

**ASSIMILATION DE DONNEES DANS LE MODELE HYDROLOGIQUE CEQUEAU :  
UTILISATION DE FILTRE PARTICULAIRE**

**Rapport de recherche R1695**

**JUIN 2016**

---

**ASSIMILATION DE DONNEES DANS LE MODELE HYDROLOGIQUE CEQUEAU :  
UTILISATION DE FILTRE PARTICULAIRE**

O. Chimi Chiadjeu<sup>1</sup>, M.-A. Boucher<sup>2</sup>

A. St-Hilaire<sup>1</sup>, F. Chebana<sup>1</sup>

**RAPPORT PRÉSENTÉ À RIO TINTO**

1. Institut National de la Recherche Scientifique

490, rue de la Couronne

Québec (Québec) G1K 9A9

2. Université du Québec à Chicoutimi

555, boulevard de l'Université

Chicoutimi (Québec) G7H 2B1

**Rapport de recherche R1695**

**JUIN 2016**

© INRS-ETE, 2016

Tous droits réservés

ISBN : 978-2-89146-874-9 (version électronique)

Dépôt légal - Bibliothèque et Archives nationales du Québec, 2017

Dépôt légal - Bibliothèque et Archives Canada, 2017

---

# TABLE DES MATIÈRES

---

<b>LISTE DES TABLEAUX</b>	<b>II</b>
<b>LISTE DES FIGURES</b>	<b>III</b>
<b>1. INTRODUCTION</b>	<b>5</b>
<b>2. FILTRE PARTICULAIRE</b>	<b>8</b>
2.1 DESCRIPTION GENERALE	8
2.2 ALGORITHME DE FILTRAGE PARTICULAIRE EXISTANT	10
<b>3. METHODOLOGIE</b>	<b>13</b>
3.1 MODELE HYDROLOGIQUE CEQUEAU	13
3.2 STRATEGIE D'ASSIMILATION DANS LE MODELE HYDROLOGIQUE CEQUEAU	13
3.3 ZONE D'ETUDE	16
<b>4. OUTILS D'ANALYSE DE PERFORMANCE</b>	<b>18</b>
<b>5. RESULTATS ET ANALYSES</b>	<b>20</b>
<b>6. COMCLUSIONS ET PERSPECTIVES</b>	<b>25</b>
<b>7. REFERENCES</b>	<b>27</b>
<b>A. ASSIMILATION INDIRECT: PERTUBATION DES DONNEES METEOROLOGIQUES</b>	<b>29</b>
<b>B. UTILISATION DU SEQUENTION IMPORTANCE RESAMPLING (SIR)</b>	<b>34</b>
<b>C. UTILISATION DU REGULARIZED PARTICLE FILTER (RPF)</b>	<b>36</b>
<b>D. ALGORITHMES</b>	<b>39</b>

# **LISTE DES TABLEAUX**

---

# LISTE DES FIGURES

FIGURE 2-1 : LE FILTRE PARTICULAIRE AVEC UN RÉ ÉCHANTILLONNAGE. L'AXE VERTICAL REPRÉSENTE LA POSITION DES PARTICULES. LA TAILLE DE CHAQUE PARTICULE REPRÉSENTE SON POIDS. L'AXE HORIZONTAL REPRÉSENTE LE TEMPS. À L'INSTANT INITIAL (T=0) TOUTES LES PARTICULES ONT LE MÊME POIDS. À T = 10, LES POIDS DES PARTICULES REPRÉSENTENT SA VRAISEMBLANCE ET LE RÉ-ÉCHANTILLONNAGE PERMET D' AVOIR L'ENSEMBLE DE POIDS ÉGAUX.....	11
FIGURE 2-2 DIFFÉRENTS ALGORITHMES DE FILTRE PARTICULAIRE .....	12
FIGURE 3-1 : DIFFÉRENTS ALGORITHMES DE FILTRE PARTICULAIRE PLUS LE MODIFIÉ AUXILIARY SAMPLING IMPORTANCE RESAMPLING. ....	15
FIGURE 3-2: BASSINS VERSANTS DU LAC-SAINT-JEAN ET NECHAKO.....	17
FIGURE 5-1 : RÉSULTATS DES SIMULATIONS SUR LES BASSINS VERSANTS DU LAC-SAINT-JEAN (GAUCHE) ET NECHAKO (DROITE) EN 1989 POUR L'ÉCHANTILLONNAGE A PRIORI SUR LES ÉTATS : A) DÉBITS OBSERVÉS, LES SIMULATIONS D'ENSEMBLE ET DÉTERMINISTE ET, LA MOYENNE DES SIMULATIONS D'ENSEMBLE. B) LES PARTICULES SIMULÉES ET LEURS MOYENNES POUR CHAQUE ÉTATS : STOCK DE NEIGE EN CLAIRIÈRE (SNC), STOCK DE NEIGE EN FORÊT (SNF), NIVEAU D'EAU DANS LE LAC ET MARRAIS (NELM), NIVEAU EAU NAPPE (NEN) ET NIVEAU EAU SOL (NES). C) LE DIAGRAMME DE RANG À GAUCHE ET LE DIAGRAMME DE FIABILITÉ À DROITE.....	21
FIGURE 5-2: POURCENTAGE DE COUVERTURE FORESTIER EN FONCTION DE L'ALTITUDE SUR LES BASSINS VERSANTS DU LAC-SAINT-JEAN ET NECHAKO. ....	22
FIGURE 5-3: BILAN HYDRIQUE (GAUCHE) ET L'ERREUR SUR LE VOLUME DE LA CRUE (DROITE) POUR UNE SIMULATION DÉTERMINISTE ET LA MOYENNE D'ENSEMBLE SUR LES DEUX BASSINS VERSANTS. ....	23
FIGURE 5-4: COEFFICIENT DE NASH-SUTCLIFFE (GAUCHE) ET MCRPS (DROITE) ÉVALUÉ SUR LA SIMULATION DÉTERMINISTE, LES MOYENNES D'ENSEMBLE POUR UN ÉCHANTILLONNAGE DIRECT SUR LES ÉTATS ET POUR CHAQUE VALEUR D'INCERTITUDE QUAND ON ÉCHANTILLONNE PAR PERTURBATIONS DES OBSERVATIONS MÉTÉOROLOGIQUES. LA VALEUR 2 CORRESPOND À UNE PERTURBATION SUIVANT UNE LOI UNIFORME DE PARAMÈTRES -2 ET 2. ....	24
FIGURE A-1 : RÉSULTATS DES SIMULATIONS SUR LES BASSINS VERSANTS DU LAC-SAINT-JEAN (GAUCHE) ET NECHAKO (DROITE) EN 1989 AVEC UNE PERTURBATION QUI SUIT UNE LOI UNIFORME DE PARAMÈTRES -2 ET 2 MM (OU°C), POUR L'ÉCHANTILLONNAGE SUR L'OBSERVATION MÉTÉOROLOGIQUE : A) DÉBITS OBSERVÉS, LES SIMULATIONS D'ENSEMBLE ET DÉTERMINISTE ET, LA MOYENNE DES SIMULATIONS D'ENSEMBLE. B) LES PARTICULES SIMULÉES ET LEURS MOYENNES POUR CHAQUE ÉTATS : STOCK DE NEIGE EN CLAIRIÈRE (SNC), STOCK DE NEIGE EN FORÊT (SNF), NIVEAU EAU LAC ET MARRAIS (NELM), NIVEAU EAU NAPPE (NEN) ET NIVEAU EAU SOL (NES). C) LE DIAGRAMME DE RANG À GAUCHE ET LE DIAGRAMME DE FIABILITÉ À DROITE. ....	31
FIGURE A-2 : RÉSULTATS DES SIMULATIONS SUR LES BASSINS VERSANTS DU LAC-SAINT-JEAN (GAUCHE) ET NECHAKO (DROITE) EN 1989 AVEC UNE PERTURBATION QUI SUIT UNE LOI UNIFORME DE PARAMÈTRES-30 ET 30 MM (OU°C),POUR L'ÉCHANTILLONNAGE SUR L'OBSERVATION MÉTÉOROLOGIQUE : A) DÉBITS OBSERVÉS, LES SIMULATIONS D'ENSEMBLE ET DÉTERMINISTE ET, LA MOYENNE DES SIMULATIONS D'ENSEMBLE. B) LES PARTICULES SIMULÉES ET LEURS MOYENNES POUR CHAQUE ÉTATS : STOCK DE NEIGE EN CLAIRIÈRE (SNC), STOCK DE NEIGE EN FORÊT (SNF), NIVEAU EAU LAC ET MARRAIS (NELM), NIVEAU EAU NAPPE (NEN) ET NIVEAU EAU SOL (NES). C) LE DIAGRAMME DE RANG (GAUCHE) ET LE DIAGRAMME DE FIABILITÉ (DROITE).....	32
FIGURE A-3:ÉVOLUTION DU MCRPS (GAUCHE)ET DU COEFFICIENT DE NASH-SUTCLIFFE (DROITE) DE LA SIMULATION DÉTERMINISTE, DE LA SIMULATION EN FONCTION DE L'INCERTITUDE APPLIQUÉE SUR LES DONNÉES MÉTÉOROLOGIQUES ET DE LA SIMULATION OBTENU AVEC UN ÉCHANTILLONNAGE DIRECTEMENT.....	33
FIGURE B-1: RÉSULTATS DES SIMULATIONS AVEC LE SIR SUR LES BASSINS VERSANTS DU LAC-SAINT-JEAN (GAUCHE) ET NECHAKO (DROITE) EN 1989 POUR L'ÉCHANTILLONNAGE DIRECT SUR LES ÉTATS : A) DÉBITS OBSERVÉS, LES SIMULATIONS D'ENSEMBLE ET DÉTERMINISTE ET, LA MOYENNE DES SIMULATIONS D'ENSEMBLE. B) LES PARTICULES SIMULÉES ET LEURS MOYENNES POUR CHAQUE ÉTATS : STOCK DE NEIGE EN CLAIRIÈRE (SNC), STOCK DE NEIGE EN FORÊT (SNF), NIVEAU EAU LAC ET MARRAIS (NELM), NIVEAU EAU NAPPE (NEN) ET NIVEAU EAU SOL (NES). C) LE DIAGRAMME DE RANG (GAUCHE) ET LE DIAGRAMME DE FIABILITÉ (DROITE).....	35
FIGURE C-1 : RÉSULTATS DES SIMULATIONS AVEC LE RPF SUR LES BASSINS VERSANTS DU LAC-SAINT-JEAN (GAUCHE) ET NECHAKO (DROITE) EN 1989 POUR L'ÉCHANTILLONNAGE L' A PRIORI SUR LES ÉTATS : A) DÉBITS OBSERVÉS, LES SIMULATIONS D'ENSEMBLE ET DÉTERMINISTE ET, LA MOYENNE DES SIMULATIONS D'ENSEMBLE. B) LES PARTICULES SIMULÉES ET LEURS MOYENNES POUR CHAQUE ÉTATS : STOCK DE NEIGE EN CLAIRIÈRE (SNC),	

STOCK DE NEIGE EN FORÊT (SNF), NIVEAU EAU LAC ET MARRAIS (NELM), NIVEAU EAU NAPPE (NEN) ET NIVEAU EAU SOL (NES). C) LE DIAGRAMME DE RANG À GAUCHE ET LE DIAGRAMME DE FIABILITÉ À DROITE..... 38

# 1. INTRODUCTION

---

Le modèle CEQUEAU (Morin & Paquet, 2007) est un modèle hydrologique qui permet de prévoir le ruissellement et les apports de débits, connaissant la physiographie du bassin versant, et les prévisions météorologiques. Les développements récents entrepris par l'industriel Rio Tinto (RT) donnent la possibilité d'intégrer de nouveau module dans le modèle. L'intégration des nouveaux modules permettra d'améliorer les prévisions faites par le modèle. Une vaste campagne d'amélioration de la prévision faite par le modèle CEQUEAU a été entreprise dans l'amélioration de la qualité de modélisation de l'évapotranspiration, de modélisation de la fonte de neige et de simulation des débits. Ce travail consiste à améliorer la qualité simulation des débits, en y intégrant un module d'assimilation de données. La première tentative d'assimilation de donnée dans le modèle CEQUEAU a été faite par Turcotte et al. en 2010 (Turcotte, et al., 2010). Assimiler les données dans le modèle CEQUEAU revient à substituer les états par les données mesurées quand elles sont disponibles. Faire cela revient à donner une crédibilité parfaite aux mesures.

