Robustness and performance of semi-distributed (SWAT) and global (GR4J) hydrological models throughout an observed climatic shift over contrasted French watersheds

Etienne Brulebois  
_Biogéosciences, UMR n°6282 CNRS University of Bourgogne Franche-Comté, etienne.brulebois@gmail.com_

Marjorie Ubertosi  
_Agroécologie, UMR n°1347 AgroSup Dijon, INRA, University of Bourgogne Franche-Comté, marjorie.ubertosi@agrosupdijon.fr_

Thierry Castel  
_Biogéosciences, UMR n°6282 CNRS, University of Bourgogne Franche-Comté, 2A2E Departement, Agrosup Dijon, thierry.castel@u-bourgogne.fr_

Yves Richard  
_Biogéosciences, UMR n°6282 CNRS, University of Bourgogne Franche-Comté, yves.richard@u-bourgogne.fr_

Sabine Sauvage  
_Ecolab, UMR n°5245 CNRS, INPT, ENSAT, University Paul Sabatier, sabine.sauvage@univ-tlse3.fr_

BYU ScholarsArchive Citation  
Brulebois, Etienne; Ubertosi, Marjorie; Castel, Thierry; Richard, Yves; Sauvage, Sabine; Sanchez Perez, José-Miguel; Le Moine, Nicolas; and Amiotte-Suchet, Philippe (2018) "Robustness and performance of semi-distributed (SWAT) and global (GR4J) hydrological models throughout an observed climatic shift over contrasted French watersheds," Open Water Journal: Vol. 5 : Iss. 1 , Article 4.  
Available at: https://scholarsarchive.byu.edu/openwater/vol5/iss1/4
Robustness and performance of semi-distributed (SWAT) and global (GR4J) hydrological models throughout an observed climatic shift over contrasted French watersheds

Cover Page Footnote
The authors thank Météo-France who provided the precipitation data, as part of an agreement with the University of Burgundy. They also thank the DREAL Bourgogne, especially Marc Philippe, who transferred the runoff time series.

Authors
Etienne Brulebois, Marjorie Ubertosi, Thierry Castel, Yves Richard, Sabine Sauvage, José-Miguel Sanchez Perez, Nicolas Le Moine, and Philippe Amiotte-Suchet

This article is available in Open Water Journal: https://scholarsarchive.byu.edu/openwater/vol5/iss1/4
Robustness and performance of semi-distributed (SWAT) and global (GR4J) hydrological models throughout an observed climatic shift over contrasted French watersheds

Etienne Brulebois¹, Marjorie Ubertosi², Thierry Castel¹, Yves Richard¹, Sabine Sauvage³, José-Miguel Sanchez Perez³, Nicolas Le Moine⁴, Philippe Amiotte-Suchet¹

¹UMR n°6282, CNRS University of Bourgogne Franche-Comté
²UMR n°1347 AgroSup Dijon, INRA, University of Bourgogne Franche-Comté
³UMR n°5245 CNRS, INPT, ENSAT, University Paul Sabatier
⁴UMR n°7619 CNRS, EPHE, University Pierre & Marie Curie

*Corresponding Author: etienne.brulebois@u-bourgogne.fr

ABSTRACT

Many studies about climate change impacts assessment are published every year. These studies commonly use a hydroclimatic modelling chain, whose principle is to feed impact models with climate models outputs. An important step in this process is to test the validity of impact models in a climate change context. However, this step is not frequently applied. The aim of this study is to test the robustness of two hydrological models with distinct conceptualizations: a global and empirical model (GR4J) and a semi-distributed and physically-based model (SWAT). They both have been calibrated and validated over climate contrasted periods. Despite a higher decrease of performance between calibration and validation for the GR4J model, both of them show relative robustness. Moreover, the stability of parameters between the two calibration periods shows that their value are not much influenced by the climate of the calibration period, and consequently remains valid during the entire projection period.

Keywords
Hydrological models, robustness, climate change, SWAT, GR4J

1. Introduction

In the past few decades, more and more studies have been dealing with climate change impacts on hydrological cycle (Huntington, 2006; Jiménez Cisneros et al., 2014). Links between global warming and hydrological cycle modification can be assessed by two main ways: comparisons between observed time series, where the authors highlight trends and/or correlation between hydrological and climate variability (Dai et al., 2009; Gedney et al., 2006; Gerten et al., 2008; Milly et al., 2005), and prospective studies, where future simulated climate data supply hydrological models to analyze their influences on water quality (Boorman, 2003; Ducharme,
2008; Rehana and Mujumdar, 2012; Whitehead et al., 2009) as well as water quantity (Arnell, 1999; Feyen and Dankers, 2009; Henrichs, T. et al., 2002; Lehner et al., 2006). Modeling approach is essential for a better understanding of interactions between the numerous processes involved in the water resources (Kroeze et al., 2012). Over France, several prospective works have studied impacts of climate change on water resources, but they mainly focused on large watersheds: Rhône river (Etchevers et al., 2002; Otlé et al., 2001), Seine river (Boë et al., 2007; Ducharne et al., 2007; Habets et al., 2011), Garonne river (Sauquet et al., 2010; Tisseuil et al., 2010, Grusson, 2016) or Loire river (Ducharne et al., 2010), or even over the entire French metropolitan territory (Boë, 2007; MEDDE, 2012b).

