Titre: Title: Developing an Automatic Model for Reconstructing Daily Flow in Ungauged Basins

Auteur: Author: Ana Hosseinpour

Date: 2014

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Citation: Hosseinpour, A. (2014). Developing an Automatic Model for Reconstructing Daily Flow in Ungauged Basins [Thèse de doctorat, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/1598/

URL de PolyPublie: PolyPublie URL: https://publications.polymtl.ca/1598/

Directeurs de recherche: Advisors: Musandji Fuamba, & Leslie Dolcine

Programme: Program: Génie civil

UNIVERSITÉ DE MONTRÉAL

DEVELOPING AN AUTOMATIC MODEL FOR RECONSTRUCTING DAILY FLOW IN UNGAUGED BASINS

ANA HOSSEINPOUR

DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET DES MINES

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION

DU DIPLÔME DE PHILOSOPHIAE DOCTOR

(GÉNIE CIVIL)

OCTOBRE 2014

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

DEVELOPING AN AUTOMATIC MODEL FOR RECONSTRUCTING DAILY FLOW IN UNGAUGED BASINS

présentée par : HOSSEINPOUR Ana

en vue de l'obtention du diplôme de: Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de :

M. LECLERC Guy, Ph. D., président

M. FUAMBA Musandji, Ph. D., membre et directeur de recherche

M. DOLCINE Leslie, Ph. D., membre et codirecteur de recherche

M. TREMBLAY Michel, Ph. D., membre

M. NGUYEN Van-Thanh-Van, Ph. D., membre

DEDICATION

I dedicate my dissertation work to my family. A special feeling of gratitude to my loving parents, my sisters and brother who always supported me.

ACKNOWLEDGEMENTS

I would like to give special thanks to Reza Khaksarfard for all his help and support.

I wish to thank Dr. Musandji Fuamba for all he taught me during these years and his countless hours of reflection, reading, and encouragement throughout this entire process. He was more than a teacher and a supervisor to me, he was also very understanding towards me regarding personal issues.

I would like to thank Dr. Leslie Dolcine for being a good guide and teacher throughout the times of uncertainty. I truly learnt a lot during the period of working with him.

Thank you Dr. Guy Leclerc, Dr. Van-Thanh-Van Nguyen, and Dr. Michel Tremblay for agreeing to serve on my committee.

Thanks also the Natural Sciences and Engineering Research Council of Canada (NSERC) by the Collaborative Research and Development Grants along with Hydro-Québec for financing the present project and providing the data presented in the study.

RÉSUMÉ

Les séries des écoulements ne sont pas mesurées directement. Leur estimation peut parfois s'accompagner d'erreurs considérables. Comme ces valeurs sont importantes dans la planification de la production hydroélectrique, il s'avère donc important de reconstruire ces séries d'écoulement avec suffisamment de précision. Différentes méthodes de reconstruction des écoulements ont été développées au cours de dernières années, et plusieurs facteurs importants doivent être analysés lors du choix de la méthode la plus appropriée. Dans cette thèse, un algorithme est proposé pour déterminer la méthode la plus appropriée dans la détermination de la série fiable des valeurs d'écoulement pour chaque étude de cas analysée. Cet algorithme permet de choisir les méthodes de calcul des séries d'écoulement à la fois pour la période de temps avant la construction du réservoir que pour la période post-réservoir, selon la disponibilité de données dans les bassins environnants.

Pour la période pré-réservoir, une nouvelle méthode basée sur le filtre de Kalman a été développée pour reconstruire la série des valeurs d'écoulement en utilisant la technique « State Fusion », lorsque les seules données disponibles pour les bassins non jaugés proviennent de bassins voisins. Les résultats de cette méthode sont par la suite comparés aux méthodes « Area Ratio », « Move type III » et régression multivariée utilisant différents indices de qualité.

Pour la période post-réservoir, une nouvelle méthode basée sur l'équation d'équilibre hydrologique est proposée pour reconstruire et filtrer les valeurs d'écoulement en utilisant une technique d'optimisation, lorsque les données hydrométriques (débit turbiné, niveau d'eau dans le réservoir et débit évacué) sont collectées dans un bassin non jaugé. Les résultats de cette méthode sont par la suite comparés aux valeurs obtenues avec la méthode classique d'équation d'équilibre hydrologique utilisant différents indices de qualité.

La stationnarité des séries écoulements reconstruites est également évaluée et l'analyse régionale réalisée pour assurer la cohérence entre le flux local et le flux régional. Enfin, les valeurs finales des séries écoulements reconstruites sont déterminées en combinant les valeurs de différentes méthodes combinées à l'aide d'une technique de pondération. Un calcul d'incertitude a été réalisé et il a permis d'évaluer la précision des séries pour la période post-réservoir.

ABSTRACT

Since flow values for basins are indirectly measured and the estimations of these values may be at times accompanied by a considerable amount of uncertainty, it is desirable to reconstruct a reliable set of flow series as these values are important for water resource management and flow prediction. Different methods of flow reconstruction have been developed during recent years. As the quality of available flow data are not the same for different time periods, different flow reconstruction methods should be selected for each different time period. In this thesis, an algorithm will be proposed in order to determine the most appropriate family of flow reconstruction method for each case study scenario. This algorithm will help to choose the best method to reconstruct flow values for both for Pre-Reservoir Construction Period (Pre-R) and Post-Reservoir Construction period (Post-R), depending on the availability of data and other factors.

A new Kalman-based method will also be developed to reconstruct the flow data series using the "State Fusion" technique for the Pre-R period, when the only available data for an ungauged basin (with no flow measurements) comes from meteorological data, the neighbouring basins' flow, and the simulated flow using a rainfall-runoff model. The results of this method will be compared to existing Area Ratio method, Maintenance of variance (Move) type III method, and Multivariable regression method using different Quality Indexes (QIs) that are designed for use on ungauged basins.

For the period when the basin has been equipped with a reservoir, the use of a new Water Balance equation (WBE) based method will be considered to reconstruct and filter the daily flow data by using an optimization technique in situations when hydrometric data (i.e. turbine flow, water level in the reservoir, and discharged flow) have been collected for an ungauged basin. The developed optimization model will be able to minimize WBE errors and flow variation. This model will be automatized using a Deterministic and a Stochastic technique to intelligently select the parameters and not require human judgment. The results of this method are then compared to the classic WBE using different QIs that are designed for use on ungauged basins.

The regional and temporal homogeneity of the reconstructed flow values are also assessed to ensure that coherence between the local flow and the regional flow, and the stationarity of flow characteristics of the basin are maintained during the sampling time period.

Finally, a Weighted Average technique will be used to calculate the final reconstructed flow series by combining the reconstructed flow values obtained from different methods. Also, the uncertainty of the final flow data series (Post-R) will be evaluated with the help of a suggested sensitivity analysis method.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABREVIATIONS

a	parameter of multivariable regression
ANFIS	Adaptive Neuro-Fuzzy Interaction System
ANN	Artificial Neural Network
AO	Arctic Oscillation
AVE	Absolute volume error
a'_i	Differences discharges of the days $i+1$ and i of the selected neighbouring basin
b	Parameter of multivariable regression
$Benefit_j$	Fitness related to j^{th} parameter set
BMA	Bayesian Model averaging
b'_i	Differences discharges of the days $i+1$ and i of the selected calculated flows
B_1, B_2	Errors of developed Kalman method
c	Optimisation model's parameter to allocate a weight to the set of variation variables (Z and Δ)
CC	Consistency Coefficient
c_j	c coefficient of j^{th} parameter set in the stochastic based model
c_n	Mother c parameter of iteration number n+1 in the stochastic based model
d	Vector of weights in deterministic based model

D	Deterministic method
$D_{n,o}$	Difference between the original and the new flow (nf_d)
dn	Parameter of POM which is related to number of days for which the optimization is being solved
dn_j	Parameter of *dn* for j^{th} parameter set in the stochastic based model
dn_n	Mother *dn* parameter of iteration number n+1 in the stochastic based model
e	Constant value of multiple regression method
E	Evaporation from the reservoir during the time Δt
$E_{\inf}$	Inferior squared WBE's error
$E_{inputdata}$	Uncertainty of flow data caused by random uncertainty of input data
$E_{\sup}$	Superior squared WBE's error
f	Probability of flow
FQ	Fitted line to values of LN3 of a 2-year return period
GA	Genetic algorithm
GLUE	Generalized likelihood uncertainty estimator
int	Interaction between stored water in the reservoir and groundwater
IR	Improvement ratio
KPSS	Kwiatkowski-Phillip-Schmidt-Shin test
m	Rank of the value

MCMC	Markov chain Monte Carlo
Move	Maintenance of variance
n	Number of days that the trend of both calculated flow and neighbouring basin flow is increasing or decreasing
N	Total number of days
NASH	Nash–Sutcliffe efficiency coefficient
NAVE	Normalized absolute volume error
n_l	Length of the short record
$n_l + n_2$	Length of the long measured data
nbf_i	Flow of i^{th} day in the neighbouring basin
nbf_s	Scaled flow from the neighbouring basin
F	Flow caused by effective rainfall
$f_{d,i}$	Flow calculated using disturbed input data for i^{th} day
$f_{D,k,i}$	Flow calculated using deterministic method for k^{th} day of segment i
F_n	Flow to reservoir number n
$f_{o,i}$	Original flow for i^{th} day
f_{RR}	Simulated flow using RR model
$f_{S,k,i}$	Flow calculated using stochastic method for k^{th} day of segment i
$F_{WA,k,i}$	Weighted average flow for the k^{th} day of i^{th} segment of a year

NN	Normalized Nash
NT	Normalized Tortuosity
Offset	Mean value of extreme volume values ($v_{\sup(n)}$ and $v_{\inf(n)}$)
p	Parameter of POM to allocate a weight to the variable of *Z*
P	Precipitation over the reservoir surface
P (ps j)	Probability of j^{th} parameter set
POM	Perreault Optimization Model
Pre-R	Pre-Reservoir Construction Period
PS	Parameter Set
Post-R	Post-Reservoir Construction Period
q	Parameter of POM to allocate a weight to the variable of Δ
q_{cali}	Reconstructed flow for day *i*
$\overline{q_f}$	Average filtered flow when $i=1,\ldots,N$
q_{fi}	Reference filtered flow for day *i*
QI	Quality Index
$QI_{ave\ S,i}$	Average QI of stochastic method for segment *i*
$QI_{ave\ D,i}$	Average QI of deterministic method for segment *i*

$q_{in,n}$	Input discharge to reservoir number *n* (regulated outflow from the upstream reservoir)
$\overline{q_{obs}}$	Average reconstructed flow
q_{obsi}	Observed flow in case-study basin in day *i*
q_{out}	Output flow from the reservoir
$q_{out,n}$	Outflow from the reservoir number *n* (the summation of turbine flow and discharged/spilled flow from reservoir number *n*)
q_{ri}	Reconstructed flow for day *i*
q_{simn}	Simulated flow by rainfall-runoff model for day *n*
$q_{sp,n-1}$	Discharged/spilled flow from reservoir number *n-1*
$q_{tr,n\ 1}$	Turbine flow from the reservoir number *n-1*
q_{WBEi}	Calculated flow using classic WBE for the day *i*
$range_{NF}$	Possible range of flow data series (total uncertainty
R_l	Predicted lower range
R_n	Reservoir number *n*
RR	Rainfall-Runoff
R_u	Predicted upper range
RV_{n}	Volume of water in the reservoir at the beginning of n^{th} day

S	Stochastic method
s	*S*urface of the basin
$s_{case\ study}$	Surface area of case study basin
s_{nb}	Surface area of neighbour basin
sf_{cal}	Specific flow calculated using reconstructed flow
SFR	*S*pecific flow ratio
sf_{WBE}	Specific flow calculated using WBE
T	Tortuosity
TM	Thornth waite-Mather
$v_{\inf\,(\min)}$	Minimum of $v_{\inf(n)}$ (in m^3/s) set when n= *starthour,…,endhour*
$v_{\inf(n)}$	Minimum of reservoir's volume measured every 5 minute during the n^{th} hour
$v_{(n)}$	Measured volume of water in the reservoir at the beginning of day n (in hm^3)
$v_{\sup\,(\max)}$	Maximum of $v_{\sup(n)}$ (in m^3/s) set when n= *starthour,…,endhour*
$v_{\sup(n)}$	Maximum of reservoir's volume measured every 5 minute during the n^{th} hour (in m^3/s)
w_1, w_1, w_1	Weight of QIs for defining the fitness in stochastic based method
$w_{D,i}$	Weight of deterministic technique for segment i
$w_{S,i}$	Weight of stochastic technique for segment i
WAM	Weighted average method

WBE	Water balance equation
WSS	Wide Sense Stationary
x_i	Measured data for day *i*
Y	Reconstrcuted flow using the Kalman filter based model
y_1	Filtered simulated flow using the Kalman filter method
y_2	Filtered neighbouring basin's flow using the Kalman filter method
$\widehat{Y}_i$	Reconstructed flow using Move III for the day *i*
Z	Flow variation during 2 consecutive days
α_1, α_2	Coefficients of developed Kalman method
γ	Parameter of POM to allocate a weight to the set of variation variables (Z and Δ)
γ_j	γ Coefficient of j^{th} parameter set in the stochastic based model
γ_n	Mother γ parameter of iteration number n+1 in the stochastic based model
Δ	Flow variation during 3 consecutive days
$\Delta_{input\ data}$	Uncertainty of flow caused by instrument uncertainty
$\Delta_{q_{in}}$	Instrument uncertainty of q_{in}
$\Delta_{q_{out}}$	Instrument uncertainty of q_{out}
Δ_{volume}	Instrument uncertainty of *volume*
Δx_i	Absolute uncertainty of measured x_i

Δs_n	Storage volume changes in the reservoir number *n*
$\widehat{\mu_y}$	Unbiased estimator of the mean of the complete extended record in Move III
$\hat{\sigma}_y$	Unbiased estimator of the variance of the complete extended record in Move III

LIST OF APPENDICES

GLOSSARY

Flow:

Flow is defined as the runoff caused by effective rainfall. Flow can be calculated using the classic Water Balance Equation (WBE) for a reservoir (as a closed system):

$$F_n = q_{out,n} - q_{in,n} + (\Delta S_n / \Delta t) \tag{i-1}$$

where:

$q_{out,n}$.....= the summation of outflow from the reservoir number n (Rn) during the sampling time period,

$q_{in,n}$ = the summation of inflow to reservoir number *n* during the sampling time period,

ΔS_n = the net change of storage volume in the reservoir number *n* during the sampling time period,

Δt = the sampling time period

F_n = the unknown flow value caused by effective rainfall (which includes all the minor terms such as evaporation, direct rainfall, and interaction between surface water and ground water) to reservoir number *n*.

As illustrated in Figure i-1, $q_{in,n}$ is equal to the regulated outflow from the upstream reservoir (if available), $q_{out,n-1}$, which is the summation of turbine flow from the reservoir number *n-1* ($q_{tr,n\ 1}$) and discharged/spilled flow from reservoir number *n-1* ($q_{sp,n-1}$).

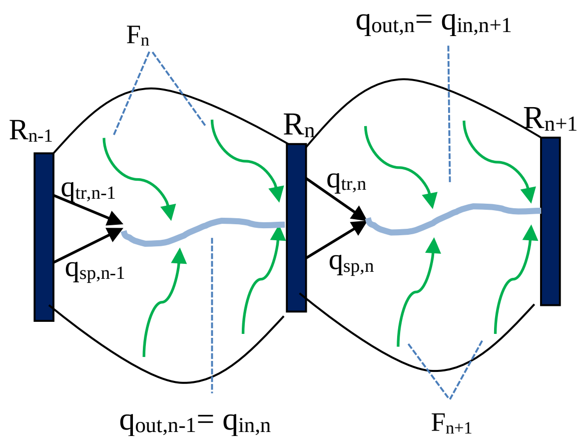


Figure i-1: Schematic of three reservoirs in series

Ungauged basin:

An ungauged basin (with or without a reservoir) is defined as a basin that does not have available or recorded river flow measurements or data.

Uncertainty:

“The component of a reported value that characterizes the range of values within which the true value is asserted to lie” (NDT resource center, 2014).

(Note: In this project, the uncertainty in a flow data series could be caused from the uncertainty of the methods’ structure (WBE, optimization model, and etc.) and/or input data uncertainty)

Noise:

Noise is defined as obvious uncertainty and expressed as the spurious variation exhibited in a data series. Usually a data series’ uncertainty cannot be detected by looking at data series’ graph; however, noise is visually distinguishable when data values are plotted on a graph.

In the work presented in this thesis, the flow data series are considered noisy, from a variation standpoint, when they clearly differ from the neighbouring basin's flow and rainfall-runoff model's flow.

Automatic model:

An automatic model is defined as a method which is independent from a human's decision and uncertainty. Automatic models could act like a type of software which takes input data and produces one or more output data. Thus, the output values will not require manual modification. Like software, the parameters of the automatic model would change depending on case study scenarios.

CHAPITRE 1 INTRODUCTION

1.1 Context of Thesis

Locally and regionally reliable flow data series (pre- and post-reservoir construction period) are essential for the purpose of analyzing flow frequency, simulating hydraulic systems, predicting the flow, designing hydraulic structures, and undertaking other activities related to water planning and management. Poor flow data records, however, may exacerbate the uncertainty of water management. For example, a lack of reliable flow data values affects water allocation studies and may result in deficiencies in hydropower production, irrigating plans, etc. As well, unreliable knowledge about flow results in poor flow prediction, leading to a lack of proper preparation for possible floods or droughts, causing irreparable financial or human loss. These examples show how unreliable flow data series affect socio-economic aspects of people's life and government's services, directly and indirectly.

The discussion of obtaining reliable flow data holds importance for Quebec (and for any other region with the same characteristic) as this province possesses 2 percent of all of the planet's freshwater, with potential use for many purposes such as agriculture, tourism, hydropower, and industry. As an example, in 2006, the province of Quebec produced 205.661 TWh of hydroelectricity, while "the clear exports of electricity have been established at 6,3 TWh for the same year" (ROBVQ, 2013). In 2004, Canada was the fourth largest producer of hydroelectricity in the world, with Quebec producing almost 50% of the total hydroelectricity in Canada (ROBVQ, 2013). Systematic management and usage of Quebec's water resources are feasible only if reliable data (including flow) is available. Flow is not measured in many of the basins of Quebec (neither the period when basins did not possess a reservoir nor the time period when they are equipped with reservoirs) because of the vastitude of province and inaccessibility of many of the catchments. Therefore flow needs to be estimated with reasonable accuracy.

A project was initiated by Hydro-Quebec entitled "reconstructing the flow data for ungauged basins of Quebec" in order to estimate the flow values of ungauged basins in Quebec. The project addressed the flow estimating methods for the Quebec's basins where the flow is not measured. It was necessary for Hydro-Quebec to conduct this project because of the claim that more reliable flow data increases the justification of water resource management and in turn will help the production of hydroelectricity by a considerable amount.

The main topic of this PhD thesis is to study the methods of reconstructing and obtaining reliable flow data when they are not readily available. This research was conducted within the framework of the Hydro-Quebec's "Reconstructing the flow data for ungauged basins of Quebec" project. The nature of the Hydro-Quebec's project has always encountered shortfalls in the existing studies and the intention of this work is to address and resolve these inadequacies. Some of the shortfalls encountered are listed below:

A) Lack of an algorithm to select the most appropriate family of the flow reconstruction method:

There exist many different methods of flow reconstruction. Selecting the appropriate method in each time period depends on different factors. For instance, the following six factors should be considered when selecting a method for basins within the province of Quebec:

1. Flexibility of method

The data reconstruction method needs to be flexible enough to be applicable for all the basins in Quebec, as most of them are ungauged.

2. Scale of reconstructed flow

The main reasons for reconstructing flow data values are to improve the processes of water resource management, flow prediction, and risk management. These goals are achievable with long-term (historical and real-time[1]) daily flow data; therefore, flow reconstruction is required to be done on a daily and long-term basis.

[1] Real-time data are the data related to the present time step. For example, the real-time daily flow is the flow of the current day.

3. Quality of reconstructed flow

As flow data are used for short term flow prediction and management, high-quality flow data is required because it may be difficult to efficiently analyze flow series affected by noise disturbing the short/long term memory of flow data (as flow analysis methods are based on the memory and behaviour of data series in time). Flow noise results in uncertainty of sometimes millions of cubic meter of estimated flow volume each day [2].

Also, high quality reconstructed flow data result in more efficient water resource planning and management in long term.

(Note: both snowmelt and evaporation are factors that can affect flow data, especially in large basins and reservoirs. Thus, accounting for these factors may improve the flow reconstruction results)

4. Applicability of method

One important objective of flow reconstruction is to predict future flow values (usually a few days in advance), which is usually possible by knowing the real-time flow data (predicting flow data is out of the scope of this study). Thus, any method employed should be applicable to both historical and real-time data.

5. Data availability

The flow reconstruction method should be selected based on available data. As most watersheds in Quebec are ungauged, the measured data obtained are limited to:

- Pre-Reservoir (Pre-R) Period: the flow data series measured in neighbouring basins to that of the studied basin (few neighbouring basins), and hydrologic data (including daily rainfall, daily minimum temperature, daily maximum temperature, daily temperature, and daily snowfall) of the studied basin.
- Post-Reservoir (Post-R) Period: the flow data series measured in neighbouring basins to that of the studied basin, hydrologic data (including daily rainfall, daily minimum

[2] For example, 15 m^3/s of flow noise is equal to (15*24*3600 = 129600) 129600 m^3 of uncertainty in volume of flow during a day.

temperature, daily maximum temperature, daily temperature, and daily snowfall) of the studied basin, as well as reservoir related data of the studied basin such as water level on the upstream and downstream side of the reservoir, gate openings, and produced electricity. These reservoir data have been used by Hydro-Quebec to calculate turbine flow, discharged flow and reservoir volume.

Thus, the flow reconstruction method may be different for Pre-R and Post-R time periods. Also, it is preferable to include as much data as possible in the flow reconstruction process because they empower the flow estimation values by taking into account different aspects of this topic.

(Note: Hydro-Quebec already used hydrologic data to simulate flow using a Rainfall-Runoff (RR) model. The results of this simulation are available for both Pre-R and Post-R periods.)

6. Quality of input data

7. It is known that the quality of input data affects the results of the flow reconstructing method. Therefore, available hydrologic data and neighbouring basin's flow data must be of high quality and reliability. On the other hand, reservoir related data (including turbine flow, discharged flow, and reservoir volume) prior to 2005[3] (considered as raw data) may contain uncertainties and noises during certain time periods.

The fundamental topic of developing a flow reconstruction and validation method for ungauged basins in Quebec that considers these six factors has always been a constant debate. However, there is no research results available on <u>how to select the appropriate</u>

---

[3] For the years after 2005, the water level on the upstream and downstream side of the reservoir, produced electricity, and system's characteristics are validated, and thus, the calculated discharged flow, turbine flow, and storage volume data for post-2005 are more reliable. However, before 2005, discharged flow, turbine flow, and storage volume data taken from Hydro-Quebec data base are calculated based on non-validated measured data. It means that these data are raw and they may contain noise and/or outlier.

method of flow reconstruction that puts into consideration the different aforementioned factors that affect choice of method for the province of Quebec (**Question # 1**).

B) *Lack of a flexible methodology in reconstructing reliable daily flow data series:*

During the past few decades, a lot of effort has been devoted to reconstructing plausible flow data series that can be used to validate flow data, and to estimate the uncertainty of hydraulic systems in Quebec (i.e. Perreault *et al*. 1996, Bennis *et al*. 1994, Bennis and Kang, 2000, and Haché *et al*. 2003). Some of these studies are listed below:

- Bisson and Roberge (1985) developed a rainfall-runoff (RR) model to simulate flow in the basins of Quebec; however, this model usually underestimates peak flows.
- Charbonneau and Berube (1987) and Berrada *et al*. (1996) proposed separate methods to remove noise from flow data series (filter flow data series) that were calculated using WBE; but ultimately, because these methods required knowledge of future data, they could not be used in real-time situations.
- Perreault *et al*. (1995) suggested a procedure to improve flow data series by combining the results from rainfall-runoff (RR) model, WBE, and neighbouring basin's filtered flow (reconstructed flow using their suggested method which has been applied in a nearby basin). However, this method may overestimate the cumulative amount of flow for a period of certain number of days (Nguyen and Bisson, 1998).
- Nguyen and Bisson (1998) suggested a regression and exponential smoothing technique to solve the problems of the Perreault's model (1995). In their method the validated flow for each day is regarded as an exponential function of previous days' flow. This method is acceptable when data are stationary; in reality, a more appropriate method is required to take into account seasonality effects on flow. Moreover, it would be preferable to calculate a more accurate flow data series by filtering the Water Balance equation (WBE) series using all available flow data of basin (such as flow from RR model and neighbouring basin, etc.).
- Lastly, Hydro-Quebec currently uses a method that adopts a mostly manual procedure for calculating and filtering flow data values. In this method, flow values are calculated

using WBE and then modified by taking into consideration regional and temporal analysis. However, the results obtained using this method may be affected by human-caused uncertainties and misjudgments, requiring a method that eliminates human error.

- Thus, Perreault (2011) suggested a WBE based optimization model (POM) which is independent from human judgment. This model reconstructs and filters hourly flow data series and attempts to minimize the variation of flow data and WBE errors (this model is described in more details in Chapter 2). Despite the fact that POM is successful in reducing noise and removing unrealistic values, it still has some deficiencies and problems. For example:

    (1) this method is dependent on volume data of 5-minute intervals (not applicable when data of 5-minute intervals are not available),

    (2) it is only applicable for an hourly time scale,

    (3) the parameters of this model are constant during the time and space,

    (4) this model does not take into account available data from rainfall-runoff model and neighbouring basins to improve results,

    (5) results are still susceptible to noise, especially for low flows.

All of the studies mentioned above were important steps towards enhancing the knowledge of flow data series in the province of Quebec. Nevertheless, all of the methodologies developed for ungauged basins in Quebec (except for the developed RR model by Bisson and Roberge, 1985) were based on WBE, and thus, only applicable for Post-R period when the input data of WBE (turbine flow, discharged flow, and reservoir volume) are available. However, reconstructing daily flow data values for Pre-R period has remained almost unexplored. Knowing the values of flow for Pre-R period is helpful in that it provides a more comprehensive understanding of historical flow data, which helps to perform more reliable flow analysis. Accordingly, the second question is <u>how to reconstruct daily flow for Pre-R period (**Question # 2**)</u>.

Even for the Post-R period, none of the research mentioned above provide a flexible methodology to reconstruct and validate reliable daily flow data series that is independent of human decisions. This gives rise to an interesting question of how to reconstruct more likely values for daily flow for Post-R period (**Question # 3**).

C) *Lack of criteria to evaluate the quality of the reconstructed flow in ungauged basins:*

An important topic related to the flow reconstruction studies is how to assess the performance of flow reconstruction methods. Usually, researchers would apply one or two quality indexes (QIs) to evaluate reconstructed data series (e.g. Mean Square Error by Gupta *et al.* (2009), temporal or spatial correlation coefficient, and Nash-Sutcliffe Efficiency by Johnston *et al.* (2009)). These methods adopt different techniques when comparing the reconstructed data with a measured data series. However, the topic is more challenging when there is no measured data series that can be used as reference data (same as current study). In the research mentioned above which applies to Quebec, traditional QIs were used in the comparison of reconstructed flow data with existing filtered flow series (the mostly manually filtered flow series are available in Hydro-Quebec's database). As the filtered flow values are less reliable prior to 2005 (as the input data may contain uncertainties before 2005), it will be advantageous to design a few different QIs to evaluate the quality of the reconstructed flow values that are independent from filtered flow values. This challenge raises the questions of how to evaluate the quality of the reconstructed flow data series for ungauged basins (**Question # 4**).

D) *Lack of a methodology to analyze the uncertainty of reconstructed flow in ungauged basins:*

Another indispensable element related to hydrological studies (aside from flow data reconstruction) is uncertainty analysis. There are many different methods available to evaluate different types of uncertainty. For instance, Generalized Likelihood Uncertainty Estimator (GLUE), Markov Chain Monte Carlo (MCMC), and most existing sensitivity analysis methods are used to define the parameter of uncertainty. The Bayesian model is also another method which factors in uncertainty in input data, output parameters, and the model's structure (Yang *et al.* 2007). Most of the existing methods are dependent on the measured data. Therefore, it is still a challenge to find out how to evaluate the uncertainty of flow data in ungauged basins (**Question # 5**).

As the previous researches could not find answers to the above mentioned questions (Questions # 1-5), the goal of this study was to find solutions to these questions. In fact, the mentioned questions formed the base of the objectives of this thesis.

1.2 Motivation

The driving force in finding solutions to the questions mentioned above (as explained in section 1.1) were because of the importance of this topic, coupled by a lack of the following tools for our research:

- ✓ an algorithm to select the most appropriate family of flow reconstruction method considering different factors,
- ✓ a flexible methodology to reconstruct daily flow data series for Pre-R period,
- ✓ an automatic flexible method (independent from a human's decision and uncertainty) for Post-R period,
- ✓ the indexes to evaluate the quality of reconstructed flow in ungauged basins, and
- ✓ a methodology to analyze the uncertainty of reconstructed flow in ungauged basins.

The desire to find solutions to the above questions also enhanced the desire of taking the next step towards reaching the final goal; the ability to estimate more reliable flow data series for ungauged basins of Quebec.

1.3 Objectives

The general objective of this research is to <u>develop a method to obtain daily flow values for ungauged basins in Quebec for both Pre-R and Post-R periods</u>. To attain this primary objective, five secondary objectives had to be completed in the course of this research. These include (Figure 1-1):

a) Introducing an algorithm for selecting the most appropriate family of flow reconstruction methods in each case study (Pre-R and Post-R periods).

The motivation for this came from the fact that currently there is no completely recognized methodology available to help researchers in selecting an appropriate method of flow reconstruction. This objective aims to find the answer to Question # 1 mentioned in Section 1.1. The developed algorithm based on this objective should consider all the 6 factors mentioned in Section 1.1 that affect the selection of flow reconstruction method in the area.

According to the suggested algorithm, WBE based methods and regression based methods are the most appropriate family of flow reconstruction methods in the current case-study for Post-R and Pre-R periods, respectively.

b) Evaluating the performance of existing methods of flow reconstruction, and defining the weaknesses in them.

   Before developing a methodology for flow reconstruction, it is necessary to assess the capabilities and weaknesses of the existing flow reconstruction methods that are currently being used.

c) Developing an optimization method based on Kalman filter[4] as a tool to combine the available data and produce flow values for Pre-R, and an automatic WBE based model to reconstruct daily flow for Post-R periods.

   This objective was formed to develop a flexible regression based methodology for the Pre-R period, and an adjustable automatic WBE based model for Post-R period. This objective addresses Questions # 2 and # 3 mentioned in Section 1.1.

d) Designing criteria applicable to ungauged basins in order to evaluate the performance of developed methods.

   After reconstructing the flow data series, an assessment of the data quality is required. This objective was designed to address Question # 4, mentioned in Section 1.1.

e) Evaluating the uncertainty of the reconstructed flow in ungauged basins.

---

[4] It is called Kalman filter based method in this thesis.

It is always helpful to define the confidence level of estimated data series. This objective was designed to analyze the uncertainty of flow data series and respond to Question # 5, mentioned in Section 1.1.

1.4 Content of Thesis

The content of this thesis is summarized in Figure 1-1:

- Chapter 1 provides an introduction to the subject, problem, motivation, and the objectives.
- Chapter 2 gives a literature review of the different methods and models of flow reconstruction for Pre-R and Post-R periods, existing methods of quality evaluation, and techniques of uncertainty analysis.
- Chapter 3 explains the case study presented in this thesis. This chapter discusses the available data and information for the area, as well as the quality of this data. In this research, the methodology and the result are divided into a few sub-methodologies and sub-results related to Pre-R and Post-R periods because the results of each sub-methodology define the approach of next step. The sub-methodologies and sub-results are presented in Chapters 4, 5, 6 and 7. Therefore, the case study is presented before them in Chapter 3.
- Chapter 4 introduces an algorithm that helps to select the most appropriate family of flow reconstruction method to different scenarios, with consideration to the applicability, advantages and disadvantages of reviewed methods in Chapter 2. This algorithm factors in all determinative parameters that can affect the selection of appropriate methods of flow reconstruction. It is then applied and tested in our case study. This algorithm and the results of its application to the case study are presented in this chapter to fulfill **Objective a**, and provide answer to **Question # 1**, stated in Section 1.1. The content of this chapter has been submitted to Canadian Water Resource Association (CWRA) in 2013 in the form of a journal paper. The submitted version of this papers is presented in appendix 1.
- Chapter 5 suggests a methodology for reconstructing daily flow for the Pre-R period (satisfying **Objective c** and answering **Question # 2**). This method is then applied in our case study and

the results are then compared to those of a few existing methods. The comparison is done using visual graphs with a few suggested QIs.

- In Chapter 6, several existing methods of flow reconstruction (classic WBE and POM) for Post-R period will be first assessed (**Objective b**). Then, an optimization model will be developed based on POM. Lastly, a sensitivity analysis will be done on the optimization method to evaluate the authenticity of its assumptions and the necessity of improving them.

  Then, the developed methodology for automatizing the suggested optimization model for Post-R period will be explained (achieving **Objective c** and answering **Question # 3**). A few QIs that are applicable to ungauged basins are then introduced to evaluate the reliability of reconstructed flow values and compare the results with classic WBE. A methodology will also be presented to evaluate the regional and temporal homogeneity of the flow data series (**Objective d**). A methodology for analysing the uncertainty and defining the range of reconstructed flow is presented in this chapter to fulfil **Objective e**.

  Chapter 7 will present the results of applying all the presented methodologies in Chapter 6 to a case study.

  Chapters 5, 6, and 7 will provide answers to **Questions # 4 and 5** mentioned in Section 1.2. The content of these chapters has been published in the form of two journal papers at Journal of Hydrologic Engineering (ASCE). The accepted version of these papers are presented in appendices 2 and 3.

Finally, conclusion and recommendations are presented at the end of the thesis. The limitations of developed methodology in this thesis are also listed in this section.

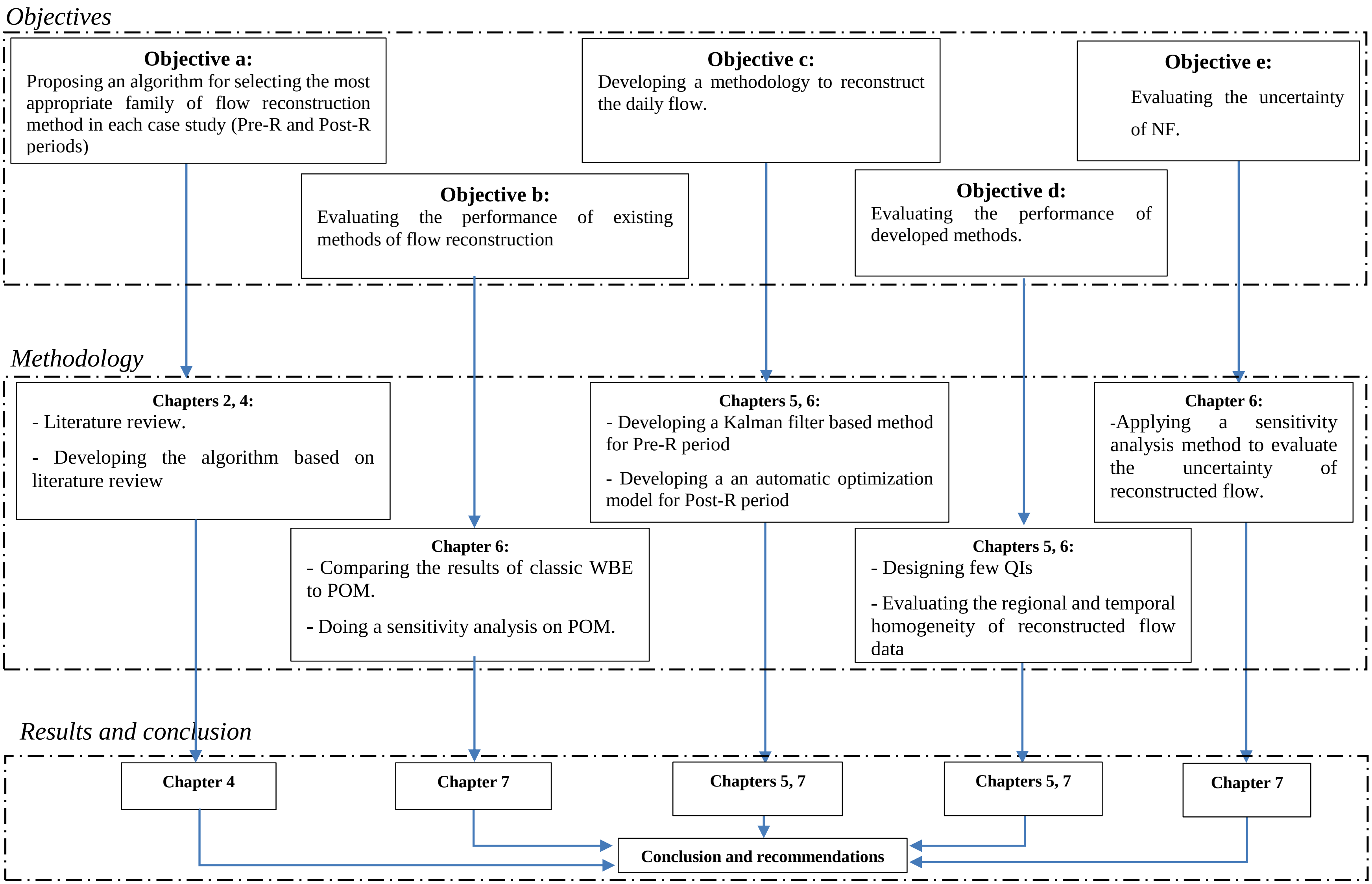


Figure 1-1: Schematic of thesis

CHAPITRE 2 LITERATURE REVIEW

2.1 Introduction

In this chapter, a general literature review is presented on different flow reconstruction methods available for Pre-R and Post-R period. This is followed by a summarized explanation of the existing methods for data series evaluation including QIs, recent methods used to assess stationarity and regional homogeneity of flow data series. A literature review is also presented on the available methods used to analyse the uncertainty found in flow data series.

2.2 Flow Reconstruction

In the past several years, many different methods and models have been developed to simulate or extend flow data in gauged or ungauged basins. In this chapter, these methods are grouped in three main categories based on their approach to simulate flow:

- hydrologic methods,
- hydraulic methods,
- and regression-based methods.

The applicable flow reconstruction methods may be different for Pre-R and Post-R period because the available data, information, and the basins' condition, are usually different for these two time periods. The categories of flow reconstruction methods (hydrologic, hydraulic, and regression-based methods) which can be applied for each Pre-R and Post-R period are explained in the following section.

2.2.1 Pre-R Period

In the case of basins without a reservoir (Pre-R period), hydrologic data, climate data, and/or catchment characteristics can be measured. Thus, hydrologic or regression based methods, which depends on these types of data, may be used for flow reconstruction.

2.2.1.1 Regression based methods

Regression-based methods are those developed based on a mathematical relationship between the flow of the study basin (as a dependent variable) and independent variables from the same or the neighbouring basins. These methods are simple and fast to use (Rezaeianzadeh *et al.* 2013).

An example of a regression-based methods is one that estimates the flow data of the interested sub-basin using available flow data of the main basin. For example, there exists one method that calculates the stream flow of an ungauged sub-basin by relating the ratio of the slope and area of that sub-basin to those of the main basin (Schreiber and Demuth, 2002).

When flow data of neighbouring basins are available, a logarithmic relation (logarithmic scaled data helps to obtain residuals that are approximately symmetrically distributed around zero) can be developed between the flow data characteristics of the neighbouring basins and applied to the basin of interest (using regression method in space). For example, Jones *et al.* (2004) applied a regression-based method to relate the logarithmic values of measured river-flow to linear combinations of soil moisture and effective precipitation. They then applied this regression equation to the ungauged basin of interest to calculate the flow data series. Also, Wen (2009) tried to reconstruct flow by relating the discharge time series to rainfall and maximum temperature. Hughes and Smakhtin (1996) explained that a potential method of extending the flow of a basin of interest would be to simply weight the observed streamflow values of one or more neighbouring gauged basins by the ratio of the catchment area of the basin of interest to the area of the gauged neighbouring basins. The problem with this method is that the flow values of adjacent basins are rarely linearly related to the catchment area of the basin of interest as they may have different hydrology and morphology. Also, it is possible to have a trend of non-stationarity in the actual stream flow data series at the sites or stations used for interpolation (Hughes and Smakhtin, 1996). Thus, it is not recommended that flow values of neighbouring basins be directly transferred to the flow characteristic of a basin of interest.

A regression method can also be developed based on the available short-term data of a basin and used to extend the flow series over a whole time interval (regression method in time) of that basin. Simple regression between a basin’s short-term flow data series and the long-term flow data series of a nearby basin (Hernandez-Henriquez *et al*., 2010, Dastorani *et al.* 2010) exemplifies this type of regression based method. In a case study by Taylor *et al.* (2006), they developed a statistically-linear model based on regression of rainfall and short-term runoff data. Since the complete rainfall data were available for their case study, a regression based method was applied to extend flow data over the whole period.

Maintenance of variance (Move) is another regression-based method for data reconstruction that preserves both mean and variance, and therefore works better than the linear-regression method (Koutsoyiannis and Efstratiadis, 2007). The Move technique reconstructs flow based on a linear regression ($\hat{y}_i = a + bx_i$) in which the variables a and b were calculated in a special way (Moog *et al*. 1999). For example, Move.I and Move.II (Hirsch, 1982) reproduce the same first and second moments when they generate the entire sequences of $\hat{y}_i$, where $i=1,...,n_1+n_2$ (n_1 is the length of the short record and n_1+n_2 is the length of the long record), compared to historical samples. However, in practice, Move is used to generate $\hat{y}_i$ with $i= n_{1+1},...,n_1+n_2$ (Vogel and Stedinger, 1985). Thus, Move.I and Move.II did not achieve its intended objective. This problem was resolved with the development of Move.III (Matalas and Jacobs 1964).

In general, regression based methods have few independent variables and usually do not take the physical characteristics or dynamics of a system into account. This reduces data authenticity, especially when they are used to reconstruct or extend short time-step and long-term flow values. Thus, it is more reliable to apply the hydrologic models that use both the hydrological and the physical data (such as characteristics of the catchment) for reconstructing flow data values.

2.2.1.2 Hydrologic methods

Hydrologic models use hydrologic data or climate data[5] (Hwang *et al.* 2005) to calculate the flow.

Different studies have been undertaken to develop a relationship between the flow and climate signals in order to identify the predictability of flow or possibility of a non-random pattern in space or time (Fortin, 2001) and in most cases, a significant statistical link has been observed (Fortin and Slivitzky, 2000). Fortin (2001) found that climate has an obvious influence on runoff; he evaluated

---

[5] In this thesis, 'hydrologic data' is referred to data such as precipitation, temperature, evaporation, and etc., but 'climate data' includes climate signals such as sea level pressure (SLP), sea surface temperature (SST), and etc.

the reliability of different climatic indices to see if there was a non-random pattern in space or time, and identified a statistically significant link between Arctic Oscillation (AO) and runoff in northern Quebec. However, the correlation was found at times to be caused by extreme climate conditions. According to Fortin and Slivitzky (2000), river flow often correlate well with the winter temperatures, suggesting that winter temperature could be an indicator of how the regional climate is affected by the global phenomenon of AO. But while the performance of climate-based flow reconstruction methods is good in some areas, some questions remain as to their level of confidence.

RR models are the main group of hydrologic models that are most often used to estimate runoff in time and space. They can estimate runoff at different time-steps in different hydraulic systems and land uses, given that limited measured flow data is available in order to calibrate the model. RR models can range from a simple relation between rainfall and runoff to complex models that also consider the hydrologic and physical characteristics of a region.

Examples of RR models are Thornthwaite-Mather (TM) for calculating monthly flow (e.g. Taylor *et al.* 2006), StormNET for calculating daily or even smaller timescale flow (e.g. Karamouz *et al.* 2011), and the Wright model to calculate mean monthly or daily flow (Adeloye and Nawaz, 1998).

Hydrotel is another RR model developed in the mid-1980s in Quebec and has been used extensively in this province for its ability to factor in the area's meteorological condition by considering snow packs and snow melt. Hydrotel is a physically-based distributed hydrological model that can be run for hourly to daily time steps. In this model, spatial variations in watershed characteristics are taken into account using GIS and remote sensing data (Fortin *et al*. 2006). However, when only meso-scale grid data are available, high resolution data should be calculated from standard meteorological station data using a disaggregation model. In addition to downscaling, calibration is another inconvenience of Hydrolet because it is very time consuming.

HSAMI is another RR model developed by Bisson and Roberge (1985) to simulate the hourly or daily flow series values in Quebec's watersheds. HSAMI is a linear reservoir-based lumped conceptual hydrological model which uses a watershed as a transfer function, whereas the meteorological conditions are used as input data, and has as its output the flow values at the outlet of the catchment. The parameters of this model include five data categories: evaporation, vertical flow (such as rainfall), horizontal flow (such as upstream flow), surface runoff and snow. It has been used for daily forecasting of natural inflows on 84 watersheds, with surface areas ranging from 160 to 69,195 km^2" (Minville *et al*. 2010). However, HSAMI does not take the effects of reservoir on flow into account.

Generally, RR models (with linear and nonlinear functions) can be classified into three distinct groups: metric (data-based, empirical or black-box), parametric (conceptual, explicit soil moisture accounting or grey box), and mechanistic (physically-based or white box) (Wagener *et al.* 2004). However, their usage is mostly restricted only to gauged basins, provided that there is flow data available (even for limited period) to calibrate the model.

Metric models commonly use basin data series (rather than the catchment behaviour) and flow to estimate model structure and parameter values, and as a result, this model is seemingly unsuitable for spatial extension of data in ungauged basin. The limitation of metric models is partially resolved by data-based mechanistic models that "constrain the degree of freedom of such models to those structures that are physically interpretable" (Wagener *et al*., 2004). Examples of metric models are the Artificial Neural Network (ANN) and Transfer Functions.

"Parametric models have a structure (defined by the modeller's understanding of the hydrological system) that is specified prior to use" (Wagener *et al.*, 2004), and is required to be calibrated to adjust to parameters that cannot all be measured independently. As a result of their dependency on flow parameters, parametric models are not easily applied to ungauged catchments. Mechanistic rainfall-runoff models attempt to relate the model parameter with catchment characteristics to avoid calibration; however, this has not been completely successful (Wagener *et al.*, 2004). Other attempts have been made to make the RR models applicable for ungauged basins by regionalization. In this method, the RR model is calibrated for as many basins as possible and the estimated parameters are then transferred to ungauged basins. Theoretically, this method should be simple enough to keep the uncertainty low, along with the number of parameters, but unfortunately it fails to capture flow behaviour with a reasonable degree of accuracy (Wan Jaafar *et al.* 2011 and Madsen, 2000).

Many studies have used regional calibration to simulate low-flows (e.g., Vezza *et al.* 2010, Schreiber and Demuth, 1997), floods (e.g. Eslamian, 2010, Wan Jaafar *et al*., 2011), and continuous flow (monthly, seasonal, and annual) data series (e.g. Singh and Singh 1996, Özçelik and

Benzeden, 2010). However, the approach to reconstruct continuous flow (such as daily) is rarely used (e.g. Kokkonen *et al.* 2003).

Generally, different criteria should be considered for selecting the appropriate RR model in each case study (Vaze *et al.* 2011). The most important of these criteria is the availability of data that will act as input data of model. Model selection also depends on the area climate; for example, if snowmelt, evaporation, or groundwater are important, the model should be able to take into account their effects. For instance, in cold regions with large snow loads, it is necessary to account for snowmelt in the flow estimation (Kim and Kaluarachchi, 2013). Land use also is another important factor that impacts the selection of RR model. For example, SWAT is mostly used to simulate the flow in rural basins (Simic *et al.* 2009) and StormNET (Boss International, 2005) is more efficient flow simulator in urban areas.

The main disadvantage of the RR models is that they most often need to be calibrated, which introduces some difficulties. Firstly, the flow data which are required in order to calibrate the model are not always available for ungauged basins. Secondly, model calibration is time consuming when flow must be estimated for several basins, as the calibration process must be repeated for each basin separately. As well, model calibration always introduces uncertainties if it is done manually.

Selecting among the families of flow reconstruction methods (regression based methods, hydrologic methods) explained in Sections 2.2.1.1 and 2.2.1.2, depends on many different factors such as data availability, expected model flexibility, expected data output reliability, etc. At this point in time, there isn't an algorithm that helps one to select the appropriate method considering these factors. This problem brings up the question of <u>how to select the appropriate method of flow reconstruction considering different mentioned factors that affect the method selection in each case study</u>. The same issue is recognized for Post-R period as well.

Another concern is that there are no recognized methods that are reliable enough to estimate flow data for Pre-R period in ungauged basins. Hydrological methods require model calibration, and thus cannot be used in ungauged basins due to lack of available data. The reliability of regression based methods is also decreased through increasing the time step. Accordingly, the second question is <u>how to reconstruct daily flow for Pre-R period in ungauged basins</u> in Quebec (or any other areas with similar conditions).

2.2.2 Post-R Period

Hydrologic data, climate signals, catchment characteristics, and/or reservoir's data could be available for the basin during the Post-R period. Thus, hydrologic, regression based, or hydraulic method, which depends on these same data, may be applicable for flow reconstruction in this period.

2.2.2.1 Regression based methods

Regression based methods were explained earlier in Section 2.2.1.1 and are presented here only to emphasize the fact that they can be also applied to reconstruct flow for Post-R period if required data are available.

2.2.2.2 Hydrologic methods

Hydrological based methods were explained earlier in Section 2.2.1.2 and is presented here only to emphasize the fact that they can be also applied to reconstruct the flow for Post-R period if required data are available.

2.2.2.3 Hydraulic methods

Hydraulic models are those based on the water cycle rule. One of the most common hydraulic model used in flow data reconstruction is the classic WBE (Equation i.1) for a reservoir which uses regulated flow data of the given reservoir to estimate the flow data. This equation is as follows:

Included in calculated flow values (*F*) of this simplified WBE is the amount of precipitation over the reservoir surface *(P)* and evaporation (*E)* from the reservoir during the time (*Δt),* as well as the interaction *(Int)* between stored water in the reservoir and groundwater. If enough information is available for these variables, they can be separated from flow. Note that the hydrograph of input flow to the downstream reservoir is not exactly the same as that of the outflow from the upstream reservoir (Das and Saikia 2013). In such cases, the flow routing equation should be used to calculate the downstream flow hydrograph in order to increase the likelihood of obtaining reliable results. For example, Smithers *et al*. (2001) used the Muskingum technique to route flows in river reaches of the Sabie River catchment in South Africa.

The WBE can be also applied to basins or rivers. Sokolov and Chapman (1974) describe a few forms of WBE and provide good information about the main components. They reported that by factoring in variables such as infiltration, evaporation (e.g. Guntner et al. 2004) and interaction between groundwater and surface water during the year (when seasonal changes can affect these variables), WBE will yield more accurate results, albeit that these variables are somewhat difficult to account for.

The great advantage of using the WBE method is that it does not need to be calibrated, allowing it to be used for ungauged basins as well. WBE is an easy and fast method which can be easily used to calculate real-time flow data. Moreover, factors such as snowmelt or evaporation can easily be considered in the equation if they are available. Unlike other flow reconstruction methods, the results of model will not be affected by model uncertainty and it will produce reliable results in cases where the input data for WBE are of acceptable quality. This model can be used for either long time step (monthly, seasonally) or short time step (daily) reconstruction, though it is mainly used for long step reconstruction.

Several studies have been conducted in Quebec to reconstruct and filter flow series. Figure 2-1 shows the hierarchy of developed flow reconstruction methods in the province. Most of these methods have been developed based on WBE as it is an easy and fast method that does not require model calibration (which is a great advantage for ungauged basins). Since the results of this method are highly affected by the quality of input data, they may include significantly noisy or unrealistic (negative) values. Thus, validating and filtering the flow data series has been always a concern in this area. For example, Charbonneau and Bérubé (1987) proposed a frequent filter based method for removing noise from flow data series calculated with WBE. This model cannot be used with real-time data because it depends on the future water level of the reservoir. It also underestimates the peak flows because it does not take into consideration short term variations in water levels. Berrada *et al.* (1996) evaluated the performance of several filtering techniques to validate the historical hydrometric data and found that all these methods cannot be used for real-time because they necessitate having pre-knowledge of future flow value. Also, filtering methods underestimate peak flow values as they do not recognise noise affected data inputs from local peaks.

Therefore, the idea of combining the results of WBE with other data series such as flow from RR model was introduced as an alternative solution. For instance, Perreault *et al*. (1995) proposed a procedure that depended on the combination of results from HSAMI, WBE, and neighbouring basin's filtered flow (reconstructed flow using their suggested method in a nearby basin). The approach is based on combined forecasts from two multiple regression models for daily flow estimation. The two applied models are the spatial and the temporal models, and their weights are function of the squared residuals. It appears that this methodology gives good results, which constitutes a considerable improvement in comparison with classical models and real-time estimation methods. This method also does not underestimate peak flows, a necessity for superior water resources management. However, it may overestimate the cumulative amount of flow for a certain number of days (Nguyen and Bisson, 1998). Also, the suggested methodology to reconstruct the flow values in each basin depends on the reconstructed flow in neighbouring basins, which may lead to an increase in the uncertainty by using this method. Implementing the suggested method by Nguyen and Bisson (1998) mitigates these problems as their method is based on regression and exponential smoothing techniques. With this method, the validated flow for each day is regarded as an exponential function of the previous days' flow. However this method is not acceptable for non-stationary data, which requires a more appropriate method that take seasonality into account.

All these studies resulted in a procedure which is presently used in Quebec (primarily by Hydro-Quebec) for calculating and validating the real-time flow based on the classic WBE. This procedure requires a daily manual task to be performed on all basins in Quebec for every day of work (or every day during the critical periods such as spring floods, strong rain events, or intense low water). How daily flow estimates are being calculated in Quebec is summarized in the following steps:

1- Manual validation of all input data to flow calculations (water level, gates' opening, produced electricity, etc.).

2- Calculating flow using classic WBE.

3- Visualizing climate- flow using the data generated and evaluating regional and temporal flow consistency, basin by basin (21 days prior values by default), after which the flow is modified or filtered if necessary[6].
    a. Raw and filtered flow for the last days, and predicted flows for the next days.
    b. Minimum, maximum, and average daily weather series of temperatures, rainfall and snow in the basin.
    c. Weather data from nearby stations.
    d. Flow data of nearby rivers.
4- Analysing the accuracy of flow according to the following considerations and filtering them (if necessary):
    a. Checking the acceptability of flow values compared to the usual raw data (check if there is excessive variation, medium jump and change, low variation, etc.).
    b. Comparing the specific flow values with that of neighbouring basins.
    c. Verifying the magnitude of evapotranspiration when flow data have doubtful coherency (if necessary).
    d. Considering the effect of wind on the variation of water levels variation and the flow calculation (if necessary).
    e. Evaluating the effect of direct rainfall over the reservoir surface on the flow calculation (if necessary).
    f. Evaluating the effect of maneuver disorders on outflow from the reservoir and the flow calculation (if necessary).
    g. Considering the effects of the upstream reservoir's variation on the downstream reservoir's flow and the flow calculation (if necessary).
    h. Considering the effect of ice cover on the water level and the flow calculation (if necessary).

[6] Decision about the necessity of performing any of optional steps is always made by the engineer who is conducting the process of flow filtering. Sometimes the engineer may believe that the estimated flow is highly affected by other factors and it is "necessary" to take them into account. There is no specific rule that can be used to provide clarification to the users if some step is necessary or not

i. Rejecting or accepting the calculated flow for the last week and/or last month based on the similarity of net calculated flow and net flow from WBE during this time period.

5- Determining if there is a need to modify the filtered flow data series of the previous days by:
   a. Preferably, filtering the flow values subjectively with consideration to the following factors:
      i. The anterior form of basin’s hydrograph in the same hydrologic condition.
      ii. The shape of the hydrograph given by the hydrological model (HSAMI).
      iii. Peak specific flow of neighbouring basins.
   b. Trying to respect the average raw flow of last days (or if is impossible last month).

6- Secondary overall work would be done once yearly, (i) to "refine" the processing steps and (ii) adjust any corrections that may have been carried out on the raw data during the year.

7- Recording any problematic basins once a year (the basins where the official flow of Hydro-Quebec is not the calculated flow but the filtered flow).

8- Repeating the procedure for validating and filtering the flow after the historical flow recalculation.

The mostly manual filtering procedure produces a smooth and realistic data series. However, it does contain some limitations and problems, some of which are listed below:

- Time consuming: This method requires considerable time and effort each day.
- Human-caused uncertainties: most of the steps in the manual flow filtering process require human decision making, and thus, it is likely that mistakes may occur due to misjudgments. The values of the filtered flow depend on the judgment of the engineer calculating the flow, which may lead to different engineers giving differing flow value for the same day. This is an obvious source of uncertainty. Also, some steps in this filtering process (such as steps 4.c to 4.h), are performed only if deemed necessary. However, one engineer may consider them important for the recording day while another may see the same steps as unnecessary for the same day. More reliable estimations of flow values require people with greater experience on deciding which factors will be considered into the flow calculations. With

inexperienced employees, the uncertainty of the filtered flow series could increase due to the lack of experience in calculation judgment calls.

- Debateable and/or insufficient information: Post-2005, the action of validating input data and the filtered flow values became a required daily task. However, for periods that were prior to 2005, the input data may contain uncertainties, putting into question the reliability of mostly manually filtered flow data. But if one were to filter the input data and use the same described mostly manual procedure for reconstructing and filtering the flow series values for the time period before 2005, a lot of accurate daily information and data would be required for this task, which is not always readily available for that period. Values such as historical water level, turbine flow, and spilled flow data series (pre-2005) contain a large amount of noise for some time periods and there is not enough information to efficiently filter them.

Therefore, a flow filtering method independent from the human decision-making and experience was missing for the time period after reservoir construction in Quebec. Thus, Perreault (2011) suggested an optimization model which reconstruct and filter flow series values by minimizing the variation of flow values and the WBE's errors. His model is explained in details in the following section.

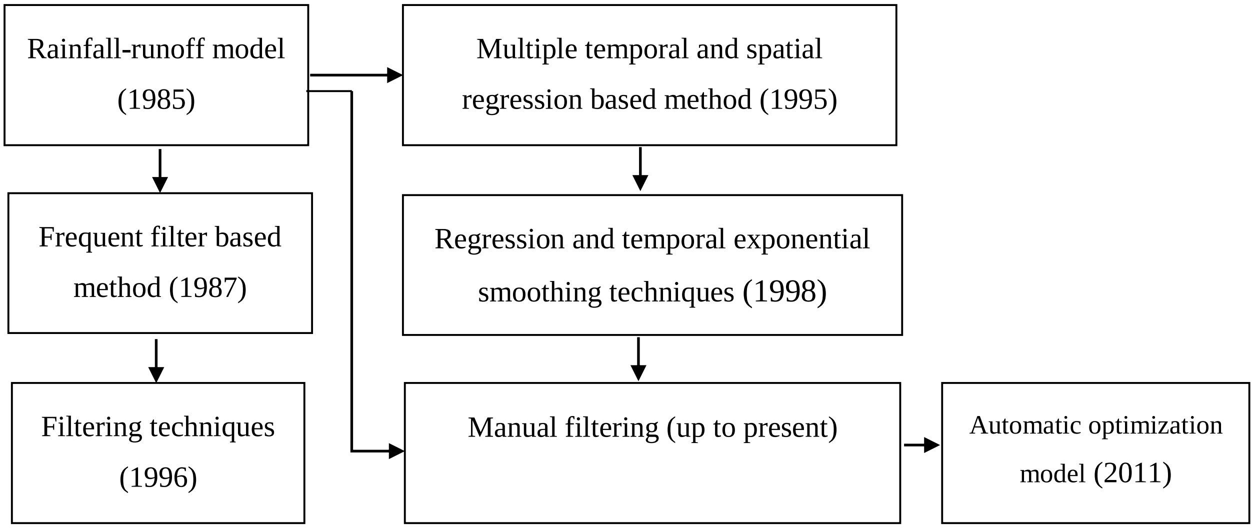


Figure 2-1: The hierarchy of developed methods for flow reconstruction and filtering in Quebec

2.2.2.3.1 Suggested method by Perreault (2011) for flow reconstruction

Perreault (2011) suggested a WBE-based optimization model (POM) to estimate hourly flow. POM uses the classic WBE to define the values of flow. At the same time, it tries to minimize WBE error and the flow variations from one hour to the next. The quadratic objective function of this optimization model is expressed by Equation 2.2:

$$Minimize \sum_{n=starthour}^{endhour} E^2_{\inf(n)} + \sum_{n=starthour}^{endhour} E^2_{\sup(n)} + \gamma \times c \times \left(p \times \sum_{n=starthour}^{endhour} Z_{(n)}{}^2 + q \sum_{n=starthour}^{endhour} \Delta_{(n)}{}^2\right) \tag{2.2}$$

where:

$E_{\inf(n)}$ and $E_{\sup(n)}$ = WBE's error during the nth hour

$Z_{(n)}$ and $\Delta_{(n)}$ = Flow variations over 2 and 3 consecutive hours, respectively.

In this equation, $E_{\inf}$, $E_{\sup}$, *Z*, and Δ are variables in m^3/s (range of these variables are defined in Equations 2.11 to 2.13), and *c*, *p*, *q*, γ, and time interval of *dn=1+end hour-start hour* represent the model's parameters (which are set as *dn*=3, *p*= *q*= γ=1, and *c*=10000 in winter and *c*=1 in rest of the year of corresponding period)

Equation 2.2 tries to minimize the summation of squared errors ($E_{\inf}^2$ and $E_{\sup}^2$) and of the variations (*Z* and Δ) for a time interval with the length of *dn* (the optimization window). In this equation, *dn* is the number of days for which the model is solved (or the length of the optimization window in which the WBE is solved). That is, the whole period is divided into several time periods with the length of *dn,* and POM is being solved for each *dn* hour.

For instance, as it is shown in Figure 2-2, a sample year (8760 hours) is divided into time periods with the length of 3 hours (*dn*=3). Thus, for the first time, the optimization model is solved for hours 1 to 3 (*start hour* =1 and *end hour* =3) and the flow values are estimated for these three hours. In the second step, the optimization model is solved for the hours 4 to 6 (*start hour* =4 and *end hour* =6) and the flow values are defined for these three hours. This means that the optimization window is non-moving.

In Equation 2.2, the parameters of *p* and *q* assign weights to the variables *Z* and Δ, and their ratio defines the importance of each of these two variables in minimizing the flow variation. The parameter *c* assigns a weight to the set of variation variables *Z* and Δ, and its magnitude defines the importance of the variation variables compared to the squared errors ($E_{\inf}$ and $E_{\sup}$). γ is another parameter which does not seem necessary for the model because it does exactly the same task as *c*.

Subject to (equations 2.3 to 2.14):

$$Volume_{(starthour)} + E_{\inf(starthour)} \geq v_{\inf(starthour)} - offset_n \tag{2.3}$$

$$Volume_{(starthour)} - E_{\sup(starthour)} \leq v_{\sup(starthour)} - offset_n \tag{2.4}$$

$$offset_n = (v_{\inf(\min),n} + v_{\sup(\max),n})/2 \tag{2.5}$$

where:

$Volume_{(starthour)}$ = the volume of the reservoir (m^3/s) at the beginning of time step (all the volume units are changed from hm to m^3/s by being multiplied by $\frac{1000000}{3600}$),

$v_{\inf(starthour)}$ = the minimum measured reservoir volume (m^3/s) during the $starthour^{th}$ hour (reservoir volume is measured every 5 min),

$v_{\sup(starthour)}$ = the maximum measured reservoir volume (m^3/s) during the $starthour^{th}$ hour,

$v_{\inf(\min),n}$ = the minimum of $v_{\inf(n)}$ set (m^3/s) where $n=starthour,\ldots, endhour$,

$v_{\sup(\max),n}$ = the maximum of $v_{\sup(n)}$ set (m^3/s) where $n=starthour,\ldots,endhour$,

offset = the mean value of extreme volume values (measured every 5 min) in the optimization window (Equation 2.5).

$Volume_{(starthour)}$ is a variable which could change in theory between 0 and reservoir's maximum capacity, but Equations 2.3 and 2.4 force it to stay between minimum and maximum measured volume. In this equation, the summation of $E_{\inf(0)}$ and $E_{\sup(0)}$ is the error of $Volume_{(starthour)}$, and *offset* is subtracted from the volume data to normalize them to around zero.

Equations 2.6 and 2.7 are the core constraints because they represent the WBE and are used to control the range of water budget.

$$Volume_{(starthour)} + \sum_{i=1}^{n} F_i + \sum_{i=1}^{n} qin_i - \sum_{i=1}^{n} qout_i + E_{\inf(n+1)} \geq v_{\inf(n+1)} - offset \tag{2.6}$$

$$Volume_{(starthour)} + \sum_{i=1}^{n} F_i + \sum_{i=1}^{n} qin_i - \sum_{i=1}^{n} qout_i - E_{\sup(n+1)} \leq v_{\sup(n+1)} - offset \tag{2.7}$$

Where:

q_{in} = the measured inflow to the reservoir (which is the released flow from the upstream reservoir),

q_{out} = the measured outflow from the reservoir (which is the summation of turbine flow and spilled flows),

F = a variable referring to flow caused by effective rainfall (m^3/s - range of this variable is defined in Equation 2.10).

In this equation, $E_{\inf}$ and $E_{\sup}$ are the inferior and superior errors of WBE. Equations 2.6 and 2.7 estimate each flow value by considering the other flow values of the optimization window.

$$F_{(n+1)} - F_{(n)} = Z_{(n)} \tag{2.8}$$

$$F_{(n)} - F_{(n-1)} - F_{(n+1)} + F_{(n)} = \Delta_{(n)} \tag{2.9}$$

$Z_{(n)}$ is the difference between flow for day number *(n+1)* and *(n)*, while $\Delta_{(n)}$ is related to the variation of flow during three consecutive days. When *n* is equal to *endhour*, $Z_{(n)}$ is assumed equal to zero and when *n* is equal to *endhour/starthour*, $\Delta_{(n)}$ is assumed equal to zero. These assumptions cause poor boundary conditions.

$$0 < F < +\infty \tag{2.10}$$

$$-\infty < E < +\infty \tag{2.11}$$

$$-\infty < \Delta < +\infty \tag{2.12}$$

$$-\infty < Z < +\infty \tag{2.13}$$

$$n = starthour, ..., endhour \tag{2.14}$$

According to Equation 2.10, flow should be greater than zero. However, *E*, *Z*, and Δ could be positive or negative (Equations 2.11, 2.12, and 2.13). Also, *n* changes between *starthour* and *endhour* for each optimization window (Equation 2.14).

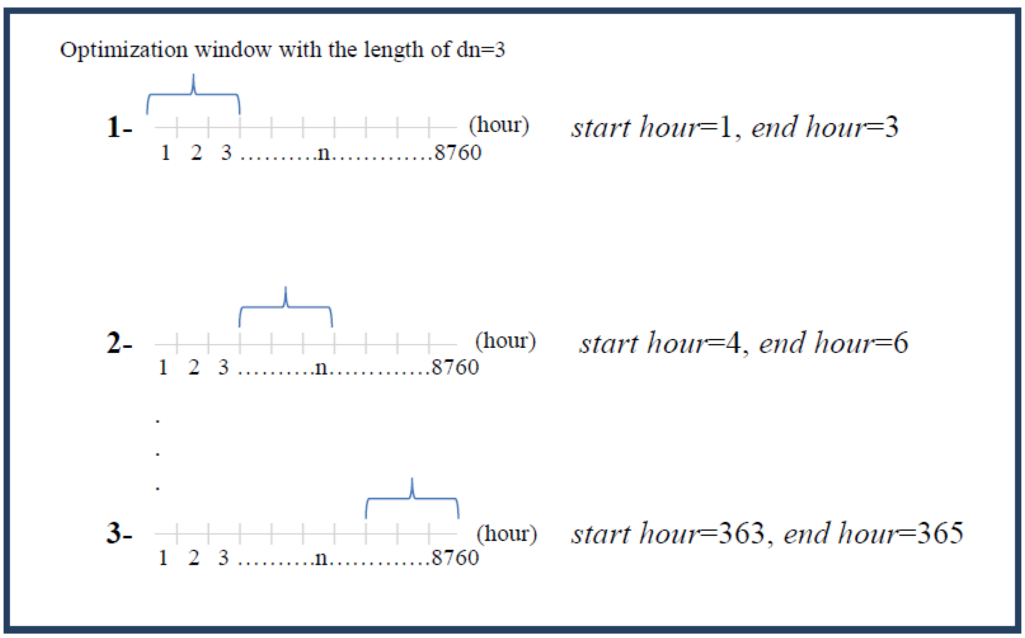


Figure 2-2: The schematic of optimization window in the POM for a hypothetical year

POM has some advantages over the simple WBE:

- It tries to minimize the flow variation (noise) and error.

- It does not allow flow to take negative values.
- It has an optimization window with the length of *dn* and flow values are estimated with consideration to the other values of flow in that window.

