



**HAL**  
open science

## Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency

Vazken Andréassian, François Bourgin, Ludovic Oudin, Thibault Mathevet,  
Charles Perrin, Julien Lerat, Laurent Coron, Lionel Berthet

► **To cite this version:**

Vazken Andréassian, François Bourgin, Ludovic Oudin, Thibault Mathevet, Charles Perrin, et al.. Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency. *Water Resources Research*, 2014, 50 (10), pp.8356-8366. 10.1002/2013WR014761 . hal-01196447

**HAL Id: hal-01196447**

**<https://hal.science/hal-01196447v1>**

Submitted on 16 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



### RESEARCH ARTICLE

10.1002/2013WR014761

#### Key Points:

- We produce a generalist list of parameter sets
- Short-list calibration is evaluated on an independent catchment data set
- With short calibration series, the generalist parameter sets give better results

#### Supporting Information:

- Supplementary Appendix 1

#### Correspondence to:

V. Andréassian,  
vazken.andreassian@irstea.fr

#### Citation:

Andréassian, V., F. Bourgin, L. Oudin, T. Mathevet, C. Perrin, J. Lerat, L. Coron, and L. Berthet (2014), Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency, *Water Resour. Res.*, *50*, 8356–8366, doi:10.1002/2013WR014761.

Received 23 SEP 2013

Accepted 25 SEP 2014

Accepted article online 30 SEP 2014

Published online 31 OCT 2014

## Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency

Vazken Andréassian<sup>1</sup>, François Bourgin<sup>1</sup>, Ludovic Oudin<sup>2</sup>, Thibault Mathevet<sup>3</sup>, Charles Perrin<sup>1</sup>, Julien Lerat<sup>4</sup>, Laurent Coron<sup>1</sup>, and Lionel Berthet<sup>5</sup>

<sup>1</sup>Irstea, UR Hydrosystèmes et Bioprocédés, Antony, France, <sup>2</sup>Université Pierre et Marie Curie, UMR Sisyphe, Paris, France, <sup>3</sup>EDF, DTG, Grenoble, France, <sup>4</sup>Bureau of Meteorology, Canberra, Australian Capital Territory, Australia, <sup>5</sup>DREAL Auvergne, Clermont-Ferrand, France

**Abstract** This paper evaluates the use of a small number of *generalist* parameter sets as an alternative to classical calibration. Here parameter sets are considered generalist when they yield acceptable performance on a large number of catchments. We tested the genericity of an initial collection of  $10^6$  parameter sets sampled in the parameter space for the four-parameter GR4J rainfall-runoff model. A short list of 27 generalist parameter sets was obtained as a good compromise between model efficiency and length of the short list. A different data set was used for an independent evaluation of a calibration procedure, in which the search for an optimum parameter set is only allowed within this short list. In validation mode, the performance obtained is inferior to that of a classical calibration, but when the amount of data available for calibration is reduced, the generalist parameter sets become progressively more competitive, with better results for calibration series shorter than 1 year.

### 1. Introduction

#### 1.1. The Ideal of a Calibration-Free World

Most hydrological models can only be used after their free parameters have been calibrated against observed data. Although most hydrologists have accepted this fact, it would be much more satisfying—from a scientific point of view—to have model parameters univocally derived from physical catchment characteristics. We would then live in a calibration-free world.

Instead of this ideal world, the real-world modeling practice remains dependent on calibration, a domain where we have been trained to hunt for optimal parameter sets in an  $n$ -dimension hyperspace ( $n$  being the number of free parameters in the hydrological model). This of course causes problems in the case of ungaged catchments, for which calibration data are not available [Bárdossy, 2007; Hrachowitz *et al.*, 2013; Oudin *et al.*, 2008; Parajka *et al.*, 2013]. However, calibration problems are not restricted to ungaged catchments: on gaged catchments, calibration is subject to numerous pitfalls that may result in miscalibration, i.e., when the parameter search fails to identify the mathematical optimum, and/or overcalibration, i.e., when, although the mathematically optimum parameter set has been identified over the calibration period, it does not remain mathematically optimum over different periods (see Andréassian *et al.* [2012] for a more detailed discussion on this topic).

In this context, we would like to reduce the dependence on classical model calibration. Here we wish to test the challenge of trading (as little as possible of) model efficiency for (as much as possible of) parameter genericity (i.e., the capacity to be not specific to a unique catchment).

#### 1.2. Why Should We be Willing to Trade Efficiency for Genericity?

There are several reasons that could lead us to look for compromises in the parameterization of our models:

1. *Ungaged catchments*—dealing with ungaged locations would be much easier if a small number of generalist parameter sets was available for any model. The recent IAHS-led decade on prediction in ungaged basins has contributed to intensifying the efforts on this topic [Blöschl *et al.*, 2013; Hrachowitz *et al.*, 2013] but also shown that the issue is far to have found a unique satisfying solution.

2. *Robustness*—naturally, we would all agree on trading efficiency in calibration for robustness, because calibration is just an intermediary step before model application on an independent data set, as is the case in validation tests. A robust parameterization is one which does not lose much efficiency (on average) between calibration and validation. This issue of robustness is often discussed along with the extrapolation capacity of hydrological models (over a range of validation periods): see the old idea of the differential split sample test of Klemeš [1986] and several more recent attempts [Coron *et al.*, 2012; Gharari *et al.*, 2013; Perrin *et al.*, 2008].
3. *Multicriteria assessment*—we would also accept to trade efficiency measured by a given criterion for improved efficiency over a set of complementary criteria: this is precisely what multiobjective calibration techniques attempt to do (see, e.g., Efstratiadis and Koutsoyiannis' [2010] review).

There is an abundant literature on robust calibration. For example, Koren *et al.* [2003] and Leavesley *et al.* [2003] argued that a good means to avoid overcalibration with distributed models was to force some a priori level of spatial and physical consistency into parameter estimates. The concept of regional calibration was also proposed by a few authors as a way to improve parameter transposability [Fernandez *et al.*, 2000; Parajka *et al.*, 2007]. One can also cite Kuzmin *et al.* [2008] who advocated local calibration approaches starting from physically relevant a priori parameter guesses.

