

RAPID applied to the SIM-France model

Cédric H. David, Florence Habets, David R. Maidment, Zong-Liang Yang

▶ **To cite this version:**

Cédric H. David, Florence Habets, David R. Maidment, Zong-Liang Yang. RAPID applied to the SIM-France model. Hydrological Processes, 2011, 25 (22), pp.3412-3425. 10.1002/hyp.8070 . hal-00641722v2

HAL Id: hal-00641722
https://hal.science/hal-00641722v2

Submitted on 18 Nov 2011

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.

RAPID applied to the SIM-France model

Cédric H. David[1,2,3] (cedric.david@mail.utexas.edu), Florence Habets[4] (florence.habets@mines-paristech.fr), David R. Maidment[2] (maidment@mail.utexas.edu) and Zong-Liang Yang[3] (liang@mail.utexas.edu)

1. Centre de Géosciences, Mines Paristech, Fontainebleau, France
2. Center for Research in Water Resources, University of Texas at Austin, Austin, Texas, USA.
3. Department of Geological Sciences, Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA.
4. UMR-7619 Sisyphe [CNRS, UPMC, Mines-Paristech], Paris, France.

Corresponding author

Cédric H. David
Department of Geological Sciences, Jackson School of Geosciences
The University of Texas at Austin
1 University Station C1160
Austin, TX 78712
cedric.david@mail.utexas.edu
+1 512 299 7191 (cell)
+1 512 471 5355 (office)
(No fax number available)

Abstract

SIM-France is a large connected atmosphere/land surface/river/groundwater modeling system that simulates the water cycle throughout metropolitan France. The work presented in this study investigates the replacement of the river routing scheme in SIM-France by a river network model called RAPID to enhance the capacity to relate simulated flows to river gages and to take advantage of the automated parameter estimation procedure of RAPID. RAPID was run with SIM-France over a ten-year period and results compared with those of the previous river routing scheme. We found that while the formulation of RAPID enhanced the functionality of SIM-France, the flow simulations are comparable in accuracy to those previously obtained by SIM-France. Sub-basin optimization of RAPID parameters was found to increase model efficiency. A single criterion for quantifying the quality of river flow simulations using several river gages globally in a river network is developed that normalizes the square error of modeled flow to allow equal treatment of all gaging stations regardless of the magnitude of flow. The use of this criterion as the cost function for parameter estimation in RAPID allows better results than by increasing the degree of spatial variability in optimization of model parameters. Likewise, increased spatial variability of RAPID parameters through accounting for topography is shown to enhance model performance.

Keywords stream flow, river network, network matrix, parameters, estimation, dam, quad-tree

1. Introduction

In the past two decades, several large scale river routing schemes have been used along with land surface models for hydrologic modeling. Among the most notable applications of large scale river routing are TRIP [Total Runoff Integrating Pathways, *Ngo-Duc, et al.*, 2007; *Oki and Sud*, 1998], RiTHM [River-Transfer Hydrological Model, *Ducharne, et al.*, 2003], the routing model of Lohmann et al. [*Lohmann, et al.*, 1996; 1998a; 1998b; 1998c; 2004; *Maurer, et al.*, 2001], that of Wetzel [*Abdulla, et al.*, 1996; *Nijssen, et al.*, 1997; *Wetzel*, 1994], and that of Olivera et al. [2000]. These approaches have been used along with land surface parameterization schemes to calculate river flow from runoff at the regional, continental and the global scale. MODCOU [Modèle Couplé, *Ledoux, et al.*, 1989] is another model with routing capabilities that differs from the previously cited models in that it has two separate networks of grid cells for horizontal routing of water on the land surface: one for overland routing and one for routing within the river system. MODCOU simulates flows throughout Metropolitan France (mainland France and Corsica) as part of the SIM-France modeling framework [*Habets, et al.*, 2008]. SIM-France (SAFRAN-ISBA-MODCOU-France) is a large connected atmosphere, land surface, river and groundwater model (see Figure 1) that involves coupling the national-scale atmospheric analysis system SAFRAN [Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige, *Durand, et al.*, 1993; *Quintana-Seguí, et al.*, 2008], with the ISBA land surface model [Interactions Soil- Biosphere-Atmosphere, *Boone, et al.*, 1999; *Noilhan and Planton*, 1989], and with the MODCOU hydrogeological model [*Ledoux, et al.*, 1989]. ISBA computes the vertical water and energy balance between the land surface and the atmosphere. The improved physics of

69 the land surface parameterization of ISBA that consist of an exponential profile for soil
70 hydraulic conductivity developed in Decharme et al. [2006] with calibration of soil
71 hydraulic conductivity and subgrid runoff over France by Quintana-Seguí et al. [2009]
72 are used in this study. Surface runoff and deep-soil drainage are computed by ISBA and
73 transferred to MODCOU which computes the horizontal flow routing on the land surface,
74 in rivers and in aquifers. Aquifers in MODCOU are modeled within the two main river
75 basins of France, the Seine and the Rhône, which together represent 30% of the land area
76 of France.
77 MODCOU handles the calculations of flow and volume of water within the river network
78 of SIM-France. This river network is made up of grid cells divided into a quad-tree
79 pattern and the calculations of MODCOU are made for groups of quad-tree cells; not for
80 each quad-tree cell separately. Using groups of cells for calculations is advantageous for
81 reducing computational costs but it limits the modularity of MODCOU. In particular, the
82 location and number of gaging stations are difficult to modify.
83 The work presented herein investigates the impact of replacing the routing module used
84 in MODCOU by a river network model called RAPID [Routing Application for Parallel
85 ComputatIon of Discharge, *David, et al.*, 2011]. RAPID uses a matrix-based version of
86 the Muskingum method to calculate flow and volume of water for each reach of a river
87 network separately and has an automated parameter estimation procedure. RAPID
88 therefore allows greater flexibility than the routing module in MODCOU with regards to
89 changing the locations of computations in an existing domain or to running SIM on a new
90 domain. RAPID was previously applied to a GIS vector river network [*David, et al.*,
91 2011], and the present study shows how it can also be applied to a quad-tree gridded river

network. In addition, RAPID is advantageous because of its ability to run in a parallel computing environment and its fine time step allowing potential comparison with river flow observations at high temporal resolution. Finally, replacing the routing module of MODCOU by RAPID has already allowed computing river flow height and helping quantify river/aquifer interactions at the regional scale [*Saleh, et al.*, 2010; 2011].

In this paper, the original river routing of MODCOU as well as that of RAPID are briefly presented followed by a ten-year application (1995-2005) of SIM-France comparing the two river routing applications for different sets of parameters used in RAPID.

2. Modeling framework

2.1. River modeling in SIM-France

The computational domain of SIM-France includes all of Metropolitan France, including Corsica. Parts of Spain, Switzerland, Germany and Belgium are also included where their drainage area flows through France, as shown in Figure 2. The total surface area of the computational domain is 610,000 km^2.

Surface routing and river routing in SIM-France are done by MODCOU [*Ledoux, et al.*, 1989]. The surface and river networks of SIM-France and their connectivity were created using a routine called HydroDem [*Leblois and Sauquet*, 2000] and consist of 193,861 surface cells and 24,264 river cells, each river cell being a particular surface cell. The surface area covered by the river cells is 65,000 km^2. The surface network uses a quad-tree structure with cell sizes of 1 km, 2 km, 4 km and 8 km. The river network has cell sizes of 1 km and 2 km. The smaller quad-tree cells are used at the conference of branches of the river network for better representation of the network connectivity and at basin boundaries for more accurate basin surface area.

The connectivity between river cells is given by a table that provides for each downstream river cell up to three upstream river cells. There are no loops or divergences in the river network of SIM-France. The connectivity between catchments and rivers is given by a table that provides for each surface cell a unique downstream cell where its runoff enters the river.

For both surface and river routing, the calculations of flow and volume of water within MODCOU are carried out using groups of cells as computing elements, therefore minimizing the amount of calculations compared to computing for all cells separately.

124 These groups of cells – or isochrone zones – are based on the notion of isochronism
125 developed by Leblanc and Villeneuve [1978]. An isochrone is a line representing a
126 constant time of travel to a reference point downstream. An isochrone zone is the area
127 between two successive isochrones. This zone is represented by a set of cells which are a
128 single computational unit in MODCOU. Both the land surface isochrones and river
129 isochrones of MODCOU have three-hour time intervals, which means that the time of
130 travel between the upstream-most and the downstream-most cell in a given isochrone
131 zone is approximately three hours. All the isochrones of a given network are determined
132 using the travel time between connected cells which is estimated based on topography
133 and on the geometry of the quad-tree mesh. For surface cells and river cells, the travel
134 time $\tau_{i,j}$ between two consecutive cells i and j is calculated using the distance
135 $d_{i,j}$ between the two cells and the slope $s_{i,j}$, as shown in Equation (1):
136
137 $$\tau_{i,j} = \alpha \cdot \frac{d_{i,j}}{\sqrt{s_{i,j}}} \quad (1)$$
138
139 where α is the inverse of a velocity. In the current version of SIM-France, a unique
140 value of α is calibrated for each major basin.
141 Figure 3 shows an example of the isochrone zones and connectivity between surface cells
142 and river cells in MODCOU for the Ardèche River Basin. Figure 3a) shows the Ardèche
143 River, its basin and three river gages. Figure 3b) shows the river isochrone zones of the
144 Ardèche River. Figure 3c) shows the surface isochrone zones corresponding to the
145 upstream-most river isochrone zone. Each surface cell belongs to a surface isochrone

146 zone, but only the isochrone zones corresponding to one river isochrone zone are shown
147 of Figure 3c) for clarity. The units used for isochrone zones are the number or
148 MODCOU 3-hour time steps to the outlet (here the Mediterranean). The quad-tree
149 structure of increasing resolution can be seen at the boundary of the basin in Figure 3c).
150 In MODCOU, the volume of water V^{out} that discharges across each isochrone line in a
151 computation time step is calculated differently for the surface network and for the river
152 network. For routing on the land surface, all the volume of water V available in the
153 isochrone zone is transferred to the downstream zone, as shown in Equation (2):
154

155 $$V^{out} = V \tag{2}$$
156

157 For routing in the river network, V^{out} is proportional to the volume of water V available
158 within the isochrone zone as shown in Equation (3):
159

160 $$V^{out} = \beta \cdot V \tag{3}$$
161

162 where $\beta \in [0,1]$ is manually calibrated and usually set constant for large basins. Equation
163 (3) can be viewed as the linear reservoir equation associated with a first-order explicit
164 development of the continuity equation. The variation of volume related to lateral inflow
165 and groundwater inflow of water are added to the volume V before calculating V^{out}. In
166 SIM-France, β has four possible values: 0.5, 0.7, 0.8 and 0.9 as shown in Figure 4.

