

Évaluation de l'utilisation des données des réanalyses ERA5 et ERA5-Land pour la modélisation de l'évapotranspiration

par

Guillaume LEVESQUE

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA MAÎTRISE
AVEC MÉMOIRE EN GÉNIE DE LA CONSTRUCTION
M. Sc. A.

MONTRÉAL, LE 18 AOÛT 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Guillaume Levesque, 2022



Cette licence Creative Commons signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette œuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'œuvre n'ait pas été modifié.

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

Mme Annie Poulin, directrice de mémoire
Département du génie de la construction à l'École de technologie supérieure

M. Richard Arsenault, codirecteur de mémoire,
Département du génie de la construction à l'École de technologie supérieure

M. Jean-Luc Martel, président du jury
Département du génie de la construction à l'École de technologie supérieure

Mme Janie Masse-Dufresne, examinatrice interne
Département du génie de la construction à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 5 AOÛT 2022

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens à remercier l'ensemble des personnes m'aillant offert conseils et soutient dans le processus de cette maîtrise. Principalement à ma directrice et mon codirecteur de mémoire ainsi que mes coloc.

Merci, Annie et Richard, d'avoir été là pour répondre à mes questions et surtout pour m'avoir proposé les projets extra curriculum à mon mémoire. Je désirais faire mon mémoire en grande partie pour approfondir mes connaissances en programmation et vivre différentes expériences. Je crois que vous avez su voir cela en moi et les projets proposés comme le petit périple au Mexique et le téléchargement de données ont bien complémenté mon temps passé à travailler sur ma recherche. Je veux souligner aussi votre force à bien cerner les élèves. J'ai adoré que vous me laissiez la plus grande latitude et que vous vous offriez surtout comme support et conseiller en lieu de dirigeant.

À mes coloc, beaucoup de discussions abordées en fin de soirée sur mon sujet, sur mes techniques, sur les processus de traitements de données, sur mes buts, sur ma motivation, merci. Fred et Chris, votre expérience et votre claire vision par rapport au processus de maîtrise m'ont permis de réaligner mes buts et de faire apparaître la lumière au bout du tunnel.

Évaluation de l'utilisation des données de réanalyses ERA5 et ERA5-Land pour la modélisation de l'évapotranspiration

Guillaume LEVESQUE

RÉSUMÉ

Cette étude se penche sur la valeur ajoutée de l'utilisation de données des réanalyses ERA5 et ERA5-Land pour l'estimation de l'évapotranspiration en modélisation hydrologiques. Les outils de modélisation (modèles et formules internes) doivent être ajustés localement et ne permettent souvent pas d'extraire des résultats intermédiaires à la réponse finale par manque de représentativité physique à l'interne même des modèles. Malheureusement, l'utilisation de ces outils se fait aussi par défaut de disponibilité suffisante en données hydrologiques et climatiques. Dans le but d'explorer le chemin vers des outils plus représentatifs, l'étude utilise de nouvelles données qui sont beaucoup plus exhaustives que les observations, permettant l'essai de formules à bases physiques pour caractériser les processus sous-jacents d'un modèle hydrologique. Plus précisément les données de réanalyses de dernière génération ERA5 et ERA5-Land. À ces données est couplé l'essai de 14 différentes estimations de l'évapotranspiration. Cet agencement permet de tester le potentiel de l'utilisation de formulation d'évapotranspiration potentielle (ETP) plus physique sur les résultats en modélisation hydrologique et donc d'ouvrir la porte vers ces changements. Afin d'assurer une généralisation la plus grande de cette étude, 1169 bassins versants sont analysés. Ils sont répartis sur l'ensemble de l'Amérique du Nord, donc au Canada, aux États-Unis et au Mexique. À cela s'ajoute l'utilisation de deux modèles hydrologiques. L'analyse des résultats se fait par l'entremise de la qualité de la modélisation, qui est jugée par la valeur du critère de performance de Kling-Gupta (KGE). Les observations restent le chef de file si l'on ne considère que les valeurs de KGE obtenues en calibration. Il y a toutefois des exceptions pour les régions montagneuses comme les rocheuses au Canada, les régions plus au nord du Canada et pour le Mexique. Dans ces cas, de meilleurs résultats en calibration avec les données de réanalyses (ERA5 et ERA5-Land) sont observables. Somme toute, la supériorité en calibration des observations ne sous-entend pas de mauvais résultats pour les données de réanalyses. Les résultats en modélisation hydrologique pour les combinaisons utilisant l'ETP à base plus physique montrent qu'il est possible d'obtenir des modèles aussi bien calibrés que par l'entremise de formulations d'ETP plus simples et plus empiriques. En général, certaines formulations plus simples génèrent de meilleurs résultats et il y a une difficulté notable en modélisation dans la région centrale des États-Unis. Somme toute, l'étude montre le potentiel en modélisation hydrologique favorable des réanalyses ainsi que le potentiel des données de réanalyses combinées à des méthodes à base plus physiques afin d'estimer l'ETP.

Mots clés : données, réanalyses, ERA5, ERA5-Land, évapotranspiration, modélisation hydrologique, représentation physique.

Evaluation of ERA5 and ERA5-Land reanalysis data utilization for hydrological modeling

Guillaume LEVESQUE

ABSTRACT

This study examines the added value of using data from ERA5 and ERA5-Land reanalyses for the estimation of evapotranspiration in hydrological modeling. The modeling tools (models and internal formulas) must be adjusted locally and often do not allow the extraction of intermediate results from the final response due to a lack of physical representativeness within the models themselves. Unfortunately, the use of these tools is also due to the lack of abundant availability of hydrological and climatic data. In an effort to explore the path to more representative tools, the study uses new data that are much more comprehensive than observations, allowing the testing of physically based formulas to characterize the underlying processes of a hydrological model. More precisely the latest generation of ERA5 and ERA5-Land reanalysis data. Coupled with this data is the testing of 14 different estimation formulas of evapotranspiration. This arrangement makes it possible to test the potential of the use of more physical potential evapotranspiration (PET) formulations on the hydrological modeling results and thus to open the door to these changes. In order to ensure the greatest generalization of this study, 1169 watersheds are analyzed. They are spread all over North America, so in Canada, the United States and Mexico. Added to this is the use of two hydrological models. The analysis of the results is done through the quality of the modeling which is judged by the value of the Kling-Gupta efficiency criterion (KGE). Observations remain the leader if only the KGE values obtained in calibration are considered. However, there are exceptions for mountainous regions such as the Rockies in Canada, more northern regions of Canada and for Mexico. In these cases, better calibration results with reanalysis data (ERA5 and ERA5-Land) are observable. All in all, the superiority in calibration of the observations does not imply poor results for the reanalysis data. The results in hydrological modeling for the combinations using the more physically based PET show that it is possible to obtain models as well calibrated as by the through simpler and more empirical PET formulations. In general, some simpler PET formulations generate better results and there is a noticeable difficulty in modeling in the central region of the United States. Overall, the study shows the favorable hydrological modeling potential of reanalyses as well as the potential of reanalysis data combined with more physical-based methods to estimate PET.

Keywords : data, reanalyses, ERA5, ERA5-Land, evapotranspiration, hydrological modeling, physical representation

TABLE DES MATIÈRES

Page

INTRODUCTION	1
CHAPITRE 1 REVUE DE LITTÉRATURE	3
1.1 Modélisation hydrologique et calibration	3
1.1.1 Modélisation hydrologique	4
1.1.2 Calibration.....	6
1.2 L'évapotranspiration	7
1.2.1 Estimation de l'évapotranspiration	8
1.2.2 Représentation physique de l'ET dans les modèles hydrologiques	10
1.2.2.1 Tendances et impacts pour le vent, l'humidité, la température et la radiation solaire.....	11
1.2.2.2 Formules d'estimation de l'ET	12
1.3 Données de réanalyses	13
1.3.1 Les réanalyses pour l'estimation de l'ET.....	14
1.3.2 Les réanalyses pour la modélisation hydrologique.....	16
1.4 Objectifs de recherche.....	19
CHAPITRE 2 DONNÉES ET SECTEURS À L'ÉTUDE.....	21
2.1 Base de données de référence	21
2.2 Bases de données météorologiques.....	22
2.2.1 Observations	23
2.2.2 Réanalyses.....	24
2.3 Secteur à l'étude.....	25
2.3.1 Zones climatiques	26
2.3.2 Domaine de l'étude	28
2.3.2.1 Caractéristiques physiographiques.....	29
2.3.2.2 Caractéristiques météorologiques	31
2.3.2.3 Caractéristiques hydrologiques	34
CHAPITRE 3 MÉTHODOLOGIE	37
3.1 Étapes générales de l'étude	37
3.2 Intrants	39
3.2.1 Choix des bassins versants.....	39
3.2.2 Données générées.....	40
3.2.3 Extraction des données	42
3.2.4 Préparation des données.....	43
3.2.4.1 Données d'observations hydriques	43
3.2.4.2 Données météorologiques de réanalyses	43
3.2.5 Données d'évapotranspiration potentielle.....	45
3.3 Modélisation hydrologique	48
3.3.1 HOOPLA	49
3.3.2 Modèles hydrologiques.....	49