However, to our knowledge, the issue of generalist parameter sets raised in this paper has only received very little interest in the hydrologic literature, which has rather focused on debating the issue of generalist and specialized model structures (see, e.g., the discussion on the FLEX modeling framework of Fenicia *et al.* [2008] and Fenicia *et al.* [2011]). We discussed this genericity question under the name “discrete parameterization” [Perrin *et al.*, 2008]: by “discrete,” we meant a parameter search that was restricted in its dimensionality from an  $n$ -dimension search for parameter values to a one-dimension search for parameter sets, i.e., the dimension of a list. However, we did not address the question of how short a list of parameters could be made: we simply showed that a random reduction of the parameter list (from 299 parameter sets to 75 sets) would only cause a modest efficiency loss.

### 1.3. Scope of the Paper

In this paper, we address a few basic questions, which we believe can help reduce some of the problems observed in the calibration of hydrological models: would it be possible to reduce calibration to a choice among a few parameter sets? What would the implications be in terms of model efficiency? How would this impact the robustness of the calibration process? Here our aim is to assess the potentiality of using only a few selected parameter sets to run a hydrological model, which implies searching for generalist parameter sets, i.e., parameter sets able to yield acceptable results on various catchments.

In section 2, we will first present the hydrological model used here (section 2.1), the three numerical criteria selected to decide on the acceptability of a parameter set (section 2.2), and the catchment set on which this study is conducted (section 2.3). Then, we will describe the approach followed to identify the most generalist parameter sets (given acceptability thresholds on the three criteria). Section 3 will discuss parameter set acceptability and section 4 the identification of a parameter short list, allowing us to reach the best average performance with the fewest generalist parameters. Last, we will investigate the advantages and limitations of calibrating our model based solely on this short list.

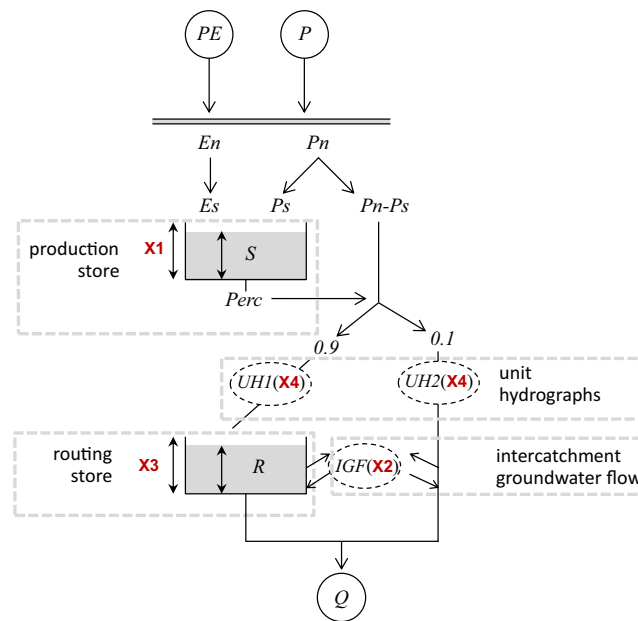
## 2. Material and Methods

### 2.1. Hydrological Model

We used the daily lumped continuous GR4J rainfall-runoff model presented by Perrin *et al.* [2003], whose structure is shown in Figure 1. The model has four parameters: the capacity of the production store ( $X_1$ , mm), the groundwater exchange coefficient ( $X_2$ , mm), the capacity of the routing store ( $X_3$ , mm), and the time base of the unit hydrograph ( $X_4$ , days). The only slight adaptation made for this study consisted in rewriting the unit hydrograph parameter  $X_4$  as a function of the catchment area  $S$ :  $X_4 = X_4^* \cdot S^{0.3}$  (tests, not presented here, showed that it does increase the genericity of the parameter). During our parameter search, we used  $X_4^*$ , which was then adjusted to each catchment using the previous equation.

### 2.2. Criteria Used to Evaluate Model Simulations

To judge the acceptability of a parameter set on a catchment, numerical criteria are needed. Here we will consider three basic and commonly used criteria:



**Figure 1.** Schematic diagram of the structure of the four-parameter GR4J model (X1–X4 are model parameters; PE: potential evapotranspiration, P: rainfall, Q: streamflow, other letters are internal state variables).

1.  $\alpha = \frac{\sigma_{obs}}{\sigma_{sim}}$ : the ratio of the standard deviation of the observed flow to the standard variation of the simulated flow. The optimal value is 1.
2.  $\beta = \frac{\mu_{obs}}{\mu_{sim}}$ : the ratio of the average observed flow to the average simulated flow (i.e., volume bias). The optimal value is 1.
3.  $\rho$ : the Pearson correlation coefficient (simulated versus observed flow). The optimal value is 1.

The three efficiency criteria  $\alpha$ ,  $\beta$ , and  $\rho$  will be aggregated into a single criterion, noted KGE, as proposed by Gupta et al. [2009]:

$$KGE = 1 - \sqrt{(1-\alpha)^2 + (1-\beta)^2 + (1-\rho)^2}$$

KGE is optimal for a value of 1.

In section 3, we will use triple thresholds ( $\alpha = \alpha_0$ ,  $\beta = \beta_0$ ,  $\rho = \rho_0$ ) to define the acceptability of parameter sets. A

parameter will be deemed acceptable if its simulation simultaneously yields criteria values above these thresholds. For the sake of simplicity, we will use in upcoming discussions four categories defined in Table 1, in order to qualify the model simulation efficiency obtained by use of the parameter set in question.

In this paper, model performance was assessed solely on the basis of the three criteria  $\alpha$ ,  $\beta$ , and  $\rho$ . However, the method presented here can be implemented with any performance criterion for which an efficiency threshold similar to those presented in Table 1 is defined.