L'assimilation de données est donc définie comme l'ensemble des techniques statistiques qui permettent d'améliorer la connaissance de l'état d'un système à partir de sa connaissance théorique et des observations expérimentales. La connaissance d'un état le plus juste possible permet de produire des prévisions les plus fiables. L'assimilation de données a ainsi pour objectif de construire cette condition initiale afin d'améliorer les prévisions et les simulations faites par un modèle. Les algorithmes d'assimilation de données moins récentes sont basés sur les approches variationnelles (3D-Var, 4D-Var). Elles consistent à minimiser une fonction de coût qui, dans la plupart des cas, mesure l'écart aux observations (Talagrand & Courtier, 1987) (Seo, Cajina, Corby, & Howison, 2009) (Reichle, 2008). L'utilisation de ces approches revient à supposer implicitement le modèle parfait et tout écart aux observations est causé par l'erreur sur les conditions initiales. D'autres approches d'assimilation de données utilisent le filtre de Kalman. L'utilisation du filtre de Kalman suppose la linéarité du modèle d'évolution des états et l'opérateur d'observation. En pratique l'hypothèse de linéarité n'est pas toujours satisfaite. Par ailleurs, si le modèle est linéarisable, le filtre de Kalman étendu (Extended Kalman Filter : EKF) qui est une forme généralisé du filtre de Kalman est utilisé. Le filtre de Kalman d'ensemble (EnKF) a été

proposé par Evensen (Evensen, 1994). Pour une description détaillée, se référer à Evensen (Evensen, 2003). Cette méthode a été présentée comme une alternative « stochastique » au filtre de Kalman étendu. Dans le cadre de système de grande taille non linéaire, l'utilisation d'une méthode de Monte Carlo a été développée pour résoudre les problèmes coût très important et de mauvaise réponse en cas de forte non-linéarité du filtre de Kalman étendu. La principale différence avec le filtre de Kalman réside dans la propagation des erreurs dans le modèle. Contrairement au filtre de Kalman où l'erreur est défini dans une matrice de covariance, dans le cadre l'EnSKF, les erreurs se propagent dans le modèle sous forme de nuage de point ou particule. Malgré l'emprunt de la notion de particule, ce filtre reste Gaussien (Clark, et al., 2006) (Papadakis, Mémin, Cuzol, & Gengembre, 2010) (Slatter & Clark, 2005). Pour s'affranchir de ces hypothèses restrictives (modèles linéaire et gaussien), il faut alors avoir recours à des approximations ou des techniques numériques basées sur des méthodes de Monte-Carlo séquentielle, communément appelé 'Sampling Importance Sampling' (SIS) ou encore filtrage particulaire. Pour plus d'information, consultez la revue de littérature de Peter Jan Van Leeuwen (Van Leeuwen, 2009). Le filtre particulaire est actuellement l'outil le plus adéquat pour l'assimilation des données dans les modèles géophysiques car, la plupart de ces modèles ne sont ni linéarisables, ni gaussiennes (Moradkhani & Hsu, 2005) (Leisenring & Moradkhani, 2011). De plus en plus utilisé dans les modèles hydrologiques, il sert aussi bien pour l'estimation des paramètres des modèles que pour, la correction des états d'un modèle hydrologique (Salamon & Feyen, 2009) (Montzka, et al., 2011). Les études récentes ouvrent les possibilités d'utilisation du filtre particulaire pour les systèmes de grande dimension (Ades & Van Leeuwen, 2015).

Concernant la méthode d'échantillonnage des particules, en hydrologie il existe principalement deux courants de pensée : La première consiste à échantillonner les particules par perturbation des observations météorologiques et la deuxième méthode est l'échantillonnage par perturbation direct des états. L'utilisation de la première méthode est liée au fait qu'on suppose que les incertitudes sont pour la plupart liée aux observations météorologiques (Abaza, Anctil, Fortin, & Turcotte, 2015) (Plaza, et al., 2012). Ici, les résultats montrent qu'une perturbation des données météorologiques ne permet pas d'introduire assez de variabilité sur les états (Annexe A). Ceci ne nous permet pas de reproduire l'état correspondant à l'état du modèle pour produire le débit observé. Par ailleurs, on note qu'une augmentation de l'incertitude sur les observations entraîne

une amélioration des simulations sur l'un de nos bassins versants. Par ailleurs, ce résultat est mitigé sur l'autre. En échantillonnant les particules directement sur les états, il y a risque pour la non-conservation le bilan hydrique. Dans la suite, l'échantillonnage de nos particules est fait directement sur les états et pour chaque période de simulation. L'impact sur le bilan hydrique sera aussi évalué pour l'échantillonnage direct.

Cette étude a pour objectif de fournir une méthodologie permettant d'assimiler les observations de débits dans le modèle hydrologique CEQUEAU. Après avoir testé les quelques approches de filtrage particulaire (Annexe B et C), nous proposons un algorithme inspiré des approches décrites dans la littérature. Dans la première partie de ce document, l'aspect général du filtre particulaire est présenté. Dans la deuxième partie, la méthodologie utilisée pour l'assimilation des données dans le modèle CEQUEAU. Dans la troisième partie les résultats obtenus avec la méthode d'assimilation choisie et un échantillonnage fait directement sur les états. Enfin, la dernière partie est réservée à la conclusion et la présentation des perspectives.

## 2. FILTRE PARTICULAIRE

---

### 2.1 Description générale

Posons  $x_t$  un vecteur d'états et  $y_t$  une suite d'observation  $t = 1, \dots, T$ , où  $T$  est le temps de simulation. Considérons qu'il existe la fonction  $f()$  qui régit l'évolution des états et la fonction  $h()$  qui permet de simuler les observations. Dans le cadre de ce travail,  $h()$  est le modèle hydrologique CEQUEAU. Posons:

$$x_t = f(x_{t-1}) + e_t \quad (1)$$

$$y_t = h(x_t) + o_t \quad (2)$$

$e_t$  est l'erreur du modèle et  $o_t$  est l'erreur sur les observations. Un filtre particulaire nous donne un estimateur des états inconnus  $x_t$  pour  $t = 1, \dots, T$  en se basant sur les données observées  $y_t$  pour  $t = 1, \dots, T$ . Les états  $x_t$  sont approximés par les distributions *a posteriori*  $P(x_1, \dots, x_t | y_1, \dots, y_t)$ . Le filtre particulaire permet l'estimation de  $P(x_1, \dots, x_t | y_1, \dots, y_t)$  ou plus précisément  $P(x_t | y_{1:t})$ .

**Rappel :** Soit deux variables A et B. Le théorème de Bayes est basé sur une connaissance de la densité de probabilité conditionnelle.

$$P(A, B) = P(A|B) * P(B) = P(B|A) * P(A) \text{ Alors } P(B|A) = \frac{P(B|A)*P(A)}{P(B)}$$

$$P(x_t | y_{1:t}) = \frac{P(y_t | x_{1:t}) P(x_t | y_{1:t-1})}{P(y_t | y_{1:t-1})} \quad (3)$$

Les particules simulées suivant la distribution *a posteriori* sont utilisées pour approcher l'intégrale par l'expression suivante :

$$P(x_{0:t} | y_{1:t}) = \sum_{i=1}^N w_t^i \delta(x_{0:t} - x_{0:t}^i) \quad (4)$$

Où le couple  $\{w_t^i, x_t^i\}$  représente le  $i^{\text{ème}}$  poids et sa particule correspondante,  $\delta()$  représente la fonction de Dirac, et  $N$  le nombre de particules. Le calcul de la densité  $P(x_t | y_{1:t})$  sert à l'évaluation d'un estimateur des états de la forme

$$E_p[f(x_{0:t})] = \int f(x_{0:t})P(x_{0:t}|y_{1:t})dx_{1:t} \quad (5)$$

Une solution pour calculer cette intégrale serait de l'approximer par la méthode de Monte-Carlo. L'approche Monte-Carlo consiste à utiliser des réalisations d'une variable aléatoire distribuée suivant la loi *aposteriori* pour le calcul des moyennes empiriques. Cette approche est intéressante dès que l'on sait échantillonner selon la loi *aposteriori*. Ceci n'étant malheureusement pas le cas. Pour résoudre ce problème, on préconise la technique d'échantillonnage d'importance. L'utilisation de cette méthode impose l'utilisation d'une autre densité de probabilité  $q(x_{0:t}|y_{1:t})$  dite densité de probabilité des observations, dont le support contient celui de  $P(x_{0:t}|y_{1:t})$  et suivant laquelle il est facile d'échantillonner.

En supposant  $q()$ , on peut écrire  $E_p[f(x_{0:t})]$  sous la forme :

$$E_p[f(x_{0:t})] = \int f(x_{0:t}) \frac{P(x_{0:t}|y_{1:t})}{q(x_{0:t}|y_{1:t})} q(x_{0:t}|y_{1:t}) dx_{1:t} \quad (6)$$

$$E_p[f(x_{0:t})] = E_q \left[ f(x_{0:t}) \frac{P(x_{0:t}|y_{1:t})}{q(x_{0:t}|y_{1:t})} \right] \quad (7)$$

Alors,  $x_{0:t}(i)$  suit la loi de  $q(x_{0:t}|y_{1:t})$   $i = 1, \dots, N$ . Cette intégrale (6) peut être approximée par :

$$E_p[f(x_{1:t})] = \sum_{i=1}^N w_t^{i*} f(x_{0:t}) \quad (8)$$

$$\text{Où, } w_t^{i*} = \frac{P(x_{0:t}^i|y_{1:t})}{q(x_{0:t}^i|y_{1:t})}$$

Les quantités  $w_t^i$  sont appelées les poids d'importance. Lorsque  $N$  tend vers l'infini, on converge presque sûrement vers  $E_p[f(x_{1:t})]$  (loi des grands nombre). Après quelques transformations et approximations (Sanjeev Arulampalam, Maskell, Gordon, & Clapp, 2002), on obtient une relation entre les poids d'importance entre les états  $t$  et  $t-1$  définie par :

$$w_t^{i*} \propto w_{t-1}^{i*} \frac{P(y_t|x_t^i)P(x_t^i|x_{t-1}^i)}{q(x_t^i|x_{t-1}^i, y_t)} \quad (9)$$

Les poids d'importance sont ensuite normalisés pour que leur somme soit égale à 1, c'est à dire  $w_t^{i*} = w_t^{i*} / \sum_{i=1}^N w_t^{i*}$ .

Le premier algorithme de filtrage particulaire est appelé ‘Séquentiel Monte Carlo’ ou communément ‘Sequential Importance Sampling’ (SIS) (Annexe A). Il consiste alors à la propagation séquentielle des poids et le support du point de chaque mesure reçue.

## 2.2 Algorithme de filtrage particulaire existant

Le principal problème lié à l’utilisation de cet algorithme et du filtre particulaire est la dégénérescence des particules. Ceci veut tout simplement dire qu’au cours des itérations, on a souvent une particule qui est privilégiée par rapport aux autres. Le poids normalisé de la particule privilégiée tend vers 1 tandis que celui des autres particules tendent vers 0. Une mesure efficace de la dégénérescence est la taille des échantillons  $N_{eff}$  défini part :

$$N_{eff} = \frac{1}{\sum_{i=1}^N (w_t^i)^2} \quad (10)$$

Où les  $w_t^i$  sont les poids des particules normalisées. Il faut noter que  $N_{eff} < N$  et la faible valeur de  $N_{eff}$  montre une dégénérescence sévère. Une solution pour réduire cette dégénérescence serait l’utilisation d’un grand nombre de particules ou d’ajouter une étape supplémentaire à l’algorithme de SIS. L’augmentation du nombre de particules ne pourra qu’exploser le temps de calcul. L’autre approche consiste à ré-échantillonner les particules à partir des particules initiales. Ce ré-échantillonnage consiste à abandonner les particules ayant les poids les moins importants ou les moins vraisemblables, et de faire proliférer les particules ayant les poids plus importants ou plus vraisemblables (Ristic, Arulampalam, & Gordon, 2004). Pour maintenir le nombre de particules constant, les particules retenues sont dupliquées par ‘Bootstrap’ (Annexe D : algorithme 2). Plusieurs méthodes de ré-échantillonnage ont été introduites dans la littérature, tel que le ré-échantillonnage multinomial, le ré-échantillonnage résiduel, le ré-échantillonnage stratifié et le ré-échantillonnage systématique (Van Leeuwen, 2009).

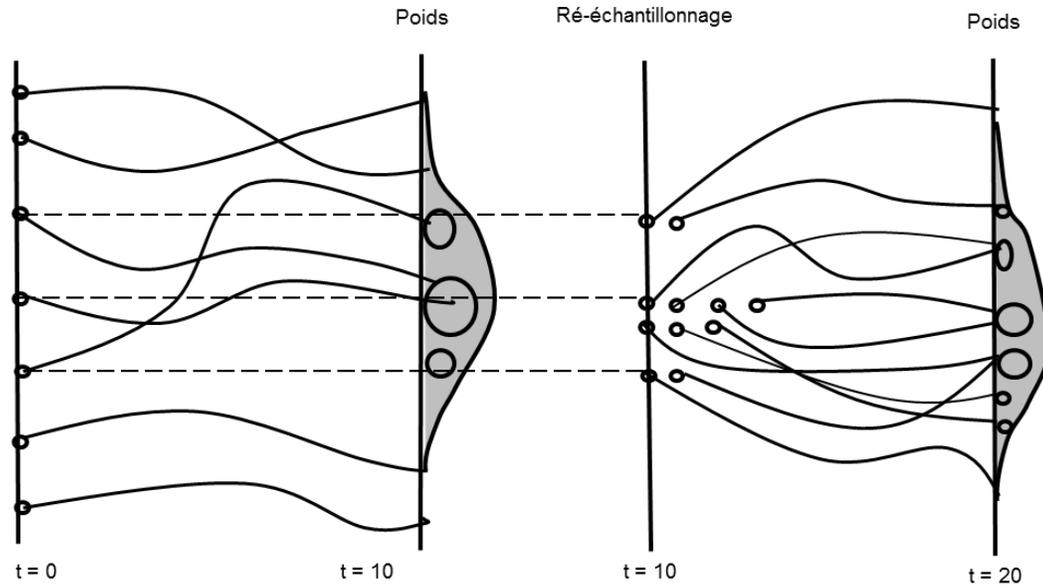


Figure 2-1 : Le filtre particulaire avec un ré échantillonnage. L'axe vertical représente la position des particules. La taille de chaque particule représente son poids. L'axe horizontal représente le temps. À l'instant initial ( $t=0$ ) toutes les particules ont le même poids. À  $t = 10$ , les poids des particules représentent sa vraisemblance et le ré-échantillonnage permet d'avoir l'ensemble de poids égaux (Van Leeuwen, 2009).