Equation (3) is applied to isochrone zones. Hence, the volume of water within each isochrone zone needs be partitioned among its several river cells before computation of the river-aquifer exchanges. This interaction depends on the aquifer head, on the river head – assumed constant – and on the volume of water in the river cell when the river infiltrates water into the aquifer. The partitioning of water volume among all cells of an isochrone zone is done using a weighted average of the total amount of water reflecting the spatial distribution of lateral inflow in each isochrone zone.

This formulation has several inconsistencies, especially when the junction between two streams lies in the interior of an isochrone zone. This can have a consequence in river-aquifer interactions, but also in the computation of river flow. Furthermore, using only one set of isochrones in each basin can lead to two gages being located in one isochrone zone (for example a zone containing a confluence with gaging stations on both sides), in which case the flow computed by MODCOU has to match the flow at two different gaging stations. In order to avoid such inconsistencies, MODCOU uses a unique set of isochrone zones for each gage, such that each gage is the downstream-most river cell in its isochrone zone. Therefore, several flow calculations can be performed for a given cell, if the given cell belongs to several isochrone zones, which is inefficient and requires time consuming processing work in case of change of number or locations or river gages. The work done herein aims at simplifying the river modeling done within SIM-France and to ensure evolution of the code as for instance the computation of river flow height [*Saleh, et al.*, 2010; 2011] and velocity.

2.2. RAPID

RAPID [*David, et al.*, 2011] is a river network model that uses a matrix-based version of the Muskingum routing scheme to calculate discharge simultaneously through a river network. RAPID was first applied to the Guadalupe and San Antonio River Basins in Texas using a vector-based river network extracted from a geographic information system dataset called NHDPlus [*USEPA and USGS*, 2007]. The governing equation used in RAPID is the following:

$$\left(\mathbf{I}-\mathbf{C_1}\cdot\mathbf{N}\right)\cdot\mathbf{Q}\left(t+\Delta t\right)=\mathbf{C_1}\cdot\mathbf{Q^e}\left(t\right)+\mathbf{C_2}\cdot\left(\mathbf{N}\cdot\mathbf{Q}\left(t\right)+\mathbf{Q^e}\left(t\right)\right)+\mathbf{C_3}\cdot\mathbf{Q}\left(t\right) \quad (4)$$

where t is time and Δt is the river routing time step. The bolded notation is used for vectors and matrices. All matrices are square. $\mathbf{I}$ is the identity matrix. $\mathbf{N}$ is the river network connectivity matrix which has a value of one in element $N_{i,j}$ if reach j flows into reach i and zero elsewhere. $\mathbf{C_1}$, $\mathbf{C_2}$ and $\mathbf{C_3}$ are parameter matrices which depend on Muskingum k, x and time step Δt. $\mathbf{Q}\left(t\right)$ is a vector of outflows from river reaches, and $\mathbf{Q^e}\left(t\right)$ is a vector of lateral inflows to these reaches from land surface runoff or groundwater inflow. The number of river quad-tree cells – here 24,264 – is used for dimension of all vectors and matrices, each element of the vectors corresponding to one river cell.

Provided with a vector of lateral inflows $\mathbf{Q^e}\left(t\right)$, RAPID calculates the flow and volume of water in all reaches of a river network, therefore allowing coupling of a river network to most land surface models and groundwater models. A different value for the parameters k and x of the Muskingum method can be assigned for each river quad-tree

211 cell, and RAPID uses two vectors $\mathbf{k}$ and $\mathbf{x}$ as input which are used to compute the values
212 of the matrices $\mathbf{C_1}, \mathbf{C_2}$ and $\mathbf{C_3}$. However, before routing with RAPID, horizontal surface
213 and subsurface routing is needed to transport runoff from a land surface cell to its
214 corresponding river cell. In the present study, this surface and subsurface routing is done
215 by MODCOU and RAPID replaces only the river modeling of MODCOU.
216 The connectivity information that already exists between the river cells in the SIM-
217 France river network is used to create the network connectivity matrix $\mathbf{N}$ needed by
218 RAPID and described in David et al. [2011].
219 RAPID uses an automated parameter estimation procedure which, given lateral inflow
220 $\mathbf{Q^e}$ everywhere in the river network, and gage measurements at some locations,
221 determines a best set of parameters based on a square error cost function. As in David et
222 al. [2011], the search for optimal vectors of parameters $\mathbf{k}$ and $\mathbf{x}$ is made by determining
223 two multiplying factors λ_k and λ_x such that:
224

225 $$\forall j \in [1, 24264] \qquad k_j^\rho = \lambda_k \cdot \frac{L_j}{c^0} \quad , \quad x_j^\rho = \lambda_x \cdot 0.1 \tag{5}$$
226

227 where j is the index of a quad-tree river cell, k_j^ρ and x_j^ρ are its Muskingum parameters,
228 L_j is the flow distance within a river cell and $c^0 = 1km \cdot h^{-1} = 0.28m \cdot s^{-1}$ is a reference
229 celerity for the flow wave. The parameters k_j^ρ and x_j^ρ are the same developed in David et
230 al. [2011] and are referred to as ρ parameters in the following. In this study, the size of
231 the side of each quad-tree river cell was used as an approximation of its flow distance.

232 The value of λ_x is bounded by the interval $[0,5]$ since the Muskingum method is stable
233 only for $x \in [0.0, 0.5]$, as shown in Cunge [1969]. The two scalars λ_k and λ_x are
234 determined such that the corresponding vectors $\mathbf{k}$ and $\mathbf{x}$ minimize the value of an
235 optimization criteria, or cost function. At the end of the optimization procedure, one
236 couple (λ_k, λ_x) is determined for a given part of the network. The values of λ_k and λ_x can
237 be determined for the entire study domain, or for sub-basins. If a sub-basin is located
238 downstream of another sub-basin, observations at a gaging station are used to provide the
239 upstream flow. Therefore, the delineation of sub-basins has to be consistent with the
240 location of available gage measurements.
241 The optimization procedure uses a line-search algorithm called the Nelder-Mead method
242 [*Nelder and Mead*, 1965] to determine the two scalars λ_k and λ_x.
243 The use of RAPID within SIM-France allows for flow and volume calculation at each
244 river cell and RAPID allows for the ready inclusion of additional river gages to be used
245 for calibration.
246

247 **3. Application of RAPID in France**

248 **3.1. Optimization of RAPID parameters**

249 This section focuses on the optimization of RAPID parameters with various options used
250 for k_j and x_j, for the optimization cost function and for the spatial variability of the
251 optimization. Two formulations are applied for computing k_j and x_j including one
252 formulation taking topography into account, two cost functions are tested, and three
253 different domain decompositions are used for optimization of parameters. In order to
254 simplify the optimization procedure and to ensure its repeatability, the parameter
255 estimation of RAPID was run uncoupled from SIM-France. Lateral and groundwater
256 inflow to the river network were obtained from a simulation using the standard version of
257 SIM-France (without RAPID) augmented with improved physics of the land surface
258 parameterization of ISBA developed in Decharme et al. [2006] and calibrated over
259 France by Quintana-Seguí et al. [2009]. Daily gage measurements from the French
260 HYDRO database [*SCHAPI*, 2008] were used for the parameter estimation as well as for
261 comparison with daily-averaged flow calculations .
262 The period of interest of the present study is August 1st 1995 to July 31st 2005. However,
263 the parameter estimation was performed using five months of the first winter (November
264 1st 1995 to March 31st 1996). As part of the first year (1995-1996) was used for
265 calibration, separate statistical results are presented for 1995-1996 and 1995-2005.
266 RAPID is run using a 30-minute time step and forced with 3-hourly lateral inflow
267 volumes; daily averages of computed discharge are compared with daily observations at
268 gage locations. There are 907 stations within the river network of SIM-France but only

493 of these have daily measurements every day during the first year (August 1st 1995 to July 31st 1996). Amongst the 493 available stations, the best 291 were utilized for optimization of RAPID parameters. The criterion used for the selection of the 291 best stations is a Nash efficiency [*Nash and Sutcliffe*, 1970] better than 0.5 in the existing SIM-France model (without RAPID) over 1995-1996. This selection excludes the gages that are affected either by dams or by water diversions, and thus avoids unrealistic model parameters due to anthropogenic modifications of river flow. Therefore, the proposed routing scheme is optimized at locations were the previous routing scheme already performed well.

The optimization is first performed on all rivers of the domain, therefore obtaining unique values of λ_k and λ_x for all 24,264 river quad-tree cells. However, such an optimization may not capture the variability between river basins and within sub-basins, due to the various slopes or soil types. Therefore, the optimization procedure was also run independently within the seven main river basins of France shown in Figure 5a) and within the twenty sub-basins shown in Figure 5b).

In order to limit the effect of the initial state of the system at the beginning of the optimization procedure, the initial flows on 01 November 1995 were estimated using a simple run of RAPID. This estimation was obtained through running the routing model from 01 August to 31 October 1995 with uniform values of λ_k and λ_x over the study domain and initial flows $0\ m^3/s$ for all river cells on 01 August 1995.