3.3.2.1	MOHYSE.....	50
3.3.2.2	GR4J	53
3.3.3	Modélisation du couvert neigeux.....	54
3.3.4	Calibration des modèles hydrologique.....	56
3.3.4.1	Méthode de calibration	57
3.3.4.2	Algorithme d'optimisation.....	58
3.3.4.3	Fonction-objectif de performance.....	59
3.4	Analyse des résultats.....	60
3.4.1	Comparaison des jeux de données	60
3.4.2	Valeurs d'ETP.....	61
3.4.3	Performance en modélisation hydrologique	61
CHAPITRE 4 RÉSULTATS		63
4.1	Analyse et comparaison des jeux des données	64
4.1.1	Variables disponibles dans les trois jeux de données	65
4.1.2	Variables disponibles que pour les réanalyses.....	67
4.2	Estimation de l'ETP	70
4.2.1	Observations et réanalyses.....	70
4.2.2	Formules d'ETP	72
4.2.3	Variables d'ETP et d'ETR des réanalyses	72
4.3	Résultats de modélisation hydrologique	74
4.3.1	Observations et réanalyses	75
4.3.2	Formulation d'estimation de l'ET.....	78
CHAPITRE 5 DISCUSSION		81
5.1	Corruption et exactitude des données de réanalyses.....	81
5.1.1	ERA5.....	81
5.1.2	ERA5-Land.....	83
5.2	Incertitude liée à la modélisation de l'ETP.....	85
5.3	Taille des bassins versants	86
5.4	Représentation physique de l'ETP en modélisation hydrologique	86
5.5	Potentiel des réanalyses en modélisation hydrologique.....	89
5.6	Limites de l'étude	90
CONCLUSION	93	
RECOMMANDATIONS	97	
ANNEXE I PARAMÈTRES ET BORNES LIMITES DES MODÈLES HYDROLOGIQUES MOHYSE ET GR4J	99	
ANNEXE II ÉVAPOTRANSPIRATION POTENTIELLE ANNUELLE.....	101	
ANNEXE III RÉSULTATS DE PERFORMANCE EN CALIBRATION	103	
ANNEXE IV CORRUPTION D'ERA5	107	

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES.....109

LISTE DES TABLEAUX

	Page
Tableau 2.1 Caractéristiques principales de ERA5 et ERA5-Land	25
Tableau 2.2 Distribution des bassins versants à travers des zones climatiques de Bukovsky.	28
Tableau 2.3 Distributions des principales caractéristiques des bassins versants	36
Tableau 3.1 Liste des variables par type de recensement	44
Tableau 3.2 Classement et description des équations d'estimation d'ETP	46
Tableau 4.1 Acronymes associés aux formulations d'ET.....	63
Tableau 4.2 Numérotation des zones climatiques.....	64
Tableau 4.3 Valeurs moyennes, pour les 1169 bassins, des coefficients de corrélation et des différences moyennes	65
Tableau 4.4 Coefficients de corrélation moyens et différences moyennes.....	68
Tableau 4.5 Résumé des statistiques sur les différences entre les KGE issus des calibrations avec OBS versus les KGE avec ERA5. (La dominance définit le jeu de données dominant pour la région, la portion représente le pourcentage de bassins pour lesquels il y a la dominance) ..	76

LISTE DES FIGURES

	Page
Figure 2.1 Localisation des bassins versants en Amérique du Nord	26
Figure 2.2 Localisation des bassins versants par rapport aux zones climatiques de Bukovsky (2011). Le code de couleur, sur la figure, réfère aux différentes régions climatiques.	27
Figure 2.3 Distributions des occupations de sols par pays. La végétation inclut la forêt, la prairie, les arbustes ainsi que les zones humides alors que l'eau comprend aussi la neige et la glace.....	30
Figure 2.4 Moyennes des précipitations totales annuelles.....	32
Figure 2.5 Évapotranspiration annuelle moyenne. Les valeurs sont estimées avec la formule de Oudin (2005)	33
Figure 2.6 Carte l'équilibre eau-énergie dans l'évapotranspiration pour chaque bassin versant	34
Figure 2.7 Carte de la dominance du régime hydrologique.....	35
Figure 3.1 Organigramme des étapes générales de l'étude.....	39
Figure 3.2 Schéma originale de production de MOHYSE	51
Figure 3.3 Schéma de production du modèle MOHYSE dans HOOPLA	52
Figure 3.4 Schéma de production de GR4J	54
Figure 3.5 Schéma de production de CemaNeige.....	55
Figure 3.6 Schéma du processus de calibration d'un modèle hydrologique.....	56
Figure 4.1 Distribution spatiale des coefficients de corrélation pour la précipitation	67
Figure 4.2 Distribution spatiale des coefficients de corrélation pour l'ETP journalière entre les réanalyses ERA5 et ERA5-Land.	68
Figure 4.3 Distribution spatiale des coefficients de corrélation pour la pression atmosphérique journalière.....	69
Figure 4.4 Boîtes à moustaches des valeurs d'ETP annuelle par formulation d'ETP	70

Figure 4.5 Boîtes à moustaches des valeurs d'ET pour la région climatique -1 (centre et sud du Mexique)	71
Figure 4.6 Comparaison saisonnière des valeurs d'ETP entre FAO-PM et E5-ETP provenant d'ERA5	73
Figure 4.7 Distribution des KGE en calibration par jeux de données et par formulation d'ET	74
Figure 4.8 Différence moyenne entre les valeurs de KGE en fonction du jeu de données et du bassin versant	75
Figure 4.9 Distribution des valeurs de KGE en calibration pour la formule d'estimation de l'ETP FAO-PM	80
Figure 5.1 Exemple de la zone de la corruption (ligne grise) de ERA5 avec la température à 2m. Tiré de (ECMWF, 2021).....	83

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

BANDAS	Banco Nacional de Datos de Aguas Superficiales
BC	Blaney-Criddle
BR	Baier-Robertson
C3S	Copernicus Climate Change Service
CONAGUA	Commission nationale de l'eau
DDS	Dynamically Dimensioned Search
DEA	Demande évaporative atmosphérique
E5-ETP	ERA5 évapotranspiration potentielle
E5-ETR	ERA5 évapotranspiration réelle
E5L-ETP	ERA5-Land évapotranspiration potentielle
E5L-ETR	ERA5-Land évapotranspiration réelle
ECMWF	European Center for Medium-Range Weather Forecasting
ERA-Interim	ECMWF Reanalysis Interim
ERA5	ECMWF Reanalysis v5
ERA5-Land	ECMWF Reanalysis v5-Land
ET	Évapotranspiration
ET0	Évapotranspiration de référence
ETA	Évapotranspiration actuelle
ETP	Évapotranspiration potentielle
ETR	Évapotranspiration réelle
FAO	Food and Agriculture Organization of the United Nations

FAO-PM	FAO Penman-Monteith
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat
GR4J	Modèle du Génie Rural à 4 paramètres Journaliers
GSIM	Global Streamflow Indices and Metadata
H	Hamon
HOOPLA	HydrOlOGical Prediction LAbaratory
HS	Hargreaves-Samini
HYDAT	Archives nationales des données hydrologiques
HQ	Hydro-Québec
HYSETS	Hydrymeteorological Sandbox — École de technologie supérieure
IPCC	Intergovernmental Panel on Climate Change
JH	Jensen-Haise
K	Kharrufa
KGE	Kling-Gupta Efficiency
KGEm	Kling-Gupta Efficiency modified
M	MOHYSE
MB	McGuiness-Bordne
MOHYSE	MOdèle Hydrologique Simplifié à l'Extrême
NARCCAP	North American Regional Climate Change Assessment Program
NRCan	Resources Naturelles Canada
NSE	Nash-Sutcliffe Efficiency
O	Oudin
OBS	Observations