Nevertheless, filling the gap of knowledge of climate change impacts over moderate-size watersheds is a major concern and more regional impact studies should be performed at this scale (Bates et al., 2008). In this context, a prospective study of climate change impacts at watershed scale has been performed over northeastern French watersheds, using spatial disaggregation of general circulation model (GCM) data ( Rossi et al., 2014) to feed hydrological models (Brulebois et al., 2015a, 2014;Legras, 2014) at daily scale throughout the entire 21st century. Two catchment hydrological models with distinct conceptualization were chosen: GR4J (global and empirical model) and SWAT (semi-distributed and physical-based model). The SWAT model has already been widely used in climate change impacts studies over a large number of watersheds (Gassman et al., 2007, 2014). GR4J model is less frequently used in climate change context, i.e, to discuss about the 2nd pathology described by Coron et al (2011).

The robustness of the GR4J model in context of climate change was less studied. Although Brigode et al., (2013) pointed out the difficulty for lumped rainfall-runoff models (such as GR4J) to simulate streamflow on periods with contrasted climate, Seiller et al. (2012) concluded, after a comparison between twenty lumped rainfall-runoff models, to the better transposability across periods of GR4J than the others. Furthermore, Le Lay et al. (2007) applied GR4J on the Upper Ouémé watershed (Benin) and highlight the relative stability of GR4J parameters across calibration on several periods. Finally, a study comparing projections of several hydrological models (including GR4J and SWAT) has been conducted by Cornelissen et al. (2013), over the Upper Ouémé and the Térou watersheds (Benin).

According to these authors, both SWAT and GR4J models were judged able to simulate future streamflow by calibrating and validating using contrasted period. But because of its lumped conceptualization, the GR4J model does not contribute to improve the knowledge of watershed processes (Cornelissen et al., 2013).

In order to test the robustness of hydrological models throughout climate contrasted periods, several testing schemes can be applied. Most of the time, the DSST method (Differential Split-Sampling Test) is used (Klemes, 1986b). It consists in calibrating and validating the model with distinct periods, and analyzing the simulation accuracy in validation.
period. However, this method does not guarantee a strong difference in climate and in its hydrological impacts between calibration and validation periods. In this paper, we propose to assess the robustness and GR4J and SWAT models throughout an observed climate shift, impacting the hydrological cycle.

The western Europe has experienced a major change during the last fifty years (Laat and Crok, 2013), which had also been detected over France (Brulebois et al., 2015b), even at Burgundy region scale (Castel et al., 2014; Richard, 2014). Over France, the shift consists in an abrupt increase in both minimal and maximal temperatures in 1987/88. At the annual scale, this increase reaches 0.92°C and 1.1°C for minimal and maximal temperatures respectively. In Burgundy, this increase can locally reach 1.37°C and 1.32°C for minimal and maximal temperatures respectively (Richard, 2014). The hydrological response to this shift has been described by Brulebois et al. (2015b) over France, on a selection of 30 watersheds. They have shown a decrease of annual streamflow of 4%, between 1988-2009 and 1969-1987 periods, although an increase of 5% of precipitations during the same time. This shift gives us the opportunity to assess the transposability of hydrological models throughout an observed climate shift. For this purpose, in this study, both SWAT and GR4J models have been alternatively calibrated and validated on two periods of 8 years, before (1980-1987) and after (1988-1995) the shift. Their robustness and parameters stability have been assessed by comparing the performance decrease between calibration and validation, as well as changes in parameters values between each calibration.

2. Materials and methods

2.1. Study area

Burgundy is located in North-East of France (Figure 1). It covers an area of about 31,500 km² and its elevation varies between 50 m in the North-West to 900 m in the Morvan mountains. It includes several geological contexts: the metamorphic and granitic basement of the Morvan mountains (the northern continuation of the Massif Central), surrounded by sedimentary (mainly...
calcareous) rocks of Jurassic and Quaternary alluvi-ums in the Saône and the Loire valleys. Furthermore, the region is divided between three hydrographical areas, which are the headwater of three main French rivers: Seine, Loire and Rhône basins. Burgundy has a well-developed hydrographical network but unequally distributed, because of the spatial heterogeneity of the geological substratum. It is characterized by a high drainage density in the Morvan mountains, with streams supplied by small and superficial groundwater, whereas the surrounding limestones plateau, which are very permeable, show almost no streams because water mainly flows throughout a well-developed karst aquifer.