These advantages are all inspired by the existing manual flow filtering process. For instance, like POM, the negative flow values are replaced by positive ones and flow values are modified by factoring in the values of previous days in the mostly manual flow filtering process. However, POM has still some deficiencies and limitations, as listed below:

1. The results of this model are still noisy, especially during periods of low flows.
2. The parameter of γ seems unnecessary.
3. It does not factor in all the available information of the basin such as simulated flow and neighbouring basins' flow to improve the quality of the reconstructed flow.
4. The designed optimization window is fixed, resulting in values near the boundary of optimization window to suffer from poor boundary condition. This imperfect boundary conditions include the following assumption about Z and Δ: if n=*endhour*, then Z=0 and if n=*starthour* or n=*endhour*, then Δ=0. Thus, in each optimization window, the first and the last flow values are affected by a definite uncertainty.
5. This model was developed for hourly flow reconstruction and its reliability decreases when larger time steps are applied, such as daily time periods (24 hours). This is because in the case of daily time-steps, $v_{\inf(n)}$ and $v_{\sup(n)}$ need to be calculated for 24 hours (instead of an hour) and this augments the range of WBE in Equations 2.6 and 2.7, which increases the uncertainty of the model output. The model may even fail to work in this situation.
6. POM is applicable only when volume data of 5-minute intervals (volume data measured every 5 minutes) are available. Otherwise, $v_{\inf(n)}$ and $v_{\sup(n)}$ should be defined based on volume data of 1-hour intervals, if they are available. This results in the model producing greater uncertainty in the data results. The model may even fail to work because the range of inequality in Equations 2.3, 2.4, 2.6, and 2.7 may be narrower. Also, when volume data

of 5-minute and hourly intervals are not available, the method is no longer applicable as $v_{\inf(n)}$ and $v_{\sup(n)}$ cannot be defined.

7. The model parameters are considered constant during the time and space. This assumption evokes the question of 'how are we to be sure that the selected parameters are the best possible when time and case-study change?' Hence, a sensitivity analysis is required in order to define if this assumption for the model is appropriate or not. This will be done in Chapter 6.

None of the aforementioned research for Post-R period provides a flexible methodology to reconstruct and validate reliable daily flow data series that is independent of human decision making. This gives rise to an interesting question of <u>how to reconstruct more likely daily flow for Post-R period in ungauged basins</u>.

2.3 Performance of Flow Reconstruction Method

It is necessary to evaluate the data quality of reconstructed flow to confirm the performance of the flow reconstruction model. This evaluation could be done using different QIs and stationarity and regional homogeneity tests.

2.3.1 Quality Indexes

The Quality Indexes are criteria that are applied to the model in order to evaluate the integrity levels of a data series. Each Quality Indexes has its own characteristics but not all of them can satisfy every user (Weglarczyk, 1998).

Some of the more popular QIs are Mean Square Error (i.e Gupta *et al.* 2009), Temporal or Spatial Correlation Coefficient (i.e. Johnston *et al.* 2009), and Nash-Sutcliffe efficiency (i.e. Villa-Alvarado *et al.* 2014, Johnston *et al.* 2009). Krause et al. (2005) compared nine different efficiency criteria, including Coefficient of Determination (r^2), Nash-Sutcliffe efficiency, Index of

Agreement, Nash-Sutcliffe efficiency with logarithmic values, and Relative Efficiency criteria applicable for hydrological model assessment. They tested these criteria by applying them on three different examples. According to their paper, each Quality Index has its own advantages and disadvantages, and could be used for specific purposes. For example, the Coefficient of Determination is not sensitive to flow over- or underestimations. Also, a large disadvantage of Nash-Sutcliffe efficiency is that the difference between measured and simulated values is calculated as a squared value. As a result, lower values are neglected while larger values are emphasized more, showing that this QI is not an appropriate index to use in evaluating the flow data series during the low flows.

Since most existing QIs compare the simulated flow values with the measured flow values, they cannot be applied for the basins with unavailable measured data series. This means that the subject of evaluating the quality flow has remained an issue in ungauged basins. This challenge raises the questions of <u>how to evaluate the quality of reconstructed flow series in ungauged basins</u>.

2.3.2 Temporal and Regional Homogeneity Tests

Temporal homogeneity (stationarity) in a data series implies that the homogeneous behaviour of a data series is time independant; its statistical properties will stay the same over time. More precisely, in the case of stationarity, the joint probability distribution of the process remains unchanged over the time. It is important to analyse whether or not the presence of stationarity in the flow data series is created artificially by the reconstruction method. Although there are some means to evaluate the stationarity of a time interval series, it is also visible at its time plot; a time series would be stationary if its time plot appears similar at different points along the time axis (Nagpaul 2005).

It is very common to employ a simple stationarity test, which indicates if the mean and variance of a data series are constant or not. Turner and Twieg (2005) divided the data series into S equal segments of size N, and then administered T-statistic with $2N-2$ degree of freedom to compare the segments' means and F-statistic with $N-1$ degree of freedom to compare the segments' variances. They defined a Wide Sense Stationary (*WSS*) index based on the results of F-test and T-test and assumed that the data series would be stationary if *WSS* exceeded 0.9.

Perreault *et al.* (1996) assessed the stationarity of average annual aggregate flow using a Bayesian procedure to detect change in mean annual series of flow, and found three different groups with three different averages. Given the limited information available in their project, it was difficult to distinguish between non-stationarity of the mean and the presence of sustained deviations caused by the autocorrelation. Despite the results obtained from the precipitation series, which support the presence of a change in the average, it is difficult to deduce the non-stationarity average in the series of annual flow.

Some of the common tests designed to evaluate the stationarity of a data series are Mann–Kendall (e.g. Cunderlik and Burn 2003), Dickey–Fuller, Augmented Dickey–Fuller (ADF) (e.g. Oh 2005), and Kwiatkowski-Phillip-Schmidt-Shin (KPSS) (Kwiatkowski *et al.* 1992). Among these, KPSS, which originated from econometrics, has been widely used in hydrological studies (i.e. Wang *et al.* 2005 and 2006). Although KPSS was originally developed for data with short memory, Lee and Schmidt (1996) deployed it to assess the stationarity of long-memory data and found it to be adequate. "KPSS tests the null hypothesis of stationarity around a deterministic trend (trend stationarity) and the null hypothesis of stationarity around a fixed level (level stationarity). If a process is not level stationary but trend stationary, it indicates that the process may be decomposed into a trend component and a stationary component" (Wang *et al.* 2005 and 2006).

On the other hand, it is expected that the catchments within the same region have homogenous flow data series (regional homogeneity) because they have more or less the same geographical and hydrological characteristics. Thus, testing of the regional homogeneity of reconstructed flow is required.

Cunnane, (1978) compared the T-year return levels with standardized empirical return levels from 14 sampled sites by plotting them against each other. They also compared the standardized empirical discharge and the fitted values by LN3 distribution in a graph to evaluate the regional homogeneity of flow. Gustard and Demuth, (2008) also used graphical methods by plotting empirical quantiles (observations) against the distribution quantiles to assess how well the homogeneity of data is according to applied distribution. They checked if the points lie close to the theoretical curve (probability plot) and the unit diagonal (P-P plot for probabilities and Q-Q plot for quantiles).

According to Albertini, 2013, some of the standard and most common plots used to assess the homogeneity of a data series and to identify unusual observations are Tukey-Anscombe plot (residuals vs. fitted values), Normal plot, Scale-Location plot, Normal Quantile-Quantile plot (Q-Q plot), and Leverage Plot (i.e. Henry, 2013., Lashermes *et al.* 2007, Sofia *et al.* 2011).

2.4 Uncertainty Analysis

Uncertainty is a component of a reported value that characterizes the range of values within which the true value is asserted to lie (NDT recourse center, 2014) "Uncertainty analysis of hydrological modeling has become an indispensable element for any hydrologic modeling and forecasting. The uncertainty of the prediction from one single model has been known to arise from input data, model structure and the process of parameter calibration" (Dong *et al.* 2013).

There are many different methods available to evaluate the different types of uncertainty found in a data series. Among those that are extensively used are the Generalized Likelihood Uncertainty Estimator (GLUE) (Beven 2007; Nott *et al.* 2012, Shen *et al.* 2012; Beven and Freer 2001), Markov Chain Monte Carlo, MCMC, (Kuczera & Parent 1998), Sensitivity Analysis (Abebe *et al.* 2010, Johnston *et al.* 2009), and the Bayesian method (Bates and Campbell 2001; Khu and Werner 2003; Li *et al*. 2010, Engeland *et al.* 2005). Comparing the different methods, GLUE is the most popular because of its conceptual uncomplicated nature, simplicity to apply, and flexibility. However, "it is computationally inefficient and can even lead to misleading results, unless a large sample may be drawn" (Han *et al.* 2014).

GLUE, MCMC, and most of the existing sensitivity analysis methods are applied to define the parameter of uncertainty. However, the Bayesian model is a method which considers uncertainty of all input data, output parameters, and model's structure (Yang *et al.* 2007) but it requires a specific likelihood function.

Although different uncertainty analysis techniques have been widely proposed and applied, most of them are dependent on the measured data. Therefore, it is still a challenge to find out <u>how to evaluate the uncertainty of flow data in ungauged basins</u>.

CHAPITRE 3 CASE STUDY

3.1 Introduction

In the research presented in this thesis, the methodology and the results are divided into several sub-methodologies and sub-results related to Pre-R and Post-R periods, which are presented in Chapters 4, 5, 6, and 7. The focus of this chapter will be on the flow reconstruction case study.

Since the topic of flow reconstruction is seen as a challenge for watersheds located in Quebec, a basin in the province of Quebec will be chosen as the subject of our case study in order to evaluate the performance of the recommended methodology for reconstructing daily flow in actual situations. Since the goal of the developed method for reconstructing daily flow in this thesis is to be suitable for use on all basins in Quebec, the Outardes basin (with its three reservoirs) has been selected as the subject of this case-study (Figure. 3-1). The reasons for the selection of the Outardes basin over other basins in the province are the following:

i) the reservoirs possess hydraulic systems with different types and numbers of gates and turbines;

ii) the basin includes reservoirs of different sizes and characteristics;

iii) the basin has a simple structure with its reservoirs arranged in series.

3.2 Case Study Description

As shown in Figure 3-1, the Outardes basin consists of three sub-basins designated as Outardes 4, Outardes 3, and Outardes 2, with each of them possessing a reservoir (see Figure 3-2 for the reservoir schematics of this basin). Outardes 3 is a small basin located downstream of Outardes 4, which is a big watershed with a large reservoir. Since Outardes 4 is located in the upstream position, the only input flow into this reservoir is the flow which is caused by rainfall). Also, Outardes 2 is a moderate sized reservoir situated downstream from Outardes 3. Among the three sub-basins of the Outardes watershed, Outardes 3 is considered the reservoir of the greatest challenge because it is small and is more susceptible to changes from the releasing outflow of the larger Outardes 4 reservoir. Any increase or decrease of discharged flow from Outardes 4 greatly affects the water level in Outardes 3, resulting in fluctuations that make it difficult to estimate flow values in this basin.

In Quebec, the operation mode is the same for all the reservoirs as they all regulate the river for short and medium term intervals. For instance:

- The reservoirs are generally emptied before spring floods for safety reasons. This allows for flood storing and flood routing inside the reservoir.
- Water depth downstream from the reservoirs is set according to the safety and economic requirements. For example, low water levels downstream of a reservoir may increase the height of the water drop, thus exceeding power production norms.
- Generally, the reservoirs are kept full (if it is safe) during the winter to maximize the hydropower production rate and meet electricity requirements during the winter season.

However, the exact daily operation (rate of water that should be taken from each reservoir) is defined based on short term (i.e. daily) decision. The decisions are usually made based on short-term predicted flow and daily measurements.

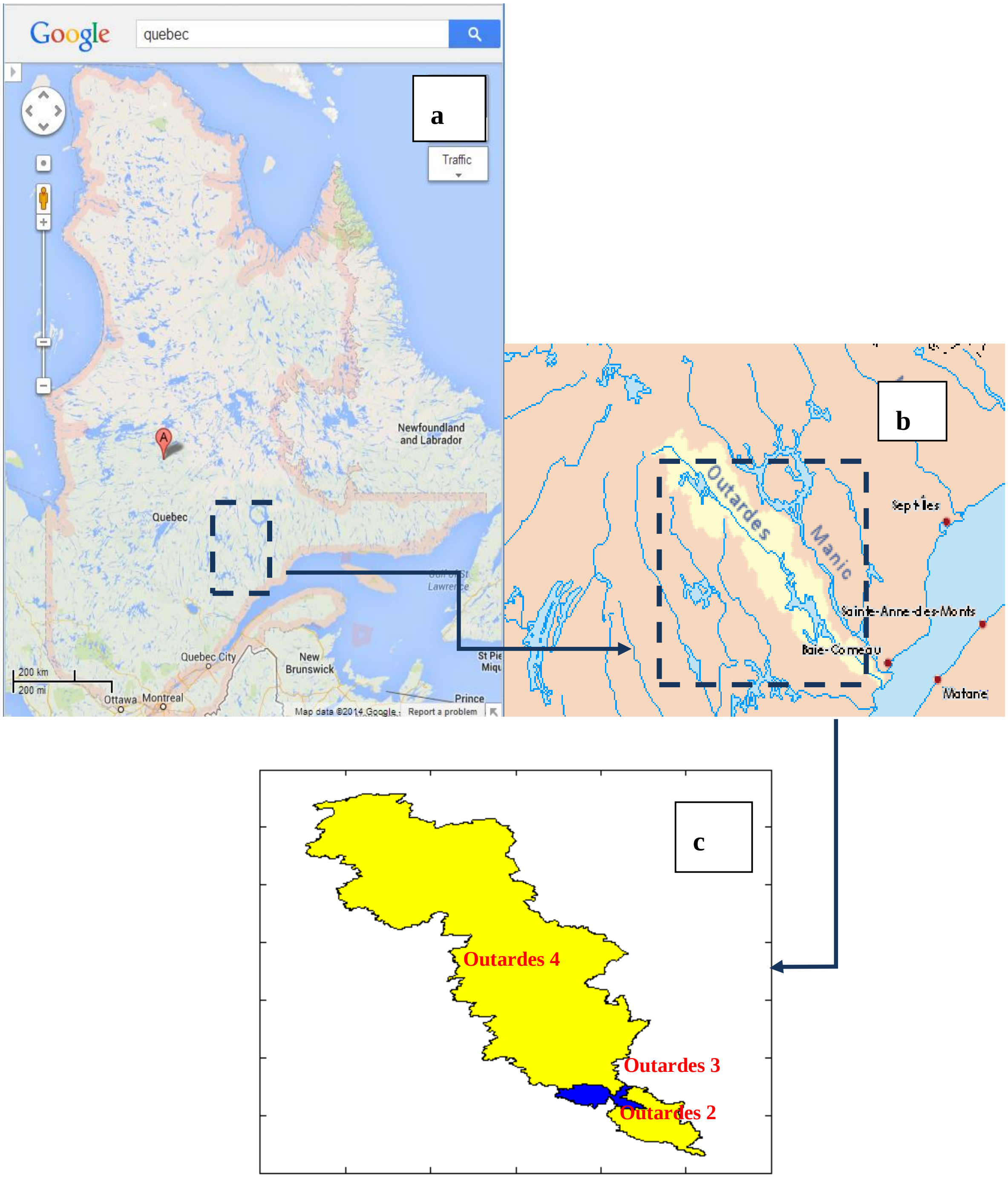


Figure 3-1: Location of the Outardes Basin in Quebec (a and b), and sub-basins of Outardes 4, Outardes 3 (in darker color), and Outardes 2 (c)

(a: https://maps.google.ca, b: www.wikipedia.org)

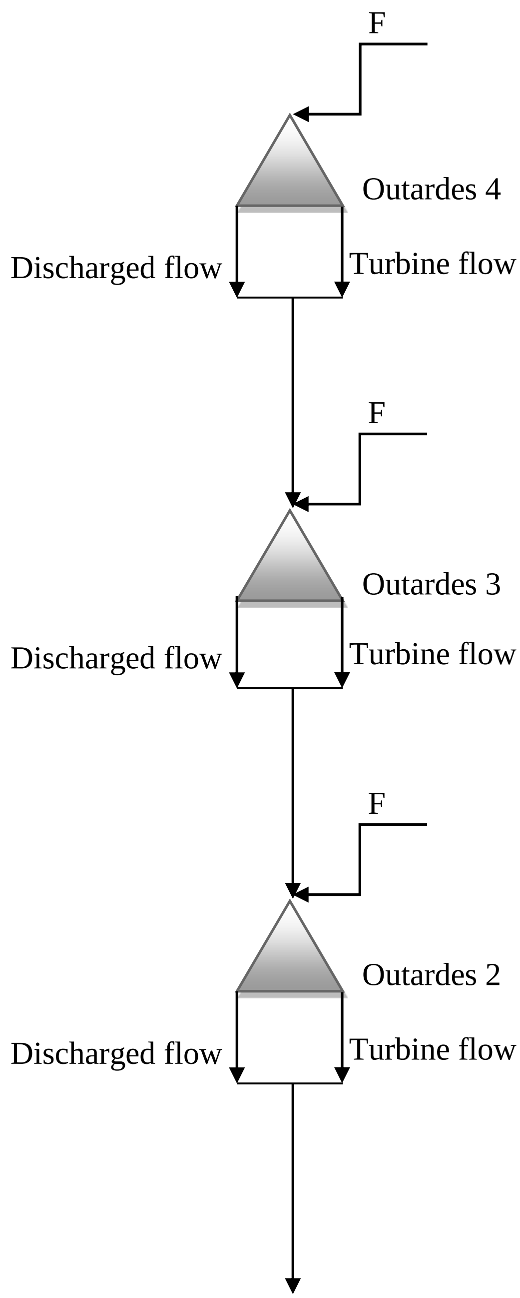


Figure 3-2: Schematic of three reservoirs of Outardes 4, Outardes 3, and Outardes 2

3.3 Available Data and Information

Table 3.1 shows the general characteristics of the Outardes sub-basins. One of the available sets of data related to the Outardes sub-basins (for both Pre and Post-R periods) is the flow data series measured from gauged basins in the surrounding area, as shown in Table 3.2.

The flow data series for these gauged basins were obtained from the CEHQ website (CEHQ, 2013) and used in the flow reconstruction process and quality evaluation procedure of the Outardes sub-basins. Thus, for each case study, two of the neighbouring basins should be selected among the measured catchment areas; one for the flow reconstruction process, and the other for quality evaluation of the reconstructed flow. The criteria used for selecting the neighbouring basins for each case study watershed are the following:

i) to be physically close to the watershed because of its possibility of sharing similar geological and hydrological characteristics,

ii) to have relatively the same surface area as the case study watershed, which increases the likelihood of both having similar flow characteristics,

iii) to have sufficient time intervals of measured flow data series,

iv) and last, but not least, should have comparable hydrographical shape with the case study watershed because of its possibility of having similar weather, drainage system, and catchment characteristics (such as watershed slope that affect the time of concentration and eventually sharpness of hydrograph).

As a result, the Moisie basin will be selected as the primary neighbouring basin for Outardes 4 and used for its flow reconstruction, while the sub-basin Romaine will be chosen as its secondary neighbouring basin and used for the purposes of reconstructed flow evaluation (Table 3.3). The sub-basin Godbout is considered the most appropriate choice as the primary neighbouring basin for Outardes 3 and Outardes 2, while Moisie will be used in a dual role as the secondary one for these two watersheds (note: the Outardes river was measured by Hydro-Quebec for a limited time before construction of Outardes 4 reservoir. The flow data of that period is used to evaluate the quality of reconstructed flow for Pre-R period in this sub-basin).

The flows from all the measured neighbouring basins are also used for regional homogeneity assessment of reconstructed flow.

Since some features of basins and data availabilities are different for Pre-R and Post-R periods, more specific characteristics for each period are defined in sections 3.3.1 and 3.3.2.

Table 3.1: Characteristics of case study basins

Characteristic of basins	Outardes 4	Outardes 3	Outardes 2
Basin area (km2/100)	171.19	4.85	13.02
Average long-term minimum temperature (°C)	-6.76	-3.71	-3.06
Average long-term maximum temperature (°C)	4.38	7.08	6.84
Average long-term rainfall (cm/day)	0.17	0.18	0.19

Table 3.2: Characteristics of gauged basins in the neighbourhood of case study

Basin Name	Latitude	Longitude	Start-End	Number of Years	Area ($km^2/100$)	Specific Flow-2007 ($L/s/Km^2$)
Godbout	49.55222	-68.09611	1975-present	39	15.77	23.13
Moisie	50.59028	-66.31861	1966-present	48	190.12	20.04
Magpie	51.14889	-64.97389	1979-present	35	72.01	22.53
Romaine	50.52167	-64.04056	1957-present	57	129.22	6.39
Natashquan	50.72083	-62.18944	1981-present	33	156.93	21.57
Saint-Paul	52.28556	-58.00306	1968-present	46	55.04	17.10
Petit Saguenay	48.31694	-70.085	1999-present	15	4.33	11.06
Chicoutimi	48.52278	-71.35278	1911-present	103	34.44	13.58
Aux Ecorces	48.30722	-72.08056	1972-present	42	11.15	20.76
Pikauba	48.57139	-71.63861	1970-present	44	4.9	22.91
Metabetchouane	48.63417	-72.65833	1978-present	36	22.12	15.79
Ouiatchouan	48.34528	-72.41	1967-present	47	5.76	13.55
Petite Peribonca	49.36167	-72.08361	1975-present	39	10.20	13.86
Ashuapmushuan1	49.14917	-72.82139	1954-present	60	155.15	16.23
Ashuapmushuan2	49.46778	-73.6	1962-2011	49	111	17.73
Mistassibi1	49.50944	-72.35472	1954-2004	50	93.2	15.51
Mistassini2	49.49056	-72.45583	1953-present	61	96.31	15.43
Manouane	50.47056	-71.55028	1980-2013	33	36.86	20.71
Valin	48.82361	-71.6275	1975-present	39	7.68	27.22
Ste-Marguerite	48.45333	-70.52167	1998-present	16	10.97	19.37

Table 3.3: Neighbouring basins selected for each case study in order to be used in the process of flow reconstruction (primary neighbouring basin) and reconstructed flow evaluation (secondary neighbouring basin)

	Outardes 4	Outardes 3	Outardes 2
Primary neighbouring basin	Moisie river	Godbout river	Godbout river
secondary neighbouring basin	Romain river Outardes river	Moisie river	Moisie river

3.3.1 Pre-R Period

The Pre-R time period includes the years between 1960 and 1968 for Outardes 4 and Outardes 3, and the years between 1960 and 1977 for Outardes 2. Available data within this time period for the Outardes basins is limited only to the hydrologic data consisting of daily rainfall, minimum temperature, maximum temperature, actual temperature, and snowfall (Hydro-Quebec's database). Because of their high reliability and quality, Hydro-Quebec has already used these sets of hydrologic data to simulate flow data series using a RR model for the basins of Quebec. Since the measured flow is not available for most basins in Quebec, the RR model has been calibrated and validated using filtered flow (the flow that was mostly filtered manually by Hydro-Quebec) instead of measured flow. Thus, for the periods that filtered flow are not reliable, the quality of these results may be questionable.

3.3.2 Post-R Period

In addition to measured hydrological data (such as rainfall, temperature, and snowfall) and simulated flow using RR model, more historical data exists for Post-R period in this area.

The Outardes 4, Outardes 3, and Outardes 2 reservoirs were built in 1969, 1969, and 1978 respectively, upon which data such as water level on the upstream and downstream of reservoir, gate openings, and produced electricity began to be measured. The collected data were then used to calculate turbine flow, discharged flow, and reservoir volume (data used in the case study were obtained from Hydro-Quebec's data base). However, the quality of measured data have always been affected by different factors such as natural phenomena (i.e. floods, ice cover), instrument disorders (i.e. gates' maneuver disorders), instrument uncertainties, neglecting flow routing, simplification of calculations, and human uncertainties. Since the uncertainty of the input data will clearly affect the quality of the flow values or any other data series generated, the filtering of input data (mostly manual) began after 2005. Table 3.4 shows list of available data and their reliability status for different time periods.

The constructed reservoirs are different in their size, number and type of gates, and turbines (which make them different for hydraulic calculations). For example, the conveying dikes to the turbines may be separate or common. Some general information about the Outardes hydraulic systems is tabulated in Table 3.5. According to this table, each reservoir has several characteristic sets for their turbines which are associated to varying characteristics of hydraulic system during the time (information is again obtained from Hydro-Quebec's database).

Table 3.4: List of available data and their validity situation for Outardes basin

	Data availability in Outardes (Pre-R)-Status of validity	*Data availability in Outardes (Post-R, before 2005)-Status of validity*	*Data availability in Outardes (Post-R, after 2005)- Status of validity*
Rainfall	available- validated	available- validated	available- validated
Temperature	available- validated	available- validated	available- validated
Snowfall	available- validated	available- validated	available- validated
Measured Neighbouring basin	available- validated	available- validated	available- validated
Simulated flow using RR model	available	available	available
Turbine flow	not available	available- calculated based on not validated data	available- calculated based on validated data
Discharged flow	not available	available- calculated based on not validated data	available- calculated based on validated data
Reservoir volume	not available	available- calculated based on not validated data	available- calculated based on validated data

Table 3.5: General characteristics of reservoirs of Outardes basin

	Outardes 4	*Outardes 3*	*Outardes2*
Construction of reservoir	1969	1969	1978
Maximum reservoir storage volume (hm3)	10940.44	14.72	16.21
Number of groups of turbines	4	4	3
Number of types of turbines	4	2	3
Number of sets of turbines' characteristic	25	15	30
Number of gates	3	3	6
Type of gates	spillway	spillway	bottom gate
Number of sets of reservoir's characteristic	2	1	1

3.4 Conclusion

This chapter presents a detailed description of the Outardes basin in Quebec, which will be used as the subject basin of our case study. The developed methodologies of daily flow reconstruction in the current project will be applied on the three catchments of this basin to evaluate their performance in actual situations. The suggested methodologies and the results of applying them on Outardes 4, Outardes 3, and Outardes 2 are presented in Chapters 4, 5, 6, and 7.

CHAPITRE 4 AN ALGORITHM FOR SELECTING THE MOST APPROPRIATE FAMILY OF METHOD FOR FLOW RECONSTRUCTION

4.1 Introduction

Having knowledge of a basin's flow data series is essential for estimating water availability, predicting extreme events, designing hydraulic structures and undertaking other activities related to water planning and management. But flow data is not always measured (even in cases of developed countries), and when recorded data is available, it may contain some level of uncertainty or gaps, which is why reconstructing flow data is necessary in these cases. The first step to achieving this is to select the appropriate family (group) of flow reconstruction methods.

Previous studies have attempted to reconstruct or complete flow data at a given reservoir in order to increase the quality of information used for water management. Regression based methods were used for years to reconstruct and extend flow data (e.g. Kevin, 1996, Rupp *et al.* 2008, Hernandez-Henriquez *et al.* 2010, Kim and Pachepsky, 2010). These methods usually relate the flow value to one or more independent variables such as rainfall and temperature. The regression based methods are still used in some cases because they are fast, simple to use, and can be developed using a minimal amount of information from the basins. Although the results of this method can be good when rough estimations are adequate, they are not highly reliable in estimating flow using short time intervals (daily or hourly).

Hydrological and hydraulic methods are the two main alternative groups of flow reconstruction methods. Generally, hydrological methods (such as rainfall-runoff models) rely primarily on climate and/or hydrological input data to simulate flow. Hydraulic methods include those based on the water cycle (such as WBE). A literature review on these methods shows that each may be appropriate in a particular case, depending on the availability and quality of data, the desired time interval measurement, the flexibility and uncertainty of the reconstructed data, climate, and length of the reconstruction period. However, there is no in-depth general study showing how to select the appropriate method for flow reconstruction which considers all these factors in one case study. Therefore, in this chapter, an algorithm will be presented in order to help select the appropriate method in each case (Figure 4-1, box # 1). This algorithm has been developed based on the characteristics, advantages, disadvantage, and applicability of existing models and methods as presented by the literature review in Chapter 2 of different methods of flow reconstruction. In fact, this algorithm achieves the first step of the current project - the choosing of the family of flow reconstruction methods - that will lead to reaching the final goal of flow reconstruction in ungauged basins.

The developed algorithm will then be used to define the appropriate family of flow reconstruction methods for Pre-R and Post-R periods in the given case study example (see also Figure 4-1). This example will confirm the capability of the algorithm to select the appropriate family of flow reconstruction method. At the conclusion of this chapter, it will be revealed that the families of WBE based models and regression based methods are respectively the most appropriate choices to be used for flow reconstruction for Post-R and Pre-R periods.

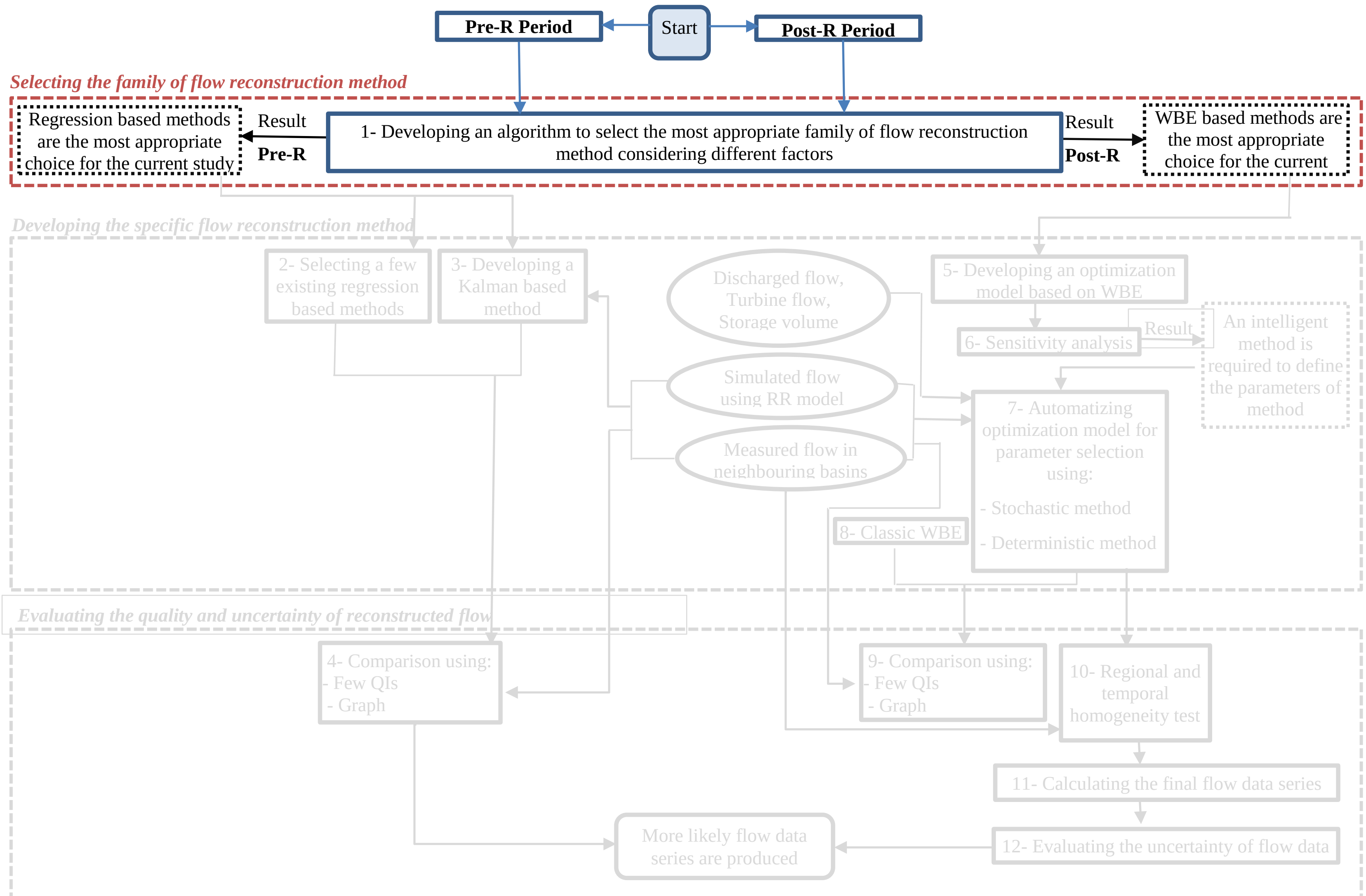


Figure 4-1: Schematic of methodology-Selecting the family of flow reconstructing

4.2 Proposed Methodology for Selecting the Appropriate Family of Flow Reconstruction Method

According to the literature review presented in Chapter 2, recent research have employed various reconstruction methods that differ according to their fundamental concepts and equations, the input data they require, their uncertainty and flexibility, and their range of application. Each method may have particular value in a given circumstance, depending on data availability and the overall objectives. However, none of the studies in the review provided a methodology for selecting a flow reconstruction method that considered all of the factors mentioned above.

A number of criteria should be considered when selecting a method of flow reconstruction. These criteria can be summarized as:

- required flexibility
- available input data
- quality of input data
- desired flow time interval and period
- desired certainty
- climate and other features of the area

The goal of this chapter is to formulate an algorithm that will help select a family of flow reconstruction methods which best responds to above criteria. The function of this algorithm is to be a step-by-step decision making process based on the factors summarized in Tables 4.1 and 4.2, with the initial step being illustrated in Table 4.1. This table suggests which model offers the most applicable methods for each case study. Then, the final decision should be made based on the available data and the advantages and disadvantage of each method summarized in Table 4.2.

Since available data and information can be different for Pre-R or Post-R periods, the first step in Table 4.1 is to define the appropriate time period (Pre-R or Post-R), after which it should be determined if the basin is gauged or ungauged. It is important to keep in mind that a gauged basin is referred to as a watershed where flow data is measured at least for short time period, making it possible to calibrate the flow reconstruction model with this data.

The two steps described in Table 4.1 are used in regard to the length and time interval of the output data (reconstructed flow), with consideration to the goal of this project.

In the work presented in this thesis, a short time step will be defined as one day or less, while time steps longer than a day will be defined as long time step. As well, a short term interval is selected if the reconstruction method is meant to produce a smaller data sample, while a long term interval is selected if a larger series of data over a longer period of time are to be reconstructed. For example, producing flow data for five days on a daily scale is considered short-term, but if the scale is hourly it would be considered long-term data reconstruction. The researcher then needs to analyse the input data and define their quality. There are different methods of assessing data quality but they are beyond the scope of this thesis. When data quality is considered unreliable, decisions about the reconstruction method become more critical. Therefore, validating input data before applying them to any flow reconstruction method is highly recommended. Finally, the applicable methods of flow reconstruction for each case are presented in the last column of Table 4.1. The suggested methods in each row of this column are selected based on capability and characteristics of each method presented in literature review found in Chapter 2. When more than one method exists for a given case, this indicates that options are available for the researcher, rather than implying that all the equations suggested are to be used. This second level selection will be done in the next step after considering available data and presented information in Table 4.2.

.

Table 4.1: Preliminary algorithm for determining the applicable methods of flow reconstruction for each case

Period	Basin	Output data time-step	Output data length	Input data quality	Applicable method
Pre-R	Gauged	Sh-ts**	Sh-t**	L-q	RR model
				H-q	RR model, Regression based method, Climate method
			L-t**	L-q	RR model
				H-q	RR model, Regression based method
		L-ts**	Sh-t	L-q	RR model, Regression based method, Climate method
				H-q	RR model, Regression based method, Climate method
			L-t	L-q	RR model, Regression based method, Climate method
				H-q	RR model, Regression based method, Climate method
	Ungauged	Sh-ts	Sh-t	L-q	- *
				H-q	Regression based method *
			L-t	L-q	- *
				H-q	Regression based method *
		L-ts	Sh-t	L-q	Regression based method *
				H-q	Regression based method *
			L-t	L-q	Regression based method *
				H-q	Regression based method *
Post-R	Gauged	Sh-ts	Sh-t	L-q	RR model, WBE
				H-q	RR model, Regression based method, Climate method, WBE
			L-t	L-q	RR model, WBE
				H-q	RR model, Regression based method, WBE
		L-ts	Sh-t	L-q	RR model, Regression based method, Climate method, WBE
				H-q	RR model, Regression based method, Climate method, WBE
			L-t	L-q	RR model, Regression based method, Climate method, WBE
				H-q	RR model, Regression based method, Climate method, WBE
	Ungauged	Sh-ts	Sh-t	L-q	WBE *
				H-q	Regression based method, WBE *
			L-t	L-q	WBE *
				H-q	Regression based method, WBE *
		L-ts	Sh-t	L-q	Regression based method, WBE *
				H-q	Regression based method, WBE *
			L-t	L-q	Regression based method, WBE *
				H-q	Regression based method, WBE *

* RR model could be used if regionalization is applicable and desired to define the parameters of model

**Sh-t = Short term, L-t = Long term, Sh-ts = Short time-step, L-ts = Long time-step

Table 4.2: Advantages and disadvantages of different groups of flow reconstruction methods

Method	Advantages	Disadvantages	Comments
WBE	• Accurate enough when all data are available for the interested basin or reservoir • Does not need to be calibrated • Simple and fast • Applicable for real-time data • Flexible enough to be applied to any case study	• Difficult to calculate water loss terms if they are not available • The uncertainty of one basin or reservoir highly affects downstream simulation	• If applied in a region where snow or evaporation are significant, these should be considered in the equation • Data validation is recommended before using the WBE if it is applied for short time step data reconstruction
RR model	• Can be used in time and space • One of the most reliable methods of flow reconstruction • Mostly have capability of considering snow, evaporation, infiltration, etc. • Could consider both hydrology and physics of flow • Applicable for real-time data	• Needs to be calibrated • It is sometimes very consuming to calibrate the model • Model parameters change from basin to basin • Requires a lot of data	• Increasing the number of parameters does not necessarily mean greater accuracy • Selecting the specific rainfall-runoff model depends on: ▪ Available data ▪ Climate of area (if evaporation or snowpack is important in the case) ▪ Land-use (urban or rural areas)
Climate model	• Climate signal data is usually available • Applicable for real-time data	• Needs to be calibrated • Model parameters change from basin to basin • Does not consider the physics of flow • Not always easy to find the climate signals which affects the flow • Needs to be updated over time with new measured data	• Climate signals needs to be downscaled
Regression based model	• Fast and simple • Few parameters need to be defined • Applicable when limited data are available • Applicable for real-time data	• Medium to low certainty • Needs to be updated over time with new measured data • Model parameters change from basin to basin • Uncertainty of model increases as the time-step decreases	• Selecting the specific regression based method depending on the available data

In the second step of the flow reconstruction method selection process, Table 4.2 is used to help in the decision of which method to use based on the strengths and weaknesses of the different methods, along with data availability. At times, data availability will be the main factor in selecting the method that will ultimately be chosen.

4.3 Applying the Proposed Algorithm to the Current Case Study

The proposed algorithm will be applied to the most appropriate family of flow reconstruction methods selected for this current project's case study. The project's main requirements on the method chosen for data reconstruction are the following:

- to be flexible enough to be applicable to all regulated rivers and basins in Quebec;
- to be applicable for daily time-step and long-term period for both Pre-R and Post-R periods;
- produces high-quality flow data;
- to be selected based on available data and their quality in Pre- and Post-R periods.

(Note: Both snowmelt and evaporation affect the results, especially in large reservoirs.)

Since there is a need to reconstruct the flow data series for both Pre-R and Post-R periods, the algorithm will be applied to define the appropriate method for each period separately.

4.3.1 Pre-R Period

Table 4.1 will be used to select the most appropriate flow data reconstruction method for Pre-R period.

The objectives of this project is to reconstruct short time step (daily) flow for long term flow for the ungauged Outardes 2 (pre-1977), Outardes 3 and 4 (pre-1969) basins. Data currently available for these basins are high quality hydrological data (such as rainfall, temperature, and snowfall) for the time period. Thus, Table 4.1 suggests the use of the Regression based method, according to the step by step process. Following up with Table 4.2, the specific regression based method should be selected in the next step after considering the reliability of the input data. For example, in the current case, a regression based method should be developed based on simulated flow using RR model, neighbouring basin's flow, rainfall, minimum temperature, maximum temperature, snowfall, and/or surface area.

4.3.2 Post-R Period

The selection of an appropriate family of flow data reconstruction method for Post-R period will be a two-steps process:

- Step 1: Define the list of appropriate methods of flow reconstruction

  According to Table 4.1., if we take into the consideration that the basins are ungauged and that the aim of the project is to reconstruct short time-step (daily) flow for long-term (from 1969 to present for Outardes 3 and 4, and from 1977 to present for Outardes 2):

  - ✓ regression based methods will be used when high quality data (rainfall, minimum temperature, maximum temperature, snowfall, surface area, and neighbouring basin's flow) are available, and
  - ✓ WBE based methods will be used if available storage volume, turbine flow, and discharged flow time-series are calculated based on non-validated data.

- Step 2: Evaluate the advantages and disadvantages of the different flow reconstruction methods by using Table 4.2:

The final method of choice for flow reconstruction method should be selected based on the advantages and disadvantages of each defined method in previous step. According to Table 4.2, regression-based methods have medium to low certainty which increases as the time step decreases. Thus, they are not the perfect choice to produce acceptable flow values when applied to the daily time-step required in the case study. On the other hand, Table 4.2 indicates that WBE is a flexible model that can be applied to any basin. This advantage could be important in the current case study, which requires a model that can be applied to all basins in Quebec. According to this table, WBE is also simple, fast, and does not require model calibration. This is another key advantage for the current project because no measured flow data are available to calibrate the model. However, WBE does presents an inconvenience in that it is necessary to calculate water loss and take evaporation and snowmelt into account as well (which can be dealt with by including these terms into the calculated flow). Moreover, Table 4.2 recommends that data validation be required for cases where input data are non-validated.

Therefore, the proposed algorithm recommends WBE as the most appropriate model for Post-R period in this case. Regression-based methods are also applicable, but are not recommended. To prove the efficiency of the algorithm, results generated by the WBE will be compared to regression-based methods during Post-R period.

4.4 Evaluating the Performance of Developed Algorithm

According to the proposed algorithm, WBE based methods are the most appropriate family of method for Post-R period in the current case-study. To evaluate the performance of the developed algorithm, the results of a WBE based method will be compared to the results of a regression based method calculation in an example basin called Outardes 4 for a Post-R period of a few years. The Classic WBE will be used as the simple common type of WBE based method, while Area Ratio will be selected as the regression based model according to the available data (as other regression based methods were tested and it was concluded that this method works better than the rest).

4.4.1 Classic WBE

Major watersheds in Quebec have their own reservoirs and currently, there is enough data available (storage volume, turbine flow from each reservoir, and discharged flow through the reservoirs' gates) for these reservoirs to allow WBE to be applied to each of them. WBE will be used to calculate flow for a reservoir (Equation 2.1), with water loss terms factored into the calculated flow.

4.4.2 Area Ratio Method

The Area Ratio method will be used to reconstruct the flow for Outardes 4 (Post-R period) because it is an example of a regression based method that can be developed independently from the measured flow data series. In this method, the flow data of each basin is related to the flow of a neighbouring basin, according to the ratio of their surface area. As explained in Chapter 2, the Moisie basin was chosen as the neighbouring basin in this case study because:

i) it is relatively close to the Outardes 4 basin and very likely to have similar characteristics to this catchment,

ii) its area is 19000 km^2, which is the approximate area of Outardes 4, increasing the likelihood that they share similar flow characteristics, and

iii) measured flow data series are available for the requested period of the case study.

4.4.3 Results

Comparison of the Area Ratio results and classic WBE methods for the time period between 2008 and 2012 are presented in Figure 4-2. To confirm the quality of the reconstructed flow values, the Nash–Sutcliffe efficiency coefficient *(NASH)* (Equation 4.1) and Absolute Volume Error (Equation 4.2) were used to compare the reconstructed flow to available filtered flow data series (flow which was mostly filtered manually and is currently the most reliable calculated flow values in all basins of Quebec) for the same time period (Table 4.3). Since the most reliable filtered flow data series is only available within the last few years, it cannot be used to calibrate long-term simulations and its applicability is limited to serving as a reference value series for recent years.

$$NASH = 1 - \frac{\sum_{i=1}^{N}(q_{fi} - q_{ri})^2}{\sum_{i=1}^{N}(q_{fi} - \overline{q_f})^2} \tag{4.1}$$

$$AVE = \frac{\sum_i | q_{fi} - q_{ri} |}{\sum_i q_{fi}} \tag{4.2}$$

where:

NASH	=	Nash–Sutcliffe efficiency coefficient,
AVE	=	absolute volume error,
q_{fi}	=	the reference filtered flow for day *i*,
q_{ri}	=	the reconstructed flow for day *i*,
$\overline{q_f}$	=	the average filtered flow when *i=1,…,N*, and N is the total number of days.

The Nash–Sutcliffe model efficiency coefficient is mostly applicable for high flow comparisons because it squares the difference and increases sensitivity to peak flows (Krause *et al.* 2005). However, in Equation 4.2, the influence of low flows and high flows are the same. Thus, both of the results can be considered as the QI. The *NASH* values vary between 1 and −∞ and the closer it is to one, the better it is. *AVE* also varies between 0 to +∞, and the closer it is to zero, the better it is.

Table 4.3: Quality index comparison of WBE and area ratio

Quality Index	WBE	Area ratio
Nash–Sutcliffe model efficiency coefficient	0.981	0.705
Absolute volume error	0.08	0.263

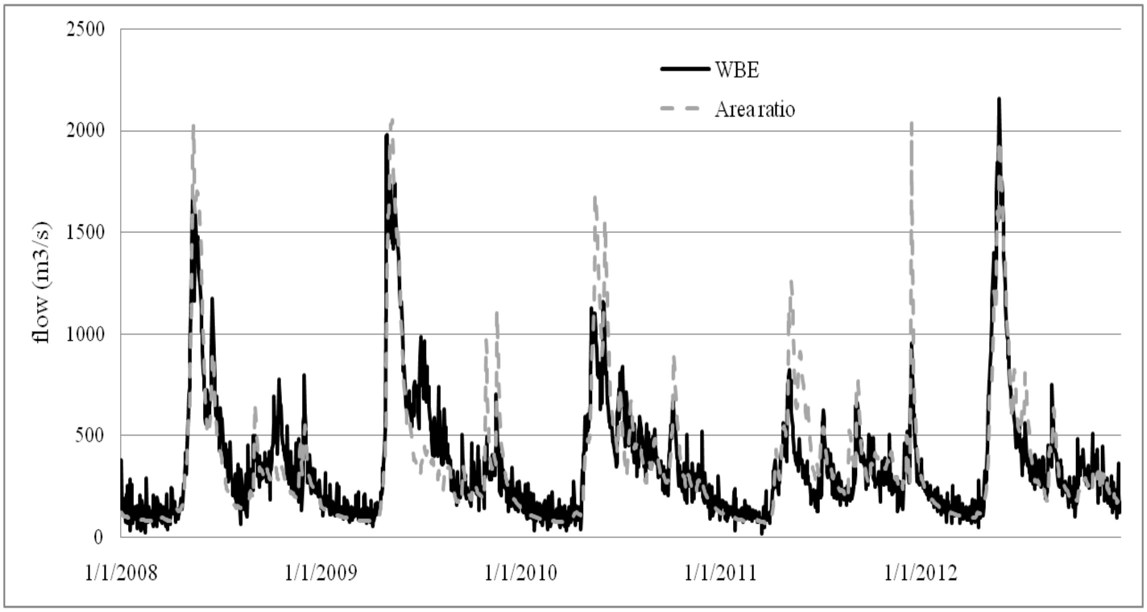


Figure 4-2: Comparison of reconstructed flow using WBE and area ratio

Comparing the results of WBE and Area Ratio methods with available filtered flow data series (filtered flow data series are not shown in Figure 4-2 to maintain confidentiality) shows that the Area Ratio method underestimated low flows and overestimated peaks, with its performance being insufficient for daily flow reconstruction. However, the over and under estimations may compensate one other and the method could prove to be a good estimation of seasonal or annual flow. On the other hand, results of the WBE method follow the same trend as filtered flow, but they are noisier, especially during low flow. This noise could be related to the uncertainty of input data. The calculated Quality Indexes for these two methods support the generated results through visual comparison. The Nash–Sutcliffe model efficiency coefficient is much better than that of WBE, indicating that this model is more successful in high flow estimations. Moreover, Absolute Volume Error is lower for the WBE, showing that this model provides greater likelihood of approximating the trend of flow data.

The greater reliability of WBE results confirms the efficiency of the proposed algorithm of method selection. The algorithm considers all aspect of flow reconstruction and clearly points out the advantages and disadvantages of different methods.

4.5 Conclusion

Knowing the flow values in each basin is important for water planning, management, and hydraulic designs. This information is also helpful for estimating water availability, designing flood-warning systems, and conducting studies based on historical flow data. Thus, flow data series should be reconstructed when previously measured data are not available. The first step of flow reconstruction is selecting the appropriate model to reach this goal.

In this chapter, an algorithm was developed to show how to select the appropriate family of flow reconstruction method in each case study. The presented algorithm will help researchers to select a family of models for each particular case with regard to different important factors such as the model's flexibility, requiring input data, output time step and uncertainty, climate, the length of the reconstruction period, and the advantages and disadvantages of each family of flow reconstruction methods.

The developed algorithm has been applied on the case studies in this thesis. The results showed that for our case study, regression based methods and WBE based methods are the most appropriate family of flow reconstruction methods for Pre-R and Post-R periods respectively (see also Figure 4-1). Lastly, the efficiency of the algorithm was tested by comparing the performance of defined appropriate method for Post-R period, WBE, with Area Ratio, a regression based method. Results confirmed the predominance of WBE to produce more likely flow values, and supported the efficiency of the algorithm.

According to the results of this chapter, a regression based method will be developed in Chapter 5 to reconstruct flow for Pre-R period, and a WBE based method will be developed in Chapter 6 to estimate flow data for Post-R period.

CHAPITRE 5 FLOW RECONSTRUCTION (PRE-RESERVOIR PERIOD, THEORY AND RESULTS)

5.1 Introduction

In the current project, the only available data for the Pre-R period are hydrological data (such as rainfall, temperature, and snowfall), neighbouring basins' flow, and simulated flow from RR models. Thus, only limited and simple methods could be used for flow reconstruction for this period. According to the results of Chapter 4, the most appropriate family of flow reconstruction methods for the Pre-R period is regression based methods. Now it is necessary to define the specific regression based method and find the answer to the question of how to reconstruct daily flow for Pre-R period. As shown in Figure 5-1, in this chapter, a new Kalman-based method will be developed as a tool to combine the available data and reconstruct the flow series for Pre-R periods (Figure 5-1, box # 3). The developed method is then compared to a few existing regression based methods (Area Ratio method, the Maintenance of Variance (Move) type III method, and the Multivariable Regression method; Figure 5-1, box # 2). Few QIs will be required to effectively evaluate the performance of the mentioned methods. However, traditional QIs cannot be used in our study because they compare the reconstructed flow with the measured flow data series (reference data), which are not available in the current case. Therefore, the question is how to evaluate the quality of reconstructed flow series for ungauged basins. In this chapter, in addition to visual comparison, some QIs that are applicable to ungauged basins are designed to evaluate the quality of flow data series for Pre-R period (Figure.5-1, box # 4).

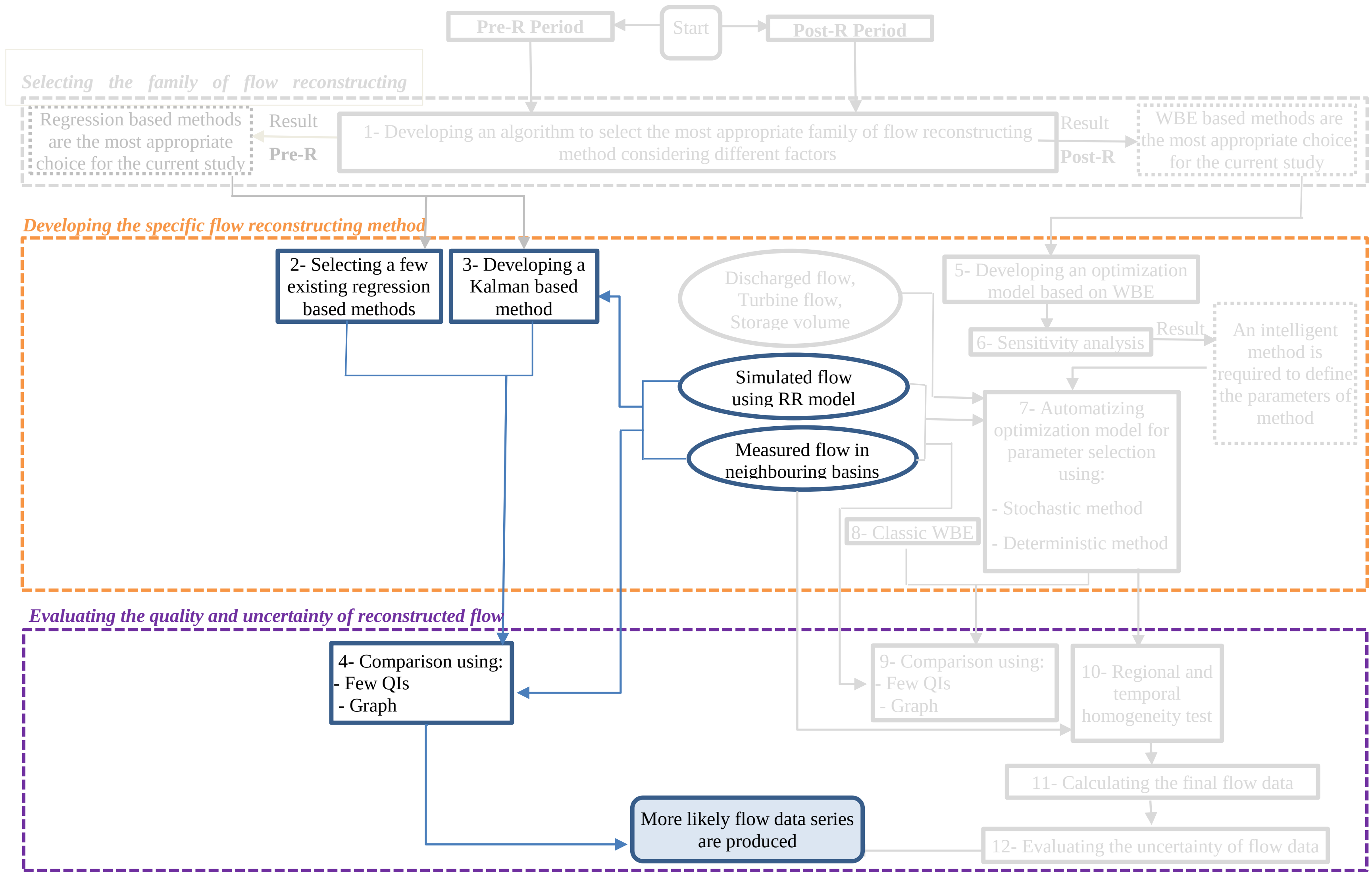


Figure 5-1: Schematic of methodology- Developing the specific flow reconstructing method and evaluating the quality and uncertainty of reconstructed flow for Pre-R period

5.2 Selecting A Few Specific Regression-Based Methods

A literature review on the different regression based methods was presented in Chapter 2, which showed that regression-based methods are fast, simple, and are mostly developed based on the data or information of neighbouring basins. For example, the flow series can be extended by weighting the observed stream flow at one or more neighbouring gauged basins. This weighted value may be the ratio of the catchment surface of the studied basin to that of the gauged basins (Hughes and Smakhtin 1996, Schreider et al. 1997). Although neighbouring basins may have different hydrological and meteorological characteristics, and thus dissimilar flow hydrographs, the Area Ratio method is one of the methods compared in the development of the Kalman-based method in this chapter.

Jones *et al.* (2004) developed a regression between the logarithms of river-flow, soil moisture and effective precipitation (which is precipitation minus actual evaporation), while Wen (2009) tried to reconstruct flow by relating discharge time series to rainfall and maximum temperature. In this project, however, no acceptable relation definable using a simple regression between a basin's flow and hydrological data (such as rainfall, temperature, and snowfall) could be found (the squared coefficient was always less than 10 percent).

A regression can also be developed between the short-term flow data series of the case study basin and the long-term flow data of a nearby basin (e.g. Hernandez-Henriquez et al. 2010, Dastorani *et al.* 2010). This method, however, does not guarantee that the flow series mean and variance will be preserved. Nevertheless, a multivariable linear regression will be developed and compared to the suggested Kalman-based method in this thesis. This multivariable regression relates the two independent variables of neighbouring basin's flow and simulated flow to the dependent variable of flow of the case study catchment (Multivariable Regression method).

Maintenance of Variance (Move) is another method used to extend flow series. This method preserves both mean and variance and has been tested by several researchers to extend the flow series (Hirsch 1982). Move III is also used to reconstruct the pre-reservoir flow data and is compared with the other methods (Move III method).

Considering that a neighbouring basins' flow and simulated flow (using RR model) are two main available data series for flow reconstruction during Pre-R period, they can be combined to give a possible set of flow values for the case study. The Kalman filter will be the method used to filter and combine data when more than one time interval series are available.

There are two types of Kalman filters, based on multi-sensor data fusion[7] (Zhou *et al.* 2010). The first one, Measurement Fusion methods, involve merging multi-sensor data or combining them based on minimum mean square error estimates, compounding the measurements, and then filtering the data series using the Kalman method. The second one, State Vector fusion, first filters the data series using a Kalman filter and then compounds them. State estimate covariance matrices are used for State Vector fusion; however, the State estimates from different estimators are usually dependent, and to account for this, a new technique of fusing the filtered data series is introduced in this thesis by combining flow data from a neighbouring basin to flow data from an RR model. This new method was established to reconstruct the flow data for the Pre-R period and its performance was compared to results from Area Ratio, Multivariate regression, and Move III methods.

Table 5.1 summarizes the list of selected regression based methods and the reason for their selection. There are a limited number of applicable methods for Pre-R period and the listed methods considered the most common. The Area Ratio is a simple and fast method that can be used, considering the available data for the current case study. This method is developed based on one available data series (flow from neighbouring basin). However, Multivariable regression model uses flow from both neighbouring basin and the RR model as the input data and benefit from both of them. This is a common, simple method that is applicable according to the limited available data for the Pre-R period. The mean and variance of the results of this method may change in time. Move III also appeared on the list because it can preserve mean and variance. However, Move III can only be developed based on the flow from neighbouring basin or from rainfall-runoff model. These comparisons show that the three mentioned methods complement each other.

The Kalman filter based method has been suggested for flow reconstruction during pre-reservoir construction period. In fact, this method is based on an optimization technique that combines the flow from neighbouring basin and from RR model filtered using Kalman technique.

---

[7] Combination of two or more measured data series of *n* values

Table 5.1: The list of applied methods for flow reconstruction during the Pre-R period

	Method	The reason of choosing method
Pre-R period	Area ratio	- simple and fast - widely used method - applicable based on available data - being used in Quebec
	Multivariable regression	- simple and fast - widely used method - applicable based on available data - benefits from both RR model and neighbouring basin's flow
	Move III	- simple and fast - widely used method - applicable based on available data - preserves mean and variance
	Kalman based method	- applicable based on available data - benefits from both RR model and neighbouring basin's flow - finds flow data, combining simulated flow and neighbouring basins' flow, using an optimization technique

5.3 Methodology

Since the available data for Pre-R period is limited to flow from an RR model and from neighbouring basins, using a simple regression based methods is unavoidable. For the work presented in this thesis, a new Kalman-based method was developed to reconstruct the flow data series using the State Vector Fusion technique. The results of this method are compared with those of selected regression based methods (Area Ratio, Multiple Regression, and Move III).

5.3.1 Area Ratio Method

In the Area Ratio method, the neighbouring basin's flow is the only flow data that is used to reconstruct flow data. In this method, presented by Equation 5.1, the flow of neighbouring basin is multiplied by the ratio of the case study basin's surface to the neighbouring basin's surface.

$$F_i = \frac{s_{nb}}{s_{case\ study}} \times nbf_i \qquad (5.1)$$

where:

s_{nb}	=	surface area of the neighbour basin
$s_{case\ study}$	=	surface area of the case study basin
nbf_i	=	flow of neighbouring basin for the i^{th} day.

5.3.2 Multivariable Regression Method

One of the flow extending methods for Pre-R periods is a Multivariable Regression method (Equation 5.2). In this method, a linear regression is developed between the flow data as a dependent variable and the logarithmically scaled flow of a neighbouring basin and the logarithmically scaled simulated flow (by RR model) as the independent variables. Here, the neighbouring basin flow is scaled using the surface area ratio.

$$NF_i = a \times nbf_i + b \times f_{RR} + e \qquad (5.2)$$

where:

f_{RR}	=	simulated flow using RR model
a and b	=	coefficients of the model
e	=	regression equation constant

This equation was developed based on the reconstructed flow data series during the Post-R period, which will be described in Chapter 6. This means that the calibration period for this method is selected from years following the Post-R period.

5.3.3 Move III

Move III (Equation 5.3) is a linear regression equation based on a specific method for calculating the slope and constant value of regression to extend a data series. The logarithmic values are used to develop this model for the Outardes basin.

$$\widehat{Y_\imath} = \acute{a} + \acute{b}(\, x_i - \overline{x_2}) \tag{5.3}$$

with:

$$\acute{a} = \frac{(n_1+n_2)\widehat{\mu_y} - n_1\overline{y_1}}{n_2} \tag{5.4}$$

$$\acute{b}^2 = \left[(n_1 + n_2 - 1)\hat{\sigma}^2{}_y - (n_1 - 1)S_{y1}^2 - n_1 \; - n_2\left(\acute{a} - \widehat{\mu_y}\right)^2\right] \times [(n_2 - 1)S_{x2}^2]^{-1} \tag{5.5}$$

$$\overline{y_1} = \frac{1}{n_1}\sum_{i=1}^{n1} y_i \tag{5.6}$$

$$S_{y1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n1}(y_i - \overline{y_1}) \tag{5.7}$$

$$S_{x2}^2 = \frac{1}{n_2 - 1}\sum_{i=n1+1}^{n1+n2}(x_i - \overline{x_1})^2 \tag{5.9}$$

$$\overline{x_2} = \frac{1}{n_2}\sum_{i=n1+1}^{n1+n2} x_i \tag{5.8}$$

$$\overline{x_1} = \frac{1}{n_2}\sum_{i=1}^{n1} x_i \tag{5.10}$$

where:

x_i	=	the available flow data for day *i*
n_l	=	the length of the short term data record
$n_l + n_2$ =		the time length of the available long term data
n_2 =		the length of reconstructed flow
$\widehat{Y_\imath}$	=	the reconstructed flow for day *i*,

$\widehat{\mu_y}$ = unbiased mean estimator of the complete extended record,

$\hat{\sigma}_y$ = unbiased variance estimator

(Vogel and Stedinger, 1985).

5.3.4 Kalman Filter-Based Method

In the developed State Vector fusion Kalman method, the measured data series are first filtered using a Kalman filter method (Equations 5.11 and 5.12). The filtered data series are then combined (Equation 5.13) using an optimization model. In the work presented in this thesis, the SSM Matlab toolbox is used to filter the flow series.

Flow filtering:

$$y_1 = Kalman(f_{RR}) \tag{5.11}$$

$$y_2 = Kalman(nbf_s) \tag{5.12}$$

where:

f_{RR} = simulated flow using the RR model

nbf_s = scaled flow from the neighbouring basin (neighbouring basin's flow multiplied by area ratio),

y_1 and y_2 = filtered flow using the Kalman filter method.

Filtered flow combining:

$$Minimize \quad (B_1^2 + B_2^2) \tag{5.13}$$

Subject to:

$$y_1 = B_1 + \alpha_1 \times Y \tag{5.14}$$

$$y_2 = B_2 + \alpha_2 \times Y \tag{5.15}$$

$$0 \le Y < \infty \tag{5.16}$$

$$[\alpha_1, \alpha_2] \in \{[1,1],[1,2],[1.2,1.8],..,[2,1]\} \tag{5.17}$$

where:

Y	=	reconstructed flow
α_1 and α_2	=	coefficients of reconstructed flow
B_1 and B_2	=	the errors values.

Equation 5.14 implies that the reconstructed flow is a multiple of filtered simulated flow (y_1) plus an error value, and Equation 5.15 shows that the reconstructed flow is a multiple of filtered Neighbouring basin's flow (y_2) plus an error value. Equation 5.16 forces the flow values to stay positive, and Equation 5.17 shows a set of coefficients of α_1 and α_2. For example, the sub-set of $[1,2]$ indicates that α_1 =1 and α_2 =2. This model tries to optimize the reconstructed flow (*Y*) by taking the most appropriate ratio of available filtered flow data series (y_1 and y_2) and minimizing the errors (B_1 and B_2).

5.3.5 Evaluating the Quality of Reconstructed Flow

An important topic that has not been discussed enough in regards to ungauged basins is the evaluation of reconstructed flow. Most of the studies have applied traditional QIs to compare the validated flow with measured flow. Even in Quebec, where measured flow data series are not readily available, reconstructed flow has been compared with filtered flow data (the flow series that mostly was filtered manually and is available in Hydro-Quebec's database) using one or two indices. However, as explained in Chapter 2, filtered flow values in Quebec are less reliable prior to 2005, and as a result, designing new Quality Indexes to evaluate the integrity of validated flow that are independent from filtered flow would be beneficial.

Since there is no proper statistical quality criterion for hydrological simulation models, more than one QI are usually used to assess more precisely the performance of the model (Weglarczyk, 1998). Therefore, it is preferable to use several QIs to evaluate the reliability of reconstructed flow data series in ungauged basins. The three drafted QIs for the Pre-R period are Normalized Nash (*NN*), Consistency Coefficient (*CC*), and Normalized Tortuosity (*NT*), as shown below by Equations 5.18, 5.20, and 5.21 respectively.

1) $$NN = 1 - \frac{(1 - NASH)}{\sqrt{(\mathrm{a} + (1 - NASH)^2)}} \tag{5.18}$$

$$NASH = \frac{\sum_{i=1}^{N}(q_{obsi} - q_{cali})^2}{\sum_{i=1}^{N}(q_{obsi} - \overline{q_{obs}})^2} \tag{5.19}$$

2) $$CC = \frac{\sum_i n_i}{N} \tag{5.20}$$

$$\begin{cases} a_i' \times b_i' > 0 \rightarrow n_i = 1 \\ a_i' \times b_i' < 0 \rightarrow n_i = 0 \end{cases} \tag{5.21}$$

$$a_i' = nbf_{i+1} - nbf_i \tag{5.22}$$

$$b_i' = q_{cali+1} - q_{cali} \tag{5.23}$$

3) $$NT = \frac{1}{1 + (\frac{T}{N})} \tag{5.24}$$

$$T = \sum | q_{cali+1} - q_{cali} | \tag{5.25}$$

where:

a	= a constant number between 0 and ∞
q_{obsi}	= the observed flow in the case-study basin in day *i*,
q_{cali}	= the reconstructed flow by developed model for day *i*,
nbf_i	= the flow of i^{th} day in the neighbouring basin,

a'_i and b'_i	= the differences of discharge between the days *i+1* and *i* for the selected neighbouring basin and calculated flows respectively
n_i	= marks the status of day number i , it is equal to one if the trend of both calculated flow (q_{cal}) and neighbouring basin's flow (nbf) is increasing or decreasing,
N	= the total number of days, and
T	= tortuosity.

(Note: simulated flow using an RR model is considered as the observed flow (q_{obsi}) to calculate *NASH*)

The QIs are all normalized by scaling them between 0 and 1. The QI values which are closest to 1 indicate the better QI.

As stated in Chapter 1, flow data series are used for different purposes such as flow prediction, flood analysis, water resource management, and flow simulation. In this project, the suggested indices are designed to fulfill the above mentioned purposes.

- *NN*: The flow data series that disregard the physics of flow and the climate of the area are considered unreliable. According to Equations 5.18 and 5.19, *NN* factor in these terms by calculating *NASH* based on the simulated flow through the hydrologic RR model (see also Figure 5-1, box # 4). Meteorological data are used as input data for RR models and thus they reflect the flow physics in flow simulation. Thus, flow physics and meteorological factors are indirectly used in *NN*. This implies that the reconstructed flow data series that best respect flow physics and the coherence between climate and flow have higher *NN* values. Such reconstructed flow data series are more reliable for calibrating hydrological models. As the difference between q_{obsi} and q_{cali} is squared in Equation 5.19, *NN* is more sensitive to peak flows. Therefore, *NN* is also a beneficial QI for flood prediction and PMF estimation.

- *CC*: Dissimilarity in variation of the reconstructed flow to that of the neighbouring basin's flow affects regional flow analysis. Therefore, *CC* is designed to compare the similarity of the trend of reconstructed flow with that of neighbouring basin's flow (Figure 5-1, box # 4) and penalizes the reconstructed flow series that do not respect regional integrity. As explained in Chapter 2, the selection of a neighbouring basin (among other measured basins in the area) for comparison with the case study watershed is based on similar physical characteristics; therefore, it is expected that the neighbouring basin shares almost the same hydrographical and flow variation with the case study basin. Any dissimilarity in variation between a neighbouring basin's flow and simulated flow causes lower *CC*.
- *NT*: Noise disturbs short term and long term flow memory and makes it difficult to analyze flow data series and thus to efficiently predict flow values. Unreliable predicted flow values will evidently lead to uncertainty in water resource management and inconveniences in flood situations. Therefore, this QI is designed to penalize noisy data series and to provide a better idea on the safety levels of predicted flow.

The names of these criteria and the reason for choosing them are listed in Table 5.2.

Table 5.2: The list of selected quality indexes for Pre-R period

	QI	**The reason of choosing QI**
Pre-R period	*CC*	- penalizes the flow data series not respecting the regional homogeneity in the sense of hydrograph shape - beneficial for regional flow analysis studies
	NN	- penalizes the flow data series not respecting meteorological factors - more sensitive to peak flows - beneficial for flood prediction and PMF estimation, and RR model calibration purposes
	NT	- penalizes noisy flow data series - useful for water management and flow prediction studies

5.4 Results

The suggested Kalman-based method was used to calculate flow on the Outardes 4, Outardes 3, and Outardes 2 sub-basins and the results were compared with the reconstructed flow using existing methods of flow reconstruction: Area Ratio method, Maintenance of Variance (Move III) method, and Multivariable Regression method. In these methods:

- The Moisie River has been selected as the neighbouring basin for Outardes 4 and the Godbout has been chosen as the neighbouring basin for Outardes 3 and Outardes 2 (Table 3.3).
- Simulated flow and neighbouring basin's flow are considered as independent variables of the Move III method.
- The calibration and validation periods of Multivariable Regression and Move III methods are 1979-2011 and 1960-1978 respectively.