PM	Penman-Monteith
SCE-UA	Shuffled Complex Evolution of the University of Arizona
T	Thornthwaite
USGS	United States Geological Survey
UTC	Temps universel coordonné

LISTE DES SYMBOLES ET UNITÉS DE MESURE

h	Heure
h/j	Heure par jour
kg/m ³	Kilogramme par mètre cube
kJ/j	Kilojoule par jour
km	Kilomètre
km ²	Kilomètre carré
kPa	Kilopascal
kPa/°C	Kilopascal par degré Celsius
m	Mètre
MJ	Mégajoule
MJ/kg	Mégajoule par kilogramme
MJ/m ² j	Mégajoule par mètre carré par jour
m/s	Mètre par seconde
m ³ /s	Mètre cube par seconde
mm	Millimètre
mm/j	Millimètre par jour
Pa	Pascal
rad	radian
Wm ⁻²	Watt par mètre carré
°	Degré
°C	Degré Celsius
°E	Degré est
°N	Degré nord
%	Pourcent

INTRODUCTION

La caractérisation du cycle de l'eau s'avère importante autant pour la sécurité de l'être humain si l'on pense à la prévention d'inondations ou encore à l'approvisionnement en eau, que pour le bon développement de la technologie comme par exemple la génération d'hydro-électricité. Ces différents usages sont tous connectés en amont par les mêmes principes qui consistent en caractériser les mouvements de l'eau, autant dans le passé que dans le présent et dans le futur. Or cette caractérisation est possible par l'entremise de la modélisation hydrologique. Cet outil est aujourd'hui utilisé pour diverses applications.

Avec l'importance que prend la projection de l'impact des changements climatiques, de plus en plus d'études pointent vers le besoin de modèles hydrologiques reposant sur des fondations plus physiques et moins empiriques (Blöschl et al., 2019). De tels modèles nécessitent par contre une quantité importante de données d'entrée permettant justement une représentation plus physique des différents processus hydrologiques.

C'est donc dans ce contexte que s'inscrit cette étude, soit de faire la lumière sur l'utilisation de nouvelles données, issues des réanalyses ERA5 (Hersbach et al., 2018) et ERA5-Land (Mañoz Sabater, 2019), pour modéliser l'évapotranspiration dans 1169 bassins versants d'Amérique du Nord, couvrant une grande diversité de conditions hydroclimatiques. Le corps de ce mémoire comprend cinq (5) chapitres en plus de cette introduction et de la conclusion. En premier, une description de la science touchée est faite dans le but de positionner la présente étude par rapport à la littérature. Les deux chapitres suivants se penchent sur la présentation des données utilisées ainsi que du secteur analysé (chapitre 2) suivi de la méthodologie mise en place pour exploiter ces outils (chapitre 3). Ces présentations sont suivies par le chapitre 4 qui contient la démonstration des résultats obtenus et du chapitre 5 qui offre une analyse et discussion de ces derniers. Pour terminer, quelques recommandations ressortant de l'étude sont présentées à la suite de la conclusion.

CHAPITRE 1

REVUE DE LITTÉRATURE

Ce chapitre présente le positionnement de l'étude par rapport à la littérature. À cette fin, la recherche est basée sur 3 axes principaux, soit (1) l'utilisation de données de réanalyses en modélisation hydrologique, (2) l'augmentation de la représentativité des processus physiques sous-jacents aux modèles hydrologiques et (3) l'intérêt de la généralisation des résultats à de grandes zones d'étude. Le chapitre se conclut par une présentation des objectifs de recherche.

1.1 Modélisation hydrologique et calibration

En hydrologie superficielle, la définition et prédiction des écoulements en rivière se fait par l'entremise de modèles hydrologiques (Arsenault, Poulin, Côté, & Brissette, 2014). Ces modèles sont une suite de formules mathématiques qui décrivent les différents processus du cycle de l'eau. À la base, cette description est générale et simple dans le but de la rendre applicable à « tout » bassin versant. L'ajustement d'un modèle, à un bassin donné, se fait alors à l'aide de paramètres libres qu'il faut calibrer. Cette étape de calage permet de compenser la simplification de la représentation des processus hydrologiques et d'ainsi fournir un modèle qui décrit bien les mouvements de l'eau localement. La définition du processus de modélisation est souvent un point tournant pour une étude puisque ce choix affecte directement les résultats.

Dallaire, Poulin, Arsenault, et Brissette (2021) montrent que le choix du modèle hydrologique est une des plus grandes sources d'incertitude dans la chaîne de modélisation qui est typiquement utilisée pour l'étude des impacts des changements climatiques sur les ressources en eau. D'autres exemples de ces sources d'incertitudes sont par ailleurs les scénarios d'émission de gaz à effets de serre, la structure des modèles de climat, les méthodes de post-traitement des données de sortie des modèles de climat, ainsi que les paramètres et les composantes des modèles hydrologiques ([Chen, Brissette, Poulin, & Leconte, 2011](#)). Se

concentrant plus spécifiquement sur la modélisation hydrologique, Seiller, Anctil, et Perrin (2012) ont proposé l'utilisation d'un ensemble de modèles hydrologiques considérant que l'ensemble performe mieux que chacun des modèles pris individuellement. L'utilisation d'un ensemble de modèles pour la simulation de débits permet d'ailleurs de réduire l'incertitude liée au choix du modèle et donc par le fait même diminuer l'impact de cette incertitude sur les résultats de simulations (Seiller et al., 2012). Chaque modèle est caractérisé par les formules qui le composent ainsi que le nombre de paramètres libres de ces formules. Ce sont des caractéristiques à considérer en raison de leur impact sur l'équifinalité en calibration, un phénomène décrivant la possibilité d'atteindre une même ou de semblables réponses par l'entremise de différents jeux de paramètres. Comme le suggèrent Her et Chaubey (2015), l'équifinalité ne signifie pas nécessairement de mauvais résultats, mais c'est un phénomène dont il faut être conscient et qui est à tenir compte dans le processus de modélisation. La présente section présente plus en détail la science entourant la modélisation hydrologique et le processus de calibration.

1.1.1 Modélisation hydrologique

Le choix du type de modèle est souvent du cas à cas, il dépend principalement de la disponibilité des intrants physiographiques et hydrométéorologiques ainsi que du temps de calcul disponible (Seiller et al., 2012). Les modèles se classent en fonction de trois critères : le processus d'écoulement, le domaine spatial et le domaine temporel. La caractérisation temporelle relève simplement de si le modèle est utilisé pour simuler des débits sur une longue période de temps (continu) ou pour un événement spécifique (événementiel). Pour cette étude, tous les modèles sont de type continu temporellement.

Pour ce qui est de la représentation des processus d'écoulement, il y a deux sous-classes. Premièrement, les modèles à base physique. Ils utilisent des formules non modifiées de représentation des processus physiques du cycle de l'eau. Ce type de modèle demande beaucoup de données et est très complexe. Par nécessité de réduction du temps de calcul et de simplification en ce qui a trait aux intrants nécessaires, des modèles conceptuels ont été

développés. Ces modèles offrent une simplification des formules et se basent souvent sur une représentation des échanges d'eau entre des réservoirs (Dallaire, 2019; Seiller et al., 2012).

La classification spatiale distingue les modèles distribués des modèles globaux. Dans un modèle distribué, le bassin versant est divisé selon une grille et pour un modèle semi-distribué, en sous-région qualifiée de sous-bassin. Généralement dans ces deux cas les caractéristiques physiographiques et hydrométéorologiques peuvent être considérées uniformes en chaque tuile de la grille ou dans chaque sous-région. De tels modèles sont avantageux pour une caractérisation du régime hydrique en différents points d'un bassin. Il faut toutefois beaucoup de données, de travail pour les mettre en place et le temps de calcul est souvent très élevé puisque les valeurs des régions en amont sont nécessaires aux calculs des régions en aval. Les modèles globaux quant à eux prennent en considération une homogénéité des données sur l'ensemble du territoire du bassin versant, alors que ce n'est vraisemblablement pas le cas. Cette action ne permet donc leur utilisation que pour des études de modélisation de débits à l'exutoire, pour lesquelles leur faible nécessité en intrants et leur rapidité à délivrer les résultats sont très avantageuses (Seiller et al., 2012). La simplicité d'utilisation d'un modèle global se fait sentir plus le bassin versant est grand. Plus la superficie du bassin est grande, plus l'utilisation d'un modèle distribué est ardue alors que l'impact de la différence dans les intrants sur le régime hydrique diminue donc le désavantage d'un modèle global diminue. En général, les modèles globaux sont aussi performants que les deux autres types (distribué, semi-distribué) pour modéliser le débit à l'exutoire. Il n'y a donc pas de valeur ajoutée à l'utilisation de modèles plus complexes dans ce cas (Smith & Gupta, 2012).