For these reasons, the water resource in Burgundy is highly fragmented, and could be seriously impacted by climate change. Burgundy climate is mainly semi-continental, with moderate Mediterranean influences in the South, and oceanic influences in the West (Chabin, 2001; Cuccia, 2008). The average annual precipitations during the 1981-2010 period are equal to 850 mm, but show strong regional differences, ranging from 600 to 700 mm in the North, 700 to 900 mm in the Saône valley, and locally rises to 1500 mm in the Morvan mountains. The average annual mini-

### 2.2 Hydrometric stations

Daily runoff data were extracted from the “Banque Hydro” database(French Ministry of Ecology, Sustainable Development and Energy) for 8 selected hydrometric stations from 1980-1995. This selection of 8 hydrometric stations (Table 1) was based on several criteria:
- Data reliability according to the station managers;
- Climatic, land-use and geological representative-ness of Burgundy;
- Distribution between the three main hydrographi-cal areas (Loire, Rhône and Seine basins)
- Data availability before and after the observed temperature shift (1987/88).

### 2.3. Hydrological models

#### 2.3.1 The SWAT model

SWAT is a physically-based and semi-distributed agro-hydrological model, operating at watershed scale, on a daily time step (Arnold et al., 1998). It allows to

<table>
<thead>
<tr>
<th>№</th>
<th>River</th>
<th>Outlet</th>
<th>Area (km²)</th>
<th>Module (m³/s)</th>
<th>Altitude (m) min/mean/max</th>
<th>Geological context</th>
<th>Main land-use</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Serein</td>
<td>Chablis</td>
<td>1119</td>
<td>8.3</td>
<td>129 / 310 / 596</td>
<td>Calcareous</td>
<td>Agricultural area</td>
</tr>
<tr>
<td>2</td>
<td>Armançon</td>
<td>Quincy-le-Vicomte</td>
<td>483</td>
<td>3.5</td>
<td>220 / 357 / 525</td>
<td>Calcareous</td>
<td>Permanent pasture</td>
</tr>
<tr>
<td>3</td>
<td>Tille</td>
<td>Arceau</td>
<td>846</td>
<td>7.9</td>
<td>225 / 396 / 591</td>
<td>Calcareous</td>
<td>Forest</td>
</tr>
<tr>
<td>4</td>
<td>Nohain</td>
<td>Saint-Martin-sur-Nohain</td>
<td>473</td>
<td>3.5</td>
<td>157 / 237 / 382</td>
<td>Calcareous</td>
<td>Agricultural area</td>
</tr>
<tr>
<td>5</td>
<td>Nièvre d’Arzembouy</td>
<td>Poiseux</td>
<td>224</td>
<td>2.5</td>
<td>88 / 282 / 599</td>
<td>Calcareous</td>
<td>Permanent pasture</td>
</tr>
<tr>
<td>6</td>
<td>Seille</td>
<td>Saint-Usuge</td>
<td>937</td>
<td>14.9</td>
<td>186 / 289 / 570</td>
<td>Alluvions</td>
<td>Agricultural area</td>
</tr>
<tr>
<td>7</td>
<td>Arroux</td>
<td>Rigny-sur-Arroux</td>
<td>2277</td>
<td>23.5</td>
<td>233 / 407 / 897</td>
<td>Metamorphic rocks</td>
<td>Permanent pasture</td>
</tr>
<tr>
<td>8</td>
<td>Bourbince</td>
<td>Vitry-en-Charollais</td>
<td>819</td>
<td>7.9</td>
<td>232 / 308 / 653</td>
<td>Schists, metamorphic</td>
<td>Permanent pasture</td>
</tr>
</tbody>
</table>

Table 1. Main characteristics of the 8 selected watersheds
simulate hydrology, weather, erosion, plant growth, nutrient cycles, land management and stream routing, at a high resolution, by dividing the simulated watershed into a large number of sub-basins, themselves are further divided into hydrological responses units (HRU). An HRU consists of a homogenous combination of land-use classes, soil types and slope classes. The model takes account of spatialized climate information, at the sub-basin scale, and needs daily precipitations (mm), minimum and maximum temperatures (°C), wind speed (m.s-1), relative humidity (%) and solar radiation (MJ.m-²). At the HRU scale and for each time-step, SWAT divides water from rainfalls between soil infiltration and surface runoff, based on the SCS-CN empirical method (USDA-SCS, 1986). The soil is considered as a multi-layer store from where water can be evaporated or transpired by plants according to potential evapotranspiration (PET) value. Among different formulations of PET proposed by SWAT, the Penman-Monteith method (Penman, 1948), which is the most commonly used one, was chosen in this study.

A percolation function brings soil water from the lowest soil layer to the shallow aquifer, where it can be re-evaporated, be drained toward the streamflow, or supply the deep aquifer.

In this study, the SWAT model has been implemented on watersheds using the Arcview GIS interface for SWAT (ArcSWAT).

### 2.3.2 The GR4J model

The GR4J model is a global rainfall-runoff model, developed by the national institute of research in sciences and technologies for environment and agriculture (IRSTEA) (Perrin, 2007). The model is few parameterized with only 4 parameters to calibrate, and includes 2 reservoirs (production and routing) and 2 transfer functions to represent the watershed processes. The model needs daily precipitation and PET to simulate streamflow at the outlet of the watershed. The four parameters are:

- The daily maximum capacity of the production store (X1, in mm);
- The groundwater exchange coefficient (X2, in mm/day), which allows water to be imported (X2>0) or exported (X2<0) from the system;
- The daily maximum capacity of the routing store (X3, in mm);
- The time base of unit hydrograph (X4, in days).

