UNIVERSITÉ LAVAL

Influence de la formulation d'évapotranspiration potentielle sur la transposabilité temporelle et le comportement du modèle SWAT

MÉMOIRE

Benoit Maranda

Maîtrise en génie des eaux
Maître ès sciences (M.Sc.)

Québec, Canada

Résumé

Le choix d'une formulation d'évapotranspiration potentielle (FETP) en modélisation hydrologique n'est pas une tâche évidente. Les hydrologues optent souvent pour une formulation leur étant familière ou pour une déjà disponible dans le modèle utilisé. Ce mémoire examine l'influence de la FETP sur la transposabilité temporelle ainsi que le comportement du modèle SWAT appliqué à un bassin versant canadien soumis à une importante crue printanière. Dans cette optique, 20 FETP sont testées en plus des trois FETP déjà incluses dans SWAT. L'étude consiste en une analyse de sensibilité de Sobol et une calibration SCE-UA sur quatre périodes ayant des caractéristiques climatiques contrastées. Les résultats ont montré que la FETP influence la transposabilité temporelle de SWAT autant que son comportement en termes de sensibilité paramétrique et de jeux de paramètre optimaux. Ces résultats soulignent l'importance de choisir une FETP appropriée aux objectifs de modélisation.

Table des matières

Liste des tableaux

Liste des figures

Liste des abréviations

ANOVA	ANalysis Of VAriance
COMP	COMPlete period (all data)
DC	Dry/Cold
DSST	Differential Split Sample Test
DW	Dry/Warm
FAST	Fourrier Amplitude Sensitivity Testing
FF	Factor Fixing
FP	Factor Prioritization
HC	Humid/Cold
HRU	Hydrological Response Unit
HW	Humid/Warm
KS	Kolmogorov-Smirnov
LH-OAT	Latin-Hypercube One-factor-At-a-Time
NSE	Nash-Sutcliffe Efficiency
NSE_{sqrt}	Nash-Sutcliffe Efficiency on square root of data
PEST	Parameter ESTimation software
PET	Potential EvapoTranspiration
PETF	Potential EvapoTranspiration Formulation
RMSE	Root Mean Square Error
$RMSE_{sqrt}$	Root Mean Square Error on square root of data
RSA	Regional Sensitivity Analysis
SA	Sensitivity Analysis
SCE	Shuffled Complex Evolution
SWAT	Soil and Water Assessment Tool

Liste des symboles

Symboles spécifiques à SWAT:

E_a	Actual evapotranspiration [mm]
δ_{gw}	Draining time of the overlying geologic formation [days]
SEV	Evaporation of the soil water [mm]
CANEV	Evaporation of water intercepted by the canopy [mm]
Q_{gw}	Groundwater flow [mm]
Q_{LAT}	Lateral (subsurface) flow [mm]
LAI	Leaf area index [m^2 m^{-2}]
canmx	Max. amount of water that may be trapped in the canopy [mm]
LAI_{mx}	Maximum leaf area index [m^2 m^{-2}]
w_{seep}	Percolation from bottom of the soil profile [mm]
PET_{LEFT}	PET left after *CANEV* [mm]
T	Plant transpiration [mm]
epco	Plant uptake compensation factor
ep_max	Potential plant transpiration [mm]
es_max	Potential soil evaporation [mm]
R_{day}	Precipitation for the day [mm]
w_{rchrg}	Recharge entering the shallow aquifer [mm]
esleft	Remaining potential soil evaporation after *SNOEV* [mm]
soil_cov	Soil cover index
Sol_Z	Soil depth [mm]
SW	Soil water content [mm]
SNOEV	Sublimation of water in the snow pack [mm]
Q_{surf}	Surface runoff [mm]
FC	Water content at field capacity [mm]
WP	Water content at wilting point [mm]

Δaq	Water content variation in the shallow aquifer [mm]
canstor	Water intercepted by the canopy [mm]
SNO	Water in the snow pack [mm]
w_{revap}	Water moving from the shallow aquifer to the soil zone [mm]
w_{deep}	Water reaching the deep aquifer [mm]

Symboles spécifiques à la méthode de SOBOL:

X_i	All sampled values of parameter i
$X_{\sim i}$	All parameters sampled values expect ones of parameter i
$V_{\sim i}$	All term of variance decomposition which does not include parameter i
$f(A)_j$	Function evaluated with the line (sample) j of the matrix A
$f(B)_j$	Function evaluated with the line (sample) j of the matrix B
$f(A_B^{(i)})_j$	Function evaluated with the line (sample) j of the matrix formed from all column (parameter) of matrix A expect the i^{th} column (parameter) which is taken from matrix B
i	Reference to a specific matrix column (or parameter)
j	Reference to a specific matrix line (or sample)
S_i	Sobol main index for parameter i [%]
ST_i	Sobol total index for parameter i [%]
k	Total number of parameters
N	Total number of samples
V_i	Variance caused by parameter i (interaction terms excluded)

Symboles spécifiques au SCE:

nopt	Number of calibrated parameters
npg	Number of member per complex
nps	Number of member per simplex

Symboles spécifiques à la régression multiple:

$R^2_{adjusted}$	Adjusted coefficient of determination
R^2	Coefficient of determination
Y	Dependent variable
$\hat{y}$	Estimates of the dependant variable with the regression
n	Number of observations
k	Number of predictor variable
r	Pearson correlation coefficient
x_i	Predictor (independent) variable i
ε	Random error with a mathematical expectation of zero
β_i	Regression coefficient associated to predictor variable i
j	Sample number (measurements of Y)
SSR	Sum of squares for the regression
SST	Total sum of squares for the dependant variable measurements

Autre symboles:

Rs	Incoming Solar Radiation [W m^{-2}]
$Q_{obs,i}$	Observed flow on day i [m^3 s^{-1}]
p-value	Probability of rejecting the null hypothesis of a statistical test
RH	Relative Humidity [%]
$Q_{sim,i}$	Simulated flow on day i [m^3 s^{-1}]
$T°$	Temperature [°C]
U	Wind speed [m s^{-1}]

Remerciements

J’aimerais remercier François Anctil pour son soutien et ses suggestions apportées tout au long du projet ainsi que Grégory Seiller pour avoir fourni du temps et des outils ayant grandement profité au projet.

Avant-propos

Ce mémoire a été rédigé sous forme d'article. Afin de permettre une meilleure diffusion des connaissances et observations effectuées lors de ce travail de recherche, il a été convenu de le rédiger en anglais. Le corps du mémoire consiste essentiellement en cet article bien que certaines modifications ont été apportées tel que la division de l'article en chapitres ainsi que l'ajout de certaines informations complémentaires utiles à la compréhension. L'article a été rédigé par l'auteur de ce mémoire (auteur principal) avec le soutien de François Anctil (second auteur). Au moment de l'écriture du mémoire, l'article n'a pas encore été soumis pour publication.

Potential evapotranspiration formulation influence on SWAT temporal transposability and behaviour: A case study

B. Maranda and F. Anctil

Article abstract

Selecting a potential evapotranspiration formulation (PETF) for hydrological modeling is not a trivial task, so hydrologists often opt for a familiar formulation or one that is already available within the model they use. This paper investigates the influence of PETF on the temporal transposability and behaviour of the SWAT model, for a Canadian catchment subject to an important spring freshet. For this matter, 20 PETF are tested in addition to the 3 available in SWAT. The analysis consists in a Sobol' sensitivity analysis and a SCE-UA calibration over four climatically contrasted periods. Results show that PETF affects SWAT temporal transposability but also its behaviour in terms of parameter influence and optimal parameter sets. These results stress the importance of selecting a PETF appropriate to one's modeling objectives.

Chapter 1

Introduction

Evapotranspiration is an important component of the water budget, as about two-thirds of the precipitation over the continents return to the atmosphere through evaporation and transpiration (Baumgartner et al., 1975) – a proportion that varies considerably from one location to the other. Potential evapotranspiration (PET) refers to the volume of water "that would be possible under ideal conditions of soil moisture and vegetation" (Thorthnwaite, 1948). This concept is often exploited in hydrological modelling for the simplicity of the PET formulations (PETF), which, for some, only require daily air temperature observations.

Many methods have been used to develop PETF, such as water balance, energy balance, aerodynamic theory and empirical regression (Rana and Katerji, 2000). In hydrological modeling, it is common to exploit PETF taking into account soil water content and plant growth. However, the diversity of the proposed approaches opens up the question of their influence on water partitioning within a hydrological model.

Hydrological model sensitivity to PETF has been addressed by Andréassian et al. (2004), which concluded that lumped hydrological models are sensitive to PETF in terms of optimal parameter sets, based on calibration of 62 catchments. Oudin et al. (2005) assessed the impact of PETF on hydrological performance – 4 lumped model, 27 PETF and 308 catchments – and concluded that simple temperature-based formulations perform better than more complex ones. Seiller and Anctil (2013) studied the relative contribution of natural climate variability, PETF, lumped

conceptual models, and snow modules on hydrological projection uncertainties and concluded that PETF is the second most important contributor to the total uncertainty, after natural climate variability. They also pointed out that exploring PETF sensitivity to different model classes (conceptual versus physical, lumped versus distributed) would be an "important complementary contribution".

These findings stress the importance of evaluating the hydrological influence of PETF selection. Indeed, the lack of measured evapotranspiration data makes it difficult to choose a particular PETF. Evaluating the impact of PETF on temporal transposability and model behaviour may thus address deficiencies in or improve current hydrological modeling practices.

1.1 Temporal transposability

Temporal transposability is the capacity of a model to adequately simulate a phenomenon in periods with different characteristics, without modifying its parameter values or internal structure. Temporal transposability of hydrological model has been recognized as a key component of model calibration for a long time (Klemeš, 1986). Nevertheless, as pointed out by Seiller et al. (2012), hydrologists often assume that parameterization of a model is transposable to periods climatologically different from their calibration period. Of course, the probability that this assumption fails increases in a climate change context. Moreover, climatic fluctuations also exist within historical data series (Koutsoyiannis, 2011) reinforcing the necessity to verify the temporal transposability of a hydrological model calibration.

Many studies already demonstrated from different perspectives the difficulty for hydrological models to transpose optimal parameter sets identified for a specific calibration period to other climatic or hydrological conditions. Wagener et al. (2003) demonstrated using a lumped model applied to an English catchment that summer and rain storm periods needed different values for parameters governing rapid water transfer. Choi and Beven (2007) implemented TOPMODEL on a South Korean catchment and concluded that optimal parameter sets obtained for some specific periods were not suitable for others. More recently, Coron et al. (2012) calculated performance losses under contrasting climate conditions using three conceptual lumped models, pointing out that "further research is needed to apply similar testing procedures with other models".

1.2 Objectives

This study aims to assess PETF influence on temporal transposability and behaviour of a SWAT implementation on a Canadian catchment. Model behaviour is regarded in terms of parameter sensitivity, water partitioning between the hydrological processes and parameter identification. Model behaviour includes problematic issues such as model overfitting to inputs (PETF) and conditions (calibration periods) not necessarily reflected by the model performance. As pointed out by Kirchner (2006), an over-parameterized model, such as SWAT, may act as a "marionette", adapting to any kind of calibration data even if the underlying premises are unrealistic. This stresses the importance of verifying the model reaction to different inputs. This concept is often referred as an equifinality problem: the possibility to obtain similar performances from dissimilar parameter sets.

To the author's knowledge, temporal transposability has never been assessed in terms of SWAT behaviour. Although PETF influence on performance has been investigated by Wang et al. (2006) for the 3 PETF included in SWAT, no study has been performed for a larger spectrum of PETF. Given the popularity of SWAT (more than 1600 publications), particularly for climate change impacts on hydrological processes and water resources, the authors advocate the need for assessing its temporal transposability and PETF sensitivity.

Chapter 2

Material and methods

The methodology of this study is separated in two steps. The first one consists in a sensitivity analysis (SA) of SWAT parameters to the model outputs. This step is essential when using an over-parameterized model such as SWAT, in order to evaluate which parameters have a more significant impact on outputs for the studied watershed. It produces more robust calibrations by reducing the number of free parameters. In this study, the influence of selecting a particular evapotranspiration formulation on parameters sensitivity is also briefly analysed.

The second step involves calibrating SWAT for 23 different PETF and for periods of dissimilar climatic properties. The former allows evaluating SWAT sensitivity to the choice of PETF and the latter addresses SWAT temporal transposability. The next sections will present a brief overview of materials used for the study.

2.1 Study area

The au Saumon catchment (Seiller et al., 2012) is located in the Province of Québec, Canada, near the border with the United States of America (Figure 1). It is a natural territory, exempt of dams controlling the flow of water, that drains 738 km^2 mostly consisting of mixed coniferous and deciduous forests over limestone, sandstone and shale soils (silt-loam) from Ordovician, Silurian and Devonian sedimentary rocks. Its altitude ranges between 277 and 1092 m, its mean annual air temperature is 4.5 °C, and its mean annual precipitation reaches 1284 mm (1975–2003), of

which 355 mm is snow, leading to a mean annual discharge of 771 mm. Its hydrological regime is characterised by a spring freshet (March to May).

The *Centre d'expertise hydrique du Québec* provided the daily hydrometeorological data: discharge (m^3/s), minimum and maximum air temperature (°C), total precipitation (mm), incoming solar radiation (W/m^2), relative humidity (%) and wind speed (m/s).

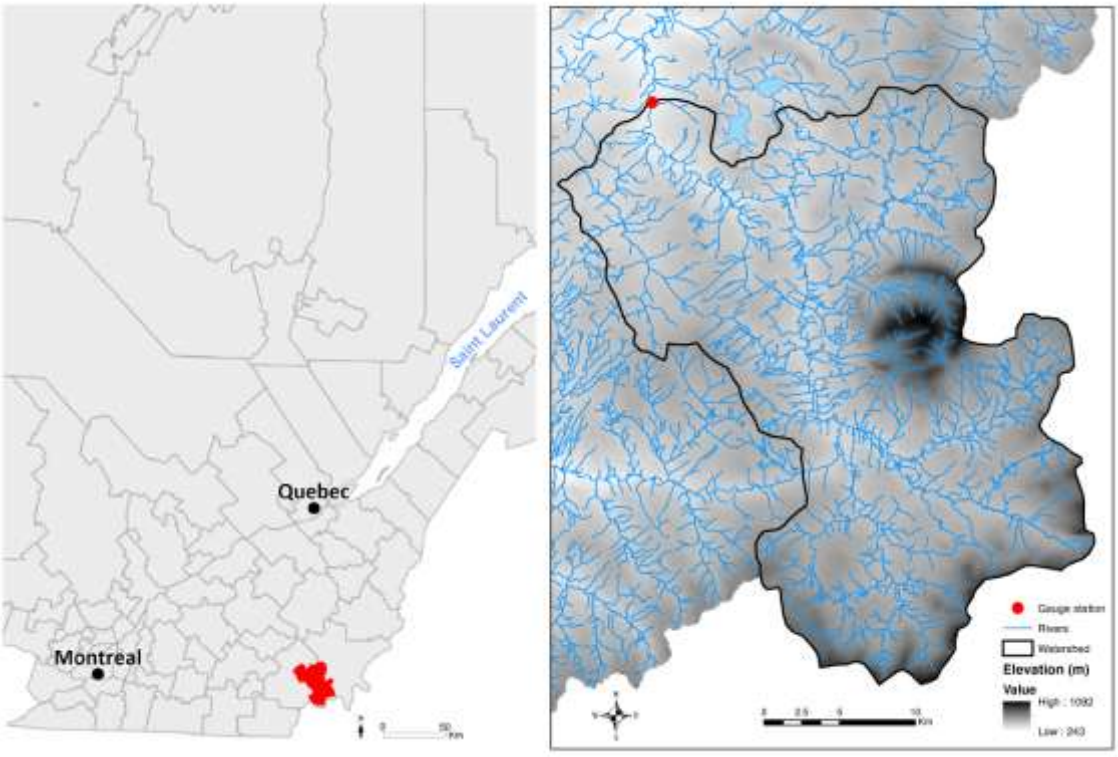

Figure 1: Location of the Au Saumon catchment (738 km²; Canada)

2.2 SWAT

SWAT (Arnold et al., 1993) is a semi-distributed, physically-based, continuous time hydrological model. It is designed to simulate water quantity and quality at the watershed and sub-watershed levels on a daily or sub-daily basis. Its physical nature provides the capacity to evaluate

many different scenarios such as anthropic modification of land uses or management practices, implementation of reservoirs or dams as well as climate change effects on hydrological processes or pollutant loadings.