La littérature scientifique recense une multitude de modèles hydrologiques. À titre d'exemple, des études ont utilisé jusqu'à 20 différents modèles (Seiller et al., 2012; Thiboult, Seiller, Poncelet, & Anctil, 2020). Tel que mentionné précédemment, plusieurs études ont montré que le choix du modèle est un des maillons pouvant générer le plus d'incertitude sur les résultats. La multitude de modèles apporte donc le fardeau du choix à l'utilisateur. Ce choix est généralement basé en fonction des données disponibles (physiographiques et hydrométéorologiques) pour le bassin versant à l'étude, des caractéristiques de l'étude à

réaliser (temps de calcul, équifinalité, etc.) et des objectifs de cette étude (Dallaire et al., 2021; Poulin, Brissette, Leconte, Arsenault, & Malo, 2011; Seiller et al., 2012).

Le choix du modèle s'accompagne aussi de celui de ses parties intrinsèques. La modularité offerte par des boîtes à outils hydrologiques comme l'« HydrOlOgical Prediction LAbaratory » (HOOPLA) (Thiboult et al., 2020) ou encore Raven (Craig et al., 2020) est avantageuse pour l'adaptabilité au secteur d'étude. Il est donc possible de choisir un modèle et d'y changer une formule calculant l'évolution du manteau neigeux par exemple. C'est ce que met de l'avant l'étude de Audrey Valéry, Andréassian, et Perrin (2014) testant six différentes routines de neige en modélisation hydrologique. Ils montrent que de tenir compte de la neige permet, bien évidemment, l'amélioration des résultats, soulignant que les résultats ne sont pas significativement influencés par la complexité de la routine.

1.1.2 Calibration

Le calage d'un modèle hydrologique se définit comme étant l'ajustement des paramètres libres dans le but de lui permettre de représenter le plus fidèlement possible les valeurs historiques observées, le plus souvent de débit. Cet ajustement se voit toutefois assez complexe considérant que les modèles peuvent comporter plusieurs paramètres, de l'ordre de la dizaine ou plus. Il est donc compliqué, voire impossible de procéder au calage manuellement et donc, il faut généralement se tourner vers des calibreurs automatiques qui consistent en des algorithmes d'optimisation (Arsenault et al., 2014; Tolson & Shoemaker, 2007).

Plusieurs algorithmes d'optimisation existent dans la littérature et chacun a ses avantages et inconvénients. Arsenault et al. (2014) ont étudié dix algorithmes afin d'identifier ceux qui sont le mieux adaptés au calage des modèles hydrologiques. Ils ont trouvé que principalement deux algorithmes se démarquent, soit le « Shuffled Complex Evolution of the University of Arizona » (SCE-UA) et le « Dynamically Dimensionned search » (DDS). Malgré que le SCE-UA est souvent le favori pour des problèmes en hydrologie en raison de sa bonne performance particulièrement avec des modèles contenant peu de paramètres, le temps de calcul est

positivement corrélé avec le nombre de paramètres à calibrer. Au contraire, la performance de DDS augmente avec la complexité du modèle puisqu'il est capable de rapidement trouver un jeu de paramètres satisfaisant, même pour des modèles qui ont dix paramètres libres et plus. (Arsenault et al., 2014). Comme le calage est un processus qui est intense en calculs, le choix de l'algorithme d'optimisation relève donc habituellement des attributs de l'étude tel que le temps disponible ainsi que la capacité de calculs disponible (Arsenault et al., 2014; Huot, Poulin, Audet, & Alarie, 2019).

La calibration d'un modèle est toujours basée sur un seul ensemble de critères de performance. Ce dernier donne une évaluation de la qualité des débits simulés en comparaison avec les débits connus, soit des observations. Un exemple générique et connu de critère de performance est le coefficient de corrélation de Pearson. En hydrologie, le critère de performance le plus utilisé dans la littérature est le « Nash-Sutcliffe Efficacy » (NSE) (Gupta, Kling, Yilmaz, & Martinez, 2009; Knoben, Freer, & Woods, 2019). Gupta et al. (2009) ont toutefois montré que cette fonction n'est pas nécessairement la meilleure pour des problèmes de calibration en hydrologie puisque le NSE tend à accorder plus de poids aux valeurs élevées de débit. Cette particularité génère des calibrations de modèles surestimant les débits de pointes et surtout les débits extrêmes, par exemple les débits printaniers résultant de la fonte de neige. De plus, lorsque décomposé en termes du coefficient de corrélation, du ratio des variances et du biais (entre débits simulés et débits observés), le NSE accorde plus d'importance au biais. Le critère « Kling-Gupta Efficiency » (KGE), qui est une version adaptée et plus équilibrée du NSE, devient de plus en plus répandu en hydrologie (Knoben et al., 2019). Ce critère de performance est meilleur pour l'ajustement à des valeurs extrêmes que le NSE puisqu'il permet une meilleure représentation de la variabilité dans les débits (Gupta et al., 2009; Kling, Fuchs, & Paulin, 2012; Knoben et al., 2019).

1.2 L'évapotranspiration

En soi, ce processus représente le transfert de la vapeur d'eau à partir de la surface et de la végétation vers l'atmosphère (Ricard & Anctil, 2019). Il est connu que l'évapotranspiration

(ET) est un processus de cycle de l'eau très important puisque l'ET peut représenter une part importante des précipitations annuelles totales. Par exemple, en Chine, ce sont environ les deux tiers des précipitations annuelles qui retournent à l'atmosphère via l'ET (Mao & Wang, 2017). Le taux d'ET réelle (ETR) est principalement géré par la disponibilité en eau dans les sols et par la demande évaporative atmosphérique (DEA) qui a comme principaux facteurs : (1) la radiation solaire, (2) la température, (3) le vent et (4) l'humidité relative (Dai, 2006; Donohue, McVicar, & Roderick, 2010; Ricard & Anctil, 2019; Wild, 2009). L'ETR est toutefois difficilement calculable avec précision, notamment en raison d'aspects liés à l'influence des types de végétation et leur stade de croissance que les quantités d'eau disponibles et prélevées via l'ET. C'est pourquoi les termes évapotranspiration potentielle (ETP) et évaporation de référence (ET_0) ont été introduites dans la littérature scientifique.

L'ETP consiste en une estimation de l'ET sans limites sur la disponibilité en eau. Cette valeur représente donc une limite supérieure sur l'ETR (Dallaire, 2019; Seiller & Anctil, 2016). Plus récemment, le terme ET_0 a été popularisé. Ce phénomène correspond à l'ET pouvant s'évaporer d'un sol sans limites en eau et pour une végétation de référence. L'ETP s'obtient par la multiplication de ET_0 par un facteur adapté à la végétation. L'ETP est un phénomène très complexe et difficile à estimer. Le grand nombre de formules d'estimation peut d'ailleurs en témoigner (Seiller & Anctil, 2016). Selon Dallaire et al., 2021, la modélisation de l'ET peut d'ailleurs contribuer à générer une part importante de l'incertitude sur la modélisation des débits d'étiage en climats historique et futur.