The results derived from the application of these methods are presented for several example years for Outardes 4 in Figures 5-2 to 5-6. Figures 5-4 to 5-6 show the results of the same models for the years when the results of classic WBE are available. These figures provide helpful information on the performance of the different models.

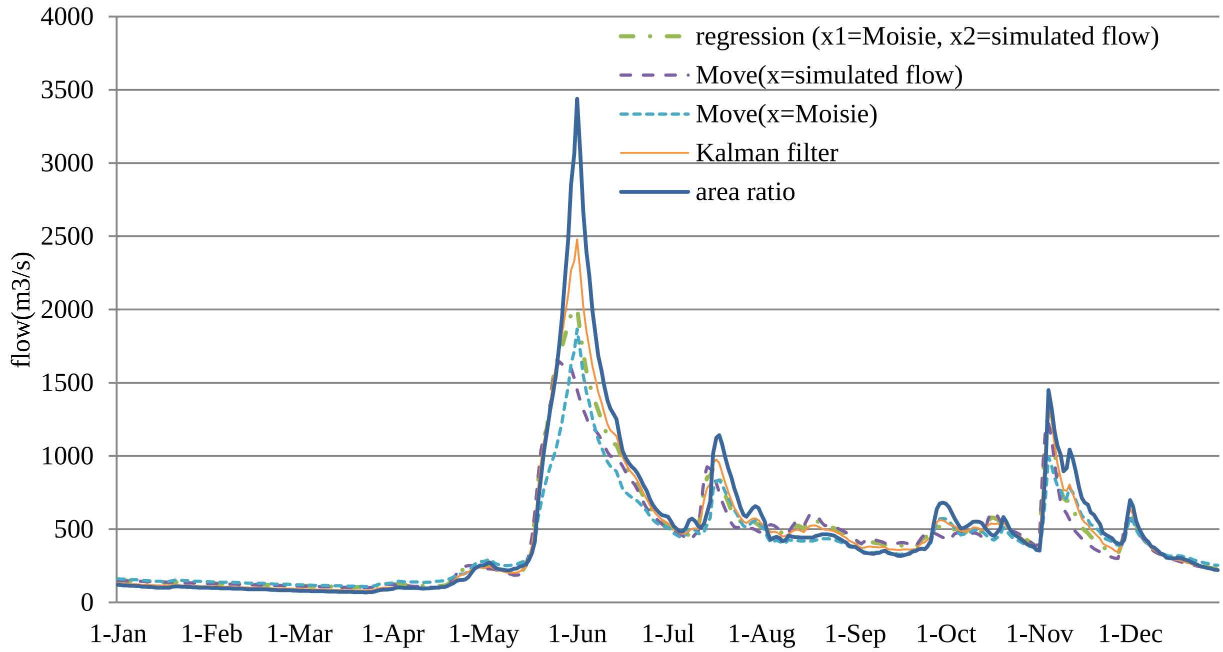


Figure 5-2: Comparison of different methods of flow reconstruction for Outardes 4 (1966)

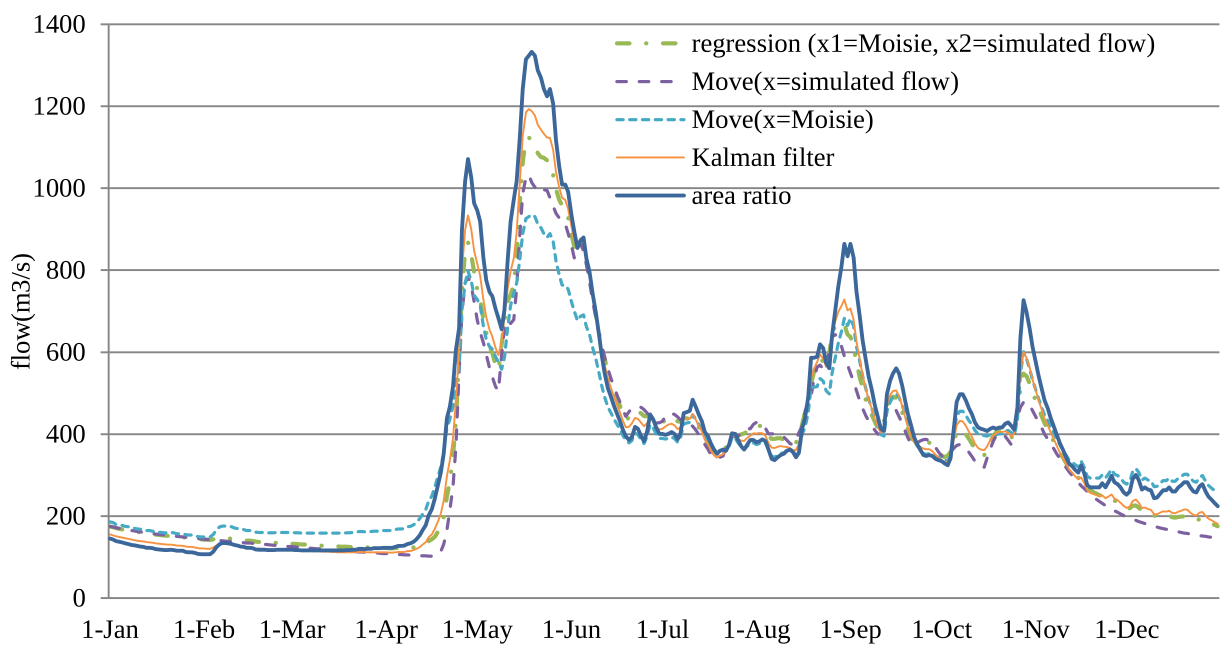


Figure 5-3: Comparison of different methods of flow reconstruction for Outardes 4 (1968)

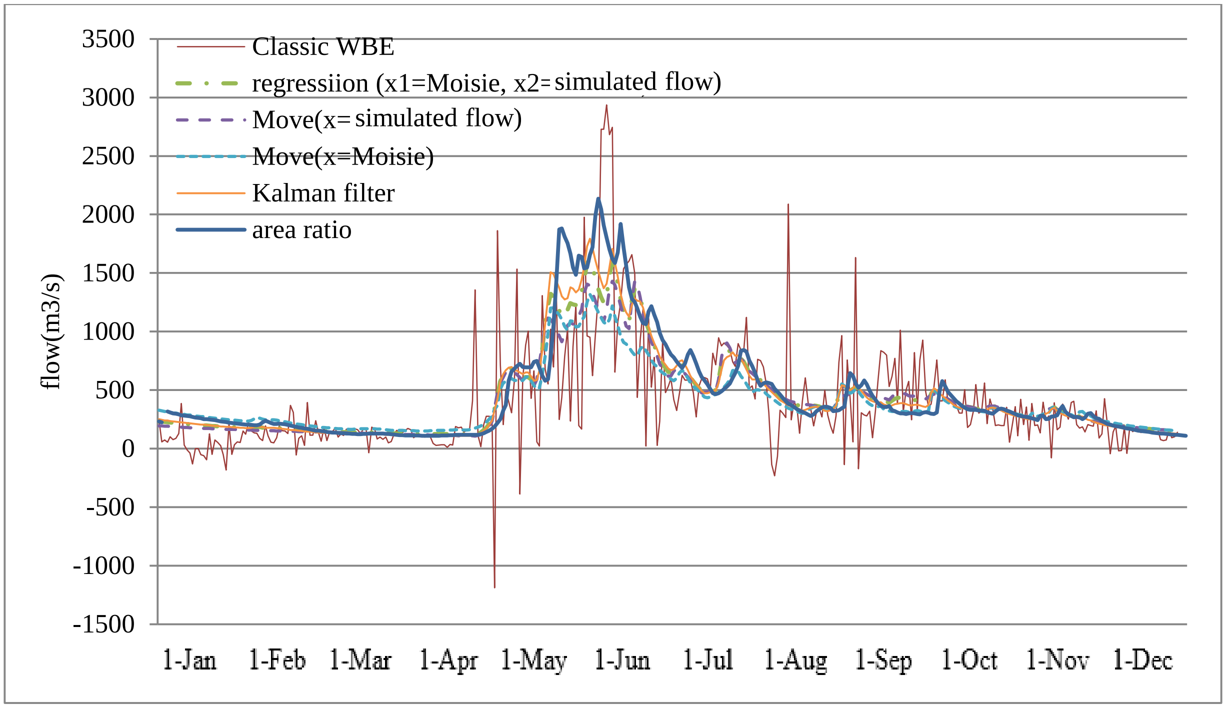


Figure 5-4: Comparison of different methods of flow reconstruction for Outardes 4 (1970)

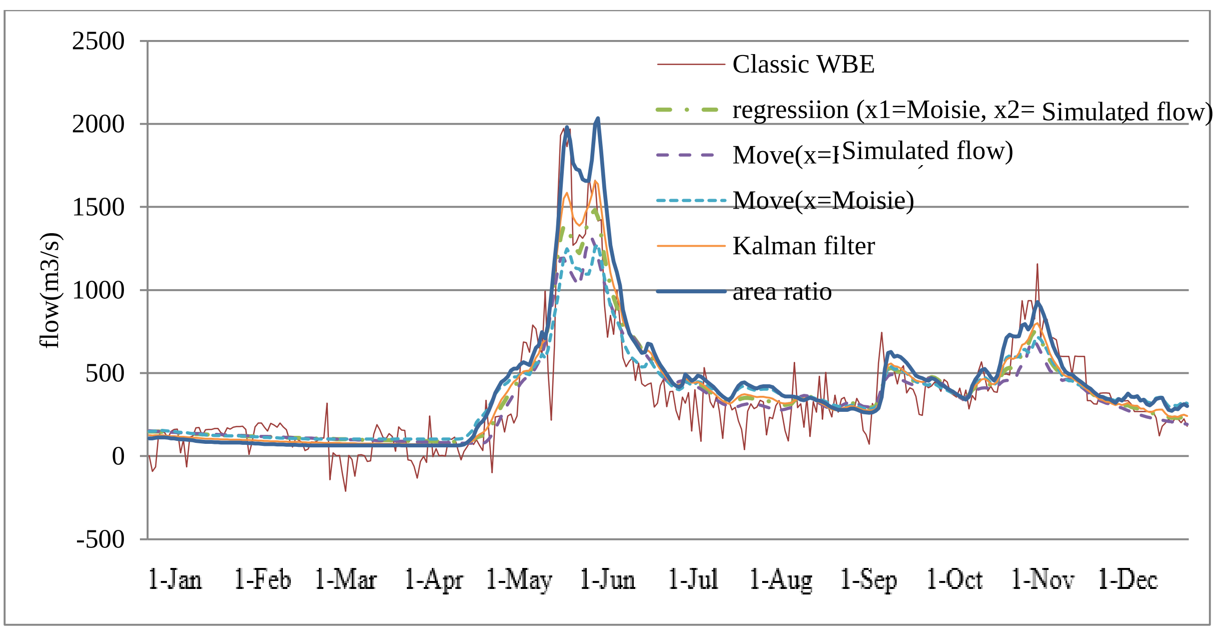


Figure 5-5: Comparison of different methods of flow reconstruction for Outardes 4 (1971)

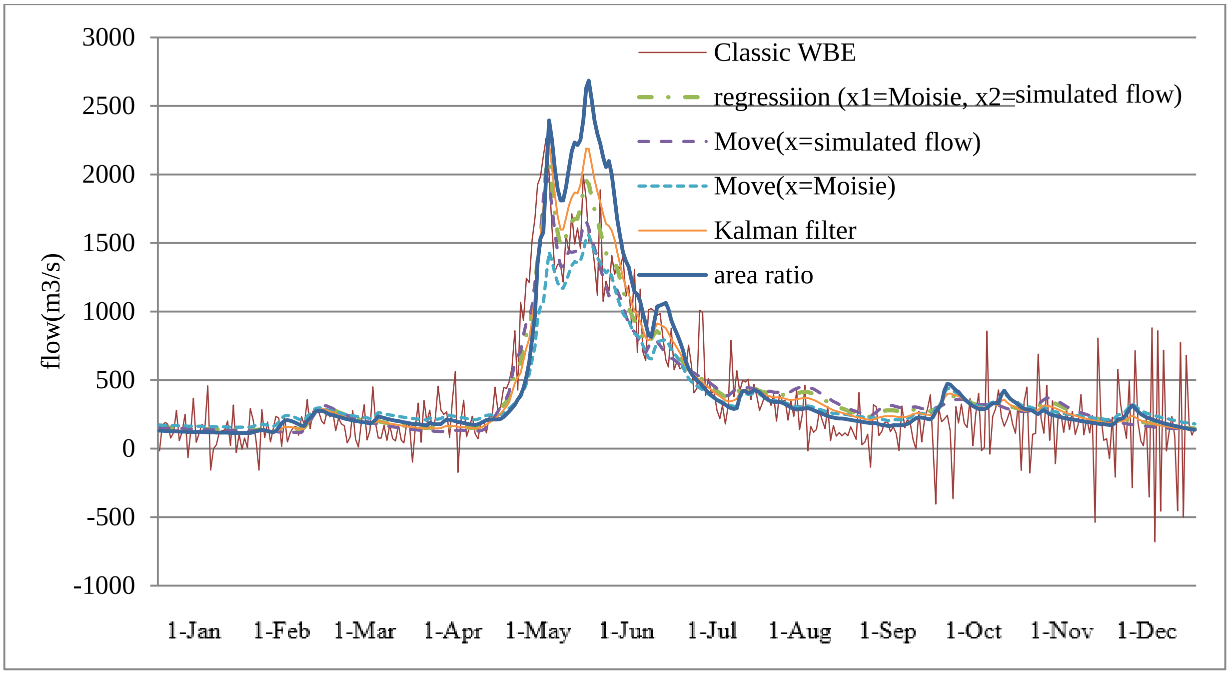


Figure 5-6: Comparison of different methods of flow reconstruction for Outardes 4 (1981)

The quality of reconstructed flows was assessed using three QIs: Normalized Nash (*NN*), Normalized Tortuosity (*NT*), and Consistency Coefficient (*CC*). To have more efficient quality evaluation, *CC* was calculated using measured flow at different neighbouring basins from those used in the process of flow reconstruction. These secondary neighbouring basins are the Romaine and Outardes rivers for Outardes 4, and the Moisie flow for Outardes 3 and Outardes 2 (Table 3.3).

The QIs calculated based on different methods of flow reconstruction for Outardes 4, Outardes 3, and Outardes 2 during the Pre-R period are presented in Tables 5.3, 5.4, and 5.5 respectively. In order to compare the performance of these methods for Pre-R period with the Post-R Period, the quality of reconstructed flow for Post-R period was reviewed using these three QIs plus Normalized Absolute Volume Error (*NAVE*) and Specific Flow Ratio (*SFR*) in Table 5.3. Chapter 6 will provide more information on *NAVE* and *SFR*.

From Figures 5-2 to 5-6 and Tables 5.3 to 5.5 (the best results are indicated on the Tables in **bold letters**), the conclusion reached was that each method has particular advantages and performs dissimilarly for different basins and years. For example:

- For Outardes 4, Kalman had the highest *CC* during the Pre-R period. However, for Post-R years, the Area Ratio method had the best *CC*. For Outardes 3 and Outardes 2, Move-III (x = neighbouring basin) and Area Ratio have respectively the highest *CC*. This means that the ability of the different methods to produce regionally coherent flow data is different for variations in time and space.

- For Outardes 4, Move III (x=simulated flow) generally produced the best individual QIs, and the best average QI for Pre- and Post-R periods. This can be also seen in Figures 5-4 to 5-6, which compare the different methods to classic WBE (although the results of classic WBE are not very reliable, they can give an idea about the general shape of hydrograph). In these figures, the reconstructed flow using Move III (x=simulated flow) had the most similar trend to classic WBE, which could be related to the quality of simulated flow data for this basin.

- Move III (x=simulated flow) produced the best *NT* values for Outardes 4, while Area Ratio produced the best *NT* values for Outardes 3, and the Kalman method for Outardes 2. This implied that smoothness of the model values are different in the three case studies.

- The best *NN* values are associated to Move (x=simulated flow) for Outardes 4, and to Multiple Regression for Outardes 3 and 2. Thus, these two methods are more adept to reflect meteorological data in comparison to other methods.

- In Figures 5-4 to 5-6, the Area Ratio method overestimates the flow in comparison with WBE. This is why this method has the highest Specific Flow (*SF*) and the lowest *NAVE* for Outardes 4 and Outardes 2 in comparison to other methods. Unlike these two basins, the highest SF is related to Move III for Outardes 3.

- The inverse relationship between *SF* and *NN* for Outardes 3 and 2 shows that the methods with higher SF (Area Ratio and Move III) are overestimating flow comparing to simulated flow. Hence, they possess low *NN*.

- For Outardes 4, all methods produce more or less the same SF. However, the range of generated SF was wider for Outardes 3 and Outardes 2.

  Comparing the SF of Pre-R and Post-R periods (Figure 7-25) shows that for Outardes 4, the four methods used for flow reconstruction for the Pre-R (Area Ratio, Multiple

Regression method, Move III, and Kalman filter based method) result in higher SF compared to the methods applied for Post-R. However, for Outardes 2, the Area Ratio and Move III (x=NB) methods shows higher SF, while for Outardes 3, the Move III method produces a higher SF than other methods. For both Outardes 3 and Outardes 2, the Kalman and Regression methods underestimate the flow in comparison to the reconstructed flow series during Post-R period.

Figure 7-25 shows that for the three sub-basins, Move III (from Pre-R period) has the most similar *SF* to the suggested flow reconstruction methods for Post-R period. This shows that Move III is probably the most reliable method for Pre-R period in this case-study. However, SF is only one of the criteria. Selecting the most appropriate method in each case-study depends on the engineering decision of researcher.

Table 5.3: Comparison of QIs for a few methods of flow reconstruction (Outardes 4)

Flow reconstruction Methods	Reservoir Name	NAVE	NT	SFR	NN	CC	Aérage QI	SF (L/s/km^2)
Kalman	Pre-RC -Outardes river		0.850		0.744	0.564	0.719	24.1
	Pre-RC -Romaine river		0.850		0.744	**0.677**	0.757	
	Post-RC -Romaine river	0.699	0.864	0.941	0.802	0.741	0.809	23
Area Ratio	Pre-RC -Outardes river		0.808		0.588	0.529	0.642	25.6
	Pre-RC -Romaine river		0.808		0.588	0.630	0.675	
	Post-RC -Romaine river	0.677	0.82	0.87	0.657	**0.997**	0.804	24.2
Regression	Pre-RC -Outardes river		0.868		0.798	0.573	0.746	23.8
	Pre-RC -Romaine River		0.867		0.650	0.592	0.703	
	Post-RC -Romaine River	**0.702**	0.876	0.925	0.850	0.791	0.829	23.0
Move III x= simulated flow	Pre-RC -Outardes river		**0.876**		**0.842**	**0.581**	**0.766**	22.6
	Pre-RC -Romaine river		**0.876**		**0.842**	0.622	**0.780**	
	Post-RC -Romaine river	0.7	**0.886**	0.97	**0.885**	0.704	0.829	21.9
Move III x=Moisie	Pre-RC -Outardes river		0.875		0.803	0.529	0.736	22.8
	Pre-RC -Romaine river		0.875		0.803	0.630	0.769	
	Post-RC -Romaine river	0.684	0.883	**0.986**	0.859	**0.997**	**0.882**	21.4

Note: (1) Pre-RC =Pre-Reservoir Construction = 1960 to1969

(2) Post-RC = Post-Reservoir Construction=1969 to 1978

(3) SF = Specific Flow

Table 5.4: Comparison of QIs for a few methods of flow reconstruction during Pre-R period (Outardes 3)

Flow reconstruction Methods	NT	NN	CC	Average QI	SF (L/s/km2)
Kalman	0.995	0.795	0.64	**0.81**	16.676
Area ratio	**0.997**	0.299	0.693	0.663	22.758
Regression	0.995	**0.802**	0.602	0.766	18.101
Move-x= simulated flow	0.985	0.422	0.502	0.64	28.080
Move-x=neighbouring basin	0.988	0.19	**0.914**	0.584	30.055

Table 5.5: Comparison of QIs for a few methods of flow reconstruction during Pre-R period (Outardes 2)

Flow reconstruction Methods	NT	NN	CC	Average QI	SF (L/s/km2)
Kalman	**0.988**	0.832	0.652	0.824	11.710
Area ratio	0.983	0.357	**0.697**	0.679	23.351
Regression	0.984	**0.988**	0.516	**0.829**	14.592
Move-x=simulated flow	0.981	0.501	0.599	0.694	16.980
Move-x= neighbouring basin	0.98	0.352	0.618	0.649	20.647

5.5 Conclusion

In this chapter, a Kalman filter based method was developed to reconstruct the flow for Pre-R years. This suggested methodology was applied to the Outardes basin and the results of that were compared with the following regression based methods: Move III, Area Ratio, and Multivariable regression using visual graphs and three different QIs. Although all the methods produced smooth and non-negative flows, they exhibited different performances in time and space. Looking at their QIs and visual comparisons, it is evident that:

- Kalman method produces a set of smooth data series (good *NT*) which is close to that of simulated flow (good *NN*). However, the results of this method are not the best compared to regional flow (moderate *CC*).
- Area Ratio method produces a smooth set of data series (good *NT*) but shows the most deviation from simulated flow (low *NN*). This method produced the best results with respect to regional flows at Outardes 2 and Outardes 4 but not at Outardes 3.
- Regression method produces a smooth set of data series (good *NT*) which most closely matches to that of simulated flow (good *NN*). The results of this method are not the best, though, with respect to regional flows in Outardes 3 and Outardes 2.
- For Outardes 3 and 2, Move III (x= simulated flow) produces a smooth set of data series (good *NT*) but does not match the data series from simulated flow (low *NN*). The results of this method are not the best also for regional flows. This method performs the best for Outardes 4.
- Move III (x= Neighbouring basin's flow) produces a smooth set of data series (good *NT*). The results of this method are comparable to regional flow (good *CC*). However, it had a very weak performance in calculating flow values close to simulated flow (low *NN*) in Outardes 3 and 2.

The general conclusion is that the results of different methods vary with different time and space periods. Selecting one method for each time period is not easy because each method has its advantages in producing flow data series. In the case that all the QIs of different methods (including SF) are very close and do not give a clue to select the most appropriate method, calculating the final reconstructed flow based on weighted flow[8] for each time period is suggested as a solution to the question of <u>how to reconstruct daily flow for Pre-R period</u>. However, in Outardes 3 and 2,

[8] The weight of each method could be defined based on the average QI of that method. In this approach, the flow with higher QI has larger weight in the final flow data series.

Move III (x= simulated flow) is the most reliable flow reconstruction method because it is the only method that produces the flow time series with the SF close to Post-R period methods. In Outardes 4, Move III (x= simulated flow) is also the most reliable flow reconstruction method because it mostly has the best QIs (including SF).

Also, the visual graphs and three QIs are possible answers the question of how to evaluate the quality of reconstructed flow series in ungauged basins for Pre-R period.

In Chapter 6, a methodology for reconstructing the flow values and evaluating the quality of estimated flow for Post-R period will be presented.

CHAPITRE 6 FLOW RECONSTRUCTION (POST-RESERVOIR PERIOD, THEORY)

6.1 Introduction

In Chapter 4, WBE-based methods were suggested as the appropriate family for flow reconstruction method for Post-R in this current research. Using classic WBE for a reservoir as a closed hydraulic system is a common method (e.g. Shiau and Lee 2005) for flow reconstruction in ungauged basins. Many researchers have attempted to estimate the water-loss parameters, which are fundamentally part of the WBE, such as infiltration (Joshi and Tambe 2010, Telis 2001) and evaporation (i.e. Yeung 2005, Gunter *et al.* 2004, Hamon 1961, Sivapragasam *et al.* 2009, and Parasuraman *et al.* 2007). However, it remained a source of uncertainty. Therefore, WBE can be simplified as Equation i.1 (classic WBE) by including water-loss parameters when calculating flow values. This simplification, along with input data uncertainty, causes noisy and even negative values of WBE calculated flow. Thus, more investigation is required on WBE-based methods in order to improve their results.

In light of the literature review presented in Chapter 2, different alternative WBE based methods have been recommended to reconstruct flow in ungauged basins in Quebec. This critical review shows that each of developed methods has its own strengths and weaknesses, and can improve the results of previous research as well. For example, the current method of flow reconstruction currently used at Hydro-Quebec has the potential of being improved by applying a WBE based optimization model, POM, as an alternative. POM improves the limitations and problems of a mostly manual filtering method (listed in Chapter 2). It is flexible enough to be applied to any reservoir where storage volume, turbine flow, and discharged flow are measured. Results of POM exhibit non-negative flow values and decreased noise. Nevertheless, this method still has some limitations (as listed in Chapter 2) that restricts its applicability and efficiency and leaves the question of how to reconstruct more likely daily flow for Post-R period unanswered. Thus, an optimization model will be developed in this chapter based on POM (Figure 6-1, box # 5) in order to address the problems stated previously. Next, a sensitivity analysis will be performed to evaluate the improved POM under the assumption that both time and space are considered as constant parameters (Figure 6-1, box # 6). Then, both automatic deterministic and stochastic techniques will be developed to intelligently estimate the parameters of suggested model (Figure 6-1, box # 7).

After, flow values have been reconstructed, it is essential to evaluate the quality of the flow values obtained. In Chapter 5, three QIs were suggested for reconstructed flow values in the Pre-R period. As more information is available for Post-R, it is desirable to develop more QIs in order to more comprehensively assess the flow data series. In current chapter, five QIs will be designed (Figure 6-1, box # 9), in addition to visual graphs, to compare the results from the developed Stochastic based models, Deterministic based models, and classic WBE (Figure 6-1, box # 8). Also, regional and temporal homogeneity of the reconstructed flow values will be evaluated in this chapter (Figure 6-1, box # 10). This part of the research is used to answer to the question of how to evaluate the quality of reconstructed flow series in ungauged basins for Post-R period.

Lastly, the final flow will be calculated using a Weighted Average method (Figure 6-1, box # 11). Evaluating uncertainty is an indispensable step in regards to flow reconstruction. There are many different methods available to evaluate different types of uncertainty, but most existing methods are dependent on previously measured data. Therefore, it is still a challenge to find out how to evaluate the uncertainty of flow data in ungauged basins. In this study, a methodology is suggested to answer this question and to estimate the probable range of the final flow data series.

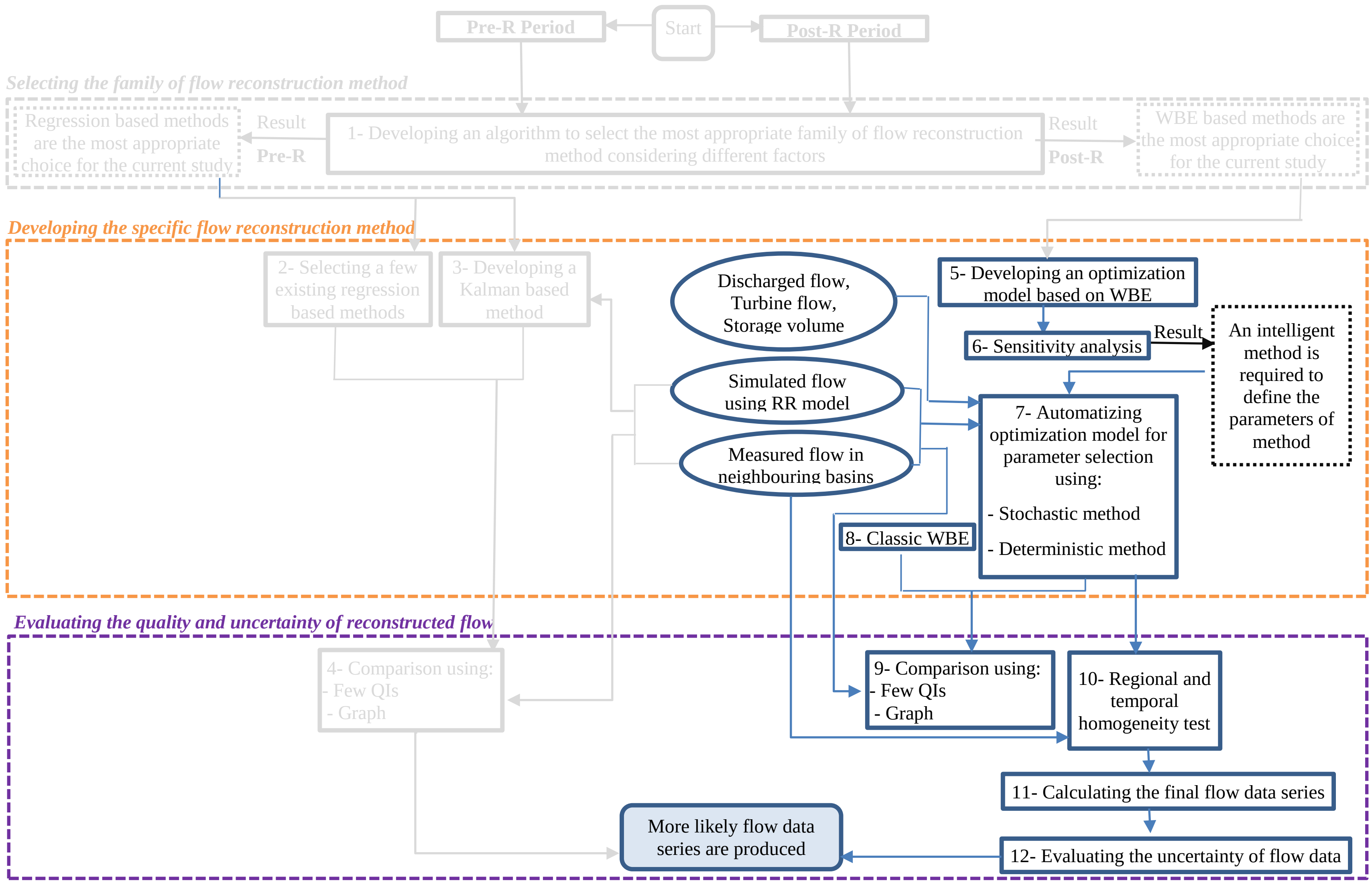


Figure 6-1: Schematic of methodology- Developing the specific flow reconstructing method and evaluating the quality and uncertainty of reconstructed flow for Post-R period

6.2 Methodology

First, an optimization model will be developed in order to address the problems of POM (described in Chapter 2). This model is called the Improved Optimization model or Improved POM in this thesis. A sensitivity analysis will then be performed to evaluate the authenticity of the model's assumption regarding constant parameters in time and space. Finally, Stochastic and Deterministic techniques are used to estimate the appropriate PS (Parameter Set) of the improved optimization model. Lastly, the final flow data series is estimated using a Weighted Average method. At the end of the chapter, a methodology will be suggested to define the probable range of flow resulting from input data uncertainty.

6.2.1 Developing an Optimization Model Based on POM

In this section, an optimization model will be developed based on POM (improved POM) in order to solve the problems of POM (mentioned in Section 2.2.2.3.1). The objective function and the constraints of the improved POM are presented in Equations 6.1 to 6.9. This optimization model:

1. considers the parameter γ as a coefficient for squared E. Equation 6.1 is a single quadratic objective function in which all the variables possess their own coefficients (solving problem number 2, Section 2.2.2.3.1).
2. has a moving optimization window with the length of *dn* to define the variables in each step. Unlike POM, only the calculated flow in the middle of the window is returned as the output of that step. For instance, as it is shown in Figure 6-2, a sample year (365 days) is divided into time periods with the length of 3 hours (*dn*=3). Thus, for the first time, the optimization model is solved for the days 1 to 3 (*start day* =1 and *end day* =3) but only the flow value of the day in the middle of window (day # 2 in this example) is saved. In the second step, the optimization model is solved for the days 2 to 4 (*start day* =2 and *end day* =4) but only the flow value of day # 3 is saved. This means that the optimization window is moving and the problem of poor boundary condition do not affect the flow values (solving problem number 4, Section 2.2.2.3.1). Calculating each flow value while

considering the other flow values of optimization window is an obvious advantage of this model over classic WBE.

3. does not depend on a 5 minute interval data sample. To make the model independent of the 5 minute interval data sample, the variable of *Volume* (Equations 2.6 and 2.7) is replaced with the measured volume data and thus no longer requires variables or data with the *inf* or *sup* subscript. Consequently,

    a. Equations 2.3 and 2.4 are removed from the model (solving problem number 5, Section 2.2.2.3.1).

    b. The inequalities of Equations 2.6 and 2.7 (WBE) are replaced by equality of Equation 7.2 (WBE). This means that the range of WBE will not be affected by time intervals anymore, and the certainty of results is not decreased with increasing time intervals.

    c. Not only is the model much simpler to use, but it is also applicable for reconstructing daily flow without any concern.

4. is applicable for any time period. Since the model is not dependent on 5 minute interval data sample anymore, it can even be applied for time periods when only hourly or daily volume data is available (solving problem number 6, Section 2.2.2.3.1).

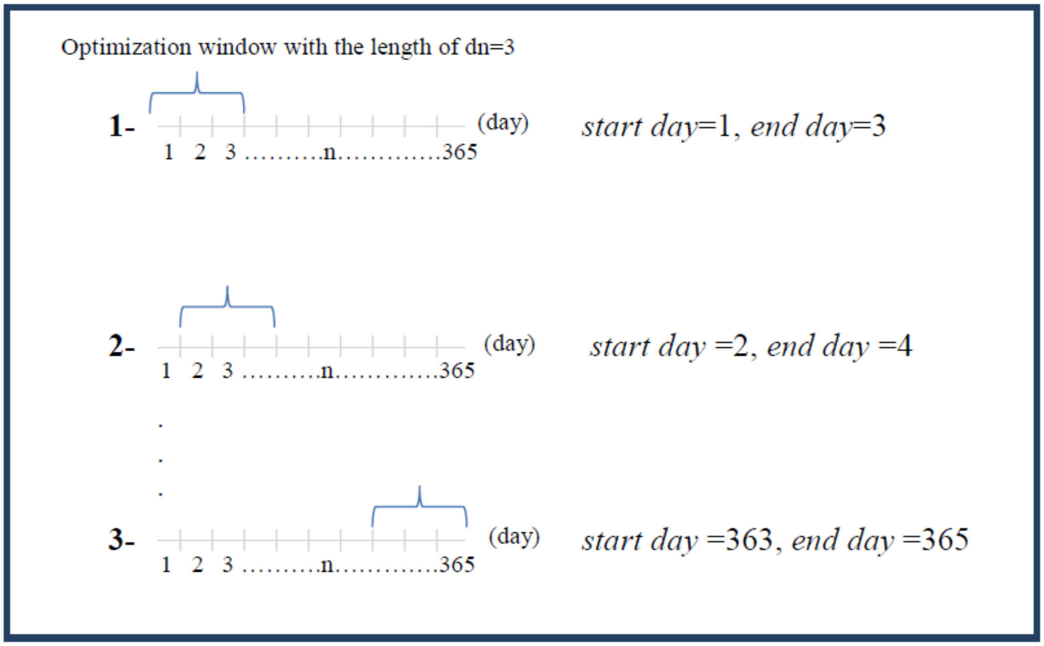


Figure 6-2: The schematic of optimization window in the improved POM for a hypothetical year

Note that the problems number 1, 3, and 7 (mentioned in section 2.2.2.3.1) are also solved using the suggested methods for parameter estimation in section 6.2.4. In order to solve the problem of constant parameters, a deterministic genetic algorithm and a stochastic probabilistic algorithm will be proposed in Section 6.2.4 in order to define the parameters automatically for each segment[9] of a year. In using these methods, the parameters could change by time and case-study (solving problem number 7, section 2.2.2.3.1).

In order to use all available data (flow from neighbouring basins and RR model), the simulated flow generated by the RR model is entered to deterministic based models as the external signal and it forces the flow values to get close to the simulated flow by a predefined weight. Also, the parameters of the stochastic based model are defined by considering the hydrograph shape of the neighbouring basins and flow from RR model (solving problem number 3, section 2.2.2.3.1).

---

[9] Each year is divided to few time periods called segment. In this thesis, each segment is defined in a way that it includes the flows of almost the same characteristics. To do so, the annual cumulative simulated flow data series is plotted and segment start points are placed where the slope of this data series shows considerable change.

The modification made in this section, in addition to the suggested methods in section 6.2.4, results in decreased noise and improved quality of reconstructed flow values (solving problem number 1, section 2.2.2.3.1).

The modified POM is represented in the following equation:

$$Minimize \quad \gamma \times \sum_{n=startDay}^{endDay} E_{(n)}^2 + c \times (p \times \sum_{n=startDay}^{endDay} Z_{(n)}^2 + q \sum_{n=startDay}^{endDay} \Delta_{(n)}^2) \tag{6.1}$$

Unlike POM, the parameter of γ is considered as a coefficient of squared E in the objective function (Equation 6.1). As it appears that the parameters of c and γ have the same role in POM, it is not necessary to consider both as the variable variation parameters of flow (Z and Δ). Therefore, it is now easier to manipulate the coefficients of different parameters in the modified objective function (Equation 6.1) and change their ratio.

Subject to (Equations 6.2 to 6.8):

$$qin_n - qout_n + F_n + E_{(n)} = (v_{(n)} - v_{(n-1)}) \times \frac{1000000}{(24 \times 3600)} \quad n = startDay,....,endDay \tag{6.2}$$

where $v_{(n)}$ is the measured volume of water in the reservoir at the beginning of day n in hm^3 (the coefficient of $\frac{1000000}{(24 \times 3600\ )}$ is to convert the unit of volume from hm to m^3/s).

In Equation 6.2, $v_{(n)}$ is the measured volume (which is a substitute for both $v_{\inf(n)}$ and $v_{\sup(n)}$ in Equations 2.6 and 2.7) that makes the model independent from 5 minute interval data sampling (data which is measured every 5 minutes) for n^{th} day. This modification seemed necessary because achieving an independent model from 5 minute interval data sampling makes the equation applicable for any time period, even if 5 minute sampling intervals are not available. Also, a model free of $v_{\inf(n)}$ and $v_{\sup(n)}$ variables is flexible enough to be applied for any time scale (hourly, daily, etc.).

By performing this alteration, the two inequality of Equations 2.6 and 2.7 are replaced by the equality of Equation 6.2 in the improved POM, and the *offset* subscript was removed from the model, with $E_{\inf,n}$ and $E_{\sup n}$ being replaced with $E_{(n)}$ (Equations 2.3 and 2.4 were removed). The rest of the model stayed the same as follows:

$$F_{(n+1)} - F_{(n)} = Z_{(n)} \qquad n = startDay, ..., endDay \qquad (6.3)$$

$$F_{(n)} - F_{(n-1)} - F_{(n+1)} + F_{(n)} = \Delta_{(n)} \qquad n = startDay, ..., endDay \qquad (6.4)$$

$$-\infty < E < +\infty \qquad (6.5)$$

$$0 < F < +\infty \qquad (6.6)$$

$$-\infty < \Delta < +\infty \qquad (6.7)$$

$$-\infty < Z < +\infty \qquad (6.8)$$

Equations 6.3 and 6.4 show the variation of flow values for two and three consecutive days respectively. These variations are minimized in Equation 6.1 to decrease the flow data noise. Although it seems that minimizing the flow variation for a period of two consecutive days (Equation 6.3) should be enough to minimize the noise, it is not able to decrease the probability of a "zigzag" data series. Minimizing the variation for a period of three sequential days (Equation 6.4) was required to give a set of smoother flow time series.

One consideration that has not been addressed in this study is the situations where there is a gap in volume, turbine flow, and/or discharged flow data series. In such cases, the suggested optimization model is not able to estimate flow values and a report value of "NA" (not available) must be entered in the records. The missing flow values, in the end, do need to be estimated. Since this situation did not occur in the current case study, the subject is not addressed in this thesis. However, different methods such as linear regression, nearest neighbour, Piecewise Cubic Spline, and Piecewise Cubic Hermite are suggested methods to be used for estimating the missing flow data for basins in Quebec.

6.2.1.1 Sensitivity analysis

A sensitivity analysis was performed to show the effects on flow values as a result of changing the parameters of the improved POM.

In the original POM, the parameters of *p*, *q*, and γ were considered equal to 1, *c* had two different values for different seasons (*c*= 10000 for winter and *c*= 1 for the remaining seasons), and *dn* was set to 3. These parameters do not change from year to year and from basin to basin in the original POM.

To understand how the results change when *c*, γ, and *dn* take on different values, few hypotheses are considered as follow:

a) The results change by changing *dn* with season:

    i. *dn* has different value in winter from rest of year

    ii. *dn* has different value in winter, summer and rest of the year

b) The results change by changing *dn* and *c* with season.

    i. *dn* and c have different values in winter from rest of year

    ii. *dn* and c have different values in winter, summer and rest of the year

c) The results are affected by the value of γ

In the hypothesis above, the changing of seasons are not dictated by a date on the calendar but on the change of the maximum and minimum climate temperature. When the minimum temperature is greater than zero, it is classified as summer, and when the maximum temperature is less than zero, it is classified as winter. All other temperature ranges are classified as spring.

6.2.2 Selecting Techniques for Defining the Parameters of Improved POM

Although POM produces more feasible flow data values than the classic WBE, it still exhibits some deficiencies. Though some of the weaknesses in the method were overcome in the improved version of POM (Section 6.2.1), nevertheless, a more intelligent method is still required to define the parameters of the model automatically.

Several methods are available in order to determine the parameters of the optimization models. Some of these are Neural Network (Cheng e*t al.* 2009, Chu 1992), Genetic algorithm (GA), and Stochastic methods (Shalev-Shwartz and Tewari 2011). Different types of GA have been widely used to solve the optimization models over the last decade (e.g. Deb 2000, Deb 2002). In this thesis, a posterior GA, which is one type of Deterministic method, will be one of the selected methods that will be used to determine the parameters of the optimization model because;

i) it automatically defines the best parameter set,

ii) it solves problems by producing multiple solutions (it gives more than a single solution to provide the possibility of engineering judgment),

iii) it is easy to understand and to apply to existing models,

iv) it searches in parallel from a population of points, which means that it is able to explore the solution space in multiple direction at the same time. Therefore, it has the ability to avoid being trapped in local optimal solution like traditional methods, which search from a single point (Marczyk, 2004),

v) it is “robust and has been proven theoretically and empirically to be able to efficiently search complex solution spaces” (Simpson *et al*. 1994).

Another method used in this work to determine the PSs is the Probabilistic Algorithm, which is one type of the Stochastic method. This method was selected because:

i) it automatically defines the best PS, and

ii) it considers the probability of different PS's to define the final PS.

In short, the applied methods to determine the PSs are GA and the Probabilistic Algorithm. However, in this thesis, the more general terms of Deterministic method and Stochastic method are respectively used to name these algorithms.

The list of suggested methods for flow reconstruction during the Post-R period and the reasons of their selection are tabulated in Table 6.1.

Table 6.1: List of applied methods to define the parameters of suggested method for flow reconstruction during the Post-R construction period

	Method	The reason of choosing the method
Post-reservoir construction period	Deterministic technique	- Defines the parameters automatically - Provides multiple solutions - Easy to understand and to be transferred to existing model - it has the ability to avoid being trapped in local optimal solution - It is able to efficiently search complex solution spaces
	Stochastic technique	- Defines the parameters automatically - Takes the probability of different parameter sets into account (for calculating the best parameter set)

6.2.2.1 Deterministic technique

As discussed in Section 6.2.1, the improved POM estimates the daily flow in ungauged basins by solving the WBE for each reservoir. In this model, a single quadratic objective function is used to minimize the error of WBE and decrease the variation of flow data series over continuous days. However, as it was stated earlier, the parameters of model cannot be constant in time and space. Since flow characteristics differ depending on the time of year (winter, spring, and summer), the PS model should be different for each sub-time period. Thus, each year is divided into few segments (sub-time period), and the most appropriate parameter set (PS) is determined for each segment. The number of segments is kept at less than 8 for each year to avoid unnecessary complexity.

In this section, a deterministic GA was developed to define the parameters of the optimization model for each segment automatically. A trial and error procedure shows that changing the objective function from Equation 6.1 to Equation 6.9 gives better results in a GA based optimization model. In this objective function (Equation 6.9), simulated flow is entered into the model as an external signal. This forces the flow values to approximate simulated flow (using RR model) by the weight of *d*. Since the RR model is developed based on meteorological data, it is able to take climate and the flow dynamics (phenomena such as winter snowfalls and spring floods) into account. Thus climate of area and the dynamic of flow are indirectly considered in the Deterministic method.

$$Minimize \quad (\gamma \times \sum_{n=startday}^{endday} E_{(n)}^2 + c \times (p \times \sum_{n=startday}^{endday} Z_{(n)}^2 + q \sum_{n=startday}^{endday} \Delta_{(n)}^2) + d \times \sum_{n=startday}^{endday} A_{(n)}^2) \tag{6.9}$$

With:

$$A = q_{sim,n} - F_n \tag{6.10}$$

where $q_{sim,n}$ is the simulated flow by RR model for day *n*, and *d* is a weight vector defined by the user.

The algorithm of the proposed methodology is presented in Figure 6-3. The algorithm benefits from posterior GA (Whitley 1994) in defining the parameters of the optimization method (bolded boxes are the steps of GA). The posterior GA is created by solving the GA for different *d* coefficients. As can be seen in Figure 6-3, the GA finds the best PS for each segment of the year and each predefined *d* coefficient (which defines the level of similarity between calculated flow and simulated flow using RR model). This approach produces more than one parameter set and thus more than one reconstructed flow series for each segment, which allows one to select among the generated results.

At the end of the process, there will be *d* estimated flow series for each segment. Among these series, the best graph is selected based on a few quality indices and trend similarity to flow calculated by classic WBE. First, the graphs with higher QIs are selected, and then the final graph is chosen visually among them. The final graph chosen should exhibit a data series that appear smoother than other graphs and should not display general over or under-estimations in comparison with classic WBE.

In the GA technique described in this section:

a) the genes include γ, *p, q, c, dn* (five parameters),

b) the fitness of each PS is the calculated QI (Nash–Sutcliffe coefficient (1970), which compares simulated flow and reconstructed flow) for the estimated flow data series based on that PS,

c) the mother PSs of the next generation are defined using the Tournament method,

d) the chance of a mutation occurring is considered at 2 percent

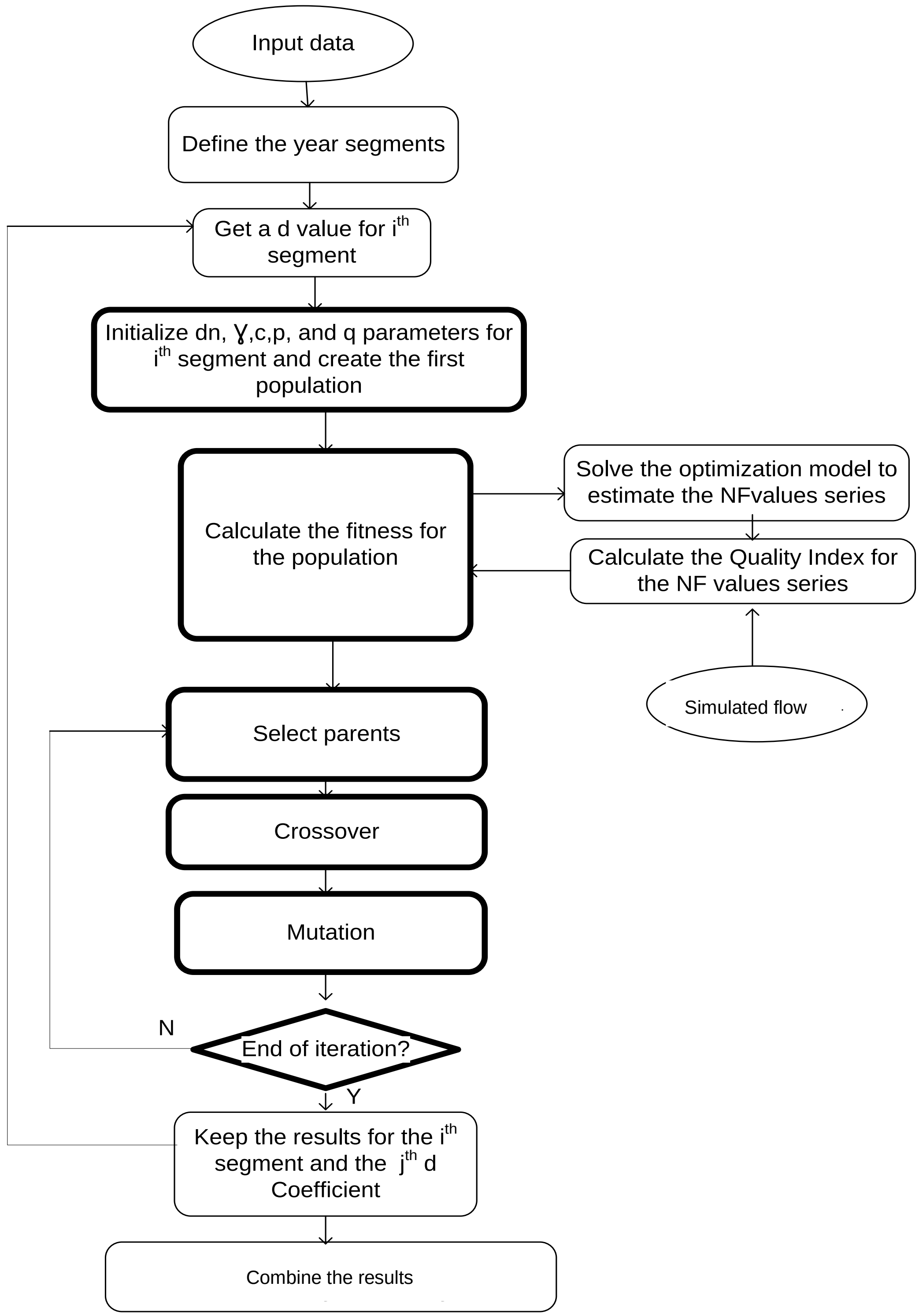


Figure 6-3: Schematic of the proposed Deterministic based optimization model

6.2.2.2 Stochastic technique

The parameters of the optimization model were also defined using a Stochastic method. In the Stochastic based optimization model, the objective function was kept as Equation 6.1.

Like the Deterministic based model, each year is divided into segments, and then, the 8-step Stochastic based optimization model is completed in order to define the model's parameters for each segment (Figure 6-4).

Step 1: One hundred random initial parameter sets are produced (Figure 6-4, Box 1).

Step 2: The optimization model is solved with each parameter set and the related flow series are produced. Thus, one hundred flow data series should be produced (Figure 6-4-box 2).

Step 3: The fitness of each reconstructed flow series from Step 2 is determined (Figure 6-4-box 3) based on weighted summation of the three normalized criteria QIs (*NN*, *NAVE*, *CC*) using Equation 6.11. *NN* compares the reconstructed flow with simulated flow from the RR model. As the RR model is developed based on meteorological data, *NN* reflects the degree of coherence between the reconstructed flow and area's climate. *CC* compares the reconstructed flow with the neighbouring basin's flow and thus reflects the level of coherence between the reconstructed flow and regional flow dynamic.

$$Benefit = w_1 \times NN + w_2 \times NAVE + w_3 \times CC \tag{6.11}$$

where:

$$NAVE = \left(\frac{1}{1 + AVE}\right) \tag{6.12}$$

$$AVE = \frac{\sum | q_{WBEi} - q_{cali} |}{\sum q_{WBEi}} \tag{6.13}$$

and:

w_1, w_2, w_3 = user defined weight coefficients

$NAVE$ = normalized absolute volume error,

q_{WBEi} = calculated flow by classic WBE for the day i

(NN, and CC are defined using Equations 5.18 and 5.20 respectively).

Step 4: Narrow down the number of parameter sets by selecting the 10 sets with the highest fitness (Figure 6-4-box 4).

Step 5: This step has 2 stages: first, the probability of the ten PSs generated from Step 4 is calculated using Equation 6.14 (Kalakrishnan et al. 2011):

$$P(ps_j) = \exp\left(-\frac{1}{Benefit_j}\right), \quad j = 1, 2, ..., 10 \tag{6.14}$$

where:

$P(ps_j)$ = the probability of j^{th} parameter set

$Benefit_j$ = the fitness related to this j^{th} parameter set (calculated in Step 3).

Afterwards, an aggregated parameter set is estimated using the 10 calculated probabilities in Equation 6.14. This aggregated PS will be used as the mother parameter set for the next iteration (Figure 6-4, Box 5). Here the parameter sets with the higher probabilities will have more weight in the mother PS definition.

$$c_n = \frac{\sum_j P(ps_j) \times c_j}{\sum_j P(ps_j)}, \; \gamma_n = \frac{\sum_j P(ps_j) \times \gamma_j}{\sum_j P(ps_j)}, dn_n = \left\| \frac{\sum_j P(ps_j) \times dn_j}{\sum_j P(ps_j)} \right\|, j=1,2,...,10 \tag{6.15}$$

where: c_n, γ_n, and dn_n form the mother parameter set of the next iteration, c_j is the c coefficient of j^{th} parameter set, γ_j is the γ coefficient of j^{th} parameter set, and dn_j is the dn parameter of j^{th} parameter set.

The optimization model is solved using this new PS and the related fitness is defined.

Step 6: If the difference between the computed fitness in Step 5 and the fitness from the last iteration is greater than a predefined value (Figure 6-4, Box 6), it indicates that the desired convergence has not been reached and we proceed to Step 7. Otherwise the iteration is stopped and user will proceed to Step 8.

Step 7: One hundred new PSs are again generated (Figure 6-4, Box 7) using the mother parameter set calculated in Step 5. To do this, each parameter is selected randomly from a specific range around the mother parameter (defined in Step 5).

Step 8: When the desired convergence is reached, the final flow data series is reconstructed based on the last parameter set defined in Step 5.

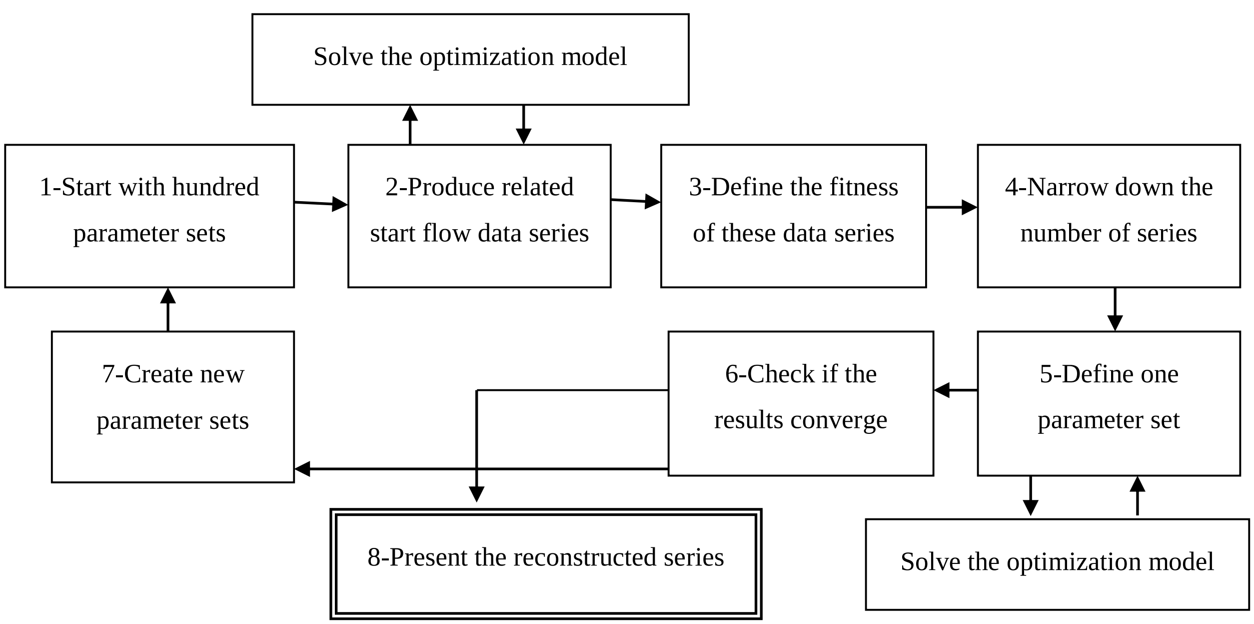


Figure 6-4: Schematic of the Stochastic based optimization model

6.2.3 Evaluating the Quality of Reconstructed Flow

The common method for assessing a data series is by comparing the generated data set with a measured data series using one or two traditional indexes. However, evaluating the quality of the data series for an ungauged basin has remained relatively ignored.

Since there are statistically no "best quality criterion" for hydrological simulation models, users tend employ more than one index to assess more efficiently the performance of the model (Weglarczyk, 1998). This project will also use different methods as well to assess the quality of the reconstructed flow. In addition to a visual evaluation, five quality indexes will be used to analyze the different aspects of the calculated flow. Also, the regional and temporal homogeneity of reconstructed flow are examined.

6.2.3.1 Quality indexes

Five QIs are designed to evaluate the degree of reliability of the reconstructed flow using the proposed Deterministic and Stochastic techniques. These five criteria are:

1) *NN* (Equation 5.18)
2) *CC* (Equation 5.20)
3) *NAVE* (Equation 6.12)
4) $$SFR = \begin{cases} 2 - \dfrac{sf_{cal}}{sf_{WBE}} & if \ \dfrac{sf_{cal}}{sf_{WBE}} > 1 \\ \dfrac{sf_{cal}}{sf_{WBE}} & if \ \dfrac{sf_{cal}}{sf_{WBE}} \leq 1 \end{cases} \quad (6.16)$$

$$SF = \frac{\sum_i flow_i}{s} \quad (6.17)$$

where:

SFR = *the specific flow ratio*

sf_{WBE} =*the specific flow calculated using the classic WBE*

sf_{cal} *=the specific flow calculated using the reconstructed flow*

s *= the surface of the basin.*

5) *NT* (Equation 5.24)

As described in Chapter 1, flow data series are used for different purposes such as flow prediction, flood analysis, water resource management, and flow simulation. In this project, the suggested indices take into consideration these purposes.

- *NAVE*: Over- or under-estimations change the pattern of flow data series and introduce some uncertainties to flow analysis studies, and thus to flow predictions. Unreliable predicted flow causes some deficiencies in water resource management and results in inefficient exploitation of natural resources. Therefore, this QI is designed to penalize over/under estimations and to provide an idea about the safety level of flow analysis.
- *SFR:* The reconstructed flow data series should be able to close the water balance budget for long time period (seasonal or annual). Otherwise, it is obvious that there is a serious problem in the results of that period. *SFR* penalizes flow series that are not successful in closing the water balance budget for a year with comparison to classic WBE. Thus, it is beneficial for the users who care about the quality of average long term flow data (*NN*, *CC*, and *NT* are described in section 5.3.5).

All of these indexes are normalized between 0 and 1, values closer to 1 indicating better quality. The name and the reason of selecting these indexes are listed in Table 6.2.

Table 6.2: The list of designed quality indexes for post-reservoir construction period

	QI	The reason of choosing QI
Post-R construction period	*CC*	- penalizes the flow data series not respecting the regional homogeneity in the sense of hydrograph shape - beneficial for regional flow analysis studies
	NN	- penalizes the flow data series not respecting meteorological factors - more sensitive to peak flows - beneficial for flood prediction and PMF estimation, and RR model calibration purposes
	NT	- penalizes noisy flow data series - useful for water management and flow prediction studies
	NAVE	- penalizes over/under estimations - sensitive to peak flow and low flows equally - beneficial for short term flow analysis
	SFR	- penalizes flow data series that does not close water balance budget - useful for evaluating the average long term flow

6.2.3.2 Regional and temporal homogeneity

Confidence in the quality of reconstructed flow would increase if the flow is homogeneous regionally and temporally. In this project, KPSS is used to evaluate the stationarity (temporal homogeneity) of the data, a method which has been widely used in hydrological studies (i.e. Wang *et al.* 2005, and 2006).

Also, a regional analysis is done to assess the coherency of the quantile of the reconstructed flow obtained from local frequency analysis with that of neighbouring basins. To do this, the reconstructed flow in the case study is compared to 19 neighbouring gauged basins. First, the annual peak flows of each basin are sorted in descending order and then the probability of each flow is defined using the Weibul formula (Equation 6.18). A lognormal distribution with three parameters is then fitted to the peak flows (*LN3*) of the flow data series.

$$f = \frac{m}{n+1} \tag{6.18}$$

where:

f = probability,

m = rank of the value

n = total number of observations.

A power regression is then developed between the basin's area and the values of *LN3* over a 2-year return period (the results are the same for other return periods) as follows:

$$FQ = as^{b} \tag{6.19}$$

where *FQ* is the fitted line to the values of *LN3* of a 2-year return period, *a* and *b* are the parameters, and *s* is the basin's area. The residuals from this equation are plotted in Quantile-Quantile plot (Q-Q plot), scale location plot, and residual versus fitted value plot to evaluate the coherence of the reconstructed peak flows with the regional peak flows.

6.2.4 Calculating the Final Flow Data Series

In this section, it is suggested that the results of the two flow reconstruction methods (Stochastic based model and Deterministic based model) be combined using a weighted average technique. This strategy is plausible because:

- none of the flow reconstruction methods is preferred over the another for the whole time period for all the basins,
- if one model is deemed as unreliable for one time period and is rejected, it does not automatically get chosen to be used as the superior model for another period,

- Each flow reconstruction method has its own strengths in factoring in different effects on flow formation. Combining the results of flow reconstruction models takes advantages of each individual simulation and thus represents a more complete representation of the likely flow characteristic of case study basin (Dong et al. 2013, Reid, 1968).

- By combining N methods, each method will contribute an uncertainty multiple of 1/N into final flow. This means that the uncertainty of N methods will not be cumulated in final reconstructed flow.

Examples of weighted average combination in hydrological research are Neural Network (Shamseldin *et al.* 1997, Xiong & O'Connor 2002), Fuzzy System (Xiong *et al.* 2001) and Bayesian Model Averaging (Ajami *et al.* 2006). Most of these methods need to be treated using measured flow. As this data is not available for many of Quebec‘s basins, the simple weighted average method (WAM) is used to combine the values of flow calculated using Stochastic and Deterministic techniques. The weighted average flow for the k^{th} day of i^{th} segment of a year ($NF_{WA,k,i}$) is calculated using Equation 6.20. In this equation, the final flow of each day of a segment is a function of the weighted reconstructed flow using Deterministic and Stochastic techniques for the same day of that segment. The weight of Deterministic and Stochastic methods for each segment is the simple average of five QIs for that segment (Equations 6.21 and 6.22) related to the Deterministic and the Stochastic methods respectively. In Equation 6.20, the method with a better average QIs has the larger role in defining the final flow.

$$F_{WA,k,i} = \frac{w_{S,i} f_{S,k,i} + w_{D,i} f_{D,k,i}}{w_{S,i} + w_{D,i}} \tag{6.20}$$

$$w_{S,i} = QI_{ave\ S,i} \tag{6.21}$$

$$w_{D,i} = QI_{ave\ D,i} \tag{6.22}$$

where:

$F_{WA,k,i}$ = the weighted average flow (combined/final flow) for k^{th} day of segment i,

$f_{S,k,i}$, $f_{D,k,i}$ = the flow calculated using stochastic and deterministic techniques for k^{th} day of segment i,

$w_{S,i}$, $w_{D,i}$ = the weight of Stochastic and Deterministic techniques for segment i,

$QI_{ave\ S,i}$, $QI_{ave\ D,i}$ = the average QI of Stochastic and Deterministic methods for segment i.

6.2.5 Uncertainty of Reconstructed Flow

Flow uncertainty is the component of reported flow that characterizes the range of values within which the true value is asserted to reside. "Uncertainty analysis of hydrological modeling has become an indispensable element for any hydrologic modeling and forecasting." (Dong *et al.* 2013). Although many uncertainty analysis techniques have been proposed and are widely applied, defining the flow uncertainty in an ungauged basin has still remained an important issue, since it is a big challenge to develop a method independent from measured data. Even the few existing applicable methods, such as sensitivity analysis, are very time consuming to perform within a whole time period. The method of sensitivity analysis is developed based on running the flow reconstruction method for hundreds of times, and it is a laborious job to do over several years.

In summary, evaluating the uncertainty of reconstructed flow for the current case study is accompanied with some inconveniences:

- Using the common methods of uncertainty analysis is not possible due to a lack of measured flow in the area.
- It is very time consuming to estimate uncertainty using traditional methods (sensitivity analysis) for the whole time period.

Thus, a methodology which is applicable and flexible for any ungauged area basin and can be executable within a reasonable time interval is required.

In this project, the input data (turbine flow, discharged flow and volume data series) of the WBE based method have been calculated based on non-validated measured data prior to 2005, and their uncertainty was obvious to us. As a result, they cause some uncertainties in the reconstructed flow. Comparing the graphs of the classical WBE pre- and post-2005 confirms that the uncertainty of flow is more related to quality of input data than anything else. Thus, a methodology towards defining the effects of input data uncertainty[10] when performing uncertainty analysis is required for the area.

The suggested methodology which fulfills these needs could be summarized in four steps (see also Figure 6-5):

1) selecting a sub-time period for which uncertainty analysis will be performed,
2) defining the daily[11] uncertainty of each input data series,
3) evaluating the uncertainty of flow for the selected sub-time period, and
4) extending the calculated uncertainty range to the whole time period.

[10] Parameter and model uncertainties will not be addressed in this chapter as they seem negligible compared to input data uncertainties.

[11] Since the reconstructed flow is daily, the estimated uncertainty of that should also be daily to give a reliable level of confidence about the range of flow data series.

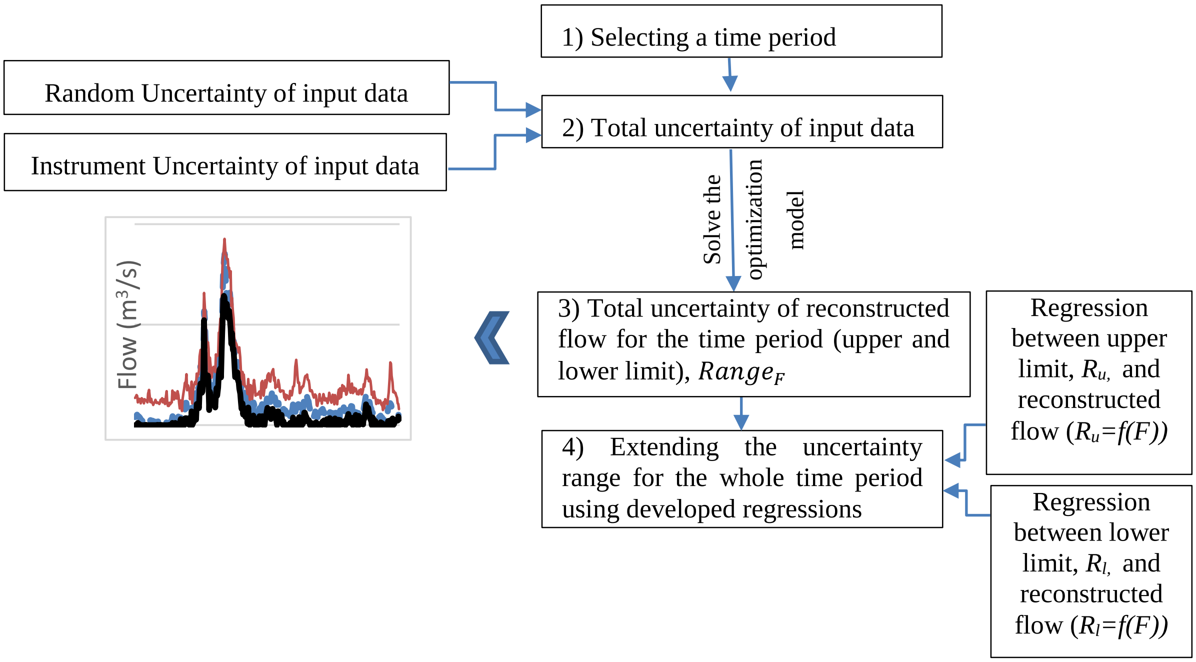


Figure 6-5: Schematic of developed methodology for flow uncertainty analysis

6.2.5.1 Selecting a sub-time period for which uncertainty analysis will be performed

The first step of uncertainty analysis process is selecting a sub-time period for which the uncertainty analysis will be performed. This the sub-time period should be at least one year to take into consideration all levels of flow (high and low flows). This one-year period should be selected carefully because it should include any usual and unusual values such as negative flows, low, medium, and high data variation. If this is the case, then the potential of calculating the uncertainty for different situation would be available. This chosen year is not representative of whole time period, but it is just a year based on which the uncertainty of whole time period is calculated.

Calculating the uncertainty for a sub-time period and then extrapolating the results to the desired time period is practical when it is very time consuming to do uncertainty analysis for a case study of multi-year time periods.

6.2.5.2 Input Data Uncertainty

Input data uncertainty includes instrument uncertainty and/or random uncertainty. The combination of these two factors forms flow uncertainty. The total uncertainty or range of flow is described by Equation 6.23:

$$Range_F = E_{input\ data} + \Delta_{input\ data} \tag{6.23}$$

where:

$Range_F$	= possible flow data series range (total uncertainty),
$E_{input\ data}$	= uncertainty of flow caused by random uncertainty of input data,
$\Delta_{input\ data}$	= uncertainty of flow caused by instrument uncertainty (while measuring input data).