In SWAT, a watershed is divided into sub-basins based on its slopes and hydrological network. Each sub-basins is next subdivided into hydrological response units (HRUs), which consist of lumped areas with unique land cover, soil attributes, and management practices (Neitsch et al., 2011). The land phase, depicted by equation (1), is calculated at the HRU level while the routing phase is calculated at the sub-basin level using Manning's equation (velocity and flow) and a variant of the kinematic wave model (water transfer).

$$SW_t = SW_0 + \sum_{i=1}^{t}\left(R_{day,i} - Q_{surf,i} - E_{a,i} - w_{seep,i} - Q_{LAT,i}\right) \qquad (1)$$

In equation 1, SW_t and SW_0 represent the soil water content at time step t and t_0 respectively, R_{day} the precipitation, Q_{surf} the surface runoff, E_a the actual evapotranspiration, w_{seep} the percolation from the bottom of the soil profile, and Q_{LAT} the lateral flow. Each term of the summation represents a different process or principle. As an example, Q_{surf} is determined by a modified version of the SCS curve number method. Following sections will detail water partitioning and evapotranspiration calculation in SWAT. A complete description of SWAT is given by Neitsch et al. (2011).

2.2.1 Water partitioning

The influence of PETF on water partitioning has been evaluated based on an extended version of equation (1). Based on Neitsch et al.

(2011)**,** the percolation from the bottom of the soil profile (w_{seep}) responds to the following relation:

$$\sum_{i=1}^{t} w_{rchrg,i} \approx \sum_{i=1}^{t} w_{seep,i} = \sum_{i=1}^{t} \left(w_{deep,i} + Q_{gw,i} + w_{revap,i} + \Delta aq_i \right) \quad (2)$$

where w_{rchrg} represents the recharge entering the shallow aquifer, w_{deep} the water reaching the deep aquifer, Q_{gw} the groundwater flow, w_{revap} the water moving from the shallow aquifer to the soil zone due water deficiencies, and Δaq the water content variation in the shallow aquifer. We also have:

$$w_{rchrg,i} = \left(1 - \exp\left[-1/\delta_{gw}\right]\right) \cdot w_{seep} + \exp\left[-1/\delta_{gw}\right] \cdot w_{rchrg,i-1} \quad (3)$$

where δ_{gw} represent the draining time of the overlying geologic formation. The latter equation shows that w_{rchrg} summed on a relatively long period will approximately be equal to w_{seep}. Combining equations (1) and (2) and rearranging gives:

$$\sum_{i=1}^{t} \left(R_{day,i} \right) = \sum_{i=1}^{t} \left(\Delta SW_i + Q_{surf,i} + E_{a,i} + w_{deep,i} + Q_{gw,i} + w_{revap,i} + \Delta aq_i + Q_{LAT,i} \right) \quad (4)$$

which represents the relative contribution of hydrological components to the total water budget. Equation (4) will be used in section 3.4 to assess PETF influence on water partitioning.

2.2.2 Evapotranspiration

SWAT proposes three different methods for calculating PET: the formulations of Priestley-Taylor, Penman-Monteith and Hargreaves. It also

allows imposing daily PET values from an external file, but imposing the same PET values to all sub-basins and HRUs.

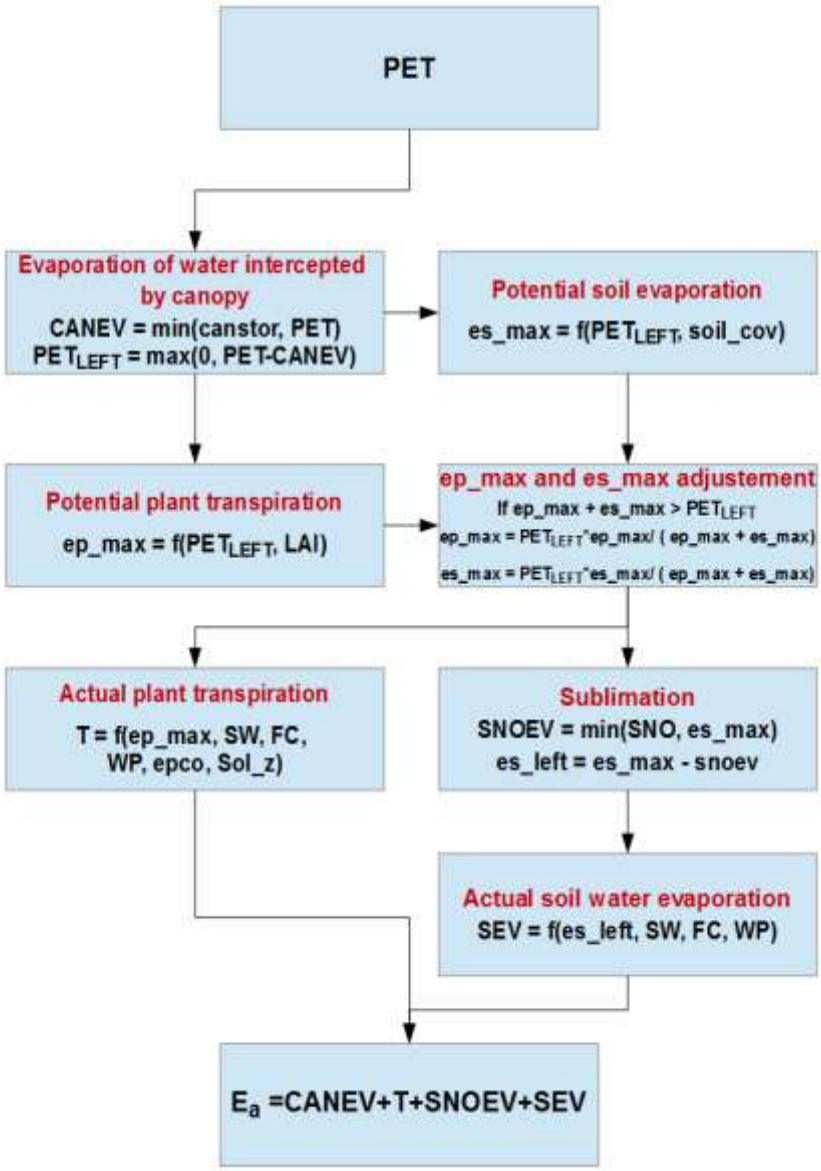

Figure 2: Schematic representation of actual evapotranspiration calculation in SWAT

Once *PET* is calculated, E_a is the sum of the evaporation of the water intercepted by the canopy (*CANEV*), the plant transpiration (*T*), the

sublimation of water in the snow pack (*SNOEV*), and the evaporation of the soil water (*SEV*). As shown in Figure 2, *CANEV* is the minimum between the water stored in the canopy (*canstor*) and *PET*. The state variable *canstor* is then calculated as a function of the leaf area index (*LAI*) on a given day, the maximum leaf area index for the plant (LAI_{mx}), and the maximum amount of water that may be trapped in the canopy (*canmx*). The remaining *PET* (PET_{LEFT}) is then the difference between *PET* and *CANEV*. The potential soil evaporation (*es_max*) is calculated next as a function of PET_{LEFT} and the soil cover index (*soil_cov*) and potential plant transpiration (*ep_max*) as a function of PET_{LEFT} and *LAI*. Those values are also adjusted to make sure that their sum does not exceed PET_{LEFT}. Subsequently, *T* is calculated as a function of *ep_max*, soil water content (*SW*), the water content at field capacity (*FC*), the water content at wilting point (*WP*), the plant uptake compensation factor (*epco*), and the soil depth (*Sol_Z*). In parallel, *SNOEV* is calculated as the minimum between the water in the snow pack (*SNO*) and *es_max*. Finally *SEV* is calculated as a function of the remaining potential soil evaporation after *SNOEV* (*es_left*), *SW*, *FC*, and *WP*.

In short, E_a calculation is mainly affected by soil cover, plant growth, and soil water balance (and PET obviously).

2.3 PET formulations

For the purpose of this study, 20 supplementary PETF have been added to the SWAT code, for a total of 23. The 20 additional formulations have been chosen following Seiller and Anctil (2013) in order to provide a large spectrum of PET series. Table 1 presents the selected PETF as well as the numbers which will be used herein for reference. Inputs are also identified in Table 1 along their mean and variance for the available 29-

year time series. More details are given in Appendix A for added formulations and in SWAT theoretical documentation (Neitsch et al., 2011) for formulations included in SWAT. Appropriate empirical coefficients have been identified by Seiller and Anctil (2013) for the au Saumon catchment. All other needed parameters have been calculated based on Allen et al. (2005).

Table 1: PETF with respective inputs, mean and variance (*Rs* = Incoming solar radiation, *RH* = Relative humidity, *T°* = temperature, and *U* = wind speed)

Formulation name	Inputs	Mean	Variance
-	-	(mm)	(mm^2)
E01-SWAT_Priestley-Taylor	*Rs, RH, T°*	1.38	3.37
E02-SWAT_Penman-Monteith	*RH, T°, U, Rs*	1.61	2.55
E03-SWAT_Hargreaves	*T°*	2.04	2.88
E04-Penman	*RH, T°, U, Rs*	1.42	1.77
E05-ASCE	*RH, T°, U, Rs*	1.75	2.01
E06-Kimberly-Penman	*RH, T°, U, Rs*	1.81	2.56
E07-Thom-Oliver	*RH, T°, U, Rs*	1.46	1.68
E08-Thornthwaite	*T°*	1.51	2.50
E09-Blaney-Criddle	*T°*	2.53	4.21
E10-Hamon	*T°*	2.06	3.20
E11-Romanenko	*RH, T°*	1.31	1.73
E12-Linacre	*RH, T°*	1.86	2.43
E13-HSAMI-HYDROTEL	*T°*	1.70	3.12
E14-Kharrufa	*T°*	1.81	4.30
E15-Wendling	*T°, Rs*	1.72	1.79
E16-Turc	*RH, T°, Rs*	2.96	7.71
E17-Jensen-Haise	*T°*	2.66	8.82
E18-McGuinness-Bordne	*T°*	2.21	4.66
E19-Doorenbos-Pruitt	*RH, T°, U, Rs*	2.56	4.90
E20-Abtew	*RH, T°, Rs*	1.97	1.54
E21-Makkink	*T°*	1.87	4.25
E22-Oudin	*T°*	1.50	2.15
E23-Baier-Robertson	*T°*	1.69	3.84

2.4 SWAT parameters

SA focuses on SWAT parameters that most affect the streamflow according to Abbaspour et al. (2007), Cibin et al. (2010), Neitsch et al. (2011), Nossent et al. (2011), and Zhang et al. (2013). 40 parameters are

thus retained and classified into 10 categories with respect to their associated process. Table 2 presents the symbol for each parameter as well as a short description, units, and range used for SA. Parameters highlighted in gray correspond to the ones used for calibration.

Table 2: SWAT parameters used in SA and calibration (in gray)

Symbol	Description	Units	min	max	Process
Canmx	Maximum canopy storage	mm H_20	0.001	10	Evapotranspiration
Epco	Plant uptake compensation factor	-	0.01	1	Evapotranspiration
Esco	Soil evaporation compensation coefficient	-	0.001	1	Evapotranspiration
Evrch	Main channel evaporation adjustement factor	-	0	1	Evapotranspiration
Alpha_Bf	Baseflow recession constant	$days^{-1}$	0.001	1	Groundwater
Gw_Delay	Delay time for aquifer recharge	days	0.001	100	Groundwater
Gwqmn	Threshold water level in shallow aquifer for base flow	mm H_20	0.001	1000	Groundwater
Rchrg_Dp	Aquifer percolation coefficient	-	0.001	1	Groundwater
Phu_plt	Potential heat units for plants	-	-50%	+50%	Plant growth
Blai	Potential maximum leaf area index for plants	-	-50%	+50%	Plant growth
Bio_init	Total plant biomass	kg/ha	-50%	+50%	Plant growth
Biomix	Biological mixing effeciency	-	0	1	Plant growth
LAI_init	Leaf area index of the canopy	-	0	10	Plant growth
Gw_Revap	Revap coefficient	-	0.02	0.2	w_{revap}
Revapmn	Threshold water level in shallow aquifer for revap	mm H_20	0.001	500	w_{revap}
Sftmp	Maximum mean air temperature for precipitation to fall as snow	°C	-3	3	Snow cover
Sno50cov	Fraction of Snocovmx that provides 50% cover	-	0	1	Snow cover
Snocovmx	Threshold depth of snow, above which there is 100% cover	mm H_20	0	500	Snow cover
Smfmn	Melt factor on December 21	mm H_20/day/°C	0.001	10	Snow melt
Smfmx	Melt factor on June 21	mm H_20/day/°C	0.001	10	Snow melt
Smtmp	Threshold temperature for snowmelt	°C	-5	5	Snow melt
Timp	Snow temperature lag factor	-	0.01	1	Snow melt
Sol_Alb	Moist soil albedo	-	0.1	1	Soil water
Sol_AWC	Available water capacity	-	-50%	+50%	Soil water

Table 2: SWAT parameters used in SA and calibration (in gray) (continued)

Symbol	Description	Units	min	max	Process
Sol_K	Saturated hydraulic conductivity of first soil layer	mm/ hour	-50%	+50%	Soil water
Sol_Z	Depth from soil surface to bottom layer	mm	-50%	+50%	Soil water
Slope	Average slope of the subbasin	m/m	-50%	+50%	Surface runoff
Cn2	Moisture condition II curve number	-	-50%	+50%	Surface runoff
Ov_n	Manning's value for overland flow	-	0.008	0.5	Surface runoff
Slsubbsn	Subbasin slope length	m	-50%	+50%	Surface runoff
Surlag	Surface runoff lag coefficient	hours	0.001	10	Surface runoff
Ch_K1	Effective hydraulic conductivity of tributary channels	mm/ hour	0.01	150	Transmission losses
Ch_K2	Effective hydraulic conductivity of main channel	mm/ hour	0.01	150	Transmission losses
Ch_N1	Manning's value for tributary channels	-	0.005	0.7	Water routing
Ch_N2	Manning's value for main channel	-	0.01	0.5	Water routing
Ch_d	Depth of water in main channel when filled to bank	m	-50%	+50%	Water routing
Ch_S1	Average slope of the tributary channels	m/m	-50%	+50%	Water routing
Ch_S2	Average slope of the main channel	m/m	-50%	+50%	Water routing
Ch_W1	Average width of tributary channel	m	-50%	+50%	Water routing
Ch_W2	Width of main channel at top of bank	m	-50%	+50%	Water routing

Since some parameters are distributed, relative changes have been used instead of absolute values for those parameters. Limit values (min and max) have been chosen according to the SWAT documentation, related studies, and personal judgement.

Even if plant growth does not directly affect channel flow, some associated parameters have been included since plant growth affects transpiration and evaporation.

Note that the w_{revap} process consists in the amount of water moving from the shallow aquifer to the soil zone due water deficiencies and may also affect evapotranspiration. The other processes should be self-explaining.

2.5 Database structuration

A structuration of the available data has been performed to identify non-continuous climatically contrasted periods and a short continuous period statistically similar to the overall database (surrogate of calibration data). Note that years 1975 and 1976 have been reserved for the warm-up of the model.