### 2.3. Catchment Sets

A set of 202 unregulated catchments was selected in this study to develop the parameter short list (Figure 2). These 202 catchments represent various hydrological conditions, given the variability in climate, topography, and geology in France. This set includes Mediterranean catchments with intense precipitation as well as larger, groundwater-dominated basins and highland catchments. We did not include catchments where snow plays a major hydrological role, since no snowmelt module was used here.

A second (independent) data set of 515 French catchments was used to test independently the parameter lists obtained on the former catchment set (the test procedure will be explicated in section 4.1). This second set only differed from the test by the period of data availability (1980–2010 for the first, 1959–2010 for the second).

Daily rainfall, runoff, and potential evapotranspiration (PE) data series were available for both catchment sets. Meteorological inputs originate from the SAFRAN database [Vidal et al., 2010]. PE was estimated using the temperature-based formula proposed by Oudin et al. [2005]. Hydrological data were extracted from the

HYDRO national archive (www.hydro.eau-france.fr). Available times series lengths range from 20 to 50 years (average length was 34 years). The characteristics of both catchment sets are given in Table 2.

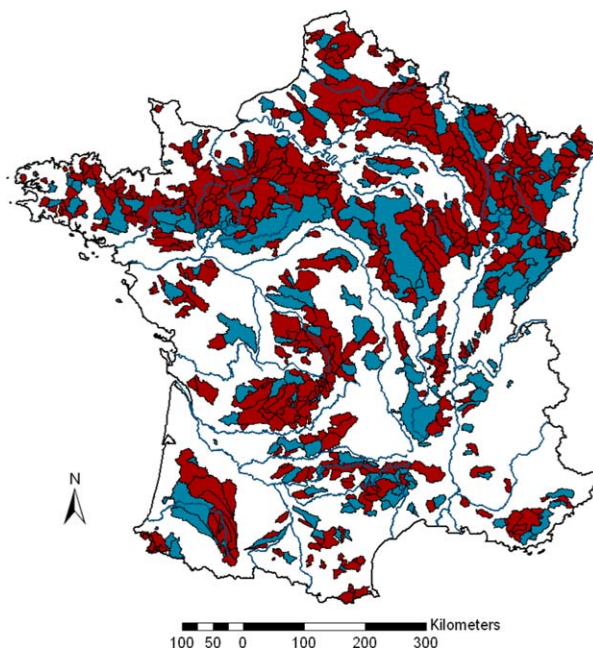
**Table 1.** Definition of the Four Qualitative Levels of Efficiency Requirements Retained for This Study<sup>a</sup>

Model Efficiency Requirement	$\alpha$	$\beta$	$\rho$
Very high	0.95 << 1.05	0.95 << 1.05	>0.9
High	0.90 << 1.10	0.90 << 1.10	>0.8
Moderate	0.85 << 1.15	0.85 << 1.15	>0.7
Low	0.80 << 1.20	0.80 << 1.20	>0.6

<sup>a</sup>A parameter set belongs to the qualitative efficiency class if it fulfills simultaneously the three conditions on  $\alpha$ ,  $\beta$ , and  $\rho$ .

### 2.4. Method Retained for Identifying and Assessing Generalist Parameter Sets

As a first step, we produced a Monte Carlo sample of  $10^6$  parameter sets, using the Latin hypercube sampling method (a structured



**Figure 2.** Location of the French catchments used for this study: in blue, the development set (202 catchments) and in red, the validation set (515 catchments).

uniform sampling method that yields a better coverage of the sampling space), between boundaries based on our experience with the GR4J model [see, e.g., Perrin et al., 2008]. For each of the  $10^6$  parameter sets, we computed the three above mentioned criteria, for each of our 202 catchments, over the 5 year time period 1996–2001. In this paper, our approach consists in analyzing this large number of model simulations:

1. First, we describe the acceptability of the parameter sets and analyze its dependency on the model's efficiency requirements.
2. Then, we investigate the parameter sets which show the best genericity.
3. Last, building on these most generalist parameter sets, we will identify a parameter short list to be used for model parameterization, as an alternative to full calibration.

### 3. On Parameter Set Acceptability

#### 3.1. Definitions

One of the key properties that this paper focuses on is the *genericity* of parameter sets; we need first to discuss it, and this will imply detailing the following properties:

1. *Parameter acceptability*: for a given catchment, we can identify the parameter sets among the  $10^6$  parameters reaching a certain efficiency threshold. These parameter sets will be qualified as **acceptable** for the given catchment. Naturally, acceptability will vary with efficiency requirements: the higher the requirements, the lower the acceptability.
2. *Catchment attainability*: similarly, for a given parameter set, we can identify the catchments where the parameter set obtains a given efficiency threshold. We say that these catchments are **attainable** by the given parameter set. Similar to acceptability, attainability will vary with the efficiency requirements: the higher the requirements, the lower the attainability.
3. *Parameter genericity*: widely acceptable parameter sets (i.e., parameter sets acceptable by more than x% of our 202 catchments) will be considered as **generalist** (at the x% level).

Note that looking for generalist parameter sets implies a pragmatic vision of hydrologic modeling: given the inherently *imperfect* nature of hydrological models, we seek neither a *perfect* fit nor the identification

**Table 2.** Summary of the Distributions of a Few Catchment Characteristics for the Development and the Test Sets

	Catchment Area (km <sup>2</sup> )	Mean Annual Rainfall P (mm)	Mean Annual Potential Evaporation E <sub>0</sub> (mm)	Mean Annual Discharge Q (mm)	P/E <sub>0</sub> Ratio	Q/P Ratio
<i>Development Set (202 Catchments)</i>						
First decile	46	768	653	168	1.05	0.21
Median	245	977	701	344	1.40	0.35
Ninth decile	2258	1391	795	830	2.04	0.60
<i>Test Set (515 Catchments)</i>						
First decile	56	723	634	160	1.03	0.22
Median	204	908	671	324	1.34	0.35
Ninth decile	1268	1194	752	616	1.77	0.53



**Table 3.** Total Number of Acceptable Parameter Sets and the Number of Attainable Catchments When Model Efficiency Requirements Rise

Model Efficiency Requirement (Model Performance Category)	Number of Parameter Sets Acceptable for at Least One Catchment (Out of 10 <sup>6</sup> )	Number of Catchments Attainable by at Least One Parameter Set (Out of 202)
Very high	2,170	125
High	23,258	189
Moderate	57,952	196
Low	115,043	202

of the unique *true* parameter set representing a given catchment. Rather, we see in generalist parameter sets the opportunity to identify clusters of catchments, which has naturally implications for similarity and classification issues.