6.2.5.2.1 Instrument uncertainty

Each instrument has its own uncertainty related to resolution, readability, and measurement method specific to that instrument; this type of uncertainty is called instrument uncertainty. Instrument uncertainty has been already defined for some of the basins in Quebec. This is explained in detail in the C1 report (Haché, *et al.* 1996), which states that the average uncertainties of turbine flow, discharged flow, input flow to the reservoir, and storage volume have been calculated for 17 sub-basins. The schematic of these basins are shown in Figure 6-6 and their characteristics are summarized in Table 6.3.

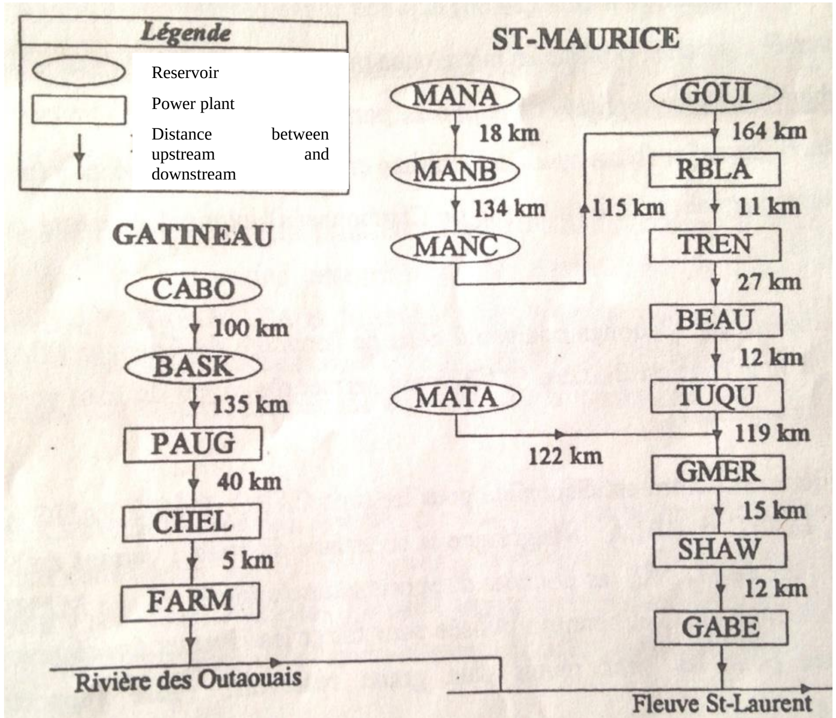


Figure 6-6: Schematic of reservoirs for which uncertainty analysis of input data is done by Hydro-Quebec

Table 6.3: Characteristics of studied basins for input data uncertainty

	Surface (km2)	Number of turbines	Number of gates	Number of water level gages
GOUI	10057	-	14	2
MANA	1507	-	7	1
MANB	733	-	7	1
MANC	842	-	7	1
RBLA	10142	6	7	1
RTREN	2719	6	9	1
BEAU	2680	6	9	1
TUQU	3711	6	8	1
MATA	4118	-	8	1
GMER	6008	9	21	1
SHAW	37	11	24	1
GABE	78	5	9	1
CABO	2659	-	10	3
BASK	13031	-	30	1
PAUG	6905	8	7	1
CHEL	1147	5	35	1
FARM	2	5	2	1

In the C1 report, the concept of uncertainty is introduced to highlight the differences between the accuracy of various quantities. Relative uncertainty is defined as the ratio of absolute uncertainty to the measured or calculated value. The absolute instrument uncertainty is defined as the possible maximum difference between an obtained value (measured or calculated) and the exact variable value. This uncertainty is a function of the measurement instrument, applied method, and the experimenter. Usually half of the smallest division of the unit is used for defining the uncertainty. Considering $y=f(x_1,x_2,..,x_n)$, the absolute uncertainty of y, Δy, is defined by:

$$\Delta y = \sum_{i=1}^{n} \left|\frac{\partial y}{\partial x_i}\right| \Delta x_i \qquad (6.24)$$

where:

$\left|\frac{\partial y}{\partial x_i}\right|$ = the partial derivative of y with respect to x_i

Δx_i = the absolute uncertainty of measured x_i.

6.2.5.2.2 Random uncertainty

Random uncertainty could also happen from any occasional natural phenomenon. For example, wind or ice cover can affect volume measurements, and maneuver disorders can influence turbine flow values. The magnitude of random uncertainty could change by season, reservoir size, number of turbines, etc. Thus, it is not easy to estimate the daily uncertainties caused by these factors. In this research, 5% of the input data value is considered as part of the random uncertainty in that data set.

In order to evaluate the sufficiency of the mentioned amounts of perturbation (5%), the sensitivity of the calculated flow (from classic WBE) to each term of q_{in}, q_{out}, and *volume* is assessed. To this end, each q_{in}, q_{out}, and *volume* data series are disturbed in a separate task (there will be three tasks in total). The amounts of perturbation are up to 5%. Then, in each task, the classic WBE is solved based on disturbed q_{in}, q_{out}, or *volume* data, and the difference between the original and the new flow data series ($D_{n,o}$) is calculated using Equation 6.25.

The 5% perturbation will be sufficient if $D_{n,o}$ shows a meaningful difference in the results of classic WBE. Since the perturbation limit could change depending on daily condition, more studies need to be performed in future research to make a better estimation of daily random uncertainty.

$$D_{d,o}\,(\%) = average\,(\frac{|nf_{d,i} - nf_{o,i}|}{|nf_{o,i}|}) \times 100 \qquad (6.25)$$

where:

$D_{n,o}$ = the difference between the original and the new flow ,

$f_{d,i}$ = the flow calculated using disturbed input data for i^{th} day,

$f_{o,i}$ = the original flow for i^{th} day.

6.2.5.3 Evaluating the uncertainty of flow for the selected time-period

6.2.5.3.1 Instrument uncertainty

To calculate instrument uncertainty, the developed optimization model should be solved based on q_{in},$\pm\,\Delta_{q_{in}}$, q_{out},$\pm\,\Delta_{q_{out}}$, and *volume*,$\pm\,\Delta_{volume}$ data series ($\Delta_{q_{in}}$, $\Delta_{q_{out}}$, and Δ_{volume} are calculated using Equation 6.24). This will give the range of reconstructed flow resulting from instrument uncertainty. Also, the range of WBE flow caused by instrument uncertainty would be simply $\pm(\Delta_{q_{in}} + \Delta_{q_{out}} + \Delta_{volume})$.

6.2.5.3.2 Random uncertainty

Instrument uncertainty always exists and can be calculated with using Equation 6.124 However, the existence and importance of random uncertainty is dependent on the conditions. Random uncertainty may affect one or more input data. Thus, few scenarios are used to evaluate the random uncertainty of each input data and their combination. This random uncertainty, then, should be added to instrument uncertainty to get the total uncertainty.

- Scenario I: Disturb q_{out} only
- Scenario II: Disturb q_{in} only
- Scenario III: Disturb volume only
- Scenario IV: Disturb all the three input data variable

The input data values are perturbed up to 5% in scenarios I to IV. Then, the developed optimization model is solved (for one hundred times) taking the randomly disturbed input data.

6.2.5.3.3 Total uncertainty

To calculate the total uncertainty, the calculated random uncertainty should be added to calculate instrument uncertainty. This gives the probable range of flow for the selected sub-time period.

6.2.5.4 Extending the calculated uncertainty range to the whole time period

The probable range of flow is calculated just for one year and it needs to be extended to the whole time period. To do this, two regressions are developed between daily reconstructed flow and daily upper limit and lower limit. These regressions, then, could be used to predict the upper and lower limit of uncertainty for each day of historical data series.

The magnitude of uncertainty depends on many random factors (such as wind and ice cover) and they may change from one year to another year. Therefore, in future studies, it is appropriate to develop a methodology that is able to consider the weight of different affective factors when estimating the range of uncertainty.

6.3 Conclusion

As described earlier, POM was recommended as the method for flow reconstruction and filtering of ungauged basins of Quebec. The limitations of this model were mitigated using the improved version of optimization model that was presented in this chapter. The improved optimization model is more reliable and practical; however, one of the questionable assumptions of this model is considering predefined constant values for the parameters. Thus, a sensitivity analysis was performed to evaluate the authenticity of this assumption. Then, two Deterministic and Stochastic techniques were suggested to select the parameters intelligently. Estimating the parameters using these automatic methods made the model easy to be reliably applied to any reservoir.

In this chapter, five Quality Indices were suggested in order to measure the reliability of the reconstructed flow. Temporal and regional homogeneity tests were also proposed to evaluate the characteristics of reconstructed flow in time and space. Then, a weighted average method was suggested to calculate the final flow data series by combining the results of Deterministic and Stochastic based. At the end, a methodology was explained to estimate the probable range of flow due to the instrument and random uncertainty. The results of applying these methods on the Outardes basin are presented and interpreted in next chapter.

CHAPITRE 7 FLOW RECONSTRUCTION (POST-RESERVOIR PERIOD, RESULTS AND DISCUSSION)

7.1 Introduction

In order to evaluate the validity of the assumption regarding constant parameters in time and space on the modified POM, a sensitivity analysis will first be conducted, after which the methods suggested in Chapter 6 for flow reconstruction will be used on three sub-basins of different sizes found in the province of Quebec. The results are then compared against those produced by the classic WBE Model. Analysis of the results will be done using different Quality Indices (Normalized Nash, Consistency Coefficient, Normalized Absolute Volume Error, Specific Flow Ratio, and Normalized Tortuosity), the Kwiatkowski-Phillip-Schmidt-Shin (KPSS) stationary test, and regional homogeneity analysis. The flow results will show the capability of the recommended methods for improving the reconstructed flow data series. The results are found to be all positive, less noisy, perfectly matched with regional flows, and are reliable enough for frequency analyses. Also, the final flow series will be calculated and the probable range of flow will be estimated using an uncertainty analysis. The results shown in this chapter will answer the questions mentioned at the beginning of Chapter 6.

7.2 Results Presentation

7.2.1 Sensitivity Analysis of the Modified POM

Classic WBE frequently produces flow data series which are noisy and highly improbable. For example, Figure 7-1 shows the results of applying this equation to the Outardes 3 basin, using 2009 as the sample year (2009 was chosen as the sample year for Outardes 3 to illustrate the sensitivity analysis results because the flow data series for this particular time period was very noisy and even included negative values). In this figure, the calculated flow values are noisy, especially during periods of low flows by exhibiting negative values, which is an improbable representation of flow. The noisy flow values could be the results of natural phenomena (i.e. floods, ice cover), instrument disorders (i.e. gates' maneuver disorders), instrument uncertainties, neglecting flow routing, simplification of calculations, and human uncertainties. According to this figure, the questionable data happened mostly during the time period when maximum temperature is less than zero (winter).

Figure 7-2 shows the results when the modified optimization model (Equations 6.1 to 6.9) was applied to the example case study of Outardes 3, using 2009 as the sample year, when all the parameters were set to one and *dn*=3. As it is clearly shown in the figure, the issue with negative flow values was resolved by applying the optimization model, causing the reconstructed flow values to be slightly less noisy. However, the problem of having noise, especially during periods of low flow, still remains an issue that needed a solution.

To understand how each parameter affects the results, different scenarios were developed. In these scenarios, the parameters of *p* and *q* are kept equal to 1 in order to evaluate the importance of *c*, *γ* and *dn* variables.

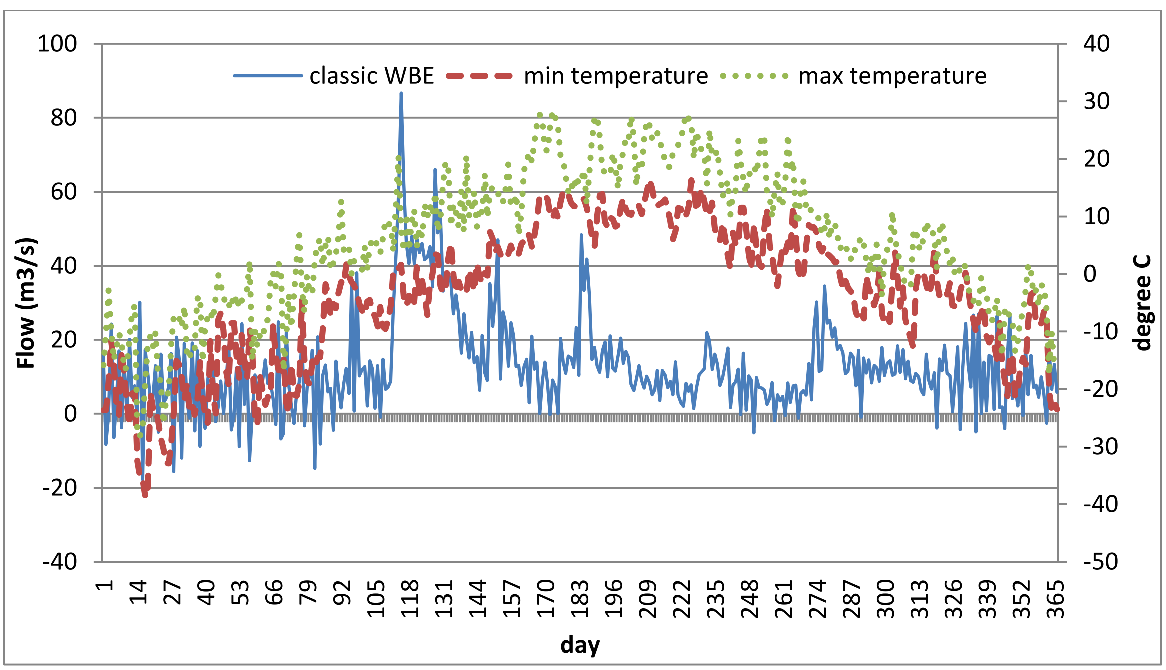


Figure 7-1: Results of applying classic WBE to Outardes 3 (2009)

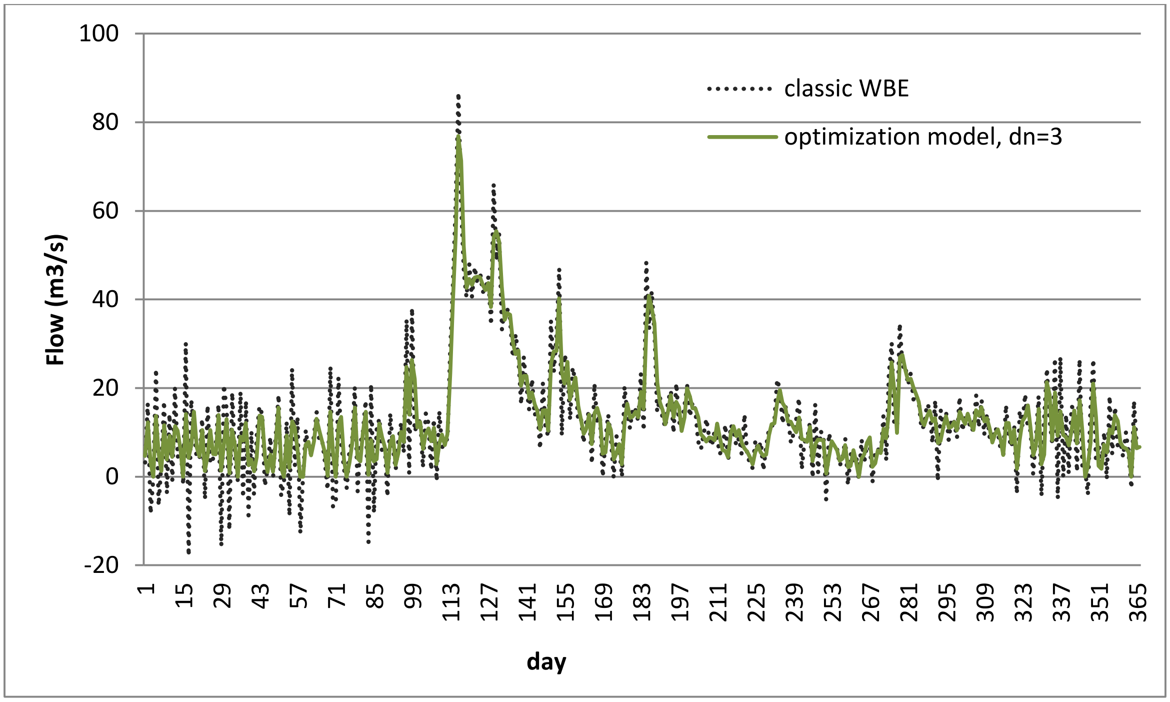


Figure 7-2: Results of applying optimization model to Outardes 3 (2009) when c=*p*=*q*=γ=1, *dn*=3

These developed scenarios are listed below:

a) The optimization results change by changing *dn* with season.

  i. *dn*=13 in winter, and for the rest of year *dn*=3 (*c*= γ =1)

  ii. *dn*=13 in winter, *dn*=7 in summer, and for the rest of year *dn*=3 (*c*= γ =1).

b) The optimization results change by changing *dn* and *C* with season.

  i. *dn*=13 and *c*=10000 in winter and for the other days *dn*=3, *c*=1 (γ =1).

  ii. *dn*=13 and *c*=10000 in winter, *dn*=7 and *c*=1000 in summer, and for the rest of year *dn*=3, *c*=1 (γ =1).

c) *dn*=13 in winter, *dn*=7 in summer, and for the rest of year *dn*=3, *c*=1 (γ =0.1,10).

Figures 7-3 and 7-4 give the results of these scenarios for the Outardes 3 basin using 2009 as the sample year. As shown in Figure 7-3-a, the noise level is decreased during periods of low flow (marked by red circles) in comparison with the results of the optimization model, when all the parameters are set to one and *dn=3* (darker line in Figure 7-3). This shows that smoother flow data series are obtained by increasing the number of days. However, increasing *dn* during the flood periods will under-estimate the peaks. Increasing *dn* from 3 to 7 during the summer in Scenario aii also gives smoother results (Figure 7-3-b in comparison with Figure 7-3-a), but again, peak flow may be underestimated.

Comparing Figure 7-3-a with 7-3-c, one can obtain good information on the effects of increasing the *C* coefficient during winter. As illustrated in Figure 7-3-c, this augmentation in *C* decreases noise and increases the data quality for sampling done during the winter season. However, increasing this parameter during the rest of year does not affect the results (comparing Figure 7-3-b to 7-3-d).

Figure 7-4 also shows the sensitivity of the optimization model towards the changing of γ. In this figure, it can be seen that the variation of γ considerably affects the results: when γ is equal to 10, the reconstructed flow is extremely affected by noise, but gets much smoother when gamma is changed to 0.1; however, some peaks are missing. Figure 7.4, in addition to Figure 7-3, demonstrates the importance of selecting the proper parameter set for each segment of year. As we concluded earlier, the optimization model helps improved flow estimations but it is still necessary to select an appropriate parameter set, which changes during time and space. Thus, it is necessary to estimate the most appropriate PS for each segment of a sampling year using intelligent techniques. In the work presented in this thesis, the suggested method for this purpose will be the Deterministic and Stochastic techniques. The results of using these techniques are presented in this chapter.

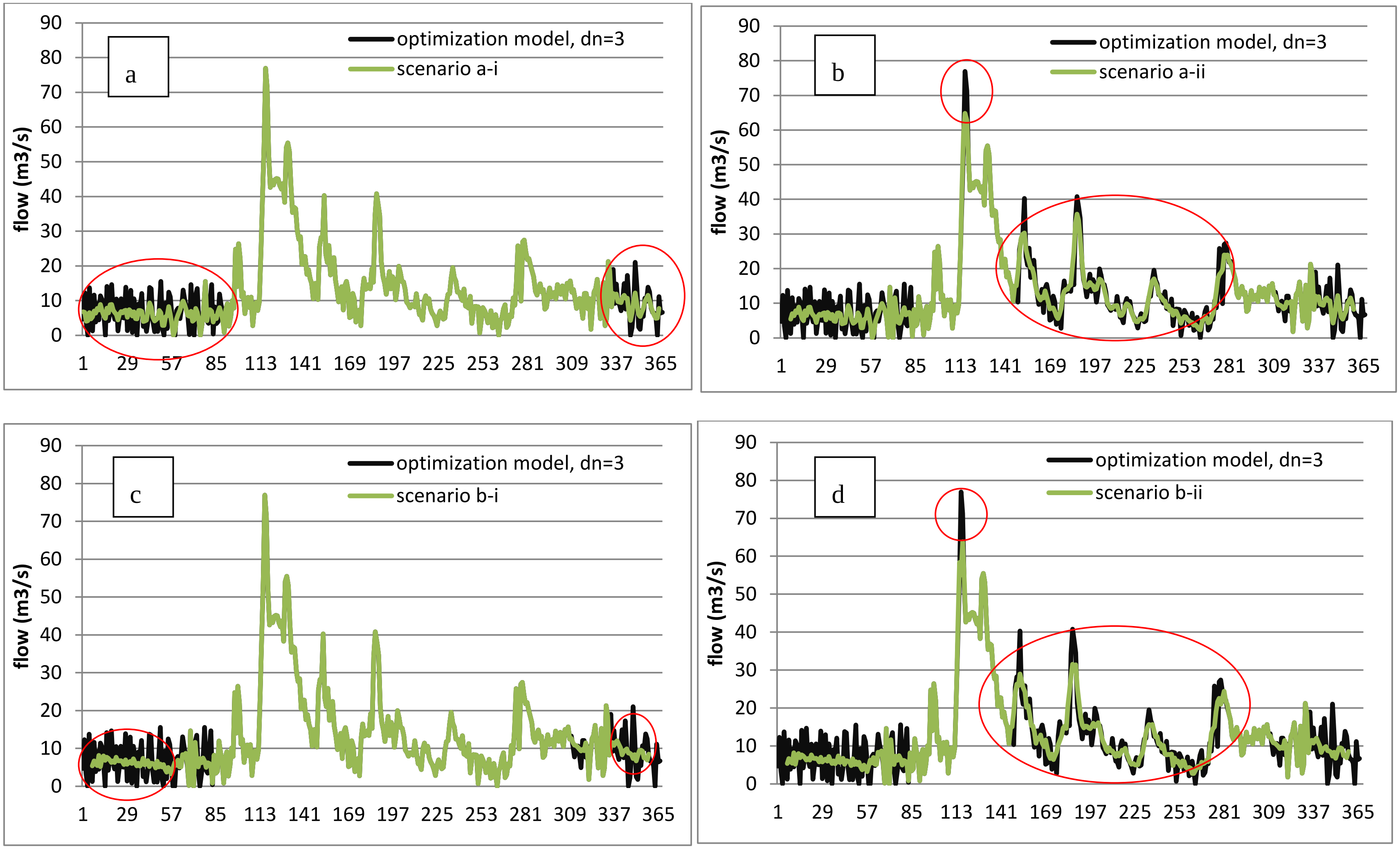


Figure 7-3: Calculated flow data series based on scenarios ai, aii, bi, and bii in comparison with optimization model, when all the parameters are set to one and *dn*=3, which is shown in darker color in the graphs (Outardes 3-2009)

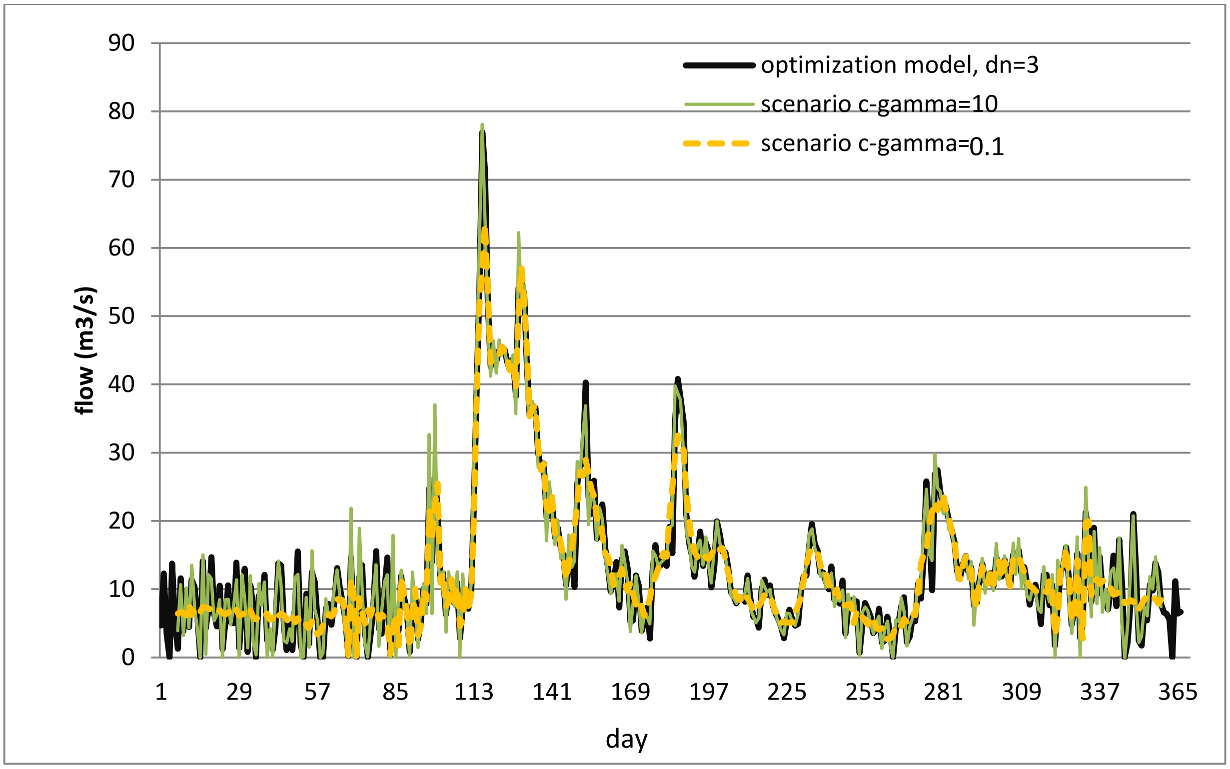


Figure 7-4: Calculated flow data series based on scenario c for two different gamma values in comparison with optimization model, when all the parameters are set to one and *dn*=3, which is shown in darkest color in the graph (Outardes 3-2009)

7.2.2 Techniques for Defining the Parameters of Improved POM

7.2.2.1 Deterministic based model

As was explained in Chapter 6, in using the Deterministic based model, each sampling year is split into several segments. Then, for each segment, the *d* coefficient takes on, one at a time, the different values of the following parameter set {150,100,80,50,20,15,10,7,5,1,0} (*d* coefficient values greater than 150 did not show any changes to the results). For each segment and each *d* value, one parameter set is selected by Genetic Algorithm (GA) and the flow series are determined based on this PS using the optimization model. In the end, eleven flow data series will be generated for each segment, each of them based on a simulated flow (using RR model) of different weight values. Note that simulated flow itself could not be considered as a flow series because it has exhibited deficiencies, especially during high flow. However, simulated flow of different weight values was factored into the process of flow reconstruction for each segment in order to benefit from their ability to increase the average QI. For example, Figure 7-5 provides a sample result from the Deterministic optimization model solved with different *d* coefficient for Outardes 4, using 2008 as the sampling year.

As illustrated, there is more than one graph for each year and the decision maker can single out the most appropriate one according to different Quality Indices (QIs) using a visual comparison against results from the regular WBE. Generally, the graphs that have considerably higher-quality indices (*CC*, *NN*, *NT*, *NAVE*, and *SFR*) are selected for each segment. The final graph is then chosen based on visual comparison (this is the graph which will be used as the Deterministic model per results in Figures 7-6 to 7-18). This graph should be smooth and follow the general trend of reconstructed flow by WBE (showing no considerable over/under estimations). It is important to keep in mind that a smoother graphical result may not necessarily be the better choice. For example, some smooth graphs may underestimate or overestimate the flow as highlighted in Figure 7-5 with red circles.

In the developed Deterministic based model, the parameter ranges ($1<dn<21$, $0<\gamma<15$, $0<c<10000$, $0<p<1000$, $0<q<1000$) are determined empirically.

7.2.2.2 Stochastic based model

As the Deterministic based model does not account for the probability of different parameter sets, a Stochastic model, which is a probabilistic algorithm, will be endorsed to determine reliable parameters for the optimization model. Unlike the Deterministic model, the Stochastic based model defines the parameters based on three QIs (*NN*, *NAVE*, and *CC*). Note that in the process of parameter definition, *CC* was calculated using Moisie as the neighbouring basin for Outardes 4 and Godbout as the neighbouring basin for Outardes 3 and Outardes 2.

The developed Stochastic based model is capable of producing one hundred random parameter sets at the beginning of first iteration in such a way that *dn* changes from 1 to 10 and each *dn* values is repeated exactly 10 times, while *c* changes from 0 to 10000, and γ changes from 0 to 10. These parameter sets are then used to solve the optimization model and are ranked according to their related QIs. In this model, more than one quality index is used to appraise the fitness of the series.

The weighted summation of *NN*, *NAVE*, and *CC* are the criteria used in ranking the parameter sets and the weights of these indices are defined through trial and error.

For our case study, w_1 & w_2=1 while w_3 changes according to the season (Equation 6.11). w_3 is usually assigned as 1 or 0.5 during low flow and 10 or more during the rest of the year. These weighted values are almost the same for all sampling years for the same basin, which makes the work in this research easier. The ten parameter sets with the highest QIs values are then drawn to define the probable "mother PS" of the next iteration. Using iteration, one hundred new parameter sets are formed based on the mother parameter set of the previous iteration. To do this, one hundred c, γ, and dn parameters are generated in the ranges of $c_n \pm 100$, $\gamma_n \pm 2$ and $dn_n \pm 2$ respectively (where c_n, γ_n, and dn_n belong to the mother parameter set). These ranges are large enough to avoid being trapped in a local optimum point and maintain parameter variety, yet be small enough to merge rapidly (in our case study, for example, the results merged after few iteration). The parameters are finalized when the difference between the fitness ("*Benefit*" calculated using Equation 6.11) of two consecutive iterations is less than 0.001. This amount of certainty would then be regarded as sufficient in this case. The results of the Stochastic based model are presented in Figures 7-6 to 7-18.

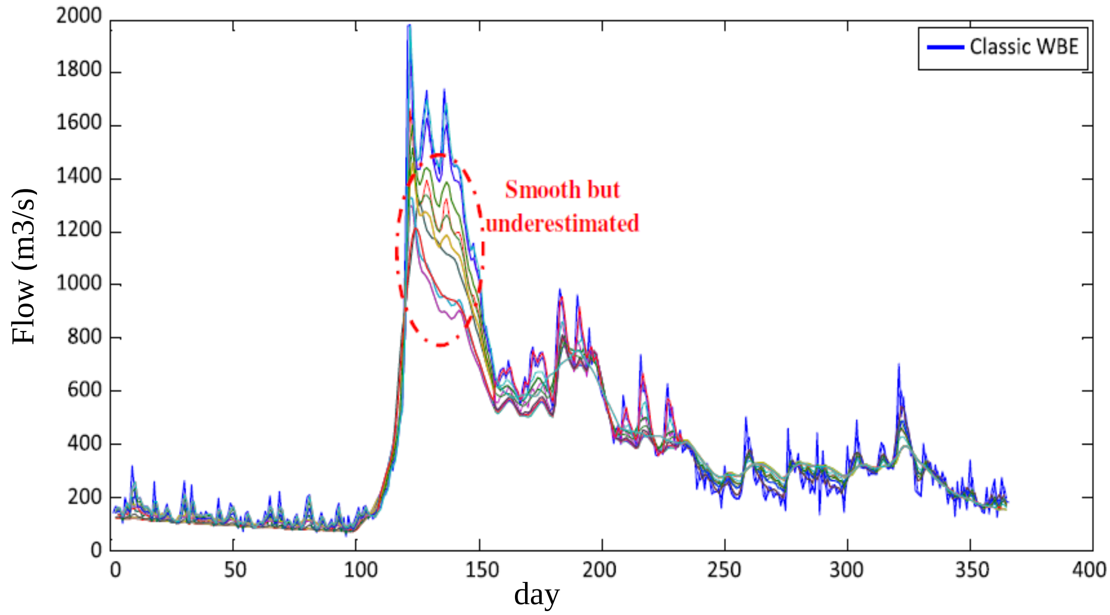


Figure 7-5: Comparison of daily flow (m^3/s) calculated by classic WBE (bold line) with deterministic based optimization model solved based on different *d* coefficient (rest of the lines)- (Outardes 4-2008)

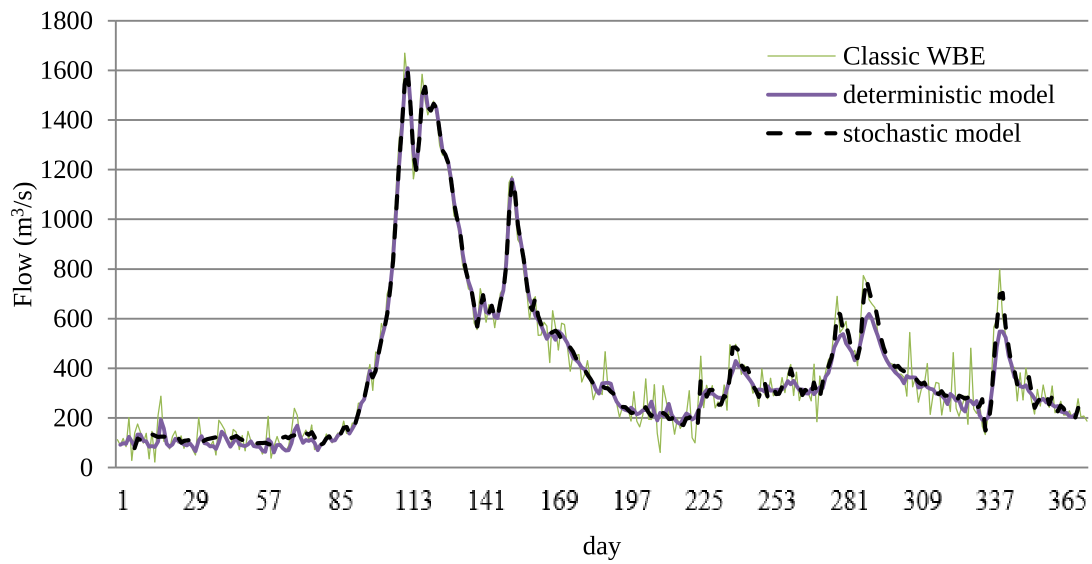


Figure 7-6: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 4-2005)

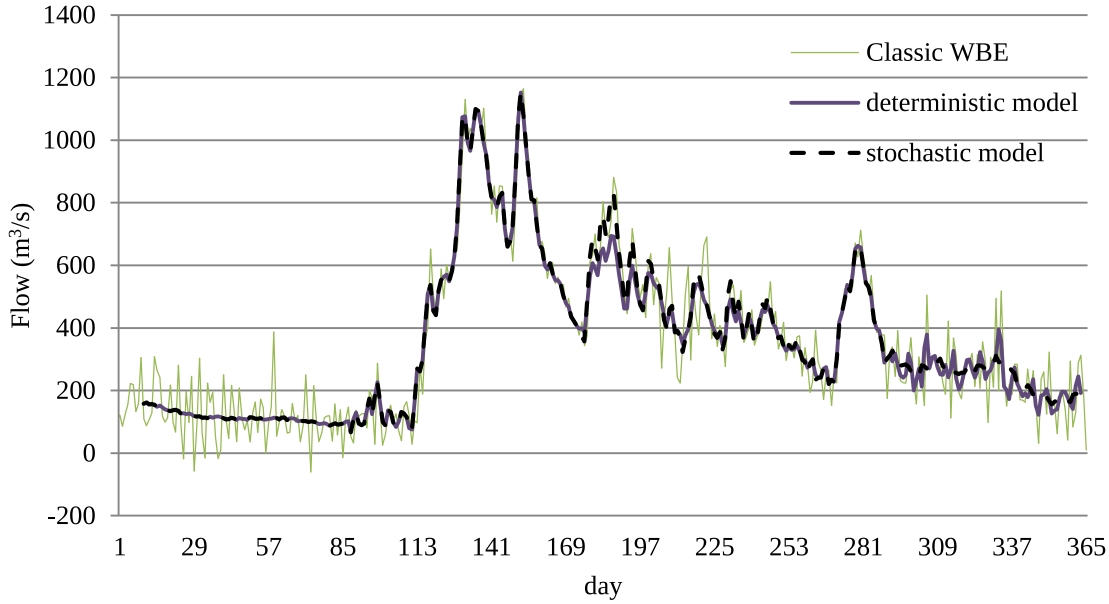


Figure 7-7: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 4-2009)

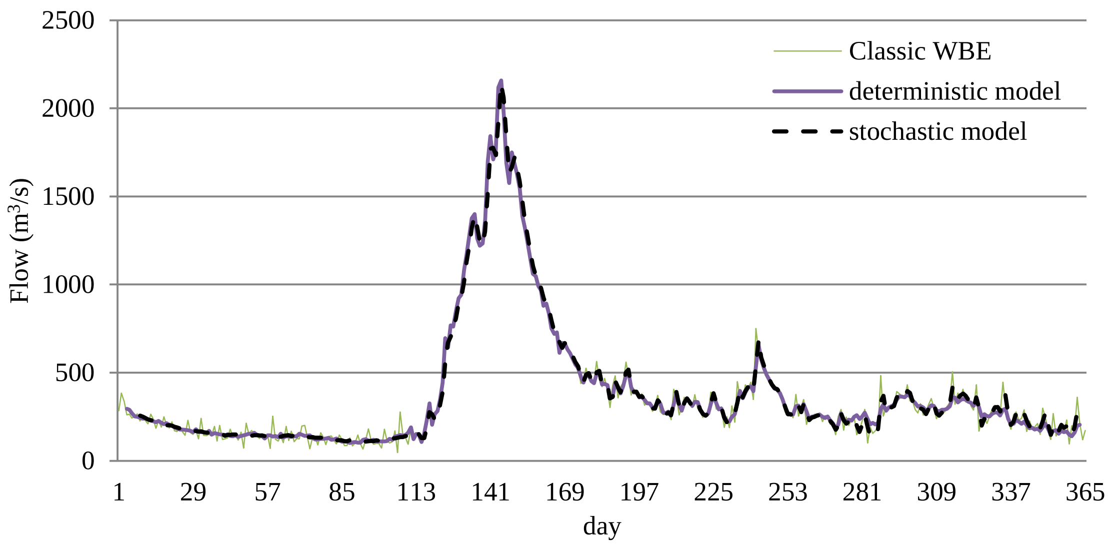


Figure 7-8: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 4-2011)

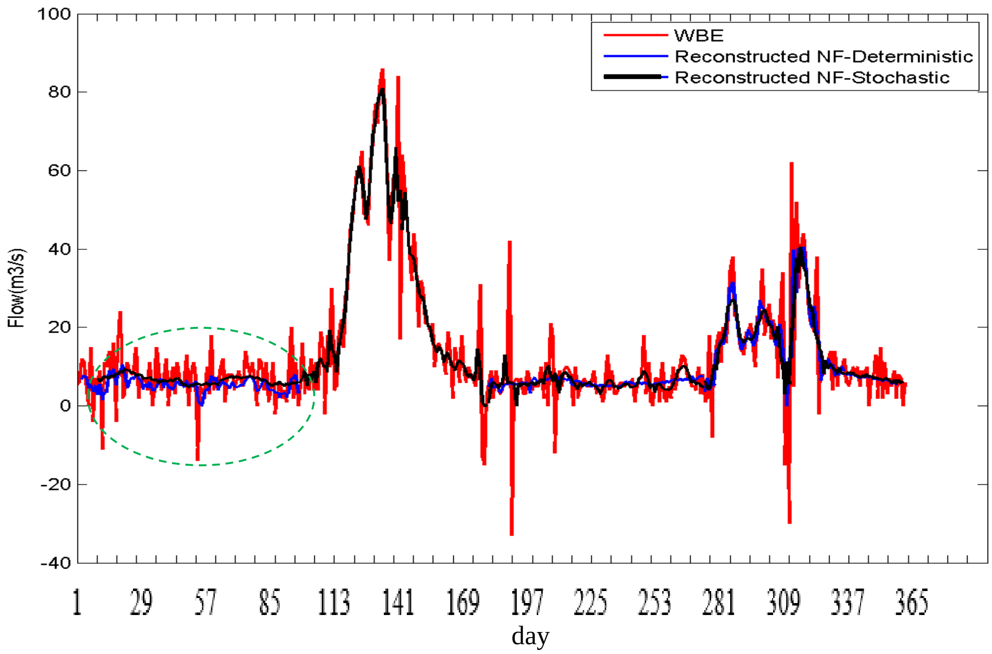


Figure 7-9: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 3-1995)

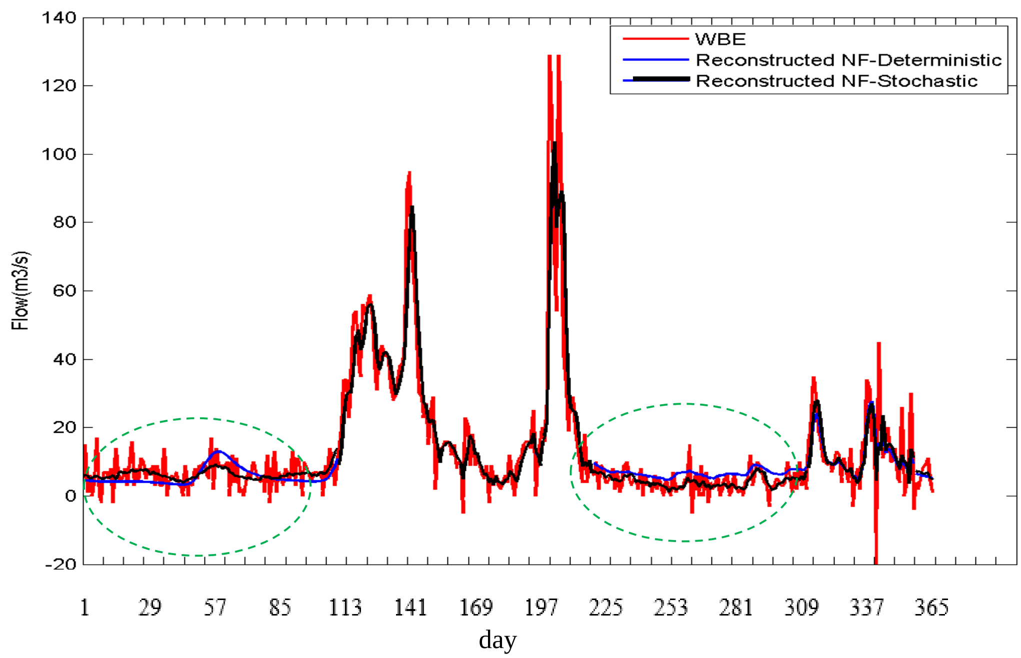


Figure 7-10: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 3-1996)

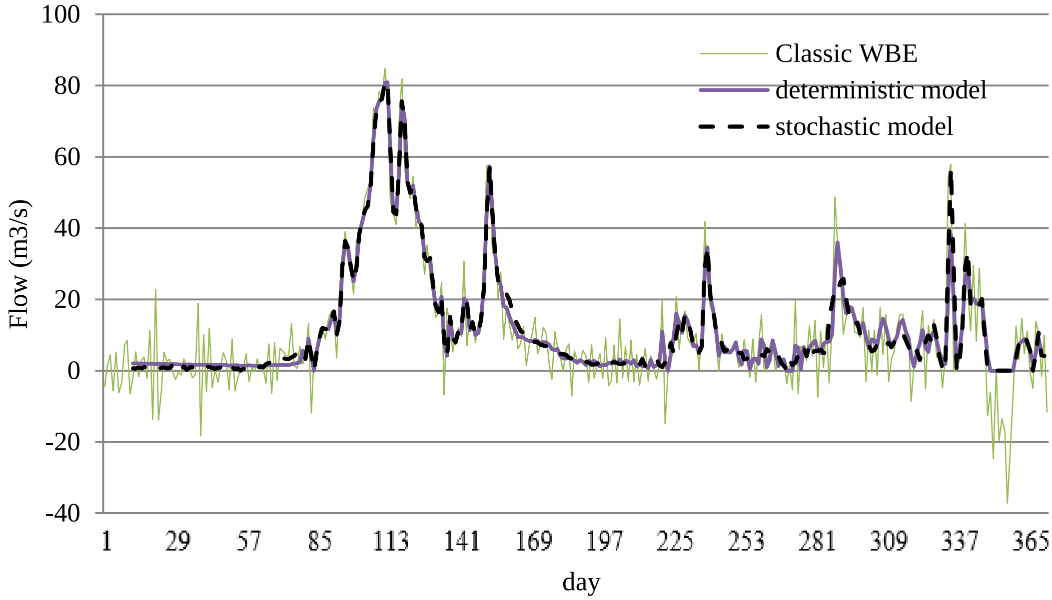


Figure 7-11: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 3-2005)

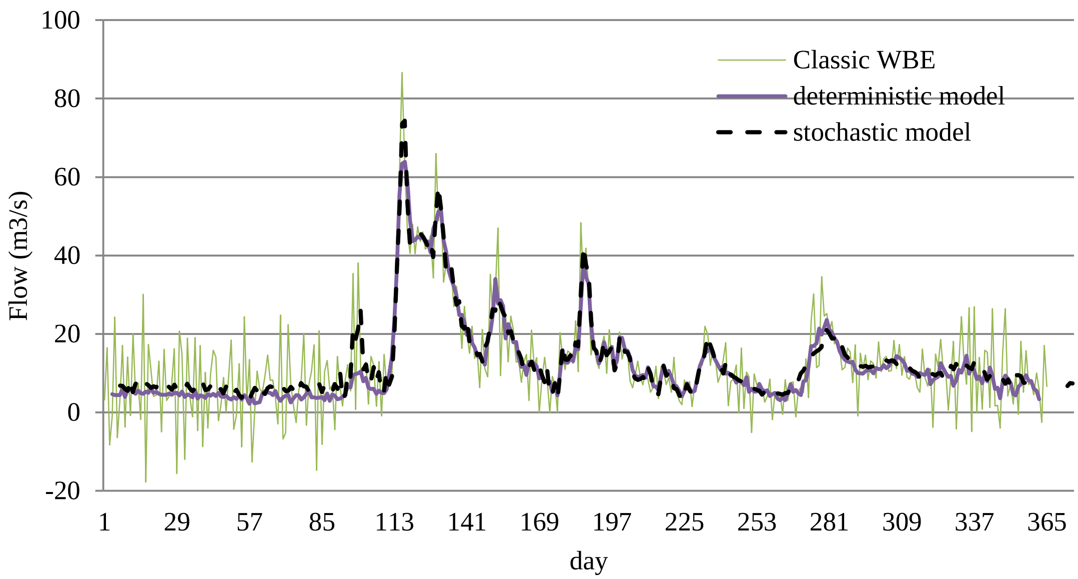


Figure 7-12: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 3-2009)

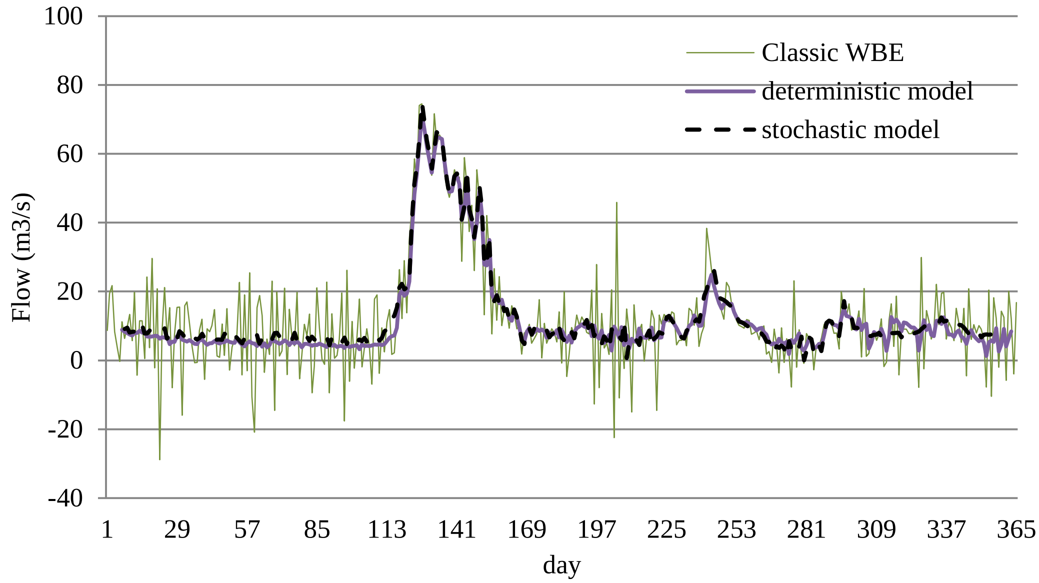


Figure 7-13: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 3-2011)

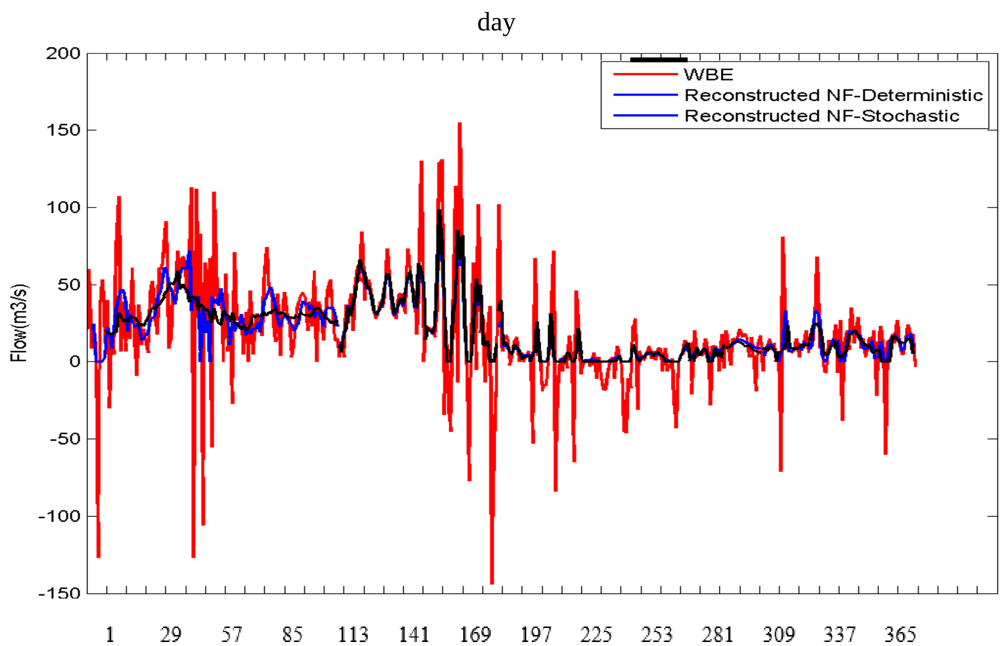


Figure 7-14: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 2-1990)

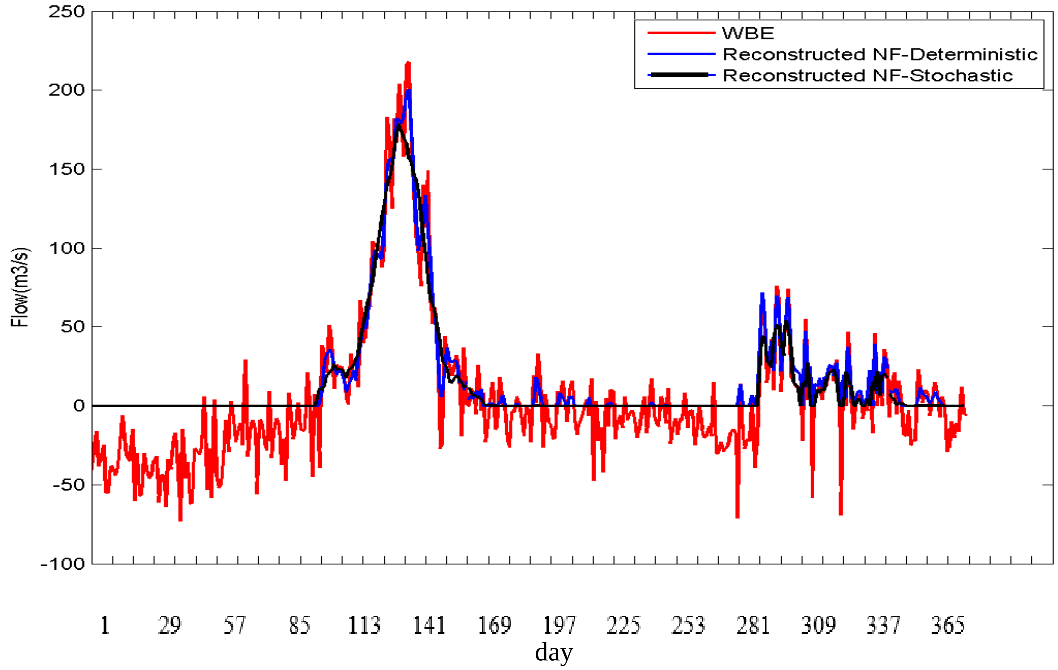


Figure 7-15: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 2-1991)

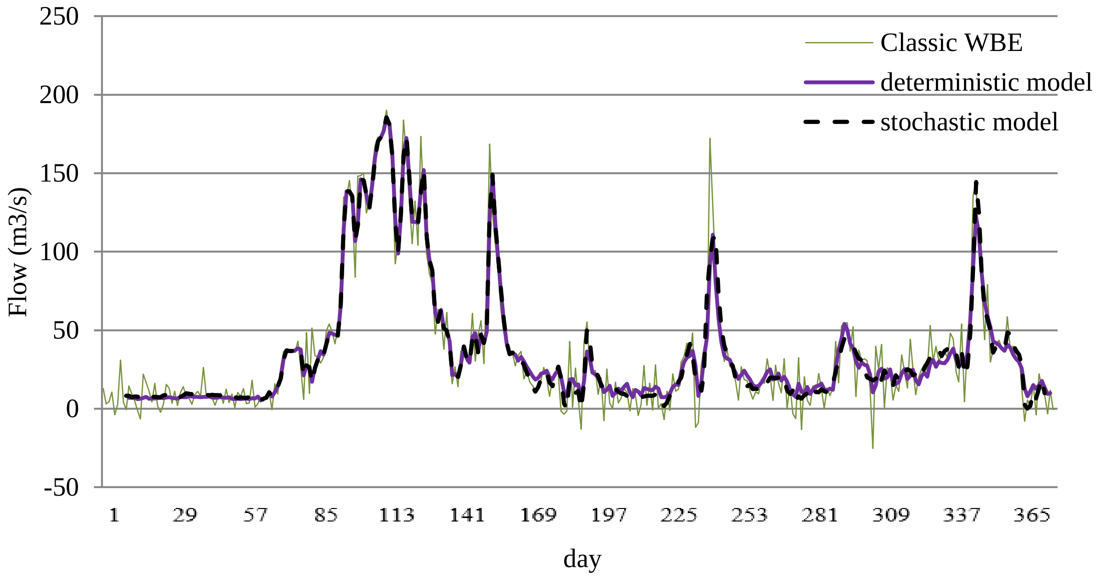


Figure 7-16: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 2-2005)

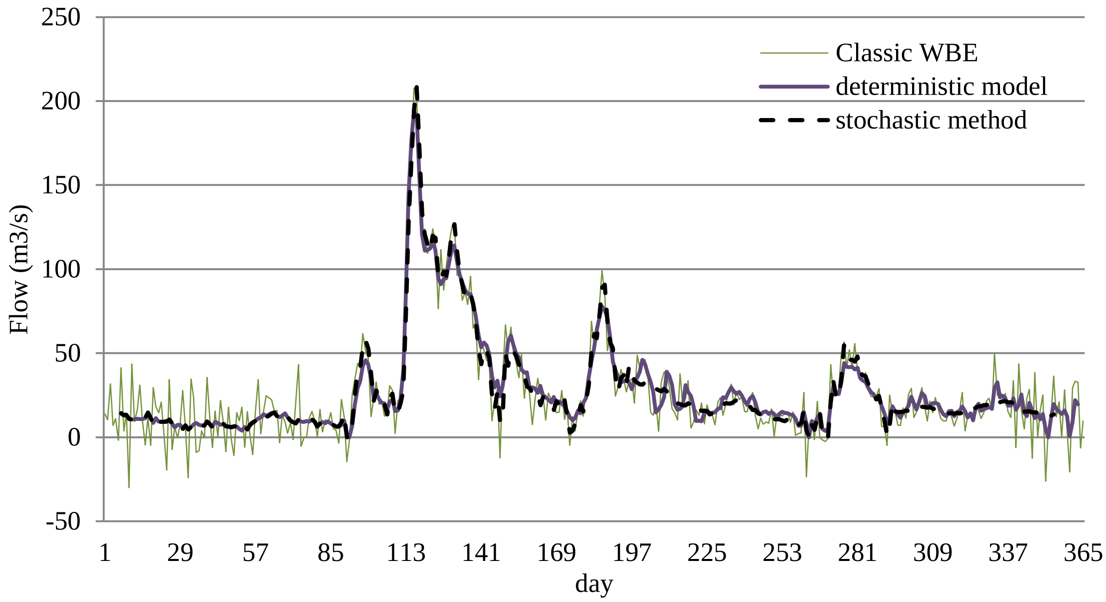


Figure 7-17: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 2-2009)

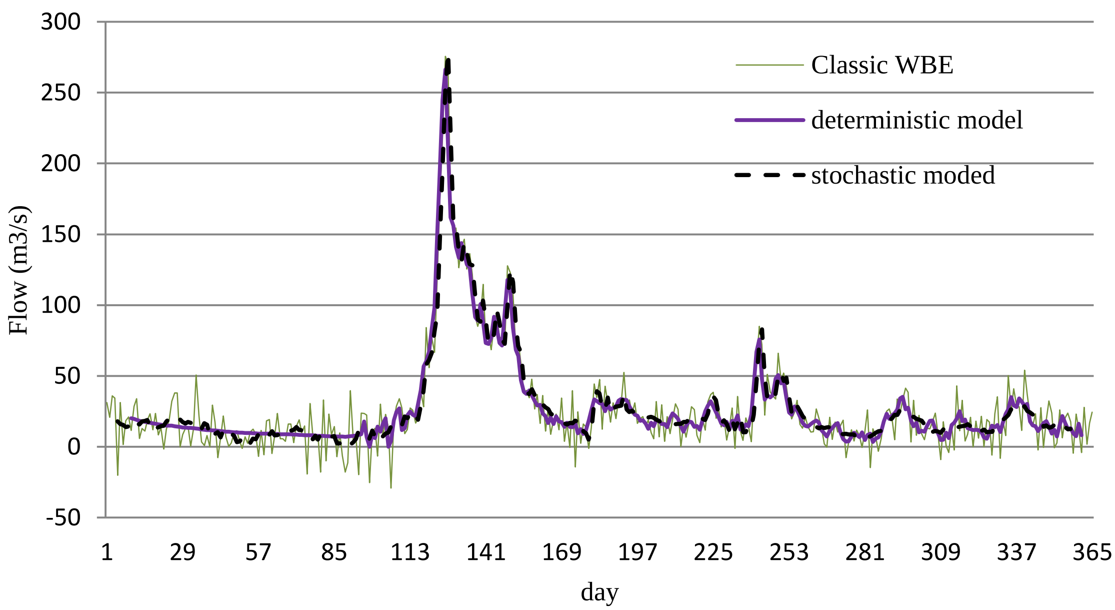


Figure 7-18: Comparison of deterministic based model, stochastic based model, and classic WBE (Outardes 2-2011)

7.2.2.3 Results discussion

The reconstructed flow series for Outardes 4, Outardes 3, and Outardes 2 for the chosen sampling years are presented in Figures 7-6 to 7-18. In these figures, the calculated flow using classic WBE is compared to reconstructed flow using the developed Deterministic and Stochastic-based optimization models. As shown, both the Stochastic and Deterministic based methods could efficiently improve the estimated flow when measured against the classic WBE. Visual comparison shows that:

- Reconstructed flow data series do not include negative values.
- Reconstructed flow data series are much smoother than classic WBE.
- Deterministic method sometimes underestimates or overestimates the flow (see Figure 7-10, the areas marked with circle) in comparison to Stochastic method. Although there is always ten other graphs which could be replaced with the first one in Deterministic method, they might exhibit noisier data (see Figure 7-10, the areas marked with circle).
- For the years after 2005, when reliable input data are available, the results of the Stochastic and Deterministic models are very similar and smooth.
- For the years prior to 2005, when input data are more uncertain, the developed optimization model could, to a large extent, compensate for input data uncertainty. For most years in this period (pre-2005), the reconstructed hydrograph is smooth and exhibit non-negative values, with acceptable graphical shape.
- For the few sample years where input data included unrealistic values for several time periods (days) in row, the calculated daily flow values using classic WBE were not acceptable for many of the chosen periods of the sample year. For example, winter flow series calculated by classic WBE show unreasonably high and low flows in Figures 7-14 and 7-15 respectively. For these years, the Deterministic and Stochastic methods failed to work. The flow should be estimated with other methods (such as manual flow validation) during these periods. This task is out of the scope of this current project.

7.2.3 Evaluating the Quality of Reconstructed Flow

7.2.3.1 Quality indexes

A more precise evaluation of reconstructed flow is possible using different QIs. Five quality indexes —*NN, CC, NAVE, SFR*, and *NT*— are intended to grade the calibre of the reconstructed flow. The results of applying these QIs on three of the flow reconstruction methods (classic WBE, Deterministic WBE, and Stochastic WBE) for the three Outardes sub-basins are summarized in Table 7.1 and Figures 7-19 to 7-25. Analysing these tables, the figures showed that:

- Classic WBE has consistently lower QIs than the other two methods (this lower quality is also clear in Figures 7-6 to 7-18, where reconstructed flows using Deterministic and Stochastic models are non-negative and are much smoother than classic WBE).
- Both Deterministic and Stochastic methods have satisfactory performance.
- The annual simple average QIs of both methods is more or less the same (see Figures 7-19 to 7-21). Referring to these figures, the average QI does not have a sharp change from one year to another one. The low average QI for the sampling years of 1990 and 1991 for the Outardes sub-basins are because of the unreliability of the input data and the failure of developed models to reconstruct the flow (see also Figures 7-14 and 7-15).
- Both Deterministic and Stochastic techniques have lower QI values for Outardes 3 in comparison with Outardes 4 and Outardes 2. This might be related to the small dimensions of the reservoir and its location downstream of a big reservoir. This reservoir is significantly affected by any changes in released flow from Outardes 4, even little ones. However, the values of Improvement Ratio (IR, the improvement achieved via applying the developed model) is considerable in Outardes 3. The Improvement Ratio is defined by Equation 7.1 as follow:

$$IR = \frac{\overline{QI}_{Deterministic\ or\ Stochastic\ model} - \overline{QI}_{WBE}}{\overline{QI}_{Deterministic\ or\ Stochastic\ model}} \times 100 \tag{7.1}$$

- Both Deterministic and Stochastic methods have almost the same average *NAVE* for the three sub-basins. This means that both methods have almost the same ability in estimating the flow trend.
- The average Improvement Ratio of *NN* is almost the same for the two Deterministic and Stochastic methods for Outardes 4 and Outardes 2. But for Outardes 3, the average Improvement Ratio of *NN* is higher for the Deterministic method, showing that this method takes meteorological factors more into account (*NN* compared the reconstructed flow with flow from RR model and RR model was developed based on meteorological data) than the Stochastic method for this basin (see also Figures 7-22 to 7-24).
- The average Improvement Ratio of *CC* is almost the same for both the Deterministic and Stochastic methods for Outardes 4, but the Deterministic method works better for Outardes 2 and Outardes 3 in producing flow with the most similar variation to its neighbouring basin (also see Figures 7-22 to 7-24).
- The average Improvement Ratio of *NT* is almost the same for the Deterministic and Stochastic methods for the three sub-basins. Please note that unlike the years after 2005, *NT* values for pre-2005 are not scaled. This is why the values of this criterion are less for the pre-2005 years, in comparison with the years after that. Thus, their values cannot be compared for the two time periods or for different basins.
- Average *SFR* of the Stochastic method is always better than the *SFR* of the Deterministic method. This means that the Stochastic method is more successful at closing the annual water budget.
- Comparing the specific flow of different methods (Pre-R and Post-R periods) shows that the specific flow of Deterministic and Stochastic methods are almost similar in all the three basins. Since a) the flow reconstruction methods used for Pre-R are less accurate than flow reconstruction methods applied for Post-R and b) the presence of reservoir affect the value of specific flow (SF), a difference between SFs for the two periods was expected. The difference between specific flows of neighbouring basins were accepted as the SF in each basin depended on different factors such as slope, soil characteristics, shape, and area of that basin.