2.5.1 Contrasted calibration data

The use of non-continuous climatically contrasted periods has first been introduced by Klemeš (1986), under the term differential split sample test (DSST), for the verification of a model under conditions dissimilar from those corresponding to the available record or used for calibration. For this study, DSST has been used to assess the temporal transposability of the model but also to verify the influence of calibration periods on PETF performance.

The identification of contrasted calibration data focussed on the air temperature and the precipitation rate because of their dominant influence on many hydrological processes but also because, according to the Intergovernmental Panel on Climate Change (IPCC, 2013), changes in temperature are virtually certain and changes in precipitation are very likely over most land areas during the 21st century. As SWAT has been developed as a continuous time model (long-term yield model, Neitsch et

al., 2011), the predicted variability of those climatic components is particularly interesting to verify the model's capacity to adapt to future climatic conditions.

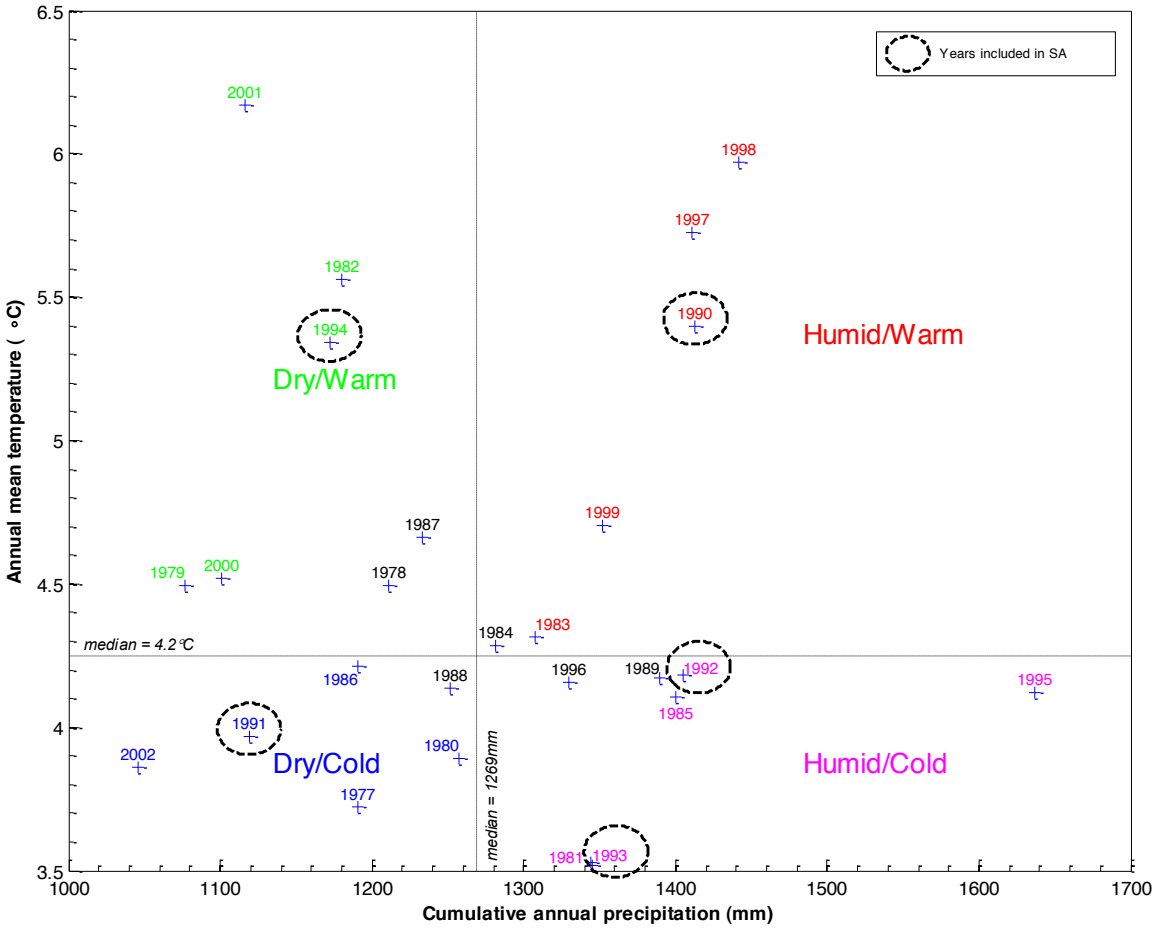

Figure 3: Database analysis for calibration in DSST. Every year are plotted according to annual mean temperature and annual cumulative precipitation. Coloured years are the one used for calibration/validation for the corresponding climatic period (green for Dry/Warm, etc.).

5 non-continuous years for which the climate (yearly average) is either Dry/Cold (DC), Dry/Warm (DW), Humid/Cold (HC), or Humid/Warm (HW) were selected. Figure 3 presents the analysis results for the au Saumon catchment.

2.5.2 Surrogate of calibration data

A short continuous period statistically similar to the overall database (surrogate of calibration data) has been identified for SA since, as it will be detailed later, 168 000 model evaluations are needed for each PETF, making it difficult to perform an analysis based on all the data (1977-2003) due to computational time issues.

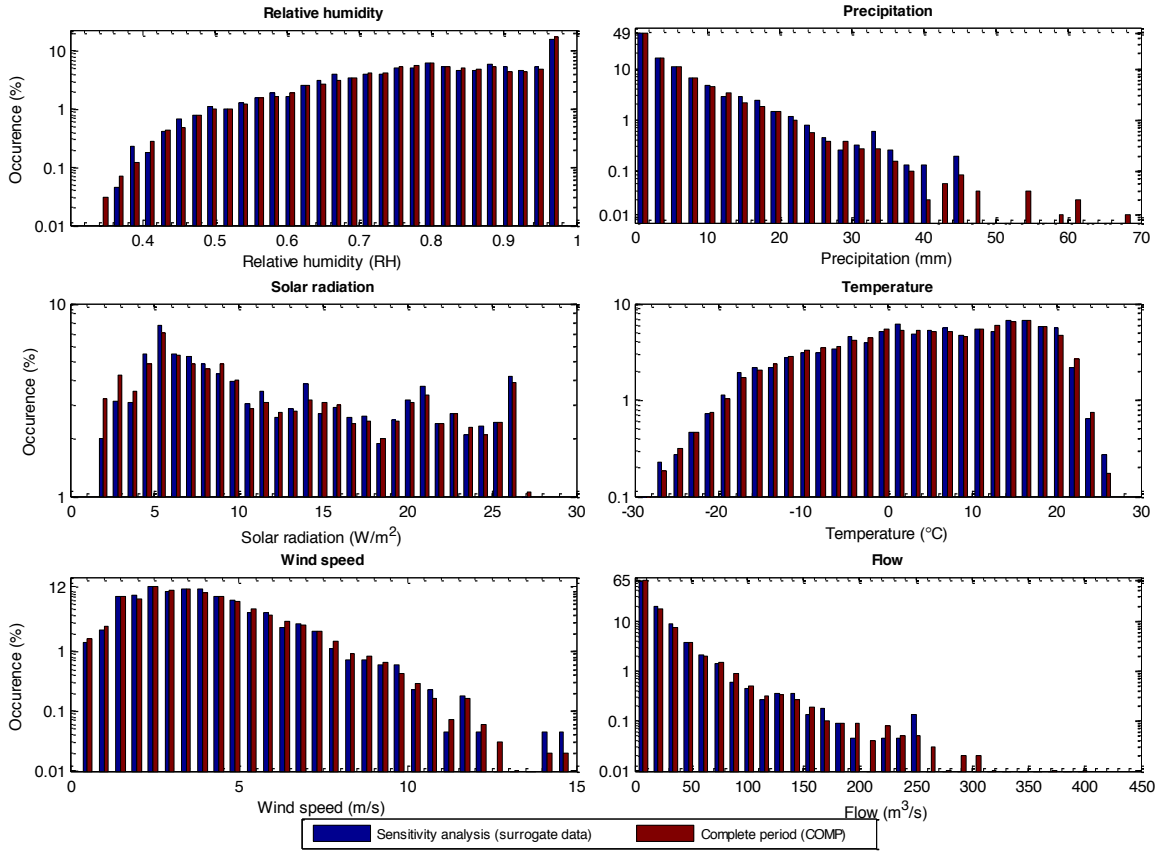

Figure 4: Comparison of climatic component between SA period (1990-1995) and complete period (1977-2003). Occurrence (%) corresponds to numbers of days for which the hydro-climatic component value is situated in the boundary represented by the respective bar.

This analysis has been carried using a Kolmogorov-Smirnov (KS) test with a confidence interval of 95% (alpha=0.05) for each of the climate component, comparing the complete period with all possible periods of 4 to

10 consecutive years. At the end, the 1990-1995 period was chosen because it included at least one year from each group of the differential split sample test (DSST, see Figure 3). Figure 4 compares the 1990-1995 period to the overall database, in percentage of occurrence for the given range of x-axis values. Log-scale y-axis allows to represent extreme values. Both databases in Figure 4 present similar histograms for every climatic component, which supports the performed selection. Only very high precipitation rate is not well depicted but those events have a relative occurrence of less than 0.2%.

This segregation methodology assumes that periods with similar statistics contain enough information to provide similar results in terms of sensitivity of the parameters. Cibin et al. (2010) showed that parameter sensitivity can be linked to climatic and hydrological characteristics of the catchment. For this reason, a histogram of the streamflow is also drawn to account for this hydrological dimension. It shows that flows are also well represented by the 1990-1995 period up to about 250 m^3/s. Higher flows are not well represented but have very low occurrence (0.15%).

2.6 Parameters sensitivity analysis

SA allows for a better understanding of the internal structure of a model and can lead to a reduction in the number of parameters used for calibration. The reduction is achieved by factor fixing (FF), freezing the value of parameters that have only a small influence on the model output, or by factor prioritization (FP), where only a limited number of the most influential parameters are kept for calibration in order to maximally reduce the output uncertainty (Saltelli et al., 2004; Nossent et al., 2011).

The present study focussed on FP, as a mean to effectively calibrate the model for each PETF without favouring one parameter over another. SA has thus been performed individually for each PETF as some recent studies revealed that SWAT parameters sensitivity is affected by the time series and the watershed (Cibin et al., 2010; Zhang et al., 2013). Section 3.1 will show that SA is indeed affected by PETF.

SWAT SA often resorts to the Latin-hypercube one-factor-at-a-time (LH-OAT) method (van Griensven et al., 2006), as the algorithm is incorporated in the SWAT code. Although this method has the advantage of being frugal from a number of simulations point of view, its results are more of a qualitative nature than of a quantitative one (Nossent et al., 2011). As pointed out by Zhang et al. (2013), LH-OAT does not provide information on the parameter interactions, with the consequence of possibly underestimating the influence of highly-interactive parameters. Moreover, Tang et al. (2007) showed that other screening methods such as parameter estimation software (PEST) and regional sensitivity analysis (RSA) also neglect parameter interactions. As a consequence, screening methods are better suited for FF than for FP (Nossent et al., 2011) especially for hydrological applications, considering their tendency for numerous parameter interactions (Tang et al., 2007).

Variance-based methods such as Fourrier Amplitude Sensitivity Testing (FAST), analysis of variance (ANOVA), and Sobol'/Saltelli (Sobol', 2001; Saltelli et al., 2010) account for parameter interactions, which make them appropriate for FP (Saltelli et al., 2010). They also have the advantage of being model independent and to potentially capture the full range of parameter values (Liburne et al., 2006).

Saltelli and Bolado (1998) compared FAST with Sobol' and concluded that the latter, although computationally more expensive, has the advantage of providing a unique way to estimate the global effect (effect which includes all interaction terms, see section 2.6.1). Tang et al. (2007) compared ANOVA and Sobol' and reported that the latter yielded more robust sensitivity rankings. Furthermore, Sobol' has been successfully applied to complex environmental models including SWAT (Cibin et al., 2010; Nossent et al., 2011; Zhang et al., 2013) and has the advantage of working well for models with a high number of parameters (Glen and Isaacs, 2012), as well as leading to more easily interpretable outputs (Nossent et al., 2011).

2.6.1 Description of Sobol''s method

The following description and notation is adapted from Sobol' (2001) and Saltelli et al. (2010). The Sobol'/Saltelli method is variance-based, calculating the relative contribution of a parameter against the total unconditional variance of the model output, which can be represented by different metrics, although the Root Mean Square Error (RMSE) and the Nash-Sutcliffe Efficiency (NSE) are customary in hydrology.

Considering the following generic model function for which input factors X are independent and scaled between 0 and 1.

$$y = f\left(X\right) = f\left(x_i, ..., x_k\right) \tag{5}$$

Using an ANOVA representation of (5) (Sobol', 2001), we get

$$f(X) = f_0 + \sum_{i=1}^{k} f_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} f_{ij} + \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{z=i+2}^{k} f_{ijz} + \ldots + f_{1\ldots k} \qquad (6)$$

where f_0 is a constant corresponding to the expectation value of (5) and every other term corresponds to functions relative to the parameters in index for which the expected value is zero (*a priori* unknown functions). Indeed, this decomposition exists and is unique only under the assumption that the terms in (6) are orthogonal. It can then be rewritten in terms of variance under the assumption that the square of (5) is integrable

$$V(Y) = \sum_{i=1}^{k} V_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} V_{ij} + \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{z=i+2}^{k} V_{ijz} + \ldots + V_{1\ldots k} \qquad (7)$$

Equation (7) comes from squaring each term of (6) and its integration on every variable from which we have:

$$V(Y) = \int f^2(X) dX - f_0^2 \qquad (8)$$

$$V_i = \int f_i^2 dx_i \qquad (9)$$

$$V_{ij} = \int f_{ij}^2 dx_i dx_j \qquad (10)$$

Dividing (7) by the total unconditional variance of (5) gives:

$$1 = \sum_{i=1}^{k} S_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} S_{ij} + \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{z=i+2}^{k} S_{ijz} + \ldots + S_{1\ldots k} \qquad (11)$$

The terms in the first summation in (11) correspond to the relative contribution of the parameters in index to the total unconditional variance

of the model ignoring interactions with other parameters. This term (S_i) is usually called the main or first order effect. Other terms correspond to the contribution of the interactions between indexed terms to the total unconditional variance. The total effect of a parameter (ST_i) corresponds to the main effect of this parameter in addition to every interaction which includes this parameter. A purely additive model would have a sum of all total effects equal to the sum of all main effects and this sum would equal 1 as equation (11) clearly shows. For a non-additive model, the sum of all main effects would be less than 1 and the total effect can theoretically range from 1 to positive infinity.

As for most hydrological models, the analytical solution is impossible to achieve, the terms for the main and total effects can be expressed in terms of variance of conditional expectation giving:

$$S_i = \frac{V_i}{V(Y)} = \frac{V\left(E\left(Y \mid X_i\right)\right)}{V(Y)} = 1 - \frac{E\left(V\left(Y \mid X_i\right)\right)}{V(Y)} \tag{12}$$

$$S_{Ti} = 1 - \frac{V_{\sim i}}{V(Y)} = 1 - \frac{V\left(E\left(Y \mid X_{\sim i}\right)\right)}{V(Y)} = \frac{E\left(V\left(Y \mid X_{\sim i}\right)\right)}{V(Y)} \tag{13}$$

In equation (12), i corresponds to a specific parameter and in equation (13) $\sim i$ corresponds to all parameters but parameter i. A Monte Carlo approximation can then be used to approximate those values. For this study, the equation proposed by Jansen (1999) (15) has been used, which is also the one advocated by Saltelli et al. (2010) for ST_i. Although estimators exist for other order terms, only the main and total effects have been calculated for this study as it is customary to compute just two sets of indices for FP.