The parameters are automatically calibrated, optimizing an objective function, such as the Nash coefficient (Nash et Sutcliffe, 1970) and range between a maximum and a minimum fixed by the user. Here, the boundaries have been provided by the developers of the model (Perrin et al., 2003).

### 2.4. Input data

Topography (25 m Digital Elevation Model), land-use information and soil types have been provided by the French Geographical Institute (IGN), the Corine Land Cover 2006 database (1/100.000) and the Infosol database from the French National Institute of the Agronomical Research (INRA, 1998) respectively.

Daily precipitations (mm) have been provided by the MeteoFrance Station Network (MFSN) on the 1961-2011 period. These observed precipitations have been then re-interpolated on a 12 km-grid, in order to get spatialized information over each watershed.

Relative humidity (%), minimal and maximal temperature (°C), wind at 2 m above the ground (m/s), and solar radiation (MJ/m²) have been provided at a daily time step. by the dynamical disaggregation of ERA-INTERIM reanalysis (Simmons et al., 2006) using the regional climate model ARW/WRF (Skamarock et al., 2008) implemented over Burgundy (Castel et al., 2010 ; Xu et al., 2012). These data have been validated on the present time and used with success for impact studies on water balance (Boulard et al., 2015).

### 2.5. Model implementation

#### 2.5.1 Calibration of SWAT and GR4J model

The first step in calibration and validation process of the SWAT model is the identification of sensitive parameters on the implemented watersheds (Arnold et al., 2012a). Such sensitivity analysis can be performed locally (changing values one at a time) or globally (changing values of all parameters). This latter requires a very large number of simulations to be done. Here, a local sensitivity analysis was performed, highlighting the influence of six parameters (Table 2) on the streamflow simulations. These parameters belong to the most sensitive (Nossent, 2012) and the most commonly used (Arnold et al., 2012a) parameters in water cycle calibration. A brief description of each parameter is given below.

The SURLAG coefficient controls the fraction of the total surface runoff which is held in a surface runoff storage before reaching the main channel.
The Curve Number depends on the soil permeability, land-use and antecedent soil water conditions, and controls the fraction of infiltrated water from precipitations. The base flow alpha factor corresponds to the groundwater flow response to changes in recharge. A low value (near 0) indicates a very slow response, while a high value (near 1) indicates a very quick response.

The groundwater delay time is the required time for the water to move from the lowest depth of the soil profile to the shallow aquifer. During this time, water can be removed from the soil by plants or soil uptake. The threshold depth of water corresponds to the water table needed in the shallow aquifer for return flow to occur. As long as the shallow aquifer has not reached this water table, there is no return flow contribution to the streamflow.

The deep aquifer percolation fraction represents the water quantity removed from the shallow aquifer to the deep aquifer (so which goes out of the system). The SURLAG parameter has been fixed for the entire project (for all watersheds) at 0.5, and the CN2 parameter has been decreased by 10% for each HRU with regard to its SWAT default value. The four other parameters have been determined by an automatic calibration using the SWAT-CUP autocalibration program with the “SUFI2” tool (Abbaspour et al., 2007a). The calibration was performed at a daily scale, on the 1977-1995 period, for the 8 watersheds independently, and with 3 years of warm-up period (performance calculate only on the 1980-1995 period). The objective function (O.F) chosen for this calibration was the Nash Efficiency (NSE) criteria (Nash and Sutcliffe, 1970). Only minimum and maximum parameters values are shown (not the values of each watershed) (Table 2).

2.5.2 Model performances in validation on the 1980-2010 period

Model performances have been tested on each calibrated watershed calibrated. SWAT simulating streamflow at sub-basin scale, 28 hydrometric stations available within selected watersheds have been used to validate simulation (Figure 2a). All of the 8 watersheds have shown satisfactory performance (NSE >0.6). At monthly scale, the mean performance (NSE) on the 8 stations reaches 0.83 with SWAT simulation and up to 0.91 with GR4J simulation (Figure 2b).