Table 7.1: Different QIs for three methods of flow reconstruction for Post-R in Outardes 4, Outardes 3, and Outardes 2 (part 1)

		Deterministic WBE model						Stochastic WBE model						Classic WBE		
		O 4	IR (%)	O 3	IR (%)	O 2	IR (%)	O 4	IR (%)	O 3	IR (%)	O 2	IR (%)	O 4	O 3	O 2
NN	1990	0.763	15.237	0.128	59.845	0.581	67.177	0.752	13.944	0.129	60.223	0.599	68.158	0.647	0.051	0.191
CC		0.519	13.475	0.400	15.184	0.479	0.620	0.521	13.791	0.392	17.332	0.458	5.272	0.449	0.460	0.482
NT		0.733	48.639	0.877	11.161	0.861	9.482	0.743	49.361	0.871	10.582	0.918	15.129	0.376	0.779	0.779
NAVE				0.561		0.179		0.861		0.557		0.196				
SFR				0.924		-1.392		0.970		0.905		-0.899				
average		0.753	25.784	0.578	28.730	0.142	25.760	0.770	25.699	0.571	29.379	0.254	29.519	0.491	0.430	0.484
NN	1991	0.918	19.071	0.631	63.902	0.717	25.147	0.928	20.014	0.555	58.984	0.735	26.922	0.743	0.228	0.537
CC		0.570	20.676	0.481	8.874	0.435	5.184	0.546	17.169	0.477	8.147	0.443	3.353	0.452	0.438	0.458
NT		0.723	46.570	0.907	5.579	0.856	2.497	0.735	47.432	0.890	3.751	0.913	8.569	0.387	0.856	0.835
NAVE				0.602		0.106		0.848		0.596		0.093				
SFR				0.883		-1.761		0.992		0.877		-1.476				
average		0.798	28.772	0.701	26.118	0.071	10.943	0.810	28.205	0.679	23.627	0.142	12.948	0.527	0.508	0.610
NN	1992	0.890	10.567	0.528	34.663	0.800	17.657	0.880	9.503	0.490	29.514	0.748	11.896	0.796	0.345	0.659
CC		0.535	9.404	0.491	3.512	0.492	3.567	0.532	8.866	0.480	1.202	0.514	7.848	0.485	0.474	0.474
NT		0.718	46.149	0.897	0.683	0.857	1.742	0.671	42.447	0.881	1.096	0.890	5.378	0.386	0.891	0.842
NAVE				0.746		0.721		0.882		0.735		0.762				
SFR				0.921		0.851		0.985		0.907		0.913				
average		0.780	22.040	0.717	12.952	0.744	7.655	0.790	20.272	0.699	10.604	0.765	8.374	0.556	0.570	0.658
NN	1993	0.949	13.900	0.437	57.511	0.650	10.581	0.933	12.455	0.370	49.800	0.668	12.956	0.817	0.186	0.581
CC		0.557	8.933	0.489	9.177	0.423	8.890	0.562	9.834	0.405	9.649	0.463	0.623	0.507	0.444	0.460
NT		0.697	42.866	0.905	5.467	0.851	1.143	0.737	45.964	0.884	3.126	0.889	3.155	0.398	0.856	0.861
NAVE				0.696		0.736		0.891		0.686		0.767				
SFR				0.901		0.853		0.993		0.892		0.904				
average		0.806	21.900	0.686	24.052	0.703	6.871	0.823	22.751	0.647	20.858	0.738	5.578	0.574	0.495	0.634
NN	1994	0.901	7.176	0.652	34.176	0.735	6.804	0.887	5.717	0.591	27.402	0.756	9.442	0.837	0.429	0.685
CC		0.570	16.349	0.496	7.262	0.462	6.146	0.535	10.868	0.460	0.114	0.466	5.302	0.477	0.460	0.490
NT		0.790	30.768	0.910	0.888	0.849	2.335	0.788	30.550	0.888	1.520	0.887	2.094	0.547	0.902	0.869
NAVE				0.768		0.773		0.925		0.761		0.795				
SFR				0.858		0.863		0.995		0.910		0.892				
average		0.821	18.098	0.737	14.109	0.736	5.095	0.826	15.712	0.722	9.679	0.759	5.613	0.620	0.597	0.681

Table 7.1: Different QIs for three methods of flow reconstruction for Post-R in Outardes 4, Outardes 3, and Outardes 2 (part 2)

		Deterministic WBE model						Stochastic WBE model						Classic WBE		
		O 4	IR (%)	O 3	IR (%)	O 2	IR (%)	O 4	IR (%)	O 3	IR (%)	O 2	IR (%)	O 4	O 3	O 2
NN	1995	0.842	5.347	0.774	14.866	0.715	0.117	0.837	4.771	0.759	13.215	0.737	3.063	0.797	0.659	0.714
CC		0.626	19.447	0.519	7.660	0.504	8.095	0.598	15.644	0.495	3.088	0.530	12.588	0.504	0.479	0.463
NT		0.800	31.179	0.910	1.633	0.859	1.999	0.820	32.882	0.896	3.246	0.894	2.024	0.550	0.925	0.876
NAVE				0.782		0.776		0.912		0.772		0.806				
SFR				0.905		0.875		0.985		0.902		0.918				
average		0.819	18.657	0.778	8.053	0.746	3.404	0.830	17.766	0.765	6.516	0.777	5.891	0.617	0.688	0.684
NN	1996	0.575	15.335	0.780	4.135	0.722	0.249	0.556	19.295	0.753	0.668	0.727	0.512	0.663	0.748	0.723
CC		0.482	6.291	0.588	18.919	0.509	4.674	0.502	2.039	0.520	8.274	0.522	7.104	0.512	0.477	0.485
NT		0.753	20.088	0.911	2.607	0.837	3.435	0.800	24.770	0.892	4.824	0.876	1.197	0.602	0.935	0.866
NAVE				0.796		0.786		0.903		0.790		0.812				
SFR				0.900		0.857		0.896		0.910		0.927				
average		0.717	13.905	0.795	8.554	0.742	2.786	0.731	15.368	0.773	4.589	0.773	2.937	0.592	0.720	0.691
NN	1997	0.904	1.358	0.886	6.079	0.821	1.379	0.916	0.064	0.846	1.738	0.824	1.744	0.916	0.832	0.810
CC		0.591	17.047	0.542	16.110	0.469	0.775	0.576	14.825	0.510	10.776	0.494	5.693	0.490	0.455	0.466
NT		0.808	22.180	0.920	1.891	0.865	3.227	0.810	22.377	0.900	4.188	0.902	1.032	0.629	0.938	0.893
NAVE				0.778		0.761		0.933		0.773		0.802				
SFR				0.893		0.851		0.947		0.906		0.921				
average		0.825	13.528	0.804	8.026	0.754	1.794	0.836	12.422	0.787	5.567	0.789	2.823	0.678	0.741	0.723
NN	1998	0.871	8.694	0.784	59.112	0.827	11.358	0.821	3.221	0.733	56.271	0.810	9.473	0.795	0.321	0.733
CC		0.501	2.044	0.486	3.065	0.428	7.638	0.532	7.836	0.447	5.361	0.420	9.678	0.490	0.471	0.460
NT		0.773	18.757	0.908	1.121	0.876	2.630	0.754	16.710	0.899	2.114	0.903	0.488	0.628	0.918	0.899
NAVE				0.729		0.745		0.941		0.695		0.786				
SFR				0.915		0.863		0.996		0.880		0.923				
average		0.803	9.832	0.765	21.099	0.748	7.209	0.809	9.256	0.731	21.249	0.768	6.546	0.638	0.570	0.697
NN	1999	0.926	1.250	0.841	43.753	0.672	6.817	0.930	1.656	0.699	32.317	0.646	3.107	0.915	0.473	0.626
CC		0.533	2.778	0.565	13.207	0.506	8.000	0.529	2.186	0.485	1.171	0.507	8.074	0.518	0.490	0.466
NT		0.749	23.625	0.917	2.366	0.868	2.365	0.747	23.429	0.898	0.264	0.907	2.075	0.572	0.896	0.888
NAVE				0.651		0.772		0.937		0.671		0.811				
SFR				0.899		0.849		0.991		0.902		0.933				
average		0.815	9.218	0.775	19.775	0.733	5.727	0.827	9.090	0.731	11.251	0.761	4.419	0.668	0.620	0.660
NN	2000	0.911	32.483	0.783	19.367	0.784	13.411	0.890	30.928	0.740	14.734	0.741	8.401	0.615	0.631	0.679
CC		0.567	16.923	0.542	11.056	0.469	0.393	0.576	18.156	0.495	2.534	0.486	3.077	0.471	0.482	0.471
NT		0.734	26.364	0.917	0.462	0.864	0.743	0.826	34.520	0.896	2.786	0.891	2.237	0.541	0.921	0.871
NAVE				1.335		0.749		0.908		1.311		0.784				
SFR				0.908		0.837		0.947		0.912		0.933				
average		0.812	25.256	0.897	10.295	0.741	4.849	0.829	27.868	0.871	6.685	0.767	4.572	0.542	0.678	0.674

Table 7.1: Different QIs for three methods of flow reconstruction for Post-R in Outardes 4, Outardes 3, and Outardes 2 (part 3)

		Deterministic WBE model						Stochastic WBE model						Classic WBE		
		O 4	IR (%)	O 3	IR (%)	O 2	IR (%)	O 4	IR (%)	O 3	IR (%)	O 2	IR (%)	O 4	O 3	O 2
NN	2001	0.774	6.443	0.691	22.071	0.761	10.717	0.770	5.888	0.642	16.059	0.759	10.575	0.724	0.539	0.679
CC		0.554	16.904	0.537	14.804	0.452	6.636	0.521	11.689	0.500	8.443	0.496	2.868	0.460	0.458	0.482
NT		0.743	26.265	0.915	1.548	0.855	3.822	0.791	30.727	0.893	4.073	0.906	2.040	0.548	0.930	0.888
NAVE				1.327		0.778		0.922		1.302		0.788				
SFR				0.925		0.877		0.991		0.912		0.921				
average		0.786	16.537	0.879	12.808	0.744	7.058	0.799	16.101	0.850	9.525	0.774	5.161	0.578	0.642	0.683
NN	2002	0.855	0.658	0.675	17.366	0.846	8.777	0.706	20.258	0.640	12.882	0.845	8.726	0.849	0.558	0.771
CC		0.615	18.943	0.514	7.811	0.447	2.907	0.527	5.321	0.480	1.202	0.458	0.487	0.499	0.474	0.460
NT				0.918		0.869		0.799		0.896		0.899		0.589	0.925	0.887
NAVE				1.333		0.743		0.934		1.308		0.768				
SFR				0.903		0.871		0.888		0.887		0.923				
average		0.826	14.929	0.868	8.655	0.755	4.598	0.771	17.306	0.842	5.777	0.779	3.519	0.646	0.652	0.706
NN	2003	0.863	7.990	0.666	31.914	0.672	7.503	0.840	5.404	0.596	23.988	0.678	8.377	0.794	0.453	0.622
CC		0.634	18.733	0.512	5.210	0.484	5.258	0.548	6.092	0.497	2.473	0.473	7.645	0.515	0.485	0.510
NT		0.813	25.491	0.909	0.555	0.855	2.758	0.797	23.989	0.887	3.074	0.888	1.067	0.606	0.914	0.878
NAVE				1.325		0.762		0.928		1.301		0.798				
SFR				0.904		0.839		0.954		0.907		0.911				
average		0.831	17.405	0.863	12.560	0.722	5.173	0.814	11.828	0.838	9.845	0.750	5.696	0.638	0.618	0.670
NN	2004	0.969	1.272	0.721	30.510	0.773	4.337	0.988	0.664	0.651	23.058	0.764	3.235	0.981	0.501	0.739
CC		0.663	21.497	0.514	6.745	0.509	2.327	0.551	5.563	0.495	3.088	0.502	3.789	0.521	0.479	0.521
NT		0.812	29.701	0.913	0.068	0.860	3.463	0.774	26.274	0.898	1.553	0.900	1.099	0.570	0.912	0.890
NAVE			0.658	1.322		0.772	8.777	0.942		1.298		0.790		0.849		
SFR			18.943	0.832		0.864	2.907	0.992		0.848		0.922		0.499		
average		0.867		0.860	12.441	0.755		0.849	10.834	0.838	9.233	0.775	2.708	0.589	0.631	0.717
NN	2005	0.885	7.4	0.821	31.4	0.828	14.0	0.860	4.4	0.798	27.7	0.806	11.0	0.824	0.625	0.726
CC		0.682	23.1	0.635	20.5	0.641	9.8	0.617	11.4	0.611	15.9	0.617	5.7	0.554	0.527	0.584
NAVE		0.900		0.808		0.801		0.905		0.661		0.797				
SFR		0.972		0.916		0.995		0.999		0.934		0.999				
NT		0.772	33.3	0.491	109.8	0.551	68.5	0.759	31.1	0.507	116.7	0.542	65.7	0.579	0.234	0.327
average		0.842		0.734		0.763		0.828		0.702		0.752				
NN	2007	0.864	7.2	0.841	114.0	0.731	28.0	0.852	5.7	0.744	89.3	0.806	41.2	0.806	0.393	0.571
CC		0.627	13.4	0.570	11.3	0.638	25.8	0.660	19.3	0.575	12.3	0.617	21.7	0.553	0.512	0.507
NAVE		0.888		0.812		0.713		0.889		0.631		0.797				
SFR		0.969		0.921		0.984		0.999		0.939		0.999				
NT		0.819	33.2	0.623	181.9	0.711	88.1	0.819	33.2	0.669	202.7	0.748	97.9	0.615	0.221	0.378
average		0.834		0.753		0.755		0.844		0.712		0.793				

Table 7.1: Different QIs for three methods of flow reconstruction for Post-R in Outardes 4, Outardes 3, and Outardes 2 (part 4)

		Deterministic WBE model						Stochastic WBE model						Classic WBE		
		O 4	IR (%)	O 3	IR (%)	O 2	IR (%)	O 4	IR (%)	O 3	IR (%)	O 2	IR (%)	O 4	O 3	O 2
NN	2008	0.808	2.0	0.650	33.2	0.875	16.7	0.801	1.1	0.676	38.5	0.811	8.1	0.792	0.488	0.750
CC		0.645	19.2	0.631	19.1	0.648	15.7	0.642	18.7	0.631	19.1	0.639	14.1	0.541	0.530	0.560
NAVE		0.920		0.884		0.783		0.924		0.700		0.773				
SFR		0.968		0.921		0.984		0.999		0.940		0.998				
NT		0.773	26.1	0.525	162.5	0.509	70.8	0.755	23.2	0.637	218.5	0.536	79.9	0.613	0.200	0.298
NN	2009	0.880	5.9	0.808	52.5	0.899	22.1	0.862	3.7	0.754	42.3	0.843	14.5	0.831	0.530	0.736
CC		0.611	13.1	0.616	22.2	0.616	14.7	0.622	15.2	0.636	26.2	0.605	12.7	0.540	0.504	0.537
NAVE		0.901		0.873		0.759		0.907		0.693		0.758				
SFR		0.969		0.921		0.984		0.999		0.939		0.997				
NT		0.775	27.5	0.628	182.9	0.646	85.1	0.786	29.3	0.605	172.5	0.640	83.4	0.608	0.222	0.349
average		0.827		0.769		0.781		0.835		0.725		0.769				
NN	2010	0.519	10.7	0.773	80.6	0.714	29.8	0.499	6.4	0.733	71.3	0.664	20.7	0.469	0.428	0.550
CC		0.551	3.6	0.559	14.5	0.625	15.7	0.559	5.1	0.600	23.0	0.674	24.8	0.532	0.488	0.540
NAVE		0.901		0.808		0.719		0.911		0.657		0.706				
SFR		0.969		0.921		0.984		0.999		0.939		0.997				
NT		0.814	24.5	0.485	142.5	0.576	94.6	0.772	18.0	0.588	194.0	0.588	98.6	0.654	0.200	0.296
average		0.751		0.709		0.723		0.748		0.704		0.726				
NN	2011	0.916	1.9	0.883	54.9	0.840	15.1	0.913	1.6	0.838	47.0	0.845	15.8	0.899	0.570	0.730
CC		0.600	10.1	0.637	15.4	0.632	17.5	0.627	15.0	0.599	8.5	0.591	9.9	0.545	0.552	0.538
NAVE		0.936		0.864		0.736		0.929		0.660		0.733				
SFR		0.969		0.921		0.985		0.999		0.938		0.998				
NT		0.773	24.5	0.611	195.2	0.628	96.9	0.763	22.9	0.624	201.4	0.644	101.9	0.621	0.207	0.319
average		0.839		0.783		0.764		0.846		0.732		0.762				

Note : O2=Outardes 2, O3=Outardes 3, O4=Outardes 4

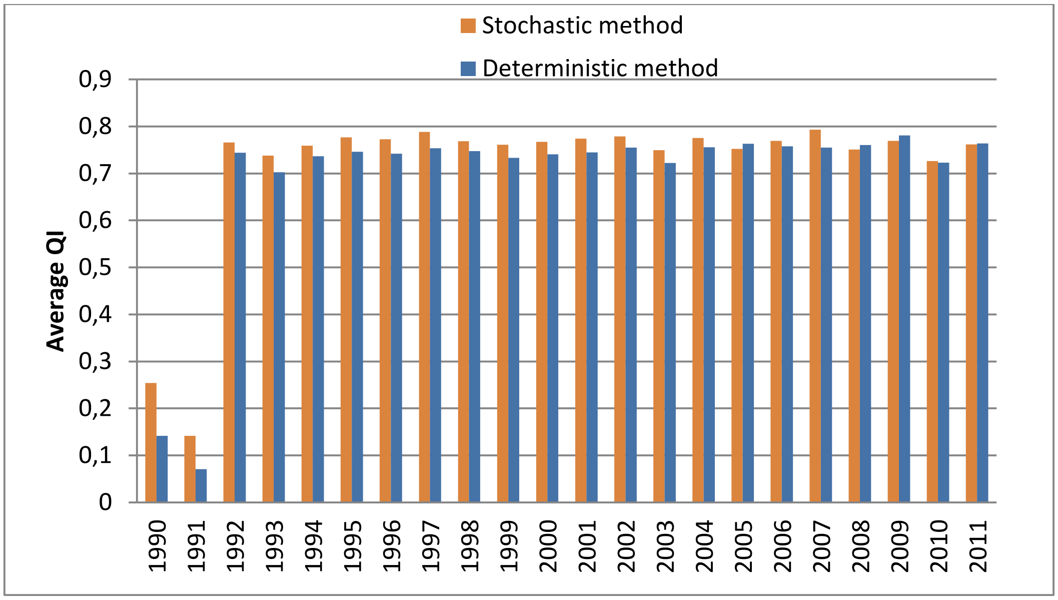


Figure 7-19: Annual average QI[12] for deterministic and stochastic methods (Outardes 2)

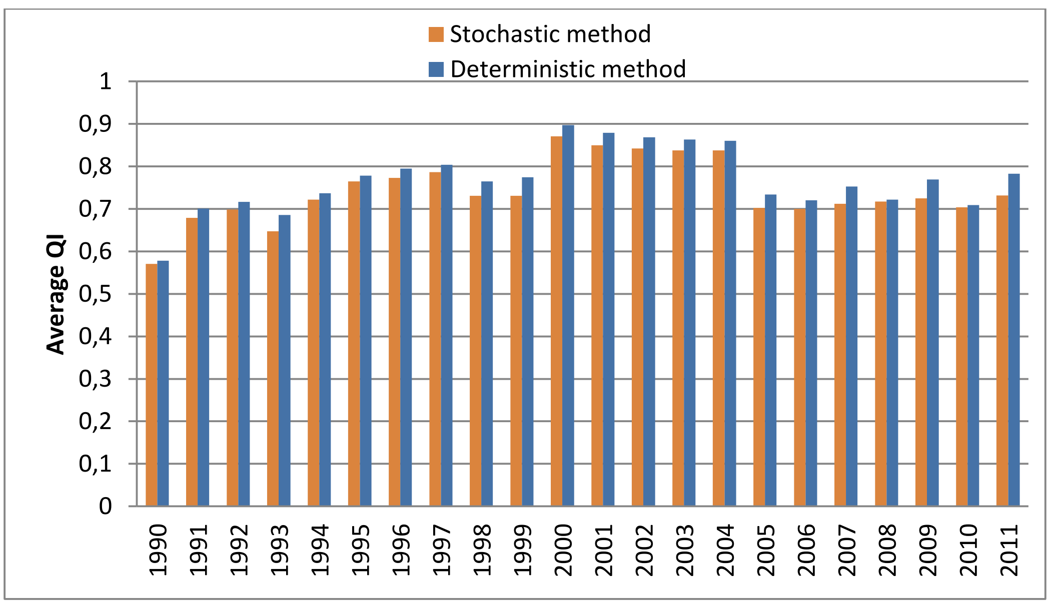


Figure 7-20: Annual average QI for deterministic and stochastic methods (Outardes 3)

[12] Simple average

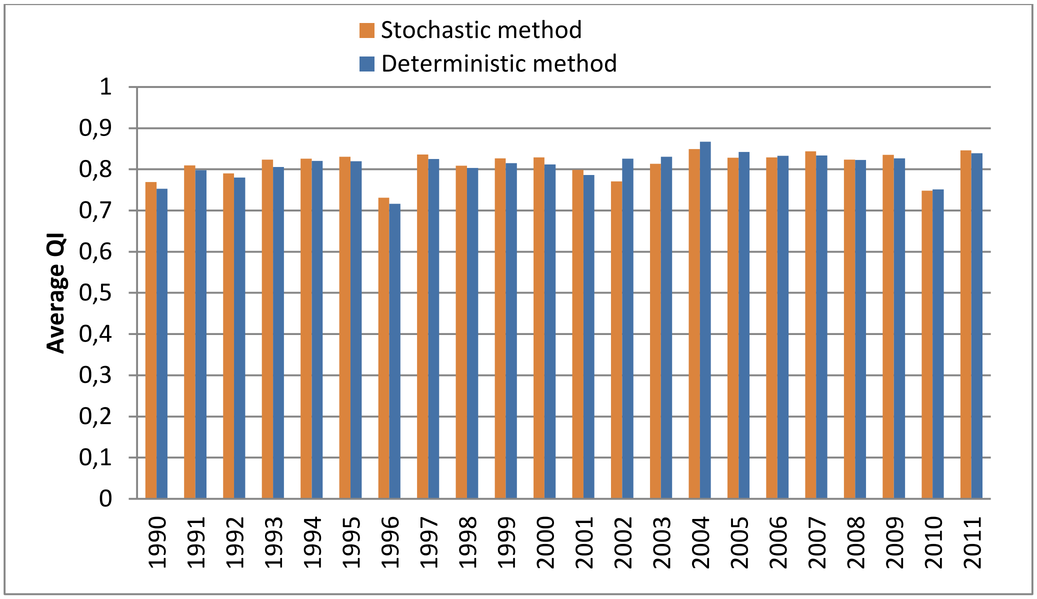


Figure 7-21: Annual average QI for deterministic and stochastic methods (Outardes 4)

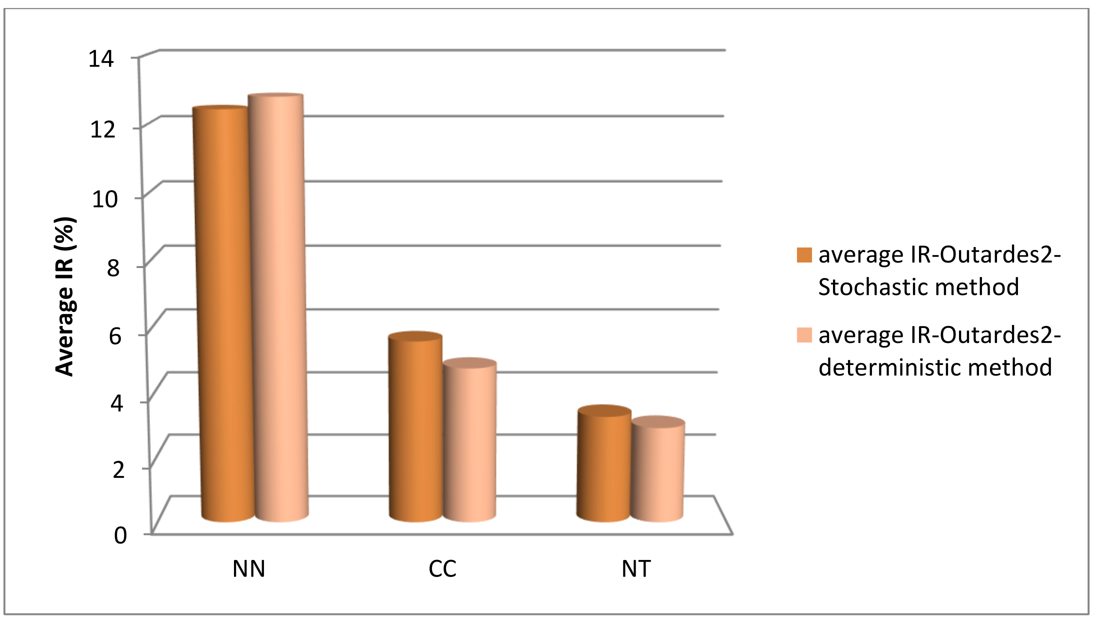


Figure 7-22: Average IR of *NN*, *CC*, and *NT* for deterministic and stochastic methods (Outardes 2)

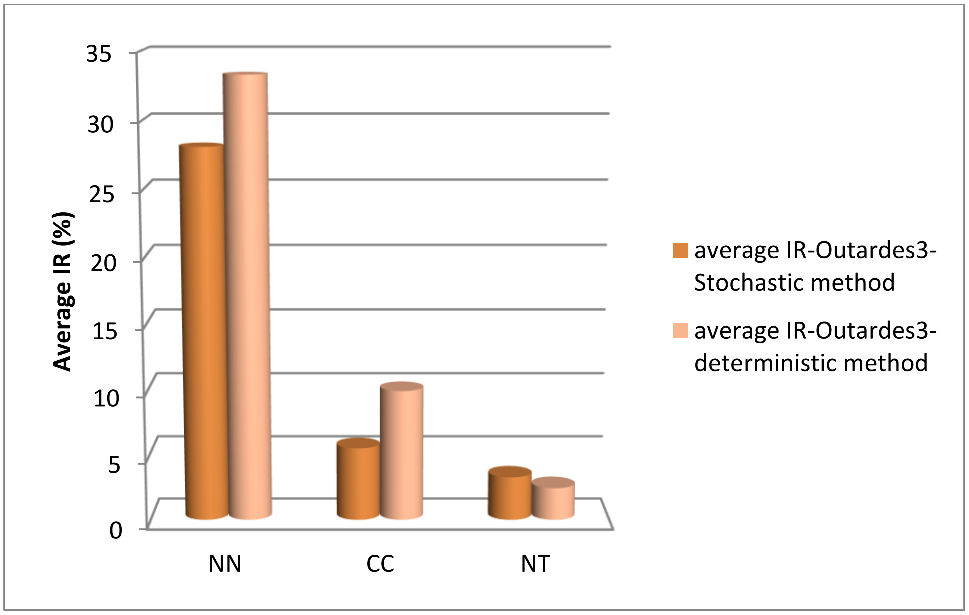


Figure 7-23: Average IR of *NN*, *CC*, and *NT* for deterministic and stochastic methods (Outardes 3)

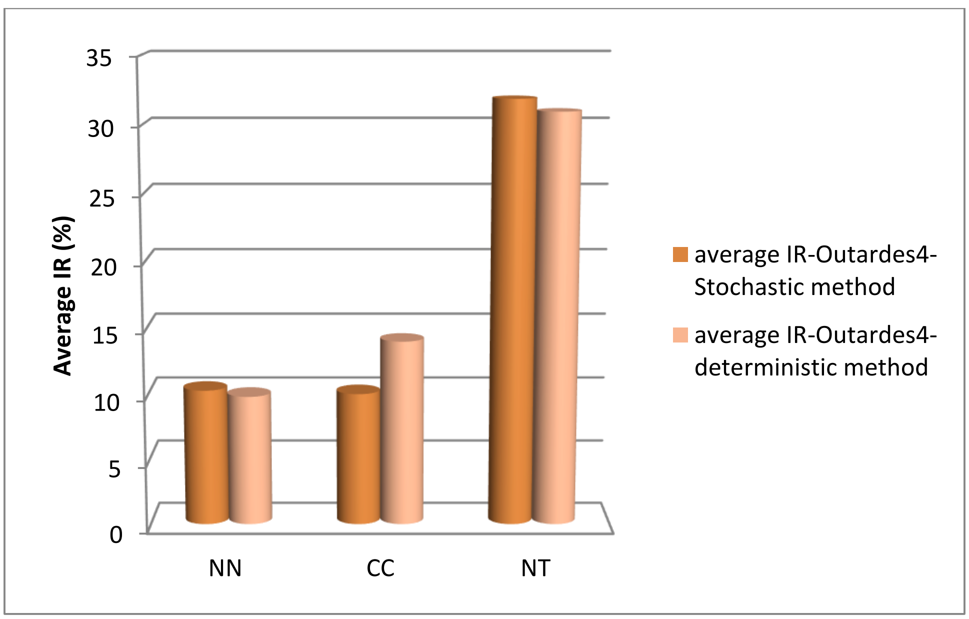


Figure 7-24: Average IR of *NN*, *CC*, and *NT* for deterministic and stochastic methods (Outardes 4)

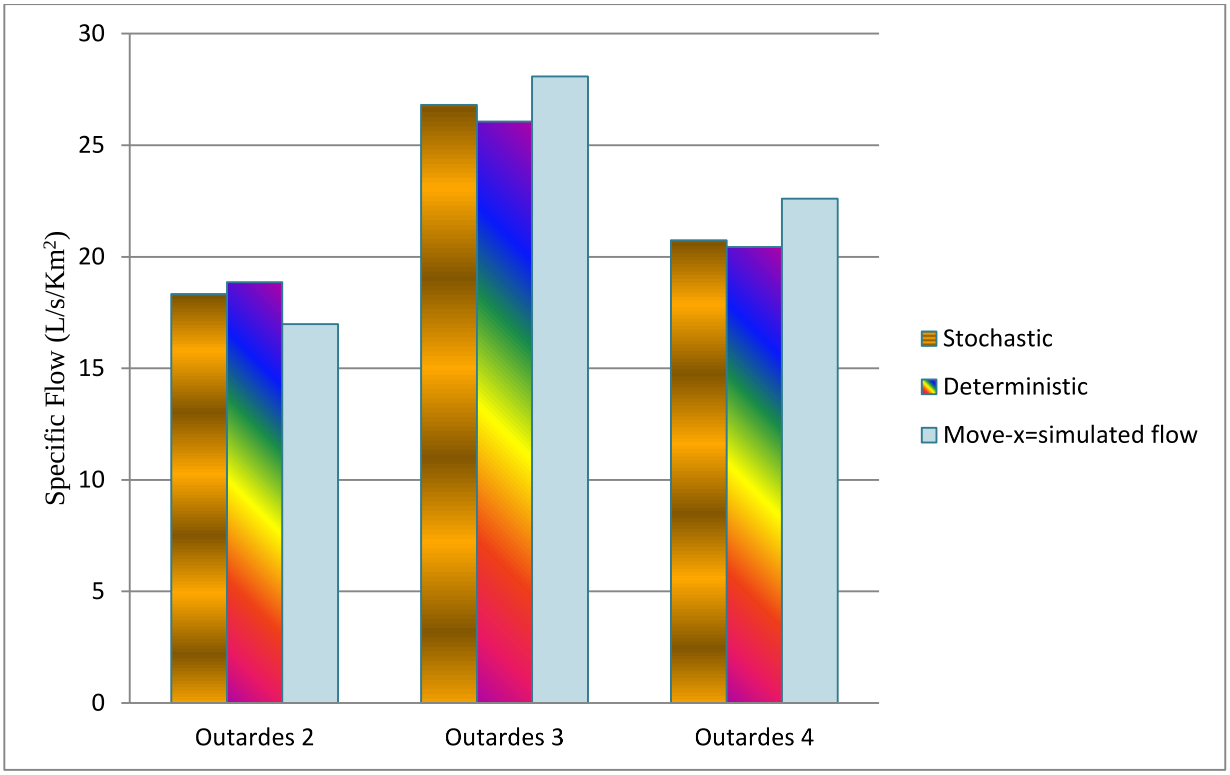


Figure 7-25: Comparison of specific flow calculated using different methods of flow reconstruction

7.2.3.2 Regional and temporal homogeneity

In order to accurately evaluate the quality of the reconstructed flow series in comparison to the regional flow, the histogram of the local reconstructed flow is first compared with the scaled regional histograms (based on surface ratio) as shown in Figure 7-26. The figure shows the scaled reconstructed flow of the three basins in this current case study (using Deterministic WBE) in comparison with those of the gauged basins for the sample year of 2007. The results show that local flow seems totally in line with regional flows: no considerable under- or over-estimations are detected in the graphs. This means that the reconstructed flow has also comparable *SF* to neighbouring gauged basins (as *SF* of a basin is equal to average annual flow on surface area of that basin).

Next, the scale location, Normal Q-Q plot, and residual versus fitted plots are graphed to evaluate regional homogeneity. To do so, a regression is developed between the basin surface area and quantiles (0.5, 0.80, 0.9, 0.95, 0.98, 0.99) of the basin's annual peak flows, after fitting a log normal distribution with three parameters as per the following equation (see also Equation 6.19):

$$FQ = 20.04s^{0.871} \ ; \ R^2 = 0.93$$

The scale location, Normal Q-Q plot, and residual versus fitted plots are then drawn (Figures 7-27 to 7-29) to complete the regional homogeneity analyses. In Figures 7-27 to 7-29, the obtained values from local frequency analysis of the three basins (Outardes 4, Outardes 3, and Outardes 2) are marked as local rivers. From these figures, it is observed that the reconstructed flow data series have comparable peak flows with its neighbouring basins. This means that the reconstructed flows values are considered as reliable data for local and regional frequency analyses because the three basins (or local rivers):

- do not disturb the linear relationship in a normal Q-Q plot (Figures 7-27).
- maintain the scattering in a scale-location plot (Figure 7-28) and does not cause any pattern in this plot.

- maintain the random scattering of residuals around zero and the constancy of residuals (Figure 7-29). This means that they do not cause residual increases or decreases in the fitted values in a pattern.

The local flow also appeared stationary using the KPSS test, which indicates that reconstructed flow has the same characteristics during this sampling time period (stationary flow). The stationarity of the reconstructed flow is valuable for frequency analyses.

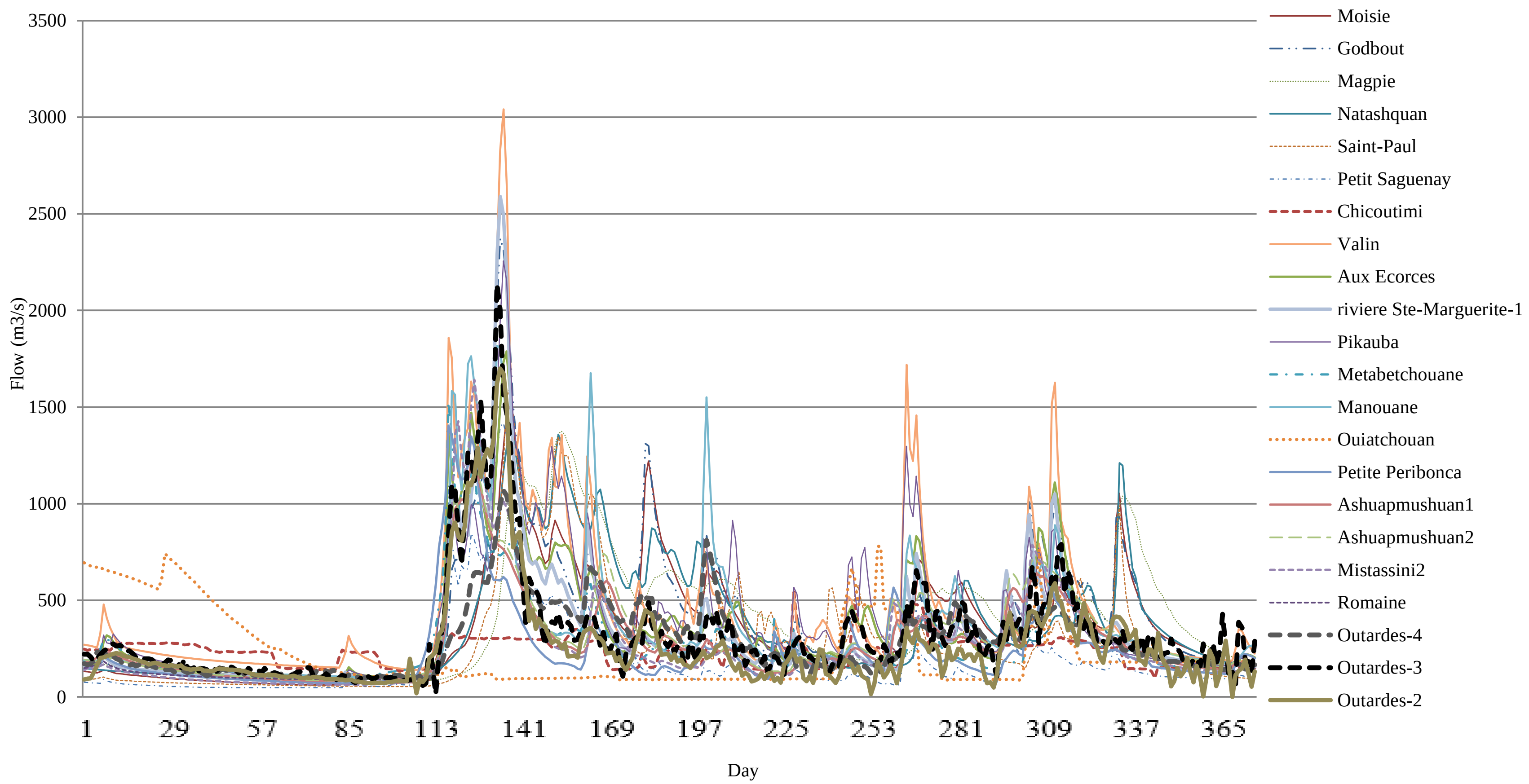


Figure 7-26: Comparing the hydrograph of scaled reconstructed flow for Outardes 4, 3, and 2 with regional flow data series

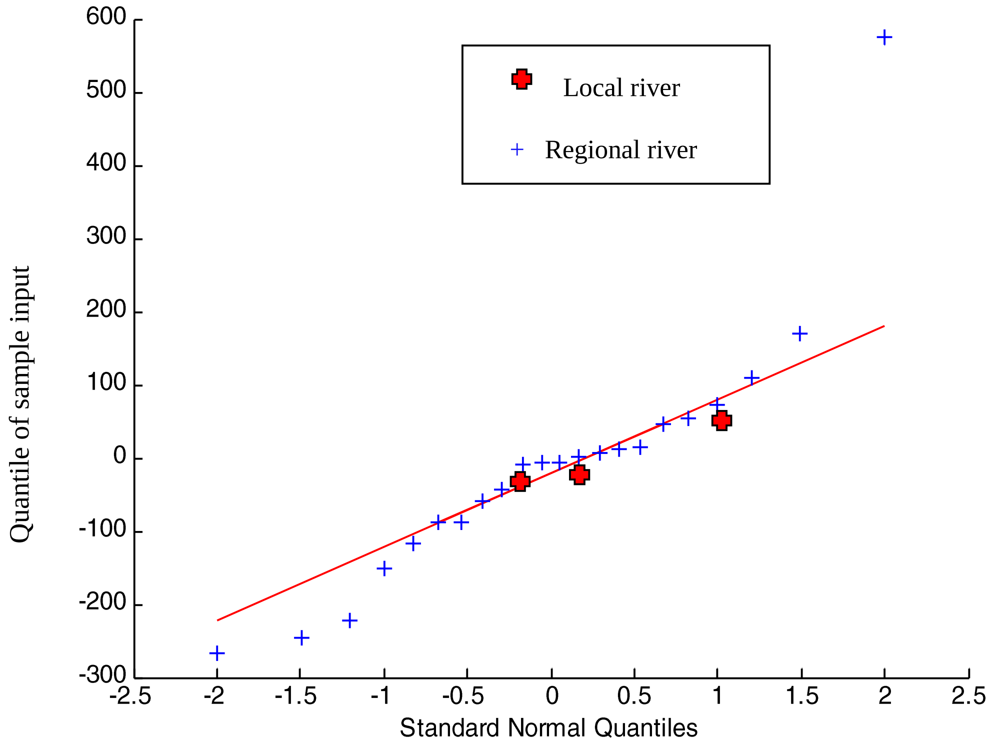


Figure 7-27: Normal Q-Q plot

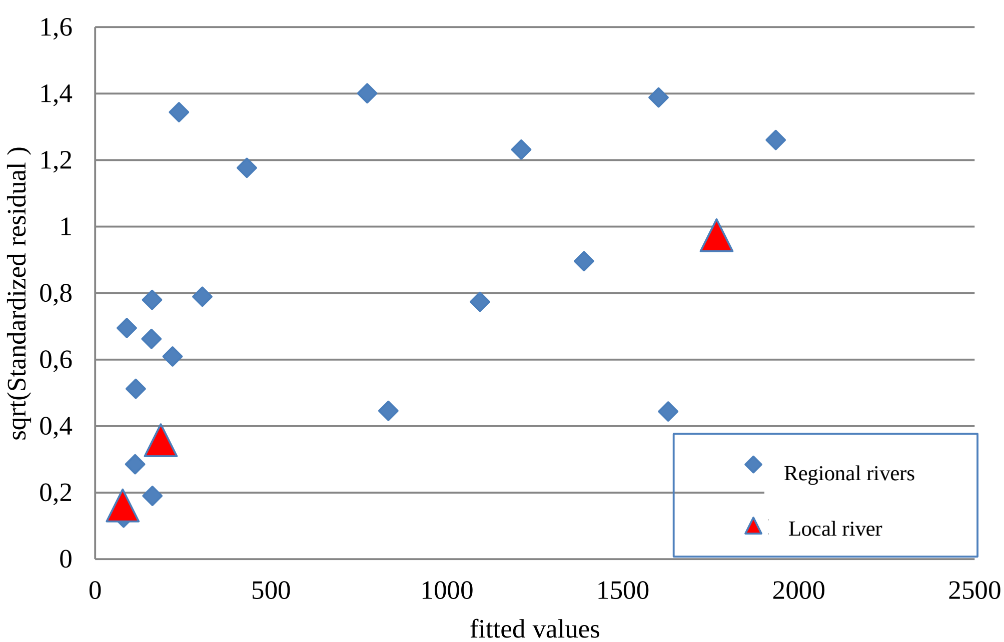


Figure 7-28: Scale location plot

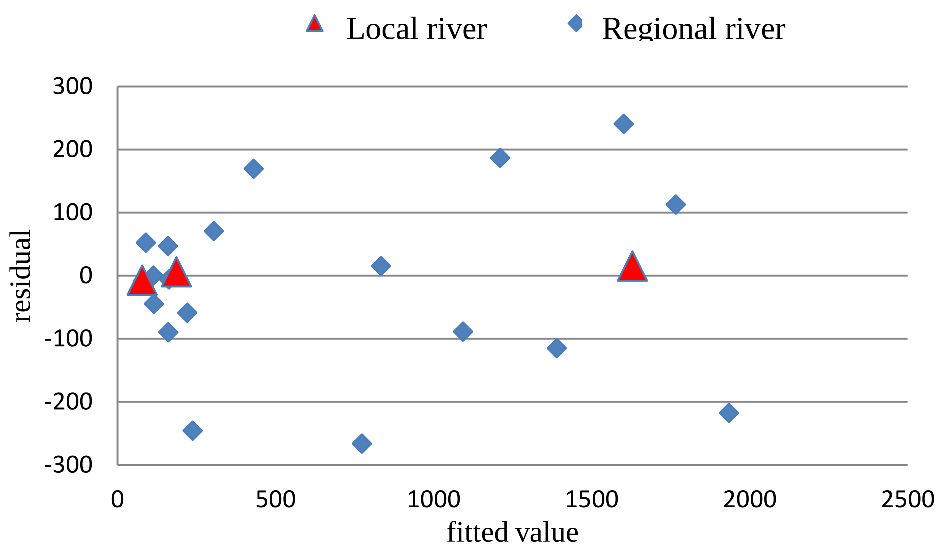
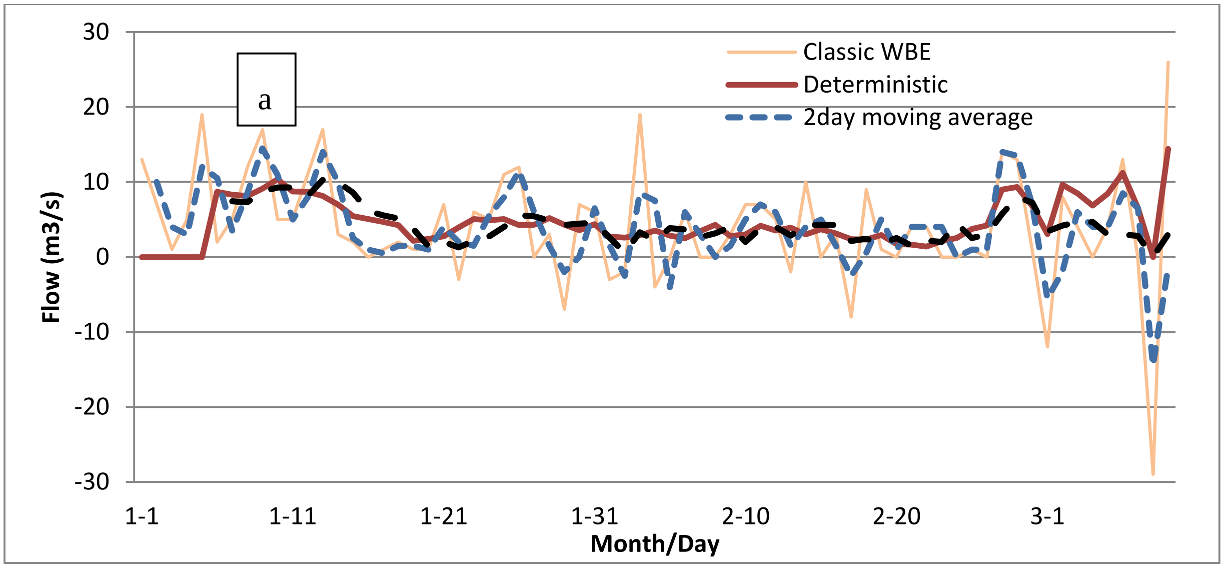


Figure 7-29: Residual versus fitted value plot

In the end, the following questions arise: Are the results of the recommended optimization model much different from that of simple average method? Is it worth it to apply an optimization model instead of using a simple method such as average method?

To answer these questions, the results of the Moving Average method are compared to those of the developed optimization model. The results of both 2-day and 7-day moving averages used on classic WBE and Deterministic optimization model for the winter and spring period using 2000 as the sample year (on Outardes 2) are presented in parts a and b of Figure 7-30 respectively. According to this figure, the 2-day moving average method still gives noisy results, while the effect of noise is noticeably less using the 7-day moving average method. However, the reliability of the 7-day moving average method decreases during local and general peak flows because it underestimates the peak flows. Since this model does not take into account the physics of flow dynamics, it disregards real peak flows and filter out them too. The 7-day moving average method causes up to 55% (10.5 m^3/s) underestimation for winter local peaks and up to 57% (82 m^3/s) underestimation for spring flows (for the example of Outardes 2 in 2000) in this example, compared to the WBE method. Also, a shift appeared in the results of the moving average method because it only uses 'past' data to calculate the flow of the sample day. To avoid this, a central moving average must be used, but it is not applicable for real-time data reconstruction.



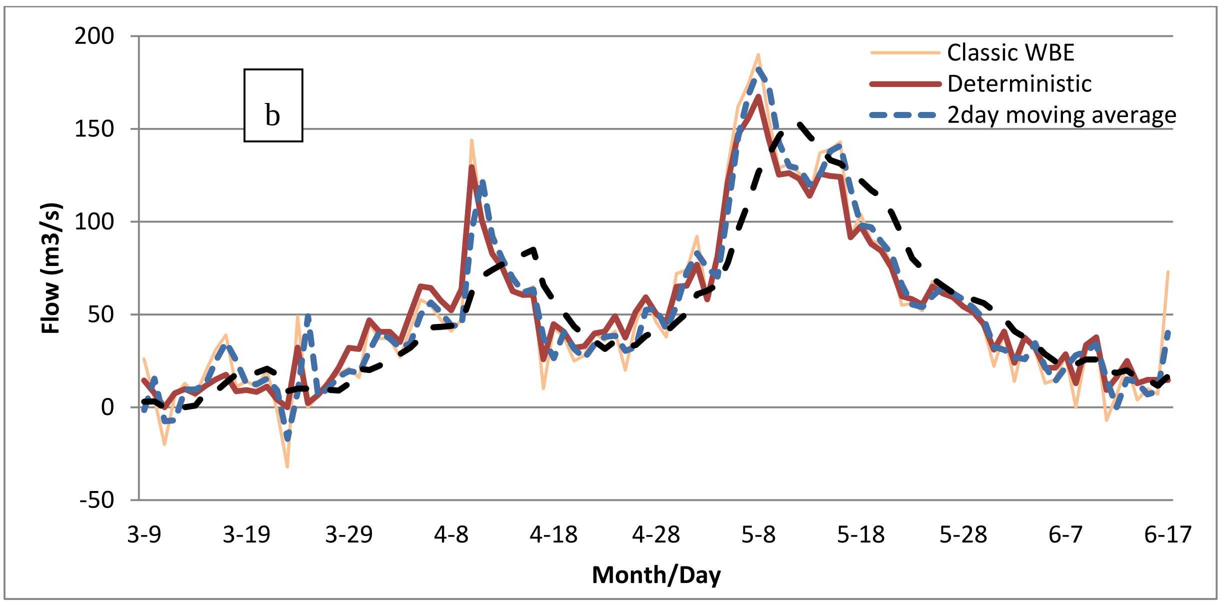


Figure 7-30: Comparing the results of moving average with deterministic based model, Outardes2 (2000). a) winter low flows, b) spring high flows

7.2.4 Calculating the Final Flow Data Series

In Chapter 6, the use of the Deterministic and Stochastic techniques was suggested as the flow data series reconstruction methods for Post-R period. As concluded in Section 7.2.3, the performance of these methods changes with time interval period during the case study. Each of these methods has their own strengths in capturing the effects of different factors on the flow data series. Thus, combining them will give a more comprehensive result. In the work for this thesis, a weighted average method will be the suggested technique to use to combine the reconstructed flow based on these two methods (the same approach can be applied for Pre-R period). The results of applying this method on Outardes 2 (2000) are shown in Figure 7-31.

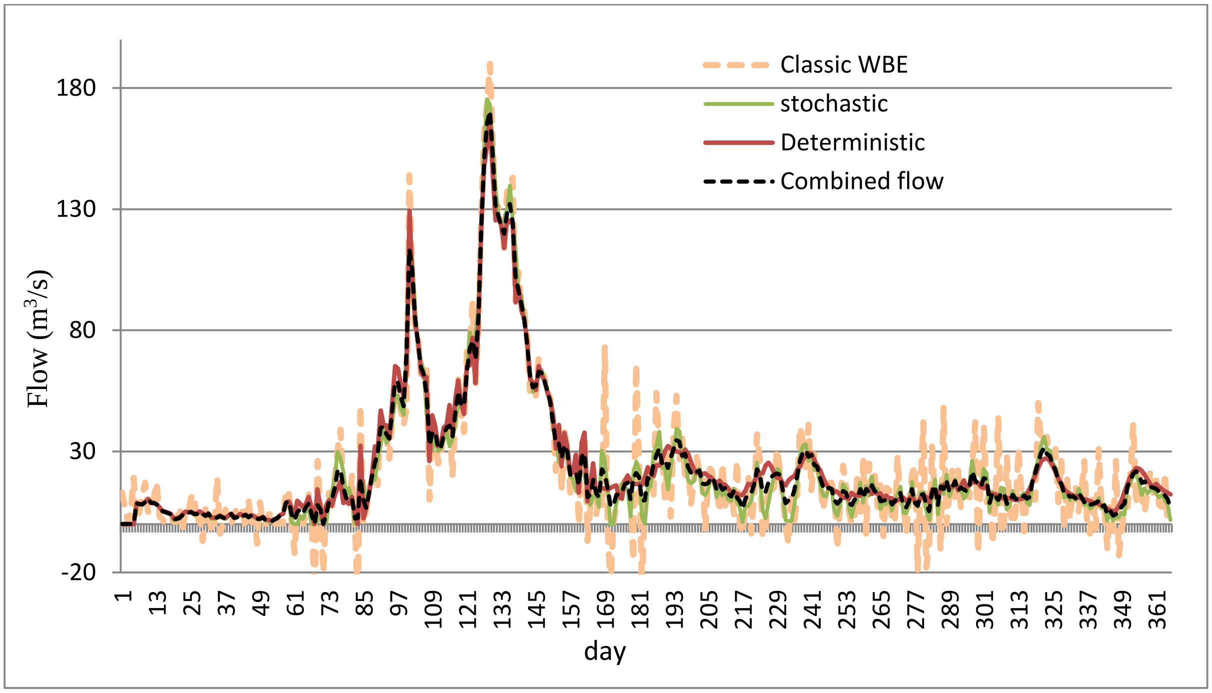


Figure 7-31: Comparison of Classic WBE, deterministic method, stochastic method, and combined flow (Outardes2-2000)

7.2.5 Uncertainty of Reconstructed Flow

As explained before in Chapter 6, several steps need to be completed to estimate the effect of input data uncertainty on the reconstructed flow series. These steps discussed in the following sections.

7.2.5.1 Selecting a sub-time period for which uncertainty analysis will be performed

Calculating the uncertainty for a sub-time period and then extrapolating the results to the desired time period is practical when it is very time consuming to do uncertainty analysis for a case study of multi-year time periods. Selecting the sub-time period is very important as the selected year should possess different flow characteristics to reflect various ranges of uncertainty. For example, in this study, the uncertainty of flow series for Outardes 2 is evaluated using year 2000 as the sample year because, according to Figure 7-32, the flow characteristics of this time period include:

- low, medium, and high variation
- negative values of flow calculated using classic WBE
- general peak flows with both similar and dissimilar values to the RR model
- local peaks with both similar and dissimilar values to the RR model
- flows with values over and under the RR model

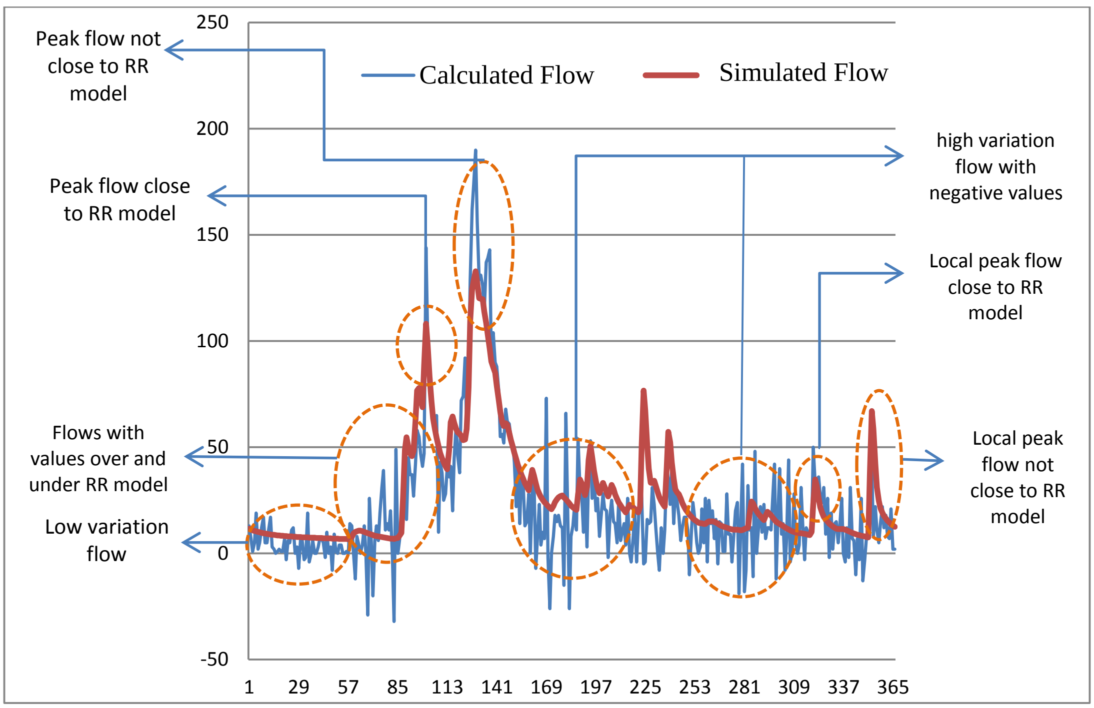


Figure 7-32: Selected year (2000) for uncertainty analysis in Outardes 2

7.2.5.2 Input Data Uncertainty

Input data uncertainty affects the values of flow reconstruction. In this section, input data uncertainties are to be factored into the reconstruction method as part of the flow results. The uncertainty of each input data series is caused either from instrument uncertainty and/or random uncertainty which are explained in the following sections.

7.2.5.2.1 Instrument uncertainty

As explained in Chapter 6, instrument uncertainty has already been defined for 17 basins in Quebec. The instrument uncertainties in the average discharged flow, the turbine flow, the storage (volume), and the input flow for these 17 basins were calculated using Equation 6.24 and are presented in Table 7.2 (report C1, Haché *et al.* 1996). Based on this table, no meaningful relationship could be developed between the uncertainty of input/output flow, reservoir volume, or surface area of basins.

Table 7.2: The average values of different terms of each 17 hydraulic system and their instrument uncertainties- q_{sp}= Discharged flow, q_{tr}= Turbine flow, q_{out} = Outflow from the reservoir, q_{out}= Inflow to the reservoir (report C1, Haché *et al.* 1996)

	Average values for the terms of hydraulic system						Uncertainty-m3/s			Uncertainty ratio-%		
Reservoir	q_{sp} (m3/s)	q_{tr} (m3/s)	q_{out} (m3/s)	q_{in} (m3/s)	flow (m3/s)	Volume (hm^3)	Q_{out}	Q_{in}	Volume	Q_{out}	Q_{in}	Volume
GOUI	172	0	172	0	171	5673.4	5.2	0	162	3.023		0.247
MANA	23	0	23	0	23	212.2	0.7	0	19.5	3.043		0.794
MANB		0		23	20	128	1.3	0.7	6		3.043	0.405
MANC	46	0	46		3	145.1	1.4	1.3	3.7	3.043		0.220
RBLA	20	366	386	218	170	379.7	6.9	66.6	8.3	1.675		0.189
RTREN	30	412	442	412	60	6.3	6.1	6.9	1.6	1.412	1.675	2.194
BEAU	31	453	484	432	41	3.4	6.8	6.1	0.6	1.408	1.412	1.525
TUQU	57	493	550	483	67	4.5	10	6.8	0.9	1.908	1.408	1.728
MATA	72	0	72	0	71	572.3	2.2	0	11.5	3.860		0.174
GMER	98	553	651	581	42	22.1	20	12.2	2.4	3.200	2.100	0.938
SHAW	125	597	722	625	49	4	14	20	0.5	2.014	3.200	1.080
GABE	95	591	686	695	-13	1.6	14.4	14	0.4	2.011	2.014	2.160
CABO	34	0	34	0	40	931.9	1.2	0	43	3.529		0.399
BASK	268	0	268	34	233	1674.6	8	1.2	28.3	2.985	3.529	0.146
PAUG	9	343	352	268	86	69.7	12.6	8	3.2	3.580	2.985	0.397
CHEL	15	338	353	352	1	3.1	12.8	12.6	1	3.626	3.580	2.787
FARM	24	333	357	353	2	1	13.2	12.8	0.1	3.697	3.626	0.864

For the basins used in the current case study, the same methodology (Equation 6.24) was applied to determine the instrument uncertainty of q_{in}, q_{out}, and *volume* (Δ_{qin}, Δ_{qout}, Δ_{volume} respectively). Unlike the C1 Report, the instrument uncertainty was calculated for each daily time period in order to have a better understanding of the daily range of flow. Then, in the Section 7.2.5.3, the optimization model will be solved using $q_{in} \pm \Delta_{q_{in}}$, $q_{out} \pm \Delta_{q_{out}}$, and *volume*,$\pm \Delta_{volume}$ data to calculate the range of flow resulting from this type of uncertainty.

The range of WBE flow caused by instrument uncertainty is simply equal to $\pm (\Delta_{q_{in}} + \Delta_{q_{out}} + \Delta_{volume})$.

7.2.5.2.2 Random uncertainty

In the work done for this thesis, the random uncertainty value of the input data was set at 5% of the data set. To evaluate the adequacy of the mentioned perturbation, the sensitivity of the calculated flow (from classic WBE) to each terms of q_{in}, q_{out}, and *volume* should be assessed. To this end, each of q_{in}, q_{out}, and *volume* data series are disturbed in a separate task (giving a total of three separate tasks) up to a 5% uncertainty value. Then, in each task, the WBE is solved based on the data generated by the disturbed q_{in}, q_{out}, or *volume*.

According to the results, the variance of the new flow data series from the original was up to 224% and 272% (Equation 6.25) when q_{in} and q_{out} are disturbed respectively. These variance values are at maximum and occur when all the values of each q_{in} or q_{out} data series disturbance is set equal to 5%. Also, the new flow data series differ from the original one up to more than 200% when the *volume* data series is disturbed. Thus, it can be concluded that a 5% perturbation of the three parameters is enough to cause a meaningful difference in the results of WBE. Since the real values of random uncertainty are different from day to day, more studies are needed to be performed in future research in order to define the exact value of daily random uncertainty.

7.2.5.3 Evaluating the uncertainty range for selected time-period

7.2.5.3.1 Instrument uncertainty

To calculate the instrument uncertainty, the improved POM optimization model (in Chapter 6) should be solved using $q_{in} \pm \Delta_{q_{in}}$, $q_{out} \pm \Delta_{q_{out}}$, and $volume \pm \Delta_{volume}$ data series. This gives the range of flow due to the instrument uncertainty. The range of WBE flow caused by instrument uncertainty is also equal to $\pm (\Delta_{q_{in}} + \Delta_{q_{out}} + \Delta_{volume})$. The instrument uncertainty calculated using WBE and optimization model for Outardes 2 using year 2000 as the sample year are presented in Figure 7-33.

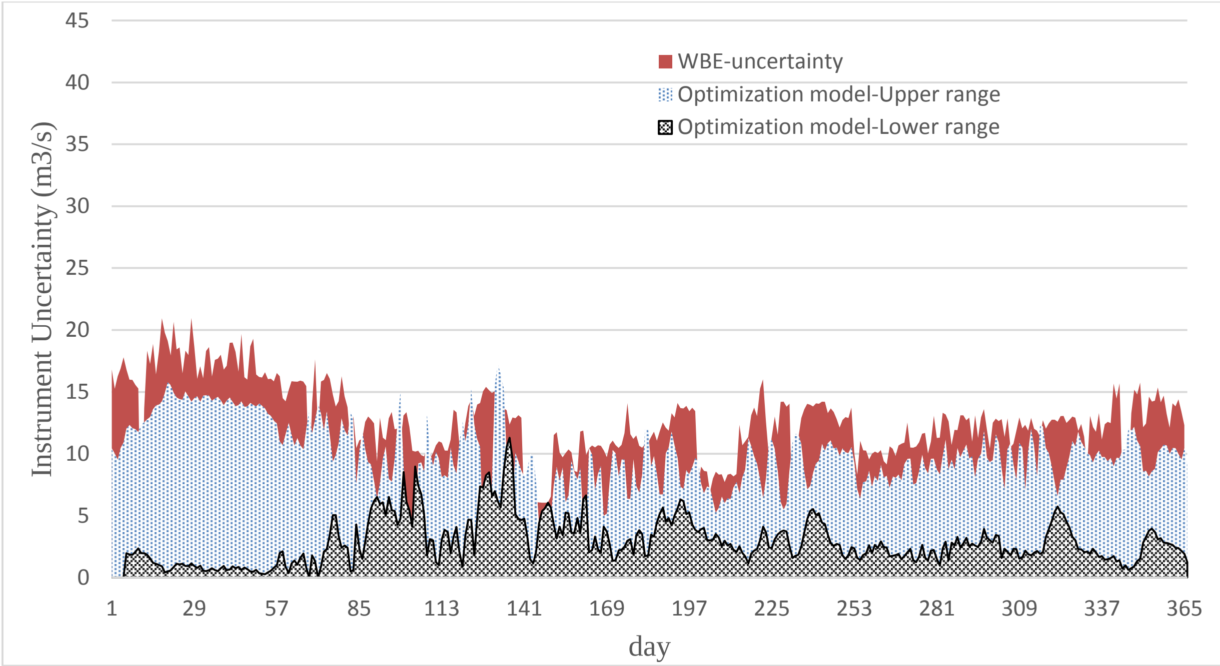


Figure 7-33: Instrument uncertainty for Outardes 2-2000

As shown in Figure 7-33, the instrument uncertainty of the WBE and that of the optimization model (upper limit) are higher during the winter season because the values of turbine flow are larger throughout this period. Increasing the instrument uncertainty by augmenting the values of turbine flow shows that improving the quality of turbine flow data or methods of measuring them could highly decrease the uncertainty. According to Figure 7-33, the average and standard deviation of instrument uncertainty for WBE are 12.7 and 3.01 m^3/s respectively, and the average and standard deviation of instrument uncertainty for optimization model are 2.9 and 1.97 m^3/s regarding to the lower limit, and 9.9 and 2.7 m^3/s regarding to the upper limit respectively. Thus, the optimization model could slightly improve the upper limit. In this figure, the lower limit of flow do not have the same shape as the upper limit. For example, the range is narrower in winter because the optimization model is not permitted to give negative flow values.

7.2.5.3.2 Random uncertainty

To evaluate the effects of random uncertainty, four different scenarios have been presented in Section 6.2.5.2.2. In these scenarios, a perturbation of 5% was applied to the input data and the effect of this random perturbation on the flow series was assessed. This gives the range of flow results stemming from random uncertainty. To do this, the optimization model is solved for each of the four scenarios in which input data are disturbed randomly, as it was explained in Section 6.2.5.2.2. The optimization model is then solved one hundred times through an iterative process. In the first scenario, it was assumed that the random uncertainty is only caused from q_{out}, which included discharged flow and turbine flow from the reservoir. In our case (Outardes 2, using 2000 as the sample year) the discharged flow is always zero; thus, the uncertainty of q_{out} is only related to the turbine flow. The results of this scenario are presented in Figure 7-34. The maximum difference between upper limit of random uncertainty and the combined flow is 40 m^3/s and the maximum difference between lower limit and the combined flow is 54.1 m^3/s in this scenario.

In the second scenario, it was assumed that the random uncertainty is only found in q_{in}, which included discharged flow and turbine flow from upstream reservoir. In our case (Outardes 2, using 2000 as the sample year), the discharged flow from the upstream reservoir is always zero; thus, the uncertainty of q_{in} is only related to the turbine flow from the upstream reservoir (Outardes 3). The results of this scenario are presented in Figure 7-35. The maximum difference between upper limit of random uncertainty and the combined flow is 41.6 m^3/s and the maximum difference between lower limit and the combined flow is 53.8 m^3/s in this scenario.

In the third scenario, it is assumed that the random uncertainty is only found in reservoir volume data. The results of this scenario are presented in Figure 7-36. The maximum difference between upper limit of random uncertainty and the combined flow is 42.5 m^3/s and the maximum difference between lower limit and the combined flow is 53.8 m^3/s in this scenario.

In the fourth scenario, it is assumed that all the input data contain random uncertainty. The results of this scenario are presented in Figure 7-37. The maximum difference between upper limit of random uncertainty and the combined flow is 37.5 m^3/s and the maximum difference between lower limit and the combined flow is 44.5 m^3/s in this scenario.

These figures, in addition to Table 7.3, show that (as was expected) the estimated range of flow is slightly higher in scenarios I and II compared to scenario III. Since these two scenarios considered the uncertainty caused by turbine flow, they result in a wider range of flow. As stated in Table 7.3, the probable range of flow is not symmetrical and the upper limit of flow is always larger than the lower limit because the lower limit is not allowed to drop below zero.

As shown in Figure 7-37, the estimated range of random uncertainty is different between low variation periods (the time period with less noise) and high variation periods. As expected, the range of uncertainty increases with variation (noise) of WBE because a noisier WBE means more uncertainty in the input data. Also, the lack of correspondence between simulated flow (using RR model) and WBE flow caused a wider range. For the periods when the values of the simulated flow are greater than that of the WBE flow, the upper range get wider and vice versa. Even global and local peak flows obey this rule. Unlike instrument uncertainty, it is not easy to determine the random uncertainty of WBE.

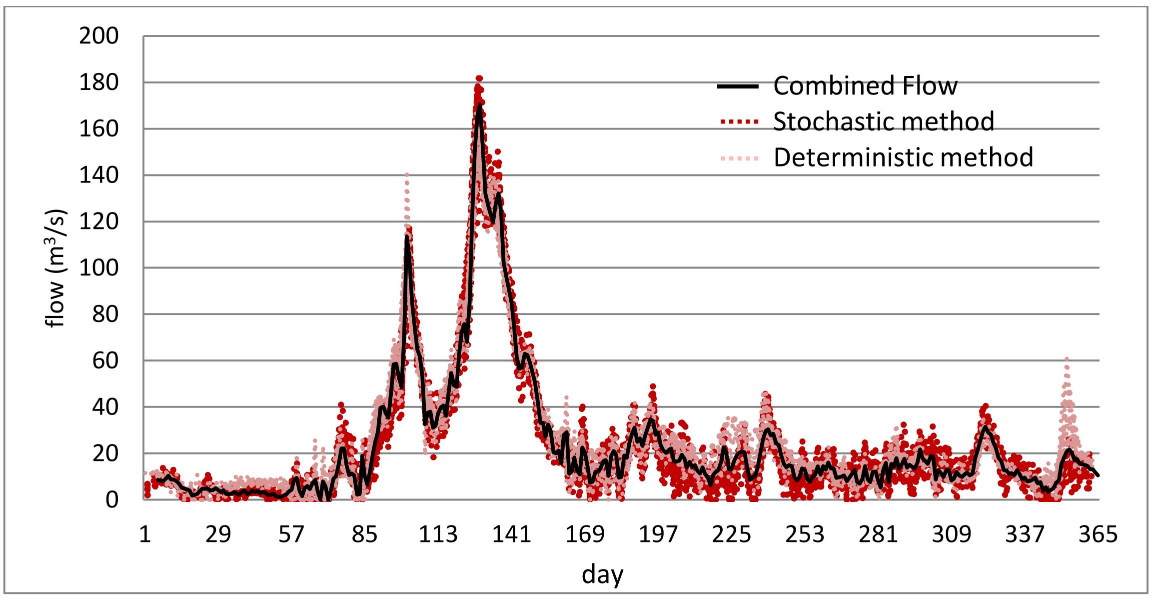


Figure 7-34: The results of uncertainty analysis for Outardes 2-2000-scenario I

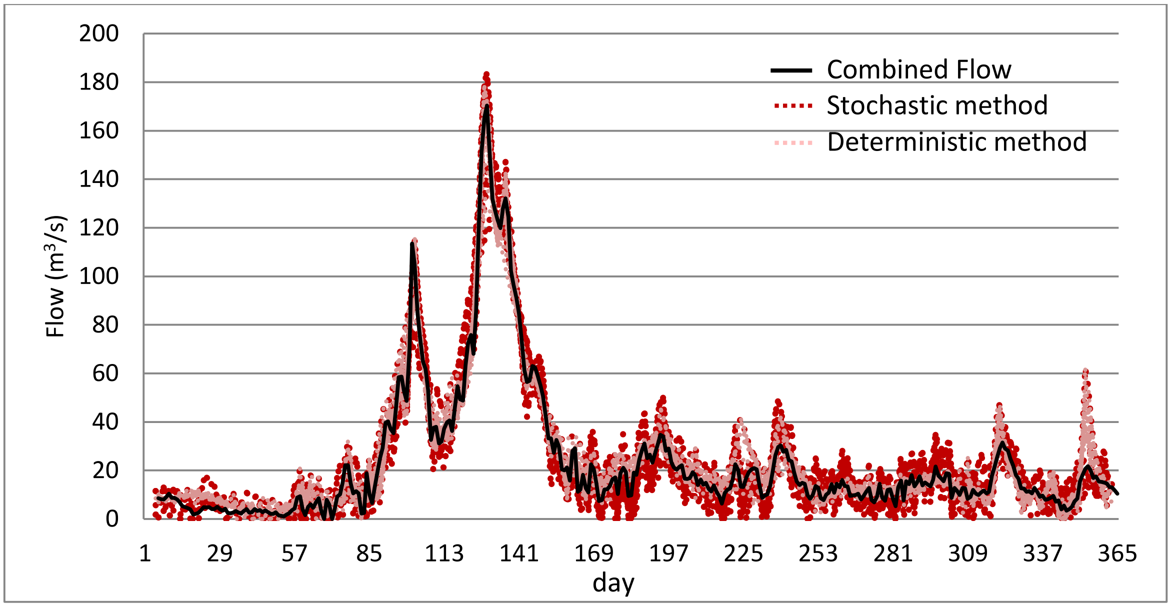


Figure 7-35: The results of uncertainty analysis for Outardes 2-2000-scenario II

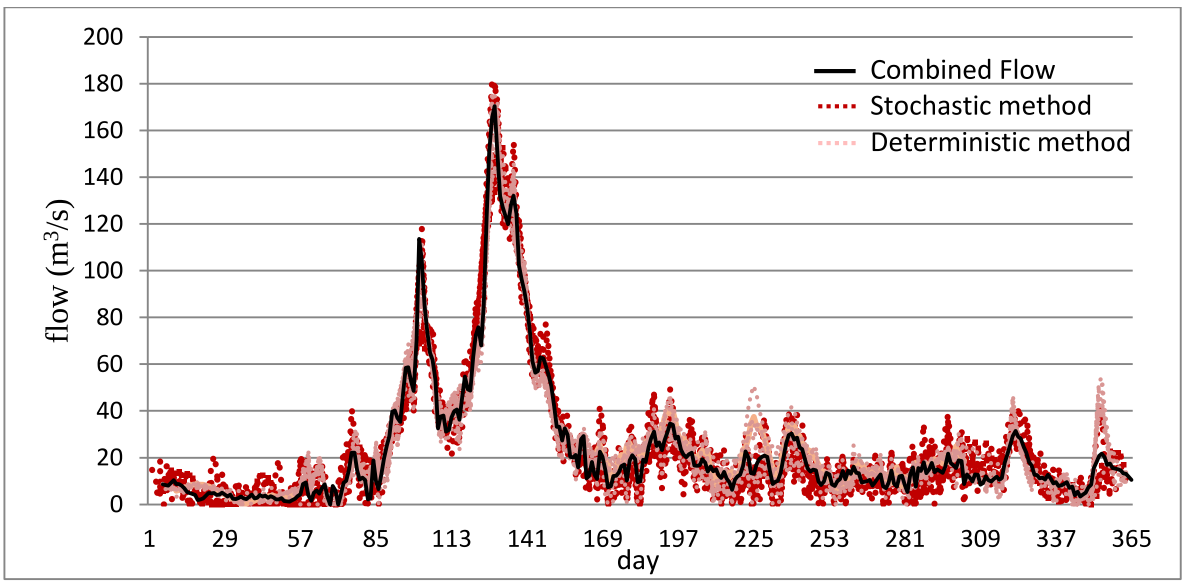


Figure 7-36: The results of uncertainty analysis for Outardes 2-2000-scenario III

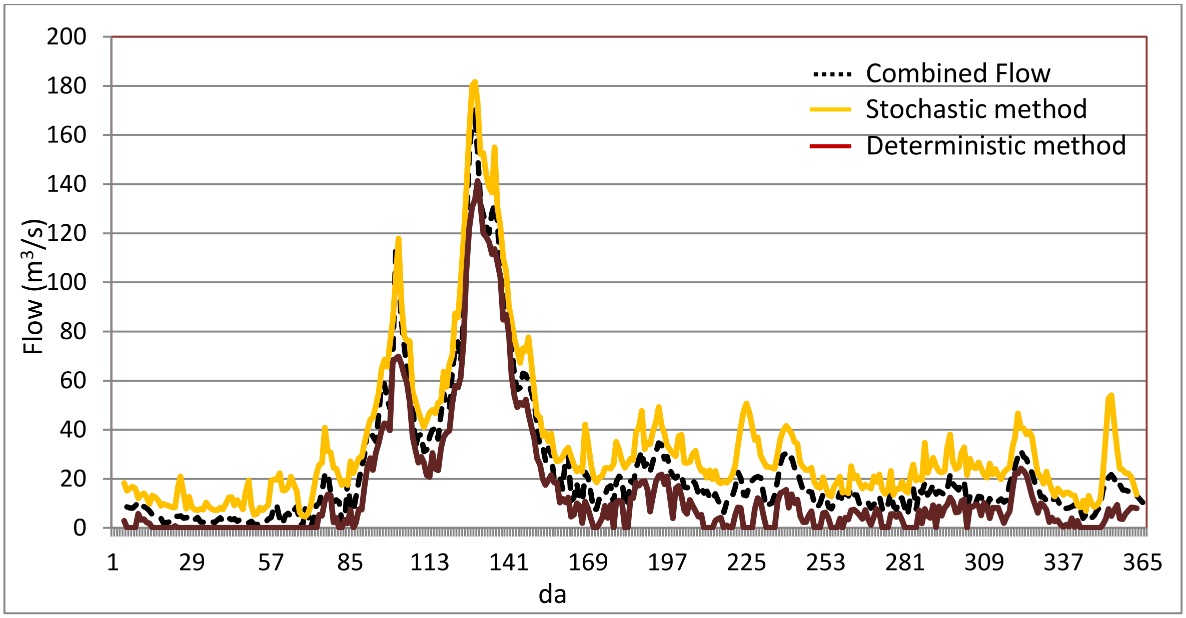


Figure 7-37: The results of uncertainty analysis for Outardes 2-2000-scenario IV

Table 7.3: The average range of estimated random uncertainty (m^3/s) based on different scenarios in Outardes2-2000

Scenario	Upper limit	Lower limit	Range
Scenario I	+ 10.24	- 9.25	19.49
Scenario II	+ 9.96	- 9.08	19.04
Scenario III	+ 8.54	- 8.04	16.58
Scenario IV	+ 9.99	- 9.10	19.09

7.2.5.3.3 Total uncertainty

In order to determine total uncertainty, the calculated random uncertainty should be added to the computed instrument uncertainty. This gives the range of flow (Figure 7-38). According to this figure, the reconstructed flow fits perfectly between the estimated flow ranges, with the upper limit exhibiting a wider range flow than the lower limit during the low flows. This indicates that if the real flow data are different from the estimated flow data, it is more probable that their value is higher than combined flow.