The results of a parameter estimation procedure depend slightly on the initial guess for the parameters. Therefore, three different sets of initial guesses for λ_k and λ_x were used:

291 $(\lambda_k, \lambda_x) = (2,3)$, $(\lambda_k, \lambda_x) = (4,1)$ or $(\lambda_k, \lambda_x) = (1,1)$. The numerical values of these three
292 sets have no particular meaning and serve to start the optimization with a different initial
293 value for **k** and **x**. Each set of initial guesses leads to slightly different results for the
294 optimal λ_k and λ_x. Out of the three sets of optimal λ_k and λ_x that are determined for
295 each sub-basin, only the best is kept. This selection is based on the set of parameters that
296 leads to the smallest value of the optimization cost function.
297 Once the optimization procedure was completed, RAPID was run with SIM-France over
298 a 10-year period, from August 1995 to July 2005. This section focuses on the first year
299 while the next section studies the ten-year run. In order to compare the overall
300 performance of both routing models on the river network, the Nash efficiency and the
301 root mean square error (RMSE) were calculated for each of the 493 gaging stations over
302 1995-1996. These criteria are sorted and comparisons between the computations of
303 MODCOU and those of RAPID are shown in Figure 6. The two graphs in Figure 6 do
304 not allow comparing both models at each gaging station since the criteria are sorted, but
305 they depict the overall relative performance of both models. Table 1 shows the average
306 Nash efficiencies and RMSEs obtained by the original version of SIM-France and with
307 RAPID using various optimization procedures. During 1995-1996, 382 stations have a
308 positive efficiency using the standard version of SIM-France. The averages presented in
309 Table 1 show the best 382 values for both efficiency and RMSE, but similar patterns are
310 found for all 493 values or the best 291 values.
311 In its original formulation, the criterion used in the optimization of RAPID is based on a
312 square error cost function ϕ_1. This function is the sum of the square errors between daily

measurements $Q_i^g(t)$ and daily-averaged $\overline{Q_i}(t)$ flow computations for several river gaging stations i and for everyday of a given period of time $[t_o, t_f]$, as shown in Equation (6).

$$\phi_1(\mathbf{k},\mathbf{x}) = \sum_{t=t_o}^{t=t_f} \sum_{i=1}^{i=291} \left[\frac{\overline{Q_i}(t) - Q_i^g(t)}{f} \right]^2 \quad (6)$$

where the summation is made daily and at river cells with active gaging stations only. t_o and t_f are respectively the first day and last day used for the calculation of ϕ_1. $i \in [1, 291]$ is the index for gaging stations. The model parameter vectors $\mathbf{k}$ and $\mathbf{x}$ are kept constant within the temporal interval $[t_o, t_f]$, and the cost function is calculated several times with different sets of parameters during the optimization procedure. f is a scalar that allows ϕ_1 to be of the order of magnitude of 10^1 which is helpful for automated optimization procedures. One can notice that, in ϕ_1, a given fractional error (5% error between modeled and measured flow for example) for two stations with different orders of magnitude for river flow influences the cost function differently. A small fractional error on a gaging station with a large flow penalizes the cost function more than the same fractional error on a gaging station with small flow. The Nash efficiency E is highly influenced by the difference between the model computation and the mean average flow, as shown in Equation (7):

333 $$E = 1 - \frac{\sum_{t=t_o}^{t=t_f} \left[Q_i^g(t) - \overline{Q_i}(t) \right]^2}{\sum_{t=t_o}^{t=t_f} \left[Q_i^g(t) - \left\langle Q_i^g \right\rangle \right]^2} \quad (7)$$

334

335 where $\left\langle Q_i^g \right\rangle$ is the average daily flow observed at the gaging station i over a long

336 interval. Therefore, the use of ϕ_1 penalizes the Nash efficiency. In order to avoid that the

337 order of magnitude of flow at each gaging station influences their weight in the cost

338 function, a new cost function ϕ_2 is created, as shown in Equation (8):

339

340 $$\phi_2(\mathbf{k}, \mathbf{x}) = \sum_{t=t_o}^{t=t_f} \sum_{i=1}^{i=291} \left[\frac{\overline{Q_i}(t) - Q_i^g(t)}{\left\langle Q_i^g \right\rangle} \right]^2 \quad (8)$$

341

342 The new cost function ϕ_2 results in the changes shown in Table 1 and Figure 6 where the

343 Nash efficiencies and RMSEs obtained with RAPID using ϕ_2 are better than with ϕ_1.

344 Overall, the Nash efficiencies and the RMSEs in RAPID are comparable while not as

345 good as those obtained with the routing scheme of the original SIM-France. Therefore,

346 the choice of the cost function is crucial to determining a set of optimal parameters.

347 In order to estimate the effect of more spatial variability in the optimization of RAPID
348 parameters, the parameter estimation was done on different basins and sub-basins. Figure
349 7 shows the sorted Nash efficiencies and RMSEs obtained with three degrees of spatial
350 variability of optimization using ϕ_2 as the cost function. These spatial variabilities
351 include "France" which has uniform parameters over the whole domain, "basins" for the
352 7 river basins of Figure 5a) (Adour, Garonne, Loire, Seine, Meuse, Rhône and Hérault)
353 and "sub-basins" where the major river basins have been divided into 20 sub-basins as
354 shown in Figure 5b). The increase in spatial variability of optimization increases the
355 efficiency while the RMSE remains almost constant, but the increase in efficiency is
356 limited compared to that triggered by a change in the cost function. The values of
357 parameters λ_k and λ_x obtained with the parameter estimation procedure using the second
358 cost function are shown in Table 2.
359 The number of gaging stations in a basin can be divided by the number or river cells in
360 the basin to calculate an observability ratio O, as done in Table 2. This ratio ranges from
361 $O = 22$ on the Ardèche River to $O = 1307$ downstream of the Seine River, showing a wide
362 spread in density of observations. The Seine River, of great interest to the French
363 community, has a higher resolution and therefore more river cells in SIM-France than any
364 other basin – all the river cells are of size 1 km – which explains the lower observability
365 ratio. Unrealistically low results are obtained for λ_k in the downstream part of the Seine
366 River and for the Ardèche River Basin. The former is explained by the limited amount of
367 stations used for optimization in the downstream part of the Seine River Basin (only one
368 station). The latter is due to the basin being small with regards to the number of gages

(leading to a low observability ratio) and therefore over-constraining the optimization procedure. The observability ratio is therefore a key metric for the quality of the optimization. These unrealistic values for λ_k may partly explain why the effect of optimization from 7 basins to 20 sub-basins is very limited. As expected, the optimization procedure converges to the largest values of the parameter λ_k for the Seine and Loire rivers which are the slowest rivers. For each of the 7 major basins, the value of λ_k is bounded by the value of λ_k for each of their corresponding sub-basins. Also, one can notice that upstream parts of a basin are usually faster (lower λ_k) than downstream parts as can be seen for the upstream part of the Loire Basin, and for the Allier Basin which are located in high topography areas. This shows that – as expected – topography plays an important role in the travel time of flow waves. This motivates a final experiment where RAPID model parameters are estimated based on topography as shown in Equation (9).

$$\forall j \in [1, 24264] \qquad k_j^{\sigma} = \lambda_k \cdot \alpha \cdot \frac{d_{i,j}}{\sqrt{s_{i,j}}} \quad , \quad x_j^{\sigma} = \lambda_x \cdot 0.1 \tag{9}$$

This formulation of k_j^{σ} is adapted from Equation (1) which is used to determine the location of isochrone zones. In the following, the parameters k_j^{σ} and x_j^{σ} of Equation (9) are referred to as σ parameters. Table 1 shows the average efficiencies and RMSEs obtained with σ parameters using ϕ_1 and ϕ_2 uniformly over France, and with

389 σ parameters over the 7 major basins using ϕ_2. Figure 8 shows the sorted efficiencies
390 and RMSEs obtained with σ parameters using ϕ_1 and ϕ_2 uniformly over France. From
391 Table 1 and Figure 8 one can conclude that regardless of the optimization cost function
392 used, σ parameters allow to obtain better results than ρ parameters. Therefore, taking
393 topography into account in the travel time of the flow wave is advantageous. Similarly,
394 regardless of the parameters used and of the spatial resolution of the optimization,
395 optimizing using ϕ_2 allows obtaining better average results than with ϕ_1. The average
396 results obtained using σ parameters and ϕ_2 are comparable (slightly better) than those
397 obtained by the original routing module of SIM-France. One should note however, that
398 the best stations with MODCOU are better than the best with RAPID, while the worse
399 stations in MODCOU are worse than the worst in RAPID. This suggests a flattening of
400 the curves most likely due to equal treatment of all stations in the ϕ_2 cost function.
401 Finally, regardless of the cost function used in optimization or the set of parameters
402 (σ and ρ) basin and sub-basin optimizations have a limited effect on overall
403 performance of RAPID. This suggests that increased inter-basin and intra-basin
404 variability of river routing parameters has little effect on efficiency or RMSE at the
405 spatial scale of France as it is modeled in SIM-France.

406 **3.2. Comparison of routing schemes over 10 years**

407 Over 1995-2005, only 3 gaging stations have a full daily record. Therefore, results
408 presented in this section are using stations with gaps in observations; efficiency and
409 RMSE are calculated only at times when measurements are available. A threshold of
410 70% of daily measurements available over 1995-2005 leads to selecting 493 gaging

411 stations. These stations differ slightly from the ones used in 1995-1996. Out of the 493
412 stations that have full daily record in 1995-1996, 436 stations are included in the 1995-
413 2005 period. Similarly, out of the best 291 stations that have full daily record in 1995-
414 1996, 261 are included in 1995-2005. During 1995-2005, 427 out of the 493 stations
415 have a positive efficiency using the standard version of SIM-France. Table 1 shows
416 average statistics for 1995-2005 for the best 427 values for both efficiency and RMSE,
417 but similar patterns are found for all 493 values. The conclusions drawn in Section 3.2.
418 regarding the sets of parameters, the cost functions and the spatial resolution of the
419 optimization are still valid for the ten-year simulation. However, one should note that
420 over ten years, MODCOU performs slightly better than RAPID using the best set of
421 options. This may be explained by the slightly different stations used for the 5-month
422 optimization and for the ten-year study. However, five months of the first year seem to
423 be sufficient to capture RAPID parameters and allow comparable performance between
424 MODCOU and RAPID over ten years. Figure 9 shows the sorted efficiencies and
425 RMSEs obtained with MODCOU and with RAPID with σ parameters optimized
426 uniformly over France using ϕ_2. Globally the two models perform comparably although,
427 similarly to 1995-1996 results, the best stations are degraded and some stations with low
428 but positive efficiency are improved. Figure 10 shows observations and modeled
429 hydrographs during 1995-1996 and 1995-2005 for the Meuse River at Stenay (the
430 location of this station is shown in Figure 11) in which MODCOU and RAPID are almost
431 indiscernible. One should note, however, that in all the hydrographs plotted (not shown)
432 the timing of events differs slightly between the two models, none of which being