2.6 Cross calibration/validation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Signification</th>
<th>Default value</th>
<th>Range of calibration</th>
<th>Fixed values (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN2</td>
<td>Curve number (no unit)</td>
<td>Calculated</td>
<td>+ / - 10%</td>
<td>-10%</td>
</tr>
<tr>
<td>SURLAG</td>
<td>Surface runoff lag coefficient (days)</td>
<td>4</td>
<td>0.5-5</td>
<td>0.5</td>
</tr>
<tr>
<td>ALPHA_BF</td>
<td>Baseflow alpha factor (days-1)</td>
<td>0.048</td>
<td>0.01-0.1</td>
<td>0.012-0.099</td>
</tr>
<tr>
<td>RCHRG_DP</td>
<td>Deep aquifer percolation fraction (no unit)</td>
<td>0.01</td>
<td>0-0.5</td>
<td>0-0.45</td>
</tr>
<tr>
<td>GW_DELAY</td>
<td>Groundwater delay time (days)</td>
<td>31</td>
<td>0.5-10</td>
<td>0.9-5.9</td>
</tr>
<tr>
<td>GW_QMIN</td>
<td>Threshold depth of water in the shallow aquifer for return flow to occur (mm)</td>
<td>1000</td>
<td>100-1500</td>
<td>500-1430</td>
</tr>
</tbody>
</table>

Table 2. Description of the most sensitive parameters of the SWAT model in our study, with their default values, range of calibration and minimum and maximum values form the semi-automatic-calibration based on 1000 simulations run of SWAT-CUP.
A Split Sample Test (SST) (Klemes, 1986), which is the most frequently used method (Thirel et al., 2015b) is processed to estimate the dependency of parameters on the climate characteristics. This method consists of a cross-calibration and validation tests of the models on two periods with distinct climate characteristics. Based on this scheme, four modalities have been created to test model performances: C1V1: calibration and validation over the P1 period (1980-1987), C1V2: calibration over P1 and validation over P2 period (1988-1995), C2V2: calibration and validation over P2, C2V1: calibration over P2 and validation over P1.

The analysis of performance decrease between modalities 1 and 2, as well as between modality 3 and 4, describes the model robustness throughout two different hydroclimatic periods. This robustness can also be assessed regarding to the error between simulated and observed streamflow.

During calibration processes, the SWAT-CUP program calculates the 95 percentage prediction uncertainty (95PPU), which represents the distribution of output variable, disallowing 5% of the very bad simulations (Abbaspour et al., 2007b). This uncertainty can be graphically represented as a band, where the thickness varies according to the number of parameters as inputs in the calibration process, and their calibration ranges.

Associated to the 95PPU, two criteria are calculated by SWAT-CUP: the P-factor, which is the percentage of observed data bracketed by the 95PPU, and the R-factor, which is the average thickness of the 95PPU band divided by the standard deviation of the observed data. The P-factor and R-factor are considered satisfactory when their values are higher than 70% and lower than 1.5 respectively (Abbaspour et al., 2007a; Moriasi et al., 2007). A good calibration must show in first satisfactory R and P-factor, then, a good efficiency (NSE or other objective function). The GR4J performance is assessed based on the performance criteria. Finally, the model behavior throughout the climate shift can also be characterized by the stability of its parameters between the two calibrations. Strong changes between parameters values (i.e. a low stability of parameters) indicate a strong dependency on the climate characteristics of the calibration period.

**Figure 2.** Monthly performance (NSE criteria) of a) SWAT model; b) GR4J model, on the selected watersheds.
3. Results

3.1 Identification of a climate shift in inputs and outputs data

Input climate data series have been divided between two sub-periods: the P1 period from 1980-01-01 to 1987-12-31, and P2 period from 1988-01-01 to 1995-12-31. Mean minimal and maximal temperature values calculated on the two sub-periods show significant (t-test) differences (table 3), with a difference about 0.9°C for both minimal and maximal temperatures. These values are very close to those described over France (Brulebois et al., 2015b). However, no significant differences can be observed for the other climate variables.

In response to these changes in temperature, an impact in streamflow is expected. Indeed, annual observed anomaly streamflow (figure 3) is about -90 mm, and except in September and November (which show few variations), all the months show decrease in streamflow, from -10% in September to 50% (July). In comparison to observation, anomalies simulated by GR4J and SWAT are fewer. Anomalies are quite similar between the two models: a decrease of about 50 mm/year. This decrease occurs during 7 months: December, January, and from April to August, and to a lesser extent, in February, while March, September and October show few changes, and November shows a significant increase of about 20%.

3.2 Cross calibration-validation of models

SWAT monthly performances (Table 4 and Figure 4) are very close between each modality (NSE from 0.72 to 0.92), while GR4J performances are more unequal (NSE from 0.27 to 0.97). In C1V1 and C1V2, sometimes SWAT is better, sometimes GR4J is. In C2V1, SWAT is systematically better than GR4J, and in C2V2, GR4J is better. The two models do not show similar behavior in calibration: the best calibration performances are obtained in C1V1 modality for SWAT and C2V2 for GR4J, and in C2V2, GR4J is better. Obviously, GR4J calibration permits to be more adapted than SWAT to the climate characteristics of the calibration period (see performance in C2V2), but in this case, the calibration is less robust across other validation period (see performance in C2V1), while SWAT calibration is relatively robust, since the monthly performance differences are very low between calibration and validation. Moreover, at monthly scale, on the eight watersheds, the mean P-factor reached 73% in C1 and 72% in C2, and the mean R-factor reached 0.73 in C1 and 0.77 in C2. From this point of view, and with monthly NSE higher than 0.80, the SWAT calibrations can be judged good.