1.2.1 Estimation de l'évapotranspiration

L'estimation de l'ETP se fait généralement à l'aide de formules empiriques, la physique du processus étant plutôt complexe à modéliser, surtout à grande échelle spatiale. Au fil du temps, une panoplie de formules ont été développées ou modifiées, lesquelles peuvent être regroupées en 4 groupes : (1) basées sur la température, (2) la radiation solaire, (3) le bilan d'énergie et (4) combinatoires (Bormann, 2011; Mao & Wang, 2017; Oudin et al., 2005; Seiller & Anctil, 2016). La littérature tente sans cesse de déterminer la meilleure formule générale, mais les

résultats se contredisent et dépendent beaucoup de l'ampleur et de la localisation du projet. Il devient donc difficile de faire un choix de formulation pour une étude. La communauté scientifique s'accorde toutefois sur le fait que la formulation de la « Food and Agriculture Organization of the United Nations » (FAO) de Penman-Monteith (FAO-PM) représente les meilleures valeurs de l' ET_0 et est celle qui permet la meilleure représentation des processus physiques dirigeant l'ET. Ce n'est toutefois pas le cas au niveau de sa performance pour calculer l'ETP (Bormann, 2011; Raza et al., 2020; Seiller & Anctil, 2016).

Donohue et al. (2010) ont étudié cinq formulations d'ETP différentes, en Australie, et ont conclu que la formule de Penman-Monteith (PM) permettait d'obtenir les valeurs les plus justes d'ETP. L'étude ne regardait toutefois que les valeurs elles-mêmes et ne considérait pas le potentiel en modélisation hydrologique. Une étude de Bormann (2011) vient opposer ses résultats sous plusieurs angles. Malgré que l'impact des différentes formules était étudié en contexte de changements climatiques, il a trouvé que la variabilité dans les résultats d'ETP était sensiblement la même à travers toutes les 18 formules testées. Cette étude est toutefois difficilement transférable puisqu'elle n'a été réalisée qu'en Allemagne.

Sur un plan plus large et touchant directement la modélisation hydrologique, une étude testant 27 formules, dans 308 bassins versants répartis sur trois continents, a montré que la simplicité des modèles d'ETP est favorable à de meilleurs résultats en modélisation hydrologique (Oudin et al., 2005). C'est-à-dire que des formules basées sur la température ou la radiation solaire sont plus performantes que les formules combinatoires. L'étude d'Oudin et al. (2005) souligne particulièrement le fait que la formule de Penman-Monteith (combinatoire) ne donne pas de bons résultats en modélisation hydrologique. Leur analyse soutient le fait que la formule demande beaucoup de données en entrée et qu'au moment de l'étude la qualité des données rares comme la vitesse du vent ou encore l'humidité relative de l'air n'était pas très bonne. La mauvaise performance des formules combinatoires serait donc peut-être renversée par l'utilisation de meilleures données. Les résultats de cette étude sont partiellement confirmés par l'étude de Seiller et Anctil (2016) testant l'impact des formules sur les projections hydrologiques en changements climatiques. Leur étude est localisée sur seulement sur deux

bassins versants, mais sur deux continents et considère 24 formules d'ETP accompagnées d'un ensemble de 20 modèles hydrologiques. Le but de cet ensemble étant de réduire l'incertitude liée au choix du modèle. Comme pour Oudin et al. (2005), les formules à base de radiation solaire offrent les meilleurs résultats. À l'opposé, Seiller et Anctil (2016) ont trouvé que les formules combinatoires étaient la classe qui générait le moins de variances dans les résultats, quoique ces formules ne soient pas associées à la production des meilleures réponses hydrologiques. Les réponses hydrologiques étaient plus constantes que pour les formules à base de température ou à base de radiation, suggérant une plus grande indépendance aux données d'entrées météorologiques.

Outre les formules d'ETP, la science tente d'évaluer de nouvelles pistes de solution dans le but de mieux estimer l'ET ou encore de mieux la comprendre. C'est le cas au Pakistan où récemment Raza et al. (2020) ont montré que d'estimer l' ET_0 à l'aide d'intelligence artificielle pouvait donner des valeurs satisfaisantes en se basant sur l'équation de FAO-PM. Il y a aussi Ruggieri et al. (2021) qui, en Italie, ont évalué avec succès les valeurs d'ETR en provenance de données satellites afin de modéliser la recharge en eau d'un aquifère.

1.2.2 Représentation physique de l'ET dans les modèles hydrologiques

La question de l'ET est un sujet amplement abordé en raison de la compréhension partielle de ce processus et de la disparité des résultats dans la littérature. Les conclusions variables auxquelles arrivent les différentes études dépendent notamment des paramètres de ces études (nombre de bassins versants, secteurs étudiés, variables prises en compte, par exemple) et du but de l'estimation de l'ET. Par exemple, en modélisation hydrologique agricole, c'est l' ET_0 qui est importante alors qu'en modélisation hydrologique de rivière, c'est l'ETP. Contrairement à la modélisation pour l'agriculture, il est accepté, pour de la modélisation de débits, que moins d'importance soit accordée à la complexité de l'ET et à la précision de la représentation de la physique sous-jacente (Blöschl et al., 2019). C'en est ainsi puisque la calibration des paramètres libres permet à un modèle de compenser pour ces élément et de quand même bien s'ajuster à la réalité même si l'ETP demeure une estimation incertaine (Milly

& Dunne, 2011; Ricard & Anctil, 2019). Jusqu'à quelques années passées, la science acceptait globalement qu'en modélisation hydrologique des débits, qu'il était important d'avoir un bon modèle calibré, moins d'importance était accordée à la représentation physique des processus. Les études de changements climatiques ont toutefois remis en question ce paradigme.

Il y a un consensus sur le changement climatique en cours. Plusieurs études et, notamment, les rapports du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), appuient ce consensus (GIEC, 2021). Par exemple, il est factuel que la température augmentera et que le cycle de l'eau subira des changements importants. Le résultat étant un cycle hydrologique plus intense et donc par le fait même une augmentation de l'ET (M. R. Allen & Ingram, 2002; Huntington, 2006; Ohmura & Wild, 2002). Pourtant, des études analysant la sensibilité et l'évolution des paramètres gérant l'ET (radiation solaire, température, humidité, vent et la disponibilité en eau) semblent sous-entendre une diminution de l'ET.

1.2.2.1 Tendances et impacts pour le vent, l'humidité, la température et la radiation solaire

Dans leur étude, Ramanathan, Crutzen, Kiehl, et Rosenfeld (2001) confirment la relation directe entre la radiation et l'ET. Ils trouvent qu'à l'échelle globale, une augmentation de la concentration en aérosols dans l'atmosphère a contribué à réduire la radiation solaire de surface dans la deuxième partie du 20^e siècle. Cette diminution a résulté en une baisse proportionnelle de l'ET et suggère un allègement du cycle de l'eau. Plus récemment par contre, Wild (2009) et Huntington (2006) ont réalisé des études globales à ce niveau. Chacune montre une augmentation de la radiation solaire et de la température de l'air dans les plus récentes années (au-delà de 1990). Cette dernière concorde avec une corrélation positive entre les deux variables. Ces phénomènes résultent en une intensification de précipitations, et donc par le fait même, du cycle hydrologique. Ces études montrent aussi l'inverse de la relation pour les bacs d'évaporation, soit l'augmentation de la radiation et la diminution de l'évaporation, ce qui va contre le consensus. Leur analyse suggère toutefois que le déclin de l'évaporation mesurée en bacs est lié à une augmentation de l'ETR et au paradoxe d'évaporation (expliqué au chapitre 5), réduisant la demande en évaporation des bacs (Huntington, 2006). C'est d'ailleurs analogue

à Wild (2009) qui a trouvé que l'augmentation en radiation affecte positivement l'ETR en régions limitées en énergie. C'est-à-dire les régions dont l'ET est principalement limitée par le manque d'énergie (radiation solaire) contrairement à limitée par la disponibilité en eau.

Le niveau de radiation solaire est aussi affecté par d'autres variables. Dai (2006) a analysé les tendances globales dans l'humidité entre 1975 et 2005. Cette étude montre une augmentation significative de l'humidité, surtout pour l'hémisphère nord. Les corrélations positives entre la température et l'humidité ainsi qu'entre l'humidité et la quantité de nuages ont également été vérifiées. Sans toutefois avoir analysé l'impact de l'humidité sur d'autres variables, il mentionne qu'au fil du temps une présence plus accrue de nuages aura un impact négatif sur le niveau de radiation solaire en surface.

