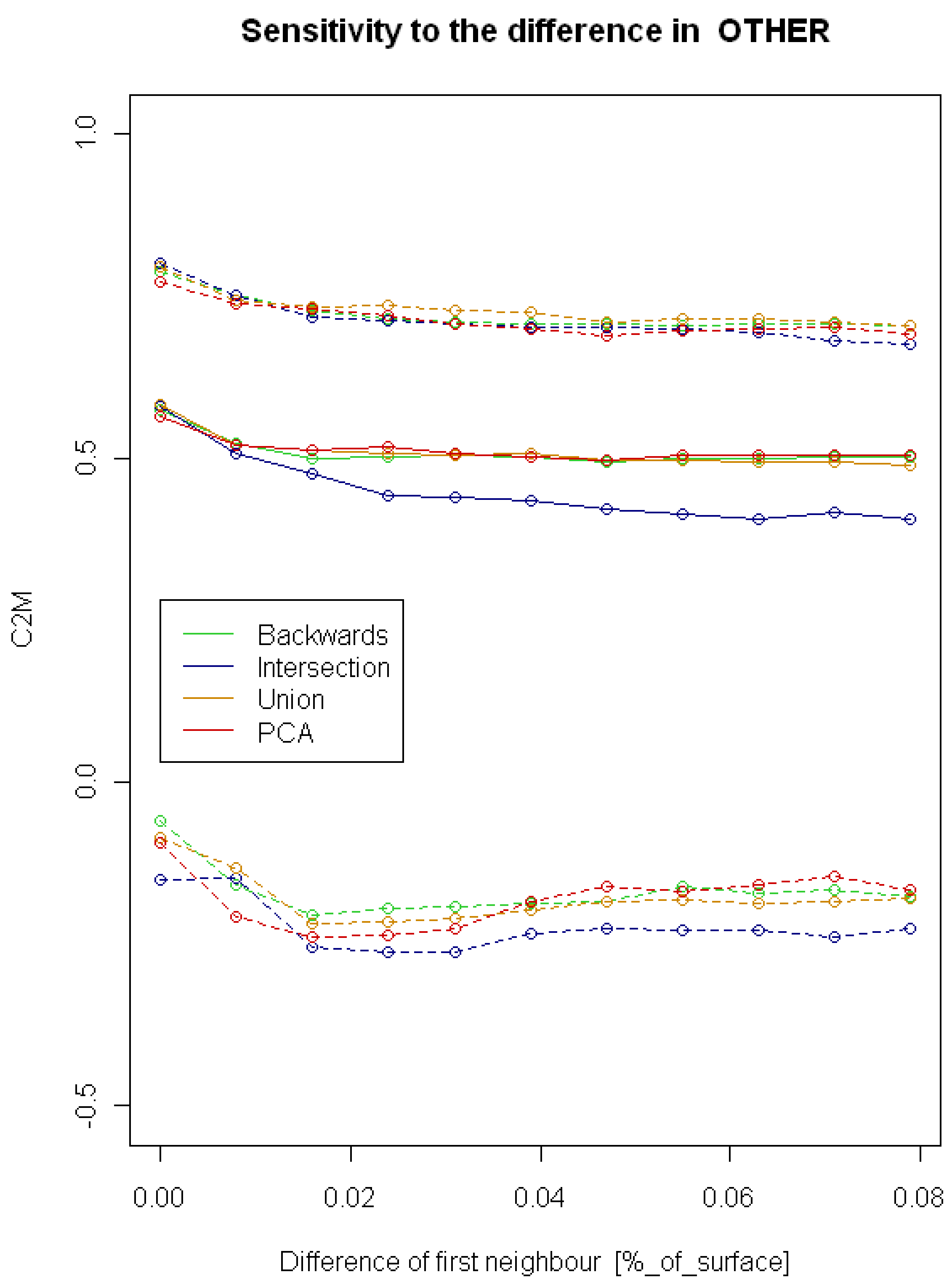
Sensitivity to the difference in AGRIC.
1.0
0.5
0.0
-0.5
C2M
Backwards
Intersection
Union
PCA
0.00
0.05
0.10
0.15
0.20
0.25
Difference of first neighbour [%_of_surface]