**3.2. Parameter Set Acceptability-Sensitivity to the Model’s Efficiency Requirements**

Table 3 presents the total number of parameter sets acceptable for at least one catchment, along with the corresponding number of attainable catchments by at least one parameter set for four typical performance categories. For the “Low” category, it is possible to find at least one acceptable parameter set for each of the 202 catchments, and on average, there are 570 acceptable parameter sets per catchment. This number decreases when requirements rise: already with the “Moderate” category there are six catchments, which cannot be attained by any parameter set. This number rises to 13 for the “High” category (still only 6% of the catchment set), and to 77 for the “Very high” category (38% of the catchment set).

**3.3. On Generalist Parameter Sets**

Here we focus on those parameter sets which—for a given triplet of efficiency requirements ( $\alpha_0, \beta_0, \rho_0$ )—are found acceptable on at least  $x\%$  of the data set (this attainability limit will be discussed later: we start with  $x = 10\%$ ). The existence and the number of generalist sets is a function of the efficiency requirements: the higher the requirements, the lower their number (Table 4).

If we look for parameter sets that can attain a large number of catchments, we will find that some of them can be particularly efficient: for “Low” efficiency requirements, the most generalist parameter set is acceptable to 76 catchments. This number decreases to 52 for “Moderate,” to 24 for “High,” and to 9 for the “Very high” efficiency requirements.

Table 5 shows how the attainability limit impacts the number of generalist parameter sets when the limit is lowered from 10 to 5%. Clearly, the number of generalist parameter sets increases rapidly when the coverage requirement is lowered.

**Table 4.** Number of Generalist Parameter Sets as a Function of Efficiency Requirements

Efficiency Requirement	Number of Generalist Parameter Sets (Attainability Limit of 10%)	Number of Catchments Attained by the Generalist Parameter Sets		
		Max	Median	Min <sup>a</sup>
Very high	0			
High	22	24	20	20
Moderate	1831	52	25	20
Low	7708	76	30	20

<sup>a</sup>The minimum is 20 by definition, because the focus here is generalist parameter sets, which attain at least 10% of the total number of catchments.

**Table 5.** Number of Generalist Parameter Sets Depending on the Genericity Requirement

Efficiency Requirement	Number of Generalist Parameter Sets for an Attainability Limit of:					
	10%	9%	8%	7%	6%	5%
Very high	0	0	0	0	0	0
High	22	53	129	245	419	735
Moderate	1831	2439	3142	4076	5336	6891
Low	7708	8758	10,089	11,505	13,383	15,734

Unavoidably, some of the generalist sets are very similar to each other. But not all of them are; some can represent distinct catchment behaviors. Table 6 presents the percentage of catchments attainable by all of the parameter sets simultaneously. Interestingly, a few generalist sets can *collectively* provide an alternative for a significant part of our catchment set:

1. The 6891 generalist sets (corresponding to a minimum attainability of 5% and a “Moderate” efficiency requirement) can collectively attain 177 catchments (88% of the catchment set).
2. The 735 generalist sets (corresponding to a minimum attainability of 5% and a “High” efficiency requirement) can collectively attain 132 catchments (65% of the catchment set).

**Table 6.** Coverage of the Catchment Data Set by the Most Generic Parameter Sets, Depending on the Attainability Limit Defining Genericity<sup>a</sup>

Efficiency Requirement	Number of Catchments Attainable by the Generalist Sets Collectively (Total Number of Catchments: 202)					
	10%	9%	8%	7%	6%	5%
Very high	0	0	0	0	0	0
High	72	78	100	108	122	132
Moderate	157	158	161	166	170	177
Low	175	178	182	183	186	187

<sup>a</sup>For each of the generalist parameter sets identified in Table 5, we give here the total number of catchments which are attainable by at least one of the parameter sets.

3. The 22 generalist sets (corresponding to a minimum attainability of 10% and a “High” efficiency requirement) can collectively attain 72 catchments (36% of the catchment set).

The high coverage reached by a very limited number of parameter sets opens the way to imagining parameterization schemes based on the use of generalist parameter sets, even if they cannot cover absolutely all catchments.

Comparing Table 6 to Table 5 shows that although the number of generalist sets rises very rapidly when we decrease the attainability limit, the proportion of the catchment set attained by a least one generalist sets rises, but much more slowly: the new parameter sets mostly attain the same catchments.

#### 4. Toward a Short List of Parameter Sets: Aiming at Maximum Model Efficiency Using Only a Minimum Number of Parameter Sets

##### 4.1. Parameter Short List and Short-List Calibration (SLC)

We now wish to look for a way to constrain our choice of parameters in calibration while losing as little as possible in terms of simulation efficiency. As our objective is precisely to recruit parameter sets, we chose the term “short list.” Parameterizing our hydrological model based on a short list not only can considerably speed up the calibration process, but it could perhaps make it more robust too, as was shown in a previous attempt [Perrin et al., 2008].

Short-list calibration (SLC) consists in replacing an unconstrained free-range calibration with a calibration restricted to the short list: this means that the search for an optimal parameter set is reduced to the few parameter sets that are part of the short list.