433 consistently better than the other regardless of the optimization options as expected from
434 results shown in Table 1 and Figure 9.
435 Figure 11 shows a spatial comparison of efficiencies obtained over France.
436 Improvements and degradations of statistical results between MODCOU and RAPID
437 have no particular spatial patterns. Overall, the discharge simulated by MODCOU and
438 RAPID are similar in RMSE and Nash efficiency. This similarity can be explained by
439 the strong dependence of discharge calculations on the lateral inflow forcing which is the
440 same for both river routing schemes. Furthermore, the routing equations used in
441 MODCOU and RAPID are comparable (the linear reservoir equation in SIM-France is a
442 simplified Muskingum equation, given x=0). The addition of RAPID to SIM-France can
443 be regarded as advantageous since RAPID provides with flow and volume of water in all
444 the cells of the river network separately and provides flexibility in the number and
445 location of river gages, which was not the case in the original version of SIM-France.
446 Also, the 30-min time step in RAPID allows potential comparisons with observations at
447 higher temporal resolution than the 3-hour time step of MODCOU. Finally, RAPID is
448 better suited than MODCOU for computation of river flow height in all grid cells of the
449 river network separately hence allowing the study of river-aquifer exchanges as shown in
450 Saleh et al. [2010; 2011].
451 **3.3. Influence of dams on river flow**
452 RAPID does not have a specific physical model for treatment of dams. However, the
453 model is designed such that observations at gaging stations can easily be substituted for
454 upstream flow. This capability is not available in the routing scheme of MODCOU and
455 is useful for a gaging station located at the outlet of a dam because the flows discharging

456 from man-made infrastructures reflect human decisions. In France, the quality of flow
457 calculations at the outlet of the Rhône River (at Beaucaire) is influenced by the dam at
458 the outlet of Lake Geneva. Figure 12 demonstrates the influence of forcing with
459 observations at Pougny (downstream of the dam) on the calculation of flow at the outlet
460 of the Rhône River Basin. The gaging station a Pougny is the outlet of the "Rhône
461 upstream" basin in Figure 5b) and is also shown on Figure 11. The first year (August 1st
462 1995 – July 31st 1996) was used for this experiment. Forcing with observations at Lake
463 Geneva increases the Nash Efficiency from 0.49 to 0.62 at Beaucaire, the outlet of the
464 Rhône basin.
465

4. Conclusions

The river routing in SIM-France is done by MODCOU which uses groups of cells called isochrone zones for its computations and does not directly compute flow and volume of water for each cell of its quad-tree river network. The use of isochrones limits the flexibility in the number and location of river gages. The work in this paper presents the replacement of the river routing module in MODCOU by the river network model called RAPID. Information on the network connectivity between the quad-tree river cells of SIM-France is readily available in tables that relate upstream and downstream cells. These tables can be used directly to create the network matrix of RAPID. A ten-year study of river flow in Metropolitan France is presented comparing RAPID and the routing module of MODCOU. An automated procedure for determining optimal model parameters is available in RAPID and various options for the estimation of the parameters are investigated. Sub-basin optimization increases model performance but its influence is much smaller than the choice of the cost function. A cost function was developed that normalizes the square-error between observations at each river gage and RAPID computations by the average flow at the gage. This cost function is found to globally increase the Nash efficiency of computed flow in all gages. We suggest that this is due to the average flow having an influence on the computation of the Nash efficiency. Therefore, the use of an appropriate criterion for quantifying the quality of river flow as the cost function for the optimization procedure helps the betterment of model computations. Also, flow wave celerities included in the temporal parameter of the Muskingum method benefit from taking into account topography when compared to a simple constant celerity formulation. Overall, the computation obtained with the addition

of RAPID are comparable to those of the original river routing module in SIM-France. We consider the addition of RAPID as advantageous since flow and volume of water is directly computed for each cell of the quad-tree river network. The formulation of RAPID allows for easily substituting observed flows for the upstream calculated flow, which is advantageous when considering a man-made infrastructure as was shown for the Rhône River.

Acknowledgments

This work was partially supported by the French *MinesParistech*, by the French *Agence Nationale de la Recherche* under the *Vulnérabilité de la nappe du Rhin (VulNaR)* project, by the French *Programme Interdisciplinaire de Recherche sur l'Environnement de la Seine (PIREN-Seine)* project, by the U.S. National Aeronautics and Space Administration under the Interdisciplinary Science Project NNX07AL79G, by the U.S. National Science Foundation under project EAR-0413265: CUAHSI Hydrologic Information Systems, and by the American Geophysical Union under a Horton (Hydrology) Research Grant.

Thank you to Pere Quintana-Seguí for discussion about and for sharing data corresponding to his work on SIM-France. Thank you to Firas Saleh and to Céline Monteil for discussions on travel time within SIM-France. The authors are thankful to the two anonymous reviewers and to the editor for their valuable comments and suggestions that helped improved the original version of this manuscript.

510 **References**

511 Abdulla, F. A., D. P. Lettenmaier, E. F. Wood, and J. A. Smith (1996), Application of a
512 macroscale hydrologic model to estimate the water balance of the Arkansas Red River
513 basin, *Journal of Geophysical Research-Atmospheres*, *101*, 7449-7459.
514 Boone, A., J.-C. Calvet, J. Noilhan, and euml (1999), Inclusion of a Third Soil Layer in a
515 Land Surface Scheme Using the Force-Restore Method, *Journal of Applied Meteorology*,
516 *38*, 1611-1630.
517 Cunge, J. A. (1969), On the subject of a flood propagation computation method
518 (Muskingum method), *Journal of Hydraulic Research*, *7*, 205-230.
519 David, C. H., D. R. Maidment, G.-Y. Niu, Z.-L. Yang, F. Habets, and V. Eijkhout (2011),
520 River network routing on the NHDPlus dataset, *Journal of Hydrometeorology*, *tentatively*
521 *accepted on 18 January 2011*.
522 Decharme, B., H. Douville, A. Boone, F. Habets, and J. Noilhan (2006), Impact of an
523 Exponential Profile of Saturated Hydraulic Conductivity within the ISBA LSM:
524 Simulations over the RhÃ´ne Basin, *Journal of Hydrometeorology*, *7*, 61-80.
525 Ducharne, A., C. Golaz, E. Leblois, K. Laval, J. Polcher, E. Ledoux, and G. de Marsily
526 (2003), Development of a high resolution runoff routing model, calibration and
527 application to assess runoff from the LMD GCM, *Journal of Hydrology*, *280*, 207-228.
528 Durand, Y., E. Brun, L. Mérindol, G. Guyomarc'h, B. Lesaffre, and E. Martin (1993), A
529 meteorogical estimation of relevant parameters for snow models, *Annals of Glaciology*,
530 *18*, 65-71.
531 Habets, F., A. Boone, J. L. Champeaux, P. Etchevers, L. Franchisteguy, E. Leblois, E.
532 Ledoux, P. Le Moigne, E. Martin, S. Morel, J. Noilhan, P. Q. Segui, F. Rousset-

533 Regimbeau, and P. Viennot (2008), The SAFRAN-ISBA-MODCOU
534 hydrometeorological model applied over France, *Journal of Geophysical Research-*
535 *Atmospheres*, *113*.
536 Leblanc, D., and J. P. Villeneuve (1978), Algorithme de schématisation des écoulements
537 d'un bassin versant 1-55 pp, Institut national de la recherche scientifique, Québec.
538 Leblois, E., and E. Sauquet (2000), Grid elevation models in hydrology. Part 2:
539 HydroDem technical note, 80 pp, Cemagref, Lyon.
540 Ledoux, E., G. Girard, G. de Marsily, J. P. Villeneuve, and J. Deschenes (1989), Spatially
541 Distributed Modeling: Conceptual Approach, Coupling Surface Water and Groundwater,
542 in *Unsaturated Flow in Hydrologic Modeling Theory and Practice*, edited by H. J.
543 Morel-Seytoux, pp. 435-454, Kluwer Academic Publishers.
544 Lohmann, D., R. NolteHolube, and E. Raschke (1996), A large-scale horizontal routing
545 model to be coupled to land surface parametrization schemes, *Tellus Series a-Dynamic*
546 *Meteorology and Oceanography*, *48*, 708-721.
547 Lohmann, D., D. P. Lettenmaier, X. Liang, E. F. Wood, A. Boone, S. Chang, F. Chen, Y.
548 J. Dai, C. Desborough, R. E. Dickinson, Q. Y. Duan, M. Ek, Y. M. Gusev, F. Habets, P.
549 Irannejad, R. Koster, K. E. Mitchell, O. N. Nasonova, J. Noilhan, J. Schaake, A.
550 Schlosser, Y. P. Shao, A. B. Shmakin, D. Verseghy, K. Warrach, P. Wetzel, Y. K. Xue,
551 Z. L. Yang, and Q. C. Zeng (1998a), The Project for Intercomparison of Land-surface
552 Parameterization Schemes (PILPS) phase 2(c) Red-Arkansas River basin experiment: 3.
553 Spatial and temporal analysis of water fluxes, *Global and Planetary Change*, *19*, 161-
554 179.

555 Lohmann, D., E. Raschke, B. Nijssen, and D. P. Lettenmaier (1998b), Regional scale
556 hydrology: I. Formulation of the VIC-2L model coupled to a routing model, *Hydrological*
557 *Sciences Journal-Journal Des Sciences Hydrologiques*, *43*, 131-141.
558 Lohmann, D., E. Raschke, B. Nijssen, and D. P. Lettenmaier (1998c), Regional scale
559 hydrology: II. Application of the VIC-2L model to the Weser River, Germany,
560 *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *43*, 143-158.
561 Lohmann, D., K. E. Mitchell, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, B. A.
562 Cosgrove, J. Sheffield, Q. Y. Duan, L. F. Luo, R. W. Higgins, R. T. Pinker, and J. D.
563 Tarpley (2004), Streamflow and water balance intercomparisons of four land surface
564 models in the North American Land Data Assimilation System project, *Journal of*
565 *Geophysical Research-Atmospheres*, *109*, 1-22.
566 Maurer, E. P., G. M. O'Donnell, D. P. Lettenmaier, and J. O. Roads (2001), Evaluation of
567 the land surface water budget in NCEP/NCAR and NCEP/DOE reanalyses using an off-
568 line hydrologic model, *Journal of Geophysical Research-Atmospheres*, *106*, 17841-
569 17862.
570 Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models
571 part I -- A discussion of principles, *Journal of Hydrology*, *10*, 282-290.
572 Nelder, J. A., and R. Mead (1965), A Simplex Method for Function Minimization, *The*
573 *Computer Journal*, *7*, 308-313.
574 Ngo-Duc, T., T. Oki, and S. Kanae (2007), A variable streamflow velocity method for
575 global river routing model: model description and preliminary results, *Hydrol. Earth Syst.*
576 *Sci. Discuss.*, *4*, 4389-4414.