Regarding performances for each watershed, the St. Martin/Nohain watershed shows the lowest NSE values and the strongest differences in performance between calibration and validation period for both GR4J and SWAT models. The performance decrease reaches 0.11 for the SWAT model, while it reaches 0.64 for the GR4J model. All the other watersheds show robust performances between calibration and validation for the SWAT model (performances decrease <0.1). Conversely, for GR4J model, all watersheds show low robustness, with a decrease performance greater than 0.1, except for the St. Usuge watershed (Table 4).

3.3 Streamflow reproduction errors

For each modality, errors between observations and simulations (in percentage) averaged on all watersheds are systematically lower for SWAT simulations than for GR4J simulations, and show a similar pattern: C1V1 modality shows the lower error, then, by increasing order, C2V2, C2V1 and C1V2. Errors are however included in the same order of magnitude (from 4.3 to 11.6% for SWAT, and from 5.5 to 17.4% for GR4J). Considering each watershed, errors in C1V1 are lower than these in C1V2 for both SWAT and GR4J models, but this is not the case between C2V2 and C2V1. Streamflow errors are not directly correlated with performance differences. We can see low or no differences in model performance between calibration and validation (SWAT performance on Arceau watershed in C1V2) associated to high streamflow errors (17.9%), and conversely, strong differences in model performance between calibration and validation (GR4J performance on Rigny/Arroux watershed in C2V1), associated to low streamflow errors (-1.7%).

3.4 Stability of parameters and induced uncertainty

The ALPHA_BF, RCHRG_DP and GW_DELAY parameters values remain stable between the 2 sub-periods and show strong and significant correlation between the two calibrations: R=0.94, 0.94 and 0.87 respectively. The GW_QMIN parameter is much less stable (R=0.45).
The ALPHA_BF values show few or no variations between the two calibrations, except for two watersheds (Figure 5a). The GW_DELAY shows an increase for three watersheds but no changes for the others. The QMIN parameter shows an increase for 5 watersheds and a decrease for St. Martin/Nohain and Chablis only. Finally, RCHRG_DP shows a slight increase for 4 watersheds and no changes for the others. All of GR4J parameters are stable between calibration, with a significant correlation coefficient significant higher than 0.75 (Figure 5b). Except for St. Martin/Nohain and Chablis watersheds, changes are slight between the two calibrations. The GR4J parameter values show quite homogenous changes between calibrations: decreases in X2 parameter, except for 2 watersheds, no changes or slight decreases in X3 parameter, few variations in X4 parameter, and very slight increases in X1 parameter, except for the Nohain watershed.

The parameters values fixed during each calibration have been set as inputs in a new SWAT-CUP.
run, to obtain the 95PPU induced by the parameters changes. This corresponds to an estimation of the uncertainty due to climate characteristics of the calibration period. At monthly scale, the R-factor averaged over all watersheds reaches only 0.1, which is much lower than that obtained during calibration processes. The two watersheds shown (Arceau and Nohain) correspond to extreme watershed behavior throughout the climate shift: the Arceau watershed shows very slight variations of its parameters fixed values between the two calibrations while St-Martin/ Nohain watershed is the one with the strongest param-

### Table 4. SWAT and GR4J monthly performances (NASH coefficient) on calibration and validation

<table>
<thead>
<tr>
<th>Watersheds</th>
<th>SWAT MODEL</th>
<th></th>
<th></th>
<th></th>
<th>GR4J MODEL</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NS C1V1</td>
<td>NS C1V2</td>
<td>NS C2V2</td>
<td>NS C2V1</td>
<td>NS C1V1</td>
<td>NS C1V2</td>
<td>NS C2V2</td>
<td>NS C2V1</td>
</tr>
<tr>
<td>Chablis</td>
<td>0.87</td>
<td>0.84</td>
<td>0.85</td>
<td>0.88</td>
<td>0.89</td>
<td>0.92</td>
<td>0.95</td>
<td>0.78</td>
</tr>
<tr>
<td>Quincy-le-Vicomte</td>
<td>0.84</td>
<td>0.88</td>
<td>0.88</td>
<td>0.86</td>
<td>0.88</td>
<td>0.92</td>
<td>0.96</td>
<td>0.79</td>
</tr>
<tr>
<td>Arceau</td>
<td>0.87</td>
<td>0.86</td>
<td>0.84</td>
<td>0.88</td>
<td>0.89</td>
<td>0.94</td>
<td>0.94</td>
<td>0.82</td>
</tr>
<tr>
<td>St-Martin/N.</td>
<td>0.83</td>
<td>0.72</td>
<td>0.77</td>
<td>0.75</td>
<td>0.69</td>
<td>0.84</td>
<td>0.91</td>
<td>0.27</td>
</tr>
<tr>
<td>Poiseux</td>
<td>0.85</td>
<td>0.80</td>
<td>0.81</td>
<td>0.87</td>
<td>0.88</td>
<td>0.97</td>
<td>0.92</td>
<td>0.79</td>
</tr>
<tr>
<td>St-Ussuge</td>
<td>0.92</td>
<td>0.87</td>
<td>0.89</td>
<td>0.89</td>
<td>0.92</td>
<td>0.90</td>
<td>0.95</td>
<td>0.86</td>
</tr>
<tr>
<td>Rigny/Arroux</td>
<td>0.82</td>
<td>0.86</td>
<td>0.84</td>
<td>0.76</td>
<td>0.86</td>
<td>0.90</td>
<td>0.91</td>
<td>0.78</td>
</tr>
<tr>
<td>Vitry-en-Charollais</td>
<td>0.84</td>
<td>0.85</td>
<td>0.84</td>
<td>0.85</td>
<td>0.80</td>
<td>0.90</td>
<td>0.92</td>
<td>0.68</td>
</tr>
<tr>
<td>Mean</td>
<td>0.86</td>
<td>0.83</td>
<td>0.84</td>
<td>0.84</td>
<td>0.85</td>
<td>0.90</td>
<td>0.93</td>
<td>0.72</td>
</tr>
</tbody>
</table>