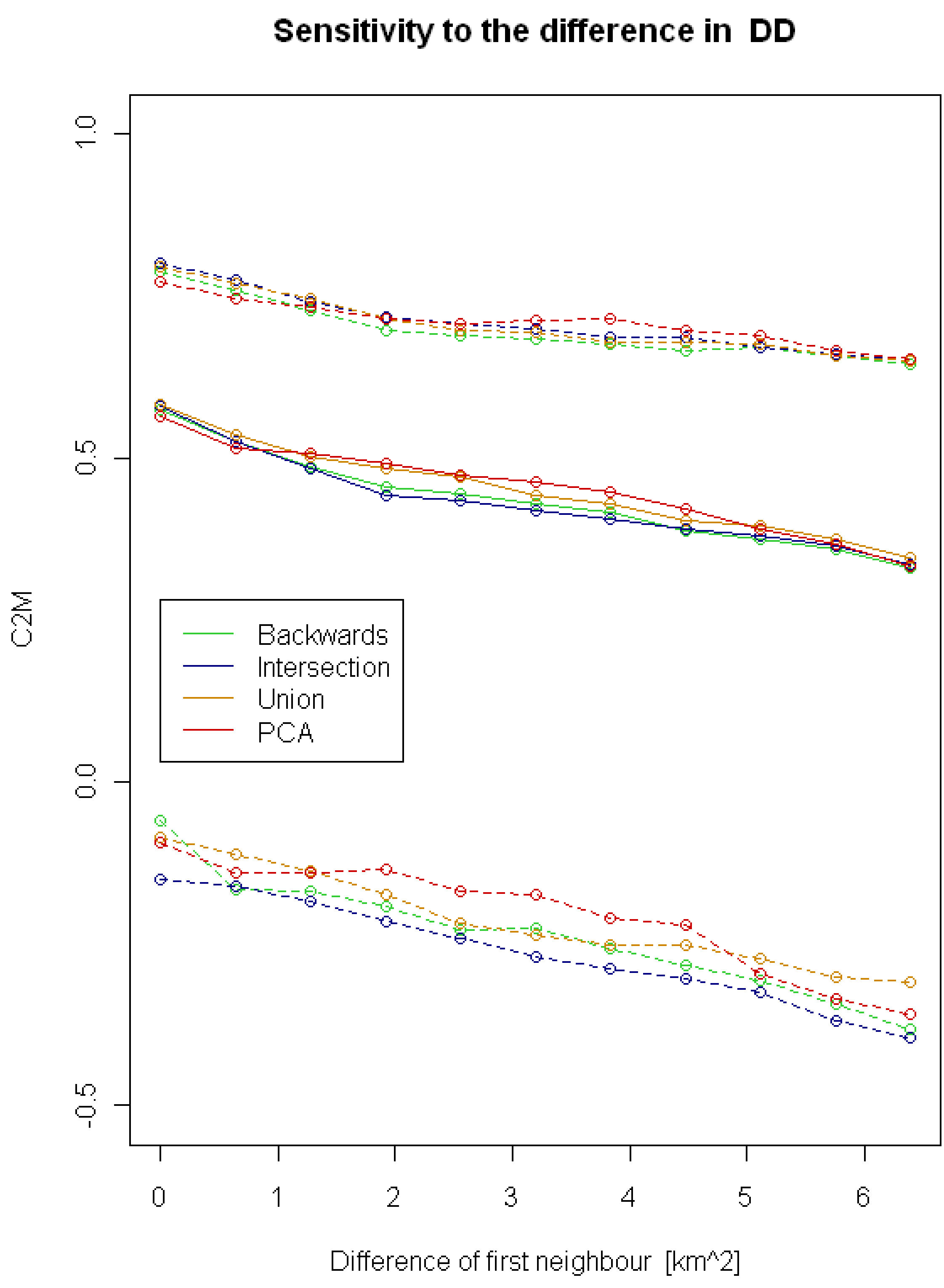


Sensitivity to the difference in FRUIT
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.01
0.02
0.03
Difference of first neighbour [%_of_surface]


Sensitivity to the difference in HYBRID
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Difference of first neighbour [%_of_surface]


Sensitivity to the difference in FOREST
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.05
0.10
0.15
0.20
0.25
Difference of first neighbour [%_of_surface]