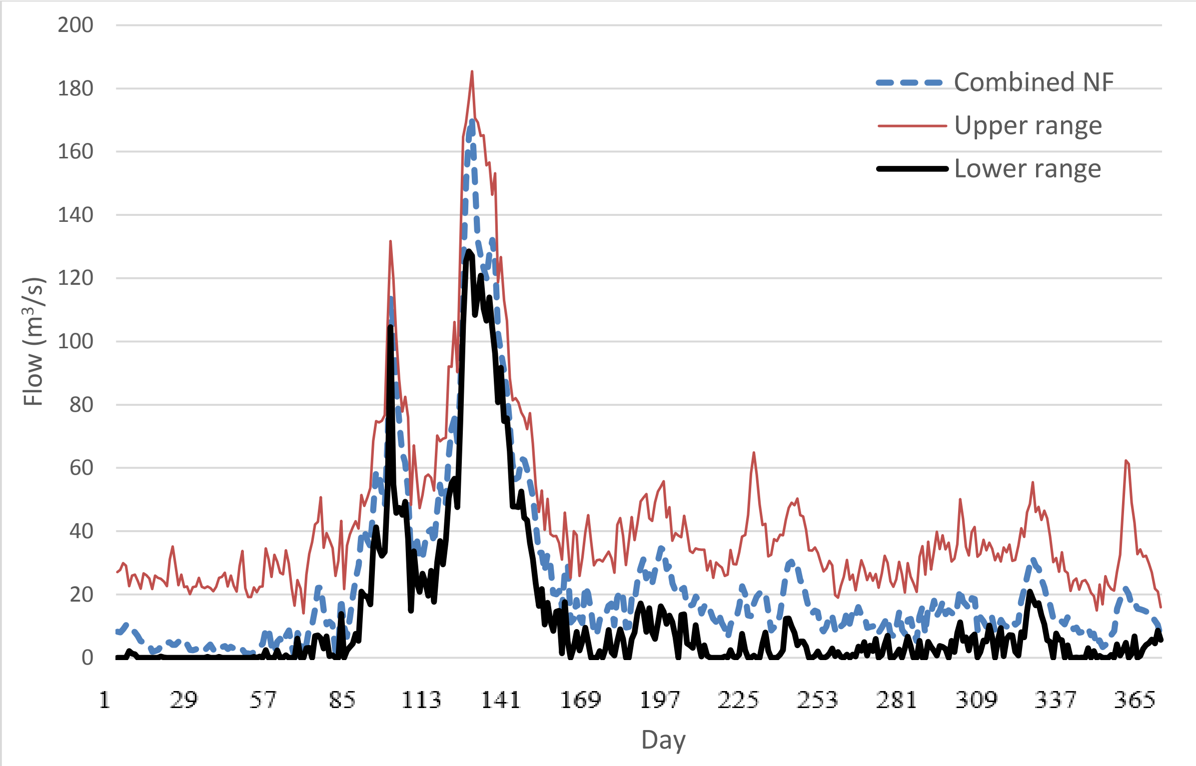


Figure 7-38: Calculated range of flow (total uncertainty)

7.2.5.4 Extending the calculated uncertainty range to the whole time period

The probable range of flow was calculated for just one sample year and should be extrapolated to the whole lifetime of the basin. To achieve this, a polynomial regression was developed between the daily flow and each daily upper limit and lower limit. The developed regressions are represented by Equations 7.2 and 7.3 respectively. These regressions, then, could be used to predict the upper and lower limits of uncertainty for each day of historical data series.

$$R_u = -0.00002F_{WA}^2 + 0.0149F_{WA} + 20.954 \quad (7.2)$$

$$R_l = -0.0003F_{WA}^2 + 0.1078F_{WA} + 3.2859 \quad (7.3)$$

where:

R_u = the predicted upper range

R_l = the predicted lower range

F_{WA} = the weighted average flow (Equation 6.20).

7.3 Conclusion

The results of the sensitivity analysis show that the improved POM suffers from poor assumption of constant PS. According to the results, the quality of flow data series is highly dependent on the PS. Thus, the results could be improved by selecting the appropriate PS in time and space.

In this thesis, the posterior Deterministic GA and the Stochastic probabilistic algorithm techniques were the suggested techniques to use in order to automatically define the parameters of the developed optimization model for ungauged basins containing a reservoir. Application of these two methods on the Outardes basin showed that they are highly capable of improving the result on the classic WBE method: the results did not contain any negative flow, they are less noisy, are more reliable, and matched perfectly with regional flows. Therefore the Deterministic and Stochastic based optimization model can be considered as the answer to the question of how to reconstruct more likely daily flow for Post-R period.

Visual graphs, in addition to five QIs designed to evaluate the performance of reconstructed flow, enabled us to compare the Stochastic based method, Deterministic based method, and classic WBE. The application of the KPSS test, Q-Q plot, scale location plot, and residual versus fitted value plot techniques were also used to confirm the regional and temporal homogeneity of reconstructed flow data. These aforementioned methods and tools answered the question of how to evaluate the quality of reconstructed flow series in ungauged basins for Post-R period.

According to the results of this chapter, the performance of the suggested flow reconstruction techniques depends on the particular case and on the time of year. Hence, a weighted average method was suggested to calculate the final flow for each segment based on the combination of results of Stochastic and Deterministic methods.

Finally, a methodology was developed to estimate the uncertainty of the final flow caused by random and instrument uncertainties on the input data. The methodology is independent from flow measurements and is based on the input data sensitivity analysis for a chosen sub-time period. This method answers the question of how to evaluate the uncertainty of flow data in ungauged basins. The results show that flow data are highly affected by the uncertainty found in the input data, especially turbine flow. Therefore, more reliable results in the flow calculation process will be gained by validating all input data before use.

CONCLUSION AND RECOMMENDATIONS

CONTRIBUTION

The main focus of this thesis has been to develop a process to reconstruct a set of smooth and realistic daily flow values for ungauged basins. The work in this thesis has obtained these results through five main contributions to achieve this goal of flow reconstruction:

i) Provide an algorithm to select the most appropriate family of the flow reconstruction methods in any case-study scenario;

ii) Develop a flexible method of daily flow reconstruction for ungauged basins for the pre-reservoir construction (Pre-R) time period;

iii) Propose a flexible automatic methodology for daily flow reconstruction and filtering for ungauged basins that are equipped with a reservoir;

iv) Develop several criteria to evaluate the quality of the reconstructed flow data for ungauged basins.

v) Propose a methodology to evaluate the uncertainty in a flow data series generated through the flow reconstruction process for ungauged basins.

The conclusions of each chapter are summarized in the “Conclusion” section. The limitations of the developed methodologies are marked by bullet points, and the recommendations on overcoming these limitations are listed at the end.

CONCLUSION

The lack of a comprehensive methodology for selecting an appropriate method of flow reconstruction necessitated an inclusive study on this subject, along with research into the development of an algorithm to assist in choosing the right family of flow reconstruction method for different case study scenarios. This led to a complete literature review on the different methods currently being used for flow reconstruction. The applicability, advantages and disadvantages of these methods have been summarized through an algorithm that helps to define the fitting family of methods in each case study scenario. This algorithm takes into account all determinative factors that may affect the selection of appropriate method of flow data series reconstruction, and was then tested for its effectiveness by applying it on the selected case study scenarios. Results of this application have shown that in this specific case study, WBE based methods and regression based techniques work better than others methods for Post-R and Pre-R periods respectively (Chapter 4).

- This algorithm puts the judgment of the input data quality and the required output data duration (long term or short term) in the hands of the researcher. If different researchers make different judgment calls on these deciding factors, the outcome of algorithm may be dissimilar. This is an area which needs to be refined in regards to the use of this developed algorithm.

Pre-R period

The existing regression based flow reconstruction methods for the Pre-R period are not considered the most precise or robust methods. However, the use of these methods are unavoidable when there is limited data available (such as flow data from neighbouring basins and RR model) for the sample time period. In this thesis, a new Kalman filter based model was developed to filter and combine the flow data from neighbouring basins and RR model for the Pre-R period. This method takes advantage of existing limited data in order to improve the results (Chapter 5). Results of developed Kalman based model were then compared with the results obtained from the use of existing models (including Area Ratio, Move III, and Multivariable regression). In addition to use of visual graphs, three QIs designed to evaluate the quality of reconstructed flow values for an ungauged basins were applied to perform the comparison.

The results show that there each method has its advantages in producing flow data series but SF is the determinative factor to select the most reliable method. In Outardes 3 and 2, Move III (x= simulated flow) is the most reliable flow reconstruction method because it is the only method that produces the flow time series with the SF close to Post-R period methods. In Outardes 4, Move III (x= simulated flow) is also the most reliable flow reconstruction method because it mostly has the best QIs (including SF).

Post-R Period

Classic WBE is the first choice from the family of WBE based methods to be used to reconstruct the flow data for the Post-R period. However, the results of using this method have been found to be noisy and unrealistic. Problems related to other alternative methods such as the manual flow reconstruction and filtering method and POM were also found. Although POM produced the much better results than classic WBE and it was independent from human decision making and experience, deficiencies and errors while using POM were still issues that needed to be resolved.

Some of these problems were solved successfully with the improved POM (Chapter 6). For example, the improved POM is applicable for a daily time sample use, no longer relying on a 5 minute volume data sample. Thus, the model is considered more flexible and can be applied towards any time sampling interval in the Post-R period. Moreover, the problem of poor boundary conditions was resolved in new version by applying a moving optimization window. Generally, altering the constraints of POM in the improved model has helped to get more realistic data series. Nevertheless, a sensitivity analysis shows that the modified POM is very sensitive to the changes in dn, C, and γ, affecting the integrity of the flow results. The results show improvement when the PSs are modified for each time interval and case study scenario (Chapters 6 and 7). The weaknesses of the modified POM are as follows:

- In this model, the WBE is simplified by including water loosing terms (evaporation, direct rainfall, and interaction between ground water and reservoir's water) into the calculated flow. This results in unrealistic values of flow when the water loss is considerable.
- Input data to the model are calculated based on non-validated measured data for the years prior to 2005. This causes some uncertainties in the reconstructed flow data series. When input data are not realistic during tens of continuous days, the model ends up failing.

The absence of a reliable technique to calculate the smooth daily flow in ungauged basins was the motivation to develop a WBE based model. This model needed to define the parameters of model intelligently; thus, an automatic optimization tool was drafted. The parameters of this model are defined by using both Deterministic and Stochastic techniques (Chapter 6). Unlike the Perreault's assumption, the parameters of this model could change depending on the sample time. Stochastic and Deterministic methods use the simulated flow (using RR model) and the neighbouring basin's flow to define the parameters of model. This approach increased the accuracy of results. These two techniques have shown different performances depending on the time and case-study (Chapter 7).

A shortage of criteria for evaluating the quality of reconstructed flow was the motivation in designing a few QIs applicable in ungauged basins. Reliability of the data was measured using five QIs during Post-R period (Chapter 6). These criteria are independent from mostly manually filtered flow data series (available in Quebec). Also, a few tests were implemented to check the regional and temporal homogeneity of reconstructed flow. The results of quality evaluation confirmed the integrity of the results (Chapter 7).

- In this thesis, a simple average method was used to define the mean QI for each sample time period. However, more investigation is still required to define the weight of different indexes in average QI.

Finally, the reconstructed flow data series were combined using a Weighted Average method to form the final Flow. The final flow data series is influenced by different sources of uncertainty, such as input data error, parameter unreliability, and/or model weaknesses. Comparing flow values before and after 2005 indicated that the uncertainty of reconstructed flow is mostly related to the random uncertainty and instrument uncertainty of the input data that may be caused by natural phenomena (i.e. floods, ice cover), instrument disorders (i.e. gates' maneuver disorders), simplification of calculations, and/or human uncertainties . Thus, an uncertainty analysis was performed to estimate the range of combined flow. Random uncertainty was defined based on four different scenarios. In each of first three scenarios, one input data was disturbed randomly (5%). In the last scenario, all the input data were perturbed together. The results of this analysis showed that the random uncertainty of the reconstructed flow is affected by the turbine flow data series more than anything else.

Also the effects of instrument uncertainties on flow reconstruction were assessed in this case study. The results of this assessment also confirmed the dependency of flow uncertainty on the quality of turbine flow data. Moreover, comparing the instrument uncertainty of the reconstructed flow with that of classic WBE showed that the quality of flow has been improved in the developed model. This means that the developed model could control part of input data uncertainty (Chapters 6 and 7).

- What requires more research is the magnitude of random uncertainty because, in practice, it could change by season, reservoir size, number of turbines, and etc. Thus, the limit random uncertainty is different on some days from the 5% introduced in this research.

RECOMMENDATIONS

Based on the present research, the following recommendations are proposed:

1- Calculating the reservoir related data (storage volume, turbine flow, discharged flow) based on validated measured data before applying them in a WBE, especially during the time periods when the quality of input data is considerably low for several days in the row.

2- Taking into consideration the basin flow routing into rivers and reservoirs to improve the quality of data (where they could make a big difference).

3- Considering the effects of wind and ice cover on water level, and that of maneuver disorders of gates on discharged flow.

4- Calculating the water loosing terms (evaporation, direct rainfall on reservoir, interaction between ground water and surface water) as part of reconstructed flow and extracting them from reconstructed flow values to have a better understanding of real flow data and their range in the area.

5- Using more QIs to evaluate the flow reconstruction more comprehensively.

6- Validating the weight of different QIs in average QI for each time period.

7- Estimating more accurate values of random uncertainty based on condition (season, reservoir size, number of turbines, and etc.)

REFERENCES

Abebe, N.A., Ogden, F.L, Pradhan, N.R. (2010). Sensitivity and uncertainty analysis of the conceptual HBV rainfall–runoff model: Implications for parameter estimation. *Journal of Hydrology*, *389* (3-4): 301–310.

Adeloye, A. J., and N. R. Nawaz. (1998). Assessing the adequacy of Wright's model reconstructed streamflow data for reservoir storage-yield-reliability analysis. *Water and Environment Journal, 12*(2):130-138.

Ajami, N. K., Duan, Q., Gao, X., and Sorooshian, S. (2006). Multimodel combination techniques for hydrological forecasting: application to distributed model intercomparison project results. *Journal of Hydrometeorology 8*, 755–768.

Albertini, S. (2013). Impact of in-stream habitat structures used in river restoration on fish populations: a feeding ecology approach. ETHZ Biology MSc, Department of Biology, Swiss federal institute of technology Zurich, Switzerland.

Bates, B. C., Campbell, E. P. (2001) A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Res. 37(4)*:937–947. doi:10.1029/2000wr900363.

Bennis, S., Cote, S., Faouzi, B., and Gagnon, L. (1994). Logiciel De Validation Des Données Hydrométriques ‚Valideb", Mathematic Hydrology Lab, And Superior Technology School, *Final Technical Report*, Hydraulic Service, Hydro-Quebec.

Bennis, S., and Kang, N. (2000). *Multivariate Technique for Validating Historical Hydrometric Data with Redundant Measurements*. Eighth International Conference on Hydraulic Engineering Software, HYDROSOFT 2000, Lisbon, Portugal.

Berrada, F., Bennis, S., and Gagnon, L. (1996). Validation Des Données Hydrométriques Par Des Techniques Univariées De Filtrage. *Canadian journal of civil engineering, 23 (4):* 872–892.

Beven, K. (2007). Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. *Hydrol Earth Syst Sci 11(1)*:460–467.

Beven, K., and Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J Hydrol 249(1–4)*:11–29. doi:10.1016/s0022-1694(01)00421-8.

Bisson, J.L., and Roberge, F. (1985). Prévision des Apports Naturels : Expérience De l'Hydro-Québec. *Report of Workshop on Flow Prediction, Downsview*. General Direction of Inshore water, Environnement Canada, Study No 5, Des Ateliers Techniques. 148 - 161.

Boss International. (2005). *StormNet and wastewater model user's manual*. http://creamdesigners.com/clients/Boss/stormnet-overview.htm.

CEHQ. (2013). *Centre d'expertise hydrique du Québec*. Ministère du Développement durable, de l'Environnement, de la Faune et des Parcs. Available @ http://www.cehq.gouv.qc.ca/ (accessed in September 2013).

Charbonneau, R., and Bérubé, R. (1987). Validation Des Données Hydrométriques; Historique De La Rivière Outaouais. *Internal Report*, Hydro-Québec.

Cheng, L., Hou, Z.G., Homma, N., Tan, M., and Gupta, M., (2009), *Solving Convex Optimization Problems Using Recurrent Neural*, Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, available at: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5178723&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5178723.

Chu, P.P, (1992), *A Neural Network for Solving Optimization Problems with Linear Equality Constraints*, International Joint Conference on Neural Networks, Baltimore, MD (p. 272 – 277). Doi: 10.1109/IJCNN.1992.226996.

Cunderlik, J.M., Burn, D.H. (2003), Non-stationary pooled flood frequency analysis, *Journal of Hydrology*, *276* (1-4), 210–223.

Cunnane, C. (1978). Unbiased plotting positions – a review. *J. Hydrol. 37*, 205–222.

Das, M. M., and Saikia, M. D. (2013). *Watershed Management*. PHI Learning Private Limited. 227 pp.

Dastorani, M.T., Moghadamnia, A., and Piri, J., and Rico-Ramirez, M., (2010), Application of ANN and ANFIS models for reconstructing missing flow data, *Environmental Monitoring and Assessment, 166* (1-4), 421-434.

Deb, K., (2000), An efficient constraint handling method for genetic algorithms, *Computer Methods in Applied Mechanics and Engineering, 186* (2-4), 311-338.

Deb, K., (2002), A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, *IEEE Transactions On Evolutionary Computation, 6* (2), 182-197.

Dong, L., Xiong, L., Zheng, Y. (2013). Uncertainty analysis of coupling multiple hydrologic models and multiple objective functions in Han River, China. *Water science and technology, 68* (3), 506-513.

England, K., Xu, C. Y., and Gottschalk, L. (2005). Assessing uncertainties in a conceptual water balance model using Bayesian methodology. *Hydrol Sci J 50(1)*:45–63. doi:10.1623/hysj.50.1.45.56334.

Eslamian, S. (2010). Flood regionalization using a modified region of influence approach. *Journal of Flood Engineering, 1* (1): 51-66.

Fortin, V. (2001). *Analyse et prévision des apports énergétiques*, technical report, Montreal, Quebec, Institut de recherche d'Hydro-Québec.

Fortin, J.P., Moussa, R., Turcotte, R., Rousseau, A.N., and Villeneuve, J.P. (2006). *The Hydrotel distributed hydrological model: a simulation and forecasting model making good use of GIS and RS data*, European Geosciences Union, Vol. 8, 08186.

Fortin, V., and Slivitzky, M. (2000). *Downward shift in runoff in Northern Quebec-Labrador in the middle eighties and possible links to the Arctic Oscillation*.

Guntner, A., Krol, M., Carlos de Araujo, J, and Bronstert, A. (2004), Simple water balance modelling of surface reservoir systems in a large data-scarce semiarid region, *Hydrological Sciences–Journal, 49*(5), 901-918.

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

Gustard, A., and Demuth, S. (2008), *Manual on low flow estimation and prediction*, World Meteorological Organisation, Operational hydrology report N.50, available at: http://www.wmo.int/pages/prog/hwrp/publications/low-flow_estimation_prediction/WMO%201029%20en.pdf.

Haché, M., Durocher, M., and Bobée, B. (2003). Modélisation Non Paramétrique De La Relation Entre Les Caractéristiques Du Vent Et La Différence De Niveaux Sur Un Grand Réservoir. *Canadian Journal of Civil Engineering, 30(4)*: 684-695.

Haché, M., Perreault, L., & Bobée, B. (1996). Validation des apports naturels historiques, Rapport final du projet C1. Chaire industrielle CRSNG - Hydro-Québec en Hydrologie statistique - INRS-EAU.

Hamon, W.R., (1961), Estimating potential evapotranspiration, Proceedings of the American Society of Civil Engineers, *Journal of Hydraulic Division, 87*, p. 107-120.

Han, J. C., Huang, G. H., Zhang, H., Li, Z., Li, Y.P. (2014). Bayesian uncertainty analysis in hydrological modeling associated with watershed subdivision level: a case study of SLURP model applied to the Xiangxi River watershed, China. Stoch. Environ. Res Risk Assess.: 28(4), 973-989.

Henry, J. C. (2013). *Impact of in-stream habitat structures used in river restoration on fish populations: a feeding ecology approach*. Thesis of Master of Science in Biological Systems Engineering, Virginia Polytechnic Institute and State.

Hernandez-Henriquez, A., Mlynowski. T.J., Dery, S.J., (2010), Reconstructing the natural stream-flow of a regulated river: a case study of La Grande Riviere, *Quebec*, *Canada*, *Canadian Water Resources Journal, 35* (3), 301-316.

Hirsch, R.M., (1982), A comparison of four streamflow record extension techniques, *Water resource research, 18* (4), 1081-1088, 1982.

Google map. (June 2014). Quebec, https:// maps.google.ca.

Hughes, D. A., and Smakhtin, V., (1996), Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves, *Hydrological Sciences -Journal- des Sciences Hydrologiques,41* (6), 851-871.

Hwang, Ch., M. F. Peng, J., Ning, J. Luo, and Ch. H. Sui. (2005). Lake level variations in China from TOPEX/Poseidon altimetry: data quality assessment and links to precipitation and ENSO. *Geophysical Journal International, 161* (1): 1–11.

Hydro-Quebec. (2014). HQ. http://www.hydroquebec.com (accessed in April 2014).

Johnston, R. K., Wang, P. F., Loy, E. C., Blake, A. C., Richter, K. E., Brand, M. C., Kyburg, C. E., Skahill, B. E., May, C. W., Cullinan, V., Choi, W., Whitney, V. S., Leisle, D. E., and Beckwith, B., (2009), *An Integrated Watershed and Receiving Water Model for Fecal Coliform Fate and Transport in Sinclair and Dyes Inlets, Puget Sound*, WA, technical report, SPAWAR, SSC Pacific, San Diego, CA 92152-5001.

Jones, P.D., Lister, D.H., and Kostopoulou, E., (2004), *Reconstructed river flow series from 1860s to present (2002)*, (Science Report SC040052/SR), Environment agency, taken from http://www.cru.uea.ac.uk/cru/data/riverflow/flowreconstruction/ReconstructedFlowsTR.pdf.

Joshi, V.U., and Tambe, D., (2010), Estimation of infiltration rate, run-off and sediment yield under simulated rainfall experiments in upper Pravara Basin, India: Effect of slope angle and grass-cover, *Journal of Earth System Science, 119* (6), 763–773.

Kalakrishnan, M., Chitta, S., Theodorou, E., Pastor, P., and Schaal, S, (2011), *STOMP: Stochastic Trajectory Optimization for Motion Planning*, International Conference on Robotics and Automation, Shanghai, China, taken from https://personalrobotics.ri.cmu.edu/courses/papers/Kalakrishnan11-stomp.pdf.

Karamouz, M., Hosseinpour, A., and Nazif, S., (2011). Improvement of Urban Drainage System Performance under Climate Change Impact: Case Study, *Journal of Hydrologic Engineering, 16* (5): 395-412.

Khu, S. T., Werner, M. G. F. (2003). Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. *Hydrol Earth Syst Sci 7(5)*:680–692.

Kim, D., J. Kaluarachchi. (2013). Predicting natural streamflows in regulated snowmelt-driven watersheds using regionalization methods. *Journal Hydrology and Earth system Sciences (HESS), 10*: 9435-9476.

Kim, J. W., and Y. A. Pachepsky. (2010). Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. *Journal of Hydrology, 394* (3-4): 305–314.

Kokkonen, T. S., A. J. Jakeman., P. C., Young, and H. J. Koivusalo. (2003). Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. *Hydrological Processes, 17*(11): 2219–2238.

Koutsoyiannis, D. and A. Efstratiadis, (2007), Uncertainty Assessment of Future Hydroclimatic Predictions: A Comparison of Probabilistic and Scenario-Based Approaches, *Journal of Hydrometeorology, 8* (3), 261–281.

Krause, P., Boyle, D.P., and Base, F., (2005), Comparison of different efficiency criteria for hydrological model assessment, *Advances in Geosciences, 5*: 89–97.

Kuczera, G., and Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology 211 (1–4)*: 69–85.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y., (1992), Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root, *Journal of Econometrics, 54* (1-3), 159-178.

Lashermes, B., Foufoula-Georgious, E., and Dietrich, W. E., (2007). Channel network extraction from high resolution topography using wavelets. *Geophysical Research Letters: 34*, L23S04, doi:10.1029/2007GL031140.

Lee, D., Schmidt, P., (1996), On the power of the KPSS test of stationarity against fractionally-integrated alternatives, *Journal of Econometrics, 73* (1), 285–302.

Li, L., Xia, J., Xu, C. Y., and Singh, V. P. (2010). Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. J Hydrol 390(3–4):210–221. doi:10.1016/j.jhydrol.2010.06.044.

Madsen, H., (2000). *Automatic Calibration and Uncertainty Assessment in Rainfall-Runoff Modelling*, Proceedings of Joint Conference on Water Resource Engineering and Water Resources Planning and Management, Minneapolis, United States, July 30-August 2, 2000.

Marczyk, A. (2004). Genetic Algorithms and Evolutionary Computation." The Talk. Origins Archive. Retrieved Oct 22, 2014 from: http://www.talkorigins.org/faqs/genalg/genalg.html.

Matalas, N. C., and Langbein., W. B., (1962), *A Correlation Procedure for Augmenting Hydrologic Data*, U. S. Geological Survey Professional Paper., 434-E, E1-E7.

Minville, M., Krau, S., Brissette, F., and Leconte. R. (2010). Behaviour and Performance of a Water Resource System in Québec (Canada) Under Adapted Operating Policies in a Climate Change Context, *Journal of Water Resource Management*, *24 (7)*:1333–1352.

Moog, D.B., P. J. Whiting., and R. B. Thomas. (1999). Streamflow record extension using power transformations and application to sediment transport. *Water resource research*, *35*(1): 243-254.

Nagpaul, P.S., (2005), Time Series Analysis in Windaus, Unesco, New Delhi, India, taken from http://portal.unesco.org/ci/fr/files/18650/11133194701TimeSeriesAnal.pdf/TimeSeries-Anal.pdf.

Nash, J. E. and Sutcliffe, J. V., (1970), River flow forecasting through conceptual models, Part I - A discussion of principles, *Journal of Hydrol.*, *10* (3), 282–290.

NDT resource center, (visited on June 6, 2014), http://www.ndt-ed.org.

Nguyen, V.T.V., and Bisson, J.L., (1998), Validation En Temps Réel Des Données Des Apports Naturels Journaliers Pour La Gestion Des Réservoirs. *Canadian Journal of Civil Engineering, 25 (6):* 1096–1102.

Nott, D., Marshall, L., and Brown, J. (2012). Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection? *Water Resources Research*, *48* (12).

Oh, C. (2005), The Contribution of Tourism Development to Economic Growth in Korean Economy, *Tourism Management*, *26* (1), 39-44.

Özçelik, C., and E. Benzeden. (2010). Regionalization approaches for the periodic parameters of monthly flows: a case study of Ceyhan and Seyhan River basins. *Hydrological Processes, 24* (22): 3251–3269.

Perreault, L. (2011). Vérification de prévisions hydrologiques probabilistes, Version 1.0, J-1809, IREQ, Diffusion Hydro-Québec.

Perreault, L., Hache, M. and Bobée, B., (1996), Montreal, QC (Rapport final du Project C2-Analyse statistique des séries d'apports énergétiques, INRS-Eau, rapport de recherché N° R-474).

Perreault, L., Roy, R., Mathier, L., and Bobée, B. (1995). La combinaison de modèles appliquée à la validation en temps réel des apports naturels aux réservoirs hydriques. *Canadian Journal of Civil Engineering*, *22*(5): 934-944.

Reid, D.J. (1968). Combing three estimates of gross domestic products. *Economica*, 35: 431–444.

Rezaeianzadeh, M., H. Tabari, A. Arabi Yazdi., S. Isik., L. Kalin. (2013). Flood flow forecasting using ANN, ANFIS and regression models. *Natural Computing and Applications*.

Rupp, D. E, Larned S. T., Arscott D. B., Schmidt J. (2008), Reconstruction of a daily flow record along a hydrologically complex alluvial river, *Journal of Hydrology, 359* (1-2):88-104.

Schreiber, P., and S. Demuth. (1997). Regionalization of low flows in southwest Germany. *Hydrological Sciences-Journal-des Sciences Hydrologiques*, *42*(6): 845-858.

Schreider, S. Y., A. J. Jakeman, J. Gallant, and W. S. Merritt. (2002). Prediction of monthly discharge in ungauged catchments under agricultural land use in the Upper Ping basin, northern Thailand. *Mathematics and Computers in Simulation*, *59*(1-3): 19-33.

Shalev-Shwartz , S., and Tewari, A., (2011), Stochastic Methods for ℓ1-regularized Loss Minimization, *Journal of Machine Learning Research, 12*, 1865-1892.

Shamseldin, A. Y., O'Connor, K. M., and Liang, G. C. (1997). Methods for combining the outputs of different rainfall-runoff models. *Journal of Hydrology, 197*, 203–229.

Shen, ZY. Chen, L., Chen, T. (2012). Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China. *Hydrol Earth Syst Sci 16(1)*:121–132. doi:10.5194/hess-16-121-2012.

Shiau, J. T., and H. C. Lee. (2005). Derivation of optimal hedging rules for a water-supply reservoir through compromise programming. *Water Resource Management, 19* (2):111–132.

Simic, Z., N. Milivojevic., D. Prodanovic., V. Milivojevic., and N. Perovic. (2009). SWAT-Based Runoff Modeling in Complex Catchment Areas-Theoretical Background and Numerical Procedures. *Journal of Serbian Society for Computational Mechanics, 3*(1): 38-63.

Simpson, A.R., Dandy, G.C., and MURPHY, L.J. (1994). Genetic Algorithms Compared To Other Techniques For Pipe Optimization. *Journal of Water Resource Planning and Management, 120 (4)*.

Singh, K. K., and Singh, S. V. (1996), Space-time variation and regionalization of seasonal and monthly summer monsoon rainfall of sub-Himalayan region and Gangetic plains of India, *Climate Research, 6* : 251-262.

Sivapragasam, C.,Vasudevan, G., Maran, J., Bose, C., Kaza, S., and Ganesh, N., (2009), Modeling Evaporation-Seepage Losses for Reservoir Water Balance in Semi-arid Regions, *Water resource management, 23* (5), 853-867.

Smithers, J. C., R. E. Schulze., A. Pike., and G. P. W. Jewitt. (2001). A Hydrological Perspective of the February 2000 Floods: A Case Study in the Sabie River Catchment. *WaterSA, 27*(3): 325-332

Sofia, G., Tarolli, P., Cazorzi, F., and Dalla Fontana, G. (2011). An objective approach for feature extraction: distribution analysis and statistical descriptors for scale choice and channel network identification, *Hydrol. Earth Syst. Sci., 15*, 1387–1402.

Sokolov, A. A., and T. G., Chapma. (1974). *Methods for water balance computations: an international guide for research and practice*. The Unesco Press.17-64 pp.

Taylor, J. C., N. Giesen, and T. S. Steenhuis. (2006). West Africa Volta discharge data quality assessment and use. *Journal of the American Water Resources Association, 42*(4): 1113–1126.

Telis, P.A., (2001), Estimation of Infiltration Rates of Saturated Soils at Selected Sites in the Caloosahatchee River Basin, Southwestern Florida, U.S. Geological Survey, Prepared in cooperation with the South Florida Water Management District (*Open-File Report 01–65*).

Turner G.H., Twieg D.B., (2005), Study of temporal stationarity and spatial consistency of fMRI noise using independent component analysis, *IEEE Transactions Medical Imaging, 24* (6), 712 – 718.

Vaze, J., Ph. Jordan, R. Beecham, A. Frost, and G. Summerell. (2011). *Guidelines for Rainfall-Runoff Modelling: Towards Best Practice Model Application.* eWater Cooperative Research Centre.

Vezza, P., C. Comoglio, M. Rosso, and A. Viglione. (2010). Low flow regionalization in north-western Italy. *Water Resource Management, 24* (14): 4049-4074.

Villa-Alvarado, C.J., Delgadillo-Ruiz, E., Mastachi-Loza, C.A., González-Sosa, E., and Norma-Maricela, R.S., (2014), A Physically Based Runoff Model Analysis of the Querétaro River Basin, *Journal of Applied Mathematics, 2014*, 1-12. Article ID 586872, doi:10.1155/2014/586872.

Vogel, R.M., and Stedinger, J.R. (1985), Minimum Variance Streamflow Record Augmentation Procedures. *Water Resources Research*, *21*(5): 715-723.

Wagener, T., H. S. Wheater, and H. V. Gupta. (2004). *Rainfall-runoff modelling in gauged and ungauged catchments:* Imperial College Press, 1-3 pp.

Wan Jaafar, W. Z., J. Liu, and D. Han. (2011). Input variable selection for median flood regionalization. *Water Resources Research*, *47*(7): W07503.

Wang, W., Van Gelder, P. H. A. J. M., and Vrijling, J. K. (2005). Detection of changes in streamflow series in western Europe over 1901–2000. Water Science and Technology: Water Supply : 5(6), 289–299.

Wang, W., Vrijling, J. K. Van Gelder, P. H. A. J. M., and Ma, J. (2006). Testing for nonlinearity of streamflow processes at different timescales. *Journal of Hydrology: 322 (1–4)*, 247–268.

Weglarczyk, S. (1998). The interdependence and applicability of some statistical quality measures for hydrological models, *Journal of hydrology*, *206*, 98-103.

Wen, L. (2009). Reconstruction natural flow in a regulated system, the Murrumbidgee River, Australia, using time series analysis. *Journal of Hydrology, 364* (3-4), 216–226.

Whitley , D., (1994), A genetic algorithm tutorial, *Statistics and Computing, 4* (2), 65-85.

Xiong, L. H. and O'Connor, K. M. (2002). Comparison of four updating models for real-time river flow forecasting. *Hydrological Sciences Journal, 47* (4), 621–639.

Xiong, L. H., Shamseldin, A. Y., and O'Connor, K. M. (2001). A nonlinear combination of the forecasts of rainfall-runoff models by the first-order Takagi–Sugeno fuzzy system. *Journal of Hydrology*, *245*, 196–217.

Yang, J., Reichert, P., and Abbaspour, KC. (2007). Bayesian uncertainty analysis in distributed hydrologic modeling: a case study in the Thur River basin (Switzerland). *Water Resource Res. 43*(10):W10401.

Yeung, C.W. (2005). Rainfall-runoff and water-balance models for management of the Fena valley reservoir, Guam: U.S. (*Geological Survey Scientific Investigations Report 2004-5287*).

Zhou, Y., Xu, J., and Jing, Y. (2010). *Comparison of Centralized Multi-Sensor Measurement and State Fusion Methods with Ensemble Kalman Filter for Process Fault diagnosis*, Control and Decision Conference (CCDC), Xuzhou, Doi: 10.1109/CCDC.2010.5498594.

APPENDIX 1 – ARTICLE 1: AN ALGORITHM FOR SELECTING THE MOST APPROPRIATE METHOD OF NATURAL FLOW RECONSTRUCTION IN HYDROPOWER RESERVOIRS

By Ana Hosseinpour, Leslie Dolcine, and Musandji Fuamba. Submitted to CWRA journal

Abstract

The quality and availability of hydrometric and hydrologic data series are a main concern in hydraulic or hydrologic research. Even in developed countries, the required data are not always measured directly, and when recorded data is available, it may have some level of uncertainties or gaps. River flow is rarely measured directly and technical error, human error, bad weather, and natural disasters can make measurements unreliable. Previous studies have attempted to reconstruct natural flow or complete flow data at a given reservoir in order to increase the quality of information used in water management. These studies have employed various methods, which differ according to their fundamental concepts and equations, the input they required, their uncertainty and flexibility, and their range of application. Each method may have particular value in a given circumstance, depending on data availability and the overall objectives. This paper describes an algorithm for selecting the appropriate method of flow reconstructing, and applies the algorithm to a Quebec reservoir case study where hydrologic and hydraulic data (but not natural flow data) are available. Based on the quality of input data and the objective of producing short time step (daily) and long-term flow data series, the algorithm led to the selection of a Water Balance Equation (WBE) approach as the most appropriate method for this specific case.

Résumé

La qualité et la disponibilité des séries de données demeurent une préoccupation majeure en recherche hydraulique ou hydrologique. Même dans les pays développés, les données nécessaires ne sont pas toujours mesurées directement, et lorsque les données enregistrées sont disponibles, elles peuvent être entachées d'importantes incertitudes. Plusieurs études antérieures ont tenté de reconstruire ou de compléter les apports naturels qui arrivent dans un réservoir donné, afin d'améliorer la qualité de l'information utilisée dans la gestion des ressources en eau. Ces études ont eu recours à diverses méthodes, qui diffèrent en fonction de leurs concepts et équations fondamentaux, de données d'entrées et leur incertitude, et de leurs champs d'application. Chaque méthode peut avoir une valeur particulière dans une circonstance donnée, en fonction de la disponibilité de données et des objectifs généraux. Cet article décrit un nouvel algorithme développé dans le but de sélectionner la méthode appropriée pour la reconstruction des séries d'apports naturels. Cet algorithme est ensuite appliqué à un réservoir au Québec où les données hydrologiques et hydrauliques sont disponibles. Sur base de la qualité de données d'entrée et l'objectif de produire des pas de temps courts (tous les jours) et les séries à long terme des apports naturels, l'algorithme développé a déterminé que c'est l'Équation du Bilan Hydrique (EBH) qui est la méthode la plus appropriée. Une procédure a donc été élaborée sur base de l'EBH pour déterminer l'information nécessaire, en vue de prendre des décisions saines de gestion de l'eau.

Introduction

Knowing the inflow and outflow in reservoirs or basins is essential for estimating water availability, predicting extreme events, designing hydraulic structures and undertaking other activities related to water planning and management. Limited measured flow data causes some uncertainty in water management policies and the design of facilities, which can have substantial financial consequences (Adeloye and Nawaz, 1998). Reconstructing or extending flow data time series can help improving the quality of water management decisions.

For years, regression-based methods were used in data reconstructing and extending (e.g. Kevin, 1996, Rupp *et al.* 2008, Hernandez-Henriquez *et al.* 2010, Kim and Pachepsky, 2010). These methods usually relate the flow value to one or more independent variables such as rainfall and temperature. The regression based methods are still used in some cases because they are fast and simple, and can be developed using a minimal amount of information from the basins. Although the results of this method can be good when rough estimates are adequate, they are not reliable to estimate natural flow (NF) with shorter time scales (daily, hourly).

Hydrologic and hydraulic methods are the two main alternative natural flow reconstructing methods to regression-based methods. Generally, hydrologic methods (such as rainfall-runoff models) are considered those that rely primarily on meteorological and hydrological input data (such as rainfall, snow, temperature) to simulate natural flow in a case study. Hydraulic methods include those based on the water cycle (such as WBE). A literature review on these methods shows that each may be appropriate in a particular case, depending on the availability and quality of data, the desired time step, the flexibility and uncertainty of the reconstructed data, and the climate and length of the reconstructing period. However, there is no complete and general study to help researchers consider all these factors when selecting a flow reconstructing method.

This paper has two main objectives. First, an approach was developed to select an appropriate NF reconstructing method for a given situation based on a literature review. The resulting algorithm was then tested through application in a case study.

Main Methods Of Natural Flow Reconstructing Or Extending

During the past several years, many different methods and models have been developed to reconstruct the natural flow at gauged or ungauged basins. These methods are grouped in three main categories based on their approach to simulate flow: i) hydrologic methods, which calculate flow values using primarily hydrologic and climatic data for the basin, ii) hydraulic methods, which estimate flow values using regulated flow and storage data, and iii) regression-based methods which define flow values based on any available and effective hydrologic, hydrometric, climatic, and physical data for the basin or neighbouring basins.

Hydrologic Models

Hydrologic models include those based on hydrologic data such as precipitation and climate (Hwang *et al.* 2005). Different studies have been undertaken to develop a relationship between flow and climate to identify the predictability of flow or possibility of a non-random pattern in space or time (Fortin, 2001) and in most areas a statistically significant link is evident (Fortin and Slivitzky, 2000). Fortin (2001) found that climate had an obvious influence on runoff. He also evaluated the reliability of different climatic indices to see if there was a non-random pattern in space or time, and identified a statistically significant link between Arctic Oscillation (AO) and runoff in northern Quebec. However, the correlation was found to be sometimes caused by extremes. According to Fortin and Slivitzky (2000), river flows often correlate well with winter temperatures. The winter temperature could be an indicator of how the regional climate is affected by the global phenomenon of AO. However, while the performance of climate-based flow reconstructing methods is good in some areas, questions remain as to their level of uncertainty. Moreover, hydrologic models that use both physical and climate characteristics of the catchment area are generally recommended, especially when flow reconstructing is being done in regulated basins (Hernandez-Henriquez *et al*., 2010).

Rainfall-runoff models are the hydrologic models most often used to estimate runoff in time and space. They can be applied to estimate runoff in different hydraulic systems and land uses where limited observed flow data are available to calibrate the model. Rainfall-runoff models can range from a simple relation between rainfall and runoff to complex models that also consider the hydrologic and physical characteristics of a region. A more complex model does not necessarily produce more accurate results, especially when it relies on spares and non-representative data (Raman *et al.* 1995).

Examples of simple rainfall-runoff models are spatial models that can be developed based on the relation between flow and climate in neighbouring basins. For example, if the rainfall data series of a basin is available but there is no flow data to calibrate the rainfall-runoff model, a simple solution is to find the relation between rainfall and runoff in a neighbouring basin for which these data series are available and then apply this relationship to the rainfall data at the basin of interest (Raman *et al*., 1995).

In addition to simple spatial rainfall-runoff models that relate regional and physiographic characteristics to temporal or spatial flow based on autoregressive techniques, some methods combine the two. Perreault *et al.* (1995) developed a method based on a combination of spatial and temporal rainfall-runoff models that considerably improved results and did not underestimate peak flows.

Rainfall-runoff models (with linear and nonlinear functions) can be classified into three distinct groups: metric (data-based, empirical or black-box), parametric (conceptual, explicit soil moisture accounting or grey box), and mechanistic (physically-based or white box) (Wagener *et al.* 2004). Metric models commonly use basin data series rather than the behaviour of catchment and flow to estimate model structure and parameter values, so do not seem suitable for the spatial extension of data in ungauged basin. Artificial Neural Network (ANN) and Transfer Functions are examples of these methods. The limitation of metric models is partially solved by data-based mechanistic models that "constrain the degree of freedom of such models to those structures that are physically interpretable" (Wagener *et al*., 2004). Parametric models define the structure before use, and need to be calibrated to adjust the parameters, which cannot all be measured independently. Their dependency on flow makes them difficult to apply to ungauged catchments.

Advantages and disadvantages of hydrologic models

Rainfall-runoff models are the main group of hydrologic models that are able to estimate the data flow of different time-steps. Examples of rainfall-runoff models are Thornthwaite-Mather (TM) for calculating monthly flow (e.g. Taylor *et al.* 2006), StormNET for calculating daily or even smaller timescale flow (e.g. Karamouz *et al.* 2011-a), and the Wright model to calculate mean monthly or daily flow (Adeloye and Nawaz, 1998).

The structure of these models affects uncertainty, but also the number of parameters included and the quality of input data. Uncertainty increases along with the number of parameters; however the models need enough parameters to take the effective factors of flow modeling into account, especially when they are used to extend short time-step or long-term data. For example, according to Adeloye and Nawaz (1998) the performance of the Wright model, which only deals with rainfall, evaporation, and soil moisture (but not the physic of runoff process) and their parameters, is especially poor at producing low-flow frequencies. The accuracy of this method decreases even more for smaller time scales.

The model that is best able to consider the climate factors (such as evaporation and snowmelt) with the greatest impact on the region's flow should be selected to simulate the flow. For example, in cold regions with large snow-loads, considering snowmelt in the flow estimation is necessary (Kim and Kaluarachchi, 2013).

One of the problems with models in this group is that they most often need to be calibrated, which presents some difficulties. First, flow data needed to calibrate the model are not available for ungauged basins. Second, calibrating the model manually takes time and, when flow must be estimated for several basins, the process must be repeated for each basin separately because using approximated parameters based neighbouring basins' parameters decreases the certainty. As well, manual calibration always introduces uncertainties.

Mechanistic rainfall-runoff models attempt to relate the model parameter with catchment characteristics to avoid calibration, however, this has not been completely successful (Wagener *et al.*, 2004). Other attempts have been made to calibrate the model on a regional basis, which makes it applicable for ungauged basins. The rainfall-runoff model is calibrated for as many basins as possible and the estimated parameters are then transferred to ungauged basins. Regional values require a rainfall-runoff model simple enough that it does not increase uncertainty along with the model parameters, but not so simple that it fails to capture the process behaviour with a reasonable degree of accuracy (Wan Jaafar *et al.* 2011 and Madsen, 2000). Many studies have used regional calibration to simulate low-flows (e.g., Vezza *et al.* 2010, Schreiber and Demuth, 1997) or floods (e.g. Eslamian, 2010, Wan Jaafar *et al*., 2011). Other studies dealing with the reconstruction of continuous flow data series using regionalization have looked at large discretization (monthly, seasonal, and annual) times (e.g. Singh and Singh 1996, Özçelik and Benzeden, 2010). Examples of rainfall-runoff models applied in regional calibration are IHACRES, which has low complexity (six parameters) (e.g. Kokkonen *et al.* 2003), for daily streamflow prediction, and different versions of the HBV (Mac-HBV, HBV light) model (e.g. Merz and Blöschl 2003, Krysanova *et al.* 1999, Samuel, Coulibaly, and Metcalfe, 2011**)** to areas with significant amounts of snow. Comparison of the lumped or semi distributed HBV model, Nordic HBV (Krysanova *et al.*, 1999), with distributed HBV models, HBV96 (Lindstrom *et al.*, 1997) and HBV-D (Krysanova *et al.*, 1999), shows that the presence of heterogeneity in distributed model versions makes them more accurate (Krysanova et al 1999). While better results are obtained with distributed models, these are also more data intensive.

Generally, the appropriate Rainfall-Runoff (RR) model should be selected based on different criteria (Vaze *et al.* 2011). The most important is data availability. Model selection also depends on the climate of area. If meteorological or geological factors such as snowmelt, evaporation or groundwater are important, the model should be able to take their effects into account. Rural or urban land use also impacts the selection of software. For example, SWAT is mostly applied to simulate the flow in rural basins (Simic *et al.* 2009) and StormNET (Boss International, 2005) is a more efficient flow simulator in urban areas.

Hydraulic Models

One of the most usual hydraulic models of data reconstructing involves applying the WBE to the reservoir or basin. This equation can be written as follows for the reservoir:

$$N_f = Q_{out} - Q_{in} + \Delta S/\Delta t \qquad (1)$$

The hydraulic data required in this model are output discharge Q_{out} from the reservoir, storage volume changes in the reservoir (ΔS), and input discharge to the reservoir Q_{in} —which is the delayed outflow from the upstream sub-basin. Precipitation over the reservoir surface P and evaporation E from the reservoir during the time Δt, and the interaction *Int* between stored water in the reservoir and groundwater are terms included in calculated NF values (N_f) in this simplified WBE. If enough information is available about these terms, they can be separated from NF. If the distance between two reservoirs in the series is long, the hydrograph of input flow to the downstream reservoir is not exactly the same as that of outflow from the upstream reservoir (Das and Saikia 2013). In such cases, the flow routing equation should be used to calculate the flow downstream hydrograph and increase the accuracy of results. For example, Smithers, Schulze, Pike, and Jewitt. (2001) used the Muskingum technique to route flows in river reaches of the Sabie River catchment in South Africa.

The WBE can also be applied to basins or rivers. The Penck-Oppokowa equation, in which water losses are usually neglected (Shiau and Lee, 2005), is a WBE applicable for basins. Sokolov and Chapman (1974) describe few forms of WBE and provided good information about the main water balance components. They described that considering infiltration, evaporation (e.g. Guntner *et al.* 2004) and interaction between groundwater and surface water during the seasons when these terms have considerable values provides more accurate results, though it is somewhat difficult.

Advantages and disadvantages of hydraulic models

WBE is a hydraulic model which can be considered for any closed hydraulic system such as a basin, river, or reservoir. The WBE is an easy method to apply and does not need to be calibrated. This method is easy and fast enough to be used for calculating real-time flow data. Moreover, factors such as snowmelt or evaporation can easily be considered in the equation if they are available (if not, they should be calculated and added to the equation, or be included in calculated NF). Also, if the distance between two reservoirs is considerable, the hydrograph of outflow from an upstream reservoir cannot be assumed as the hydrograph of inflow to the downstream reservoir (Das and Saikia 2013). This assumption could significantly affect the results of the flow calculation.

In cases where almost all the required data for the WBE are of good quality, the result will be reliable. Although this model is mostly applied for long time-step reconstructing (monthly, seasonally) it can also be used for short time steps (daily).

Regression-based Methods

Regression-based methods apply linear or nonlinear regression to relate flow to hydrologic data, hydraulic data or physical characteristics (of the basin in question or neighbouring basins) that are available and impact on flow rate. These methods have some overlap with hydrologic models, which are based on the simple regression that relates runoff to ratio of rainfall.

Where flow data for neighbouring basins are available, a logistic relation can be developed to relate this flow data to some characteristics of that basin, this relationship can then be applied to the basin of interest (flow reconstructing regression in space). For example, Hughes and Smakhtin (1996) explained that a probable method to extend the natural flow for an area could be simply weighting the observed streamflow at one or more gauged basins by the ratio of the catchment areas of the basin of interest to the area of the gauged basins. Jones, Lister, and Kostopoulou (2004) applied a regression-based method to relate the values of the logarithms of river-flow to linear combinations of data on soil moisture and effective precipitation (precipitation minus actual evaporation). Wen (2009) tried to reconstruct flow by relating discharge time series to rainfall and maximum temperature.

The regression method could be developed based on available short-term data and applied to extend the flow series over the whole time period (flow reconstructing regression in time). Simple regression between a basin's short time-flow data series and the long-term flow of a nearby basin (Hernandez-Henriquez *et al*., 2010, Dastorani *et al.* 2010) exemplifies this type of regression based method. In a case study, Taylor *et al.* (2006) developed a statistically-linear model based on regression of rainfall and short-term runoff. Since complete rainfall data were available in this case, the regression could be applied to extend flow data over the whole period.

Other examples of regression-based methods are those that estimate the flow data at ungauged sub-basins using flow data available for the main basin. For example, one method calculates the stream flow in ungauged sub-basins by relating the ratio of slope and area of that sub-basin to those of the larger basin (Schreiber and Demuth, 2002).

Maintenance of variance (Move) is another regression-based method of data reconstructing that preserves both mean and variance, and thus works better than the linear-regression method (Koutsoyiannis and Efstratiadis, 2007). The Move technique (Hirsch, 1982) reconstructed flow based on a linear regression ($\widehat{y_i} = a + bx_i$) in which a and b were calculated in a special way (e.g Moog, Whiting, and Thomas. 1999).

Advantages and disadvantages of regression-based methods

Regression-based methods are developed based on a mathematical relationship between flow in a basin and independent variables from the same or neighboring basins. These methods have few independent variables and do not usually take the physics of the system into account, which reduces the certainty, especially when they are used to reconstruct or extend short time-step and long-term data.

Common regression-based methods include the normal ratio method and the correlation method. Comparing the results of these traditional methods with Artificial Neural Networks (ANN) and Adaptive Neuro-Fuzzy Interaction System (ANFIS) showed that ANN and ANFIS, especially ANFIS, was better able to reconstruct the missing data (Dastorani *et al.* 2010).

The area ratio method is another regression-based method of flow reconstructing and extension. The problem with this method is that even adjacent basins are rarely linearly related to the catchment area. Also, it is possible to have a trend or non-stationary in the actual stream flow data series at the site or stations used for interpolation (Hughes and Smakhtin, 1996). For areas where hydrology and morphology are different even in neighbouring basins, it is not recommendable to directly transfer their parameters to the basin of interest.

Nevertheless, regression-based methods are simple and fast (Rezaeianzadeh *et al.* 2013) and can be developed easily when available data are limited.

The advantages and disadvantages for all the three groups are summarized in Table 2.

Proposed Methodology For Choosing The Appropriate Method For NF Reconstructing

A number of criteria should be considered when selecting a flow reconstructing method. These can be summarized as:

- required flexibility
- requiring input data
- quality of input data
- desired flow time-step and period
- desired certainty
- climate and other features of the area

In this paper, an algorithm has been developed to select the flow reconstructing method that best responds to above criteria. There is no absolute best method and the choice depends on the features and objectives of a given case study. A step-by-step decision making algorithm is proposed based on factors summarized in Figure 1 and Tables 1 and 2.

The initial step is illustrated in Figure 1. A group number is chosen for the case study. First the proper time-step required for data reconstructing is selected. In this paper short time step refers to daily time steps or less and every time step longer than that is called a long time step. Then, one of the short-term or long-term boxes is selected. Short term is selected if the reconstructing method is meant to produce a small number of data, and long term is selected if long series of data are to be reconstructed. For example, producing flow data for five days on a daily scale is considered short-term data, but if the scale were hourly it would be considered long-term data reconstructing. The researcher then needs to analyse the available data, define its quality and determine the group number, which is the last row of Figure 1. There are different methods of assessing data quality but they are not beyond the scope of this paper. When data quality is low, decisions about the reconstructing method are more critical. It is highly recommended to validate input data before applying them in any flow reconstructing method.

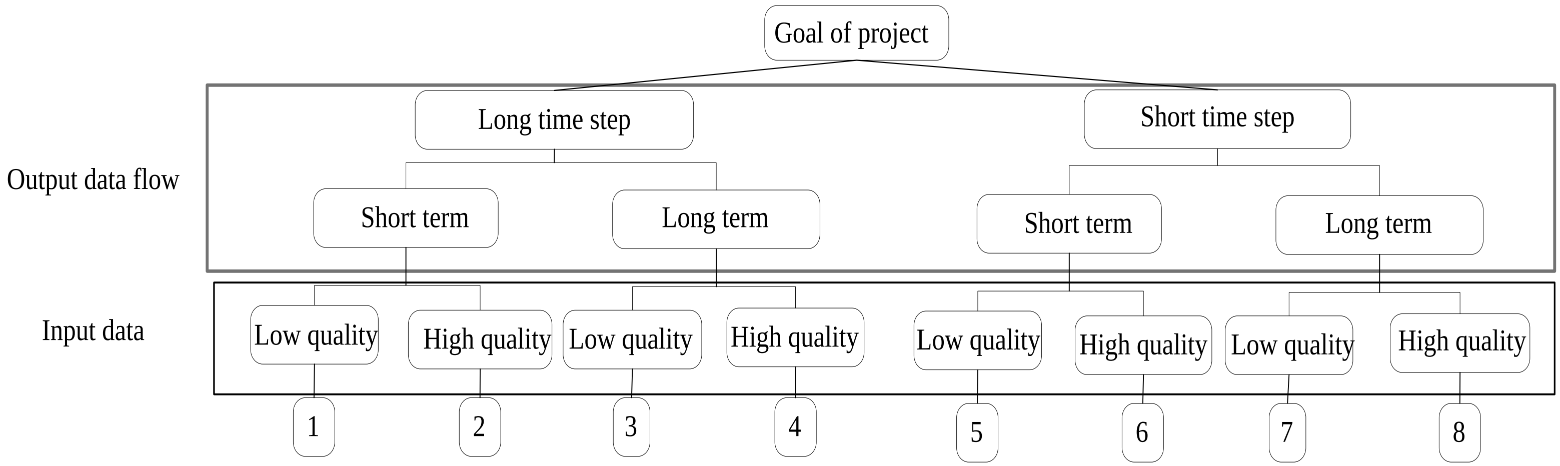


Figure 1: Preliminary algorithm for determining the group number of appropriate method for data reconstructing

Table 1: Determination of the appropriate method code for the identified group number from Figure 1 to be applied in Table 2

Group number from Figure 1	Method group	Method Type	Method Code
7,5	Hydraulic	WBE	*I*
	Hydrologic	Rainfall-Runoff Models	*II*
8	Hydraulic	WBE	*I*
	Hydrologic	Rainfall-Runoff Models	*II*
	Regression based		*IV*
1, 2, 3, 4, 6	Hydraulic	WBE	*I*
	Hydrologic	Rainfall-Runoff Models	*II*
		Climate model	*III*
	Regression based		*IV*

Table 2: Advantages and disadvantages of different groups of natural flow reconstructing methods

Method code from Table 1	Advantages	Disadvantages	Comments
I	• Simple and accurate when all data are available for the basin or reservoir in question • No need for calibration • Results can be calculated quickly • Flexible enough to be applied in all basins	• Difficult to calculate losses if they are not available • The uncertainty of one basin or reservoir highly affects those downstream	• If applied in a region where snow or evaporation are significant, these should be considered in the equation • Data validation is recommended before using the WBE if it is applied for short time step data reconstructing
II	• Can be used in time and space • Are some of the most reliable methods of flow reconstructing • Most have special capability in considering snow, evaporation, infiltration, etc.	• Metric and mechanistic models require that data be calibrated • It is time consuming to model different basins of an area • Models need to be calibrated for each basin separately • Require a lot of data	• Increasing the number of parameters does not necessarily mean greater accuracy • Selecting the rainfall-runoff model depends on: ▪ Available data ▪ Climate of area (if considering evaporation or snowpack is important for the case) ▪ Land-use (usually RR models are designed either for urban or rural areas)
III	• Climate signal data is usually available	• Model parameters change from basin to basin (not flexible) • Does not consider the physics of flow • Not easy to find the climate index which affects flow • Needs calibration	• Climate data sometimes needs to be downscaled
IV	• Results can be calculated quickly • Few parameters need to be defined • Simple to apply • Applicable when limited data are available	• Medium to low certainty • Needs to be updated over time with new data • Requires finding the parameters for each basin. Uncertainty of model increases as the time-step decreases	• Among five regression models of runoff coefficient, single linear regression, monthly linear regression, monthly linear regression with stochastic description for residuals, and a double regressed model, the monthly linear regression model with stochastic description for the residuals has the best results (Raman et al. 1995) • Some other methods like ANN and ANFIS are recommended as alternatives

In the second step, the codes for methods that are likely applicable to a determined group number are taken from Table 1. When more than one method exists for a given group number, this does not mean that they can all be applied to the study area, but rather present the options the researcher should consider. This second level of selection will be done in next step using Table 2, which presents the advantages and disadvantages of each method.

The final step involves using Table 2 to find out more about the strengths and weaknesses of different methods, and make a decision based on these and on data availability. Sometimes data availability is the main driver in the method selection. For example, methods requiring calibration can only be used in gauged basins for which at least short-term time data are available.

Case Study

The proposed algorithm was then applied to a real case study to demonstrate its ability to select the most appropriate method for flow reconstructing. The case-study basin was part of a project aimed at reconstructing daily natural flow in all regulated rivers of the Quebec province in Canada.

Important requirements of this study include:

- The data reconstructing method needs to be flexible enough to be applicable to all regulated rivers in Quebec.
- Data reconstructing is required for daily and long-term scales.
- High-quality flow data is required.
- The daily reconstructing method should be selected based on available data in the area, which is rainfall, minimum temperature, maximum temperature, snowfall, surface area, water level-volume curve of each reservoir, turbine flow from each reservoir, discharged flow through the reservoir gates, and water level upstream and downstream from the reservoir. Natural flow data are available for a limited number of basins.
- Hydrologic data (rainfall, minimum temperature, maximum temperature, snowfall) is of good quality. However, hydrometric data (water level-volume curve of each reservoir, turbine flow from each reservoir, discharged flow through the reservoir gates, and water

levels upstream and downstream of the reservoir) are sometimes noisy and may contain uncertainties.

- Both snowmelt and evaporation affect the results, especially in large reservoirs.

One sub-basin in Quebec was selected as the case study for this paper. This sub-basin, the Outardes 4, has an area of 17119 Km^2 (Figure 2). It has its own reservoir and is located in the upstream basin, thus the only input flow into this reservoir is NF (the flow which is caused by rainfall).

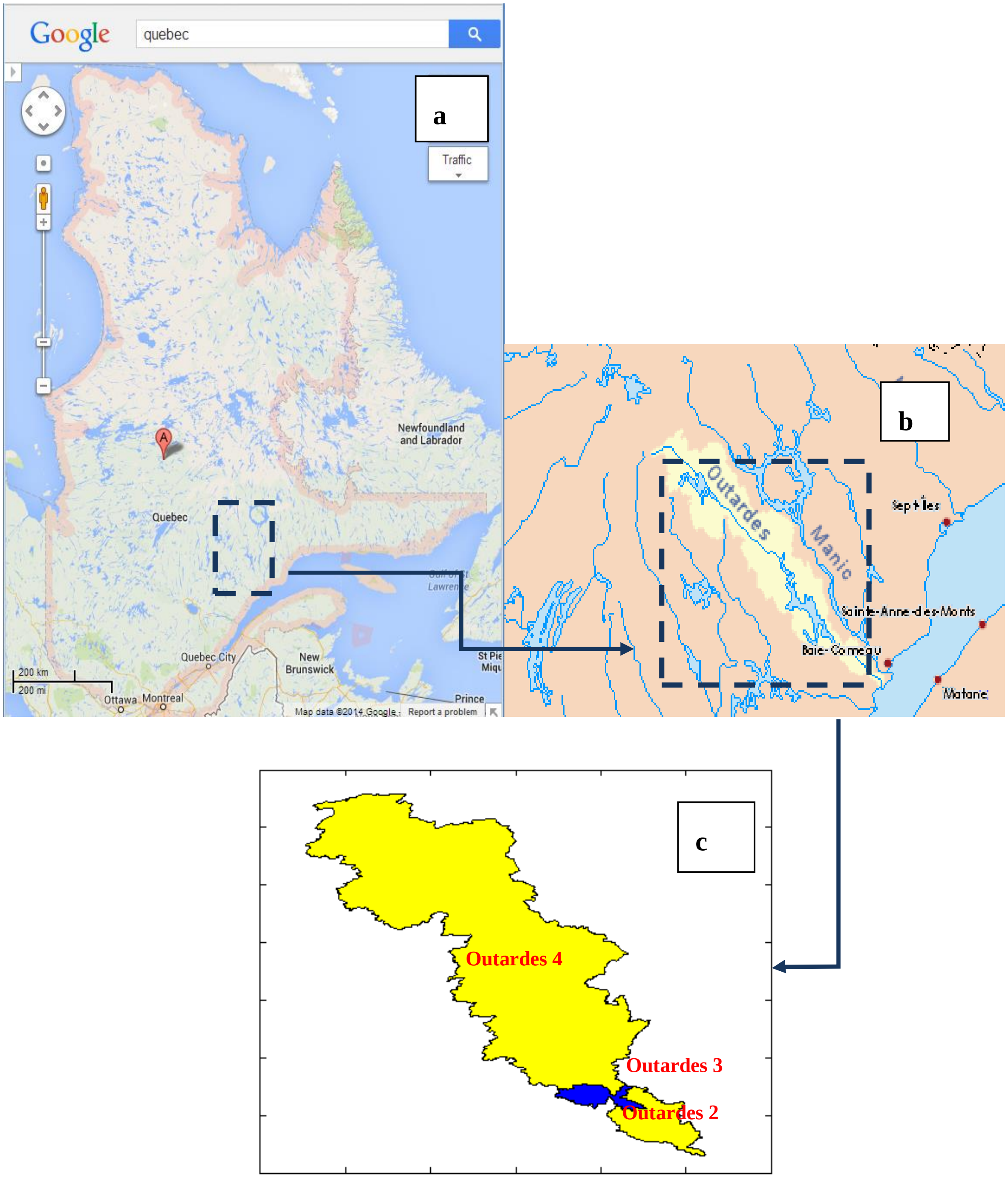


Figure 2: Location of the Outardes Basin in Quebec (a and b), and sub-basins of Outardes 4, Outardes 3 (in darker color), and Outardes 2 (c) (a: https://maps.google.ca, b: www.wikipedia.org)

Applying the proposed algorithm to the case study

The appropriate flow data reconstructing method will be selected in three steps:

1- Defining the group number for appropriate methods of NF reconstructing according to Figure 1.

Considering that the aim of the project is to reconstruct short time-step (daily) and long-term flow, group number 7 or 8 should be selected. If working with hydrometric data of lower quality, the group number will be 7; if hydrological data has been validated or hydrologic data is of higher quality, the group number will be 8. Initially, both groups will be considered because the final decision were made using Tables 1 and 2, which present the advantages, disadvantages and data required with each method.

2- Defining the appropriate method code for the group number using Table 1:

According to Table 1, all hydraulic models (WBE), hydrologic models (rainfall-runoff models and climate models), and regression-based methods are applicable for group numbers 7 and 8. This table shows that the WBE, RR models, and regression-based methods have the group codes of *I*, *II*, and *IV* respectively.

3- Find out more about the advantages and disadvantages of different flow reconstructing methods, referring to Table 2:

The final choice of flow data reconstructing method should be based on the advantages and disadvantages of each. According to Table 2, the WBE with a group code of *I* is a flexible model that can be applied to any basin. This advantage could be important in the current case study, which requires a model that can be applied to all basins in Quebec. It is also simple and fast and does not require calibration. This advantage is also important in the current case study because no measured flow data are available to calibrate the model. However, it will present the inconvenience of necessity of calculating losses and consider evaporation and snow, which can be dealt with by considering these terms in the calculated NF. Moreover, Table 2 recommends data validation for the case study, in which input data are noisy.