La recherche des méthodes pour résoudre le problème de dégénérescence a donné naissance à plusieurs algorithmes de d'assimilation de données utilisant le filtre particulaire. Pour résoudre le problème de dégénérescence, plusieurs algorithmes ont été implémentés. Ils découlent pour la plus part des modifications faites sur le SIS. La Figure 2-2 présente les différents algorithmes du filtre particulaire. Le 'Sequential Importance Resampling' (SIR) est une extension du SIS dans la mesure où, au SIS on y a ajouté l'étape de ré-échantillonnage. Malgré cette étape de 'Resampling', on observe une dégénérescence des particules après quelques itérations (Sanjeev Arulampalam, Maskell, Gordon, & Clapp, 2002). L'Auxiliary Sampling Importance Resampling (ASIR) qui est une version évoluée du SIR dans la mesure qu'au SIR, après l'étape de ré-échantillonnage, on a ajouté une étape supplémentaire qui consiste à échantillonner les nouvelles particules. L'échantillonnage est fait après le 'Resampling' et selon la connaissance *a priori* des particules précédents. Le 'Generic Particle Filter' (GPF) a été proposé comme une alternative à l'APF. Elle diffère du SIS par une étape supplémentaire dite de régularisation ajoutée après le Resampling si le test de dégénérescence est positif. Le 'Regularized Particle Filter' (RPF) est le GPF avec une étape supplémentaire qui consiste à tirer de nouvelles particules par la méthode de noyau (Noh,

Tachikawa, & Kim, 2011). Le noyau ici n'est rien d'autre que la loi de probabilité associée aux poids des particules.

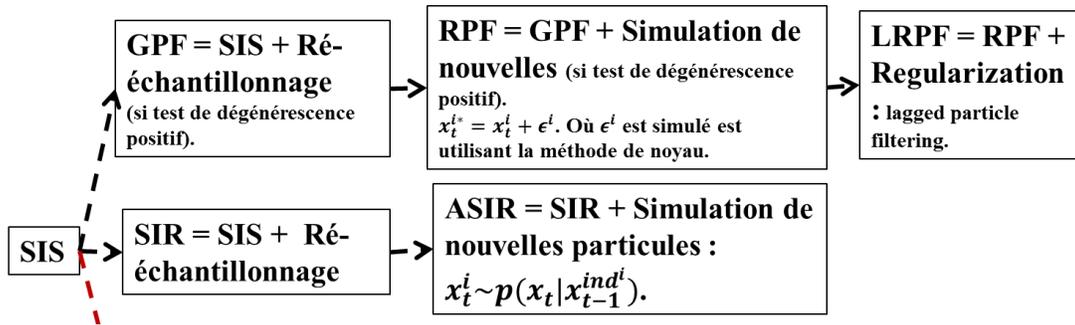


Figure 2-2 Différents algorithmes de filtre particulaire

## 3. METHODOLOGIE

---

### 3.1 Modèle hydrologique CEQUEAU

Le modèle CEQUEAU est un modèle hydrologique déterministe qui est couramment utilisé pour les prévisions de quantité et de qualité de l'eau (Morin & Paquet, 2007). Il prend en compte les données météorologiques, la physiographie du bassin versant qu'il découpe en carreaux et en sous carreaux. Les carreaux sont les carrés de même taille et un sous carreau est une superficie définie par la ligne de partage des eaux et incluse dans un carreau. Contrairement aux carreaux, leurs tailles ne sont pas identiques. Le modèle CEQUEAU est séparé en deux grandes parties : les fonctions de production qui prennent en entrée la température et la précipitation (solide et liquide) et renvoie la lame d'eau sur chaque sous carreau et une fonction de transfert qui gère le transfert des eaux entre les sous carreaux et l'écoulement jusqu'à l'exutoire du bassin versant.

### 3.2 Stratégie d'assimilation dans le modèle hydrologique CEQUEAU

Soit  $N_x$  le nombre de variables d'états. Pour chaque pas de temps  $t$ , la particule  $x_t = (x_t^1, \dots, x_t^{N_x})$  est un vecteur d'état et  $y_t$  une observation. Dans le cadre de notre projet, les variables constituant les particules du modèle et l'observation sont :

Particules  $x_t$  :

- Stock neige en clairière (SNC)
- Stock neige en forêt (SNF)
- Niveau de l'eau dans le réservoir nappe (NEN)
- Niveau de l'eau dans le réservoir sol (NES)
- Niveau de l'eau dans les Lacs et Marrais (NELM)

Observations  $y_t$  :

- Débit.

Notre but est d'assimiler les observations  $y_t$  et d'utiliser les filtres particulaires pour mettre à jour

les états du modèle CEQUEAU. On suppose comme dans les études qui font une assimilation de données par filtrage particulaire que la distribution d'échantillonnage  $q(x_t|x_{t-1}, y_t) = P(x_t|x_{t-1})$ . Alors, l'équation (9) s'écrit:

$$w_t^* \propto w_{t-1}^* P(y_t|x_t). \quad (14)$$

Supposons que l'erreur entre le débit simulé et observé ( $o_t$ ) suit une loi normale de moyenne nulle et d'écart-type  $\sigma$  tel que.

$$P(y_t|x_t) = P(h(x_t)|x_t) = G(h(x_t) - y_t) \quad (15)$$

où  $G$  est la densité de probabilité d'une gaussienne de moyenne nulle et d'écart-type  $\sigma_t = \alpha y_t + \beta$ . Cette définition assure l'hétérogénéité de l'écart-type.  $\alpha$  est inversement proportionnelle à  $y_k$  et  $\beta$  est utilisé comme incertitude pour les périodes de faible débit. Après avoir testé plusieurs valeurs, nous avons fixé  $\alpha = 0.01$  et  $\beta = 25 \text{ m}^3/\text{s}$ . L'écart-type définit dans la littérature (Salamon & Feyen, 2009), (Noh, Tachikawa, & Kim, 2011) est utilisé avec les valeurs de 0.125 et 0.1 pour  $\alpha$  et  $5 \text{ m}^3/\text{s}$  pour  $\beta$ . La différence entre leurs valeurs et le nôtre est due au débit d'eau produit par les bassins versants que nous avons étudiés.

Après avoir testé les algorithmes proposés dans la littérature, nous sommes arrivés à la même conclusion (Sanjeev Arulampalam, Maskell, Gordon, & Clapp, 2002) : Le RPF, GPF et ASIR ne sont pas adaptés pour les modèles dont les incertitudes sur les états sont grandes. Nous proposons un algorithme inspiré d'ASIR : Modified Auxiliary Sampling Importance Resampling (MASIR). Nous avons d'une part intégré le test le ré-échantillonnage dans notre algorithme et d'autre part, échantillonné uniquement les particules dupliquées. Les particules dupliquées pendant le RESAMPLING sont remplacées par de nouvelles particules échantillonnées suivant la loi  $p(x_t|\mu_t)$  où,  $\mu_t$  correspond à la moyenne des états pondérés par leur poids. L'algorithme 3 représente l'aspect général de l'algorithme d'assimilation de données dans le modèle CEQUEAU (Annexe D). On initialise le premier état moyen par la particule initiale et son débit est le débit correspondant à cet état.

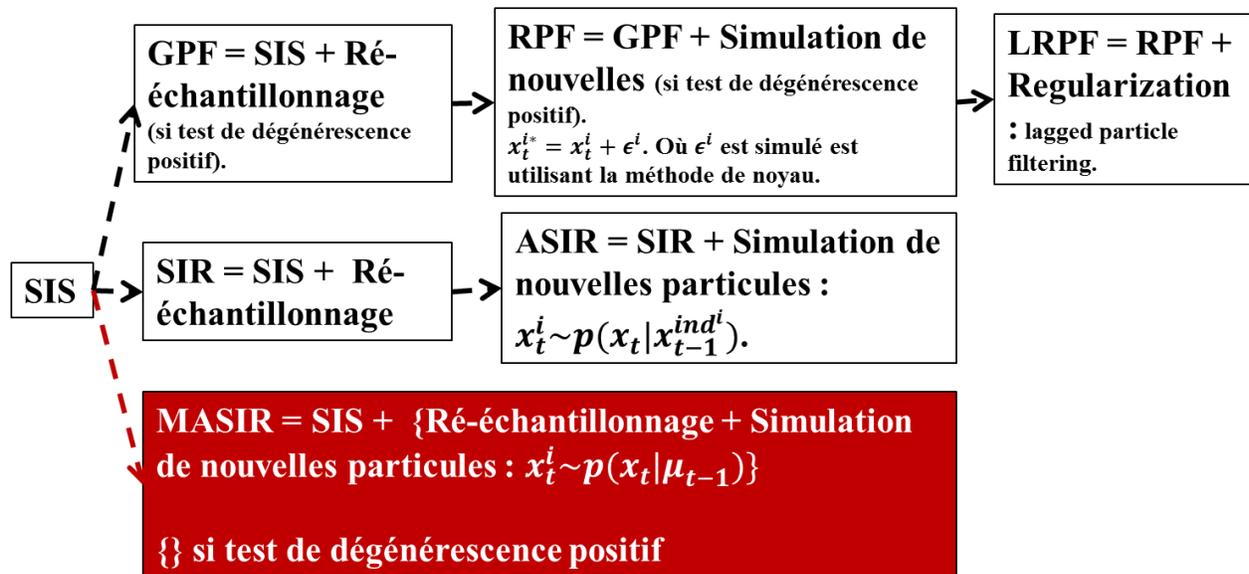


Figure 3-1 : Différents algorithmes de filtre particulaire plus le Modified Auxiliary Sampling Importance Resampling.

L'échantillonnage des particules est fait directement sur les états. À chaque particule on assigne un poids identique et on simule le débit correspondant. Pour chaque instant, on effectue une assimilation de données à l'aide de MASIR. Cette fonction prend en entrée les particules à l'instant  $t$  et leur poids respectif, l'état moyen et le débit observé. Il retourne les particules à l'état  $t+1$  et leur poids respectif, le débit simulé pour chaque particule à l'instant  $t+1$  et l'état moyen à l'instant  $t+1$ . À partir de cet état moyen, on calcule le débit correspondant.

L'algorithme 4 présente le module d'assimilation de données (Annexe D). C'est dans cette fonction qu'on assimile les observations. Pour chaque particule, on évalue le débit par le modèle CEQUEAU et on calcule le poids correspondant par l'équation (12). Après avoir normalisé les poids, si le test de dégénérescence est positif, on utilise la fonction RESAMPLE. Cette fonction sélectionne les particules ayant les poids moins importants et les remplace par les particules ayant les poids les plus importants ou plus vraisemblables. Ceci est fait de façon autonome par l'algorithme 3 défini à l'annexe D: On ne fixe ni nombre de particules à sélectionner ni un seuil qui permet de sélectionner un nombre de particules. Ensuite, on calcule l'état moyen pondéré des poids des particules. Toutes les particules dupliquées sont remplacées par de nouvelles particules qui, sont échantillonnées autour de l'état moyen. On assigne au nouvel ensemble de particules un

même poids.

### **3.3 Zone d'étude**

Les bassins versants utilisés pour notre étude sont : le bassin versant du Lac Saint Jean et le bassin versant Nechako (Figure 3-2). Sur ces bassins, les stations météorologiques et hydrométriques de RT permettent la collecte des données utilisées par le modèle CEQUEAU.

Le bassin versant du Lac Saint Jean est situé à l'Est du Canada dans la province de Québec. Il a une superficie de 73800 km<sup>2</sup> avec 90% de surface forestier, et le reste couvert de surface agricole, d'habitation et d'eau. Sur ce bassin versant l'élévation de surface varie entre 88 m et 792 m. Sur ce bassin, on a enregistré une précipitation moyenne d'environ 960 mm par an. Le bassin versant Nechako quant à lui est situé à l'Ouest du Canada dans la province de Colombie Britannique. Il a une superficie de 47 200 km<sup>2</sup> avec 63% de surface forestier et le reste en surface agricole et habitation. Sur ce bassin versant, l'élévation de surface varie entre 986 m et 1893 m. Sur ce bassin, on enregistre une précipitation moyenne d'environ 727 mm par an.

Les deux bassins versants ont été choisis à cause de leurs positions géographiques, leurs conditions météorologiques et leur importance pour la production d'hydroélectricité. Par-ailleurs, le bassin versant Nechako est particulièrement intéressant pour la neige. Nous disposons des données physiographiques des deux bassins versants et, les données météorologique et hydrologiques proviennent des stations situées sur et autour des bassins versants de 1985 à 2013 avec une fréquence journalière. Ne disposant pas de station météorologique sur chaque carreau, une interpolation par la méthode de polygone de Thiessen est nécessaire pour calculer les valeurs de température et précipitation sur chaque carreau.