Sensitivity to the difference in OTHER
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0.00
0.02
0.04
0.06
0.08
Difference of first neighbour [%_of_surface]


Sensitivity to the difference in DD
C2M
1.0
0.5
0.0
-0.5
Backwards
Intersection
Union
PCA
0
1
2
3
4
5
6
Difference of first neighbour [km^2]

15 GR4J

Figure 40 shows the structure of the GR4J model. GR4J is a two-storage, four-parameter, daily-step conceptual RR model that has proven to give good results on a wide range of French catchments.
The four parameters have the following functions:
-X1: depth of the routing store
-X2: depth of the production store
-X3: leaks and gains (sub-surface exchanges with neighboring catchments and/or deep aquifer systems)
-X4: unitary hydrograph UH1 base time

For more details on GR4J, see Perrin et al. (2003).

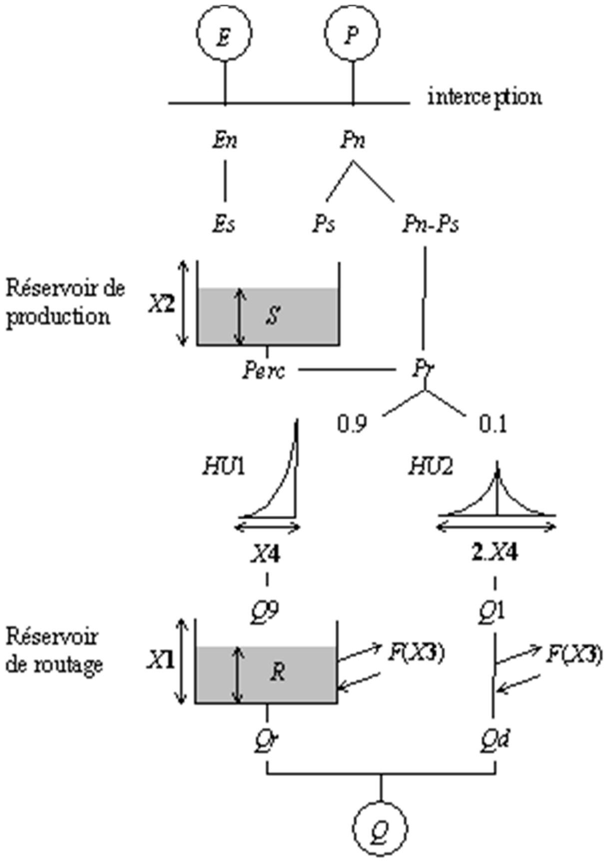


Figure 40: Scheme of the four-parameter GR4J model

Index

Acknowledgements **3**

Résumé **4**

Abstract **5**

1 Introduction **7**

Part 1 – Methods, databases and literature **11**

2 Databases used in this thesis **13**

2.1 Why should we use a large set of catchments? 14

2.2 How our dataset was made 14

2.3 How can we characterize a catchment? 15

2.4 Distribution of a few key descriptors and flow characteristics over our dataset 19

3 Methodological aspects **23**

3.1 General principles of the comparative testing of alternative regionalization methods 24

3.1.1 Jack-knife approach to cross validation 24

3.1.2 There is no absolute truth in this world: we need benchmarks 24

3.1.3 Specificities of different models: how the regionalization exercise differs for a statistical and for a rainfall-runoff model 25

3.2 Catchment selection: differential approach for the donor and the receiver pool 26

3.3 Further methodological requirements to assess the robustness of a regionalization method27

3.3.1 Why this question makes sense? 27

3.3.2 Assessing the impact of the density of neighbors: metrological desert generation vs random network reduction 28

3.4 Synthesis of the methodological choices 32

4 Literature review on the regionalization of rainfall-runoff models **33**

4.1 All agree more or less on a definition for an ungaged basin 34

4.2 For a hydrological fundamentalist, there is no special problem with ungaged basins… 34

4.3 With ungauged basins, the solution lies in "putting more physically measurable" parameters in the model in order to reduce (suppress?) the dependency on calibration 35

4.4 With ungauged basins, the solution lies in finding *a posteriori* a relationship between calibrated parameters and relevant physiographic and climatic descriptors (or geographical coordinates) 36