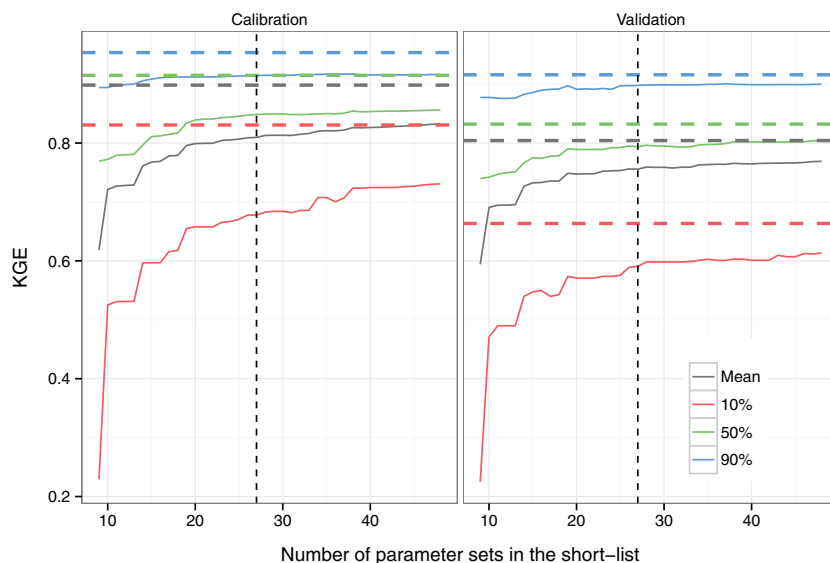
##### 4.2. Stepwise Selection of Efficient Short Lists

An efficient short list is one that yields the best model efficiency for a given number of parameter sets. To identify these lists, we proceeded in a stepwise manner, alternating forward selection and backward elimination (the pseudo-code presented in the Supporting Information provides a formal description). The specificity of this approach, however, lies in the fact that the search was started with the generalist parameter sets identified in the previous section.

The selection comprises the following steps:

1. Initialization: the short list search is initialized with the 22 generalist parameter sets corresponding to the “High” efficiency requirement and the 10% attainability limit (see Table 5).
2. Elimination: we first test whether one of the parameter sets is redundant. The objective function used is the average  $KGE_{av}$  value for the whole catchment set (noted  $KGE_{av}$ ). Here a parameter set is redundant when it can be removed without significantly affecting  $KGE_{av}$  (significant meaning a decrease of more than 0.001 in  $KGE_{av}$ ). We renew the test as long as a parameter set can be removed.
3. Selection: after all redundant parameters have been removed, we add an additional parameter set selected among suitable candidates (see below). The candidate selected is chosen by evaluating its added value to the SLC process.

During the selection process, parameter sets were explored by order of decreasing genericity (as defined in Table 6). This means (figures refer to Table 5) that after exploring the group of 22 high-efficiency parameter sets at an attainability level of 10%, we move to the 53 “High efficiency–9% genericity,” etc. until we have visited all the categories of Table 5. Each time we changed a category (from “High” to “Moderate” efficiency, from “Moderate” to “Low”), we included a single search through the entire (10<sup>6</sup>)



**Figure 3.** Efficiency of the short list calibration (SLC) on the independent catchment set, as a function of the number of parameter sets retained. Different colors represent different quantiles. Solid lines represent the efficiency of the short lists and dashed lines show the reference: full calibration.

collection of parameter sets to make possible the selection of more specific (i.e. less generic) parameter sets, provided they yield a larger improvement of  $KGE_{qv}$ .

This stepwise selection approach yielded lists of progressively increasing length, which can be used for SLC. For the sake of good scientific practice, SLC will be tested on an independent data set.

### 4.3. Validation of the Short List on an Independent Data Set

In order to identify a parameter list offering the best compromise between efficiency and length, we tested the suitability of the SLC process on a validation data set of 515 catchments (see section 2.3).

With each of the short lists successively, GR4J was calibrated on half of the period available and validated on the second half. Then we inverted the role of each time period. Figure 3 shows the results in calibration and validation for average KGE efficiency, as well as for three quantiles (10, 50, and 90%), in order to illustrate the variability among the 515 validation catchments.

Figure 3 shows that SLC provides model efficiencies that progressively approach that of full calibration. With long streamflow series, SLC remains inferior to full calibration, but if a short list of around 27 members (vertical dashed line in Figure 3) is selected, we see that we need to trade little efficiency in order to use generalist parameter sets (note in validation mode, the distance between full calibration and SLC is almost halved in comparison to the calibration mode).

Table 7 further analyzes these results by looking at the four different efficiency criteria: again, it demonstrates the good performance and robustness of SLC. Although full calibration remains slightly superior, we see that the difference with SLC in validation is limited (0.04 of KGE, 0.03 of correlation, very little difference for  $\alpha$  and  $\beta$ ): SLC is only inferior to full calibration for the correlation criterion.

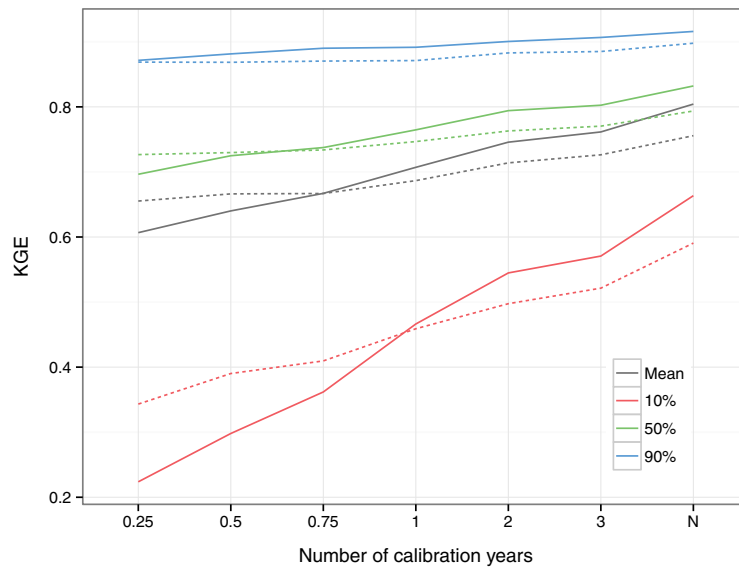
These results confirm those of Perrin *et al.* [2008] but with a much shorter list (they investigated lists varying between 75 and 299 parameter sets).