577 Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood (1997),
578 Streamflow simulation for continental-scale river basins, *Water Resources Research*, *33*,
579 711-724.
580 Noilhan, J., and S. Planton (1989), A Simple Parameterization of Land Surface Processes
581 for Meteorological Models, *Monthly Weather Review*, *117*, 536-549.
582 Oki, T., and Y. C. Sud (1998), Design of Total Runoff Integrating Pathways (TRIP) - A
583 Global River Channel Network, *Earth Interactions*, *2*, 1-35.
584 Olivera, F., J. Famiglietti, and K. Asante (2000), Global-scale flow routing using a
585 source-to-sink algorithm, *Water Resources Research*, *36*, 2197-2207.
586 Quintana-Seguí, P., P. Le Moigne, Y. Durand, E. Martin, F. Habets, M. Baillon, C.
587 Canellas, L. Franchisteguy, and S. Morel (2008), Analysis of near-surface atmospheric
588 variables: Validation of the SAFRAN analysis over France, *Journal of Applied*
589 *Meteorology and Climatology*, *47*, 92-107.
590 Quintana-Seguí, P., E. Martin, F. Habets, and J. Noilhan (2009), Improvement,
591 calibration and validation of a distributed hydrological model over France, *Hydrol. Earth*
592 *Syst. Sci.*, *13*, 163-181.
593 Saleh, F. S. M., N. Flipo, F. Habets, A. Ducharne, L. Oudin, M. Poulin, P. Viennot, and
594 E. Ledoux (2010), CONTRIBUTION OF 1D RIVER FLOW MODELING TO THE
595 QUANTIFICATION OF STREAM-AQUIFER INTERACTIONS IN A REGIONAL
596 HYDROLOGICAL MODEL, paper presented at XVIII International Conference on
597 Computational Methods in Water Resources (CMWR 2010)
598 CMWR 2010, Barcelona.

599 Saleh, F. S. M., N. Flipo, F. Habets, A. Ducharne, L. Oudin, P. Viennot, and E. Ledoux
600 (2011), Modeling the impact of in-stream water level fluctuations on stream-aquifer
601 interactions at the regional scale, *Journal of Hydrology*, *accepted*.
602 SCHAPI (2008), Banque HYDRO, Service Central d'Hydrométéorologie et d'Appui à la
603 Prévision des Inondations available online at http://www.hydro.eaufrance.fr/index.php.
604 USEPA, and USGS (2007), NHDPlus User Guide, available online at
605 http://www.horizon-systems.com/nhdplus/documentation.php.
606 Wetzel, S. (1994), A hydrological model for predicting the effects of climate change, 85
607 pp, Princeton University, Princeton.
608
609

610 Table 1 Average efficiencies and average root mean square errors computed for MODCOU and for RAPID with 7 different
611 sets of parameters. The best 382 values are used for 1995-1996 and the best 427 values are used for 1995-2005.

				1995-1996		1995-2005	
				Best 382 values		Best 427 values	
Vector of parameters used in optimization	Optimization cost function	Spatial optimization	Model	Average efficiency	Average RMSE (m^3/s)	Average efficiency	Average RMSE (m^3/s)
N/A	N/A	N/A	MODCOU	0.617	8.37	0.650	12.67
ρ	ϕ_1	France	RAPID	0.581	8.85	0.614	13.64
ρ	ϕ_2	France	RAPID	0.611	8.39	0.638	13.07
ρ	ϕ_2	7 basins	RAPID	0.615	8.40	0.640	13.06
ρ	ϕ_2	20 sub-basins	RAPID	0.615	8.38	0.637	13.04
σ	ϕ_1	France	RAPID	0.602	8.63	0.632	13.25
σ	ϕ_2	France	RAPID	0.620	8.37	0.647	12.91
σ	ϕ_2	7 basins	RAPID	0.624	8.32	0.646	12.92

612

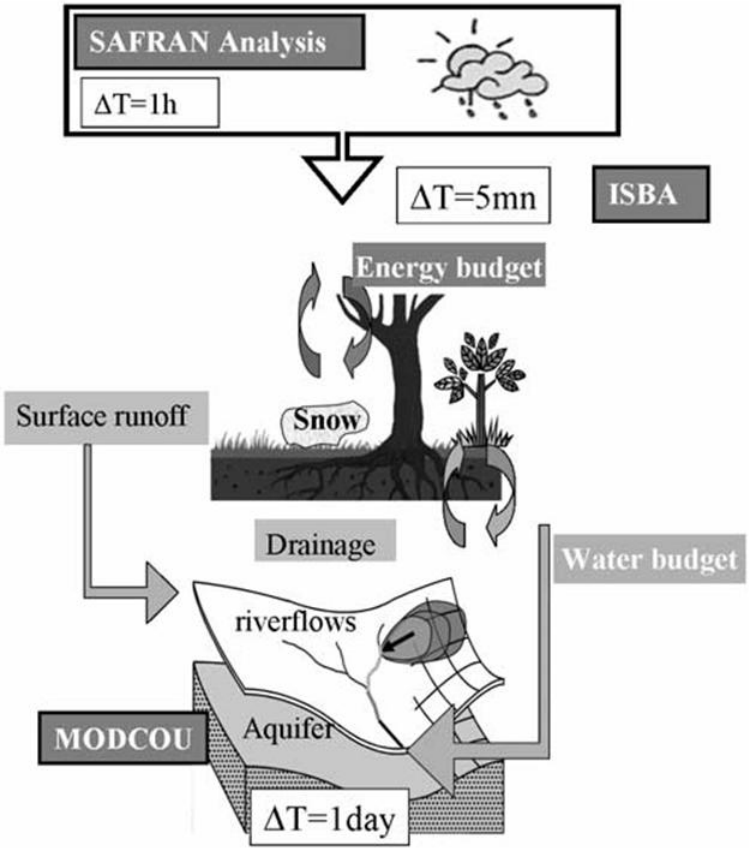
613 Table 2 Results of optimization procedure using ρ parameters and the ϕ_2 cost function

Basin	Sub-basin	Number of river cells	Number of stations	Observa-bility ratio	Optimized λ_k	Optimized λ_x	Basin	Sub-basin	Number of river cells	Number of stations	Observa-bility ratio	Optimized λ_k	Optimized λ_x
France	all basin	24264	291	83.4	0.366	0.237	Loire	Loire downstream	1763	25	70.5	0.436	0.091
Adour	all basin	666	9	74.0	0.375	0.313	Seine	all basin	5115	41	124.8	0.531	0.234
Garonne	all basin	2985	58	51.5	0.294	0.009	Seine	Seine upstream	2919	30	97.3	0.579	0.145
Garonne	Garonne upstream	558	5	111.6	0.160	0.420	Seine	Oise	889	10	88.9	0.469	3.766
Garonne	Tarn	356	8	44.5	0.152	0.674	Seine	Seine downstream	1307	1	1307.0	0.031	4.984
Garonne	Lot	369	10	36.9	0.394	0.113	Meuse	all basin	832	3	277.3	0.383	0.059
Garonne	Dordogne	431	12	35.9	0.356	0.056	Rhône	all basin	3426	51	67.2	0.256	0.118
Garonne	Garonne downstream	1271	23	55.3	0.375	0.313	Rhône	Saône	1043	32	32.6	0.236	0.007
Loire	all basin	4138	88	47.0	0.414	0.197	Rhône	Ardèche	66	3	22.0	0.000	0.156
Loire	Vienne	706	20	35.3	0.386	0.145	Rhône	Rhône upstream	279	1	279.0	0.500	4.750
Loire	Allier	458	17	26.9	0.308	2.670	Rhône	Rhône downstream	2038	15	135.9	0.403	0.076
Loire	Loire upstream	541	12	45.1	0.391	0.305	Hérault	Hérault	101	3	33.7	0.375	4.813
Loire	Loir	670	14	47.9	0.453	0.273							