4.4.1 Absolute relationships 36

4.4.2 Relative relationships 37
4.5 With ungauged basins, the solution lies in finding one or more similar catchment(s) (in order to transfer parameters from them) 38
4.5.1 Methods focusing on spatial proximity 38
4.5.2 Methods focusing on physical similarity 39
4.5.3 How to define similarity? 41
4.5.4 Concerning possible complementarities between spatial proximity and physical similarity 42
4.6 With ungauged basins, the solution lies in using a previously made statistical regionalization to guide us in the choice of model parameters 43
4.7 My opinion (before I started this work), how it evolved, and how the solutions I tried to implement relate to the literature 45

Part 2 – Studies relative to flow statistics and their regionalization 47

5 Linking flow statistics to physiographic descriptors 49

5.1 Brief review of the literature on the regionalization of flow statistics 50
5.2 Regression as a conceptual model of the relationship between physiographic properties, climate and streamflow 52
5.2.1 Nation-wide vs local formulations 52
5.2.2 Selecting relevant descriptors 53
5.3 Streamflow statistics considered and results 54
5.3.1 Streamflow statistics considered 54
5.3.2 List of physiographic descriptors 54
5.3.3 Results 55
5.3.4 Review of the dependence of the selected statistics on each descriptor 59

6 Using neighbour catchments residuals to improve the efficiency of flow statistics regionalization 63

6.1 Residual's spatial structure as a descriptor of overlooked or not observable properties 64
6.1.1 IDW interpolation 64
6.1.2 Results 65
6.2 Constraints on the surface of donor catchments 67
6.3 Accounting for nested donor catchments 70
6.4 Excluding outliers from the donors' list 73
6.5 Final considerations on the results obtained for the regionalization of flow statistics 87

Part 3 – Regionalization of rainfall-runoff models – direct methods 89

7 Physiographic similarity regionalization 91

7.1 Introduction 92

7.1.1 Common points of the tested regionalization methods 92

7.2 Method based on Principal Component Analysis 93

7.2.1 Preliminary selection of explanatory variables 94

7.2.2 Principal Component Analysis as a tool to overcome the issue of correlated descriptors 95

7.2.3 Results 96

7.3 Backwards sorting method 99

7.3.1 Variable selection algorithm 99

7.3.2 Results 100

8 Joining spatial proximity and physiographic similarity 103

8.1 Introduction 104

8.2 An intersection-based method 105

8.2.1 description 105

8.2.2 Results 106

8.3 A union-based method 107

8.3.1 Description 107

8.3.2 Results 108

8.4 Comparison of the tested regionalization approaches 110

9 Sensitivity analysis of regionalization methods: how do they react to the lack of similar catchments? 113

9.1 Introduction 114

9.1.1 Results of the elimination neighboring donors 114

9.2 Sensitivity of regionalization methods to the lack of similar catchments 116

9.2.1 Results 116

9.3 Sensitivity of regionalization methods to thresholds of model efficiency 117

9.3.1 Results 117

Part 4 – Regionalization of rainfall-runoff models – the indirect path 121

10 Direct and indirect regionalization 123

10.1 Introduction 124

10.1.1 Why could an indirect regionalization be advantageous? 124

10.2 Review of the relevant scientific literature 125

10.2.1 How does the work presented in this chapter relate to the existing literature? 125

10.3 Issues of concern for implementing an indirect regionalization scheme 125
10.3.1 How does the first level of regionalization affect the second? 125
10.3.2 How to constrain the initial choice of possible parameter sets? 126
10.3.3 Can such a method be robust? 126
10.4 Method 127
10.4.1 General choices 127
10.4.2 Criterion used to further constrain the choice of parameter sets 129
10.4.3 Three benchmark comparisons. 129
10.5 Discussion of results 129
10.5.1 Number of parameter sets to be retained 130
10.5.2 Impact of statistics' regionalization quality on the following regionalization of RR model parameter sets 131
10.5.3 Could it be advantageous to constrain the choice of parameter sets with an additional criterion? 135
10.5.4 Robustness of the method: application of the metrological desert test 136
11 How the choice of an efficiency criterion impacts our vision of the 'best' regionalization method 139
11.1 What is the best regionalization method when we adopt an FDC-based performance criterion? 140
11.1.1 Performance criterion used for calibration 140
11.1.2 Regionalization results 140
11.2 Use of the Gupta et al. decomposition of NSE as diagnostic tool : where do lie the differences between C2M and the FDC-based criterion used in this chapter? 142
11.2.1 Detail of the NSE decomposition used 142
11.2.2 Difference in calibrated parameter sets 142
11.2.3 Difference in regionalized parameter sets 145
12 Conclusion 149
13 References 153
Part 6 – Appendices 157
14 Sensitivity to the elimination of similar donors: graphic results 157
15 GR4J 193
Index 195
List of Figures 200
List of Tables 204