- Hydrometric Station
- ▲ Meteorological Station

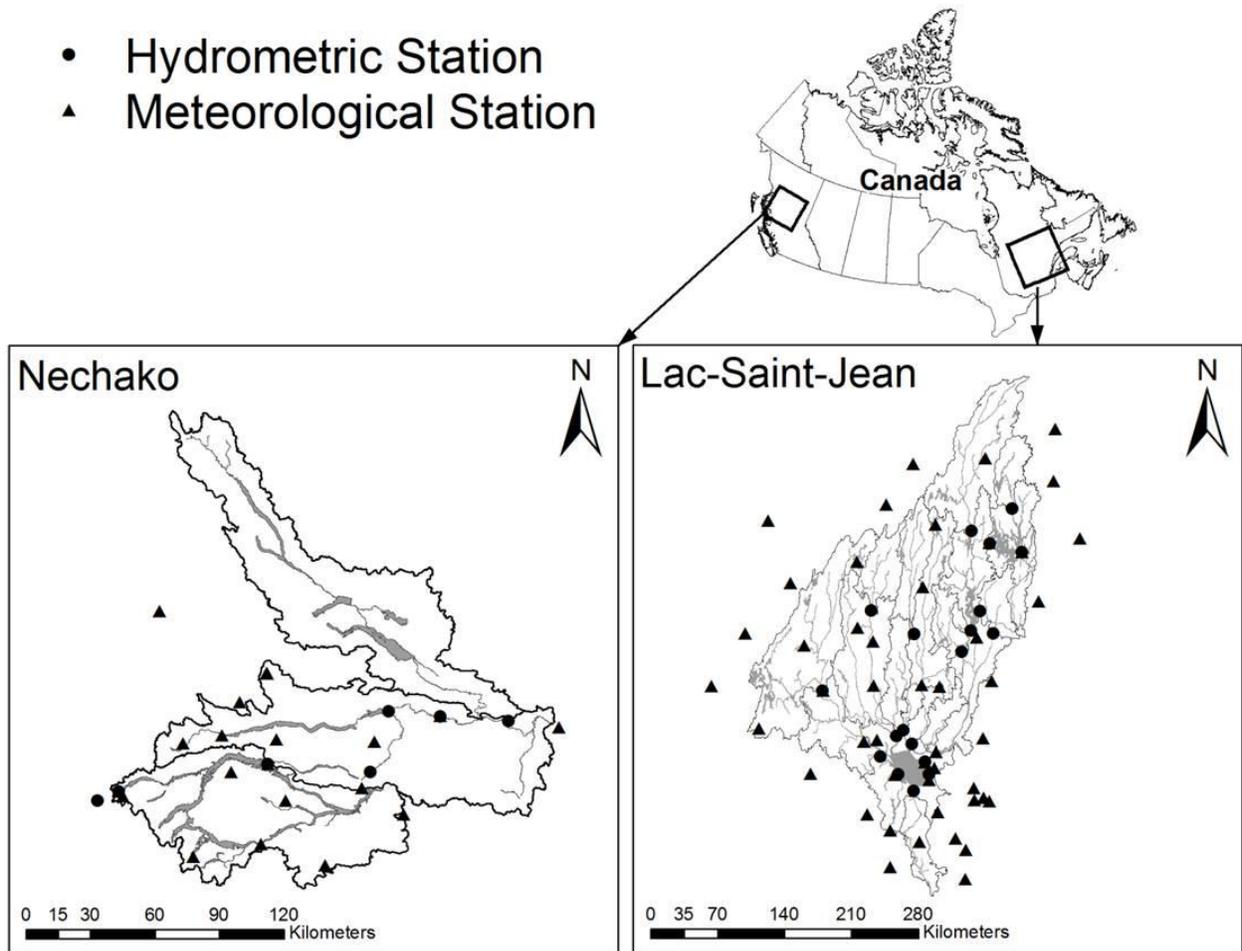


Figure 3-2: Bassins versants du Lac-Saint-Jean et Nechako

## 4. OUTILS D'ANALYSE DE PERFORMANCE

---

Pour évaluer la performance des simulations, nous allons utiliser en plus du coefficient de Nash - Sutcliffe, l'erreur sur le volume de la crue, le 'Continuous Ranked Probability Score' (CRPS), l'histogramme de rang et le diagramme de fiabilité.

Soit  $Q_t^{Obs}$ ,  $Q_t^{Sim}$  respectivement le débit observé et simulé à l'instant  $t$ , et  $T$  la durée de simulation. Le coefficient de Nash -Sutcliffe (NSE) est défini par :

$$NSE = 1 - \frac{\sum_{t=1}^T (Q_t^{Obs} - Q_t^{Sim})^2}{\sum_{t=1}^T (Q_t^{Obs} - \bar{Q}^{Obs})^2}. \quad (12)$$

Plus le coefficient de Nash-Sutcliffe est proche de 1, meilleur sera la simulation. On définit l'erreur sur le volume de crue (ErrVolCrue) par :

$$ErrVolCrue (m^3/j) = \sum_{t=1}^T (Q_t^{Obs} - Q_t^{Sim})/24 \times 3600. \quad (13)$$

Plus l'erreur sur le volume est proche de 0, meilleur sera la simulation.

Le CRPS permet d'évaluer la qualité d'une simulation ou d'une prévision d'ensemble en termes d'incertitude et de comparer les résultats à ceux de la simulation déterministe. Le calcul repose sur une fonction de densité de probabilité qui est construite sur la base des membres de l'ensemble.

$$CRPS(F_t, Q^{Obs}) = \int_{-\infty}^{+\infty} (F_t(Q) - H(Q > Q^{Obs}))^2 d_Q \quad (14)$$

$F_t(Q)$  est la fonction de distribution cumulée à l'instant  $t$ ,  $Q$  la variable prédite (débit simulé par chaque particule), et  $Q^{Obs}$  est la valeur observée correspondante (débit observé). La fonction  $H$  est la fonction de Heaviside ce qui équivaut à 0 pour des valeurs prédites plus petites que la valeur observée, 1 sinon. Le CRPS partage la même unité que la variable  $Q$  prédite. La moyenne des CRPS (MCRPS) nous donne une idée sur la qualité de simulation d'un ensemble de particules. Cette grandeur très pertinente en termes de simulation d'ensemble, permet de comparer les effets d'une quelconque modification sur l'algorithme. Exemple : étude de performance en fonction de la fréquence d'assimilation.

L'histogramme de rang est un outil graphique qui permet d'évaluer l'erreur de chaque membre par la simulation des sources d'incertitude. A chaque pas, on définit un vecteur contenant les simulations et les observations. Ensuite, on range les éléments de ce vecteur dans l'ordre décroissant. On retient le rang de l'observation à chaque pas de temps. Comme son nom l'indique, l'histogramme de rang est l'histogramme du rang des observations. Si l'histogramme est uniforme (tous les rangs ont la même fréquence d'apparition) alors, l'ensemble de particules représente bien toute les sources d'incertitude. Si l'histogramme est concave alors, les sources d'incertitude sont sous-estimées. Si l'histogramme est convexe, les sources d'incertitude sont surestimées. Si l'histogramme est asymétrique alors, le processus de simulation hydrologique comporte un biais.

Le diagramme de fiabilité illustre la fiabilité des performances et les caractéristiques d'un système de simulation probabiliste. C'est le tracé de la fréquence d'apparition des évènements simulés en fonction de la probabilité d'occurrence. Par exemple, pour évaluer la fréquence d'apparition dans un intervalle de 50%, on compte le nombre de fois que l'observation se situe entre 45% et 55% sur la fréquence cumulée des simulations. Ceci est fait sur chaque intervalle entre 0 et 100%. Dans le meilleur des cas, cette droite est identique à une droite de pente et d'ordonnée à l'origine 0. Si le tracé est au-dessus respectivement en dessous de la droite de pente 1, on peut dire qu'il y a une sur-dispersion respectivement sous-dispersion des erreurs sur les états.

Les deux premiers critères seront utilisés pour évaluer la qualité d'une simulation déterministe (donc sans incertitude) formée de la moyenne des particules. Les autres critères nous permettront d'évaluer les incertitudes sur la simulation de l'ensemble des particules. Ils nous permettront de savoir si la méthode d'échantillonnage couvre entièrement les sources d'erreur sur les états du modèle et la dispersion de ces erreurs

## 5. RÉSULTATS ET ANALYSES

---

A titre d'illustration, nous allons présenter les résultats obtenus sur l'année 1989. Pour les autres années de 1986 à 2013, les résultats seront résumés par les critères de Nash-Sutcliffe, le MCRPS et l'erreur sur le volume de la crue. En faisant varier le nombre de particule de 20 à 500 par pas de 20 sur le bassin versant du Lac-Saint-Jean, le MCRPS nous a permis de fixé le nombre optimum de particules à 220. À cause de la fréquence d'acquisition des données et le temps de réponse du modèle, l'assimilation de données est faite chaque jour.

La Figure 5-1 présente les résultats obtenus en 1989 sur le bassin versant du Lac-Saint-Jean et le bassin versant Nechako. Pour les deux bassins versants, les moyennes des débits simulés sont très proches des débits observés (Figure 5-1-a). La Figure 5-1-b présente les différents états simulés et leur moyenne pondérée par les poids des particules. En échantillonnant directement sur les états, nous avons assuré une grande diversité de particules au cours des itérations. Ceci nous a permis de déterminer un état proche de l'état permettant d'observer le débit mesuré. On peut se rendre compte ici que sur le bassin versant du Lac-Saint-Jean le stock de neige en clairière devient nul avant le stock de neige forêt alors que, sur le bassin versant Nechako c'est l'inverse. La Figure 5-2 montre la répartition du pourcentage de forêt en fonction de l'altitude sur le Lac-Saint-Jean et Nechako. La forêt est uniformément répartie sur toutes les altitudes sur le bassin versant du Lac-Saint-Jean alors que, sur le bassin versant Nechako, le pourcentage de forêt diminue quand l'altitude augmente. Au final, la faible valeur du taux de fonte de neige en forêt par rapport au taux de fonte en clairière explique l'évolution du stock de neige en forêt et en clairière sur la Lac-Saint-Jean. Sur le bassin versant Nechako, en altitude il y a moins de forêt et plus de clairière mais, les températures sont moins élevées en altitude. Ceci explique le fait que le stock de neige en forêt devient nul avant le stock de neige en clairière. La Figure 5-1-c présente l'histogramme de rang et le diagramme de fiabilité pour les deux bassins versants. L'incertitude des états est sous-estimée, légèrement biaisé et sous-dispersé dans les deux cas. Bien que la perturbation introduite sur les états permette de bien simuler le débit, il n'est pas suffisant pour représenter les incertitudes. Au vu des résultats obtenus les autres années, les conclusions tirées pour l'analyse des résultats pour l'année 1989 sur les deux bassins versants sont valables pour les autres années et pour les deux bassins versants.