Les études de Breña-Naranjo, Laverde-Barajas, et Pedrozo-Acuña (2017) ainsi que de T. R. McVicar et al. (2012) montrent que le vent est l'un des facteurs les plus importants, c'est-à-dire ayant le plus d'impact sur les valeurs d'ET. Les deux études ont observé, globalement, une diminution de la vitesse du vent. Cette réduction coïncide avec une diminution en évaporation mesurée en bacs, mais aussi en ET₀ (Breña-Naranjo et al., 2017; T. R. McVicar et al., 2012). Le vent joue un rôle important dans la demande évaporative atmosphérique puisque les mouvements de l'air permettent de réduire le niveau de saturation en vapeur d'eau de l'air. L'estimation de l'ET peut donc souffrir de biais puisque la majorité des formulations ne considèrent pas le vent. C'est pourquoi T. R. McVicar et al. (2012) suggèrent de procéder avec précaution dans l'estimation de l'ET et de toujours considérer les quatre variables dirigeant la demande évaporative de l'atmosphère.

1.2.2.2 Formules d'estimation de l'ET

La littérature soutient majoritairement une estimation de l'ET avec des formules les plus simples possible, entre autres parce que les formulations plus complexes demandent beaucoup de données, augmentant l'incertitude des résultats et par le fait même réduisant la performance des modèles hydrologiques. Quelques études tentent toutefois de paver une nouvelle voie vers

l'utilisation plus systématique de formulations combinatoires comme celle de FAO-PM dans le but d'offrir une représentation physique plus juste de l'ET.

McVicar, Roderick, Donohue, et Van Niel (2012) montrent, dans leur étude globale, que la réduction de la vitesse du vent peut avoir un effet variable sur le régime hydrique. Pour des régions très humides, une augmentation des débits se fait ressentir. Pour les régions arides, l'impact ne se fait pas ressentir puisque l'ET est conduit par les précipitations alors que pour les régions équilibrées, l'impact est difficilement discernable. Ils suggèrent que des méthodes plus physiques d'estimation de l'ETP seraient nécessaires.

Donohue et al. (2010) ont étudié la capacité de cinq formules d'ETP à capturer la variabilité et l'interaction des variables de vent, d'humidité, de radiation solaire et de température en Australie. Les résultats montrent que la seule équation considérant les quatre variables, la formule de PM, permet de bien capturer les tendances et de donner les résultats les plus vérifiques. L'étude souligne aussi que l'augmentation en température n'est pas directement et positivement corrélée avec la quantité d'ET, contredisant les études de Huntington (2006) et Wild (2009).

En soi, les dernières études touchantes plus spécifiquement la méthode d'estimation de l'ETP et les valeurs en résultant sont en accord sur le fait que les quatre variables conduisant la demande évaporative de l'atmosphère doivent être considérées (Donohue et al., 2010; T. McVicar et al., 2012). Cette suggestion est directement alignée avec les 23 problèmes à résoudre en hydrologie soulevés dans Blöschl et al. (2019) soulignant que des méthodes à bases plus physiques sont nécessaires afin de permettre une plus grande transférabilité spatiale et temporelle des modèles hydrologique.

1.3 **Données de réanalyses**

Une réanalyse est une méthode scientifique pour développer un enregistrement complet de la façon dont le météo et le climat changent au fil du temps. Des observations et un modèle

numérique qui simule un ou plusieurs aspects du système terrestre sont combinés objectivement de façon à générer une estimation synthétisée de l'état du système. Une réanalyse s'étend généralement sur plusieurs décennies ou plus et couvre l'ensemble du globe, depuis la surface de la Terre jusqu'à bien au-dessus de la stratosphère (<https://reanalyses.org/>). Une grande quantité de variables atmosphériques et terrestres décrivant la météo et le climat sont ainsi produites.

Il existe cinq générations de réanalyses et les processus qui permettent de les générer ainsi que les données de sortie se sont évidemment améliorées d'une génération à l'autre. La littérature contient un bon nombre d'études des réanalyses, plus précisément sur leur précision (Mostafa Tarek, François P. Brissette, & Richard Arsenault, 2020). Par exemple, Bosilovich (2013) a étudié la variabilité des réanalyses MERRA et ERA-Interim (de deuxième génération) pour les variables de températures et précipitations. L'étude montre que la température est bien représentée alors que c'est l'opposé pour les précipitations. Ce résultat est en accord avec Manzanas, Amekudzi, Preko, Herrera, et Gutiérrez (2014) qui ont analysé la variabilité et la tendance interannuelles des précipitations des réanalyses ERA40 (première génération), ERA-Interim (deuxième génération) et NCEP (première génération). Leur étude montre qu'au Ghana, les réanalyses surestiment la variabilité et donc que ces produits sont à utiliser avec grande précaution. Récemment, la communauté scientifique a commencé à étudier le potentiel des réanalyses sous de nouvelles facettes puisque leur précision s'améliore et les biais et incertitudes à leur utilisation diminuent (Hersbach et al., 2020). Il y a notamment de maintes études analysant le potentiel d'estimation de l'ET. De plus, alors que peu d'études sur l'utilisation des données de réanalyses en modélisation hydrologique étaient présentes, les chercheurs voient de plus en plus leur potentiel et leur utilité (M. Tarek, F. P. Brissette, & R. Arsenault, 2020).

1.3.1 Les réanalyses pour l'estimation de l'ET

Sur le plan de l'ET, la littérature tente de situer l'aptitude des réanalyses à capter les tendances de l'ET, et par le fait même, voir s'il y a un potentiel pour estimer l'ET à partir de ces dernières.

Feng, Su, Zhi, Tu, et Ji (2019) ont analysé la variabilité et les tendances dans les valeurs de l'ET globalement entre 1979 et 2015. Dans cette étude, sept réanalyses ont été utilisées pour générer les séries d'ET dont NCEP-NCAR (première génération), ERA-Interim, MERRA et CFSR (ces trois dernières de deuxième génération). Les résultats montrent que ce ne sont pas toutes les réanalyses qui permettent de bien modéliser les tendances historiques en raison de biais, NCEP-NCAR étant la moins performante. Ils recommandent donc d'utiliser ces sources de données avec précautions.

Quelques études ont jugé l'utilisation des données de réanalyses dans le but de créer des valeurs de ET_0 et d'alimenter des formulations basées sur PM (PM, PM modifiée, PM-FAO). Ces études sont très localisées et donc leur généralisation n'est pas possible. Srivastava, Han, Ramirez, et Islam (2013) ont utilisé ERA-Interim et NCEP-NCAR sur un bassin versant en Angleterre. Les résultats montrent que l' ET_0 générée avec ERA-Interim était comparable à celle en provenance d'observations. Pour NCEP, les résultats n'étaient pas concluants. Ces résultats sont partiellement en concordance avec l'étude de Martins et al. (2017) dans la région de la péninsule ibérique. Les résultats avec NCEP-NCAR sous-estimaient les valeurs de ET_0 . La différence se trouve au niveau des résultats avec ERA-Interim qui surestimaient les valeurs et donc ne pouvaient être considérés valables.

Mao et Wang (2017) ont étudié quatre réanalyses, soit deux réanalyses de base ainsi que leur version terrestre (ERA-Interim, ERA-Interim Land, MERRA et MERRA Land), en Chine. Il est montré qu'en général les réanalyses ne sont pas aussi bonnes que les observations pour le calcul de l' ET_0 . Complémentairement, Mao et Wang (2017) ont étudié le potentiel d'utiliser les valeurs d'ET sortant directement de ces mêmes réanalyses. Malheureusement, les résultats n'étaient pas concluants. Les variables d'ET de ces réanalyses n'étaient pas en mesure de bien capter les variabilités et tendances interannuelles. À la connaissance de l'auteur, Mao et Wang (2017) est la première étude à avoir jugé le potentiel des variables d'ET en provenance de réanalyses comme valeurs d' ET_0 .

Dernièrement, Pelosi, Terribile, D’Urso, et Chirico (2020) ont présenté le même genre d’analyse, mais en Italie et en comparant deux réanalyses de surface, dont ERA5-Land (cinquième génération). Il est montré que ERA5-Land offre des données de bonne qualité et permet de générer des valeurs de ET₀ satisfaisantes. Cette étude confirme ainsi la possibilité d’utiliser les données de ERA5-Land en remplacement des données observationnelles manquantes si nécessaire.