**Table 7.** Comparison of the Efficiency of SLC and Full Calibration in Validation Mode: Values Are Averages Obtained on 515 Independent Catchments, and in Parentheses, the 80% Confidence Interval Has Been Added<sup>a</sup>

	$\beta$	$\alpha$	$\rho$	KGE
Full calibration	1.01 (0.86–1.18)	1.00 (0.84–1.16)	0.90 (0.82–0.95)	0.80 (0.66–0.92)
Short-list calibration	0.99 (0.82–1.18)	0.99 (0.80–1.19)	0.87 (0.78–0.94)	0.76 (0.59–0.90)

<sup>a</sup>The objective function in calibration was KGE.





**Figure 4.** Impact of the availability of calibration data on the relative efficiency of full calibration (solid lines) and short-list calibration (dashed lines). The efficiency is computed in Kling-Gupta efficiency in validation mode (the same validation period is used for all lengths). N stands for the entire period.

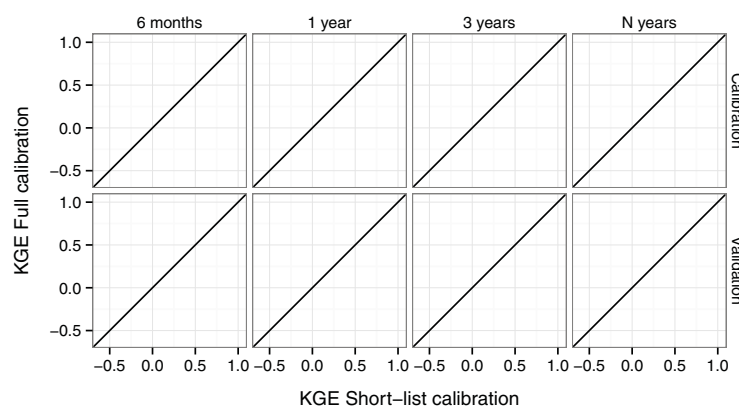
**4.4. Testing Short-List Calibration With Short Calibration Periods**

Here the analysis is pushed further by looking at the impact of the amount of calibration data on the relative performance of full and short-list calibrations (Figures 4 and 5).

Figures 4 and 5 show that under 1 year of calibration data, SLC becomes more efficient on almost all quantiles, first on the 10% efficiency quantile, then on the mean and median results. The additional constraint of calibrating only on the short list, although it reduces the efficiency when long calibration periods are available, has clear advantages when only short periods are available. Figure 5 illustrates this point: in calibration mode, full calibration cannot be beaten, but in validation mode, the scatterplot shows a progressive shift toward the 1:1 line. This confirms the results of Perrin et al. [2008].

**4.5. Description of the Compromise Short List (27 Parameters)**

Table 8 presents the compromise short list: out of the corresponding 27 parameter sets, 22 are generalist sets belonging to the “High” efficiency requirement, two are generalist sets belonging to the “Low” efficiency requirement, and three sets do not belong to any genericity group. Note that from the initial 22 parameter sets (“High-efficiency”-10% attainability), only two remain.

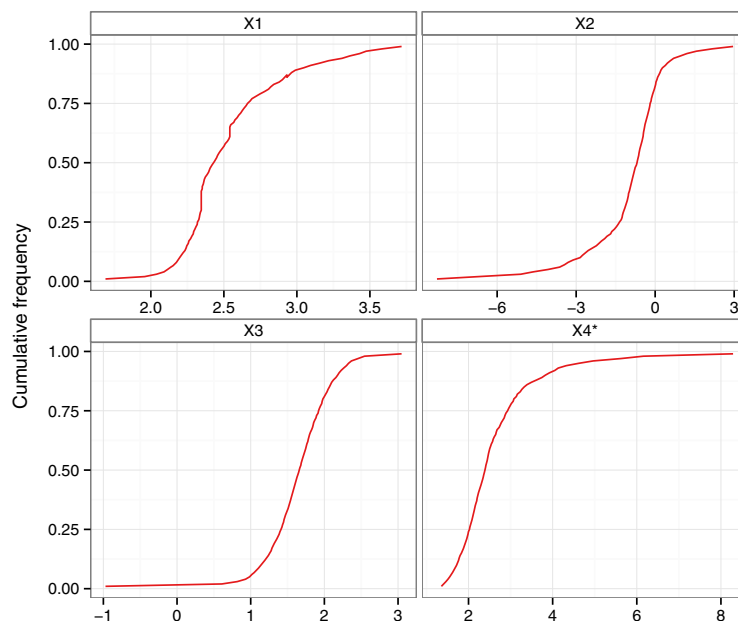


**Figure 5.** Full calibration versus short-list calibration, as a function of the calibration period length. (top row) Calibration results; (bottom row) validation results.

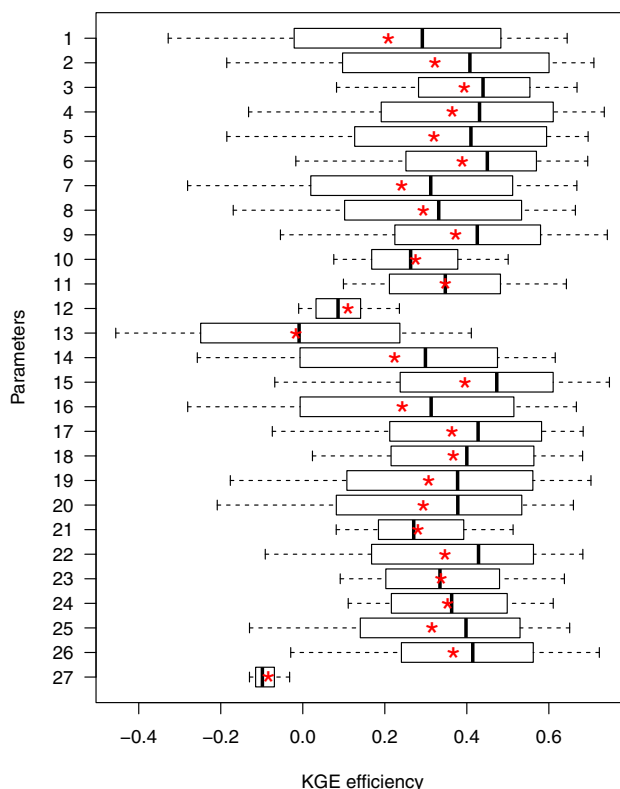
**Table 8.** Characteristics of the Compromise Short List