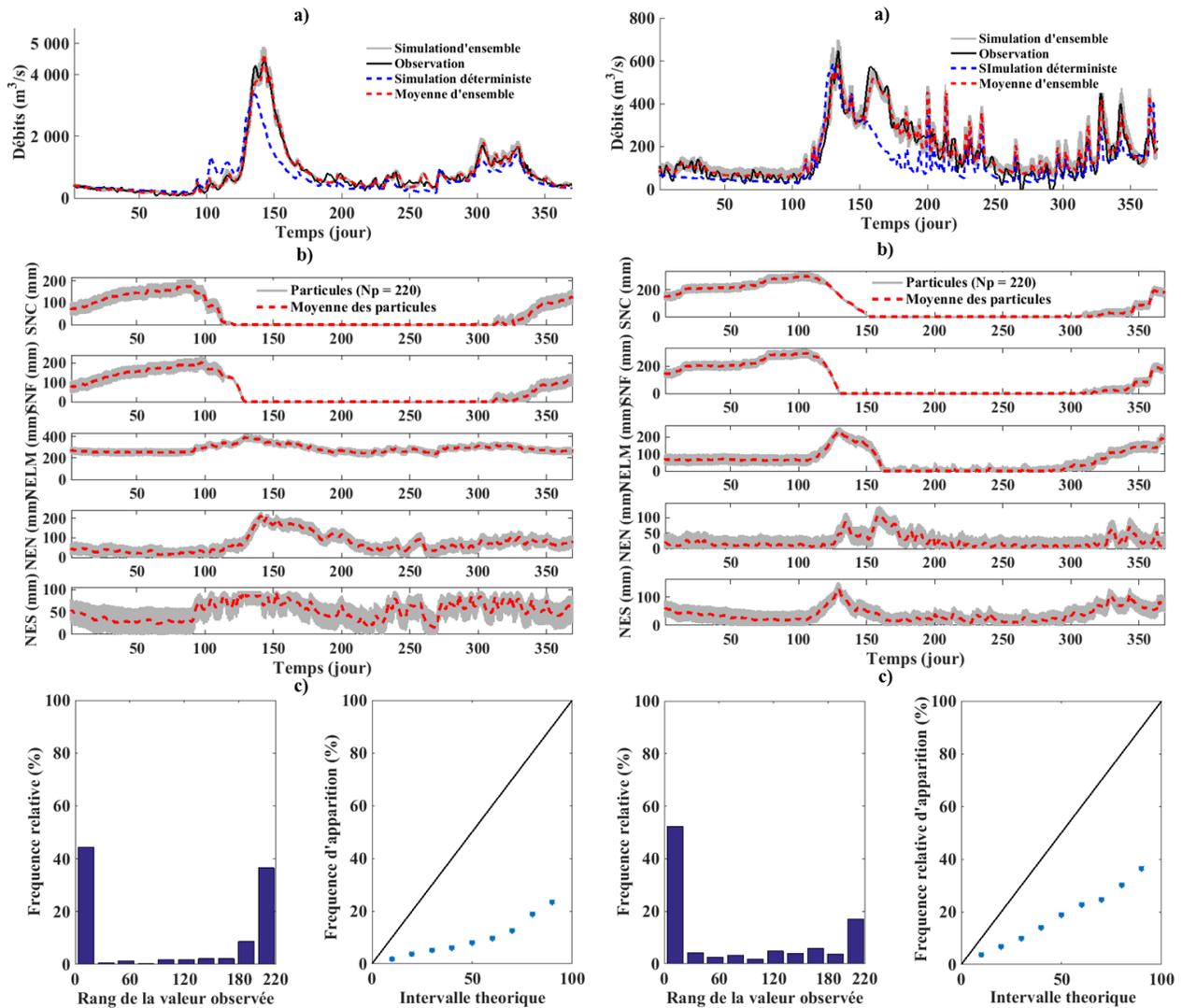


Figure 5-1 : Résultats des simulations sur les bassins versants du Lac-Saint-Jean (gauche) et Nechako (droite) en 1989 pour l'échantillonnage a priori sur les états : a) Débits observés, les simulations d'ensemble et déterministe et, la moyenne des simulations d'ensemble. b) les particules simulées et leurs moyennes pour chaque états : Stock de neige en clairière (SNC), stock de neige en forêt (SNF), Niveau d'eau dans le lac et marais (NELM), Niveau eau nappe (NEN) et niveau eau sol (NES). c) le diagramme de rang à gauche et le diagramme de fiabilité à droite.

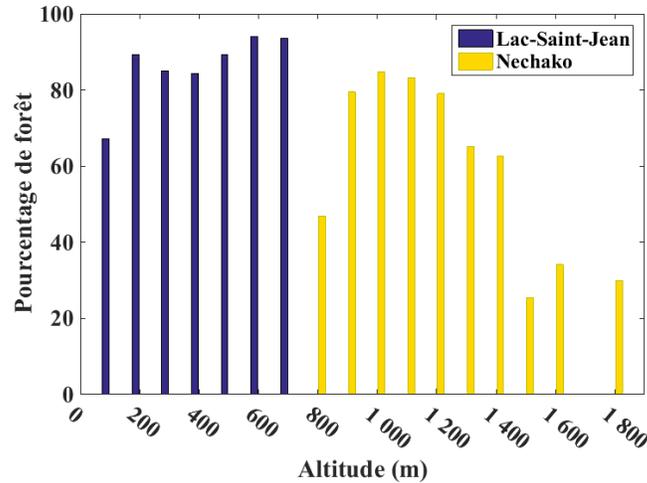


Figure 5-2: Pourcentage de couverture forestier en fonction de l'altitude sur les bassins versants du Lac-Saint-Jean et Nechako.

Pour évaluer l'effet d'un échantillonnage direct sur les bilans hydriques, nous avons comparé les bilans obtenus pour une simulation sans assimilation et une simulation avec assimilation dans le cas d'un échantillonnage direct sur les états. Sur la Figure 5-3, nous avons comparé les bilans hydriques par année avec et sans assimilation de données d'une part, et d'autre part les erreurs sur le volume de la crue sur chaque année de 1986 à 2013. Pour les deux bassins versants, la plus grande différence maximale entre les deux bilans hydriques est de 0.22mm. On peut donc conclure qu'échantillonner directement sur les états a un effet négligeable sur le bilan hydrique et ceux pour les deux bassins versants. Pour les deux bassins versants, l'erreur sur le volume de crue obtenue avec assimilation pour chaque année est très proche de zéro. Pour les 28 années, on a respectivement pour le bassin versant du Lac-Saint-Jean et Nechako une moyenne de  $-0.01 \text{ m}^3/\text{j}$  et  $-0.03 \text{ m}^3/\text{j}$  et un écart-type de 0.02. Par rapport à une simulation déterministe, on note une amélioration 8% sur la moyenne des erreurs du volume de la crue.

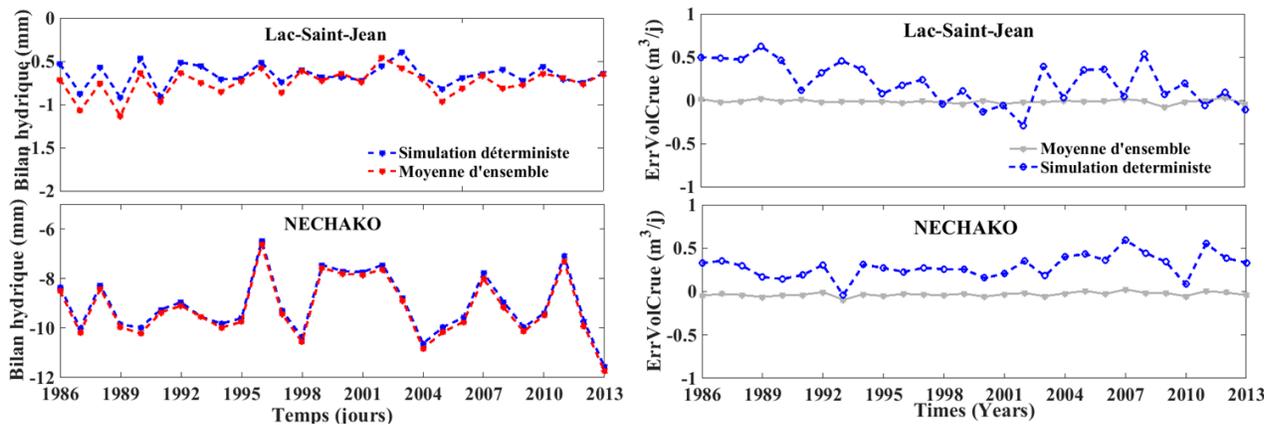


Figure 5-3: Bilan hydrique (gauche) et l'erreur sur le volume de la crue (droite) pour une simulation déterministe et la moyenne d'ensemble sur les deux bassins versants.

La Figure 5-4 présente respectivement les coefficients de Nash-Sutcliffe et les MCRPSs obtenus chaque année, ceci pour une simulation déterministe et d'ensemble. Sur la même figure, on compare les simulations avec échantillonnage direct sur les états et ceux par perturbation des observations météorologiques. Sur le bassin versant Nechako, lorsque l'incertitude sur les observations météorologiques augmente. Ceci est moins marqué sur le bassin versant Lac-Saint-Jean. Ceci peut être dû à la différence en termes de pourcentage de couverture de forêt et d'altitude entre les deux bassins versants. Sur les deux bassins versants, les valeurs de coefficient de Nash-Sutcliffe sont pour toutes les années très proches de 1 pour un échantillonnage direct sur les états et, supérieur aux valeurs de Nash obtenus par perturbation des observations météorologiques quel que soit l'incertitude appliquée sur les observations météorologiques. Le coefficient de Nash-Sutcliffe nous permet de dire que, dans le cadre d'une analyse déterministe, on améliore la qualité de simulation. Cette amélioration est plus marquée pour un échantillonnage direct sur les états. Sur les 28 années, en moyenne, le coefficient de Nash-Sutcliffe passe de  $0.87 \pm 0.06$  pour une simulation sans assimilation de données à  $0.96 \pm 0.02$  pour une simulation avec assimilation de données et échantillonnage direct sur les états pour le Lac-Saint-Jean et, de  $0.22 \pm 0.28$  à  $0.89 \pm 0.06$  pour Nechako.

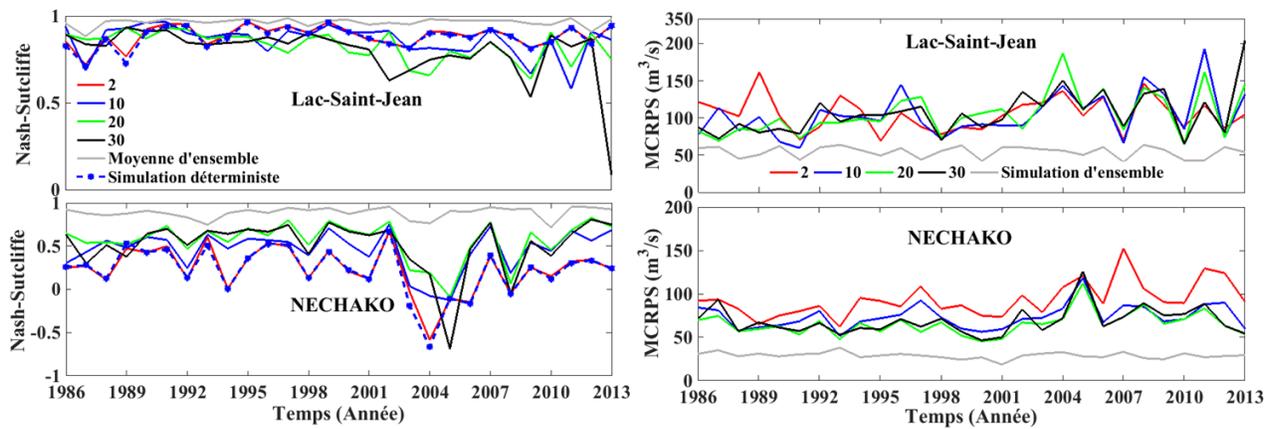


Figure 5-4: Coefficient de Nash-Sutcliffe (gauche) et MCRPS (droite) évalué sur la simulation déterministe, les moyennes d'ensemble pour un échantillonnage direct sur les états et pour chaque valeur d'incertitude quand on échantillonne par perturbations des observations météorologiques. La valeur 2 correspond à une perturbation suivant une loi uniforme de paramètres -2 et 2.

Contrairement au coefficient de Nash-Sutcliffe, l'augmentation de l'incertitude sur les observations météorologiques entraîne une baisse du MCRPS. Ceci montre une amélioration en termes de simulation d'ensemble pour un échantillonnage par perturbation des observations météorologiques. Mais ceci n'est pas suffisant pour égaler les simulations faites par échantillonnage direct sur les états. Les valeurs de MCRPSs obtenues pour un échantillonnage direct sur les états sont largement inférieures aux valeurs de MCRPSs des valeurs obtenus pour un échantillonnage par perturbation des observations météorologiques, quelles que soit l'ampleur de la perturbation appliquée sur les observations météorologiques. La grande différence entre les valeurs de MCRPS entre un échantillonnage direct sur les états et l'échantillonnage par perturbation des observations météorologiques permettent de trancher en faveur de l'échantillonnage direct sur les états. C'est la raison pour laquelle nous avons porté une attention particulière à un échantillonnage direct sur les états.

## 6. COMCLUSIONS ET PERSPECTIVES

---

L'objectif principal de cette étude était d'améliorer les simulations de débits par le modèle CEQUEAU en y implémentant un algorithme d'assimilation de données utilisant le principe de filtrage particulaire. Une revue de la littérature nous a permis d'une part de nous rendre compte de la diversité d'algorithme de filtrage particulaire, d'autre part de soulever les points d'ombre sur l'utilisation de filtrage particulaire. Exemple le principe d'échantillonnage, le choix de la fonction d'importance. Comme préconisé dans la littérature, en méconnaissance de la distribution d'importance, nous avons supposé que la distribution d'importance est égale à la loi *a priori*  $P(x_t^i/x_{t-1}^i)$ . Faire de la sorte reviendrait à dire que la distribution d'importance ne dépend pas des observations. Nous avons par ailleurs supposé que l'erreur sur les observations suit une loi normale et avons définis une fonction qui assure l'hétéroscédasticité de l'écart-type. Pour résoudre le problème de dégénérescence lié à l'utilisation des filtres particulaires plusieurs outils sont proposés dans la littérature telle que le SIR, RPF et APF. Ces algorithmes ont été testés et les résultats présentés en annexe montrent ici que le SIR, l'APF et RPF ne sont pas les outils adéquats pour l'AD dans le modèle CEQUEAU. Nous avons définis un nouvel algorithme inspiré des algorithmes précédemment cités.