If one imagines two independent matrixes *A* and *B* (*N* lines by *k* columns, where *N* stands for the number of samples and *k* for the number of parameters), we can create *k* other matrices composed of every column of A except the i^{th} ones, which is taken from *B*. It is then possible to calculate the main and total effects of every parameter using:

$$E\left(V\left(Y \mid X_i\right)\right) = \frac{1}{2N}\sum_{j=1}^{N}\left(f\left(B\right)_j - f\left(A_B^{(i)}\right)_j\right)^2 \tag{14}$$

$$E\left(V\left(Y \mid X_{\sim i}\right)\right) = \frac{1}{2N}\sum_{j=1}^{N}\left(f\left(A\right)_j - f\left(A_B^{(i)}\right)_j\right)^2 \tag{15}$$

Table 3: Estimator of total variance for total and main indices

Estimator of *V(Y)* for S_i	Estimator of *V(Y)* for ST_i
$\hat{V}_{Si}\left(Y\right) = \frac{1}{2}\left(\hat{V}_{Si}\left(f\left(A\right)\right) + \hat{V}_{Si}\left(f\left(B\right)\right)\right)$	$\hat{V}_{STi}\left(Y\right) = \frac{1}{2}\left(\hat{V}_{STi}\left(f\left(A\right)\right) + \hat{V}_{STi}\left(f\left(A_B^{(i)}\right)\right)\right)$
$\hat{V}_{Si}\left(f\left(A\right)\right) = \frac{1}{N}\sum_{j=1}^{N} f^2\left(A\right)_j - f_{0x}^2$ $\hat{V}_{Si}\left(f\left(B\right)\right) = \frac{1}{N}\sum_{j=1}^{N} f^2\left(B\right)_j - f_{0x}^2$	$\hat{V}_{STi}\left(f\left(A\right)\right) = \frac{1}{N}\sum_{j=1}^{N} f^2\left(A\right)_j - f_{0A}^2$ $\hat{V}_{STi}\left(f\left(A_B^{(i)}\right)\right) = \frac{1}{N}\sum_{j=1}^{N} f^2\left(A_B^{(i)}\right)_j - f_{0i}^2$
$f_{0x}^2 = \frac{1}{N}\sum_{j=1}^{N} f\left(A\right)_j f\left(B\right)_j$	$f_{0A}^2 = \left[\frac{1}{N}\sum_{j=1}^{N} f\left(A\right)_j\right]^2$ $f_{0i}^2 = \left[\frac{1}{N}\sum_{j=1}^{N} f\left(A_B^{(i)}\right)_j\right]^2$

Estimators used in this study for the total unconditional variance (*V(Y)*) are grouped in Table 3. Considering that *2N* model evaluations are needed for *f(A)* and *f(B)* and *Nk* for $f(A_B^{(i)})$, the total number of model evaluations for this method is *N(k+2)*.

2.6.2 Sampling

The two matrixes *A* and *B* were sampled independently using Latin-hypercube sampling (McKay et al., 2000), as it has been used with success in hydrology (Tang et al., 2007; Cibin et al., 2010; Zhang et al., 2013). This stratified sampling method consists in dividing the range of the *k* parameters into *N* strata of equal probability (*1/N*) and randomly sampling each stratum. The strata are then shuffled for each parameter, leading to *N* semi-random samples. Note that the different parameter ranges are first linearly scaled between 0 and 1. For matrix $A_B^{(i)}$, radial sampling was used as it surpassed the winding stairs sampling using the estimators reported in (14) and (15) (Saltelli et al., 2010).

Sample size is crucial to the Sobol'/Saltelli method as it determines the precision and reliability of the results. Cibin et al. (2010) and Zhang et al. (2013) used a sample size of 2000 for the SA of 13 and 28 SWAT parameters. Nossent et al. (2011) concluded that a sample size of 2000 was needed to attain a stable rank for influential parameters out of 26 SWAT candidates. As the number of parameters used for this study is higher (40), a sample size of 4000 was selected, which asks for 168 000 model evaluations for each PETF (3 864 000 in total).

2.6.3 Bootstrapping

Bootstrapping was applied to calculate the 95% confidence intervals of the main (S_i) and total (ST_i) effects. All samples were resampled 5 000 times with replacement. The percentile method was then used to obtain the 95% confidence intervals. This method requires a high number of samples to achieve a reliable estimate of the intervals which explains why 5000 (re)samples were used.

2.7 Autocalibration

Autocalibration consists in applying an algorithm to find the global or local minima of an objective (cost) function. For this study, a slightly modified version of the Shuffled Complex Evolution (SCE) algorithm (Duan and Gupta, 1992; Duan et al., 1994) has been used, which already showed to be robust (Wang et al., 2010) – this method is also imbedded in the SWAT code (Green and van Griensven, 2008). The SCE is a global optimization algorithm based on the evolutionary concept. It proceeds in five steps that consists in (1) sampling a certain number of points in the parameter space and computing the objective function at those points, (2) sorting the points according to their objective function values, (3) partitioning all points in *ngs* complexes (sub-populations), (4) evolving each complex independently, and finally (5) shuffling every complex members. The last four steps are repeated until the convergence criterion is attained.

Calibration has been done on each of the DSST periods as well as on the complete period, resulting in 5 calibrations for each PETF. The calibration has been performed on the 21 selected parameters (see section 3.1 hereafter) using the surrogate database. Even if it would have been ideal to perform SA for each DSST period, using different parameters would have made it impossible to compare parameter values obtained with the different calibrations.

The maximum number of iterations for SCE has been set to 20 000 and the number of complexes (*ngs*) to 10. The relatively high number of maximum iterations is justified by the high number of parameters and the necessity to minimize the impact of the calibration process on the results.

The number of members per complex (*npg*) and the number of members per simplex (*nps*) have been estimated with the following equations:

$$npg = 2 \cdot \min\left(nopt, 16\right) + 1 \tag{16}$$

$$nps = \min(nopt, 16) + 1 \tag{17}$$

where *nopt* represents the number of calibrated parameters. The limitation on the number of members per complex and simplex is based on the incorporated module of SCE in the SWAT code (an external module has been used for this study).

2.8 Metrics

Two metrics have been retained for this study. The first one is the RMSE applied to the square root of the values, used as an objective function for SA and calibration – use of the square root of the values provides a more equal weight to all flows (Chiew and McMahon, 1994; Oudin et al., 2006):

$$RMSE_{sqrt} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(\sqrt{Q_{sim,i}} - \sqrt{Q_{obs,i}}\right)^2} \tag{18}$$

where N is the number of observations, $Q_{sim,i}$ is the simulated flow on day i and $Q_{obs,i}$ the observed flow on the same day.

The second metric is the NSE also applied to the square root of the values for the same reasons:

$$NSE_{sqrt} = 1 - \frac{\sum_{i=1}^{N}\left(\sqrt{Q_{sim,i}} - \sqrt{Q_{obs,i}}\right)^2}{\sum_{i=1}^{N}\left(\sqrt{Q_{obs,i}} - \overline{\sqrt{Q_{obs,i}}}\right)^2} \quad (19)$$

It ranges from negative infinity to 1 for a model perfectly simulating the observations. The NSE is affected by the variance of the observed flows and thus provides higher scores when the variance is high and lower ones when the variance is low. This makes it difficult to compare values for dissimilar periods. However, the use of square root values lowers this impact by buffering differences in the variance.

2.9 Multiple regression analysis

Multiple regression analysis has been performed on some results. This method consists in predicting or explaining the variation of a dependent variable from a set of predictor variables (independent variables) (Salkind, 2007) in such a way that:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \quad (20)$$

where Y corresponds to the dependent variable, x_i to a predictor variable, k to the number of predictor variables, β_i to regression coefficients (unknown), and ε to a random error with a mathematical expectation of zero. For this study, multiple regression analysis was used to evaluate the part of the variability in the dependent variable that could be explained by predictor variables. This is evaluated by the well-known coefficient of determination (R^2) evaluated for multiple predictor variables and given by:

$$R^2 = \frac{SSR}{SST} = \frac{\sum_j \left(y - \bar{y}\right)^2}{\sum_j \left(y_j - \bar{y}\right)^2} \tag{21}$$

where *SSR* corresponds to the sum of squares for the regression, *SST* to the total sum of squares, *j* to the number of samples (measurements of *Y*), y to the estimates of the dependant variable with the regression, $\bar{y}$ to the mean value of *Y*, and y_j to samples of *Y*. R^2 represents the part of the variance of the dependant variable which can be attributed to the variance in the predictor variables. As R^2 values tend to always increase when new explanatory variables are added even if they only help slightly to explain the dependent variable, an adjusted form of R^2 is used instead. Unlike the common R^2, the adjusted R^2 may decrease when new independent variables are considered. The adjustment consists in weighting the R^2 value by a function of the number of samples and the number of predictor variables such that:

$$R^2_{adjusted} = 1 - \left(\frac{n-1}{n-k}\right)\left(1 - R^2\right) \tag{22}$$

where *n* corresponds to the number of observations and *k* to the number of predictor variables.

The *t-statistic* is also evaluated for every predictor variable to verify their pertinence. This statistic consists in testing the null hypothesis that the coefficient value is zero against the hypothesis that the coefficient value is different from zero. The regression is judged satisfactory when related *p-values* for every predictor variable is below 0.05, meaning that there is only a 5% probability that the predictor variable has no effect on the dependant variable.

It is important to note that the Pearson correlation coefficient (*r*) is used when only one predictor is available, to report the nature (sign) of the correlation (positive or negative). The Pearson correlation coefficient is calculated by:

$$r = \frac{\sum_j \left(x_j - \bar{x}\right)\left(y_j - \bar{y}\right)}{\sqrt{\sum_j \left(x_j - \bar{x}\right)^2}\sqrt{\sum_j \left(y_j - \bar{y}\right)^2}} \tag{23}$$

Chapter 3

Results

This section groups results of the SA and of the evaluation of the influence of PETF on SWAT performance, temporal transposability, water partitioning, and parameterization.

3.1 Sobol''s SA

Figure 5 illustrates the SA results for ST_i (blue) and S_i (red). The central mark represents the median value of the 23 PETF, the box edges the 25th and 75th percentiles, while the whiskers extend to the extreme values. Parameters (abscissa) are ordered according to their median ST_i value and the dotted line illustrates the confidence interval for every PETF. Results clearly show that an important portion of the influence of the parameters originates from interactions between the parameters. Indeed, the sum of ST_i for all parameters ranges from 1.49 to 1.84 while the sum of S_i ranges from 0.36 to 0.65, indicating that the model is non-additive, which is consistent with findings from Nossent et al. (2011).

Figure 5 also shows that the aquifer percolation coefficient (*Rchrg_dp*) is by far the most influential parameter. Actually, the expected reduction in variance if this parameter would be fixed reaches 40% (S_i) of the total unconditional model variance and the expected variance that would be left, if all parameters but this one would be fixed, reached 57% (ST_i). The difference between the former and the latter comes from the interactions between *Rchrg_dp* and the other parameters. Most SWAT SA reported in the literature identified instead the curve number (*Cn2*) as the

most influential (van Griensven et al., 2006; Feyereisen et al., 2007; Nossent et al., 2011) while it came second in our analysis. This is certainly linked to the au Saumon geology and climate and possibly to the objective function based on the square root of the stream flows. During low flows, the proportion of the discharge attributed to Q_{GW} will be higher than for the high flows, since low flows are normally derived from groundwater storage (Smakhtin, 2001). To the opposite, surface runoff (Q_{Surf}) will gain more importance during high flows, which generally coincide with wet or melting periods. *Cn2* is function of soil permeability, land use, and antecedent soil moisture conditions, and mostly serves to calculate Q_{Surf} from SCS curves. *Rchrg_dp* corresponds to the portion of aquifer recharge which is lost to the deeper aquifer and does not contribute to groundwater flow (Q_{GW}). High sensitivity for the latter is plausible for the au Saumon watershed as recharge tends to come primarily from the upland water source. Another explanation could be that low flows are more frequent then high flows and that $RMSE_{sqrt}$ tends to give a more equal importance to every magnitude of flows, thus attributing more importance to the Q_{GW} component of the total flow.

Rchrg_dp is actually the most influential parameter for every PETF except for E-16, which provided the highest mean PET (see Table 1). However, for this formulation, *Cn2* is the most influential parameter. Calculating Pearson correlation coefficient (r) between the mean PET and *Cn2* ST_i and S_i gave 0.88 and 0.93 respectively, which shows that *Cn2* sensitivity is highly influenced by the choice of a PETF, more specifically by the quantity of PET. Higher PET leave less water available for runoff as the E_a component is applied before the distribution of precipitation between Q_{surf} and the water entering the soil profile. The higher sensitivity to *Rchrg_dp* for low PET (r = -0.67) could be attributed to the fact that more water will then reach the aquifer.

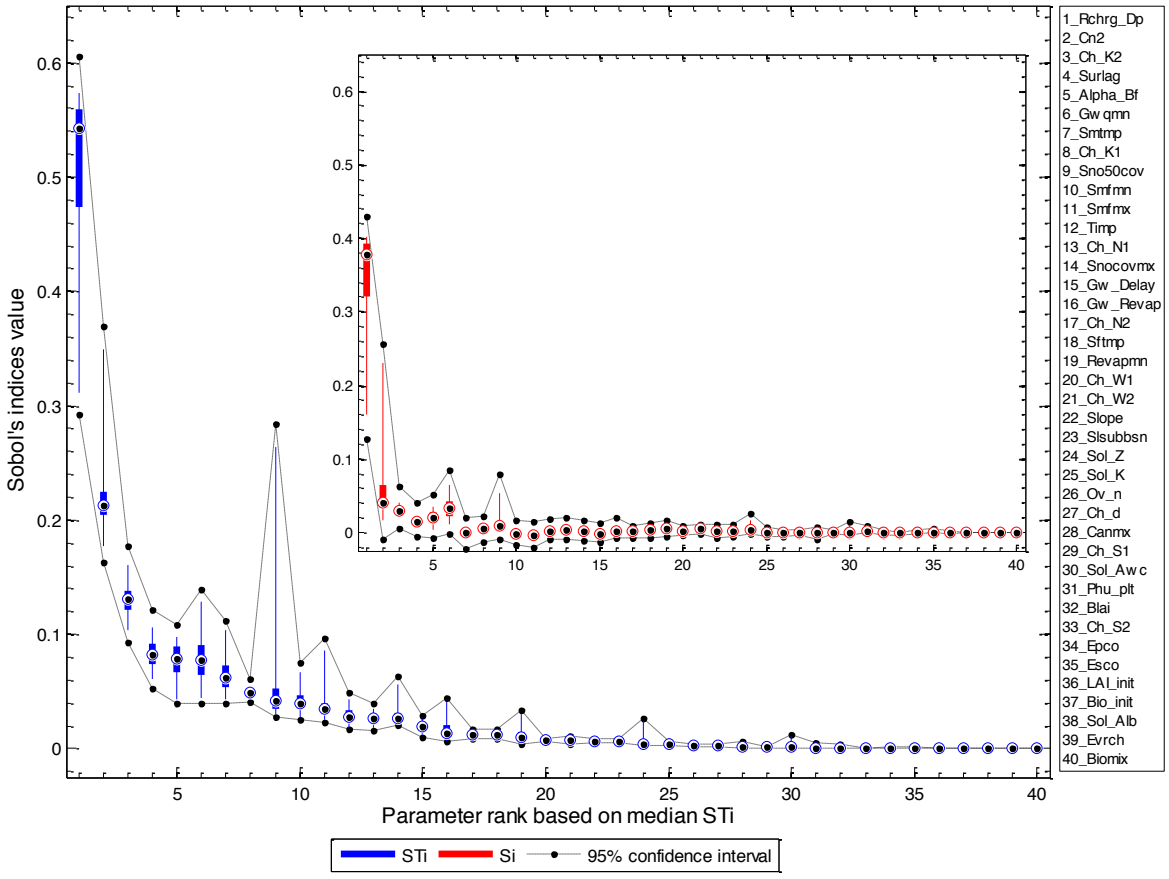

Figure 5: Main (S_i) and total (ST_i) effect values for every parameters and PETF. Boxes include values obtained with every PETF for the corresponding parameter. Dotted lines correspond to the 95% confidence interval for every PETF (interval for which we are confident at 95% that the indices value of the corresponding parameter is situated in this interval for any PETF).