1.3.2 Les réanalyses pour la modélisation hydrologique

Originalement les études sur l’utilisation des données de réanalyses pour faire de la modélisation hydrologique n’étaient pas très répandues. Pour celles existantes, elles sont majoritairement à petite échelle, possiblement en raison de la grande incertitude et du biais qu’apportait leur utilisation. Les données analysées, dans toutes les études citées dans la présente section, sont les précipitations et les températures au sol issues de réanalyses. Par exemple, Choi, Kim, Rasmussen, et Moore (2009) ont étudié la réanalyse NARR (deuxième génération) sur trois bassins versants dans la province du Manitoba au Canada. Cette étude a trouvé que les données de sortie de la réanalyse fournies à un modèle hydrologique calibré par des observations résultait en de mauvaises performances et un important biais dans les débits. Cela indique une faible fiabilité de NARR¹ qui, similairement aux observations, aurait dû produire de bons débits. Il est toutefois montré que le biais pouvait être retiré des débits simulés et ainsi atteindre une aussi bonne performance que les observations en utilisant un modèle calé avec la réanalyse même. Aujourd’hui la littérature contient de plus récentes études permettant une meilleure généralisation des résultats et par le fait même d’avoir une meilleure idée du potentiel des données de réanalyses. Essou, Sabarly, Lucas-Picher, Brissette, et Poulin (2016) et Essou, Brissette, et Lucas-Picher (2017) en sont d’ailleurs de grands contributeurs.

¹ Il est à noter que la réanalyse NARR, contrairement à toutes les autres, est régionale. Elle ne couvre que le territoire de l’Amérique du Nord. Elle avait aussi la particularité d’assimiler des données de précipitations provenant de stations météo, alors que toutes les précédentes n’assimilaient que des données atmosphériques.

En premier lieu Essou et al. (2016) ont étudié quatre réanalyses différentes, et ce, sur 370 bassins versants aux États-Unis. Les réanalyses étudiées étaient ERA-Interim, CFSR, MERRA et NARR. Dans tous les cas, la température était fiable et elle captait bien les variabilités et les tendances. C'est plus au niveau de la précipitation qu'il y avait un biais positif, surtout présent pour les régions plus humides à l'est du pays. En modélisation, les réanalyses ont bien performé, malgré qu'inférieurement aux observations. Le seul bémol dans les débits était à l'est du pays, où le biais dans les données de précipitations était le plus grand.

Essou et al. (2017) ont poursuivi avec une étude au Canada, et ce, sur 316 bassins versants. Elle portait sur l'utilisation de données de réanalyse comme données d'entrée en modélisation hydrologique, mais basée sur la densité des stations météorologiques. C'est-à-dire que le but était de vérifier la performance des réanalyses contre celles des observations en régions comportant un différent nombre de stations météorologiques. L'étude utilisait comme données de réanalyses celles de ERA-Interim, MERRA et CSFR. Cette dernière montre que même si en général les observations performaient mieux que les réanalyses, les débits simulés avaient une précision satisfaisante. C'est pour des bassins versants ayant moins d'une station par 1000 kilomètres carrés que les réanalyses se démarquent des observations. Les réanalyses permettent une meilleure définition de la variabilité des précipitations dans ces cas. Plus particulièrement, ERA-Interim et MERRA performent significativement mieux dans les zones montagneuses du Canada (Essou et al., 2017).

L'arrivée de nouvelle génération de réanalyse de ECMWF, ERA5 (cinquième génération), a poussé Tarek et al. (2020) à définir son potentiel et vérifier si elle est meilleure que ERA-Interim, son prédécesseur. Cette étude est d'autant plus intéressante en raison de sa transférabilité. Elle couvre 3138 bassins versants répartis au Canada et aux États-Unis. Ils ont trouvé que ERA5 comporte encore quelques biais dans les températures et précipitations, particulièrement à l'est des États-Unis pour les précipitations. Ces biais sont toutefois beaucoup moindres que ceux dans ERA-Interim. En modélisation hydrologique, il est démontré qu'en général, la performance des débits obtenus par ERA5 était similaire à celle des données d'observation. Pour la première fois, aux connaissances de l'auteur, une étude soutient

positivement l'utilisation des données de réanalyses pour effectuer de la modélisation hydrologique.

Tarek et al. (2020) ont poussé cette dernière étude un peu plus loin en testant le potentiel des données de réanalyse comme données de référence pour des études en changements climatiques. Cette suivante analyse a conservé la même couverture spatiale, mais en ajoutant 7 nouvelles sources de données dont au moins 3 autres réanalyses et conservant ERA5. Au niveau de la température, l'ensemble des jeux de données performe similairement. Il n'est pas possible d'en dire autant pour les précipitations qui induisent un biais dans les résultats et donc ne permettent pas une aussi bonne performance pour une étude en changements climatiques. Les recommandations soulignent par contre que malgré qu'il serait conseillé d'utiliser les données d'observation, ERA5 est une alternative à considérer dans les régions pour lesquelles les données d'observations de qualité se font rares.

Dans les études touchant les réanalyses, la plus profonde est probablement celle de Ricard et Anctil (2019). Dans cette analyse, des données d'observations et des données de réanalyses sont combinées afin d'estimer l'ETP avec la formulation de PM. Les données de base comme la température proviennent des observations alors que l'humidité, le vent et la radiation solaire proviennent des réanalyses. Quatre réanalyses sont utilisées, soit MERRA-2 (troisième génération), ERA-Interim, CFSR et JRA-55 (ces trois dernières de deuxième génération). Les résultats montrent une performance comparable des débits simulés avec la combinaison de données versus ceux générés avec des modèles d'ETP empiriques basés sur la température et n'utilisant que des observations. Ce sont des résultats importants puisque l'utilisation de la formulation de PM permet une meilleure fidélité que des formules d'ETP à base de température au niveau de la représentation physique sous-jacente aux modèles. L'analyse montre aussi que la performance des résultats par utilisation des réanalyses peut être significativement améliorée en procédant à une correction de biais sur les données de réanalyses avant leur utilisation. Malheureusement, leurs résultats manquent de possibilité de généralisation puisque l'étude n'a été faite que sur six bassins versants, le tout concentré dans le sud du Québec au Canada. Aux connaissances de l'auteur, cette étude est la seule de son genre dans la littérature.

1.4 Objectifs de recherche

Le but de l'étude est à caractère exploratoire. Elle étudie le potentiel des réanalyses de dernier cri ERA5 et ERA5-Land (cinquième génération) dans un cadre de modélisation hydrologique entièrement alimenté par les données de réanalyses. Cela veut donc dire que si les variables sont disponibles, elles sont utilisées. Le but avec cette exploration est de contribuer à la généralisation des résultats. C'est pourquoi elle comprend 1169 bassins versants répartis sur l'ensemble de l'Amérique du Nord : au Canada, aux États-Unis et au Mexique. Il s'agit donc d'une des premières études du potentiel de modélisation hydrologique avec données de réanalyses couvrant une aussi grande région de la terre et utilisant autant de variables en provenance des réanalyses.

Dans le même ordre d'idée que le problème touchant la représentativité des processus physiques sous-jacent aux modèles présentés dans Blöschl et al. (2019), l'auteur teste différentes formulations de l'estimation de l'ETP, particulièrement celle de FAO-PM qui est reconnue comme celle donnant les valeurs de références. À cela s'ajoute l'analyse du potentiel des variables d'ETR et d'ETP que fournissent ERA5 et ERA5-Land. Elles sont calculées par les modèles de réanalyse météorologique et devraient donc dépendre plus des processus physiques en interaction dans le climat, un aspect qui s'aligne avec le désir de l'auteur d'augmenter la représentativité physique des modèles hydrologiques. C'est d'ailleurs, à la connaissance de l'auteur de la littérature, la première fois que ce type de variables est analysé sous l'angle du potentiel en modélisation hydrologique. L'étude se rapprochant le plus de cet aspect est celle de Ricard et Anctil (2019). Cette dernière ne force seulement que l'humidité, le vent et la radiation solaire pour alimenter les formulations d'ETP en modélisation hydrologique et ne tient pas compte des variables d'ETP et d'ETR issues de réanalyses.

À la lumière de ces buts et du côté exploratoire de l'étude, les objectifs principaux pour l'analyse sont (1) de définir le potentiel des réanalyses, pour la modélisation de l'évapotranspiration, en fonction des caractéristiques physiographiques et

hydrométéorologiques des bassins versants afin de permettre la meilleure généralisation des résultats et (2) de déterminer l'impact de l'augmentation de la représentation des processus physiques au sein d'un modèle hydrologique sur la performance des débits simulés.