614

Captions to illustrations

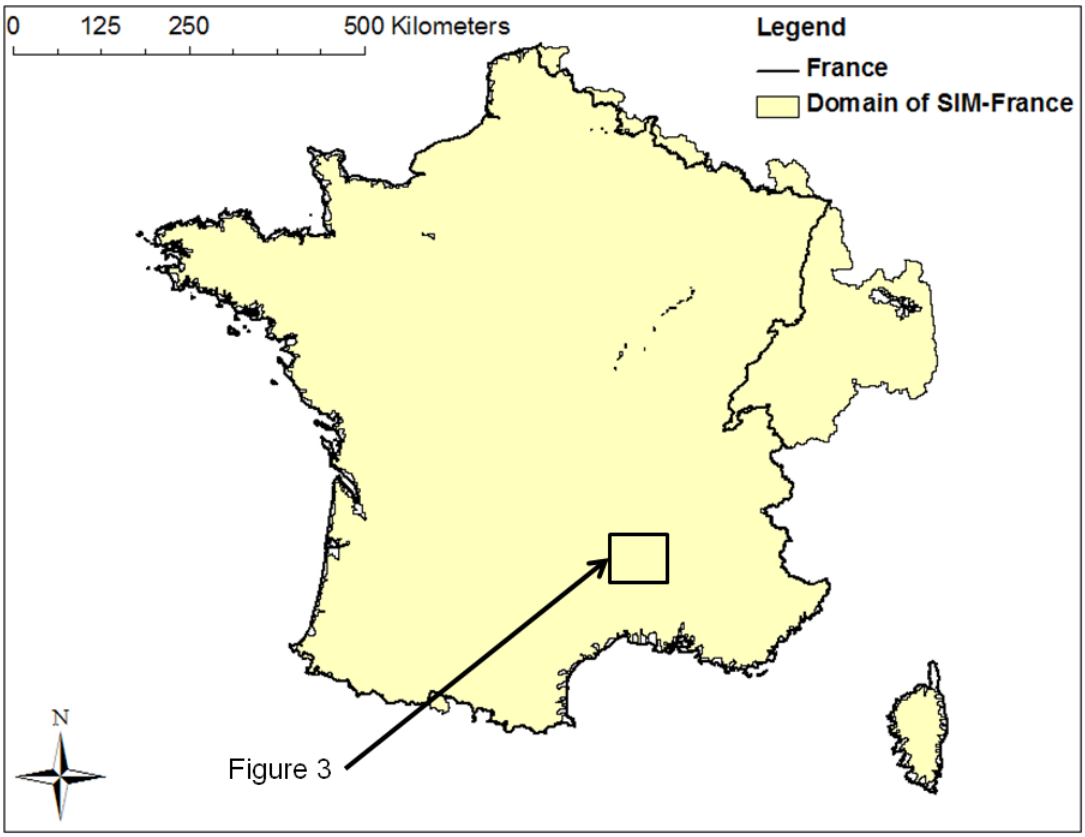
618



619

620 Figure 1 Structure of SIM-France, from Habets et al. [2008]

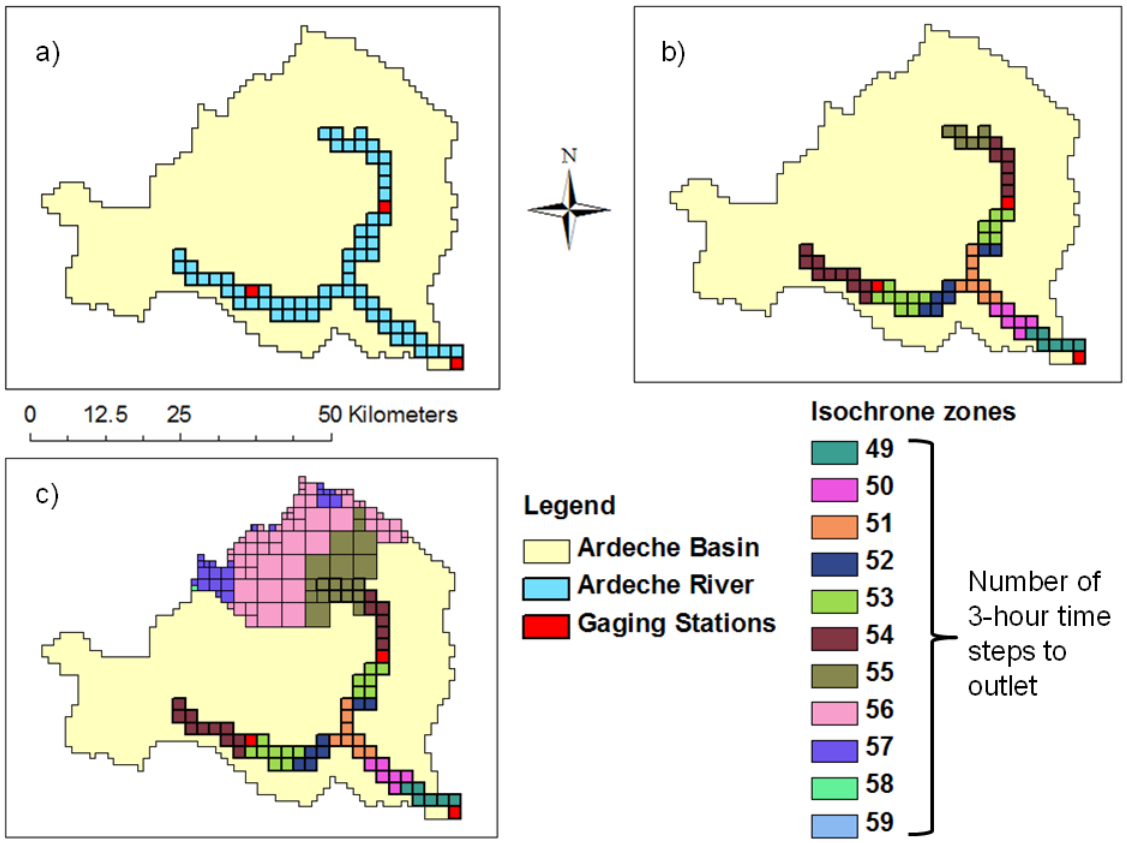
621



622

623 Figure 2 France and computational domain of SIM-France

624



625

626 Figure 3 Surface and river isochrone zones in Ardèche Basin in MODCOU within

627 SIM-France

628

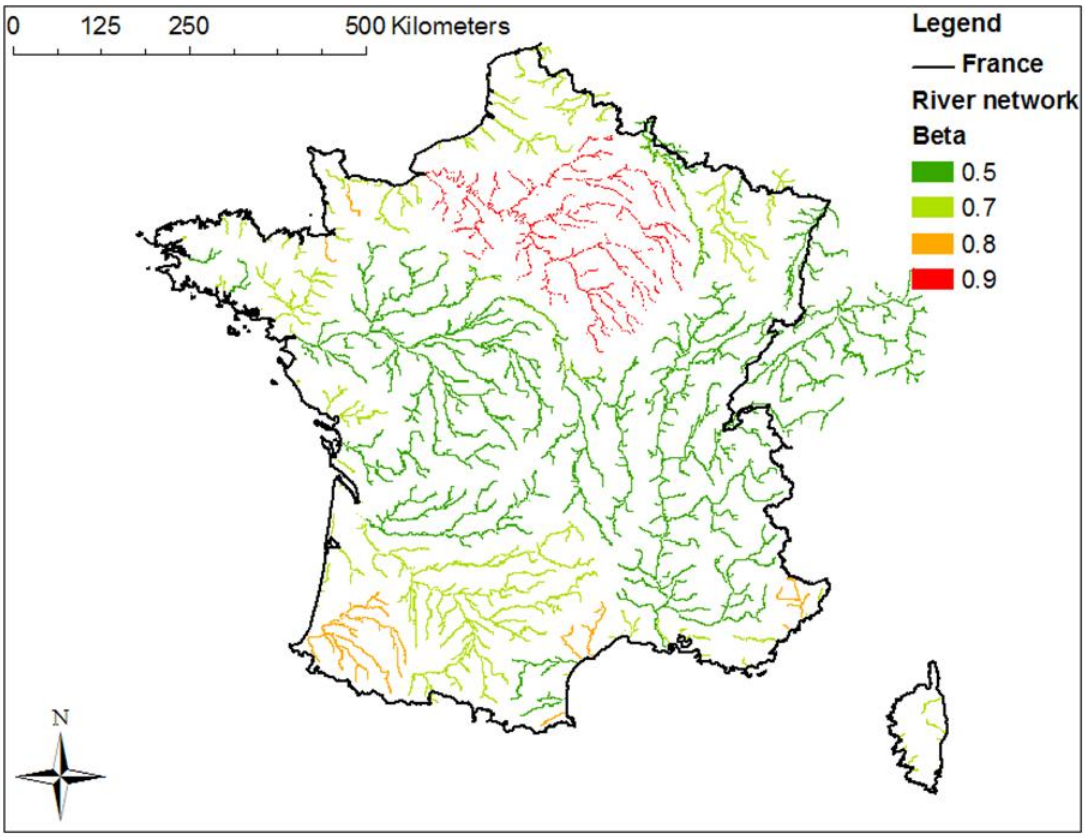
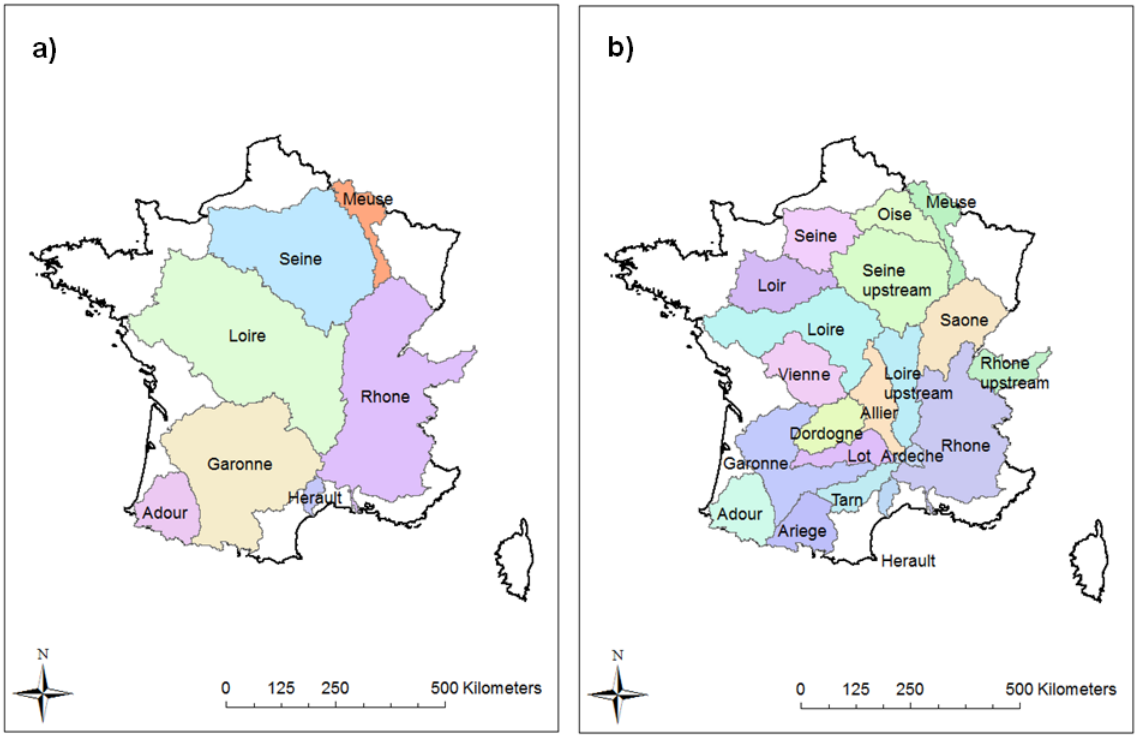


Figure 4 Map of the parameter β used for river routing in MODCOU within SIM-France