Other influential parameters, for which sensitivity (ST_i) is correlated to the mean PET, are *Revapmn* (r = 0.99), *Gw_Revap* (0.99), *Gwqmn* (0.97), *Alpha*_bf (-0.94), *Ch_N1* (-0.94), *Ch_N2* (-0.92), *Surlag* (-0.92) and *Ch_K2* (-0.91). These results show that PETF impact the hydrological process in the model. Note that the routing parameters (*Ch_N1* and *Ch_N2*) and the transmission losses (*Ch_K2*) are negatively correlated to the mean PET, which could be explained by the fact that a higher PET reduces the availability of water before it is partitioned between soil percolation and surface runoff. The sensitivity (ST_i) of parameters related to ω_{revap} process, which influences the water table height in shallow aquifer, and the

sensitivity of *Gwqmn*, which identifies the threshold water table level for the base flow to occur, are positively correlated to the mean PET showing that dry spells are dominated by groundwater processes. These results stress the importance of an adequate representation of E_a, as PETF directly impacts the relative importance of hydrological processes in the model.

Figure 5 also reveals that the au Saumon SWAT implementation is mainly affected by surface runoff (*CN2, surlag*), groundwater and soil processes (*Rchrg_Dp, Alpha_bf, Gwqmn, Gw_Delay, Gw_Revap, Revapmn, Sol_Z, Sol_Awc*), transmission losses (*Ch_K2, Ch_K1*), snow-related processes (*Smtmp, Sno50cov, Smfmn, Smfmx, Timp, Snocovmx, Sftmp*), and routing (*Ch_N1, Ch_N2*). On the other hand, parameters linked to the channel dimensions/routing (*Ch_W1, Ch_W2, Ch_d, Ch_S1, Ch_S2*), to plant growth (*Phu_plt, Blai, LAI_init, Bio_init, Biomix*), and to E_a calculation (*Canmx, Epco, Esco*) only account for a small portion of the total variance. The fact that parameters linked to plant growth and E_a calculations are less sensitive, and will thus be excluded from the calibration, leaves less flexibility for the model to adapt to different PETF, as they directly affect E_a calculations. This does not discriminate our analysis as a different study in which SA would have been applied before calibration would also have excluded those parameters (for this watershed obviously).

The criteria used here for FP is to preserve at least 90% of the total variance of the model for every PETF, in order not to favour one formulation over another, while keeping a small number of parameters for the calibration. This way, a parameter may be preserved even if only one PETF requires it for its sum of ST_i to exceed 90%. At the end, 21 parameters were kept (Table 2).

3.2 Simulated hydrographs

Figure 6 presents the daily inter-annual hygrographs obtained after calibration for each DSST series. Flow has also been smoothed by using 7-days mean flow to improve visibility. The blue and red lines represent, respectively, the observed flow and simulated flow for the PETF with the best results in terms of NSE_{sqrt}, which corresponds to E-01 for DC and DW, E-21 for HC and E-23 for HW. The dark gray shading represents hygrograms of the corresponding calibration period for every PETF and the light gray shading includes hygrograms of all calibration periods for every PETF (validated in the corresponding period). This figure shows that the model adequately represents the observed hygrogram. One can also see that the highest flows occur during the HC period, which makes sense as peaks flow are always attained for this basin during the spring freshet. Indeed, HC favours higher snow water equivalent levels.

An interesting result from this figure is that simulated flows seem more sensitive to PETF on dry periods than on wet periods which is notable by the larger flow envelop for dry periods. This could be due to the fact that during dry periods, a good distribution of water between the different hydrological processes is necessary considering the limited quantity of water available. During wet periods when water is less limited and when aquifer recharge culminates, the model can, for example, counterpart the effect of higher (lower) PET values by letting less (more) water going to the deep aquifer recharge, which is not possible when the water is limited. Moreover, one can see that for the DW and DC periods, the larger sensitivity to PETF seem to occur during the melting period (April) as well as in November, which might be counter-intuitive as higher PET values are obtained during summer. This observation confirms the impact of the variation of calibration parameterization between the

different PETF having a yearlong effect. We can also see that the period of calibration (light gray shading) also seems to produce the highest variability during the melting period, showing a high sensitivity of the snow melt module to the period of calibration.

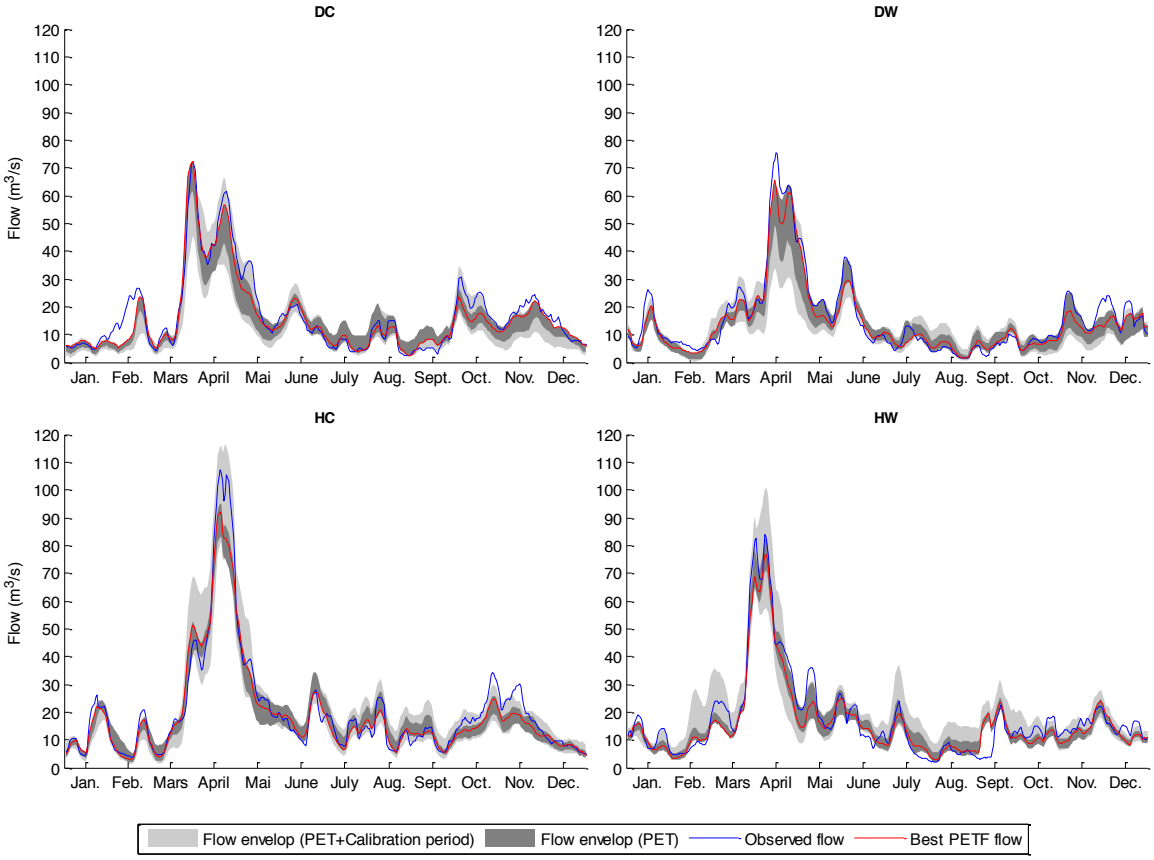

Figure 6: Observed and simulated flow for each validation period. Blue line is the observed flow and red line is the simulated flow for the PETF with the best performance in the corresponding validation period. Dark gray envelop includes flow series for every PETF in the corresponding validation period and light gray envelop includes flow series for every calibration periods and PETF in the corresponding validation period.

The following sections present results from calibration on each DSST periods from different perspectives. The influence of the PETF and calibration period on performance is discussed first, followed by the influence on hydrological process. The influence of PETF and calibration period on parameterization of the model is presented last.

3.3 Temporal transposability

Calibration has been performed for all PETF and five datasets: 4 DSST and all observations (COMP). Validation performances (NSE_{sqrt}) are plotted per DSST period in Figure 7 where the best performance is achieved, as expected, when validating over the same DSST period for which calibration was made. Figure 7 hence illustrates the performance lost due to the temporal transposability of the model and selection of a "suboptimal" PETF. The blue inverted triangles correspond to the calibration in DC, the red circles to DW, the green diamonds to HC, the black squares to HW and the pink right-oriented triangles to COMP. The gray bars represent the variation of performance due to the calibration periods for a specific PETF and validation period.

As it would be very demanding to evaluate a confidence interval for each calibration period and for each PETF, it has been calculated only for E-01 and on the entire dataset (see Figure 7). The selection of this period is based on the assumption that a higher number of points (days) will result in a less robust calibration (for the same calibration method). The E-01 formulation has been chosen because it gave good results for each period of calibration and because it showed the highest level of interaction between parameters in SA (ST_i = 1.84). We assume that a higher level of interaction between the parameters leads to a higher potential of equifinality and thus potentially to a less robust calibration. The confidence interval has been calculated based on 27 calibrations for the E-01 formulation on the entire period. The 27 calibrations have been done by randomly changing the random seed number as to get different calibration pathways. The confidence interval corresponds directly to the difference between the maximum and minimum values for those

calibrations. The horizontal dotted line in Figure 7 corresponds to this interval.

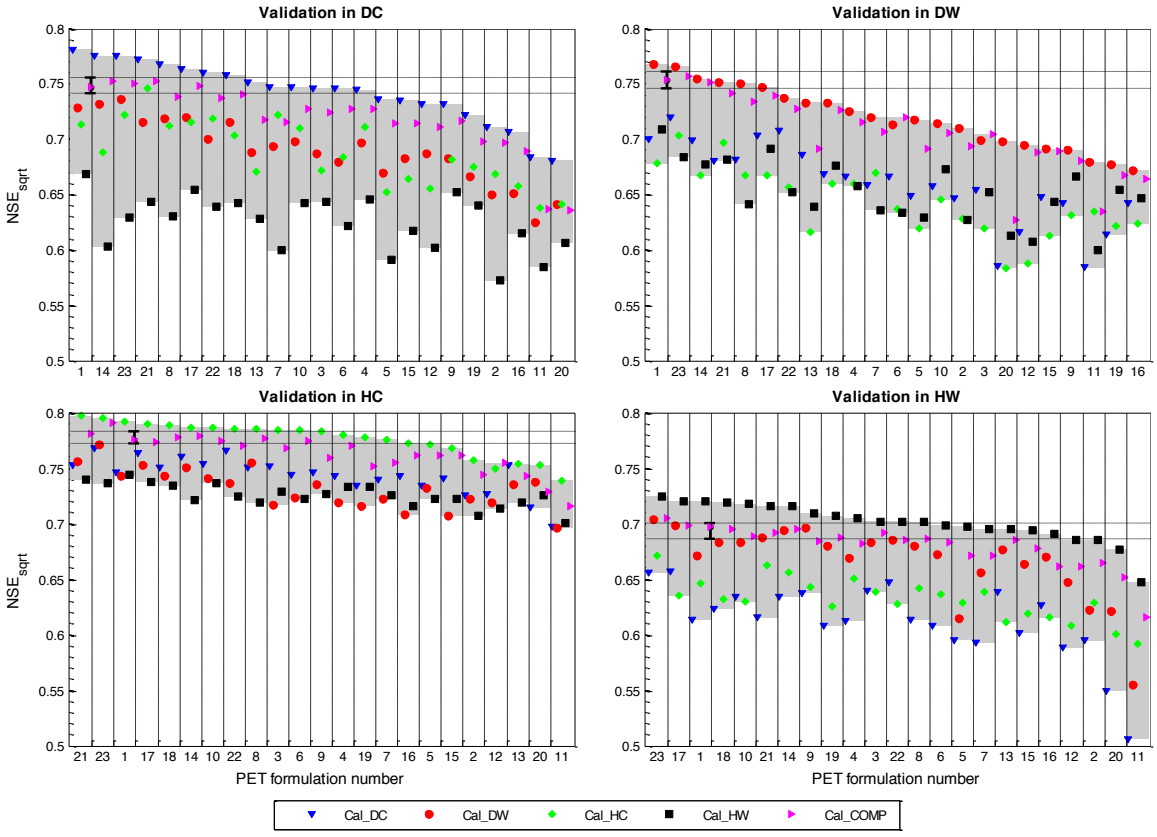

Figure 7: Values of NSE_{sqrt} for every PETF and validation periods. Every type of points correspond to a calibration period (blue inverted triangles refer to calibration in Dry/Cold (DC), etc.). Error bar extended with horizontal dotted line corresponds to 27 calibrations with E-01 on the complete period (1977-2003). PET formulation numbers refer to those from Table 1.

3.3.1 Performances in calibration

Globally, the best performances (at least twice on the four best ranks) are achieved by E-01, E-14, E-17, E-21 and E-23 and the worst (at least twice on the four worst ranks) by E-02, E-11, E-12, E-16 and E-20 when validating on the same periods as the calibration. It is interesting to see that 4 out of the 5 formulations which show globally the best performances are only temperature dependant which is consistent with

findings from Oudin et al. (2005). Also, 3 out of 5 formulations which show globally the worst performances need at least three inputs (*RH, T°, Rs*) and the two others, two inputs (*RH, T°*). It is although reassuring that the formulations giving the best performances for a given calibration/validation period also give good performances for the other calibration/validation periods, showing that relative PETF performance is not much affected by calibration periods (when validating in the same period).

3.3.2 Performances in validation

As expected, the best results are achieved when the calibration period is the same as the validation period and the worst (in most cases) when the calibration period is the opposite climate in terms of DSST. This confirms the relevance of performing a DSST prior to a climate projection in order to assess the uncertainty due to the possible modification in climate statistics. On the other hand, we can also see that the calibration on the COMP period gives almost every time the second best calibration, which means that choosing a sufficiently long period for calibration can effectively result in a more robust calibration and possibly to more reliable results in a projection context. Also, as in Figure 6, we can see that, globally, the sensitivity to PETF is similar to the sensitivity to the chosen calibration period. Indeed, the mean difference of performance between the best PETF and worst PETF for the different validation periods is 0.084 and the mean difference between the best and worst performance of a given PETF for all PETF and all validation periods is 0.087. In comparison, the mean value of the difference between the best and worst performance for the 27 calibrations of E-01 when validating over the four DSST periods gives 0.014. This clearly shows that the variability in performance is mainly due to the PETF and the calibration periods and not from the calibration process.

Even if the sensitivity to the calibration periods and PETF is similar, we can see that it varies from one validation period to another. Indeed, it is clear from the results illustrated in Figure 7 that the variation in NSE due to the calibration period for the validation in HC is smaller compared to other validation periods. Based on the mean difference between the maximum and minimum values of the NSE_{sqrt} for each PETF, we get values of 0.12, 0.08, 0.05 and 0.09, respectively for DC, DW, HC, and HW. This means that the HC period is indeed less sensitive to the calibration periods than the other validation periods. This might be due to the bad transposability of parameterization issued from the calibration in the HW period. Even for the HC and DW periods, this calibration period often performs as bad or even in some case worse than the calibration in the complete opposite climate in terms of DSST. Moreover, when validating with the COMP period (not shown here), the HW period performs almost always the worst. As DW and HC still have more affinity with the HW period (temperature or precipitation being statistically closer) than the DC period, this could explain the lower sensitivity to calibration for the DW and HC periods. Parameterization results (section 3.5) will effectively show that the HW period shows extreme values (relative to other periods) for many parameters, particularly for those related to groundwater and snow module. It is necessary to nuance these results by the fact that the NSE criterion is weighted by the variance of the observed flow. Indeed, a very large variance of the observed flow will tend to tamper the variability when calculating the NSE, since a higher numerator will be needed to obtain the same absolute variation in NSE.