En échantillonnant directement sur les états, il y a un risque sur la non-fermeture du bilan hydrique. Nous avons montré par comparaison des bilans hydriques sans assimilation et avec assimilation que, l'échantillonnage direct sur les états a un effet négligeable sur le bilan hydrique. En termes de simulation déterministe et d'ensemble, échantillonner directement sur les états donne de meilleurs résultats; car sur toutes les années d'assimilation, le coefficient de Nash-Sutcliffe moyen passe de 0.87 sans assimilation à 0.96 avec assimilation sur le bassin versant Lac-Saint-Jean et de 0.22 sans assimilation à 0.88 avec assimilation sur le bassin versant Nechako. Sur les deux bassins versants, l'erreur sur le volume de la crue passe en moyenne sur les 28 années de  $0.2 \text{ m}^3/\text{j}$  (écart-type de 0.2) à  $-0.03 \text{ m}^3/\text{j}$  (écart-type de 0.02) et de  $0.3 \text{ m}^3/\text{j}$  (écart-type de 0.1) à  $-0.03 \text{ m}^3/\text{j}$  (écart-type de 0.02) respectivement sur les bassins versants de du Lac-Saint-Jean et Nechako. On montre aussi que le MCRPS passe de  $158 \text{ m}^3/\text{s}$  à  $81 \text{ m}^3/\text{s}$  et de  $93 \text{ m}^3/\text{s}$  à  $29 \text{ m}^3/\text{s}$  respectivement pour les bassins versants du Lac-Saint-Jean et Nechako.

Ces résultats montrent que l'assimilation de donnée par filtrage particulière est un outil efficace pour améliorer la qualité de simulation de débit dans la modèle CEQUEAU. Cet outil est d'autant meilleur lorsque l'échantillonnage des particules est fait directement sur les états. Bien que pour les bassins versants utilisés cette méthode d'échantillonnage n'a pas d'effet sur le bilan hydrique, il est important de vérifier après assimilation si la conservation du bilan hydrique est maintenue. À cause des températures négatives ou inférieures aux températures de fonte, on peut supposer que pendant la période hivernale le stock de neige au sol n'est pas important pour l'assimilation de données. Pour confirmer cette hypothèse, il est important d'étudier l'influence des états sur une année et pour différent saison de l'année. Ceci pourra nous permettre de réduire le nombre de nos états et indirectement le temps de calcul.

## 7. RÉFÉRENCES

---

- Abaza, M., Anctil, F., Fortin, V., & Turcotte, R. (2015). Exploration of sequential streamflow assimilation in snow dominated watersheds. *Advances in Water Resources*, 80, 79-89.
- Ades, M., & Van Leeuwen, P. J. (2015). The equivalent-weighted particle filter in high-dimensional system. *Quarterly Journal of the Royal Meteorology Society*, 485-503.
- Clark, M. p., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., & Leavesley, G. H. (2006). Assimilation of snow covered area information into hydrologic and land-surface models. *Advances in Water Resources*, 1209-1221.
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of geophysical research*, 99, 10143-10162.
- Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation. *Ocean dynamics*, 53, 343-367.
- Leisenring, M., & Moradkhani, H. (2011). Snow water equivalent prediction using Bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*, 253-270.
- Montzka, C., Moradkhani, H., Weihermüller, L., Hendricks Franssen, H.-J., Canty, M., & Vereecken, H. (2011). Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. *Journal of Hydrology*, 410-421.
- Moradkhani, H., & Hsu, K.-L. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, 41(5).
- Morin, G., & Paquet, p. (2007). *Modèle hydrologique CEQUEAU*. Rapport de recherche, INRS-ETE, Université du Québec, INRS-Eau, Terre et Environnement.
- Noh, S. J., Tachikawa, Y., & Kim, S. (2011). Applying sequential monte carlo methods into a distributed hydrologic model: lagged particle filtering approach with regularization. *Hydrology and Earth System Sciences*, 15, 3237-3251.
- Papadakis, N., Mémin, E., Cuzol, A., & Gengembre, N. (2010). Data assimilation with the weighted ensemble Kalman filter. *Tellus A*, 673-697.
- Plaza, D. A., De Keyser, R., De Lannoy, G. J., Giustarini, L., Matgen, P., & Pauwels, V. R. (2012). The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. *Hydrology and Earth system sciences*, 16, 375-390.
- Reichle, R. H. (2008). Data assimilation methods in Earth sciences. *Advances in Water Resources*, 1411-1418.
- Ristic, B., Arulampalam, S., & Gordon, N. (2004). Beyond the Kalman Filter: Particle Filters for Tracking Applications. *Artech House*.

- Salamon, P., & Feyen, L. (2009). Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of Hydrology*, 428–442.
- Sanjeev Arulampalam, M., Maskell, S., Gordon, N., & Clapp, T. (2002). A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *Signal Processing, IEEE Transactions*, 50(2), 174-188.
- Seo, D.-J., Cajina, L., Corby, R., & Howison, T. (2009). Automatic updating for operational streamflow forecasting via variational data assimilation. *Journal of Hydrology*, 255-275.
- Slatter, A. G., & Clark, M. P. (2005). Snow Data Assimilation via an Ensemble Kalman Filter. *Journal of Hydrometeorology*, 7(3), 478-493.
- Talagrand, O., & Courtier, P. (1987). Variational assimilation of meteorological observation with the adjoint vorticity equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, 113(478), 1311-1328.
- Turcotte, R., Fortier Fillion, T. C., Lacombe, P., Fortin, V., Roy, A., & Royer, A. (2010). Simulation hydrologique des derniers jours de la crue de printemps: le problème de la neige manquante. *Hydrological Sciences Journal—Journal des Sciences Hydrologiques*, 55(6), 872-882.
- Van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Mathematical advances in data assimilation*, 4089-4114.

## A. ASSIMILATION INDIRECTE: PERTURBATION DES DONNÉES MÉTÉOROLOGIQUES

---

Ici, nous allons présenter les résultats obtenus en utilisant l'échantillonnage par perturbation des observations météorologiques. Cela consiste à modifier la précipitation et la température pour créer des états. Nous avons échantillonné suivant une loi uniforme une suite de 220 variables aléatoires de précipitation et de température avec des perturbations suivant une loi uniforme de paramètres suivantes :  $[-2,2]$ ,  $[-10,10]$ ,  $[-20,20]$  et  $[-30,30]$ . Chaque couple de variable (température et précipitation) est utilisé comme une donnée d'entrée au modèle CEQUEAU pour simuler les états. Nous avons ainsi pour les 220 couples de données météorologiques, simulées 220 jeux d'états. En utilisant l'algorithme d'assimilation de données présentées dans ce document, nous avons ainsi évalué l'effet des perturbations appliquées sur les données météorologiques sur la simulation d'ensemble des débits pour les bassins versants du Lac-Saint-Jean et Nechako. Pour illustrer nos résultats, nous allons présenter les résultats obtenus en 1989 pour une perturbation qui suit une loi uniforme de paramètres  $[-2,2]$  et  $[-30,30]$  mm (ou °C) sur les observations météorologiques. Les autres résultats seront résumés par le MCRPS et le coefficient de Nash-Sutcliffe.

La figure A1 présente les résultats obtenus en 1989 sur les bassins versants du Lac-Saint-Jean et Nechako avec une perturbation qui suit une loi uniforme de paramètres  $-2$  et  $2$  mm(ou °C) sur les observations météorologiques. Les ensembles simulés et leur moyenne sont très proches de la simulation déterministe pour les deux bassins versants (figure A1-a). Il en est de même pour les particules et leur moyenne (figure A1-b). L'incertitude appliquée sur les observations météorologiques n'a pas un effet significatif ni sur les débits simulés, ni sur les états. En début d'année, la surface est gelée et les températures sont négatives. Donc, une légère variation de la température ne peut entraîner ni la fonte de neige, ni l'écoulement de l'eau dans le sol et la nappe. C'est la raison pour laquelle on n'observe pas de variation des états. Avec une légère augmentation du coefficient de Nash-sutcliffe et une légère diminution de MCRPS on peut dire que pour cette année on a une légère amélioration du débit simulé. La figure A1-c présente les histogrammes de rang et les diagrammes de fiabilité pour les deux bassins versants. Les

distributions sont biaisées et l'incertitude est sous-estimée. À cause de la différence entre les deux bassins versants, la sous-dispersion et la sous-estimation des incertitudes sont moins marquées sur le bassin versant Nechako.

Pour améliorer l'échantillonnage des états et qualité de simulation, on suppose qu'il faudrait augmenter le niveau d'eau au sol, dans la nappe et dans les lacs et marais. Au jour 1 (1<sup>er</sup> janvier), une augmentation considérable de la fonte de neige, ne peut se faire qu'en supposant une température moyenne supérieure à la température de fonte. La Figure A2 présente les résultats obtenus avec une perturbation qui suit une loi uniforme de paramètres -30 et 30 mm(ou°C) sur les observations météorologiques. Bien que ce cas ne soit pas réaliste, il entraîne de variation plus importante des débits simulés avec assimilation en début de simulation (figure A2-a). Une telle modification permet d'obtenir une plage plus importante des variations des états assimilées (Figure A2-b). Autour du 140<sup>ième</sup> jour, on note une augmentation brusque des débits simulés par AD. Ceci est due aux particules simulées par une température négative au nulle et une précipitation élevée entraînant une accumulation du stock de neige au sol et en forêt. Car, quand la température observée est supérieure à la température de fonte, le surplus de neige accumulé fond. Quant aux grandes variations aussi bien des débits simulés que des états (particules), elles sont dues aux particules simulées avec les températures supérieures à la température de fonte. La figure A2-c présente les histogrammes de rang et les diagrammes de fiabilité pour les deux bassins versants. Comme dans le cas précédent, on peut conclure qu'il y a une sous-estimation et une sous-dispersion des erreurs sur les états du modèle. Néanmoins, elles sont légèrement moins marquées que dans le cas précédent et moins encore sur le bassin Nechako.

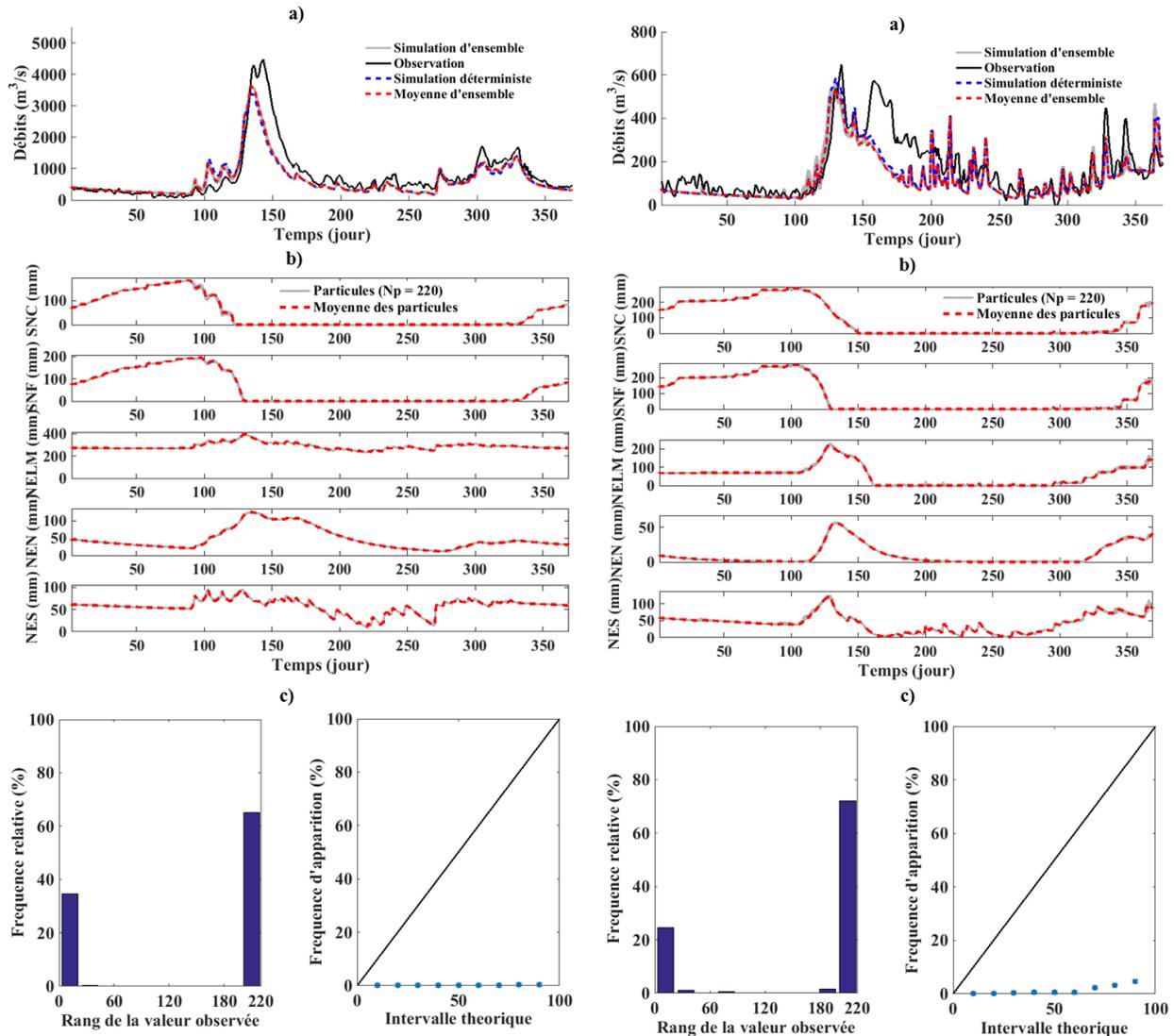


Figure A-1 : Résultats des simulations sur les bassins versants du Lac-Saint-Jean (gauche) et Nechako (droite) en 1989 avec une perturbation qui suit une loi uniforme de paramètres -2 et 2 mm (ou°C), pour l'échantillonnage sur l'observation météorologique : a) Débits observés, les simulations d'ensemble et déterministe et, la moyenne des simulations d'ensemble. b) les particules simulées et leurs moyennes pour chaque états : Stock de neige en clairière (SNC), Stock de neige en forêt (SNF), Niveau eau lac et marais (NELM), Niveau eau nappe (NEN) et niveau eau sol (NES). c) le diagramme de rang à gauche et le diagramme de fiabilité à droite.