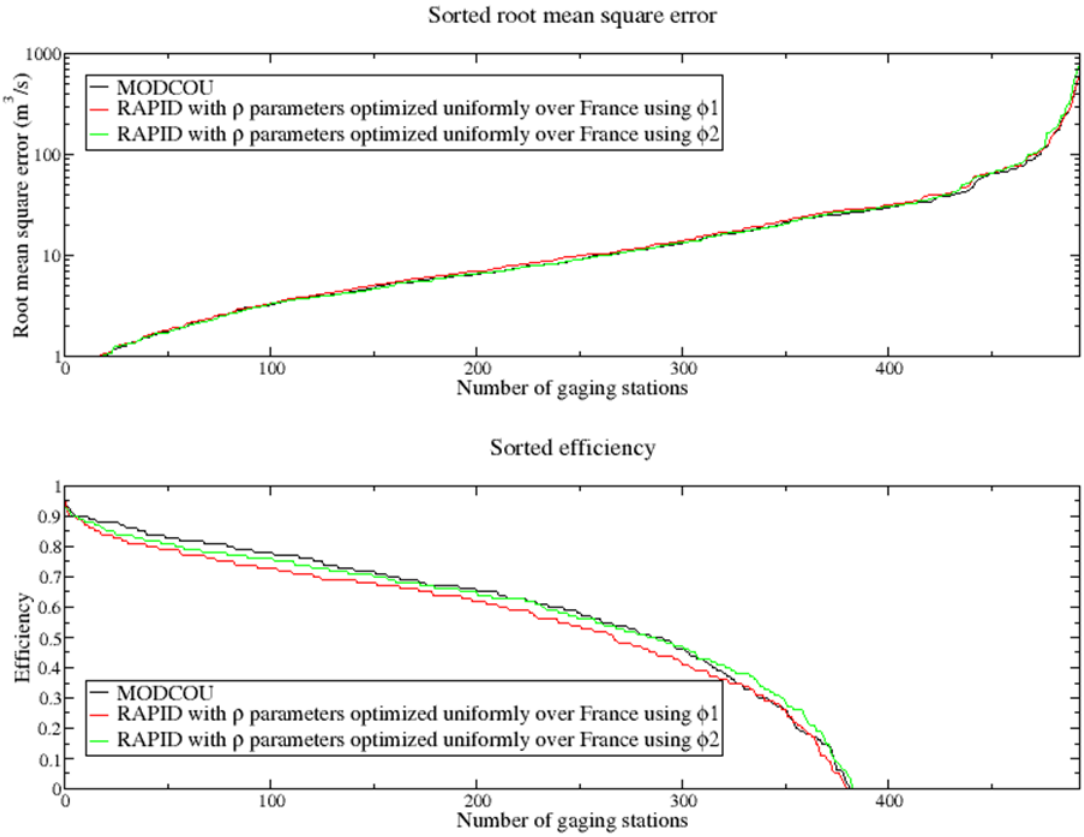


633

634 Figure 5 Basins treated independently during optimization of RAPID parameters in

635 SIM-France. a) Seven major river basins. b) Twenty sub-basins

636



637

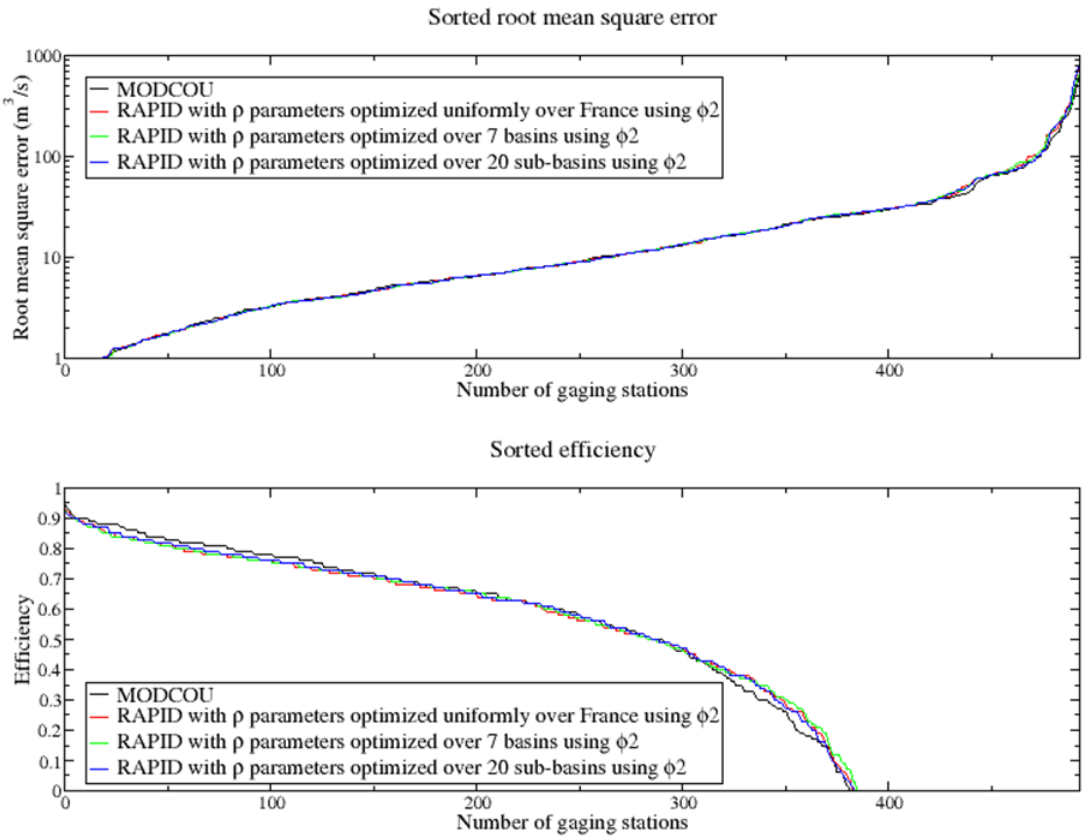
638 Figure 6 Comparison of sorted RMSEs and efficiencies for the year 1995-1996