The sensitivity to PETF was found higher for dry periods than for humid ones. Indeed, the differences between the maximum and minimum NSE_{sqrt} for a given calibration and validation period and for all PETF are 0.14, 017, 0.06 and 0.08, respectively for DC, DW, HC and HW. This

higher sensitivity to PETF in dry periods may be explained by the limited quantity of water during those periods, resulting in the need for a better representation of E_a. Indeed, we think that when water is somehow limited, the model has less adaptability as, trivially, no process creates water, thus no parameter can virtually add water to the model when needed.

In terms of robustness (inter-PETF worst performance comparison for a given validation period), we can see that the formulations giving the best results when validating in the same period as the calibration period are not necessarily the ones showing more robustness. As an example, the E-09 formulation ranks 18th for the calibration and validation in DC, but occupy the 3rd rank for the calibration on HW and validation on DC, the latter rank corresponding to the robustness of E-09. This shows that the choice of PETF is non-trivial as the relative performance of a formulation in calibration does not necessarily reflect the relative performance of this formulation when validating on an opposite period in terms of DSST.

3.3.3 Hypothesis concerning inter-PETF performance variability

In an attempt to explain the performance variability between PETF, Figure 8 presents the daily mean inter-annual values of PET, E_a as well as every component of E_a as described in Figure 2. The line colors go from green for the best PETF in terms of performance to red for the worst one. Note that the y-axis scale is different for the PET-related graphic than for the other components in order to improve readability. Based on Figure 8, the poor performance of some PETF is attributed to winter PET production. Indeed, we see that the best formulations are those who produce low winter PET. As this figure shows, winter PET translates into winter E_a, especially *CANEV* and *SEV*, which lowers summer water availability to *TRANSP,* hence producing not enough E_a in summer. We can see that

some PETF producing a lot of PET in summer and giving mediocre results actually produce relatively low E_a in summer. Figure 8 clearly shows that this lower E_a is due to lower values of transpiration (T) for those PETF, which may result from lower soil moisture content or a lower *LAI* resulting from water stress (water stress is correlated to the ratio of T and *ep_max*). Another cause for lower performance may come from the fact that some formulations produce summer PET that is too low for the model to correctly reproduce summer flows.

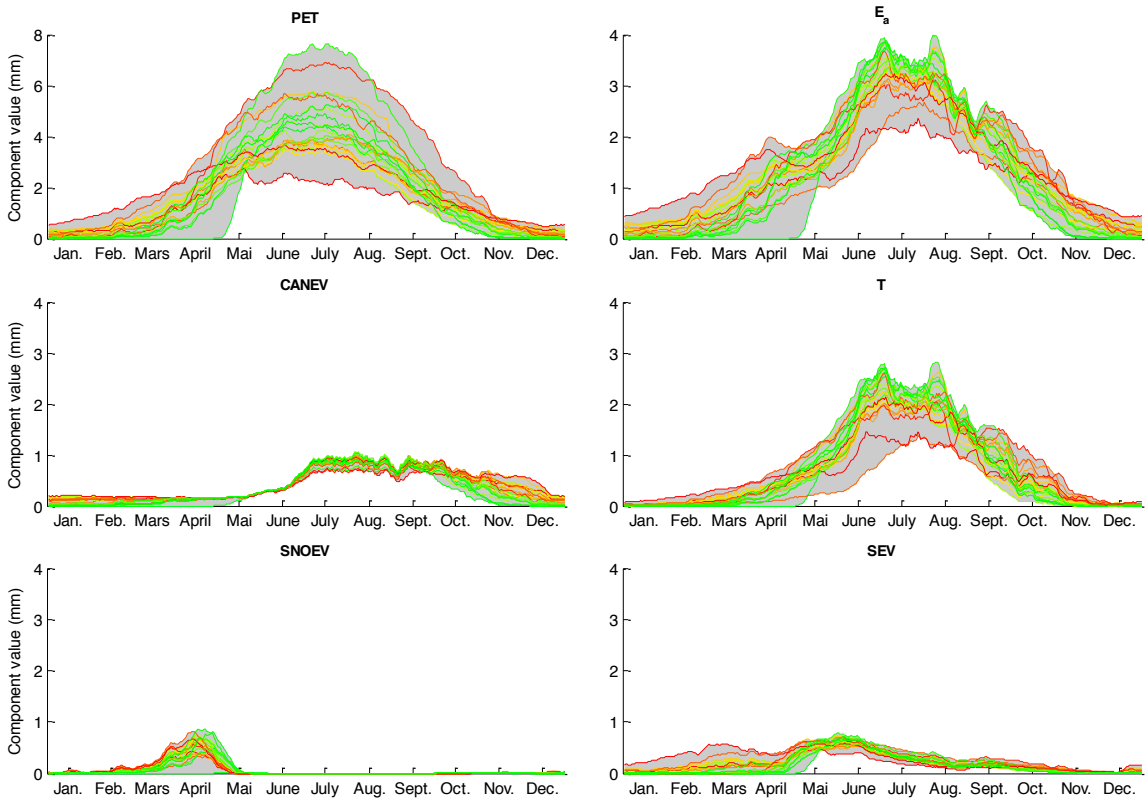

Figure 8: Daily inter-annual mean component values of E_a for every PETF (one line per PETF) on the complete period (1977-2003). Formulation performances go from green for best PETF to red for worst PETF (NSE_{sqrt}).

3.4 Water partitioning

Figure 9 illustrate water partitioning based on equation (4) in COMP (1977-2003) for every PETF (x-axis) and DSST. PETF are drawn in

descending order of performance while the white dotted line illustrates the PETF relative performance from 0 (worst) to 1 (best). The panels of Figure 9 correspond to different periods of calibration although the water balance is always calculated on COMP. The panel entitled *Calibration in COMP (E1 only)* groups the 27 repetitions of the E-01 calibration on COMP, illustrating the influence of the calibration process on water partitioning. The sum of every component on total precipitation ranged from 1.0001 to 1.0003 for every formulation and DSST, confirming that the balance is effectively closed.

Figure 9 shows the clear influence of PETF and DSST on the partition of water. Indeed, for a given calibration period, the absolute difference in Q_{surf} contribution to the total water balance between PETF is 6% for the best case (DC) and 10.8% for the worst case (DW). Relative to the mean value of the fraction of this component for the same periods (shown in parenthesis thereafter), we get a value of 15% for the best case (DC) and 28% for the worst case (DW). As a comparison, the same component gives a difference of 3.6% (8.9%) for the 27 calibrations on the same PETF. For Q_{gw}, which is another important contributor to total flow, we get 5% (40.8%) in the best case (COMP) and 10.6% (69.8%) in the worst case (DW), while the same component gives 3% (24.8%) for the 27 calibrations on the same PETF. This shows that the uncertainty arising from the PETF is not negligible. Repeating the same exercise, but considering the effect of the calibration period, led to a maximum difference of 13.5% (34.5%) for Q_{surf} and 15.7% (129.1%) for Q_{gw}.

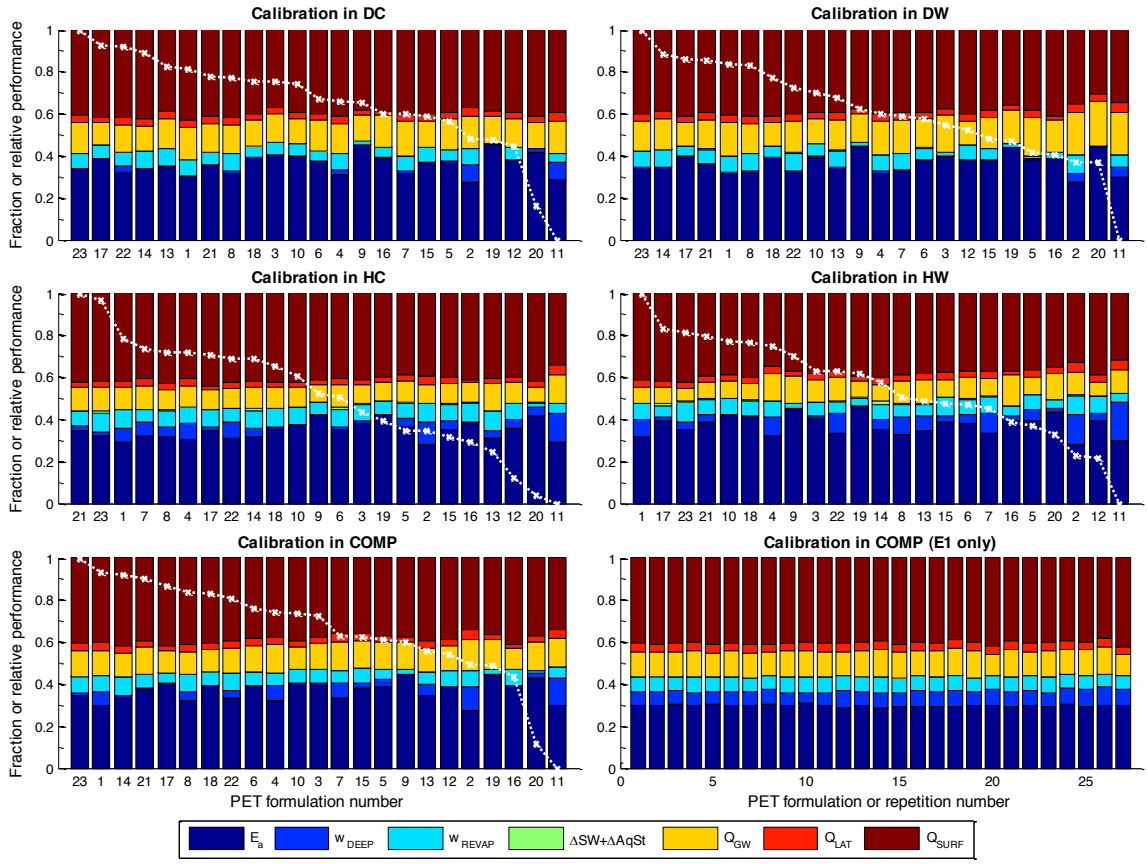

Figure 9: Hydrological component fraction for different calibration periods with relative performance (white dotted line) of every PETF. Bar colors (stacked) correspond to specific hydrological component as a fraction of total precipitation in the complete period (1977-2003). The lower right-end graph corresponds to results of the 27 calibrations with E-01 formulation in the complete period.

Figure 9 also reveals that the relative importance of the Q_{gw} component is higher when calibrating over dry periods than for humid ones: the groundwater flow gain in importance for dry periods. Calibrating the model in those periods seems to result in a parameterization favouring the groundwater flows.

Finally, Figure 9 shows that there is no clear relation between the volume of evapotranspiration and the performance of the PETF. The performance of PETF is thus mainly linked to the production of PET in winter and not to its volume. Calculating the adjusted coefficient of

determination ($R^2_{adjusted}$) between the NSE_{sqrt} and the contribution of each component to the water budget shows that, for DC ($R^2_{adjusted}$=0.64) and DW ($R^2_{adjusted}$=0.71), performance is mostly correlated to w_{deep} and w_{revap}, while for HC performance, it is mostly correlated to Q_{LAT} and Q_{surf} ($R^2_{adjusted}$=0.51) and, for HW, to Q_{surf} ($R^2_{adjusted}$ =0.52). This, again, supports the idea that dry events are more driven by groundwater flow and wet events are more driven by surface flow.

3.5 Parameterization

The influence of PETF and DSST on parameterization explains much of the presented results. Figure 10 illustrates the values of the calibrated parameters for all conditions tested. Again, the blue inverted triangles identifies DC calibration, the red circles DW, the green diamonds HC, the black squares HW and the pink right-oriented triangles COMP. The gray bars illustrate the DSST variation for each PETF. Finally, red lines illustrate the variability arising from the 27 calibrations of E-01 on COMP.

3.5.1 PETF influence

Figure 10 reveals a very large variability in parameter values, although some are more constant. Calculating the variance of values scaled from 0 to 1 with their respective boundaries and taking the mean value of this variance for every calibration identifies parameters with the highest inter-PETF variability: *Gw_revap* (9.0×10^{-2}), *Gw_Delay* (7.8×10^{-2}), *Ch_N1* (5.1×10^{-2}), and *Revapmn* (4.8×10^{-2}). On the other hand, parameters showing lower variability are *Alpha_Bf* (5.1×10^{-6}), *Ch_K1* (1.0×10^{-3}), *Surlag* (1.9×10^{-3}), *Smfmx* (2.1×10^{-3}), and *Smtmp* (2.4×10^{-3}). Higher variability does not necessarily mean that this parameter is more influenced by PETF as it might as well be caused by the calibration process due to a lesser

sensitivity of the parameter, to an imperfect calibration or to interaction between parameters (equifinality).

To evaluate the impact of the calibration process on the variance of the parameters, a KS test has been performed between the series of parameter values from the 27 calibrations of E-01 with the COMP period and the 23 calibrations on different PETF for the same period. Results from this test, in terms of *p-values*, are grouped in Table 4. From the four parameters showing the highest variability, only *Gw_revap*, *Gw_Delay*, and *Revapmn* are significantly (*p-values* inferior to 0.05) linked to the PETF. *Gw_revap* is associated to w_{revap}, *Gw_Delay* to the time delay in aquifer recharge, and *Revapmn* to the water level threshold of shallow aquifer for w_{revap} to occur. This result is quite interesting as *Gw_revap* and *Revapmn* are the only parameters directly used to calculate w_{revap} and also showed to be highly correlated to mean PET in the SA, meaning that it is highly influenced by the PETF. This result is expected as this process is associated to the volume of water moving into the soil in response to water deficiencies that will consequently impact the water availability in the soil layers. Higher PET might thus cause higher water deficiencies favouring w_{revap} but also limit the percolation from the soil layers, limiting w_{revap} as the aquifer recharge will be lower (w_{revap} only occurs if the amount of water in aquifer is higher than *Revapmn*). Calculating Pearson correlation coefficient (r) between the mean PET and the mean w_{revap} for different PETF over the COMP period gives -0.54, suggesting that the second effect of PETF on w_{revap} (percolation limitation followed by w_{revap} limitation) is more important than the first effect (higher deficiencies followed by higher w_{revap}) for this case. No explanation has been found for the relation between *Gw_Delay* and PETF.