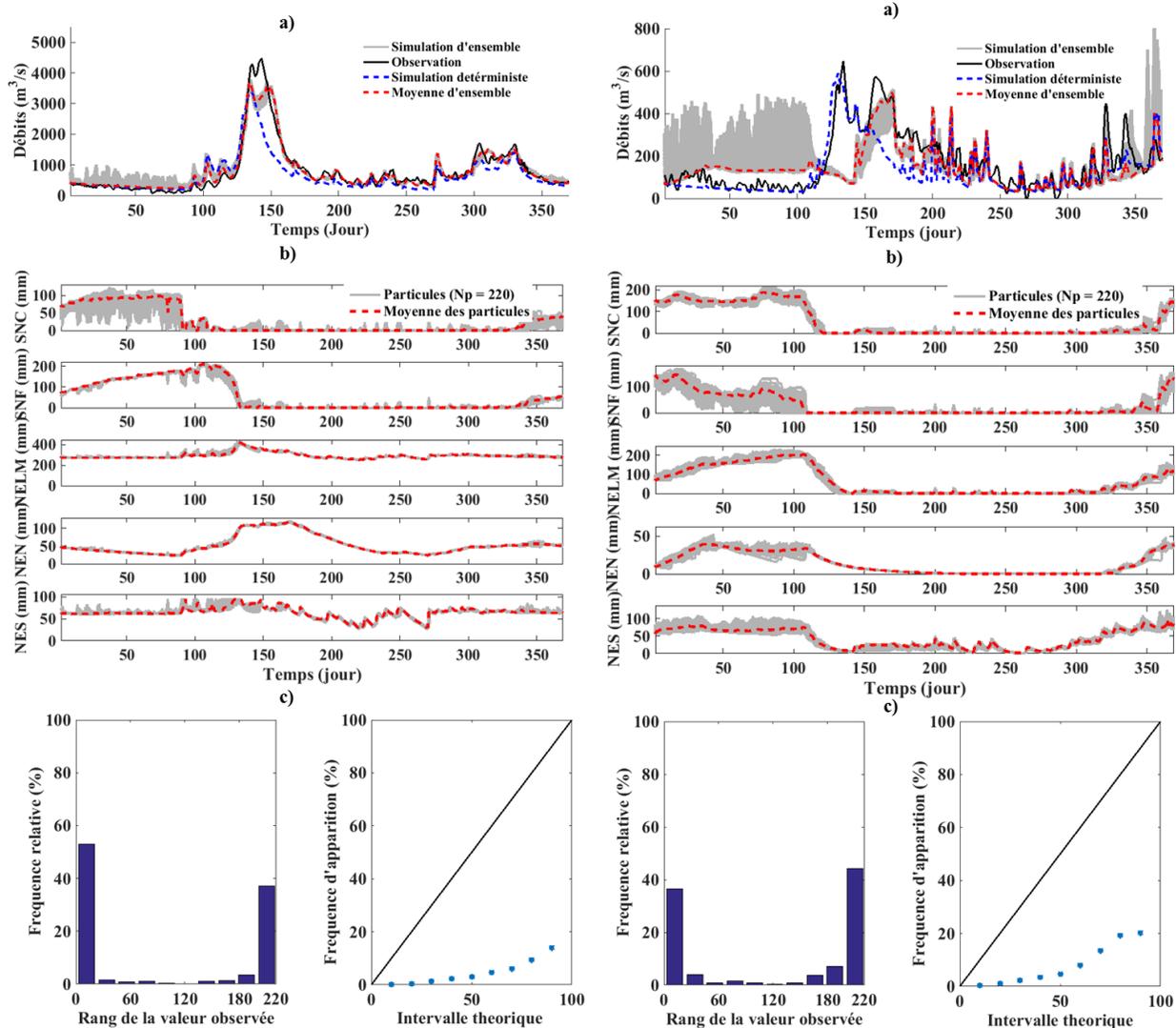


Figure A-2 : Résultats des simulations sur les bassins versants du Lac-Saint-Jean (gauche) et Nechako (droite) en 1989 avec une perturbation qui suit une loi uniforme de paramètres-30 et 30 mm (ou°C), pour l'échantillonnage sur l'observation météorologique : a) Débits observés, les simulations d'ensemble et déterministe et, la moyenne des simulations d'ensemble. b) les particules simulées et leurs moyennes pour chaque états : Stock de neige en clairière (SNC), stock de neige en forêt (SNF), niveau eau lac et marais (NELM), niveau eau nappe (NEN) et niveau eau sol (NES). c) le diagramme de rang (gauche) et le diagramme de fiabilité (droite).

Sur la figure A3, nous avons tracé les MCRPSs obtenus pour chaque perturbation sur les observations météorologiques et celui obtenu avec un échantillonnage faite avec une connaissance a priori des états ceci, pour les deux bassin versant de 1986 à 2013. Lorsqu'on augmente l'incertitude sur les observations météorologiques, la qualité de simulation augmente sur le bassin versant Nechako car le MCPRS diminue. Ce n'est pas le cas sur le bassin versant du

Lac-Saint-Jean. Le niveau de perturbations qui donne la plus petite valeur médian de MCRPS est, 10 pour le Lac-Saint-Jean et de 20 pour le bassin versant Nechako. Le MCRPS obtenu pour l'échantillonnage avec une connaissance a priori sur les états est nettement inférieur aux MCRPSs pour l'échantillonnage à partir des observations météorologiques sur les deux bassins versants.

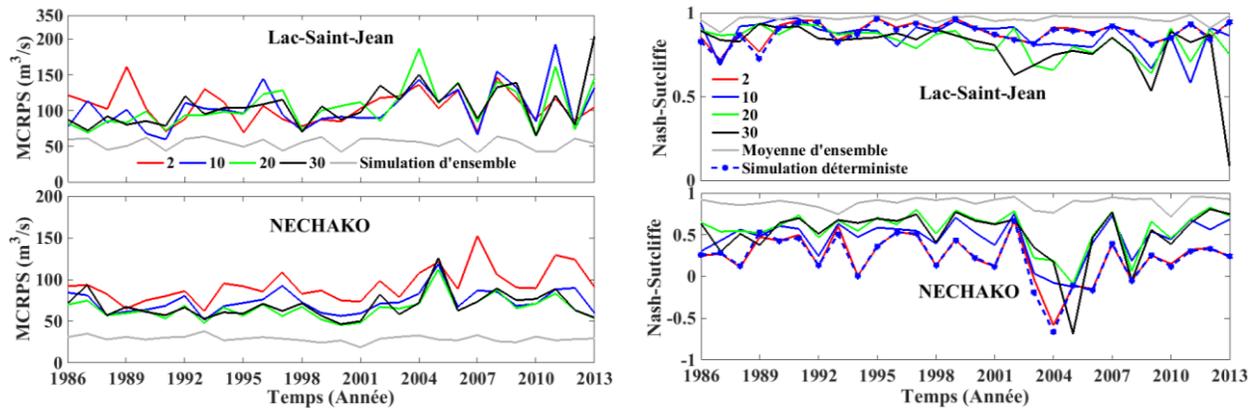


Figure A-3:Évolution du MCRPS (gauche)et du coefficient de Nash-Sutcliffe (droite) de la simulation déterministe, de la simulation en fonction de l'incertitude appliquée sur les données météorologiques et de la simulation obtenu avec un échantillonnage direct.

Le coefficient de Nash confirme une amélioration de la qualité de simulation en fonction de l'incertitude sur les observations météorologiques sur le bassin versant Nechako. En comparaison avec la simulation déterministe, l'assimilation des données permet d'améliorer la qualité de simulation. Cette amélioration est plus marquée lorsque l'échantillonnage est fait directement sur les états. Un échantillonnage par perturbation des observations météorologiques de  $\pm 2$  n'a aucun effet sur les états et les débits simulés. On montre que les coefficients de Nash-Sutcliffe sont identiques. Lorsqu'on augmente l'incertitude sur les observations météorologiques, on améliore la qualité de simulation sur le bassin versant Nechako. Ceci n'est pas toujours le cas sur le bassin versant du Lac-Saint-Jean

## B. UTILISATION DU SEQUENTION IMPORTANCE RESAMPLING (SIR)

---

Pour appliquer le SIR, nous avons utilisé l'algorithme 3 en remplaçant le MASIR par l'algorithme 1 (SIS) et nous avons ajouté l'étape supplémentaire du RESAMPLING définie par l'algorithme 2. L'échantillonnage est fait directement sur les états. Nous allons vérifier l'effet de la modification des états sur le bilan hydrique. Cette approche a été testée sur le bassin versant du Lac Saint Jean et du Nechako avec les observations météorologiques obtenu en 1989.

La Figure B1 présente les résultats obtenus sur le bassin versant du Lac-Saint-Jean et de Nechako en 1989 pour un échantillonnage direct sur les états avec le 'Sequential Importance Resampling'. On note une légère différence entre l'ensemble débits simulés avec AD et sans AD (figure B1-a). Cette différence est plus prononcée pendant la période de dégel. Passé cette période, les débits simulés avec et sans assimilation sont presque identique. Contrairement à l'échantillonnage par perturbation des observations météorologiques, ici on a plus d'échantillon différent à l'instant initial. À l'étape de ré-échantillonnage, le SIR fait qu'on remplace les particules les moins plausibles par les particules les plus plausibles. Alors, au fils des itérations on se retrouve avec des particules peut différentes les unes des autres (Figure B1-b).Cela entraine également la convergence des débits simulés avec AD par chaque particule vers le débit simulé sans AD.

On peut dire ici que la modification des états au cours de l'échantillonnage n'a pas d'effet notable sur le bilan hydrique, car la différence entre les bilans hydriques avec et sans AD est de 2%.Pour la même année et dans les mêmes conditions, Le coefficient de Nash augmente légèrement. Quant au MCRPS, il reste inférieur au MCRPS obtenu par la méthode proposée sur les deux bassins versants. L'histogramme de rang et le diagramme de fiabilité (figure B3-c) permettent de dire que les sources d'erreur liée aux états sont sous-estimées. À l'aide tous ces outils on peut clairement dire qu'il est nécessaire d'introduire dans l'algorithme une étape supplémentaire après le ré-échantillonnage.

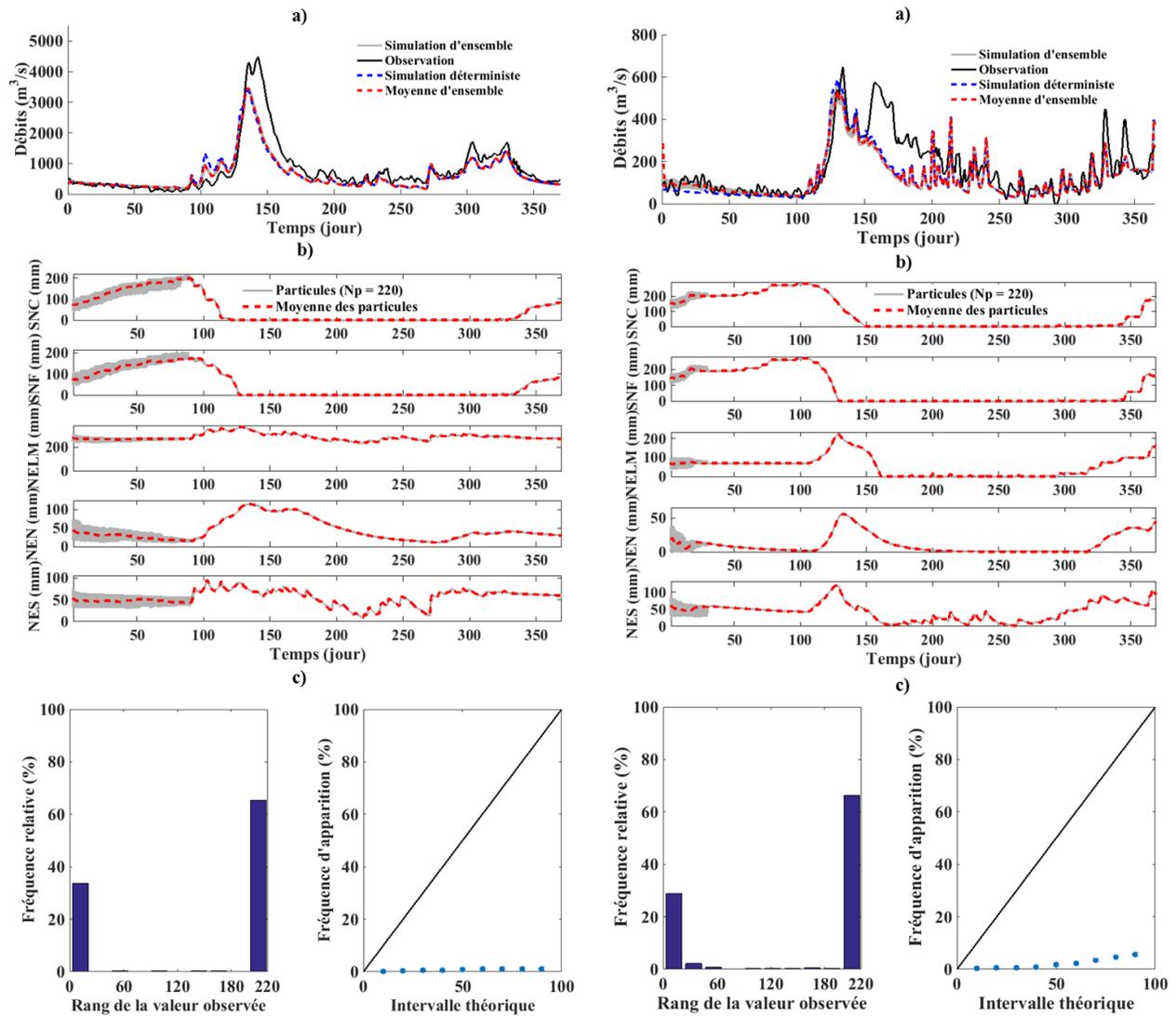


Figure B-1: Résultats des simulations avec le SIR sur les bassins versants du Lac-Saint-Jean (gauche) et Nechako (droite) en 1989 pour l'échantillonnage direct sur les états : a) Débits observés, les simulations d'ensemble et déterministe et, la moyenne des simulations d'ensemble. b) les particules simulées et leurs moyennes pour chaque états : Stock de neige en clairière (SNC), stock de neige en forêt (SNF), niveau eau lac et marais (NELM), niveau eau nappe (NEN) et niveau eau sol (NES). c) le diagramme de rang (gauche) et le diagramme de fiabilité (droite).