**Figure 4.** SWAT (left) and GR4J (right) performances (Nash Efficiency) of the watersheds for the four modalities
eter changes (Figure 6). Therefore, the R-factors calculated on these two watersheds (based on the 95PPU obtained in Figure 6) are quite different: it equal to 0.03 for Arceau, and reaches 0.20 for St. Martin/Nohain.

Nevertheless, the thickness of the 95PPU band obtained from the parameters values fixed on the two calibrations period remains very slight, even for watersheds showing the strongest changes in its parameter values (Figure 5a).

4. Discussion

The behavior of SWAT and GR4J model throughout the climatic shift shows both similarities and differences. Concerning the performance decrease between calibration and validation, the two models are opposite. The low decrease of SWAT performance between calibration and validation provided us the proof of the ability of SWAT model to simulate correctly streamflow throughout a climatic shift (table 4). For GR4J performance, however, strong decreases have been observed over some watersheds (Vitry-en-C. or Rigny/Arroux for example), but these decreases are associated to low variation in parameters values (Table 4 and Figure 6). This fact indicates that the better adaptation of GR4J model on the second calibration period is not associated with strong changes in parameters values. This stability of parameters (although some differences, especially for X2 parameter), already highlighted (Brigode et al., 2013), means, in agreement with Le Lay (2007), that changes in parameter values are not always a good indicator of changes in watershed behavior.

Errors between simulated and observed streamflow showed similar patterns between SWAT and GR4J (Table 5). Error values are systematically higher for C1V2 that for C1V1, showing that models calibrated during period 1 are adapted to a wetter climate. When models calibrated on P1 are applied on P2, simulated stream flows are over-estimated compared to the observed ones. Conversely, models calibrated on P2 show stream flows more underestimated compared to the observed ones. This fact is consistent with the changes in both SWAT and GR4J parameters values.

Indeed, they mainly consist in an increasing of GW_DELAY, GW_QMN and RCHRG_DP for SWAT parameters, and a decrease of X2 and a slight increase of X1 for GR4J parameters. These changes correspond to an adaptation of the model to a drier period: they allow water to meet more easily soil and plant water demand: with an increase in store capacity (increase of X1, GW_QMN and GW_DELAY), and to be exported out of the system (by a decrease of X2 or an increase of RCHRG_DP).

However, for some watersheds, these recharge parameters (both RCHRG_DP and X2) are close between the two calibrations, or even change in the

<table>
<thead>
<tr>
<th></th>
<th>C1V1</th>
<th>C1V2</th>
<th>C2V1</th>
<th>C2V2</th>
<th>C1V1</th>
<th>C1V2</th>
<th>C2V1</th>
<th>C2V2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chabis</td>
<td>5.4</td>
<td>18.1</td>
<td>-6.0</td>
<td>5.6</td>
<td>1.9</td>
<td>23.2</td>
<td>-12.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Quincy-le-Vicomte</td>
<td>5.7</td>
<td>10.6</td>
<td>3.7</td>
<td>9.2</td>
<td>7.7</td>
<td>22.5</td>
<td>-8.2</td>
<td>5.5</td>
</tr>
<tr>
<td>Arceau</td>
<td>10.8</td>
<td>17.9</td>
<td>10.6</td>
<td>18.1</td>
<td>-1.2</td>
<td>10.8</td>
<td>-4.0</td>
<td>10.9</td>
</tr>
<tr>
<td>St-Martin/Nohain</td>
<td>-0.7</td>
<td>16.7</td>
<td>-11.8</td>
<td>4.1</td>
<td>-7.9</td>
<td>14.2</td>
<td>-30.4</td>
<td>-6.3</td>
</tr>
<tr>
<td>Poiseux</td>
<td>0.7</td>
<td>8.5</td>
<td>-15.3</td>
<td>-10.6</td>
<td>4.6</td>
<td>18.8</td>
<td>-9.1</td>
<td>6.6</td>
</tr>
<tr>
<td>St-Usuge</td>
<td>-3.1</td>
<td>12.2</td>
<td>-5.3</td>
<td>3.3</td>
<td>8.3</td>
<td>25.3</td>
<td>-7.3</td>
<td>8.5</td>
</tr>
<tr>
<td>Rigny/Arroux</td>
<td>-7.4</td>
<td>8.5</td>
<td>7.3</td>
<td>7.9</td>
<td>6.9</td>
<td>19.9</td>
<td>-1.7</td>
<td>10.9</td>
</tr>
<tr>
<td>Vitry-en-Charollais</td>
<td>-0.4</td>
<td>0.7</td>
<td>0</td>
<td>0.2</td>
<td>-5.3</td>
<td>4.5</td>
<td>-7.2</td>
<td>5.7</td>
</tr>
<tr>
<td>Mean</td>
<td>4.3</td>
<td>11.6</td>
<td>7.5</td>
<td>7.4</td>
<td>5.5</td>
<td>17.4</td>
<td>10.0</td>
<td>7.9</td>
</tr>
</tbody>
</table>