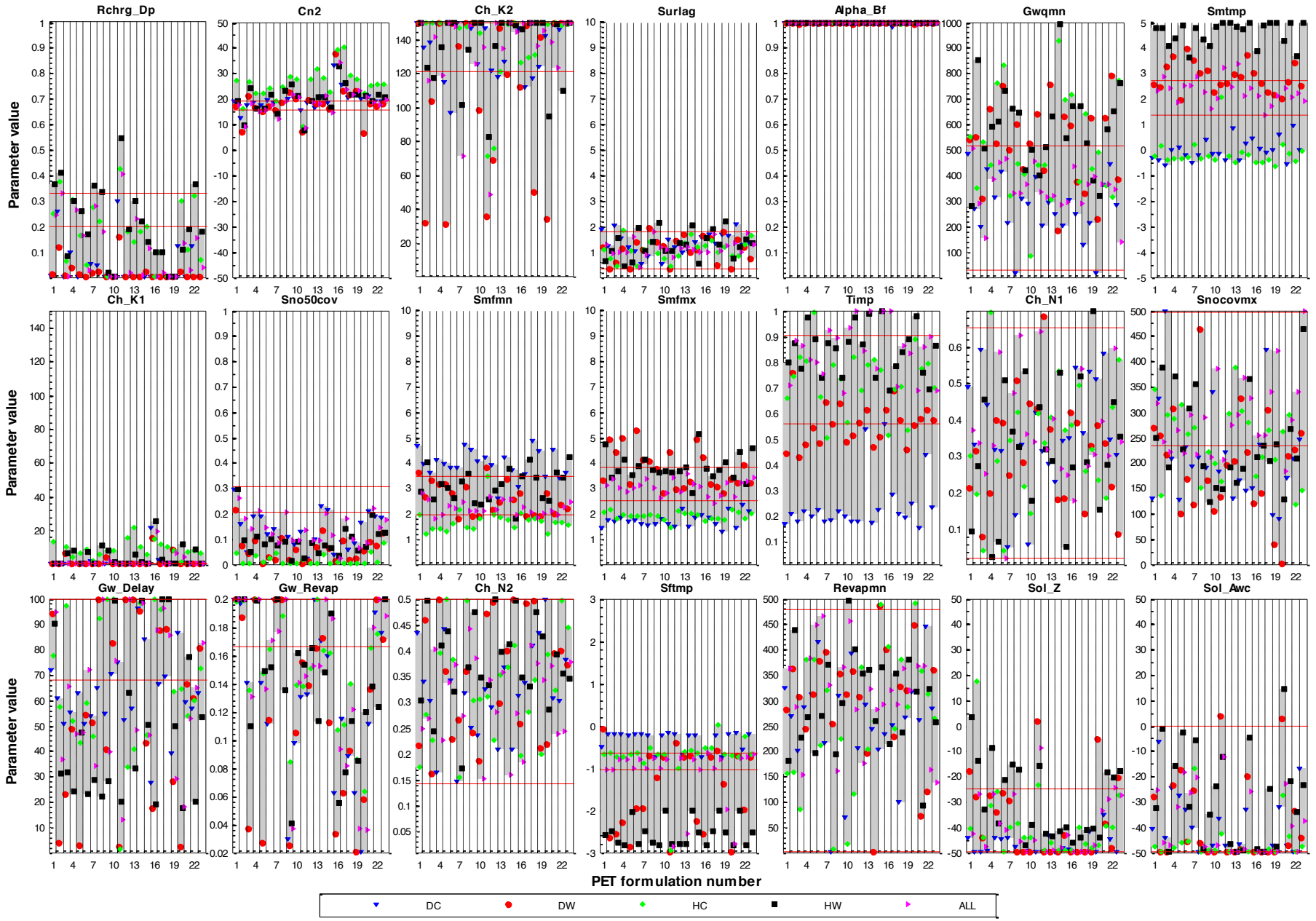

Figure 10: Parameter values comparison for every calibration period and every PETF. Each type of point corresponds to a specific calibration period (red circle for calibration in Dry/Warm (DW), etc.).

Other parameters that do not show very high variability but are significantly influenced by PETF are *Rchrg_Dp* ($p = 7.78 \times 10^{-6}$), *Sno50cov* (2.4×10^{-10}), and *Timp* (2.4×10^{-4}). As explained earlier, *Rchrg_Dp* directly affects the fraction of the aquifer recharge that is transmitted to the deep aquifer. It is thus expected that this parameter is influenced by PETF, as higher (lower) PET results in less (more) water for deep aquifer recharge. Indeed, the Pearson correlation coefficient (r) between this parameter values and the mean PET for the COMP period gives -0.67, showing that higher (lower) PET effectively results in lower (higher) *Rchrg_Dp* values, and thus to a lower (higher) portion of aquifer recharge transmitted to the deep one. Influence of PETF on *Sno50cov* and *Timp* is difficult to assess as those parameters are snow-related.

Cn2 is also correlated to the mean PET ($r = 0.89$). As explained in section 3.1, PETF producing more PET leave less water available to partitioning, as E_a is applied beforehand. A positive correlation between *Cn2* and mean PET is thus expected since, when less water is available to the partitioning, higher *Cn2* values will be needed to produce a similar surface runoff.

3.5.2 DSST influence

Table 4 also identifies the highest significant difference for all DSST pairs (if any). For example, 6 KS tests have been performed for *Rchrg_Dp*, i.e. all combinations of the 4 DSST. For this parameter, the lowest *p-value* (2.20×10^{-6}) occurs with DW/HW, showing that this parameter is mainly affected by precipitation. Indeed, as shown in Figure 10, this parameter culminates for almost all PETF in HW, while it is the opposite in DW. The parameter is also almost always higher in HC than in DC.

Table 4: Statistical analysis of parameter values for the different calibration periods and PETF. Every *p-value* corresponds to results from Kolmogorov-Smirnov tests. *P-values* of second column correspond to comparison between associated parameter values obtained with the 27 calibrations of E-01 and with all PETF in the complete period. *P-values* of the fourth column correspond to comparison between associated parameter values for the pair of calibration period showing the largest significant difference. See section 2.9 for $R^2_{adjusted}$ meaning.

Parameter name	*p-value* comparing E-01 calibrations to all PETF ones	DSST Pair showing the largest significant difference	*p-value* for this pair	$R^2_{adjusted}$ for the evaluation of interaction between parameters	Number of parameter used in $R^2_{adjusted}$	Correlated parameters
Rchrg_Dp	7.78×10^{-07}	DW/HW	2.20×10^{-06}	0.64	3	Cn2, Ch_N1, Sol_Awc
Cn2	1.21×10^{-02}	DC/HC	4.18×10^{-07}	0.75	4	Ch_K2, Ch_N1, Gw_Delay, Sol_Z
Ch_K2	2.42×10^{-02}	-	9.58×10^{-02}	0.72	4	Cn2, Ch_N1, Gw_Delay, Gw_Revap
Surlag	1.43×10^{-01}	-	9.58×10^{-02}	0.98	2	Ch_N1, Gw_delay
Alpha_Bf	5.76×10^{-02}	HC/DW	4.28×10^{-02}	0.00	0	-
Gwqmn	1.25×10^{-01}	DC/HW	4.18×10^{-07}	0.60	2	Cn2, Revapmn
Smtmp	5.95×10^{-02}	HC/HW	2.48×10^{-11}	0.94	4	Smfmn, Smfmx, Snocovmx, Timp
Ch_K1	3.17×10^{-01}	DC/HC	4.18×10^{-07}	0.26	2	Surlag, Ch_N1
Sno50cov	2.40×10^{-10}	DC/HC	1.13×10^{-08}	0.85	4	Gw_Delay, Smfmx, Gw_Revap, Snocovmx
Smfmn	1.75×10^{-03}	DC/HC	2.10×10^{-10}	0.91	4	Smtmp, Smfmx, Snocovmx, Timp
Smfmx	7.14×10^{-01}	HC/HW	2.48×10^{-11}	0.91	4	Smfmn, Smtmp, Snocovmx, Timp
Timp	2.40×10^{-04}	DC/HW	2.48×10^{-11}	0.73	4	Smfmn, Smtmp, Smfmx, Snocovmx
Ch_N1	8.17×10^{-01}	-	3.60×10^{-01}	0.98	4	Surlag, Gw_delay
Snocovmx	3.17×10^{-01}	DC/HC	4.28×10^{-02}	0.92	4	Smfmn, Smtmp, Smfmx, Timp
Gw_Delay	2.40×10^{-04}	DC/HW	6.55×10^{-04}	0.79	4	Cn2, Ch_K2, Ch_N1, Sol_Awc
Gw_Revap	7.35×10^{-04}	-	1.95×10^{-01}	0.73	3	Rchrg_Dp, Ch_K2, Sol_Awc
Ch_N2	3.54×10^{-01}	DC/HW	4.28×10^{-02}	0.00	0	-
Sftmp	1.04×10^{-02}	DC/HW	2.48×10^{-11}	0.33	3	Smtmp, Smfmx, Sol_Z
Revapmn	8.56×10^{-03}	-	9.58×10^{-02}	0.61	2	Gwqmn, Sol_Z
Sol_Z	4.04×10^{-02}	DC/HW	2.20×10^{-06}	0.64	3	Cn2, Ch_N1, Sol_Awc
Sol_Awc	4.61×10^{-02}	HC/DW	6.55×10^{-04}	0.56	3	Gwqmn, Gw_Delay, Sol_Z

This is explained by the fact that, during dry periods, less water is available for deep aquifer recharge, resulting in lower *Rchrg_Dp*. During wet events, more water is available and favours higher *Rchrg_Dp*. This also explains why, when calibrating on a dry period and validating on COMP, w_{deep} correlation to performance increases. Indeed, COMP/DW and COMP/HW give *p-values* of 6.4×10^{-3} and 9.6×10^{-2} respectively, showing that only DW values are significantly different than those for the COMP period.

A similar analysis for *Cn2* reveals why, for humid periods, Q_{surf} is more correlated to the performance than any other water budget component. Indeed, as already mentioned, *Cn2* is associated to permeability, land use, and antecedent soil moisture conditions, and is mainly used to calculate the surface runoff (Q_{Surf}). In this case, the lowest *p-value* (4.18×10^{-7}) occurs with DC/HC, showing that this parameter is, as *Rchrg_dp*, mainly affected by precipitation. COMP/DC and COMP/HC give *p-values* of 2.0×10^{-1} and 2.2×10^{-6} respectively, stating the importance of selecting a calibration period statistically similar to the projection one. Otherwise, parameterization might need to be dynamically updated as a function of the climate conditions, as already pointed out by Cibin et al. (2010).

The impact of air temperature is also shown in Figure 10, notably for snow parameters *Smtmp*, *Smfmx*, *Timp*, and *Sftmp*. *Smtmp* and *Sftmp* are the snowmelt threshold temperature and mean temperature at which precipitation is equally likely to be liquid or solid. *Smfmx* and *Timp* are the melt factor for June 21st and snow temperature lag factor.

As an example, for *Smtmp*, values are always higher for warmer periods than for colder ones, indicating that snow melt occurs at lower

temperature for colder periods than for warmer ones. Indeed, looking back at Figure 6, we notice that for warm periods higher flows are obtained in February and Mars when using calibration of other periods (pale gray envelop) which could be attributed to lower *Smtmp* values. It is although difficult to explain and demonstrate how snow melt parameters act on hygrographs because of the high level of interaction (as will be showed below) between them. It is nevertheless clear that calibrating on different climatic periods affects snow parameters.

Gwqmn is also influenced by the calibration period – the lowest *p-value* (4.18×10^{-7}) occurs with DC/HW. It is the threshold water level in the shallow aquifer for base flow (Q_{gw}) to occur. Its lower DC values explain why Q_{GW} account for a higher proportion of the water budget than for humid periods. High values are also obtained in DW, but this could be explained by other parameters which tend, in DW, to favour shallow aquifer recharge. Indeed, *Cn2* values are then relatively lower and favour recharge. We can detect in DW that *Rchrg_Dp* is very low for most PETF, which diminishes the water transferred to the deep aquifer and favours higher water content in the shallow aquifer, explaining higher *Gwqmn*, even if a high Q_{GW} contribution to the water budget is observed.

3.5.3 Equifinality

As shown above, evaluating the impact of PETF and calibration periods is relatively easy but evaluating calibrated parameter values impact on hydrological processes, to get a better understanding of those impacts, is way more complicated. This is partially due to the fact that parameters interact with each other, which has been clearly shown in SA. These interactions can also cause higher variability in parameter values as interactive parameters can compensate each other. As the variability in

parameterization is very high for the 27 COMP calibrations of E-01 (Figure 10), the part of the variability which may be attributed to parameter interactions has been evaluated by calculating the adjusted coefficient of determination ($R^2_{adjusted}$) for combinations of up to 4 parameters (i.e. regression of a parameter value with up to 4 predictor variables) for parameter values issued from these 27 E-01 calibrations. The limitation of 4 predictors is based on Knofczynski and Mundfrom (2008), which suggested that, for a good prediction level with R^2 of at least 0.7, the number of samples should be 25 (27 samples used for this study). Combinations with the highest $R^2_{adjusted}$ have been retained. Also, only the combinations with *p-values* lower than 0.05 (*t-statistic*) for every predictor have been kept.

Table 4 shows the $R^2_{adjusted}$ for every parameter and for the best combination and also presents the number of parameters used in the regression. For parameters with $R^2_{adjusted}$ lower than 0.7, the maximum number of parameters used for the regression has been lowered. Finally, parameters with $R^2_{adjusted}$ values lower than 0.6 have been judged nonsignificant as a higher number of samples would then be required ($R^2_{adjusted}$ in red). Parameters with $R^2_{adjusted}$ = 0 have no combination of predictors resulting in *p-values* lower than 0.05 (*t-statistic*). It is important to note that the part of variability of the parameters which is not explained by interactions, may be attributed to an imperfect calibration process.

These results will not be discussed in details here as this is not directly linked to our objectives, but are presented to justify and explain the high variability in parameterization resulting from the 27 COMP calibrations of E-01, as well as to justify the difficulty putting in relation parameterization as well as water partitioning. It is indeed interesting to see that, for many parameters, at least 70% of the variability can be

attributed to interactions with other parameters. Moreover, we can see that snow-related parameters have very high $R^2_{adjusted}$ especially *Smtmp* (0.94), *Smfmn* (0.91), and *Smfmx* (0.91). This might seem trivial as those parameters are all used in snowmelt routines but it shows that calibrating multiple parameters with interactions between them and used for the same purpose (evaluating snow melt, for example) creates an equifinality problem making it difficult to assess the impact of calibration periods or PETF for such a complex model. Another interesting result taken from Table 4 is the strong correlation between *Surlag* and *Ch_N1*. In fact, calculating the correlation only between these two parameters gives an R^2 value of 0.96. This result shows how much parameters can compensate one another. This strong correlation is explained by the fact that *Surlag* (surface runoff lag coefficient) affects the release delay of surface runoff to the main channel, while *Ch_N1* is the Manning coefficient used to calculate the time of concentration of tributary channels. Both parameters thus have a similar effect on the hydrological processes and may thus compensate each other. Finally, one may note again that *Gwqmn* is correlated with *Cn2*, confirming the hypothesis of interaction stated earlier.

3.5.4 Partial conclusion

This section showed how difficult it can be to decipher a complex model like SWAT. Indeed, the impact of PETF and calibration periods is easy to assess in terms of performance but the problem of equifinality makes it difficult to identify how and why these impacts occur. For example, it has been possible to relate that calibrating on a dry season produces a higher groundwater flow contribution to the water balance than when calibrating on wet years (for a validation in the same period), but explaining this finding based on the parameterization turned out difficult because of the length of the causal chain and the fact that many

factors included in this chain compensate each other (for example, *Cn2* affects precipitation partitioning, many parameter influence the volume of water percolating through the shallow aquifer, others dictate how much water '*revap*', more affect groundwater flow, etc.). This equifinality problem also shows how a physical model can, in some ways, act as a black box as some influential parameters which are difficult or even impossible to measure interact with other parameters, making it hard to prove that we obtain good results for the right reasons and that processes simulated by the model really represent the physical reality of the catchment.

Chapter 4

Conclusion & recommendations

4.1 Conclusion

4.1.1 Sensitivity analysis

SA has been performed on each of 23 PETF using Sobol' method. The results showed that:

- PETF directly influences the parameter sensitivity such that, for formulations leading to lower mean PET, the model relies more on surface processes, while it relies more on groundwater processes for the opposite;
- the au Saumon catchment is mainly influenced by parameters related to surface runoff, groundwater and soil processes, transmission losses, snow-related routines, and routing. On the other hand, parameters linked to channel dimensions, plant growth, and E_a calculation account for a much smaller percentage of the total model variance.