According to Table 2, RR and climate models with group code of *II* needs to be calibrated to the basin, which is not possible in the present case study because short-term measured flow data would be required to calibrate these models before applying them to simulate long-term flow.

Although enough data are available to apply regression-based methods under group code *IV*, it has medium to low certainty and uncertainty increases as the time step decreases (Table 2). This method is unlikely to produce certain results when applied to the daily time-step required in the case study.

To summarize, the algorithm for choosing a method of flow reconstructing finds WBE to be the most appropriate model in this case. Regression-based methods are also applicable, but are not recommended. To prove the efficiency of the algorithm, results of the WBE and regression-based methods are then compared.

Results

WBE

Most of the watersheds in Quebec have their own reservoir and there are enough data available (water level-volume curve of each reservoir, turbine flow from each reservoir, discharged flow through the reservoirs' gates, and water level upstream and downstream from the reservoirs) about these reservoirs to allow the WBE to be written for each of them. The general WBE for a reservoir is calculated by Equation 1:

$$N_f = Q_{out} - Q_{in} + \Delta S/\Delta t$$

Regression-based method

The area ratio method, in which the flow data for each basin is related to the flow of neighbouring basin according to the ratio of their surface area, is selected to reconstruct the NF because it can be developed independent of flow data series from Outardes 4. The Moisie basin is chosen as neighbouring basin in this case because: i) it is close to the case study basin and probably has very similar characteristics to the Outardes 4 basin, ii) its area is 19000 km^2, which is close to the area of Outardes 4 and increases the likelihood that they share similar flow characteristics, and iii) measured flow data series are available. The flow data series for this basin were obtained from CEHQ website (CEHQ, 2013).

A comparison of results from the area ratio and WBE methods for the years 2008 to 2012 is presented in Figure 3. To decide on the quality of reconstructed flow, the Nash–Sutcliffe model efficiency coefficient *(NASH)* (Equation 2) and absolute volume error (Equation 3) quality indexes are used to compare the reconstructed flow to available filtered flow data series for the last few years (Table 3). This filtered flow data series is the most reliable flow data for the area, which is calculated using the WBE and filtered manually. Since this data series is only available for the last few years, it cannot be used to calibrate long-term simulations and its applicability is limited to serving as a reference value series for recent years.

$$\mathrm{NASH} = 1 - \frac{\sum_{i=1}^{N}(Q_{fi} - Q_{ri})^2}{\sum_{i=1}^{N}(Q_{fi} - \overline{Q_f})^2} \tag{2}$$

$$AVE = \frac{\sum |Q_{fi} - Q_{ri}|}{\sum Q_{fi}} \tag{3}$$

where *AVE* is the absolute volume error, Q_{fi} is the reference filtered flow for day *i*, Q_{ri} is the reconstructed flow for day *i*, and $\overline{Q_{obs}}$ is the average reconstructed flow. The Nash–Sutcliffe model efficiency coefficient is most applicable for high flow comparisons because the squared difference in this equation increases sensitivity to peak flows (Krause *et al.* 2005). However, in Equation 3, the influence of low flows and high flows are the same. The *NASH* values vary between 1 and $-\infty$ and the closer it is to one, the better. *AVE* also varies between 0 to $+\infty$, and the close it is to zero the better.

Table 3: Quality index comparison for WBE and area ratio

Quality Index	WBE	Area ratio
Nash–Sutcliffe model efficiency coefficient	0.981	0.705
Absolute volume error	0.08	0.263

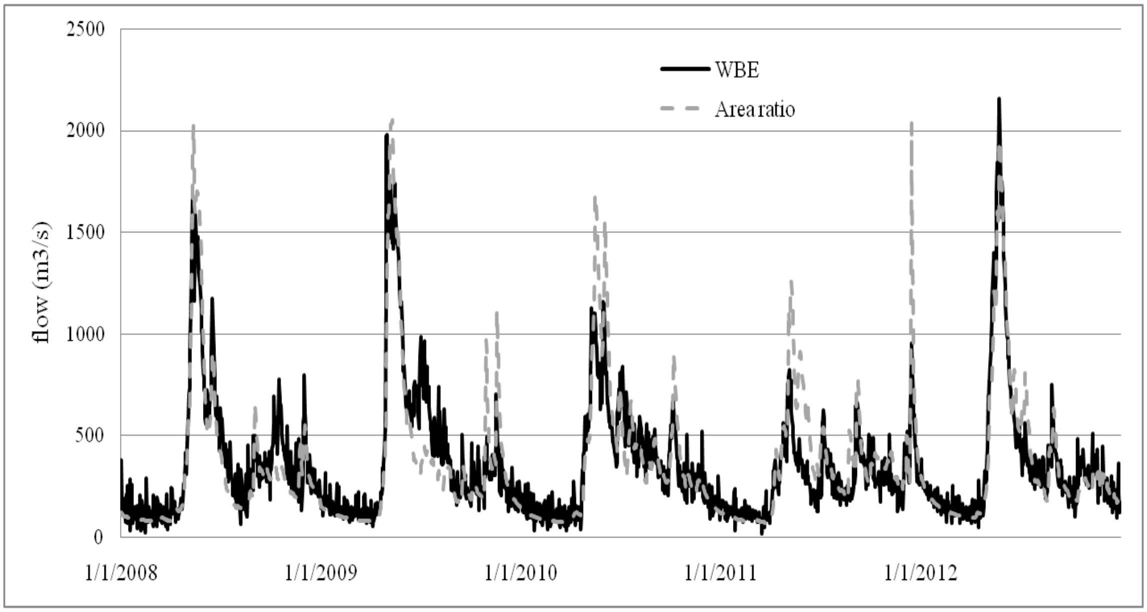


Figure 3: Comparison of reconstructed natural flow using WBE and area ratio

Comparing the results from the WBE and area ratio methods with available filtered flow data series shows that the area ratio method underestimated low flows and overestimated peaks. This shows that while the over and under estimations may compensate each other and the method provides a good estimation of annual flow, the method's performance is not good enough for daily NF reconstructing. Results of the WBE method follow the same trend as filtered flow, but are noisier, especially during low flow. This noise could be related to applying raw input data instead of validated input data in the WBE method. The calculated quality indexes for two methods support the results of visual comparison. The Nash–Sutcliffe model efficiency coefficient is much better for WBE, indicating that this model is more successful in high flow estimations. Moreover, absolute volume error is lower for the WBE, showing that this model provides greater certainty in flow data trend approximation.

The greater reliability of WBE results confirms the efficiency of the proposed method selection algorithm. The algorithm considers all aspect of NF reconstructing and clearly points out the advantages and disadvantages of the different methods.

Conclusion And Recommendations

Knowing the values of surface flow in each basin is important for water planning and management and for hydraulic design. This information is also helpful in estimating water availability, designing flood-warning systems, and conducting studies based on historical flow data. Limited or low quality flow data results in noticeable uncertainty in water management plans and hydraulic design. Methods are required to increase the quality of data through data reconstructing and decrease the likelihood of error in management and design.

The literature review accomplished in the first part of this paper revealed important differences in flow reconstructing methods with regard to their flexibility, requiring input data, output time step and uncertainty. Also, different factors such as climate and the length of the reconstructing period were found to be important factors in the appropriateness of different flow reconstructing methods. However, none of the studies in the review provided a methodology for selecting a flow reconstructing method that considered all these factors. This paper thus undertook to categorize the different methods into three groups — hydrologic models, hydraulic methods and regression-based methods — and then into subgroups. An algorithm was developed to support a method selection based on the factors mentioned and on the advantages and disadvantages of each subgroup of NF reconstructing methods. This algorithm helps to select a proper data reconstructing method in each particular case. Lastly, the applicability of this procedure was tested by applying it to a case study. Results showed that the WBE was the most appropriate method of flow reconstructing for the Outardes basin, given the requirements of the case study. Comparison of the results achieved with this method and with a regression-based method (area ratio) confirmed WBE’s ability to produce more reliable results and supported the efficiency of the algorithm.

References

Adeloye, A. J., and N. R. Nawaz. 1998. Assessing the adequacy of Wright's model reconstructed streamflow data for reservoir storage-yield-reliability analysis. *Water and Environment Journal* 12(2):130-138.

Boss International. (2005). "StormNet and wastewater model user's manual. http://creamdesigners.com/clients/Boss/stormnet-overview.htm.

Das, M. M., and M. D. Saikia. 2013. *Watershed Management*. PHI Learning Private Limited. 227 pp.

Dastorani, M.T., A. Moghadamnia, J. Piri, and M. Rico-Ramirez. 2010. Application of ANN and ANFIS models for reconstructing missing flow data. *Environmental Monitoring and Assessment* 166 (1-4) :421-434.

Eslamian, S. 2010. Flood regionalization using a modified region of influence approach. *Journal of Flood Engineering* 1 (1): 51-66.

Fortin, V., and M. Slivitzky., 2000. *Downward shift in runoff in Northern Quebec-Labrador in the middle eighties and possible links to the Arctic Oscillation*.

Fortin, V. 2001.*Analyse et prévision des apports énergétiques*, Institut de recherche d'Hydro-Québec., technical report.

CEHQ. 2013. *Centre d'expertise hydrique du Québec*. Ministère du Développement durable, de l'Environnement, de la Faune et des Parcs. http://www.cehq.gouv.qc.ca/ (accessed in September 2013).

Guntner, A., M. S. Krol, J. C. de Araujo, and A. Bronstert. 2004. Simple water balance modelling of surface reservoir systems in a large data-scarce semiarid region. *Hydrological Sciences Journal* 49(5): 901-918.

Hernandez-Henriquez, A., T.J. Mlynowski., and S.J. Dery.2010.Reconstructing the natural streamflow of a regulated river: a case study of La Grande Riviere. Quebec, Canada. *Canadian Water Resources Journal* 35(3): 301–316.

Hirsch, R.M., 1982. A comparison of four streamflow record extension techniques. *Water resource research* 18 (4): 1081-1088.

Hughes, D. A., and V. Smakhtin. 1996. Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. *Hydrological Sciences -Journal- des Sciences Hydrologiques* 41(6): 851-871.

Hwang, Ch., M. F. Peng, J., Ning, J. Luo, and Ch. H. Sui.2005. Lake level variations in China from TOPEX/Poseidon altimetry: data quality assessment and links to precipitation and ENSO. *Geophysical Journal International* 161 (1): 1–11.

Jones, P. D., D. H. Lister, and E. Kostopoulou. 2004. *Reconstructed river flow series from 1860s to present (2002).* Science Report SC040052/SR, Environment agency.

Karamouz, M., A. Hosseinpour, and S. Nazif. 2010. Improvement of Urban Drainage System Performance under Climate Change Impact: Case Study. *Journal of Hydrologic Engineering* 16 (5): 395-412.

Kim, D., J. Kaluarachchi. 2013. Predicting natural streamflows in regulated snowmelt-driven watersheds using regionalization methods. *Journal Hydrology and Earth system Sciences (HESS)* 10: 9435-9476.

Kim, J. W., and Y. A. Pachepsky. 2010. Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. *Journal of Hydrology* 394 (3-4): 305–314.

Kokkonen, T. S., A. J. Jakeman., P. C., Young, and H. J. Koivusalo. 2003. Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. *Hydrological Processes* 17(11): 2219–2238.

Koutsoyiannis, D., and A. Efstratiadis. 2007. Uncertainty Assessment of Future Hydroclimatic Predictions: A Comparison of Probabilistic and Scenario-Based Approaches. *Journal of Hydrometeorology* 8 (3): 261–281.

Krysanova, V., A. Bronstert., and D. I. Muller-Wohlfeil. 1999. Modelling river discharge for large drainage basin: from lumped to distributed approach. *Hydrological Sciences Journal* 44(2): 313-331.

Lindstrom, G., B. Johansson., M. Persson., M. Gardelin., S. Bergstrom. 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology* 201: 272-288.

Madsen, H., 2000. Automatic Calibration and Uncertainty Assessment in Rainfall-Runoff Modelling. *Proceedings of Joint Conference on Water Resource Engineering and Water Resources Planning and Management*, Minneapolis, United States, July 30-August 2, 2000..

Merz, R., and G.B löschl. 2003. Regionalisation of catchment model parameters. *Journal of Hydrology* 287(1-4): 95-123.

Moog, D.B., P. J. Whiting., and R. B. Thomas. 1999. Streamflow record extension using power transformations and application to sediment transport. *Water resource research* 35(1): 243-254.

Özçelik, C., and E. Benzeden. 2010. Regionalization approaches for the periodic parameters of monthly flows: a case study of Ceyhan and Seyhan River basins. *Hydrological Processes* 24 (22): 3251–3269.

Perreault, L., R. Roy, L. Mathier, and B. Bobée. 1995. La combinaison de modèles appliquée à la validation en temps réel des apports naturels aux réservoirs hydriques. *Canadian Journal of Civil Engineering 22(5)*: 934-944

Raman, H., S. Mohan, and P. Padalinathan. 1995. Models for extending streamflow data: a case study. *Hydrological Sciences Journal* 40 (3): 381-393.

Rezaeianzadeh, M., H. Tabari., A. Arabi Yazdi., S. Isik., L. Kalin. 2013. Flood flow forecasting using ANN, ANFIS and regression models. *Natural Computing and Applications*.

Samuel, J., P.Coulibaly, and R.A.Metcalfe. 2011.Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. *Journal of Hydrologic Engineering* 16 (5): 447-459.

Simic, Z., N. Milivojevic., D. Prodanovic., V. Milivojevic., and N. Perovic. 2009. SWAT-Based Runoff Modeling in Complex Catchment Areas-Theoretical Background and Numerical Procedures. *Journal of Serbian Society for Computational Mechanics* 3(1): 38-63.

Schreiber, P., and S. Demuth. 1997. Regionalization of low flows in southwest Germany. *Hydrological Sciences-Journal-des Sciences Hydrologiques* 42(6): 845-858.

Schreider, S. Y., A. J. Jakeman, J. Gallant, and W. S. Merritt. 2002. Prediction of monthly discharge in ungauged catchments under agricultural land use in the Upper Ping basin, northern Thailand. *Mathematics and Computers in Simulation* 59(1-3): 19-33.

Shiau, J. T., and H. C. Lee. 2005. Derivation of optimal hedging rules for a water-supply reservoir through compromise programming. *Water Resource Management* 19 (2):111–132.

Singh, K. K., and S. V. Singh. 1996, Space-time variation and regionalization of seasonal and monthly summer monsoon rainfall of sub-Himalayan region and Gangetic plains of India, *Climate Research*6 : 251-262.

Smithers, J. C., R. E. Schulze., A. Pike., and G. P. W. Jewitt. 2001. A Hydrological Perspective of the February 2000 Floods: A Case Study in the Sabie River Catchment. *WaterSA* 27(3): 325-332.

Sokolov, A. A., and T. G., Chapma. 1974. *Methods for water balance computations: an international guide for research and practice*. The Unesco Press.17-64 pp.

Taylor, J. C., N. Giesen, and T. S. Steenhuis. 2006. West Africa Volta discharge data quality assessment and use. *Journal of the American Water Resources Association* 42(4): 1113–1126.

Vaze, J., Ph. Jordan, R. Beecham, A. Frost, and G. Summerell. 2011. *Guidelines for Rainfall-Runoff Modelling: Towards Best Practice Model Application.* eWater Cooperative Research Centre.

Vezza, P., C. Comoglio, M. Rosso, and A. Viglione. 2010. Low flow regionalization in north-western Italy. *Water Resource Management* 24 (14): 4049-4074.

Wagener, T., H. S. Wheater, and H. V. Gupta. 2004. *Rainfall-runoff modelling in gauged and ungauged catchments:* Imperial College Press, 1-3 pp.

Wan Jaafar, W. Z., J. Liu, and D. Han. 2011. *Input variable selection for median flood regionalization. Water Resources Research*, 47(7): W07503.

Wen, L. 2009. Reconstruction natural flow in a regulated system, the Murrumbidgee River, Australia, using time series analysis. *Journal of Hydrology* 364 (3-4), 216–226.

HydroQuebec.1996. HQ. http://www.hydroquebec.com (accessed in September 2013).

APPENDIX 2 - ARTICLE 2: NATURAL FLOW RECONSTRUCTING IN UNGAUGED BASINS USING NEW KALMAN FILTER AND WATER BALANCE BASED METHODS (PART 1: THEORY)

By Ana Hosseinpour, Leslie Dolcine, and Musandji Fuamba. <u>Published at Journal of Hydrologic Engineering, ASCE</u>

Abstract

Abstract: Because natural flow (NF) values are either not directly measured or have the potential to contain considerable error, when deemed necessary, the reconstruction of a reliable NF series is ostensibly important. Selecting the appropriate method depends on available data. For a time period before reservoir construction (pre-reservoir construction period), the only available data for ungauged basins came from the neighboring basins and simulated flow used in a rainfall-runoff model. A new Kalman-based method developed in this paper looks to reconstruct the NF series using the state fusion technique, which is then compared with the area ratio method, the maintenance of variance (Move) type III method, and the multivariable regression method using different quality indexes (QIs). In the perspective of the post-reservoir construction period, when hydrometric data (i.e., turbine flow, water level in the reservoir, and discharged flow) is collected in an ungauged basin (with no flow measurements), a new water balance equation (WBE)-based method is recommended for reconstructing and filtering the NF data using an optimization technique that would then be compared with the classic WBE that implements different QIs. DOI: 10.1061/ (ASCE)HE.1943-5584.0000977. © 2014 American Society of Civil Engineers.

Author keywords: Water balance equation; Stochastic method; Deterministic method; Natural flow estimation; Data filtering; Ungauged basins; Kalman filter; Maintenance of variance; Area ratio; Multivariable regression.

Introduction

Since Natural Flow (NF) values are either not directly measured or have the potential to contain considerable error, when deemed necessary, the reconstruction of a reliable NF series is evidently important. In fact, a locally and regionally reliable reconstructed NF would be helpful for flow prediction, water resource management and planning, frequency analyses, system simulation, and climate change along with extreme event studies.

Although, rainfall-runoff models are the most usual models to reconstruct the NF in gauged basins, the options are much more limited in ungauged basins. Logistic models (i.e. area ratio, Move, Regression method) are the simple regression based methods which are mostly developed on the premise of available flow data from neighbouring basins. However, their reliability varies because even neighbouring basins are not identical in terms of climate and physical characteristics. Classic WBE is also a simple method which can be applied in the ungauged basins where hydrometric data are available. This equation can be written for the basin or reservoir as a closed hydraulic system. Since the results of this model are very sensitive to quality of input data, they are usually noisy and uncertain. In this paper, a new Kalman filter method added with stochastic and deterministic WBE based methods are recommended to reconstruct and filter NF values. These methods are developed to remove the noise from NF data series and improve their certainty.

Thus, the objectives of the present project can be summarized as: i) determining the NF reconstructing methods which can be applicable under real conditions in ungauged basins, and ii) improving the classic WBE and Kalman filter methods in order to remove the noise from NF values and predict the results (the uncertainties caused by wind are not considered in this paper). The definition of NF in this paper is the runoff caused by effective rainfall. This NF can be calculated using classic WBE as follow:

$$NF=Q_{out}- Q_{in}+ (\Delta S/\Delta t) \tag{1}$$

where $Q_{out,n}$ is the output discharge from the reservoir number n (R_n), ΔS the storage volume changes in the reservoir number n, $Q_{in,n}$ the input discharge to reservoir number n, and NF_n the unknown NF value (which includes all the losing terms) to reservoir number n.

As illustrated in Figure 1, $Q_{in,n}$ is the regulated outflow from the upstream reservoir (if there is one), $Q_{out,n-1}$, which is the summation of turbine flow from the reservoir number *n-1* ($Q_{tr,n\ 1}$) and discharged/spilled flow from reservoir number *n-1* ($Q_{sp,n-1}$).

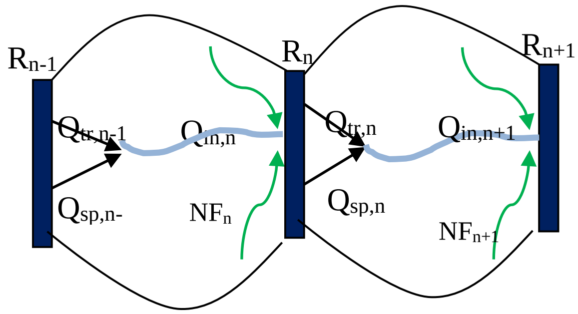


Figure 1: Schematic of three reservoirs in series

Literature review

Depending on the goal of project and available data, several methods have been developed to reconstruct the NF. The selection of method highly depends on available data which can be different before or after the existing of reservoir in the basins. Thus, the study of these models is performed for the two Pre- and Post-Reservoir Construction periods. Nevertheless, some methods such as rainfall-runoff methods (Karamouz *et al.* 2011a, Karamouz, *et al.* 2011b, Hernandez-Henriquez *et al.* 2010) are applicable independent of the reservoir existence but usually need to be calibrated. HSAMI (Nicol 2010 and Bisson 1995) is a rainfall-runoff model, which allows the physical process of weather in the watershed to be reproduced based on WBE. The input data are weather parameters including five categories: evaporation, vertical flow, horizontal flow, surface runoff and snow. From the input data, this model simulates the daily flow at the outlet of the catchment.

Pre-Reservoir Construction Period

For the pre-reservoir time period, WBE can be written for the whole basin as a closed system if there is available data. In cases where limited data are available, the use of simple flow reconstructing methods is unavoidable. For example, the flow series can be extended by weighting the observed stream flow at one or more neighbouring gauged basins. This weight is the ratio of the catchment areas of the basin of interest to the gauged basins (Hughes and Smakhtin 1996, Schreider et al. 1997). Jones *et al.* (2004) developed a regression between the logarithms of river-flow, soil moisture and effective precipitation, which is precipitation minus actual evaporation, and Wen (2009) tried to reconstruct flow by relating discharge time series to rainfall and maximum temperature. This regression can also be developed between short-term flow data series from the basin and long-term flow data from a nearby basin (e.g. Hernandez-Henriquez et al. 2010, Dastorani *et al.* 2010). All these methods can be put in a group called regression based methods. Considering the available data in our case study, area ratio and linear multiple regression are relevant, and were compared to other methods (Kalman and Move III).

Maintenance of variance (Move) is another method of extending flow measurements. This method preserves both mean and variance, and is thus superior to linear regression models, where variance and mean are important in uncertainty estimation (Koutsoyiannis and Efstratiadis, 2007). The Move technique (Hirsch 1982), developed to solve the problem of variance underestimation in linear least squares regression of the logarithmic flows, has been tested by several researchers to extend the flow series. For example, Moog and Whiting and Thomas (1999) used Move to extend the flow series of the Snake River (U.S), and replaced the logarithmic transformation by the more general Box-Cox scaled power transformation to generate more linear, constant-variance relationship for the Move extension. The results of their study show some improvements in flow estimation, especially during low flow.

Different types of Move method are explained by Vogel and Stedinger (1985). All reconstructing of the flow is based on a linear regression ($\widehat{y_i} = a + bx_i$) in which a and b are calculated in a special way. Move type III takes a different approach to defining a and b (Matalas and Jacob 1964) by forcing the mean and variance of produced values to equal the expected values, given $y_1, \ldots y_{n1}$, and $x_1, \ldots x_{n1+n2}$ (where n_l is the length of the short record, and $n_l + n_2$ is the length of the long record). This technique is also implemented in this paper to reconstruct the pre-reservoir NF and compare with the other methods.

Kalman filter is the other method that can be instituted when data series are noisy (Noriega 1992). When more than one data series is available, two types of multi-sensor data fusion (combination of two or more measurement data series of n values) of the Kalman filter are relevant (Zhou *et al.*

2010). One of them begins by fusing the measurements and then filters the fused data series using the Kalman method. Common measurement fusion methods involve simply merging the multi-sensor data or combining them based on minimum mean square error estimates. The other one first filters the data series using a Kalman filter and then fuses them. State estimate covariance matrices are used in state-vector fusion; however, the state estimates from different estimators are usually dependent. A different method of fusing the filtered data series is developed in this paper. This new method was established to reconstruct the NF for the pre-reservoir period and its performance was compared to results from other methods.

Post-Reservoir Construction Period

Classic WBE for a reservoir as a closed hydraulic system is a common method (e.g. Shiau and Lee 2005) of NF reconstructing in ungauged basins. Although many researchers have attempted to estimate the losing terms such as infiltration (Joshi and Tambe 2010, Telis 2001) and evaporation (i.e. Yeung 2005, Gunter *et al.* 2004, Hamon 1961, Sivapragasam *et al.* 2009, and Parasuraman *et al.* 2007), they remain a source of uncertainty. Thus, this equation can be simplified using Equation 1. This simplification, along with input data uncertainty, causes noisy and even negative values of reconstructed NF. Filtering and validating the reconstructed flow, however, would improve the quality of reconstructed NF. Perreault (2011) suggested a WBE-based optimization model estimating the hourly NF and minimizing noise and different errors. This optimization model was expressed as follows:

$$Minimize \sum_{n=starthour}^{endhour} \varepsilon_{\inf(n)}^2 + \sum_{n=starthour}^{endhour} \varepsilon_{\sup(n)}^2 + \gamma \times C \times \left(p \times \sum_{n=starthour}^{endhour} \xi^2 + q \sum_{n=starthour}^{endhour} \delta^2\right) \tag{2}$$

Subject to:

$$volume_{(0)} - \varepsilon_{\inf(0)} \geq V_{\inf(0)} - Offset$$

$$volume_{(0)} - \varepsilon_{\sup(0)} \leq V_{\sup(0)} - Offset$$

$$volume_{(0)} + \sum_{i=0}^{n} NF_i + \sum_{i=0}^{n} Qin_i - \sum_{i=0}^{n} Qout_i + \varepsilon_{\inf(n+1)} \geq V_{\inf(n+1)} - Offset$$

$$volume_{(0)} + \sum_{i=0}^{n} NF_i + \sum_{i=0}^{n} Qin_i - \sum_{i=0}^{n} Qout_i - \varepsilon_{\sup(n+1)} \leq V_{\sup(n+1)} - Offset$$

$$NF_{(n+1)} - NF_{(n)} = \xi_{(n)}$$

$$NF_{(n)} - NF_{(n-1)} - NF_{(n+1)} + NF_{(n)} = \delta_{(n)}$$

$$Offset = (V \inf_{\min} + V \sup_{\max})/2$$

$$n = starthour, ..., endhour$$

$$-\infty < \varepsilon < +\infty$$

$$0 < NF < +\infty$$

$$-\infty < \delta < +\infty$$

$$-\infty < \xi < +\infty$$

where Q_{in} is the input flow to the reservoir (which is the released flow from the upstream reservoir), Q_{out} the output flow from the reservoir (which is the summation of turbine and spilled flows), $V_{\inf(n)}$ minimum volume of the reservoir during the n^{th} hour, and $V_{\sup(n)}$ maximum volume of the reservoir during the n^{th} hour. *NF*, ε, ξ, δ, and $volume_{(0)}$ are variables. $volume_{(0)}$ signifies the volume of the reservoir at the first hour, *NF* points to the NF entering the reservoir during n^{th} hour, ε is WBE error, and ξ, δ are the variation of NF during the 2 and 3 consequent days respectively. In Equation 2 *C*, *p*, *q*, γ, and time interval of *dn=1+end hour-start hour* represent the model's parameters. In this equation *C, p, q,* and γ are the parameters to allocate the weight to the variables and dn marks the number of days for which the WBE is solved. Thus, the objective function is minimizing the NF variation and WBE error, the first two constrains force the $volume_{(0)}$ to stay between minimum and maximum measured volume and the second two constrains control the range of WBE.

Although the Perreault model produces better measurement of NF than classic WBE, it still presents some deficiencies. Results are still noisy during low flow. Also, the model was developed to calculate hourly flow and is not very good at estimating longer time steps such as daily flow. Moreover, two fixed values of *C* coefficient were selected for winter and summer, and were deemed the best *C* coefficients for each season regardless of the year and the reservoir. Still, Perreault's model presents a large advantage over the classic WBE. The model includes windows with the length of *dn* and all flow values in a window are estimated considering that set of flow values. But these windows do not move and the days near the boundary of windows are affected by poor boundary conditions, including the assumption of variables like δ and ξ equal to zero for the first and last time step of each window. Also, the length of the window is considered as a fixed value, which begs the question about how sure can we be that this window size is best for all seasons, all years, and all reservoirs.

Several methods are available to determine the parameters of optimization models. Some of these are neural network (Cheng e*t al*. 2009, Chu 1992), Genetic algorithm (GA), and stochastic methods (Shalev-Shwartz and Tewari 2011). Different types of GA have been widely used to solve the optimization models over the last decade (e.g. Deb 2000, Deb 2002).

In this paper, a posterior deterministic GA is one of the methods engaged to determine the parameters of the optimization model. This posterior GA is capable of automatically selecting a few best parameter sets (PSs) likely to be the most realistic. The other method used in this paper to choose parameter sets is a stochastic method that considers the probability of best PSs to define the final PS.

The main objectives of this paper can be stated as follows:

1) Develop a new state fusion Kalman filter to reconstruct the pre-reservoir NF values;
2) Compare the Kalman method we developed to the Move III, area ratio, and multivariate regression;
3) Reconstruct and filter the daily post-reservoir NF using the Modified Perreault model to change the time scale from hourly to daily, while preserving the concept of minimizing different errors;
4) Define the parameters of the model automatically using deterministic and stochastic methods.

5) Compare the model with classic WBE.

Data stationarity

It is important to analyse whether or not the presence of stationarity in the flow series is created artificially by the reconstruction method. Stationarity of a data series implies homogeneity in the sense that the series behaviour is not dependent on time and its statistical properties stay the same over time. More precisely, in the case of stationarity, joint probability distribution of the process remains unchanged over time. However, it is common to employ a simple stationarity test, which tells if the mean and variance of a data series are constant or not. Although there are some means to evaluate the stationarity of a time series, it is also visible at its time plot: A time series would be stationary if its time plot appears similar at different points along the time axis (Nagpaul 2005).

Perreault *et al.* (1996) assessed the stationarity of average annual aggregate flow using a Bayesian procedure to detect change in mean annual series of flow, and found three different groups with three different averages. Given the limited information available in their project, it was difficult to distinguish between non-stationarity of the mean and the presence of sustained deviations caused by the autocorrelation. Despite the results obtained from the precipitation series, which support the presence of a change in the average, it is difficult to deduce the non-stationarity of the average in the series of annual flow.

Turner and Twieg (2005) divided the data series into S equal segments of size N, and then administered T-statistic with $2N\text{-}2$ degree of freedom to compare the segment means and F-statistic with $N\text{-}1$ degree of freedom to compare the segment variances. They defined a Wide Sense Stationary (*WSS*) index based on the results of F-test and T-test and assumed that the data series would be stationary if *WSS* exceeded 0.9.

Mann–Kendall (e.g. Cunderlik and Burn 2003), Dickey–Fuller, and Augmented Dickey–Fuller (ADF) (e.g. Oh 2005) are some of the tests designed to evaluate the stationarity of data. Kwiatkowski-Phillip-Schmidt-Shin (KPSS) is another method developed for data with short memory, though Lee and Schmidt (1996) deployed it to assess the stationarity of long-memory data and found it to be adequate.

In this paper, a state fusion Kalman filter is applied to estimate the pre-reservoir NF data series. This method is compared with some logistic methods for the same time period. Also, a deterministic and stochastic based WBE is formed to reconstruct the NF after reservoir construction. The quality of reconstructed flow is then evaluated using different QIs before and after reservoir construction. Then the stationarity of data is checked and regional analysis accomplished to ensure coherence between local flow and regional flow.

Methodology

Pre-Reservoir Construction Period

Since WBE is not applicable before a reservoir is built and the only available data for this period comes from neighbouring basin flow and flow simulated by HSAMI (the simulated flow values using this model are available for the current project), few regression based methods are set up for this time period. The applied regression based methods are common simple methods of flow reconstructing. We laid out the new Kalman-based method to reconstruct the NF series using the state vector fusion method.

Multivariable regression

One method for extending NF into the period before reservoir construction is through multivariate regression. In this model, a linear regression is arranged between the flow as a dependent value and logarithmic scaled neighbouring basin flow and logarithmic simulated flow by HSAMI as the independent variables. Here the neighbouring basin flow is scaled using the surface area ratio. This regression was utilized based on the calibration period (the post-reservoir years when NF is reconstructed using methods described under the title "Post-Reservoir Construction Period") and then used to reconstruct the flow for the pre-reservoir years.

Area ratio

In the area ratio method, the neighbouring basin's flow is the only flow that has been used to reconstruct NF. In this method the flow of neighbouring basin is multiplied by a ratio of the basin's surface to the neighbouring basin's surface.

Move III

Move III is a linear regression using a specific method for calculating the slope and constant value of regression to extend a data series. The logarithmic values are used to develop this model.

Kalman Filter

In the state vector fusion Kalman method we developed, measured data series are first filtered using a Kalman filter (Equations 3 and 4) method; the filtered data series is then fused (Equation 5). In this work, the SSM Matlab toolbox is used to filter the NF series and an optimization model is fashioned to fuse the filtered data series. The formulation of this model is as follows:

$$Y_1 = Kalman(y_{HSAMI}) \tag{3}$$

$$Y_2 = Kalman(y_{nb}) \tag{4}$$

$$Minimize(\beta_1^2 + \beta_2^2) \tag{5}$$

Subject to:

$$Y_1 = \beta_1 + \alpha_1 \times y$$

$$Y_2 = \beta_2 + \alpha_2 \times y$$

$$0 \le y < \infty$$

$$\alpha_1, \alpha_2 \in \{[1,1],[1,2],[1.2,1.8],..[2,1]\}$$

where y_{HSAMI} is simulated flow by HSAMI, y_{nb} refers to the scaled flow from the neighbouring basin (neighbouring basin's flow multiplied by area ratio), Y_1 and Y_2 are filtered flow using the Kalman filter, y is shown as the fused flow (reconstructed flow), α_1 and α_2 are the coefficients of fused flow, and β_1 and β_2 are the errors. This optimization model tries to minimize the difference between fused flow (y) and available filtered flow data series (Y1 and Y2).

The neighbouring basin which is adopted in this method is selected from watersheds for which measured flow series are available. The basin is chosen based on the similarity of its size and its flow trend to case-study (similarity to simulated flow by HSAMI).

The three quality indices to evaluate the soundness of reconstructed pre-reservoir flow are Normalized Nash (*NN*), Consistency Coefficient (*CC*), and Normalized Tortuosity (*NT*) as follows:

$$\mathrm{NN} = 1 - \frac{(1-\mathrm{NASH})}{\sqrt{(\mathrm{a} + (1-\mathrm{NASH})^2)}} \tag{6}$$

$$\mathrm{NASH} = 1 - \frac{\sum_{i=1}^{N}(Q_{obsi} - Q_{cali})^2}{\sum_{i=1}^{N}(Q_{obsi} - \overline{Q_{obs}})^2}$$

$$CC = \frac{\sum n}{N} \tag{7}$$

$$\begin{cases} a' \times b' > 0 \rightarrow n = 1 \\ a' \times b' < 0 \rightarrow n = 0 \end{cases}$$

$$a' = NBF_{i+1} - NBF_i$$

$$b' = Q_{cali+1} - Q_{cali}$$

$$NT = \frac{1}{1 + (\frac{T}{N})} \tag{8}$$

$$T = \sum | Q_{cali+1} - Q_{cali} |$$

where Q_{obsi} is the observed flow in case-study basin (simulated flow by HSAMI is considered as observed flow to calculate NASH) in day i, Q_{cali} the reconstructed flow by developed model for day *i*, NBF_i the flow of *i*th day in the neighbouring basin, terms a' and b' are the differences discharges of the days *i+1* and *i* of the selected neighbouring basin and calculated flows respectively, *n* marks the number of days that the trend of both calculated flow (Q_{cali}) and neighbouring basin flow (*NBF)* is increasing or decreasing, *N* counts the total number of days, and *T* represents tortuosity. The neighbouring basin is selected from watersheds for which measured flow series are available. This selection is based on i) the similarity of flow trend (from HSAMI) and neighbouring basin and ii) the size similarity of neighbouring basin to the interested basin. The reasons for selecting these QIs are explained later under the title of "Quality evaluation of reconstructed NN".

.

Post-Reservoir Construction Period

Deterministic method

In this paper, the Perreault optimization model is modified to calculate the daily NF and minimize noise and errors. Like Perreault's model, the new optimization estimates the NF of ungauged basins by solving the WBE for each reservoir, and a single quadratic objective function is used to minimize the error of WBE and decrease the variation of NF over continuous days. One of the strengths of this model over the classic WBE is that it calculates and filters flow values at the same time. The modified Perreault model is written as follows:

$$Minimize(\gamma \times \sum_{n=startday}^{endday} \varepsilon_{(n)}^2 + C \times (p \times \sum_{n=startday}^{endday} \xi^2 + q \sum_{n=startday}^{endday} \delta^2) + d \times \sum_{n=startday}^{endday} a^2) \tag{9}$$

Subject to:

$$a = Q_{simn} - NF_n$$

$$\mathrm{Q_{inn} - Q_{outn} + NF_n} + \varepsilon_{\mathrm{n}} = (\mathrm{RV_{n+1} - RV_n}) \times (\frac{1000000}{24 \times 3600})$$

$$NF_{(n+1)} - NF_{(n)} = \xi_{(n)}$$

$$NF_{(n)} - NF_{(n-1)} - NF_{(n+1)} + NF_{(n)} = \delta_{(n)}$$

$$n = startday, ..., endday$$

$$-\infty < \varepsilon < +\infty$$

$$-\infty < \xi < +\infty$$

$$0 < NF < +\infty$$

$$-\infty < \delta < +\infty$$

where Q_{simn} is the simulated flow by rainfall-runoff model for day *n, d* is a vector of weights defined by the user, and RV_n is the volume of water in the reservoir at the beginning of n^{th} day. *NF*, ε, ξ, a, and δ are the variables and *C, p, q*, γ, and time interval of *dn=end day-start day* are parameters. Unlike the Perreault model, here the time interval is a moving window. Every time the optimization problem is solved, the NF is calculated for all days in the window, but the only flow that is kept is from the day located in the middle of window. In this way, every reconstructed NF value is calculated considering the effects of nearby days. This represents a significant advantage over classic WBE, which calculates each day's flow value independently, and the Perreault model, which considers a non-moving window and estimates uncertain flow values for the first and last time step of each interval. As can be seen in the objective function (Equation 9), Q_{simn} is entered as an external signal to the model. That is, the model tries to get close to this simulated flow (by HSAMI) by the weight of *d*.

To evaluate the effects of five parameters of the optimization model on results, the model has been solved using several combinations of parameters and it is concluded that the model is very sensitive to *dn*, *C*, and γ changes. Moreover, it is understood that these parameters are not constant during a year. They depend on the season or, more precisely, on the temperature.

To select a proper parameter set for each time period, each year is split into segments. The cumulative HSAMI flow data series is calculated and segment starts are considered as the points where the slope of the NF series shows considerable change. The number of segments is kept at less than 8 for each year to avoid unnecessary complexity. Then, a GA model is applied to select the most reliable parameters for each segment.

The algorithm of the proposed method is presented in Figure 2. It provides an automatic method for calculating and filtering daily flow data. The algorithm benefits from GA (Whitley 1994) to define the parameters of the optimization method (bolded boxes are the GA's steps numbered from 1 to 9). Since the NFs are calculated for ungauged basins, it is difficult to define a reliable quality index to determine the certainty of reconstructed NF. Thus, a GA with a posterior approach was used to define all parameters. The posterior GA is created by solving the GA for different *d* coefficient. As can be seen in Figure 2, for each segment (which is a specific time period in a year) and each predefined *d* coefficient (defining the level of similarity between calculated NF and simulated flow using HSAMI), the GA finds the best coefficient sets. This approach allows for more than one parameter set and thus more than one reconstructed NF series and can select among those.

Step 1 defines the preliminary parameters (genes). GA has five genes of , *p*, *q*, c, *dn* (five parameters) initialized for each segment and each *d* value (Box 1). Step 2 defines the fitness of each parameter set (Box 2). To do so, the optimization model must be solved using the parameters selected in Step 1 to estimate the NF series and then estimate their quality indices and consider their fitness. The selected quality index is the Nash–Sutcliffe coefficient (1970), which compares measured and calculated data. Since no measured flow is available for the area, the simulated flow by HSAMI is considered as the observed flow. Step 3 selects half the initial population with higher fitness as parents of next generation using the tournament method (Box 3). Step 4 completes cross-over between each random pair of parents (Box 4), and Step 5 undertakes mutation by the chance of 2 percent (Box 5).

Steps 3 to 5 are then performed *m* times (*m*=number of iterations defined by the user) to optimize the parameters. The result of the current segment and *d* are saved and Steps 1 to 6 will be done for the next *d* value of the current segment. At the end there will be *d* estimated NF series for each segment. Considering all combinations of these *d* series in *n* segments gives $s=d^n$ reconstructed NF series for each year. Among these *s* series, the graph that best matches the regional flow or the graph that is most logical according to upstream and downstream basin flow, (if there is one) can be selected. Here, the best graph is selected based on five quality indices and trend similarity to

NF calculated by classic WBE. First, few graphs with higher quality indices are selected and the final graph will be selected visually from among them. The final graph should be smoother than other graphs and should not display over or under-estimations in comparison with classic WBE. These five quality indices are the same ones explained later under the title of "Quality evaluation of reconstructed NN".

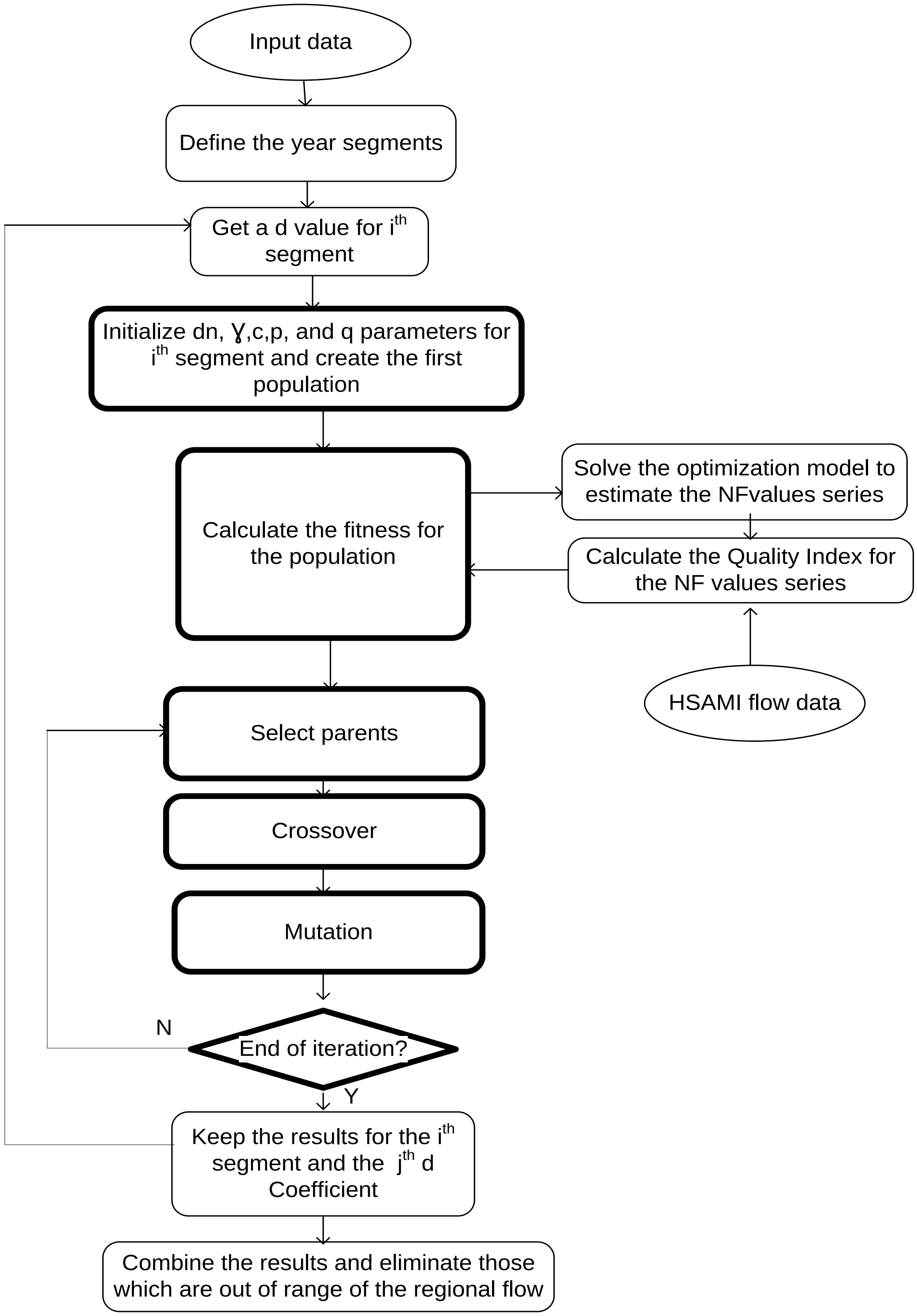


Figure 2: Schematic of the proposed Deterministic based Method to reconstruct daily NF for the Post-Reservoir Construction Period

Stochastic method

The parameters of the optimization model were also defined using a stochastic method. This time, the objective function (Equation 9) was changed by removing parameters p and q, the external signal ($Qsim_n$) and, consequently, the last item of the objective function. Equation 10 presents this modification.

$$Minimize\ \ (\gamma \times \sum_{n=startday}^{endday} \varepsilon_{(n)}^{2} + C \times (\sum_{n=startday}^{endday} \xi^{2} + \sum_{n=startday}^{endday} \delta^{2})) \tag{10}$$

A stochastic multi objective model method is then developed to define the parameters of this model (*γ, C, dn =end day-start day).* The advantage of this method is that more than one index is applied to evaluate the soundness of the data series and the probability of different parameter sets is taken into account.

As in the deterministic method, each year is split into segments and the model's parameters are defined for each segment. Then, the 8-step stochastic based WBE is completed as in Figure 3 for each segment.

Step 1: One hundred random initial parameter sets are produced (Fig 3, Box 1).

Step 2: The optimization model is solved with each parameter set and related NF series are produced. Thus, at the end of this step, one hundred flow data series are in hand (Fig 3-box 2).

Step 3: The fitness of each reconstructed NF series from Step 2 is defined (Fig 3-box 3). Since the model is multi-objective, more than one index is applied to evaluate the soundness of the series. The weighted summation of three normalized criteria is considered as a single index in this model and is maximized .This index is as follows:

$$benefit = W_1 \times NN + W_2 \times NAVE + W_3 \times CC \tag{11}$$

where:

$$NAVE = (\frac{1}{1 + AVE}) \tag{12}$$

$$AVE = \frac{\sum |Q_{WBEi} - Q_{cali}|}{\sum Q_{WBEi}}$$

where *W1*, *W2*, *W3* are the weights defining by the user, *NN*, and *CC* are defined using Equations. 6 and 7 respectively, *NAVE* is normalized absolute volume error, and Q_{WBEi} is the calculated flow by WBE for the day *i*. This index has the advantage of considering different factors in defining the fitness of each data series. The reason and advantages of selecting each of these QIs are explained under the title "Quality evaluation of reconstructed NN".

Step 4: To narrow down the number of parameter sets, the 10 sets with higher fitness are selected (Fig 3-box 4).

Step 5: First the probability of 10 parameter sets from Step 4 are defined as follow (Kalakrishnan et al. 2011):

$$Ps_j = \exp\left(-\frac{1}{benefit_j}\right),\ j = 1,2,...,10 \tag{13}$$

where *P (ps j)* is the probability of j^{th} parameter set, and $benefit_j$ is the fitness related to this parameter set (from Step 3).

Then, one parameter set is estimated using these 10 probabilities (Equation 13) as a mother parameter set of the next iteration (Fig 3, Box 5). Here the parameter sets with higher probabilities have more weight in the mother parameter set definition.

$$C_n = \frac{\sum P(ps_j) \times C_j}{\sum P(ps_j)},\ \gamma_n = \frac{\sum P(ps_j) \times \gamma_j}{\sum P(ps_j)},\ dn_n = round\left(\frac{\sum P(ps_j) \times dn_j}{\sum P(ps_j)}\right),\ \text{j=1,2,...,10} \tag{14}$$

where Cn, γ_n, and dn_n form the parent parameter set of the next iteration, Cj is the C coefficient of j^{th} parameter set, γ_j is the γ Coefficient of j^{th} parameter set, and dn_j is the dn parameter of j^{th} parameter set. The optimization model is solved using this new parameter set and the related fitness is defined.

Step 6: If the difference between defined fitness in Step 5 and fitness from the last iteration is greater than a predefined value (Fig 3, Box 6), it is assumed that the desired conversion has not been reached and we proceed to Step 7. Otherwise the iteration is stopped by going straight to Step 8.

Step 7: One hundred new parameter sets are produced (Fig 2, Box 7) based on the mother parameter set calculated in Step 5. To do this, each parameter is selected randomly from a specific range around the mother parameter (defined in Step 5).

Step 8: When the desired conversion is reached, the final flow data series is reconstructed based on the last parameter set defined in Step 5.

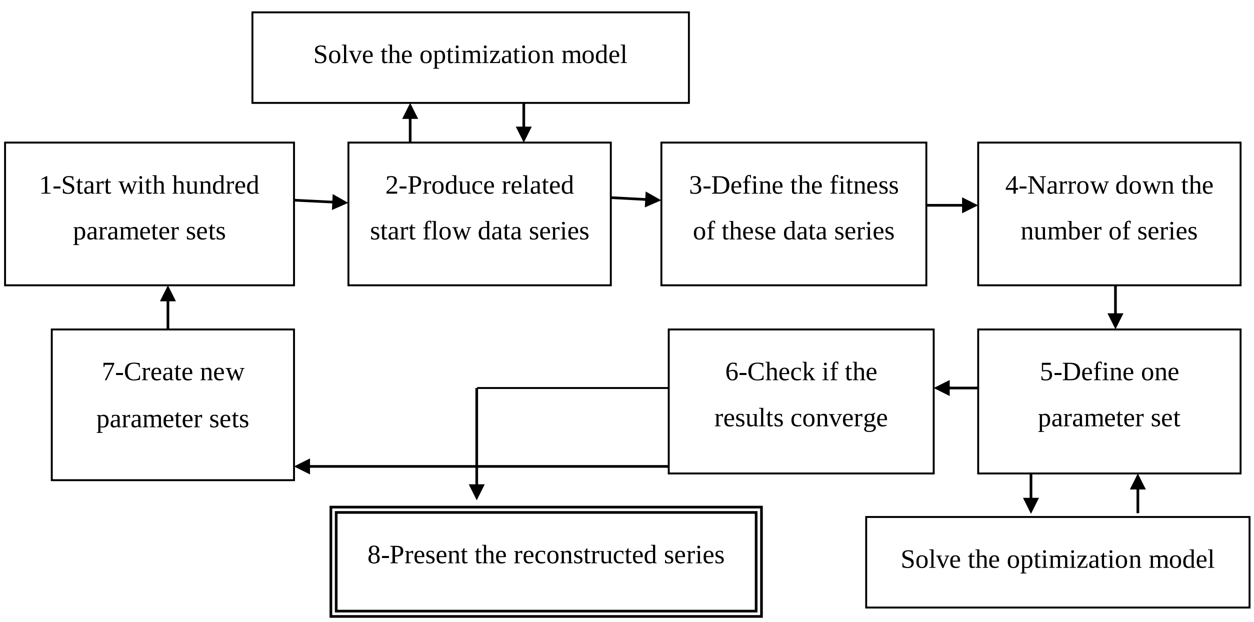


Figure 3: Schematic of the Stochastic based WBE Method

Quality evaluation of reconstructed NF

Five QIs are used to evaluate the degree of excellence of reconstructed flow with the proposed methods. These five criteria are:

6) NN (Equation 6)
7) CC (Equation 7)
8) NAVE (Equation 12)
9) $$SFR = \begin{Bmatrix} 2 - \dfrac{SF_{cal}}{SF_{WBE}}, if\ \dfrac{SF_{cal}}{SF_{WBE}} \leq 1 \\ \dfrac{SF_{cal}}{SF_{WBE}}, if\ \dfrac{SF_{cal}}{SF_{WBE}} > 1 \end{Bmatrix} \tag{15}$$
$$SF = \frac{\sum flow_i}{A}$$
10) NT (Equation 8)

where SFR is specific flow ratio, SF_{WBE} and SF_{cal} are specific flow calculated using WBE and reconstructed NF respectively, and A is the surface of the basin.

These criteria have been developed to evaluate the quality of the reconstructed flow to meet the needs of different users. These needs include: calibration of deterministic hydrologic model for flow forecasting or probable maximum flood estimation (PMF); and local and regional flood analysis for dam safety. NN is calculated based on simulated flow by the hydrologic model HSAMI and considers the effects of meteorological factors on flow. This means that those Q_{cal} series that best respect the coherence between climate and flow have a higher NN index and are more reliable for flow predictions and PMF estimation. NAVE penalizes any over- or under-estimation. SFR penalizes NF series that are not successful in closing the water balance budget, and NT penalizes noisy data series. Also, CC compares the similarity of the reconstructed flow trend and the neighbouring basin's flow trend and penalizes reconstructed flow that does not respect regional integrity. This QI can be useful for those concerned with flood safety in the area. All of these indices vary between 0 and 1 and values closer to one indicate better quality.

Confidence in quality can be increased by checking whether the reconstructed NF is stationary and homogeneous. The KPSS stationarity test is applied here. This test is a null hypothesis that a data series is stationary around a deterministic trend (Kwiatkowski *et al.* 1992). Also, a regional analysis is done to assess the coherency of reconstructed series quantiles obtained from a local frequency analysis with the quantiles of neighbouring basins. To do this, the reconstructed flow in the case study is compared to 19 neighbouring gauged basins. First, the annual peak flows of each basin are sorted in descending order and then the probability of each flow is defined using the Weibul formula (Equation 16). A lognormal distribution with three parameters is then fitted to peak flow (*LN3*) of the NF series.

$$f = \frac{m}{n+1} \tag{16}$$

where f is the probability, m is the rank of the value and n is the total number of observations.

A power regression is then developed between the basin's area and the values of *LN3* of a 2-year return period as follows:

$$FQ = aA^b \tag{17}$$

where *FQ* is fitted line to values of *LN3* of a 2-year return period, a and b are parameters, *A* is the basin's area. The residuals from this equation are plotted in Q-Q plot, scale location plot, and residual versus fitted plot to evaluate the coherence of reconstructed peak flows with regional ones.

Conclusion

NF reconstructing is done in two parts, pre-reservoir and post-reservoir time period, using different methods. A Kalman-based method is developed to reconstruct the flow for pre-reservoir years and is compared with regression based methods: Move III, area ratio, and Multivariable regression.

For the post-reservoir years, however, the deterministic and stochastic based WBE are developed to improve the results of classic WBE. The quality of reconstructed flow before and after reservoir construction is measured using different quality indices and regional coherence is evaluated based on 19 gauged basins in the area. Part 2 of this paper is presented as a Case Study paper dealing with the execution of these new methods for three sub-basins of different sizes in Quebec (Canada).

References

Bisson, J.-L. (1995). “Crue maximale probable de la rivière Gatineau”. Service Prévisions, *Internal Technical Report*, Hydro-Quebec.

Cheng, L., Hou, Z.G., Homma, N., Tan, M., and Gupta, M., (2009), “Solving Convex Optimization Problems Using Recurrent Neural” *Proc. International Joint Conference on Neural Networks*, Atlanta, Georgia, USA.

Chu, P.P, (1992). “A Neural Network for Solving Optimization Problems with Linear Equality Constraints”. Neural Networks, JCNN, *Proc. International Joint Conference*, vol.2, 272 – 277.

Cunderlik, J.M., Burn, D.H. (2003), “Non-stationary pooled flood frequency analysis”, *Journal of Hydrology* 276 (1-4): 210–223.

Dastorani, M.T., Moghadamnia, A., and Piri, J., and Rico-Ramirez, M., (2010). “Application of ANN and ANFIS models for reconstructing missing flow data”. 166:1-4, DOI: 10.1007/s10661-009-1012-8,421-434.

Deb, K., (2000). “An efficient constraint handling method for genetic algorithms”. *Computer Methods in Applied Mechanics and Engineering*, 186(2-4): 311-338.

Deb, K., (2002). “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II”. *IEEE Transactions on Evolutionary Computation*, 6(2): 182-197.

Gunter, A., Krol, M., Carlos de Araujo, J, and Bronstert, A. (2004). “Simple water balance modelling of surface reservoir systems in a large data-scarce semiarid region”. *Hydrological Sciences–Journal*, 49(5): 901-918.

Hamon, W.R. (1960). “Estimating potential evapotranspiration”. *Massachusetts Institute of Technology, Department of Civil and Sanitary Engineering*, 150 pp.

Hernandez-Henriquez, A., Mlynowski. T.J., Déry, S.J., (2010). “Reconstructing the natural stream-flow of a regulated river: a case study of La Grande Riviere, Quebec, Canada”. *Canadian Water Resources Journal*, 35(3): 301-316.

Hirsch, R.M. (1982). “A comparison of four streamflow record extension techniques”. *Water Resources Research*, 18(4): 1081-1088.

Hughes, D. A., and Smakhtin, V. (1996). “Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves”. *Hydrological Sciences Journal*, 41(6): 851-871.

Jones, P.D., Lister, D.H., and Kostopoulou, E. (2004). “Reconstructed river flow series from 1860’s to present (2002)”. *Science Report* SC040052/SR, Environment Agency.

Joshi, V.U., and Tambe, D. (2010). “Estimation of infiltration rate, run-off and sediment yield under simulated rainfall experiments in upper Pravara Basin, India: Effect of slope angle and grass-cover”. *J. Earth Syst. Sci*. 119(6): 763–773.

Kalakrishnan, M., Chitta, S., Theodorou, E., Pastor, P., and Schaal, S, (2011). “STOMP: Stochastic Trajectory Optimization for Motion Planning”. *Int. Conf. on Robotics and Automation*, Shanghai, China.

Karamouz, M., Hosseinpour, A., and Nazif, S. (2011-a). “Improvement of Urban Drainage System Performance under Climate Change Impact”. Case Study. *Journal of Hydrologic Eng*., 6(5): 395-412.

Karamouz, M., Noori, N., Moridi, A., Ahmadi, A. (2011-b). "Evaluation of Floodplain Variability Considering Climate Change Impacts.” *Journal of Hydrological Processes*. 25(1): 90-103.

Koutsoyiannis, D. and A. Efstratiadis, (2007). “Uncertainty Assessment of Future Hydroclimatic Predictions: A Comparison of Probabilistic and Scenario-Based Approaches”. *Journal of Hydrometeorology*, 8 (3), 261–281.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y., (1992). “Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root”. *Journal of Econometrics* 54(1992): 159-178.

Lee, D., Schmidt, P., (1996). “On the power of the KPSS test of stationarity against fractionally-integrated alternatives”. *Journal of Econometrics* 73(1996): 285–302.

Matalas, N. C., and Jacobs, B., (1964). "A correlation procedure for augmenting hydrologic data". Statistical Studies in Hydrology, Geological Survey Professional Paper 434-E, Nolan, T.B., Director.

Moog, D.B., and Whiting, P.j., (1999). "Streamflow record extension using power transformations and application to sediment transport", *Water Resource Research*, 35(1): 243-254.

Nagpaul, P.S., (2005), "Time Series Analysis in WinIDAMS", Unesco, New Delhi, India.

Nash, J. E. and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models, Part I - A discussion of principles". *J. of Hydrol.*, 10(3) : 282–290.

Nicol, P. (2010)."Étude de crue pour la sécurité du barrage de Carillon, Calibration du Modèle HSAMI & Montage du modèle de laminage". *Internal Technical Report*, Hydro-Quebec.

Noriega, G., (1992), "Application of Kalman Filtering to Real-Time Processing of Geophysical Data", IEEE Transaction on Geoscience and Remote Sensing, Vol, 30, No. 5.

Oh, C. (2005). "The Contribution of Tourism Development to Economic Growth in Korean Economy". *Tourism Management*, 26(1): 39-44.

Parsuraman, K., Elshorbagy, A., Carey, S., (2007), "Modelling the dynamics of the evapotranspiration process using genetic programming", *Hydrological Sciences–Journal*, 52(3): 563-578.

Perreault, L., Hache, M. and Bobee, B., (1996). "Rapport final du Project C2-Analyse statistique des séries d'apports énergétiques", INRS-Eau, Research Report N° R-474.

Perreault, L. (2011). "Vérification de prévisions hydrologiques probabilistes, Version 1.0". J-1809, IREQ-2010-0000, Diffusion Hydro-Québec.

Schreider, P., and Demuth, S. (1997). "Regionalization of low flows in southwest Germany". *Hydrological Sciences Journal*, 42(6): 845-858.

Shalev-Shwartz , S., and Tewari, A., (2011). "Stochastic Methods for ℓ1-regularized Loss Minimization". *Journal of Machine Learning Research*, 12(2011): 1865-1892.

Shiau, J.T., Lee, H.C., (2005). “Derivation of optimal hedging rules for a water-supply reservoir through compromise programming”. *Water Resource Management*, 19(2005): 111–132.

Sivapragasam, C.,Vasudevan, G., Maran, J., Bose, C., Kaza, S., and Ganesh, N., (2009). “Modeling Evaporation-Seepage Losses for Reservoir Water Balance in Semi-arid Regions”. *Water Resources Management*, 23(5): 853-867.

Telis, P.A., (2001). “Estimation of Infiltration Rates of Saturated Soils at Selected Sites in the Caloosahatchee River Basin, Southwestern Florida”. U.S. GEOLOGICAL SURVEY, Open-File Report 01–65, Prepared in cooperation with the South Florida Water Management District.

Turner G.H., Twieg D.B. (2005). “Study of temporal stationarity and spatial consistency of fMRI noise using independent component analysis”. *Medical Imaging, IEEE Transactions*, 24(6): 712–718.

Vogel, R.M., and Stedinger, J.R. (1985). “Minimum Variance Streamflow Record Augmentation Procedures”. *Water Resources Research*, 21(5): 715-723.

Wen, L., (2009). “Reconstruction natural flow in a regulated system, the Murrumbidgee River, Australia, using time series analysis”. *Journal of Hydrology* 364(3-4): 216–226.

Whitley , D., (1994). “A genetic algorithm tutorial”. *Statistics and Computing*, 4(2): 65-85.

Yeung, C.W., (2004). “Rainfall-runoff and water-balance models for management of the Fena valley reservoir”. *Guam. Scientific Investigations Report*, U.S. Department of the Interior, Prepared in cooperation with the U.S. Department of the Navy.

Zhou, Y., Xu, J., and Jing, Y., (2010), “Comparison of Centralized Multi-Sensor Measurement and State Fusion Methods with Ensemble Kalman Filter for Process Fault diagnosis”, *Proc. of Control and Decision Conference (CCDC)*, 26-28 May 2010, Xuzhou (China), 3302-3307.

APPENDIX 3 – ARTICLE 3: NATURAL FLOW RECONSTRUCTING IN UNGAUGED BASINS USING NEW KALMAN FILTER AND WATER BALANCE BASED METHODS (PART 2:CASE STUDIES, RESULTS, AND DISCUSSION)

By Ana Hosseinpour, Leslie Dolcine, and Musandji Fuamba. Published at Journal of Hydrologic Engineering, ASCE

Abstract

During the pre-reservoir years, when limited data (neighbouring basin flow and simulated flow) are available for ungauged basins, simple methods are relevant to reconstruct natural flow (NF). On the other hand for the post-reservoir years, when hydrometric data (water level in the reservoir, turbine flow, and discharged flow) are available for ungauged basins, the classic Water Balance Equation (WBE) method is usually employed to determine NF. However, applying classic methods still produce noisy and unreliable flow data series.

In an attempt to produce more reliable NF data, new methods are proposed and described in Part 1 of this paper: the State Fusion Kalman Filter method (Pre-Reservoir Construction Period) and the Deterministic WBE Method and the Stochastic WBE Method (Post-Reservoir Construction Period). This paper which is presented as Part 2 deals with the execution of these new methods for three sub-basins of different sizes in Quebec (Canada). Obtained results are then compared against those produced by the use of the current methods of NF reconstructing: Area Ratio Method, Maintenance of variance (Move) type III Method, and Multivariable Regression Method (Pre-Reservoir Construction Period), and the classic WBE Model (Post-Reservoir Construction Period). This comparative analysis shows the capability of proposed methods for improving the constructed data series with improved results that do not contain any negative flow. They are less noisy, perfectly matched with regional flows, and are reliable enough for frequency analyses. Since these NF values include evaporation, infiltration, and other losing terms of a closed hydraulic system, with more studies needed in order to define the contribution of each of them.

Introduction

This is Part 2 of the paper on Natural Flow Reconstructing Using Kalman Filter and Water Balance Based Methods. It presents three case studies, results and discussion in accordance to the application of proposed methods largely described in Part 1.

Methodology

To begin, few classic methods like Multiple Regression (Hughes and Smakhtin 1996, Schreider and Demuth. 1997, Jones *et al.* 2004), Area Ratio, and Move Type III (Vogel and Stedinger 1985) are introduced in this work to reconstruct the NF for the Pre-Reservoir construction period. The classic WBE method is utilized to reconstruct the NF for the post-reservoir period. Secondly, the proposed new Kalman Filter method is employed in the reconstructing of the NF for the Pre-Reservoir construction period as well as the proposed stochastic WBE method and deterministic WBE method which are also used for the Post-Reservoir construction. Details of these new methods are presented in Part 1 (Hosseinpour *et al.* 2013).

The used case studies are selected in the Outardes basin that is located in the Province of Quebec. The main objective of this application is to determine the capability of proposed methods to improve the quality of reconstructed and filtered NF series. This quality would then be evaluated using different Quality Indexes (Normalized Nash, Consistency Coefficient, Normalized Absolute Volume Error, Specific Flow Ratio, and Normalized Tortuosity). Finally, the Kwiatkowski-Phillip-Schmidt-Shin (KPSS) test was chosen to evaluate the stationarity of reconstructed NF values.

Case Studies Description

The Province of Quebec is one of the main sources for available surface water in Canada found in numerous lakes, reservoirs and rivers. The method for reconstructing NF developed in this paper is meant to be suitable to all basins in the province. However, the Outardes basin (Fig. 1), with its three reservoirs has been selected as the subject of this case-study. The reason behind this selection is that in comparing with the other basins of area, i) this basin has a simple structure by having its reservoirs arranged in series; ii) the reservoirs possess less complicated hydraulic systems from the numbering and types of the gates and turbines perspective; and iii) the basin includes reservoirs of different sizes and characteristics.

As shown in Figure 1, Outardes envelopes three sub-basins designated as Outardes 4, Outardes 3, and Outardes 2. Outardes 3 is a small basin located downstream of Outardes 4, which is a big watershed with a large reservoir, and Outardes 2 is situated downstream from Outardes 3. Among the three sub-basins of the Outardes watershed, Outardes 3 is the most challenging reservoir because it is the smallest, and is highly affected by any changes in released flow from the large reservoir at Outardes 4. Any increase or decrease of discharged flow from Outardes 4 would seriously affect the water level in Outardes 3 resulting in fluctuations that make it difficult to estimate NF values in this basin. The data is thus deemed to be very noisy. Characteristics of each of these basins are shown in Table 1. All the hydrologic and hydraulic data relating to these reservoirs, including the simulated flow using HSAMI (Nicol 2010 and Bisson 1995), are taken from the HQ's database.

Table 1: Characteristics of case study basins

Characteristic	Outardes 4	Outardes 3	Outardes 2
Basin area (km^2)	17119	485	1302
Maximum reservoir volume (hm^3)	10940.44	14.72	16.21
Average long-term minimum temperature (°C)	-6.76	-3.71	-3.06
Average long-term maximum temperature (°C)	4.38	7.08	6.84
Average long-term rainfall (cm/day)	0.17	0.18	0.19

Pre-Reservoir Construction Period

The four methods of Multiple Regression, Area Ratio, Move Type Three and Kalman Filter are exercised in a Case Study (Fig 1) to reconstruct and filter the NF series for the time period before reservoir construction in the Outardes 4 basin. This time period includes the years of 1960 to 1968; and the reservoir which was built in 1969. Among the gauged basins of area, Moisie is most similar to Outardes 4 and is named to evaluate the quality of reconstructed flow. Two other gauged rivers — Romaine and Outardes — are used as secondary neighbouring basins for the purposes of the same type of evaluation. Moisie and La Romaine’s flow data series were obtained from CEHQ (2013), while the flow data series for the Outardes River for the years 1960 to 1969 was made available from the Hydro-Quebec database.