Means are calculated from absolute values.

Table 5. Errors (in percentage) between observed and simulated interannual module. Means are calculated from absolute values.
other way (Vitry-en-C., Rigny/Arroux, Arceau). These watersheds are also those with the lowest differences between C1V2 and C2V1 streamflow errors. On the other hand, St-Martin/Nohain and Chablis watersheds, which show strong differences for both SWAT and GR4J recharge parameters, have the highest differences in streamflow errors between C1V2 and C2V1.

This fact leads us to think that when the recharge parameters are used to adapt the model to the climate characteristics of the calibration period, the model is less robust. The question is why some watersheds have differences in their recharge parameters, while other watersheds keep their parameters stable between the two calibrations. The answer is maybe in the way the climate shift is taken into account in the model. The GR4J model uses directly potential evapotranspiration values from ERA_INTERIM spatial disaggregation; and we can assume that the PET is correctly reproduced by SWAT model: mean PET values calculated on the 1980-1987 period is 688 mm, which is close to the value from ERA_INTERIM (Table 3).

Another possibility is the bad reproduction of ET (effective evapotranspiration). The ET simulated by SWAT on the 1980-1987 period is 486 mm, so about 70% of the PET, what we think is a credible value.

Here, we argue that the parametrisation during calibration process is the origin of the mistake. Instead of modify store capacities (which allow more water to meet soil or plant demand), calibration processes result in changes in recharge parameters. From our point of view, these changes are not the good solution for the model to be adapted at contrasted climate periods. The integration of another parameter in the calibration process (linked with evapotranspiration) could improve the cross calibration/validation in climate-contrasted periods. This problem highlights the fact that the automatic calibration process cannot replace the expertise of the user concerning the choice of parameter to be changed or not.

Figure 5. Correlation of a) SWAT and b) GR4J calibrated parameter values between the 2 calibrations. Stars show significance.
5. Conclusion and perspectives

The existence of an observed shift in air temperature over France in 1987/88 (Brulebois et al., 2015b) allowed us to test the robustness of two hydrological models throughout this shift. The aim of this study is the assessment of the ability of both SWAT and GR4J models to simulate observed discharge during post-shift period, when the model has been calibrated before the shift, and conversely. Such assessment is an essential step in a prospective study on climate change impacts (Thirel et al., 2015a).

First, both models (GR4J and SWAT) were able to reproduce correctly streamflow of the 8 selected watersheds during the reference period (1980-2010), with P-factor equal to 0.74, R-factor equal to 1.02, and NSE equal to 0.83 at monthly scale for SWAT simulations, and NSE equal to 0.91 for GR4J simulations.

Regarding to models robustness throughout contrasted climate periods, performance decreases observed between calibration and validation showed that GR4J model can be more efficient in calibration, but also less robust during validation. Conversely, SWAT showed homogenous performance and lower errors in streamflow simulations for each modality tested (C1V1, C1V2, C2V2, C2V1). The integration of more parameters in calibration could improve the efficiency in calibration but also reduce robustness.

Both models showed a good stability of their parameters between the two calibrations, especially on GR4J parameters (correlation coefficients are higher than those on SWAT parameters).

Despite this stability, some changes have been observed in parameters value between the two calibrations. The changes in recharge parameters (RCHRG_DP and X2) characterize an adaptation to a drier calibration period, but could be a “maladaptation” of the model in a climate change context.

Finally, the uncertainty induced by these changes remains slight; the watershed with the strongest changes (Nohain at St-Martin/Nohain) showed a very thin band of uncertainty (based on the 95PPU analysis).

These results let us conclude to the validity of model parameters throughout contrasted climate periods. Despite a lower robustness for GR4J model than for SWAT, it appears to be reliable in climate change context. But it does not eliminate the need to analyze jointly the SWAT and GR4J results in order to be more confident in the simulations with a specific focus on soil water content and evaporation transpiration calculation.

References


Penman, H.L. (1948). Natural Evaporation from open water, bare soil and grass. 120–145.