List of Figures

Figure 1: Distribution densities of six physiographic distributors, compared with normal distributions (dashed lines). Note the semi-log scales for Area and Drainage Density (lognormal distributions were employed in these two cases) 20

Figure 2: Distributions of five flow characteristics. 21

Figure 3: Distribution of the distance of the closest neighbouring catchments over our dataset (distance calculated between catchments' centroids) 27

Figure 4: The catchments of our dataset are presented. Red-circled catchments do not have a neighboring basin closer than 20 km 29

Figure 5: Random density reduction (left) VS metrological desert generation (right). Blue represents the ungauged catchment, beige the authorized donor catchments, red the discarded donor catchments. In both examples, 20 donors have been discarded. 30

Figure 6: Impact of a progressive reduction of the number of donor catchments on the efficiency of a regionalization exercise (here showed: the median efficiency of a regionalized RR model). a- random reduction of the density of donors, b- creation of a metrological desert. 31

Figure 8: Scatterplots of empirical and regression-calculated values for flow statistics $Q_5$to Q_{50} 57

Figure 9: Scatterplots of empirical and regression-calculated values for flow statistics Q_{60} to Q_{95} 58

Figure 10: Scatterplots of empirical and regression-calculated values for flow statistics S_1, S_2, S_3 59

Figure 11: Map of the "hybrid" land cover class, expressed as fraction of the catchment's surface occupied by it. Most of the catchments that are rich with this land cover are climatically influenced by the Atlantic ocean. 62

Figure 12: Scatterplots of empirical and estimated flow statistics, with a regression model (left column) or with regression and IDW interpolation of the residuals (right column) 66

Figure 13: Simple IDW scatterplots (left) confronted with size-constraint method (right). Low to high: Q_{95}, Q_{50}, Q_5 69

Figure 14: Simple IDW scatterplots (left) confronted with nested donor weighting method (right). Low to high: Q_{95}, Q_{50}, Q_5 72

Figure 15: PCA-based Regionalization performances. Top left, median efficiency per number of donor catchments used. Top right, distribution of efficiencies compared to a random

selection of donors (dashed line) and calibrated model (solid grey line). Bottom left, performance in a "metrological desert" situation. 97
Figure 16: Time-series of observed, regionalized and simulated (with prior calibration) streamflows on three example catchments, of good (H301010) "median" (K2363010) and poor (H6402030) performances 98
Figure 17: Backwards-sorting Regionalization performances. Top left, median efficiency per number of donor catchments used. Top right, distribution of efficiencies compared to a random selection of donors (dashed line) and calibrated model (solid grey line). Bottom left, performance in a "metrological desert" situation. 101
Figure 18: Performances of spatial proximity and physiographic similarity methods (dashed grey and dashed black lines) confronted with the performance of an ideal method perfectly combining the strenghts of the two approaches (solid black line) 105
Figure 19: Combining spatial proximity and physical similarity, results of the intersection regionalization method. Top left, median efficiency per number of donor catchments used. Top right, distribution of efficiencies compared to a random selection of donors (dashed line) and calibrated model (solid grey line). Bottom left, performance in a "metrological desert" situation. 107
Figure 20: Combining spatial proximity and physical similarity, results of the union regionalization method. Left, distribution of efficiencies compared to a random selection of donors (dashed line) and calibrated model (solid grey line). Right, performance in a "metrological desert" situation. 109
Figure 21: Distribution of the performances of the tested direct regionalizations, compared to two benchmarks: random donor selection (dotted grey line), calibrated model (solid grey line) 110
Figure 22: Comparison of the performances of the tested direct regionalizations under the "metrological desert" robustness test 111
Figure 23: Performances of several regionalization approaches in a "metrological desert" situation. Upper dashed line: 0.9 quantile of the performance distribution. Continuous line: median. Lower dashed line: 0.1 quantile 115
Figure 24: Sensitivity of several regionalization approaches to the lack of well-modeled donors. Upper dashed line: 0.9 quantile of the performance distribution. Continuous line: median. Lower dashed line: 0.1 quantile 119
Figure 25: Sensitivity of several regionalization approaches to the exclusion of badly modeled donors. Upper dashed line: 0.9 quantile of the performance distribution. Continuous line: median. Lower dashed line: 0.1 quantile 120