639 between MODCOU and RAPID with ρ parameters optimized uniformly over France

640 using the original cost function ϕ_1 and using the new cost function ϕ_2

641

642

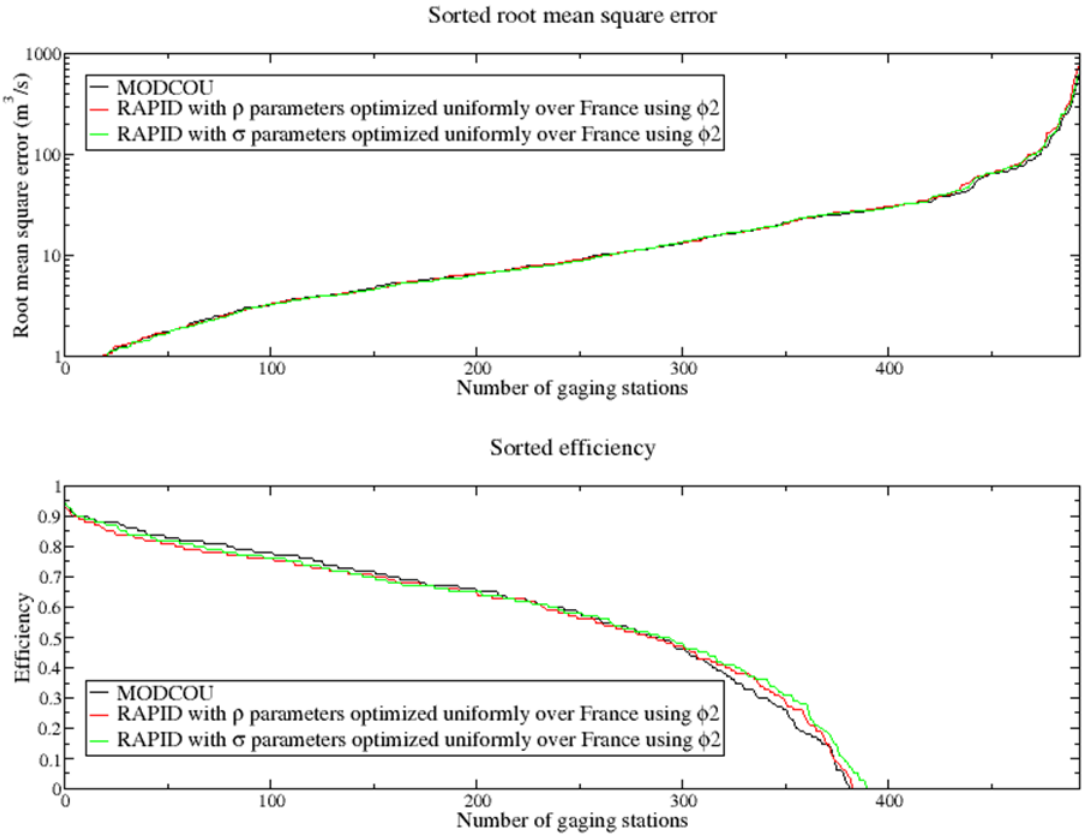


643

644 Figure 7 Effect of sub-basin optimization for RAPID on RMSEs and efficiencies

645 for the year 1995-1996 with ρ parameters using the new cost function ϕ_2

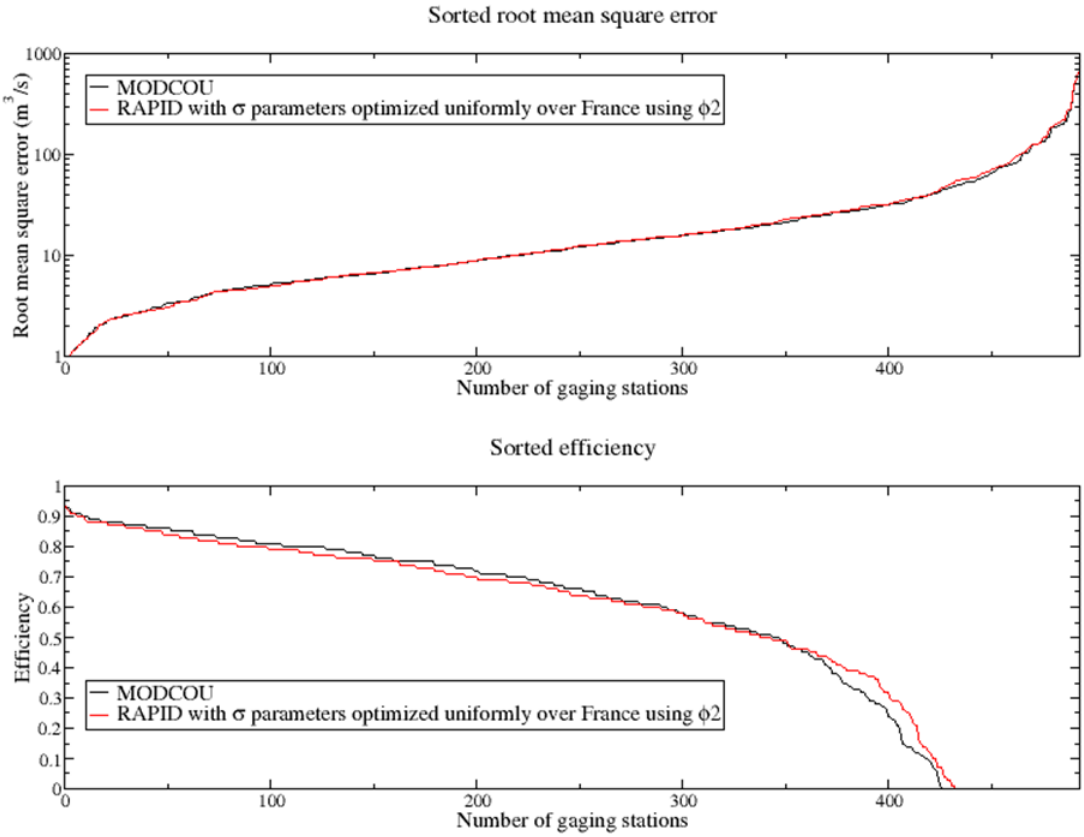
646



647

648 Figure 8 Effect of set of parameters ρ and σ for RAPID on RMSEs and efficiencies

649 for the year 1995-1996 using the new cost function ϕ_2 uniformly over France

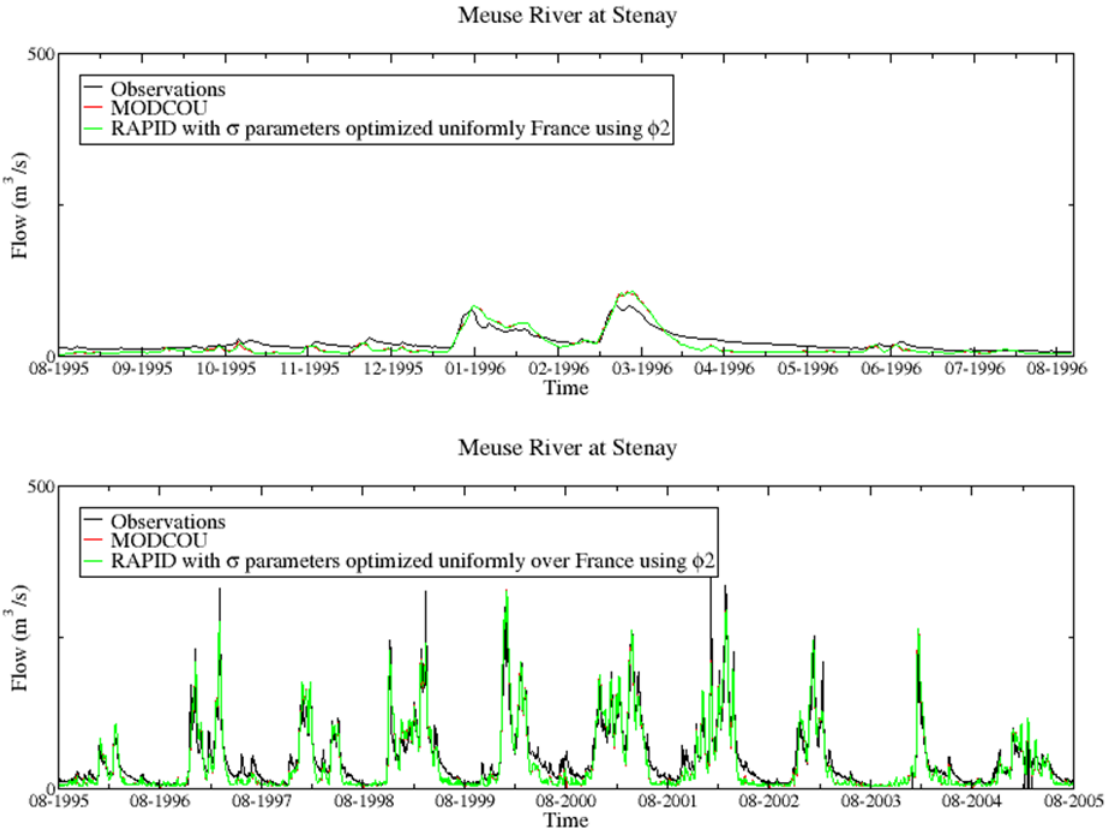


650

651 Figure 9 Comparison of sorted RMSEs and efficiencies for the years 1995-2005

652 between MODCOU and RAPID with σ parameters optimized uniformly over France

653 using the new cost function ϕ_2

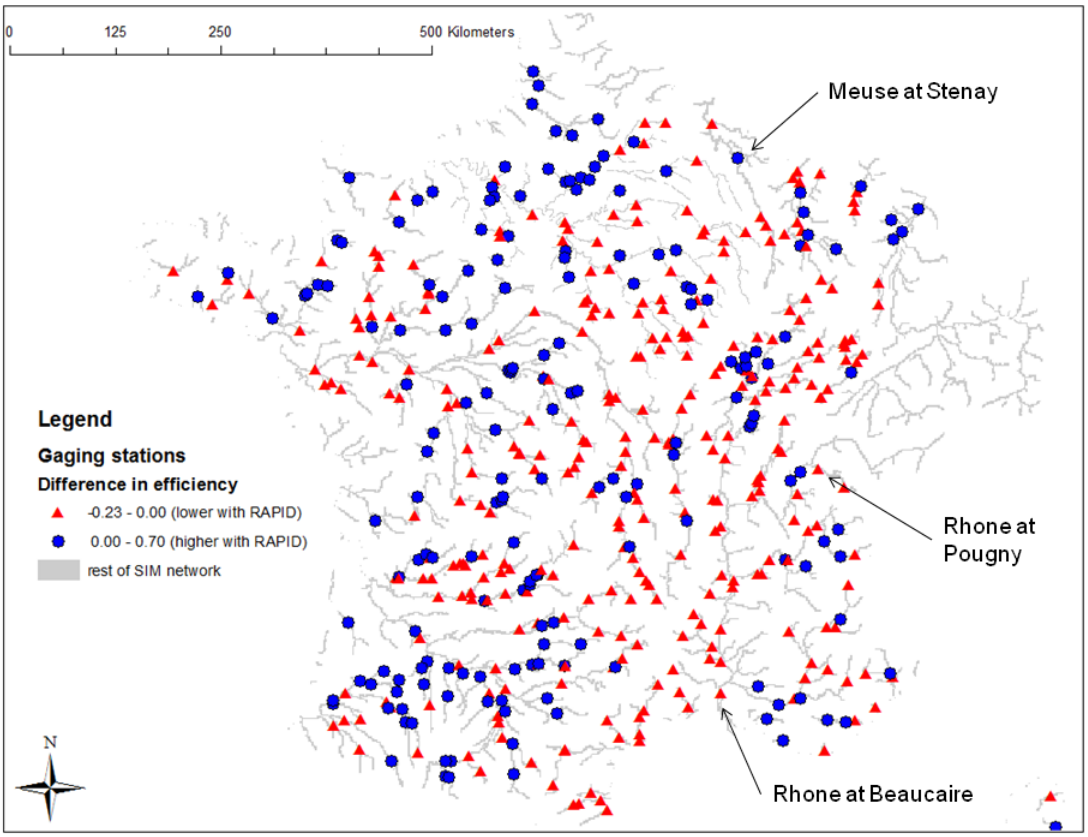


654

655 Figure 10 Comparison of 1995-1996 and 1995-2005 hydrographs for the Meuse

656 River at Stenay obtained by MODCOU and RAPID with σ parameters optimized

657 uniformly over France using the new cost function ϕ_2



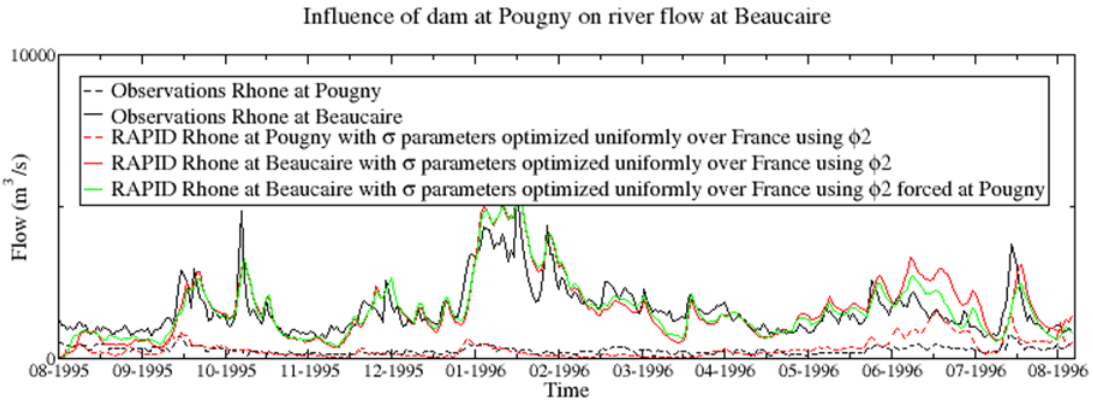
658

659 Figure 11 Spatial difference of efficiencies obtained for the years 1995-2005

660 between RAPID using σ parameters optimized uniformly over France using the new cost

661 function ϕ_2 and MODCOU

662



663

664 Figure 12 Comparison of RAPID discharge calculation at the outlet of the Rhône

665 River (at Beaucaire) with and without forcing at the outlet of Lake Geneva (at Pougny).