Parameter Set ID	X1 (mm)	X2 (mm)	X3 (mm)	X4* (days)	Parameter Set Properties	
					Selection Procedure on 202 Catchments: Efficiency Requirement and Attainability Level	Validation Test on 515 Catchments: Attainability Level (%)
1	174.9	-0.6	19.0	2.82	High, 6%	8.9
2	253.0	-0.6	45.8	1.95	High, 8%	6.3
3	351.2	-2.6	102.1	1.60	High, 6%	5.9
4	200.1	-1.5	76.2	2.36	High, 9%	5.5
5	180.9	-1.5	35.3	2.43	High, 7%	5.5
6	530.8	-0.4	64.6	1.43	High, 9%	5.0
7	198.2	-0.7	12.8	2.66	High, 7%	4.6
8	291.0	0.3	109.1	2.06	High, 5%	4.5
9	354.2	-0.5	103.0	2.49	High, 8%	4.4
10	836.9	-3.4	68.2	1.75	Low, 6%	4.4
11	826.5	-1.1	90.2	1.40	High, 5%	4.2
12	4005.6	-4.6	318.3	2.03		4.0
13	144.3	0.8	41.0	2.59		4.0
14	126.0	-0.2	169.2	1.25	High, 6%	3.8
15	275.7	-1.3	78.5	2.25	High, 10%	3.8
16	196.8	-0.7	28.6	1.84	High, 5%	3.7
17	327.1	-1.4	34.0	1.67	High, 6%	3.3
18	238.5	-2.9	267.7	2.03	High, 6%	3.1
19	215.9	-0.4	102.7	2.15	High, 7%	2.9
20	281.0	-0.7	8.2	2.69	High, 5%	2.5
21	346.5	-5.5	61.5	2.10	Low, 7%	2.4
22	328.4	-1.0	15.9	2.53	High, 6%	1.8
23	586.6	-1.5	154.4	2.67	High, 5%	1.7
24	364.4	-3.9	185.7	2.16	High, 5%	1.6
25	281.1	-1.2	15.1	2.28	High, 5%	1.2
26	342.5	-0.9	77.1	3.56	High, 10%	0.8
27	452.5	-54.5	149.9	1.24		0.3

Table 8 also provides the attainability levels for each set over the 515 independent catchment sets. Interestingly, two of the nongeneralist sets are selected by around 4% of the data set each, while one of the highly generalist sets (“High”—10%) is in fact chosen by less than 1% of the data set, but overall, the hierarchy of genericity is preserved.



**Figure 6.** Cumulative frequency of the parameters used over the 515 catchments when considering the full calibration (continuous lines) and the short-list calibration (discrete points). Note that a logarithm transformation is used for X1 and X3. Quantiles ranging from 0.01 to 0.99 are used here.



**Figure 7.** Range of efficiency of each of the parameter sets part of the compromise short list presented in Table 8: the values characterize the distribution of efficiencies computed on the 202 catchments of the calibration data set.

Figure 6 shows the cumulative distribution functions of the parameter sets used in full calibration and with the short-list calibration. One can notice that the ranges of variation are well captured by the 27 parameter sets of the short list, although some of the extreme values obtained by optimization are not represented.

In order to give an overview of the generality of each parameter retained in the compromise short list, we present in Figure 7 the distribution of each of the parameters efficiency (expressed in KGE) over the 202 catchments of the calibration data set. It shows that if most of the parameter sets have a rather high median efficiency, sets 12, 13, and 27 show a different behavior: they seem to represent an extremely specific behavior, and to give rather low performance on most of the other catchments. This difference was expected, as these three parameter sets are those which do not belong to one of the “high-efficiency” groups.

### 5. Conclusion and Perspectives

This paper aimed to identify generalist parameter sets for the GR4J hydrological model. We assembled a list of 27 parameter sets and validated it by relying exclusively on it for the calibration of GR4J. We call this approach short-list calibration (SLC).

We obtained mixed results: on one side, SLC efficiency remains inferior to that of full calibration. But on the other side, as the length of the calibration time period becomes shorter than 1 year, SLC takes progressively the advantage over full calibration. In terms of computing efficiency, this means that a calibration can be obtained with only 27 model runs, where a classical calibration would require at least several hundred. Even in the context of a full calibration, the use of the short list allows for a quick initial prescreening to initialize a local search procedure.

Several options could be explored: it would be instructive to test SLC with less parsimonious models (in their comparison, Perrin *et al.* [2008] showed that the more heavily parameterized model benefitted more from discrete calibration and this could be the same for SLC). And since in this paper, we based our short list selection on a single aggregated efficiency criterion (KGE), we could also test other criteria in an explicit multiobjective setting.

But more importantly, we see in this exploration of generalist parameter sets new perspectives for ungaged catchments (with ensemble simulations in an attempt to address the difficulty of parameterizing hydrological models) and for poorly gaged catchments (because we demonstrated that the short list was a powerful constraint when streamflow data were scarce).

Last, parameter short lists of different length offer a potential for catchment classification, based on behavioral similarity (i.e., under the assumption that the similarity of the model’s parameter sets obtained on two different catchments reflects the similarity of their behavior with respect to rainfall-runoff transformation, see the discussion by Oudin *et al.* [2010]). Further tests are needed in order to study the geographic

coherence of generalist parameter sets, possible relationships to physical catchment characteristics, and the geographic transferability of the proposed short list.

### Acknowledgments

The authors would like to acknowledge three anonymous reviewers who contributed to improve this paper through their constructive critics. Hydrometrical data used for this paper can be accessed through the French National Hydrometric Archive (<http://www.hydro.eaufrance.fr/>). Climatic data can be accessed through Météo France climatological archive (Publithèque: <http://publithèque.meteo.fr/okapi/accueil/okapiWebPubli/index.jsp>).