## C. UTILISATION DU REGULARIZED PARTICLE FILTER (RPF)

---

Dans ce cas, l'échantillonnage va être fait directement sur les états. Ceci étant, nous allons chaque fois vérifier l'effet de la modification des états sur le bilan hydrique. Pour tester l'utilisation du RPF, nous avons utilisé l'algorithme général présenté par l'algorithme 3. À l'étape de ré-échantillonnage des particules, nous avons utilisé la méthode de noyau. Ceci assure une diversité de particules au cours des itérations et, permet de réduire le problème de dégénérescence en ajoutant un bruit sur les particules après le RESAMPLING. Le principe est le suivant : à chaque instant, on calcul une matrice de corrélation entre les états et les poids de chaque particule et on tire un bruit aléatoire  $\varepsilon^i$  à partir du noyau estimer sur tous les poids. Ici, le noyau est la loi de probabilité estimée sur les poids  $\omega_t$ . Les nouveaux états ré-échantillonnés sont :

$$x_t^{i*} = x_t^i + h_{opt} D_t \varepsilon^i \quad (C1)$$

Avec  $h_{opt} = A n^{-\frac{1}{n_x+4}}$  avec  $A = \left[ 8 c_{n_x}^{-1} (n_x + 4) (2\sqrt{\pi})^{n_x} \right]^{\frac{1}{n_x+4}}$ .  $n_x$  est la dimension du vecteur d'état,  $c_{n_x}$  est le volume de la sphère de rayon unitaire  $R^{n_x}$ ,  $D_t$  est la racine carré de la matrice de corrélation entre les états et les poids de chaque particule,  $\varepsilon^i$  est un bruit aléatoire obtenu à partir du noyau.

La figure C1 présente les résultats obtenus sur le bassin versant du Lac-Saint-Jean et de Nechako en 1989 pour un échantillonnage direct sur les états avec le RPF. Sur les deux bassins versant, les résultats sont similaires. On note une légère différence entre les débits simulés avec AD et le débit simulé sans AD (figure C1-a). Celle-ci est plus marquée pendant la période de fonte. Si on s'intéresse à l'évolution des états, on note une convergence de ces derniers vers un état ou un groupe d'état peut différent (figure C1-b). Ceci est dû à la méthode de ré-échantillonnage car, les valeurs  $h_{opt} D_t \varepsilon^i$  qui sont ajoutés aux états sont très faibles devant la valeur des états. Cette méthode n'est pas efficace dans ce cas, car les bruits liés aux états ne sont pas dans cet ordre de grandeur. Néanmoins, on note une amélioration du coefficient de NASH. Le MCRPS obtenue ici

est nettement supérieur au MCRPS obtenu par la méthode proposée. L'histogramme de rang et diagramme de fiabilité (figure C1-c) permettent de dire que les sources d'erreur liées aux états sont sous-estimées. Cette approche est pertinente quand le bruit des états est faible. Cette remarque rejoint le résultat de M. Sanjeev et al. (2002) qui préconise l'utilisation de cette méthode pour les états ayant de faible bruit.

L'échantillonnage direct ne pourrait pas avoir d'effet sur le bilan hydrique car, la différence entre les bilans hydriques avec et sans AD est de 3%. Il est nécessaire de préciser que la vérification du bilan hydrique n'est pas exhaustive. La comparaison entre les hydrogrammes observés et simulés, nous a permis de conclure que la méthode RPF n'est pas adaptée pour notre étude.

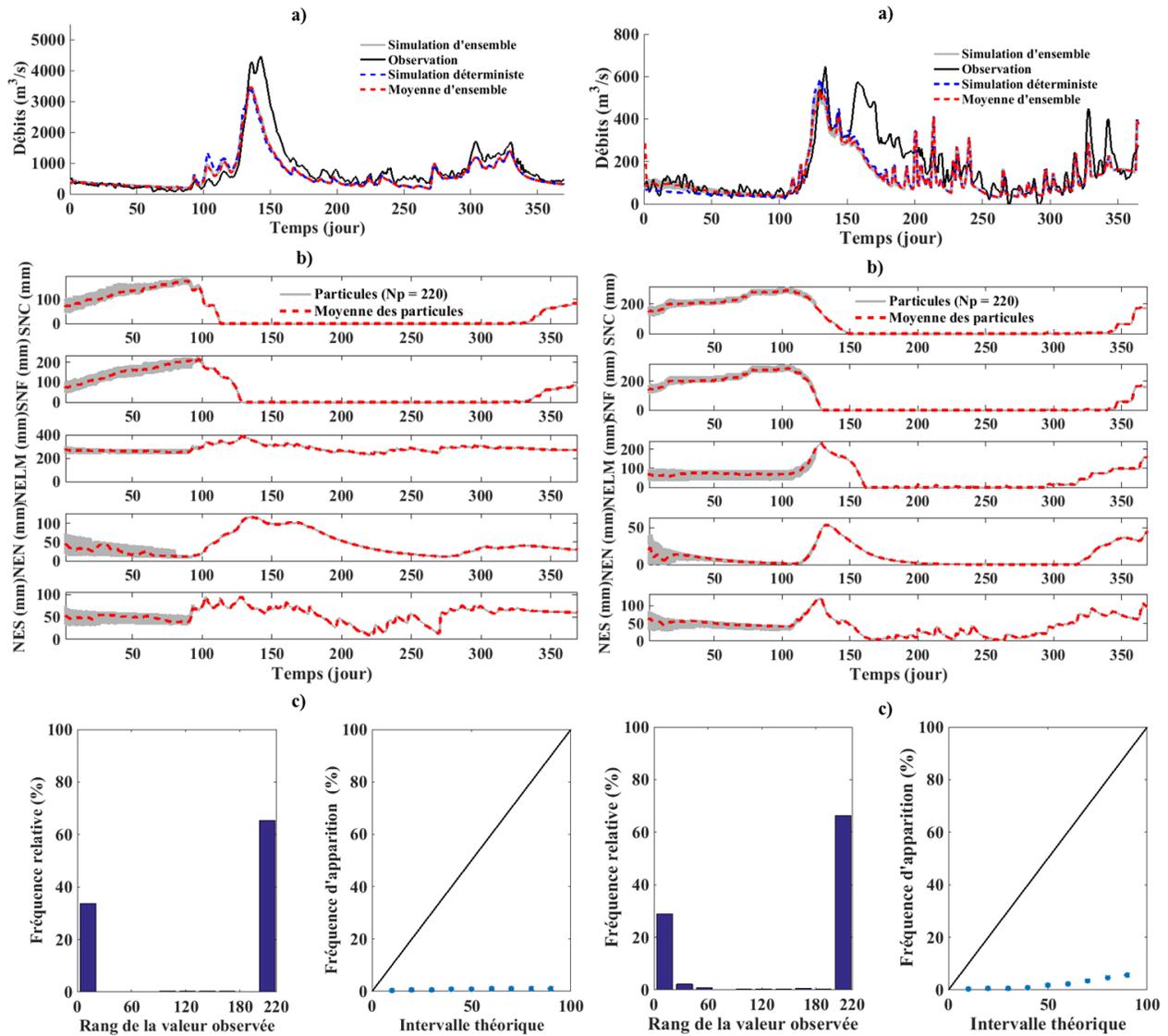


Figure C-1 : Résultats des simulations avec le RPF sur les bassins versants du Lac-Saint-Jean (gauche) et Nechako (droite) en 1989 pour l'échantillonnage l'a priori sur les états : a) Débits observés, les simulations d'ensemble et déterministe et, la moyenne des simulations d'ensemble. b) les particules simulées et leurs moyennes pour chaque états : Stock de neige en clairière (SNC), stock de neige en forêt (SNF), Niveau eau lac et marais (NELM), Niveau eau nappe (NEN) et niveau eau sol (NES). c) le diagramme de rang à gauche et le diagramme de fiabilité à droite.

## D. ALGORITHMES

---

Algorithme 1: SIS

$$[\{x_t^i, w_t^i\}_{i=1}^N] = SIS[\{x_{t-1}^i, w_{t-1}^i\}_{i=1}^N, y_t]$$

- POUR  $i=1:N$ 
  - Tirer  $x_t^i \sim q(x_t/x_{t-1}^i, y_t)$
  - Calcul du poids de chaque particule  $w_t^i$  (equation 9)
- FIN POUR

---

Algorithme 2: Resampling

$$[\{x_t^i, w_t^i, Ind^i\}_{i=1}^{N_p}] = \text{RESAMPLE}[\{x_t^i, w_t^i\}_{i=1}^{N_p}]$$

$$c_1 = 0$$

- POUR  $i = 2 : N_p$ 
  - $c_i = c_{i-1} + w_t^i$
- FIN POUR
- $i = 1; u_1 \sim U[0, N_p^{-1}]$
- POUR  $j = 2 : N_p$ 
  - $u_j = u_1 + N_p^{-1}(j-1)$
  - TANT QUE  $u_j > c_i$ 
    - $i = i + 1$
  - FIN TANT QUE
  - $x_t^{j*} = x_t^i, w_t^j = N_p^{-1}, Ind^j = i$
- FIN POUR

---

Algorithme 3: Algorithme général

- Initialisation:  $\mu_1 = x_1$  et  $y_1^\mu = h(\mu_1)$
- Échantillonner  $N_p$  particules
- $\{w_1^i = N_p^{-1}\}_{i=1}^{N_p}, \{y_{s,1}^i = h(x_1^i)\}_{i=1}^{N_p}$

- POUR  $t = 1 : T$ 
  - $\left[ \{x_{t+1}^i, w_{t+1}^i, y_{s,t+1}^i\}_{i=1}^{N_p}, \mu_{t+1} \right] = \text{MASIR} \left[ \{x_t^i, w_t^i\}_{i=1}^{N_p}, \mu_t, y_t \right]$
  - $y_{t+1}^\mu = h(\mu_{t+1})$
- FIN POUR

Algorithme 4: ASIR Modifié (MASIR)

$$\left[ \{x_{t+1}^i, w_{t+1}^i, y_{s,t+1}^i\}_{i=1}^{N_p}, \mu_{t+1} \right] = \text{MASIR} \left[ \{x_t^i, w_t^i\}_{i=1}^{N_p}, \mu_t, y_t \right]$$

- POUR  $i = 1 : N_p$ 
  - Évaluer  $x_{t+1}^i$  et  $y_{s,t+1}^i$
  - $w_{t+1}^i \propto w_t^i p(y_t / x_t^i)$
- FIN POUR
- $sw = \sum_{i=1}^{N_p} w_{t+1}^i$
- POUR  $i = 1 : N_p$ 
  - $w_{t+1}^i = sw^{-1} w_{t+1}^i$
- FIN POUR
- $N_{eff} = \left( \sum_{i=1}^{N_p} (w_t^i)^2 \right)^{-1}$
- SI  $N_{eff} \leq N_p$ 
  - $\left[ \{x_t^i, w_t^i, Ind^i\}_{i=1}^{N_p} \right] = \text{RESAMPLE} \left[ \{x_t^i, w_t^i\}_{i=1}^{N_p} \right]$
  - $\mu_{t+1} = \sum_{i=1}^{N_p} x_t^i w_t^i$
  - Retirer les particules dupliquées:  $N_r = N_p - N_d$  (où  $N_r$  est le nombre de particules restant et  $N_d$  nombre de particules dupliquées)
  - Échantillonner  $N_r$  particules
  - $\left\{ w_t^i = N_p^{-1} \right\}_{i=1}^{N_p}$
- SINON
  - $\mu_{t+1} = \sum_{i=1}^{N_p} x_t^i w_t^i$

- FIN SI