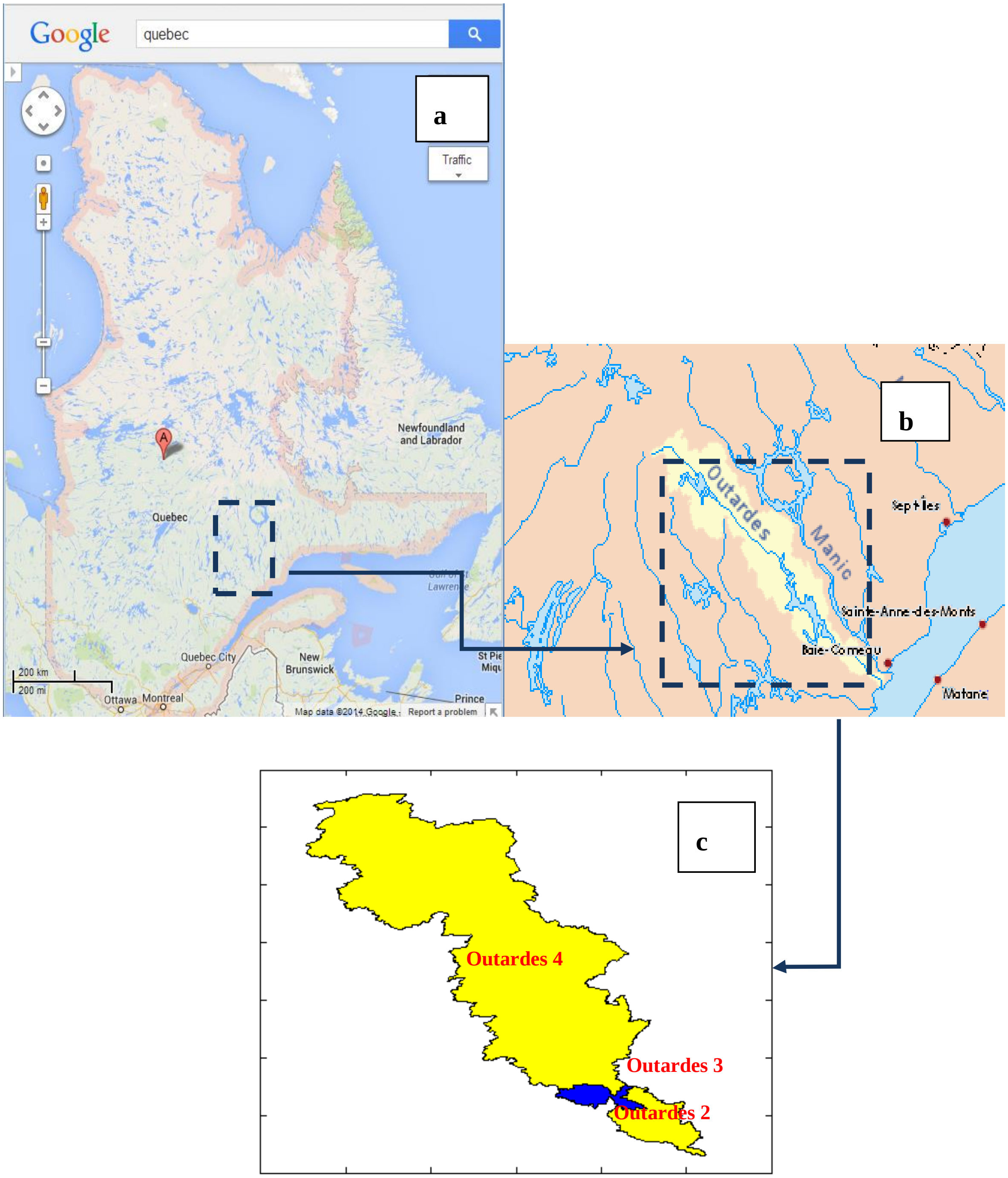


Figure 1: Location of the Outardes Basin in Quebec (a and b), and sub-basins of Outardes 4, Outardes 3 (in darker color), and Outardes 2 (c) (a: https://maps.google.ca, b: www.wikipedia.org)

Post- reservoir period

Both deterministic and stochastic methods are significant to reconstruct and filter the NF series and to be compared with classic WBE and HSAMI for the time period after reservoir construction in Outardes 4, 3, and 2 basins. Among the gauged basins in the area, Moisie is most similar to Outardes 4, while Godbout is most similar to Outardes 3 and Outardes 2. These basins are explored in assessing the quality of reconstructed NF in the Outardes basins. The daily flow data for these stations are taken from CEHQ (2013).

In order to perform regional analyses, flow data from the gauged basins are gathered and compared to data from Outardes 4, Outardes 3 and Outardes 2. The information from these gauged data is presented in Table 2. The flow data series for these rivers are all accessible from CEHQ (2013).

Table 2: Characteristics of gauged basins used for regional analyses

Basin Name	Latitude	Longitude	Start-End	Year Number	Area (km^2/100)
Godbout	49.55222	-68.09611	1975-2007	2	15.7
Moisie	50.59028	-66.31861	1965-2007	41	190
Magpie	51.14889	-64.97389	1979-2007	28	72.3
Romaine	50.52167	-64.04056	1957-2006	49	130
Natashquan	50.72083	-62.18944	1982-2007	25	156
Saint-Paul	52.28556	-58.00306	1968-2007	39	66.3
Petit Saguenay	48.31694	-70.085	1975-1998	23	7.36
Chicoutimi	48.52278	-71.35278	1950-2000	50	33.9
Aux Ecorces	48.30722	-72.08056	1972-2000	28	11.1
Pikauba	48.57139	-71.63861	1970-2000	30	4.95
Metabetchouane	48.63417	-72.65833	1965-2000	35	22.8
Ouiatchouan	48.34528	-72.41	1967-2000	33	5.62
Petite Peribonca	49.36167	-72.08361	1975-2000	25	10.9
Ashuapmushuan1	49.14917	-72.82139	1963-2000	37	153
Ashuapmushuan2	49.46778	-73.6	1963-2000	37	111
Mistassibi1	49.50944	-72.35472	1961-2002	41	93.2
Mistassini2	49.49056	-72.45583	1963-2000	37	98.7
Manouane	50.47056	-71.55028	1980-2000	20	17.17
Valin	48.82361	-71.6275	1975-2000	25	7.46
Ste-Marguerite-1	48.45333	-70.52167	1977-1997	20	11

Results Presentation

Pre-reservoir period

Four methods of reconstructing the pre-reservoir flow are compared in this paper, including multiple regression, area ratio, Move type three, and Kalman filter. These methods are germane to reconstruct the NF for the Outardes 4 reservoir from 1960 to 1968, when the only available data originated from the neighbouring basin flow and simulated flow by HSAMI. The multiple regression method operates with a linear regression based on the NF as a dependent variable, and logarithmic scaled neighbouring basin flow and simulated flow by HSAMI as the independent variables for the calibration period. The regression is then carried out to reconstruct the flow for the years prior to 1969 validation period (1960-1978). Two thirds of the period for which calculated NF using classic WBE is available (1969-2011) is considered as the calibration period (1979 to 2011) and the rest (1960-1978) is regarded as validation period.

Both the area ratio and Move III methods work in the same calibration and validation period. For the Move III method, HSAMI flow and neighbouring basin (Moisie) flow are considered in turn as independent variables. The results derived from the application of these methods over five years are presented in Figures 2 to 6. Figures 4 to 6 show results of the model following reservoir construction, but provide helpful information on the performance of the different models, as their value as a reference flow can be compared to flow calculated by classic WBE.

The quality of reconstructed pre-reservoir flows is assessed at the Moisie basin and tested using three QIs: normalized Nash (NN), normalized tortuosity (NT), consistency coefficient (CC). To avoid bias in this quality evaluation, CC is calculated by the means of the measured flow in the neighbouring Romaine and Outardes rivers. Assuming that the models that work better after reservoir construction would also perform better before the reservoir has been built; the quality of reconstructed NF post reservoir was reviewed engaging these three QIs plus normalized absolute volume error (NAVE) and specific flow ratio (SFR). The calculated QIs based on different methods of pre-reservoir NF reconstructing are presented in Table 3.

Table 3 shows that each method has particular advantages. For example, Kalman has high CC before 1969, area ratio has shown good consistency after 1969, and multiple regression method has high NAVE post-reservoir construction. However, Move III generally has the best individual

QI, and best average pre- and post-reservoir QI. This can also be recognized in Figures 4 to 6, which compare the models to regular WBE. In these figures, area ratio method overestimates the NF in comparison with WBE; this also happens with the Kalman method in some years (see Figure 4) and causes lower NAVE. This is why these two methods have higher Specific Flow (SF) and lower SFR than other methods. According to Table 3, all the methods have lower SF post reservoir, which could be related to changes of precipitation rate during the time. According to rainfall and snow data records in the area, after 1969 daily average rainfall decreased 0.7 percent and daily average snowfall decreased 17.2 percent.

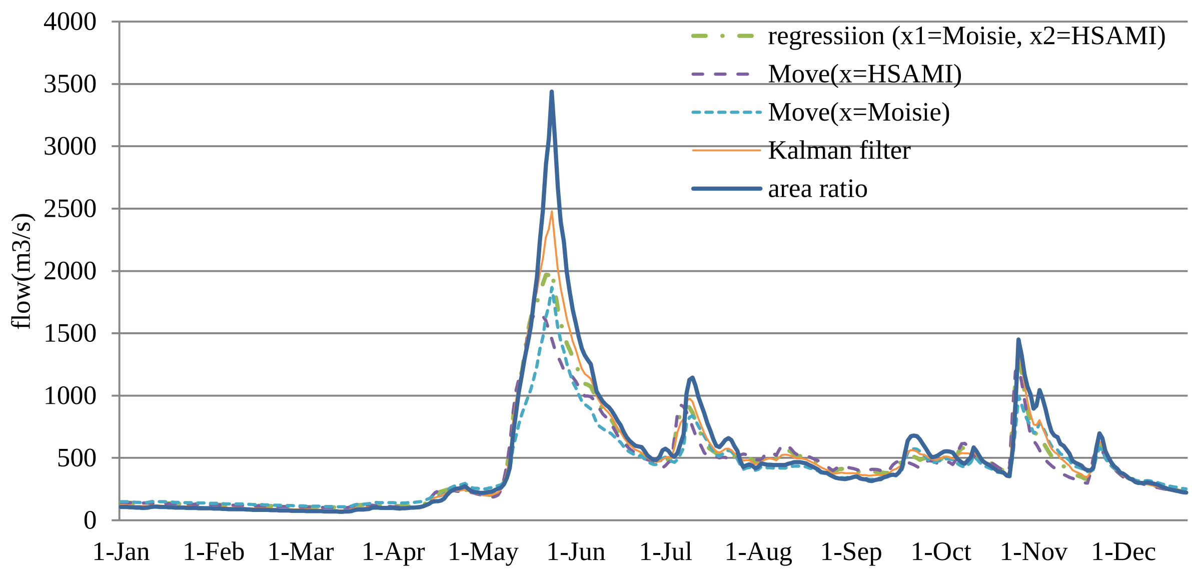


Figure 2: Comparison of different methods of NF reconstructing for Outardes 4 (1966)

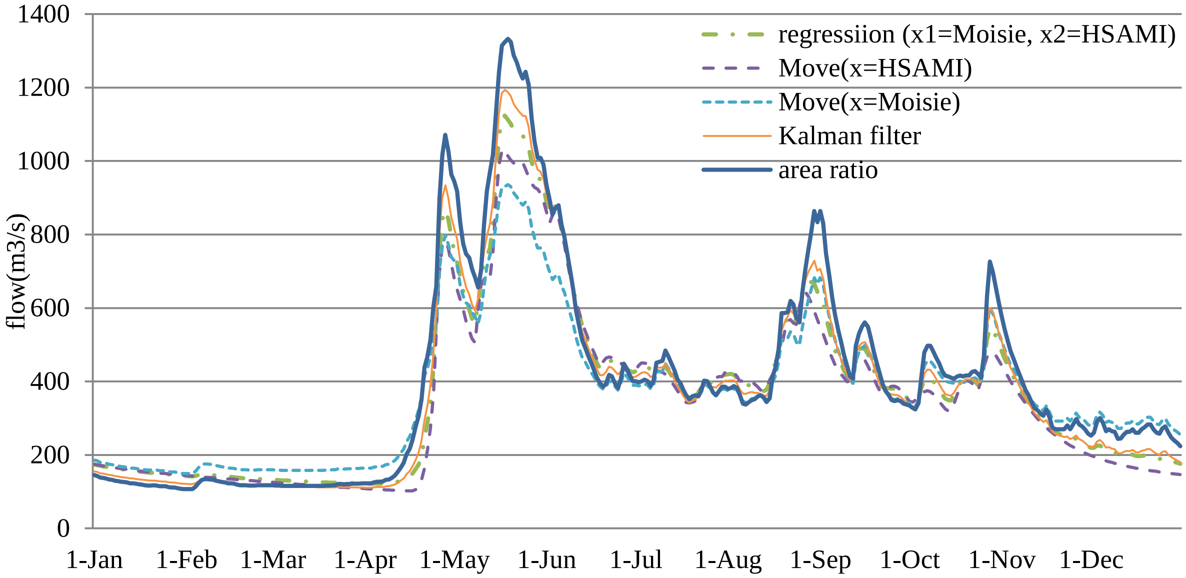


Figure 3: Comparison of different methods of NF reconstructing for Outardes 4 (1968)

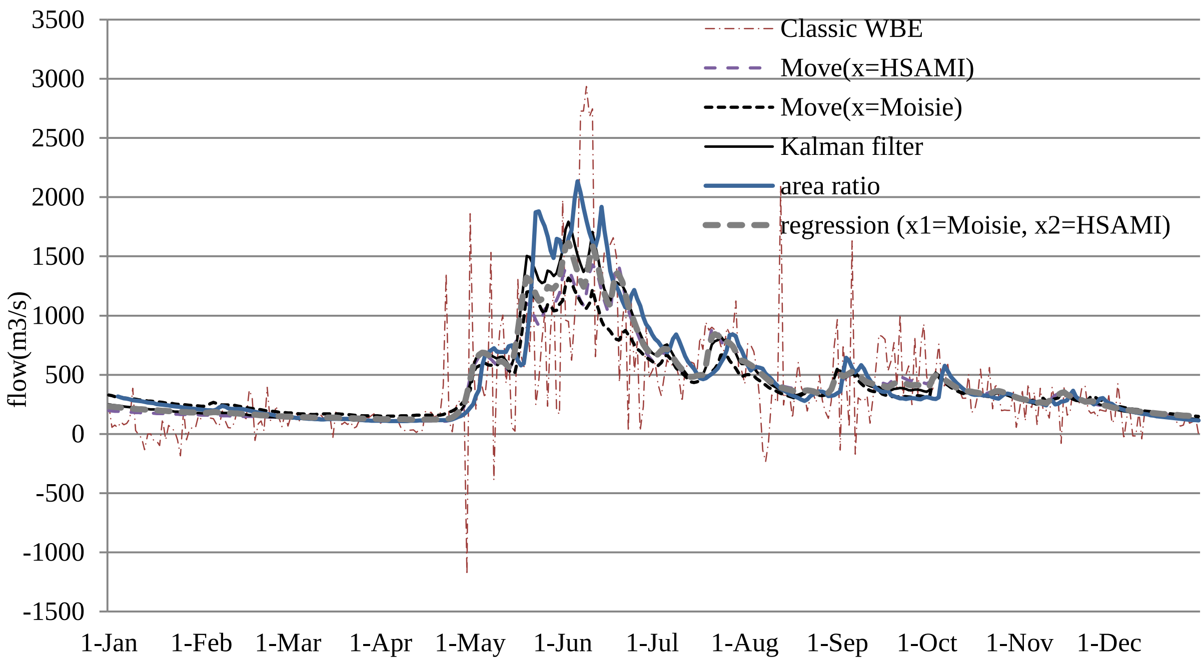


Figure 4: Comparison of different methods of NF reconstructing for Outardes 4 (1970)

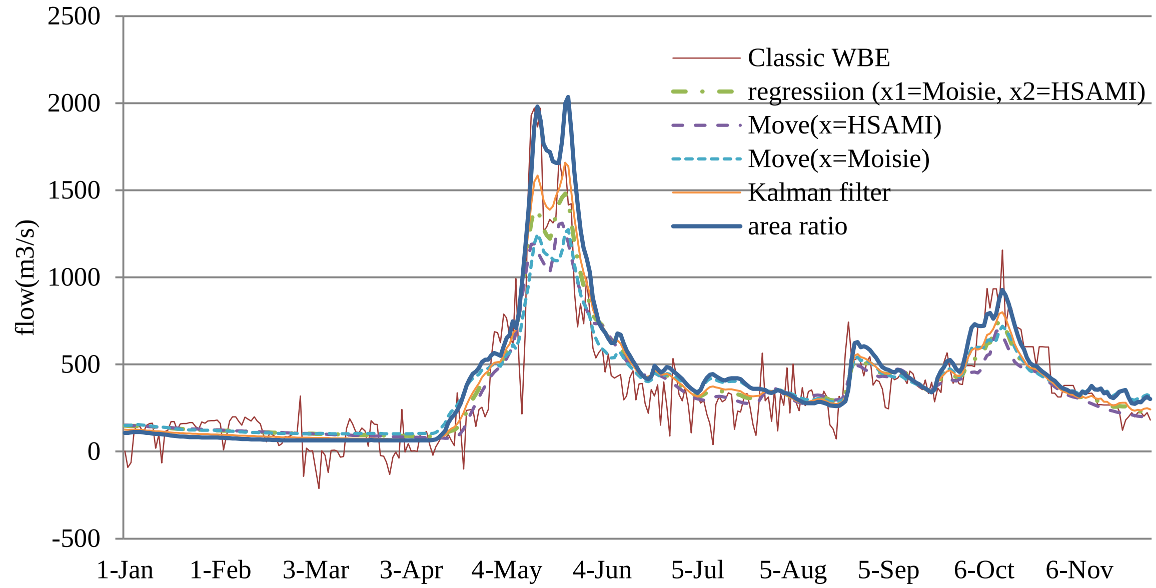


Figure 5: Comparison of different methods of NF reconstructing for Outardes 4 (1971)

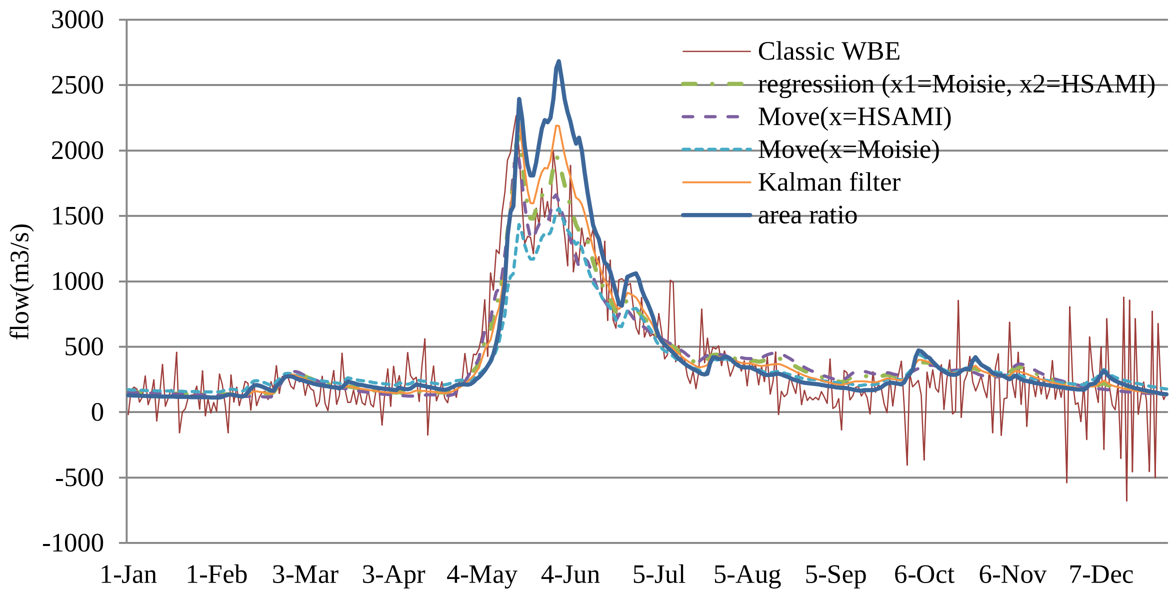


Figure 6: Comparison of different methods of NF reconstructing for Outardes 4 (1981)

Table 3: Comparison of QIs for a few NF models during Pre/Post-Reservoir Construction Periods

NF Methods	Reservoir Name	AVE	NT	SFR	NN	CC	Average QI	SF ($m^3/s/km^2$)
Kalman	Pre-RC -Outardes river		0.850		0.744	0.564	0.719	0.0241
	Pre-RC -Romaine river		0.850		0.744	**0.677**	0.757	
	Post-RC -Romaine river	0.699	0.864	0.941	0.802	0.741	0.809	0.023
Area Ratio	Pre-RC -Outardes river		0.808		0.588	0.529	0.642	0.0256
	Pre-RC -Romaine river		0.808		0.588	0.630	0.675	
	Post-RC -Romaine river	0.677	0.82	0.87	0.657	**0.997**	0.804	0.0242
Regression	Pre-RC -Outardes river		0.868		0.798	0.573	0.746	0.0238
	Pre-RC -Romaine river		0.867		0.650	0.592	0.703	
	Post-RC -Romaine river	**0.702**	0.876	0.925	0.850	0.791	0.829	0.0230
Move III x=HSAMI	Pre-RC -Outardes river		**0.876**		**0.842**	**0.581**	**0.766**	0.0226
	Pre-RC -Romaine river		**0.876**		**0.842**	0.622	**0.780**	
	Post-RC -Romaine river	0.7	**0.886**	0.97	**0.885**	0.704	0.829	0.0219
Move III x=Moisie	Pre-RC -Outardes river		0.875		0.803	0.529	0.736	0.0228
	Pre-RC -Romaine river		0.875		0.803	0.630	0.769	
	Post-RC -Romaine river	0.684	0.883	**0.986**	0.859	**0.997**	**0.882**	0.0214

Note: (1) Pre-RC =Pre-Reservoir Construction = 1960 to1969

(2) Post-RC = Post-Reservoir Construction=1969 to 1978

(3) SF = Specific Flow

Post-reservoir construction period

Deterministic WBE model

As was expounded upon in the first part of this paper (Part 1, Figure 2), each year was split into no more than 8 segments. Then, the *d* coefficient was collected from {150,100,80,50,20,15,10,7,5,1,0} because a *d* coefficient amounting to more than 150 did not change the results. Thus, for each segment and each *d* value, one parameter set was selected by genetic algorithm (GA) and NF values series were determined based on this parameter set.

Therefore, at the end, 11 NF values series were available for each segment, each of them based on a different weight of Q_{sim}. Note that Q_{sim} itself could not be considered as a NF series because it showed deficiencies, especially during high flow. Weighted values were observed during different time steps of the year to take advantage of periods that displayed good quality.

Defining the parameters of the model using GA for *d* coefficients in each segment required a few steps. Step 1 involved initializing the parameters of the model, which was done for all GAs (n=10) using random values in a predefined range. The ranges of these parameters ($1<dn<21$, $0<\gamma<15$, $0<c<10000$, $0<p<1000$, $0<q<1000$) were determined empirically.

In Step 2, the optimization model was solved using the initial parameters and Steps 3 to 6 were repeated to improve and finalize these initial parameters.

Figure 7 provides a sample of results from the deterministic optimization model solved with different *d* coefficient for Outardes 4-2008. As illustrated, there is more than one graph for each year and the decision manager can single out the most appropriate one according to different quality indices (QIs) and visual comparison with regular WBE. Generally, the few graphs which had considerably higher-quality indices (CC, NN, NT, NAVE, and SFR) were elected. The graph that was smoother and more reliable was then chosen based on visual comparison (this is the graph which is named deterministic in Figures 8 to 16). This graph should follow the general trend of reconstructed flow by WBE. It is important to keep in mind that the smoother graph is not necessarily the better flow series. For example, some smooth graphs underestimate or overestimate the flow, while noise is completely removed, they should not be positioned as the final graph. An example of smooth but underestimated flow is highlighted with red circles in Figure 7.

Stochastic WBE model

Since the deterministic model does not favour the probability of different parameter sets and defines the parameters based on only one quality index (Nash), a stochastic model was also endorsed to solve these problems and determine more reliable parameters for the optimization model. This stochastic method defined the best parameter set founded on the probability of different sets and the soundness of each parameter set based on three quality indices (NN, NAVE, and CC). Note that CC was calculated using Moisie as the neighbouring basin for Outardes 4 and Godbout as the neighbouring basin for Outardes 3 and Outardes 2.

This stochastic model is capable of producing one hundred random parameter sets in such a way that *dn* changes from 1 to 10 and each *dn* values is repeated exactly 10 times, *C* changes from 0 to 10000, and γ changes from 0 to 10. These parameter sets were then made to solve the optimization model (Part 1, Equation 10) and were ranked according to their related QIs. Since the model is multi-objective, more than one quality index is stated to appraise the soundness of the series. The weighted summation of NN, NAVE, and CC are the criteria used in ranking the parameter sets. The weights of these indices were defined through trial and error. For our Case Study $w1=w2=1$ (Part 1, Equation 11) and $w3$ changes according to the season. It is usually 1 or 0.5 during low flow and 10 or more during the rest of the year. These weights are almost the same for each year at one basin, which makes the job easier. Ten parameter sets with higher QIs are then drawn to define the probable mother parameter set of the next iteration. In each iteration, one hundred new parameter sets were formed on top of the mother parameter set of the previous iteration. To do this, one hundred *C*, γ, and *dn* parameters were generated in the ranges of $C_n \pm 100$, $\gamma_n \pm 2$ and $dn_n \pm 2$ respectively (C_n, γ_n, and dn_n belong to the mother parameter set). These ranges are big enough to avoid being trapped in a local optimum point and maintain parameter variety, yet small enough to merge rapidly. The parameters were finalized when the QI ("benefit" calculated using Part 1, Equation 11) of two consequent iterations was less than 0.001. This certainty would be regarded sufficient variable in this case.

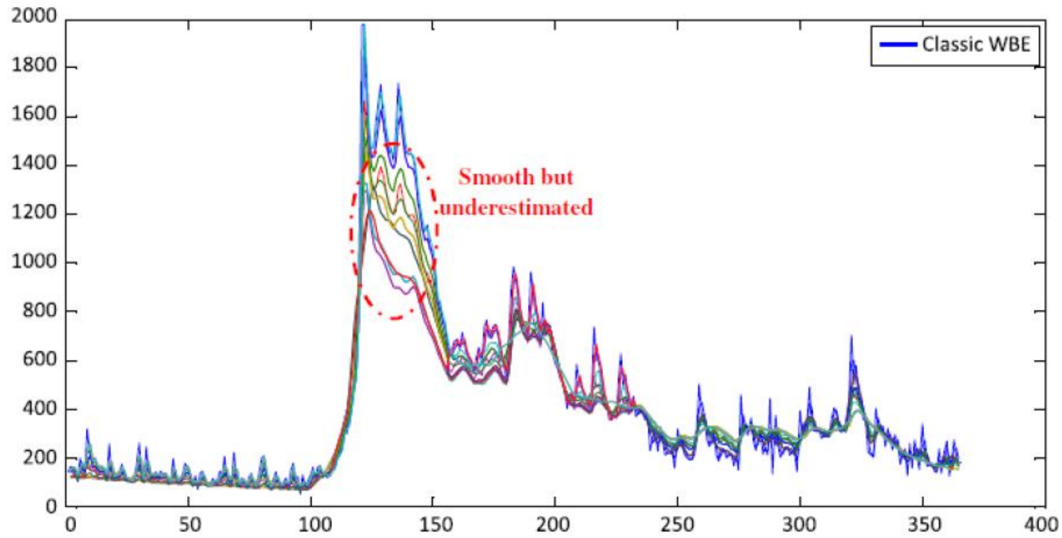


Figure 7: Comparison of daily NF (m3/s) by classic WBE and deterministic based optimization model solved with different d coefficient for Outardes 4-2008

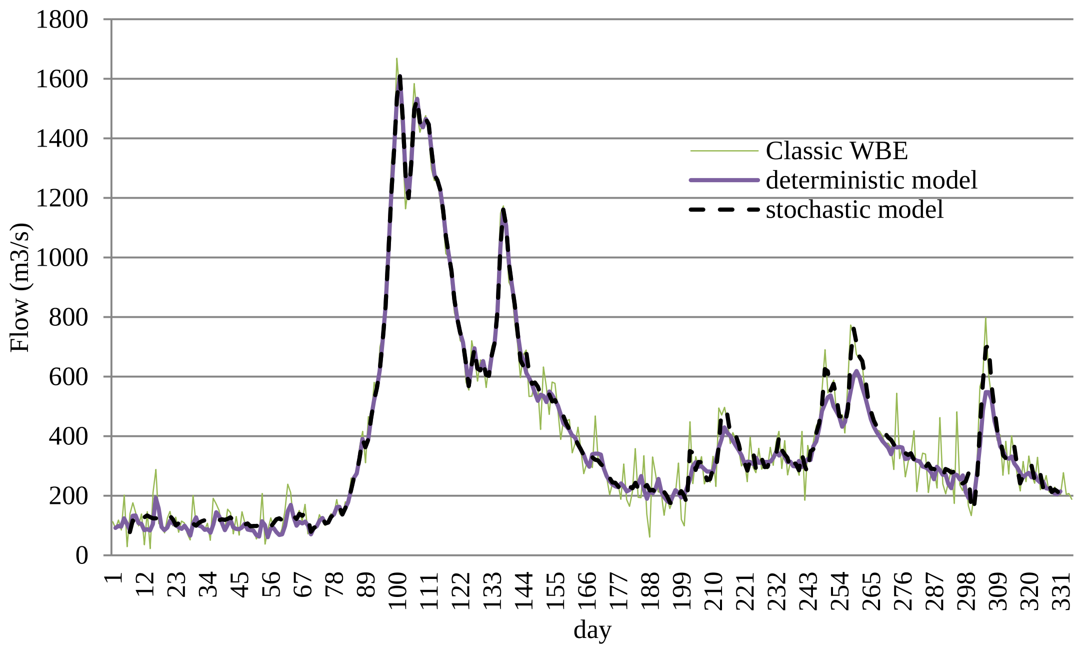


Figure 8: comparison of deterministic WBE, stochastic WBE, and classic WBE (Outardes 4-2005)

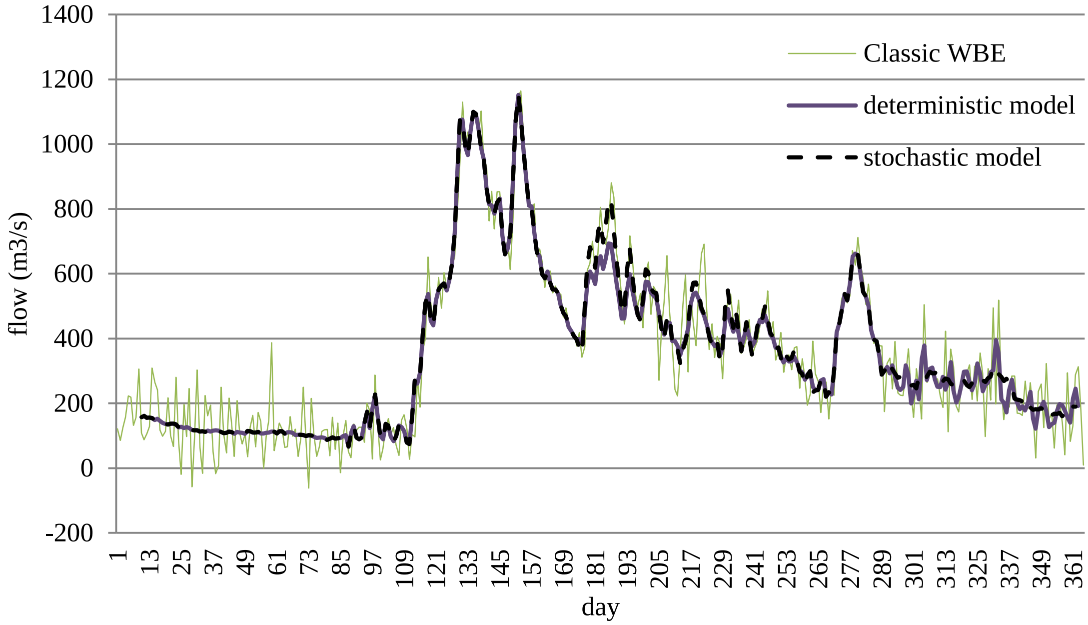


Figure 9: comparison of deterministic WBE, stochastic WBE, and classic WBE (Outardes 4-2009)

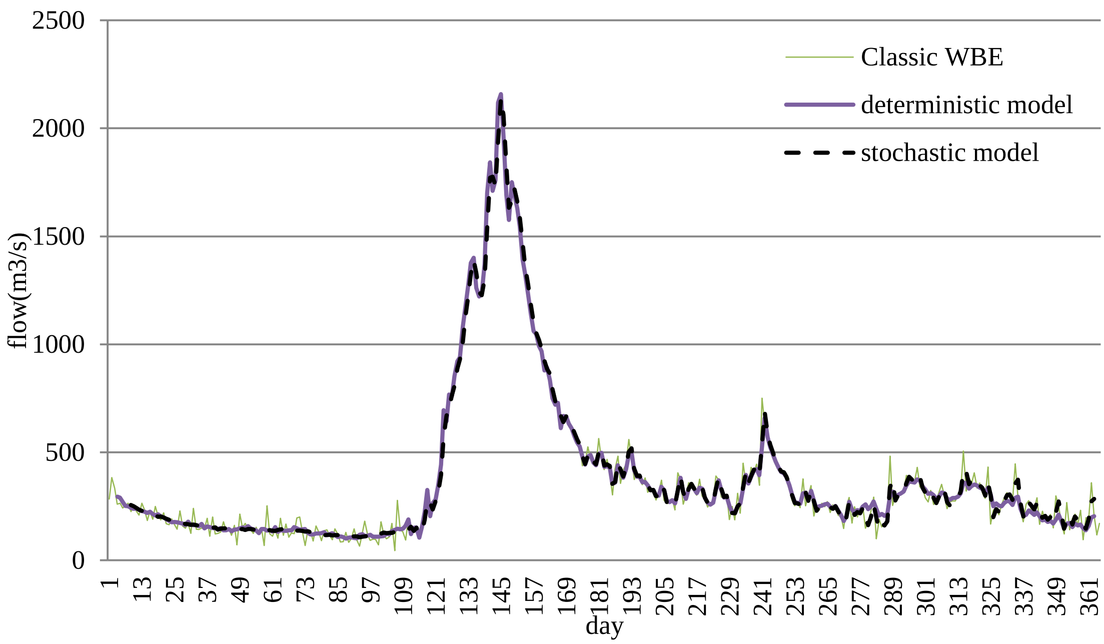


Figure 10: Comparison of deterministic WBE, stochastic WBE, and classic WBE (Outardes 4-2011)

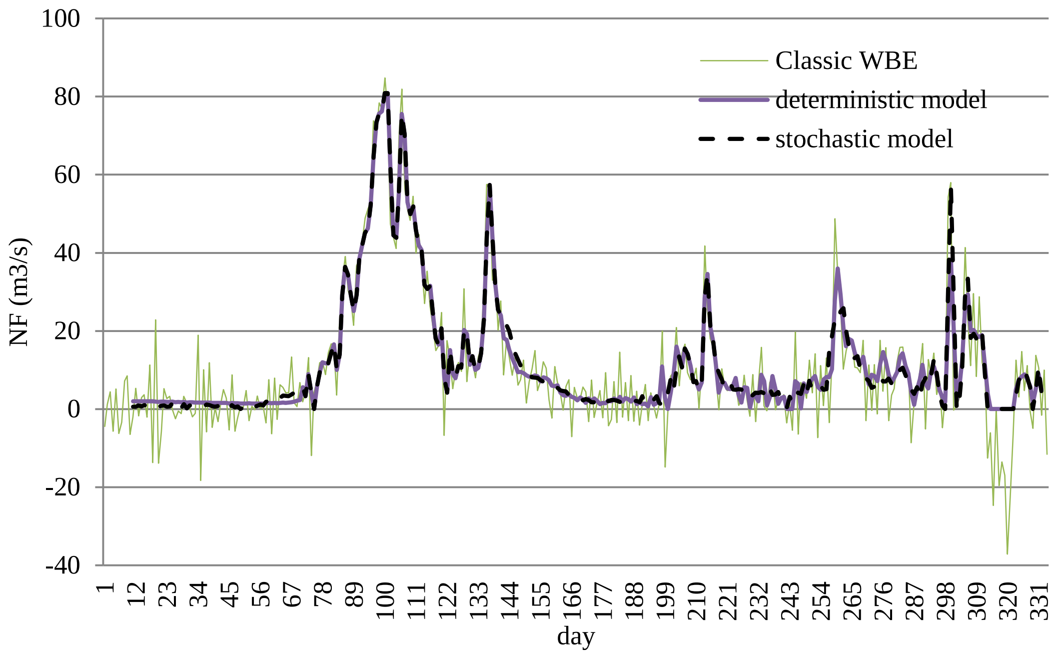


Figure 11: Comparison of deterministic WBE, stochastic WBE, and classic WBE (Outardes 3-2005)

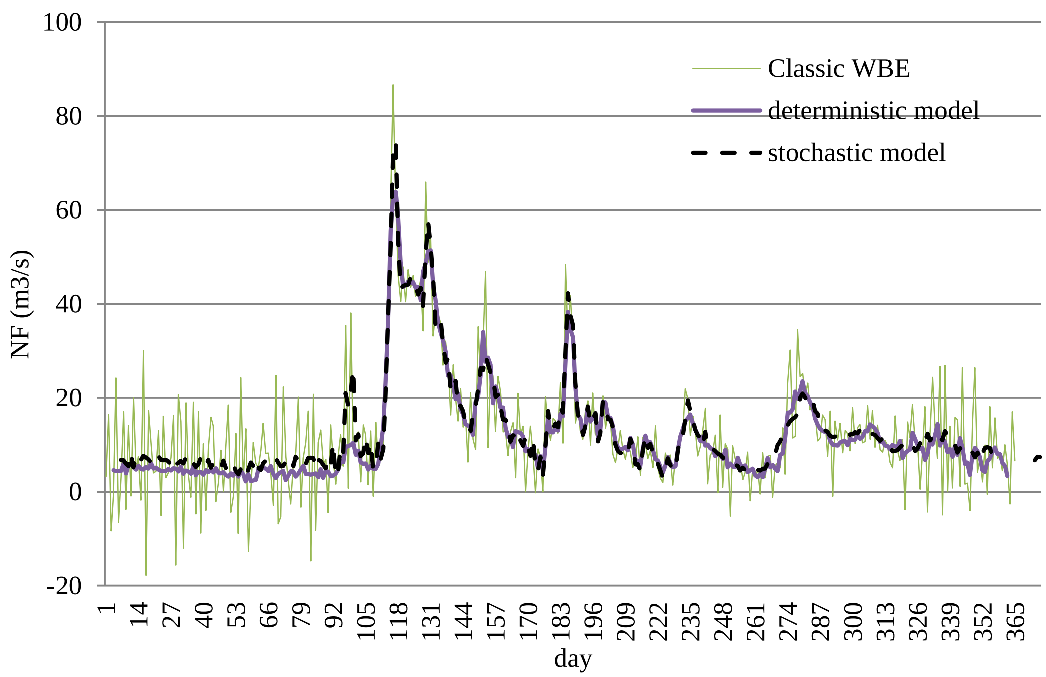


Figure 12: Comparison of deterministic WBE, stochastic WBE, and classic WBE (Outardes 3-2009)

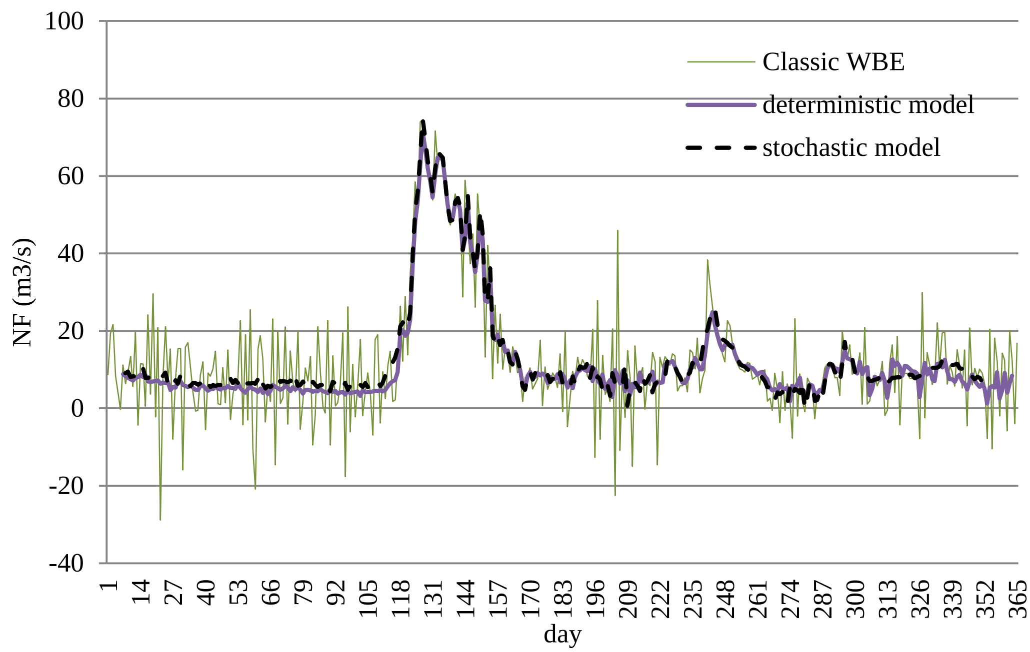


Figure 13: Comparison of deterministic WBE, stochastic WBE, and classic WBE (Outardes 3-2011)

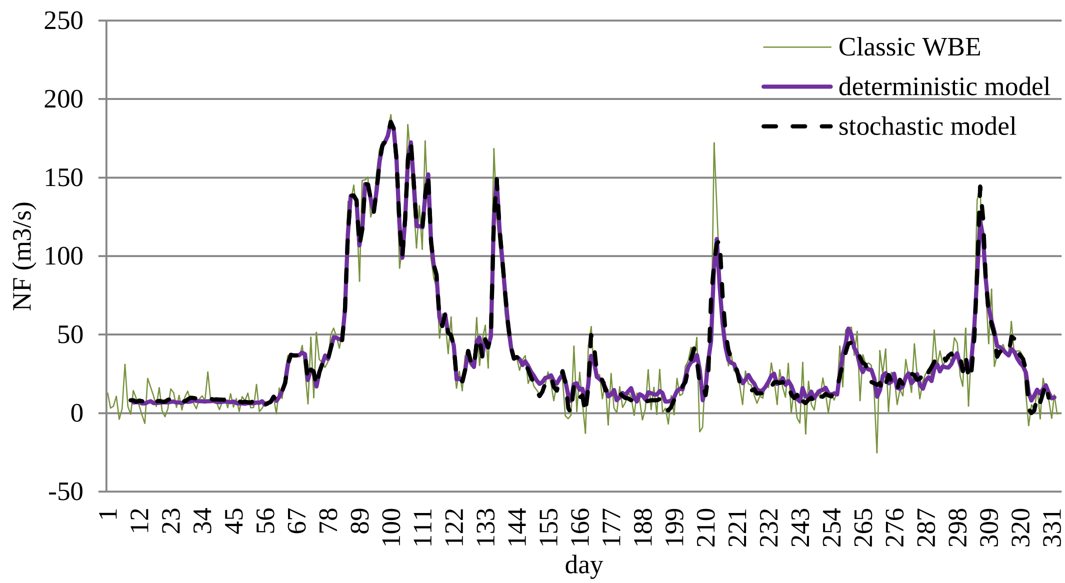


Figure 14: Comparison of deterministic WBE, stochastic WBE, and classic WBE (Outardes 2-2005)

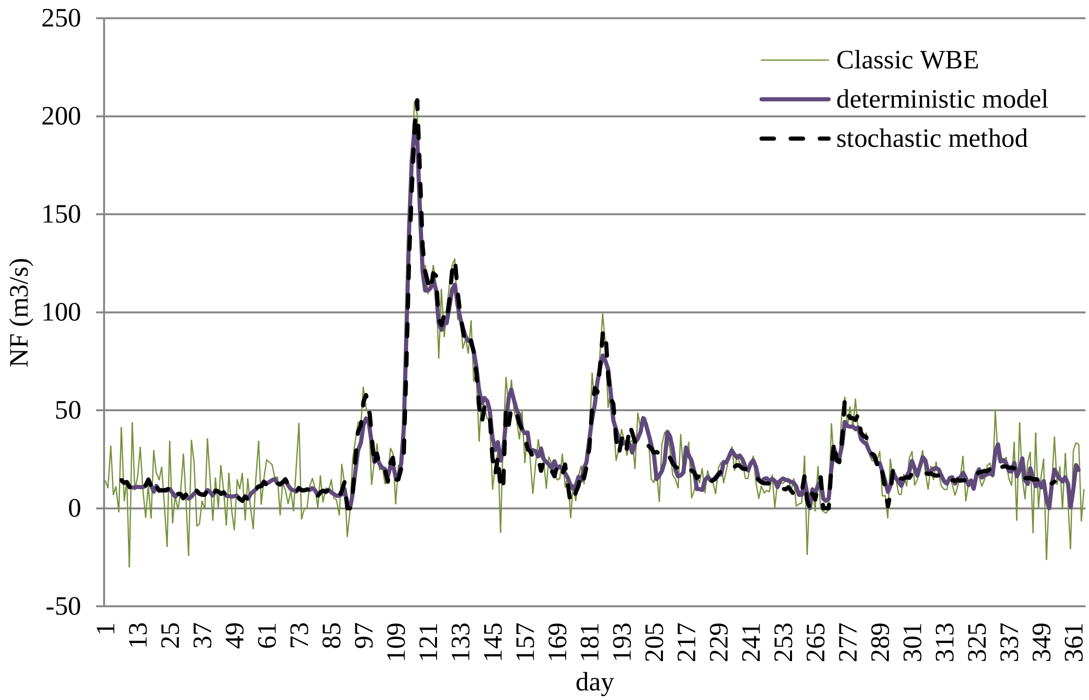


Figure 15: Comparison of deterministic WBE, stochastic WBE, and classic WBE (Outardes 2-2009)

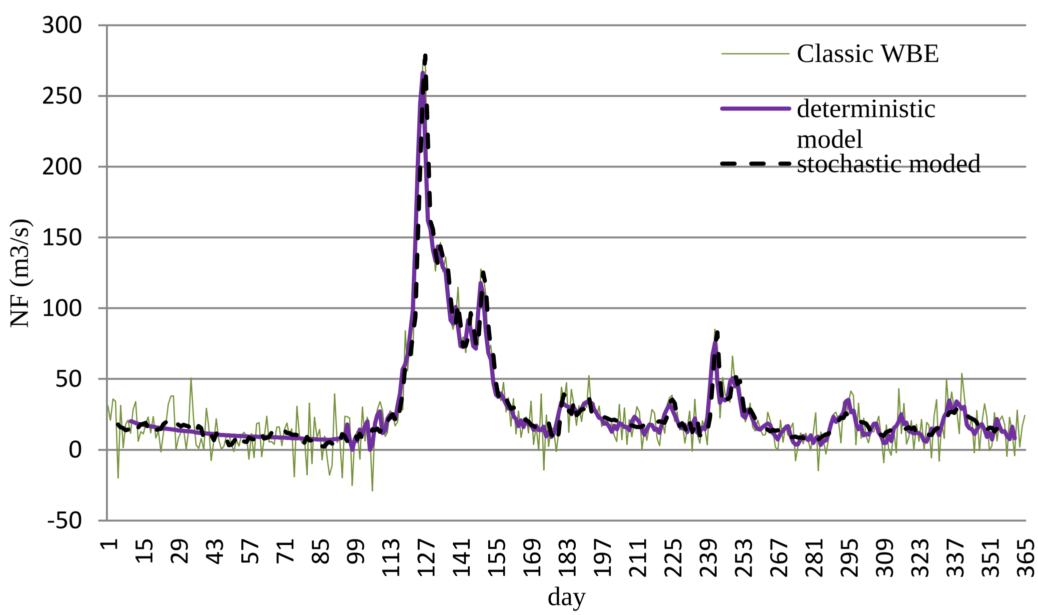


Figure 16: Comparison of deterministic WBE, stochastic WBE, and classic WBE (Outardes 2-2011)

Results Discussion

The reconstructed NF series for Outardes 4, Outardes 3, and Outardes 2 for certain recent years are presented in Figures 8 to 16. The calculated flow using classic WBE is compared to reconstructed flow using developed deterministic and stochastic-based optimization models. As shown in these figures, the methodology (both stochastic and deterministic method) efficiently improved the NF estimation when measured against the classic WBE. No negative value is seen for NF reconstructed using developed methods, and they are much smoother than classic WBE.

Five quality indexes — NN, CC, NAVE, SFR, and NT — were also intended to grade the excellence of reconstructed NF. The QIs on three methods of flow reconstructing (classic WBE, deterministic WBE, and stochastic WBE) for the past six years in the three sub-basins of Outardes are summarized in Table 4. WBE has consistently lower QIs than the other two. This lower quality is also clear in Figures 8 to 16, where reconstructed flows using deterministic and stochastic models are much smoother and more realistic (i.e. they include no negative flow) than classic WBE.

On average, the deterministic and stochastic methods improved the NN, CC, and NT with the same rate for Outardes 4 and Outardes 2. In Outardes 3, however, these QIs are respectively improved 61, 17, and 162 percent under the deterministic method and 52, 17, and 184 percent under the stochastic method. Essentially, this means that the stochastic model produces smoother NF series and the deterministic model better compliments meteorological conditions, which is not surprising given that Q_{sim} is entered as the external signal in the deterministic model (Part 1, Equation 6).

Comparing the quality of two deterministic and stochastic methods shows that in Outardes 3 the average NAVE is 0.84 and 0.66, respectively, supported in these methods. The two models also have almost the same average NAVE in Outardes 4 and Outardes 2. The average SFR is always better when using the stochastic method, reflecting its success in closing the water balance budget.

It can also be concluded from Table 4 that both the deterministic and stochastic methods have lower QIs in Outardes 3 than in Outardes 4 and Outardes 2, which may be related to the small dimensions of the reservoir and its location downstream from a big reservoir. This reservoir is significantly affected by any little change in released flow from Outardes 4. However, the greatest improvement resulted from bringing to bear the developed model in Outardes 3. As seen in the

QIs listed in Table 4, the Improvement Ratio (IR) is higher in Outardes 3 than in the two other sub-basins. The improvement Ration is defined by Equation 1 as follow:

$$IR = \left(\frac{\text{QI Deterministic model or Stochastic model - QI WBE}}{\text{QI Deterministic model or Stochastic model}} \times 100\right) \qquad (1)$$

Generally, the difference between stochastic and deterministic method is more substantial during the years when the NF data series is noisier (for example Outardes 3-2009) which mostly coresponds to the years before 2000.

In order to accurately evaluate the quality of reconstructed NF, compared to regional flow, the histogram of local reconstructed NF is first compared to regional histograms as shown in Figure 17. The latter shows the scaled reconstructed NF (using deterministic WBE) of the three basins of the Case Study compared with the gauged basins of the area for the year 2007. Results show that local flows seem totally in line with regional flows: no considerable under- or over-estimation is seen in the graphs. The local NF also appeared stationary in KPSS test, which means that reconstructed flow has the same characteristics during this time. This stationary data in reconstructed flow is valuable for frequency analyses.

A regression is then developed between the basin surface and quantiles (0.5, 0.80, 0.9, 0.95, 0.98, 0.99) of annual peak flows of basins after fitting a log normal distribution with three parameters as (see Part 1, Equation 17):

$FQ = 20.04A^{0.871}$, $R^2 = 0.93$

Lastly, scale location, Normal Q-Q plot, and residual versus fitted plots are drawn (Figures 18 to 20) to complete the regional analyses. In Figures 18 to 20, the obtained values from local frequency analysis of the three basins (Outardes 4, Outardes 3, and Outardes 2) are exhibited with red points. From these figures, it can be interpreted that the local basins have peak flows totally comparable to those of neighbouring basins. The reconstructed flows are reliable for local and regional frequency analyses because:

- The three basins (red points) do not disturb the linear relationship in normal Q-Q plot (Figure 18).
- They maintain the scattering in scale-location plot (Figure 19) and do not cause any pattern in this plot.
- They maintain the random scattering of residuals around zero and the constancy of residuals. This means they do not cause residual increases or decreases in the fitted values in a pattern.

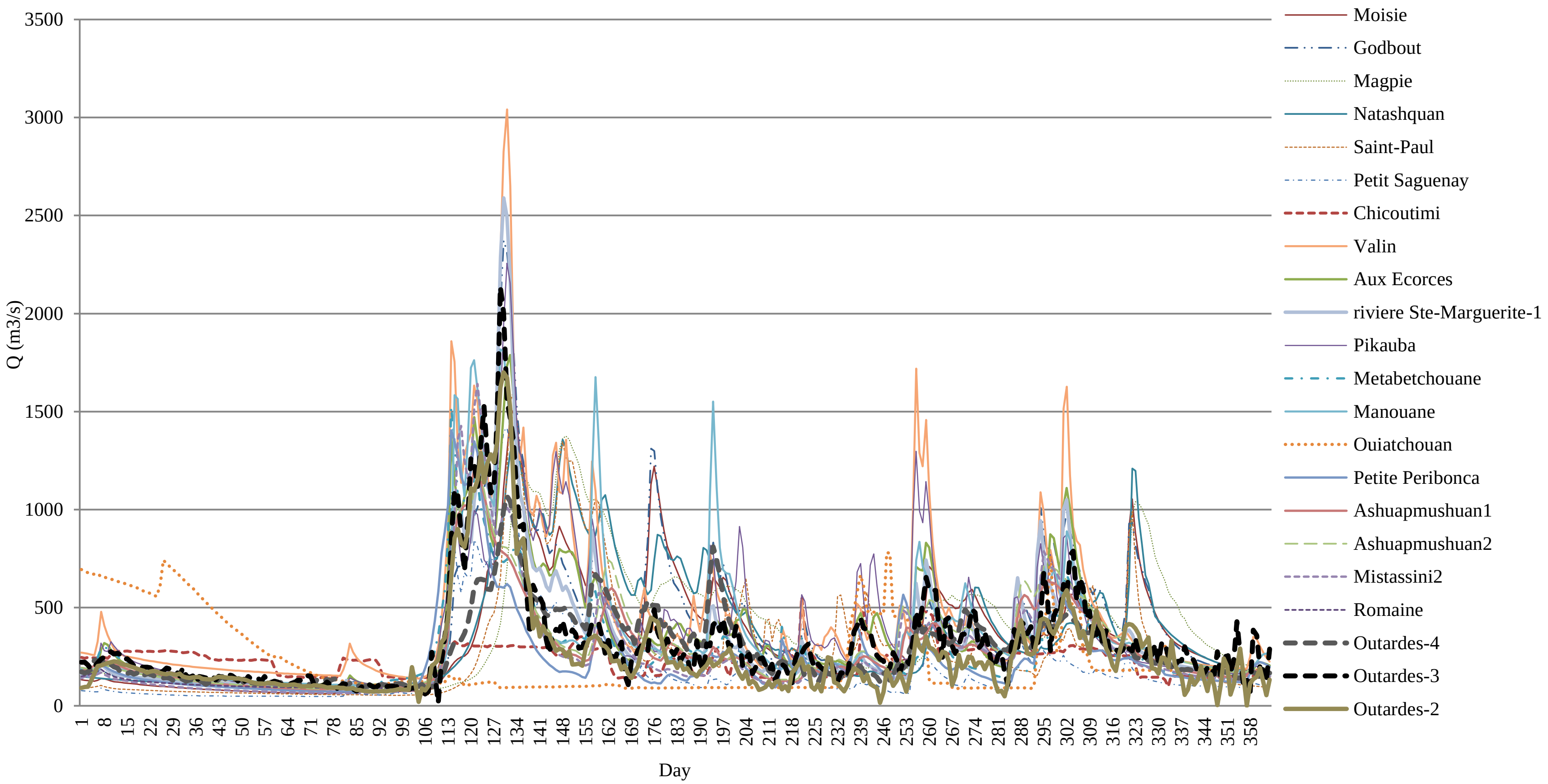


Figure 17: Comparison of hydrograph of local scaled reconstructed flow with regional flow

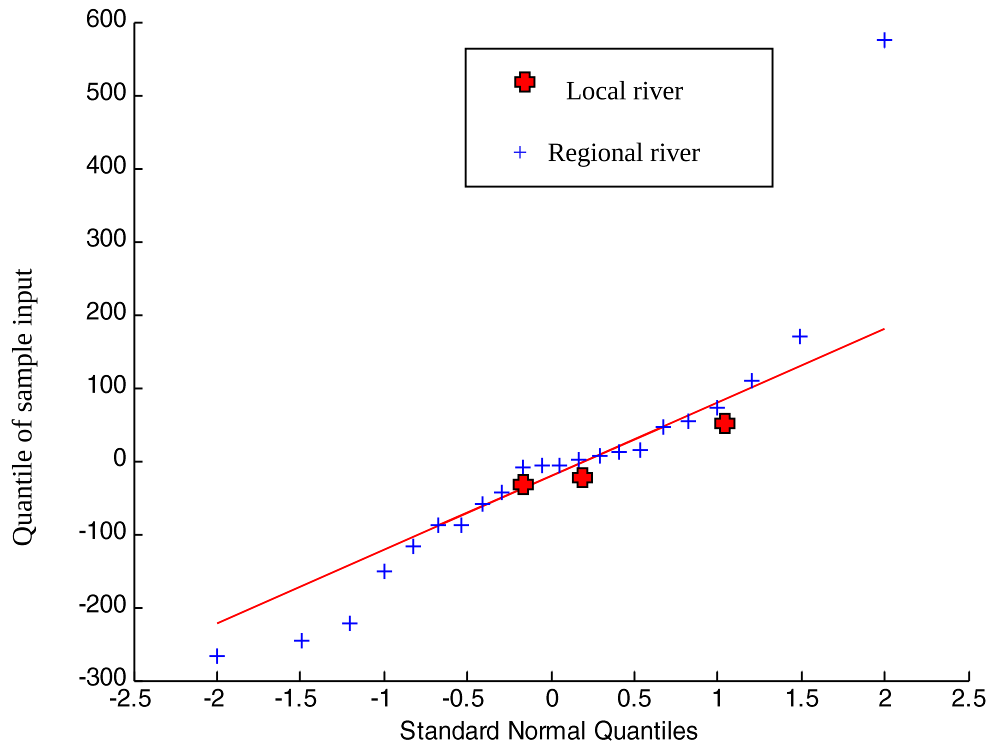


Figure 18: Normal Q-Q plot

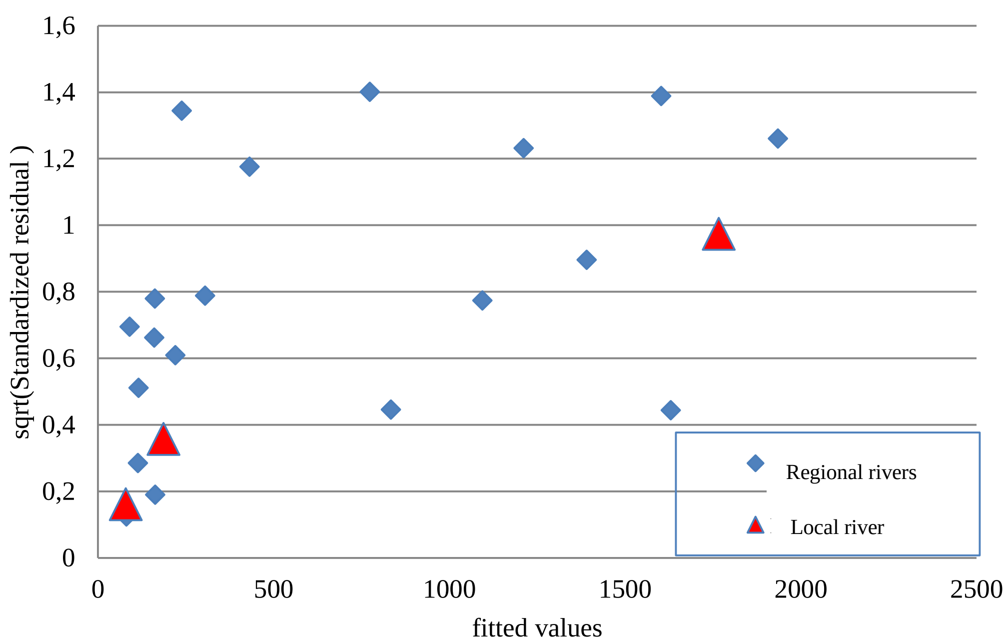


Figure 19: Scale location plot

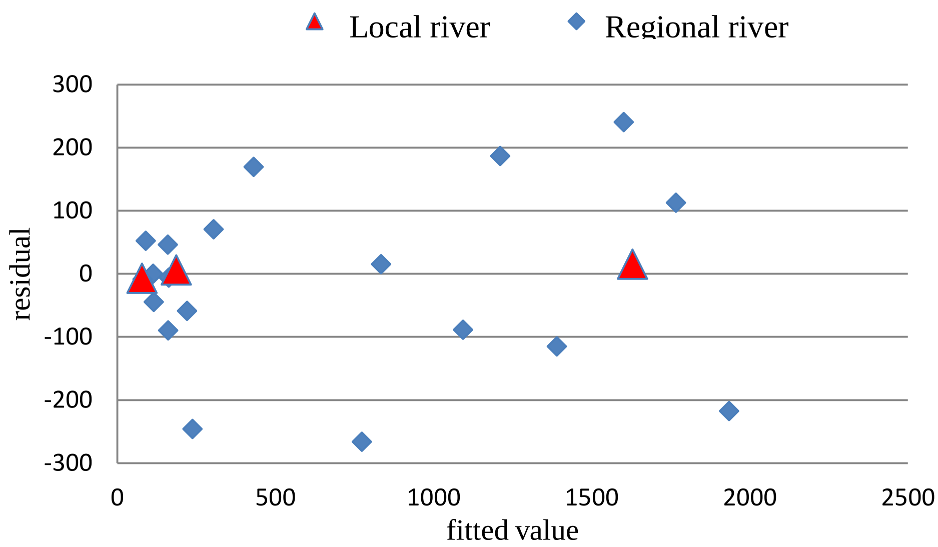


Figure 20: Residual versus fitted value plot

Conclusion and Recommendations

Since limited information is available about natural flow in the pre-reservoir period, the challenge is to select the best method of flow reconstructing between Kalman, Move III, area ratio and multivariable regression. Although each method produces smooth and nonnegative flows, they do vary in quality. Looking at their QIs and visual comparisons, it is evident that area ratio is a weak model for data reconstruction. The other methods, however, perform differently at different times of the year. Generally, Move III (x=HSAMI) has higher QIs than the others but this does not hold true year-round nor does it in every year. For example, the Kalman method produced better results for the year 1971. The best way to calculate natural flow for the pre-reservoir period is to compare the quality of reconstructed NF for each time segment and then apply the method that performed best for that segment. An alternative is to calculate the weighted NF for each segment based on the defined QI for each method.

In this paper, a new optimization model has been proposed and employed in the reconstruction of daily NF of ungauged basins containing a reservoir. The new optimization model has been solved using a posterior deterministic GA and a stochastic method that automatically define the model's parameters. Application of this method to the Outardes Basin (Quebec, Canada) shows that it is highly capable of improving the result of the classic WBE Model. The results do not contain any negative flow, they are less noisy, perfectly matched with regional flows, and are reliable enough for frequency analyses. Table 3 and Figures 8 to 16 enable us to conclude that the performance of the model depends on the particular case and on the time of year. Hence, as with the pre-reservoir period, two approaches are relevant to finalizing the NF: choosing the best method for that segment based on the average calculated QI for that segment, or calculating the weighted NF for each segment based on the defined QI for each method.

Although the proposed model works much better than the classic WBE Model for every reservoir size, there are still some deficiencies. These problems are more visible in small reservoirs (Outardes 3) and those affected by upstream reservoirs changes. These problems are caused by:

- Model uncertainty: This model does not determine specifically the effects of wind on the reservoir. Also, the effects of lamination are not contemplated for calculating the delayed flow from upstream reservoir in Outardes 3 and 2.
- Input data uncertainties: Since the input data is not validated, some uncertainties in the results cannot be avoided.

Since these NF values include evaporation, infiltration, and other losing terms of a closed hydraulic system, more studies are required so as to define the contribution of each of them.

Table 4: Different QIs for three methods of flow reconstructing after reservoir in Outardes 4, Outardes 3, and Outardes 2

		Deterministic WBE model						Stochastic WBE model						Classic WBE		
		O 4	IR	O 3	IR	O 2	I R	O 4	IR	O 3	IR	O 2	IR	O 4	O 3	O 2
NN	2005	0.885	7.4	0.821	31.4	0.828	14.0	0.860	4.4	0.798	27.7	0.806	11.0	0.824	0.625	0.726
CC		0.682	23.1	0.635	20.5	0.641	9.8	0.617	11.4	0.611	15.9	0.617	5.7	0.554	0.527	0.584
NAVE		0.900		0.808		0.801		0.905		0.661		0.797				
SFR		0.972		0.916		0.995		0.999		0.934		0.999				
NT		0.772	33.3	0.491	109.8	0.551	68.5	0.759	31.1	0.507	116.7	0.542	65.7	0.579	0.234	0.327
average		0.842		0.734		0.763		0.828		0.702		0.752				
NN	2007	0.864	7.2	0.841	114.0	0.731	28.0	0.852	5.7	0.744	89.3	0.806	41.2	0.806	0.393	0.571
CC		0.627	13.4	0.570	11.3	0.638	25.8	0.660	19.3	0.575	12.3	0.617	21.7	0.553	0.512	0.507
NAVE		0.888		0.812		0.713		0.889		0.631		0.797				
SFR		0.969		0.921		0.984		0.999		0.939		0.999				
NT		0.819	33.2	0.623	181.9	0.711	88.1	0.819	33.2	0.669	202.7	0.748	97.9	0.615	0.221	0.378
average		0.834		0.753		0.755		0.844		0.712		0.793				
NN	2008	0.808	2.0	0.650	33.2	0.875	16.7	0.801	1.1	0.676	38.5	0.811	8.1	0.792	0.488	0.750
CC		0.645	19.2	0.631	19.1	0.648	15.7	0.642	18.7	0.631	19.1	0.639	14.1	0.541	0.530	0.560
NAVE		0.920		0.884		0.783		0.924		0.700		0.773				
SFR		0.968		0.921		0.984		0.999		0.940		0.998				
NT		0.773	26.1	0.525	162.5	0.509	70.8	0.755	23.2	0.637	218.5	0.536	79.9	0.613	0.200	0.298
average		0.823		0.722		0.760		0.824		0.717		0.751				
NN	2009	0.880	5.9	0.808	52.5	0.899	22.1	0.862	3.7	0.754	42.3	0.843	14.5	0.831	0.530	0.736
CC		0.611	13.1	0.616	22.2	0.616	14.7	0.622	15.2	0.636	26.2	0.605	12.7	0.540	0.504	0.537
NAVE		0.901		0.873		0.759		0.907		0.693		0.758				
SFR		0.969		0.921		0.984		0.999		0.939		0.997				
NT		0.775	27.5	0.628	182.9	0.646	85.1	0.786	29.3	0.605	172.5	0.640	83.4	0.608	0.222	0.349
average		0.827		0.769		0.781		0.835		0.725		0.769				
NN	2010	0.519	10.7	0.773	80.6	0.714	29.8	0.499	6.4	0.733	71.3	0.664	20.7	0.469	0.428	0.550
CC		0.551	3.6	0.559	14.5	0.625	15.7	0.559	5.1	0.600	23.0	0.674	24.8	0.532	0.488	0.540
NAVE		0.901		0.808		0.719		0.911		0.657		0.706				
SFR		0.969		0.921		0.984		0.999		0.939		0.997				
NT		0.814	24.5	0.485	142.5	0.576	94.6	0.772	18.0	0.588	194.0	0.588	98.6	0.654	0.200	0.296
average		0.751		0.709		0.723		0.748		0.704		0.726				
NN	2011	0.916	1.9	0.883	54.9	0.840	15.1	0.913	1.6	0.838	47.0	0.845	15.8	0.899	0.570	0.730
CC		0.600	10.1	0.637	15.4	0.632	17.5	0.627	15.0	0.599	8.5	0.591	9.9	0.545	0.552	0.538
NAVE		0.936		0.864		0.736		0.929		0.660		0.733				
SFR		0.969		0.921		0.985		0.999		0.938		0.998				
NT		0.773	24.5	0.611	195.2	0.628	96.9	0.763	22.9	0.624	201.4	0.644	101.9	0.621	0.207	0.319
average		0.839		0.783		0.764		0.846		0.732		0.762				

Note : O2=Outardes 2, O3=Outardes 3, O4=Outardes 4

References

Bisson, J.-L. (1995). “Crue maximale probable de la rivière Gatineau”. Service Prévisions, *Internal Technical Report*, Hydro-Quebec.

Caluparo (2013). Pourvoirie – Outfitter. Le Québec et ses régions administratives. Website visited on October 2013 : <http://clauparo.com/photos%20site/Carteprov%20copier.jpg>

CEHQ (2013). Centre d’Expertise Hydrique du Québec. Website visited on October 2013: http://www.cehq.gouv.qc.ca/

Hosseinpour A., Dolcine, L. and Fuamba, M. (2013). “Natural Flow Reconstructing Using Kalman Filter and Water Balance Based Methods. Part 2: Case Studies, Results and Discussion”. Paper submitted to the *ASCE Journal of Hydrologic Eng*.

Hughes, D. A., and Smakhtin, V. (1996). “Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves”. *Hydrological Sciences Journal*, 41(6): 851-871.

Jones, P.D., Lister, D.H., and Kostopoulou, E. (2004). “Reconstructed river flow series from 1860’s to present (2002)”. *Science Report* SC040052/SR, Environment Agency.

Nicol, P. (2010)."Étude de crue pour la sécurité du barrage de Carillon, Calibration du Modèle HSAMI & Montage du modèle de laminage". *Internal Technical Report*, Hydro-Quebec.

Schreider, P., and Demuth, S. (1997). “Regionalization of low flows in southwest Germany”. *Hydrological Sciences Journal*, 42(6): 845-858.

Vogel, R.M., and Stedinger, J.R. (1985). “Minimum Variance Streamflow Record Augmentation Procedures”. *Water Resources Research*, 21(5): 715-723.