4.1.2 DSST performance

Five calibrations, one for each of the 4 climatologically contrasted periods (DSST) and one for the complete data set (1977-2003), have been performed for each PETF (23) resulting in a total of 115 calibrations. The analysis of the performance brought the following observations:

- four out of five PETF leading to the best performance are temperature-based while the worst ones need two inputs or more;
- the variability in performance due to PETF (average of 8.4% in NSE) and calibration period (average of 8.7% in NSE) are similar and about 6 times higher than the one associated to the calibration process;
- validating in HC and DW showed less sensitivity to calibration periods in terms of performance than when validating in DC and HW;
- dry periods are more sensitive to PETF than humid ones in terms of performance;
- the relative performance of a PETF in calibration does not necessarily reflect the relative performance of this formulation when validating on an opposite DSST period;
- the reason of a PETF showing a worse performance seems to be mainly correlated to the winter PET production, which results to higher E_a in winter and lower transpiration (T) in summer;
- performances for all DSST reveal that, for the au Saumon catchment, E-01 and E-23 PETF should be preferred and E-11 and E-20 formulations should be avoided. It is interesting to note that E-23 was developed specifically for Canadian catchments based on linear regressions, showing the pertinence of using locally developed formulations for evapotranspiration modelling. Studies on other Canadian catchments would be necessary to support this conclusion, although Seiller et al. (2012) arrived at a similar conclusion for the same watershed, but resorting to 20 lumped hydrological models.

4.1.3 Water partitioning

An equation has been derived from SWAT theoretical documentation to assess the influence of PETF on water partitioning. Looking at the relative contribution of different hydrologic processes to the total water budget, the following observations were made:

- the relative importance of the groundwater flow to the total water budget is higher when calibrating on dry periods than on humid ones (for the same validation period);
- PETF causes an absolute variation in surface runoff (Q_{surf}) and base flow (Q_{gw}) contributions to the total water budget ranging respectively from 6% to 10.8% and from 5% to 10.6% depending on the calibration periods, from which about 3.6% (Q_{surf}) and 3% (Q_{gw}) could be attributed to the calibration process. Including the impact of the calibration periods gives a maximum absolute variation of 13.5% and 15.7% respectively for Q_{surf} and Q_{gw};
- performance is more linked to groundwater components (w_{deep} and w_{revap}) for dry periods and more linked to surface component (Q_{surf}) for humid ones.

4.1.4 Parameterization

Optimal parameter sets issued form the 115 calibrations have been analyzed. The main conclusions from this analysis were that:

- PETF affects groundwater-related parameters (*Gw_Delay, Revapmn, Gw_revap, Rchrg_Dp*), snow-related parameters (*Sno50cov, Smfmn, Timp, Sftmp*), and *Cn2*;

- the period of calibration affects parameterization: *Rchrg_Dp* and *Cn2* are more influenced by precipitation while *Smtmp*, *Smfmx*, *Timp*, and *Sftmp* by temperature;
- the high level of interactions is reflected in the high correlation between parameters of the 27 COMP calibrations with E-01. Indeed, the fact that there is such a high variation in parameters and small variation in the objective function shows that the effects of many parameters compensate one another. This equifinality problem shows how a physical model can, in some ways, act as a black box, as some influential parameters that are difficult or even impossible to measure interact with others, making it hard to ascertain that we end up with good results for the right reasons and that the processes simulated by the model represent the physical reality of the catchment.

4.2 Recommendations

PETF influences performance and the hydrological processes simulated by the model. Fluctuation in the parameterization suggests that for a good representation of the actual evapotranspiration, the model must be reliable in terms of water partitioning. This might translate in developing novel methods for actual evapotranspiration calculations and going through extensive testing to guarantee the ability to correctly simulate the evapotranspiration processes under dissimilar climatic conditions (spatial and temporal). Applying a similar methodology to other catchments could also provide a deeper understanding of the reasons why some PETF perform better than others.

Many results illustrate that parameterization identified for specific conditions (calibration periods) shows specificity related to those

conditions, suggesting the necessity of dynamically adapting the model. This could be performed by applying DSST parameter sets as a function of the climatic conditions encountered during simulation in order to reduce uncertainties due to the specificity of the calibration dataset. As we normally calibrate a model to reflect the specificity of a watershed, it seems also logical to dynamically adapt model parameters as a function to the specificity of the simulated periods.

Acknowledgements

The authors acknowledge NSERC, Ouranos, and Hydro-Québec for support, as well as partners of the QBIC^3 (Quebec-Bavaria International Collaboration on Climate Change) project.

References

Abbaspour KC, Yang J, Maximov I, Siber R, Bogner K, Mieleitner J, Srinivasan R. 2007. Modelling hydrology and water quality in the pre-alpinealpine Thur watershed using SWAT. Journal of Hydrology 333, 413-430.

Allen RG, Walter IA, Elliott R, Howell T, Itenfisu D, Jensen M. 2005. The ASCE standardized reference Evapotranspiration equation.

Andréassian V, Perrin C, Michel C. 2004. Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. Journal of Hydrology, 286(1-4), 19–35.

Arnold JG, Allen PM, Bernhardt G. 1993. A comprehensive surface-groundwater flow model. Journal of Hydrology, 142, 47-69.

Baumgartner A, Reichel E, Lee R. 1975. The world water balance: mean annual global, continental and maritime precipitation, evaporation and run-off.

Chiew F, McMahon T. 1994. Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments. Journal of Hydrology, 153(1-4), 383–416.

Cibin R, Sudheer KP, Chaubey I. 2010. Sensitivity and identifiability of stream flow generation parameters of the SWAT model. Hydrological Processes, 24, 1133–1148.

Choi HT, Beven K. 2007. Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. Journal of Hydrology, 332, 316-336.

Coron L, Andreassian V, Perrin C, Lerat J, Vaze J, Bourqui M, Hendrickx F. 2012. Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. Water Resources Research, 48, W05552.

Duan Q, Gupta V. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources, 28(4), 1015–1031.

Duan Q, Sorooshian S, Gupta V. 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology, 158, 265–284.

Feyereisen GW, Strickland TC, Bosch DD, Sullivan DG. 2007. Evaluation of SWAT Manual Calibration and Input Parameter Sensitivity in the Little River Watershed. American Society of Agricultural and Biological Engineers, 50(3), 843-855.

Glen G, Isaacs K. 2012. Estimating Sobol sensitivity indices using correlations. Environmental Modelling & Software, 37, 157-166.

Green CH, van Griensven A. 2008. Autocalibration in hydrologic modeling: Using SWAT2005 in small-scale watersheds. Environmental Modelling & Software, 23, 422-434.

IPCC. 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jansen MJW. 1999. Analysis of variance designs for model output. Computer Physics Communications, 117, 35-43.

Kirchner JW. 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42, W03S04.

Klemeš V. 1986. Operational testing of hydrological simulation models. Hydrological Sciences, 31(1), 13–24.

Knofczynski GT, Mundfrom D. 2008. Sample Sizes When Using Multiple Linear Regression for Prediction. Educational and Psychological Measurement, 68(3), 431-442.

Koutsoyiannis, D. 2011. Hurst-Kolmogorov dynamics and uncertainty. Journal of the American Water Resources Association (JAWRA), 47(3), 481-495.

Liburne L, Gatelli D, Tarantola S. 2006. Sensitivity analysis on spatial models: new approach. Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.

McKay MD, Beckma RJ, Conover WJ. 2000. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code. Technometrics, 42, 55-61.

Neitsch SL, Arnold JG, Kiniry JR, Williams JR. 2011. Soil and Water Assessment Tool- Theoretical Documentation - Version 2009 TWRI Report (Vol. TR-406). Texas: Texas Water Resource Institute, College Station.

Nossent J, Elsen P, Bauwens W. 2011. Sobol' sensitivity analysis of a complex environmental model. Environmental Modelling & Software, 26, 1515-1525.

Oudin L, Hervieu F, Michel C, Perrin C, Andreassian V, Anctil F, Loumagne C. 2005. Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of Hydrology, 303(1-4), 290–306.

Oudin L, Andréassian V, Mathevet T, Perrin C, Michel C. 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. Water Resources Research, 42, W07410.

Rana G, Katerji N. 2000. Measurement and estimation of actual evapotranspiration in the field under Mediterranean climate: a review. European Journal of Agronomy, 13, 125-153.

Salkind NJ (Ed.). 2007. Encyclopedia of Measurement and Statistics. Thousand Oaks, CA: Sage Publications, Inc., pp.648-651

Saltelli A, Bolado R. 1998. An alternative way to compute Fourier amplitude sensitivity test (FAST). Computational Statistics & Data Analysis, 26, 445-460.

Saltelli A, Tarantola S, Campolongo F, Ratto M. 2004. Sensitivity analysis in practice. Italy: John Wiley & Sons, Ltd.

Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications, 181, 259–270.

Seiller G, Anctil F, Perrin C. 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrology and Earth System Science, 16, 1171-1189.

Seiller G, Anctil F. 2013. Climate change impacts on the hydrologic regime of a Canadian river: comparing uncertainties arising from climate natural variability and lumped hydrological model structures. Hydrology and Earth System Science Discussion, 10, 14189-14227.

Smakhtin VU. 2001. Low flow hydrology: a review. Journal of Hydrology, 240, 147-186.

Sobol' IM. 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 55, 271-280.

Tang Y, Reed P, Wagener T, van Werkhoven K. 2007. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrology and Earth System Science, 11, 793–817.

Thornthwaite CW. 1948. An Approach toward a Rational Classification of Climate. Geographical Review, 38(1), 55-94.

van Griensven A, Meixner T, Grunwald S, Bishop T, Diluzio M, Srinivasan R. 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of Hydrology, 324, 10-23.

Wagener T, McIntyre N, Lees MJ, Wheater HS, Gupta HV. 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrological Processes, 17, 455-476.

Wang X, Melesse AM, Yang W. 2006. Influences of Potential Evapotranspiration Estimation Methods on SWAT's Hydrologic Simulation in a Northwestern Minnesota Watershed. American Society of Agricultural and Biological Engineers, 49(6), 1755-1771.

Wang YC, Yu PS, Yang TC. 2010. Comparison of genetic algorithms and shuffled complex evolution approach for calibrating distributed rainfall–runoff model. Hydrological Processes, 24, 1015-1026.

Zhang C, Chua J, Fu G. 2013. Sobol's sensitivity analysis for a distributed hydrological model of Yichun River Basin, China. Journal of Hydrology, 480, 58-68.

Appendix A

E04 : Penman (1948)	$PET = \frac{\Delta R_n + \gamma(e_s - e_a).2.62(1+0.537U)}{\lambda(\Delta+\gamma)}$
E05 : FAO56 P-M (ASCE, 2005)	$PET = \frac{0.408.\Delta R_n + \gamma\left(\frac{900}{T_a+273}\right).U(e_s - e_a)}{\Delta+\gamma(1+0.34U)}$
E06 : Kimberly-Penman (1982)	$PET = \frac{\frac{\Delta}{\Delta+\gamma}R_n + 2.62\left(\frac{\gamma}{\Delta+\gamma}\right).(e_s - e_a).K}{\lambda}$ $with\ K = \left[0.4 + 0.14\ \exp\left(-\left(\frac{J_D - 173}{58}\right)^2\right)\right] +$ $\left[0.605 + 0.345\ \exp\left(-\left(\frac{J_D - 243}{80}\right)^2\right)\right]U$
E07 : Thom and Oliver (1977)	$PET = \frac{\Delta R_n + 2.5\gamma(e_s - e_a).2.62(1+0.536U)}{\lambda\rho\left(\Delta+\gamma\left(1+\frac{r_s}{r_a}\right)\right)}$
E08 : Thornthwaite (1948)	$PET = 16\left(\frac{DL}{360}\right)\left(\frac{10T_a}{I}\right)^K$ $with\ I = \sum_{n=1}^{12}\left(\frac{T_k}{5}\right)^{1.514}$ *where* T_k *is the mean monthly temperature and* $K = 0.49239 + 1.792.I.10^{-2} - 0.771.I^2.10^{-4}$
E09 : Blaney-Criddle (1950)	$PET = k\left(100.\frac{DL}{365\times12}\right).(0.46.T_a + 8.13)$ *with k varying from 0.45 to 1.2 according to season and vegetation*
E10 : Hamon (1961)	$PET = \left(\frac{DL}{12}\right)^2 \exp\left(\frac{T_a}{16}\right)$
E11 : Romanenko (1961)	$PET = 4.5\left(1+\frac{T_a}{25}\right)^2\left(1-\frac{e_a}{e_s}\right)$

E12 : Linacre (1977)	$$PET = \frac{\frac{500(T_a + 0.006z)}{100 - Lat} + 15(T_a - T_d)}{80 - T_a}$$ *with Lat the latitude of the station and z the altitude (m)*
E13 : Hydro-Québec - HSAMI	$$PET = 0.02978(T_{max} - T_{min}) \left(9\left(\frac{9}{5}(T_{max} + T_{min}) + 64\right)\right)$$
E14 : Kharrufa (1985)	$$PET = 0.34\left(100.\frac{DL}{365 \times 12}\right).T_a^{1.3}$$
E15 : Wendling – WASim (1975)	$$PET = 100R_s(1.1 - \alpha) + 93.f_k.\left(\frac{T_a + 22}{150(T + 123)}\right), with\ f_k = 0.8$$
E16 : Turc (1955)	$$PET = 1000.\frac{1}{\lambda\rho}.\left(\frac{T_a}{T_a + 15}\right).\frac{(R_s(1 - \alpha) + 24)}{1.3}.c$$ $$with\ c = 1 + \frac{50 - RH}{70}\ \text{for}\ RH < 50\%$$ $$c = 1\ for\ RH \geq 50\%$$
E17 : Jensen and Haise (1963)	$$PET = 1000.\frac{R_e}{\lambda\rho}.\frac{T_a}{40}$$
E18 : McGuinness and Bordne (1972)	$$PET = 1000.\frac{R_e}{\lambda\rho}.\frac{T_a + 5}{68}$$
E19 : Doorenbos and Pruitt (1977)	$$PET = K.\frac{1}{\lambda}.\frac{\Delta R_s}{\Delta + \gamma} - 0.3$$ $$with\ K = 1.066 - 0.13\frac{RH}{100} + 0.045U - 0.02\frac{RH}{100}U - 0.00315\left(\frac{RH}{100}\right)^2 - 0.0011U^2$$
E20 : Abtew (1996)	$$PET = 0.53\frac{R_s}{\lambda\rho}(1 - \alpha).1000$$
E21 : Makkink (1957)	$$PET = \left(0.61.\frac{1}{\lambda}.\frac{\Delta}{\Delta + \gamma}.\frac{R_s}{58.5} - 0.012\right) \times 100$$
E22 : Oudin (2005)	$$PET = 1000.\frac{R_e}{\lambda\rho}.\frac{T_a + 5}{100}$$
E23 : Baier and Robertson (1965)	$$PET = 0.157T_{max} + 0.158(T_{max} - T_{min}) + 0.109R_e - 5.39$$

PET	Potential Evapotranspiration (mm.day^{-1})
Δ	Slope of vapor pressure curve (kPa.°C^{-1})
λ	Latent heat of vaporization (MJ.kg^{-1})
ρ	Water density (=1000 kg.L^{-1})
d_a	Air density (kg.m^{-3})
γ	Psychrometric constant (kPa.°C^{-1})
e_s	Saturation vapour pressure (kPa)
e_a	Actual vapour pressure (kPa)
r_a	Aerodynamic resistance (s.m^{-1})
r_s	Surface resistance of reference crop (= 70 s.m^{-1})
U	Wind speed 2 m above soil surface (m.s^{-1})
T_a	Air temperature (°C)
T_d	Dew point temperature (°C)
T_{max}	Maximum air temperature (°C)
T_{min}	Minimum air temperature (°C)
R_e	Extraterrestrial radiation (MJ.m^{-2}.day^{-1})
R_s	Global short-wave radiation (MJ.m^{-2}.day^{-1})
R_n	Net incoming solar radiation (MJ.m^{-2}.day^{-1})
RH	Relative humidity (%)
DL	Day length (h.day^{-1})
α	Surface albedo (-)
J_D	Julian day (ordinal date)
C_p	Air specific heat capacity (= 1.013.10^{-3} MJ.kg^{-1}.°C^{-1})