### References

- Andréassian, V., N. Le Moine, C. Perrin, M. H. Ramos, L. Oudin, T. Mathevet, J. Lerat, and L. Berthet (2012), All that glitters is not gold: The case of calibrating hydrological models, *Hydrol. Processes*, 26(14), 2206–2210.
- Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, 11, 703–710.
- Blöschl, G., M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije (2013), *Runoff Prediction in Ungauged Basins. Synthesis Across Processes, Places and Scales*, 465 pp., Cambridge Univ. Press, Cambridge, U. K.
- Coron, L., V. Andréassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, and F. Hendrickx (2012), Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, *Water Resour. Res.*, 48, W05552, doi:10.1029/2011WR011721.
- Efstratiadis, A., and D. Koutsoyiannis (2010), One decade of multi-objective calibration approaches in hydrological modelling: A review, *Hydrol. Sci. J.*, 55(1), 58–78.
- Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2008), Understanding catchment behavior through stepwise model concept improvement, *Water Resour. Res.*, 44, W01402, doi:10.1029/2006WR005563.
- Fenicia, F., D. Kavetski, and H. H. G. Savenije (2011), Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, *Water Resour. Res.*, 47, W11510, doi:10.1029/2010WR010174.
- Fernandez, W., R. M. Vogel, and S. Sankarasubramanian (2000), Regional calibration of a watershed model, *Hydrol. Sci. J.*, 45(5), 689–707.
- Gharari, S., M. Hrachowitz, F. Fenicia, and H. H. G. Savenije (2013), An approach to identify time consistent model parameters: Sub-period calibration, *Hydrol. Earth Syst. Sci.*, 17, 149–161.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377(1–2), 80–91.
- Hrachowitz, M., et al. (2013), A decade of predictions in ungauged basins (PUB): A review, *Hydrol. Sci. J.*, 58(6), 1198–1255.
- Klemes, V. (1986), Operational testing of hydrologic simulation models, *Hydrol. Sci. J.*, 31(1), 13–24.
- Koren, V., M. Smith, and Q. Duan (2003), Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models, in *Calibration Watershed Models*, *Water Sci. Appl.*, vol. 6, edited by Q. Duan et al., pp. 239–254, AGU, Washington, D. C.
- Kuzmin, V., D. J. Seo, and V. Koren (2008), Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search, *J. Hydrol.*, 353, 109–128.
- Leavesley, G., L. E. Hay, R. J. Viger, and S. L. Marstrom (2003), Use of a priori parameter estimation methods to constrain calibration of distributed-parameter models, in *Calibration Watershed Models*, *Water Sci. Appl.*, vol. 6, edited by Q. Duan et al., pp. 255–266, AGU, Washington, D. C.
- Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Ancil, and C. Loumagne (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, *J. Hydrol.*, 303(1–4), 290–306.
- Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, W03413, doi:10.1029/2007WR006240.
- Oudin, L., A. Kay, V. Andréassian, and C. Perrin (2010), Are seemingly physically similar catchments truly hydrologically similar?, *Water Resour. Res.*, 46, W11558, doi:10.1029/2009WR008887.
- Parajka, J., G. Blöschl, and R. Merz (2007), Regional calibration of catchment models: Potential for ungauged catchments, *Water Resour. Res.*, 43, W06406, doi:10.1029/2006WR005271.
- Parajka, J., A. Viglione, M. Rogger, J. L. Salinas, M. Sivapalan, and G. Blöschl (2013), Comparative assessment of predictions in ungauged basins—Part 1: Runoff-hydrograph studies, *Hydrol. Earth Syst. Sci.*, 17, 1783–1795.
- Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, 279(1–4), 275–289.
- Perrin, C., V. Andréassian, T. Mathevet, and N. Le Moine (2008), Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments, *Water Resour. Res.*, 44, W08447, doi:10.1029/2007WR006579.
- Vidal, J.-P., É. Martin, L. Franchistéguy, M. Baillon, and J.-M. Soubeyroux (2010), A 50-year high-resolution atmospheric reanalysis over France with the Safran system, *Int. J. Climatol.*, 30, 1627–1644, doi:10.1002/joc.2003.

# 1. Appendix 1 (supplementary material)

2

3 We reproduce below a pseudo-code for the stepwise selection of the most efficient  
4 Short-List, as an aid in following the explanation given in section **Erreur ! Source du**  
5 **renvoi introuvable..**

6

7  $N_{SL}$  : Number of parameter sets in the Short List  
8  $N_{tot}$  : Total number of parameter sets ( $10^6$ )  
9  $i_{EG}$  : Index of the efficiency group (see Table 5 : 1 for “Very High”, 2 for “High”, etc.)  
10  $j_{AL}$  : Index of the attainability limit level (see Table 5 : 1 for 10%, 2 for 9%, etc.)  
11  $N(i, j)$  : List of the parameters listed in Table 5 under efficiency group  $i$  and  
12 attainability limit  $j$

13

14 Program

15  $N_{SL}=0$

16  $Crit_{ref}=0.$

17

18 do  $i_{EG} = 1, 4$

19     do  $j_{AL} = 1, 6$

20         do  $k_{Candidate} = 1, N(i_{EG}, j_{AL})$

21             call Test( $k_{Candidate}$ )

22             enddo

23         enddo

24         do  $k_{Candidate} = 1, N_{tot}$

25             call Test( $k_{Candidate}$ )

26             enddo

27     enddo

28 End

29

30 Subroutine Test( $k_{Candidate}$ )

31     call Add\_to\_the\_Short\_List( $k_{Candidate}$ )

32      $N_{SL} = N_{SL}+1$

```
33   call Evaluate_the_Short_List(crit)
34   if(crit.gt.critref) then
35       critref = crit
36       do i = 1, NSL
37           crit = Evaluation_of_the_Short_List_without(i)
38           if(abs(crit-critref).lt.0.001) then
39               call Remove_from_the_Short_List(i)
40               NSL = NSL-1
41           endif
42       enddo
43   else
44       call Remove_from_the_Short_List(kCandidate)
45       NSL = NSL-1
46   endif
47 end subroutine Test
48
49
50
```