Figure 26: Optimal number of donors for an indirect regionalization scheme. Black line: statistics regionalized with a regression and an IDW interpolation of the residuals. Grey dashed line: statistics regionalized using a regression. Grey dotted line: "true" statistics (cheat) 130
Figure 27: Performance of the proposed "plain regression approach". The black line presents the results obtained with flow statistics obtained through a "plain regression approach", confronted with three benchmark comparisons: random donor (dotted grey), spatial proximity (dashed grey) and calibrated model (solid grey) 132
Figure 28: Performance of the proposed "regression + residuals interpolation". The black line presents the results obtained with flow statistics obtained through a regression approach combined with an IDW-based interpolation of residuals. It is confronted with three benchmark comparisons: random donor (dotted grey), spatial proximity (dashed grey) and calibrated model (solid grey) 132
Figure 29: Performance of an ideal case where the flow statistics could be regionalized with no errors (black dashed line). It is confronted with three benchmark comparisons: random donor (dotted grey), spatial proximity (dashed grey) and calibrated model (solid grey) 134
Figure 30: Performance of an ideal case where the flow statistics could be regionalized with no errors (black solid line). It is confronted with two benchmark comparisons: calibrated model (grey solid line) and a “cheating” method that selects a-posteriori the best possible donor among the catchments available in the database (grey dashed line). 135
Figure 31 Performance of the proposed approach when selecting 5 parameter sets out of the the first 10 neighboring catchments (solid black line). It is confronted with three benchmark comparisons: random donor (dotted grey), spatial proximity (dashed grey) and calibrated model (solid grey) 136
Figure 32: "Metrological desert" test. The median efficiency of an indirect regionalization method using regression-estimated flow statistics (black line) is confronted with the optimal physiographic-similarity method identified in Part 3 (grey line). 137
Figure 33: Performance distribution of regionalization results according to an FDC-based criterion. Continuous grey line: calibration performance. Dashed grey line: random regionalization........ 141
Figure 34: distribution of correlations for C2M-calibrated and FDC-calibrated parameters 143
Figure 35: distribution of bias for C2M-calibrated and FDC-calibrated parameters........ 144
Figure 36: distribution of relative variability (alpha) for C2M-calibrated and FDC-calibrated parameters 145

Figure 37: distribution of correlations for direct and indirect regionalization methods 146
Figure 38: distribution of bias for direct and indirect regionalization methods 147
Figure 39: distribution of relative variability for direct and indirect regionalization methods 148

List of Tables

Table 1: Essential characteristics of the 865 catchment data set 14

Table 2: List of catchment descriptors available for this study 16

Table 3: Matrix of correlations between descriptors (p<0.05 significance in bold characters) 18

Table 4: coefficient of determination and RMSE for the regressions between flow statistics and catchment descriptors (calculated on log-transformed values). Av_Q stands for average annual runoff. 55

Table 5: Rankings of the significance of available descriptors for each flow statistic (threshold at p=0.05). 60

Table 6: Regression coefficients for each descriptor and flow statistic.................................. 61

Table 7: Comparison in the results between regression-estimated statistics and regression with IDW interpolation of the residuals.. 65

Table 8: RMSE on log-values for simple IDW and IDW with area-ratio constraint 68

Table 9: RMSE on log-values of flow statistics when using simple IDW or IDW giving more weight to nested catchments. The third column shows the esponent "a" presented at point c) 71

Table 10: Average efficiencies obtained when using only one physiographic descriptor to define site-similarity 95

Table 11: List of discarded descriptors at each iteration 100