CHAPITRE 2

DONNÉES ET SECTEURS À L'ÉTUDE

Ce chapitre présente les bases de données utilisées ainsi que le domaine de l'étude. La plupart des données proviennent d'une base de données de référence (HYSETS). Sa provenance et ce qui en est utilisé sont présentés, suivis d'une plus ample description des différents types de données externes à HYSETS ainsi que des caractéristiques physiographiques et géographiques du secteur à l'étude.

2.1 Base de données de référence

La base de données de base pour l'étude est la récente « Hydrometeorological Sandbox — École de technologie supérieure » (HYSETS) (Arsenault et al., 2020). Les données de HYSETS peuvent être téléchargées gratuitement sur le site <https://osf.io/rpc3w/>. Elle offre une multitude de types de données autant hydrométéorologiques que physiographiques. Elle comprend 14 425 bassins versants distribués sur l'ensemble de l'Amérique du Nord, incluant le Canada, les États-Unis et le Mexique. Pour chacun de ces bassins, il est plus précisément possible d'obtenir des données d'observations de débits, des données météorologiques d'observation et de réanalyses ainsi que leurs caractéristiques physiographiques.

Les données sur les débits d'eau journaliers proviennent de différentes sources dépendamment du pays de provenance (Arsenault et al., 2020). Pour les bassins versants localisés au Canada, les débits sont tirés de la base de données « Archives nationales des données hydrologiques » (HYDAT) d'Environnement et Changement Climatique Canada. Cette dernière donne l'information sur les débits journaliers enregistrés par le réseau de stations accompagné des métadonnées sur les bassins versants associés au cours d'eau, par exemple l'aire de drainage. Dans le cas de tout ce qui est localisé aux États-Unis, les données proviennent du service web

« National Water Information System » de l’« United States Geological Survey » (USGS). Pour terminer, les données en provenance du Mexique sont fournies via la Commission nationale de l’eau (CONAGUA) par l’entremise de la base de données « Banco Nacional de Datos de Aguas Superficiales » (BANDAS). Un premier filtre a été appliqué sur les données du Mexique en raison de la qualité variable des données. Seulement les stations comportant des séries temporelles avec moins de 40% de données manquantes ont été conservées. Un second filtre est appliqué sur la liste de stations, cette fois-ci pour les trois pays, dans le but de retirer tous les cours d’eau comportant une structure de contrôle du régime pouvant affecter significativement les débits (barrage hydroélectrique, réservoir, etc.) (Arsenault et al., 2020).

Les données physiographiques nécessaires à l’étude sont l’aire de drainage, le contour des bassins ainsi que leur localisation. Pour la majorité des bassins, l’aire de drainage est fournie dans les fichiers de métadonnées distribués en paire avec les données hydriques par les entités respectives mentionnées plus haut. Il en est de même pour ce qui est des contours de bassins qui viennent sous forme d’un fichier « shapefile ». Dans certains cas, la géométrie des bassins n’est pas disponible. Elles ont donc été définies à partir du « Global Streamflow Indices and Metadata » (GSIM) qui fournit une délinéation estimative (Do, Gudmundsson, Leonard, & Westra, 2018; Gudmundsson, Do, Leonard, & Westra, 2018). Pour les bassins en provenance de GSIM, leurs métadonnées (aire, localisation) sont extraites à l’aide de l’application QGIS 3.4 (Arsenault et al., 2020). Pour l’étude, les métadonnées d’origine (fournies par les organisations) sont préférées aux données extraites à partir de la géométrie de GSIM. Comme GSIM est une estimation, il est jugé que les métadonnées d’origine sont plus fiables.

2.2 **Bases de données météorologiques**

Les données météorologiques sont de deux différents types. Il y a des données d’observations ainsi que deux différentes réanalyses. Peu importe leur origine, chacun des types de données est toutefois fourni sous la même structure. Le format représente une grille de points en longitude et latitude, suivant une résolution spatiale. Chaque point de la grille comporte une

série temporelle de données climatiques. Plus de détails sont donnés dans les sous-sections qui suivent.

2.2.1 Observations

Les données d'observations proviennent directement de HYSETS. Comme il n'y a pas de grille couvrant l'ensemble du territoire de l'Amérique du Nord, deux produits distincts sont exploités en fonction de la localisation des bassins. La première source de grilles d'observations est la base de données Livneh et la seconde est la grille de Ressources Naturelles Canada (NRCan).

Les grilles climatiques Livneh (Livneh et al., 2015) couvrent la superficie du Mexique, des États-Unis et du sud du Canada. Les données sont fournies en grilles ayant une résolution spatiale de $0,0625^\circ$ par $0,0625^\circ$ en longitude et latitude avec une couverture temporelle journalière de 1915 à 2015. Dans HYSETS, seulement les années 1950 à 2015 sont conservées. Les données de Livneh sont distribuées par NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, de leur site web au <https://www.esrl.noaa.gov/psd/>. Les grilles climatiques de NRCan (McKennedy et al, 2011), quant à elles, couvrent la superficie du Canada allant jusqu'à une latitude d'environ 84°N avec une résolution spatiale d'environ 10 km par 10 km en longitude et latitude et une couverture temporelle journalière de 1950 à 2010. Dans les deux cas, chaque point de grille résulte d'une interpolation entre des données de stations météorologiques sur la précipitation, la température minimale et la température maximale journalières (Arsenault et al., 2020).

Les données dans HYSETS sont distribuées sous forme de séries temporelles singulières pour chaque bassin. Une agglomération des points de grilles, pour chaque bassin, est nécessaire pour atteindre ces résultats. Les données provenant de Livneh et NRCan ont subi la même conversion. C'est-à-dire que pour chaque bassin, une agglomération spatiale en longitude et latitude est appliquée sur la totalité des points de grilles se retrouvant à l'intérieur du contour du bassin afin de ne conserver qu'une seule série de données représentant les valeurs moyennes journalières pour l'ensemble du bassin (Arsenault et al., 2020). Pour l'étude, la priorité est

accordée aux données de Livneh, donc pour les bassins versants canadiens qui sont couverts par des données de Livneh et de NRCAN, celles de Livneh sont utilisées. La motivation de ce choix touche le désir d'avoir le plus d'uniformité possible dans les sources d'intrants.

2.2.2 Réanalyses

Pour l'étude, deux réanalyses sont utilisées. En provenance du « European Center for Medium-Range Weather Forecasting » (ECMWF), les produits sont l'« ECMWF Reanalysis v5 » (ERA5) (Hersbach et al., 2018), et la « ECMWF Reanalysis v5-Land » (ERA5-Land) (Mañoz Sabater, 2019). En soi, les deux produits sont similaires. Toutes deux sont fournies à un pas de temps horaire sous la base du fuseau horaire « temps universel coordonné » (UTC). Les deux réanalyses se différencient toutefois en ce qui a trait à leur résolution spatiale et à leurs étendues temporelle et spatiale. Leurs principales caractéristiques sont montrées dans le tableau 2.1. ECMWF travaille conjointement avec le « Copernicus Climate Change Service » (C3S) pour le développement de ces produits. Leur accessibilité se fait d'ailleurs via le « C3S Data store » au [https://cds.climate.copernicus.eu/#!/home](https://cds.climate.copernicus.eu/#!/).

ERA5 est la réanalyse de cinquième génération de ECMWF qui remplace « ECMWF Reanalysis Interim » (ERA-Interim). Elle est générée en utilisant comme état initial la simulation climatique Cy41r2 du « Integrated Forecast System » et une routine d'assimilation de données d'observations à quatre dimensions (Hersbach et al., 2020). La réanalyse offre de multiples variables, plus précisément 205 différentes variables au niveau de la surface et un total de 281 en comptant les variables spectrales et celles atmosphériques. La réanalyse est en constant développement et sa date de fin avance dans le temps de sorte que mensuellement, de nouvelles données sont disponibles, et ce, avec un délai d'environ 5 jours (Hersbach et al. 2020).

ERA5-Land est produite conjointement avec ERA5. Elle équivaut à une version de ERA5 avec une résolution de grille plus fine et modélise seulement au-dessus de la terre, ignorant les étendues d'eau. Elle est générée directement par une simple simulation avec les paramètres de

surface d'ERA5 en données d'entrée (Hersbach et al., 2020). Sa couverture temporelle est un peu plus petite puisqu'elle commence plus tard et elle se termine en temps presque réel avec un délai de quelques mois. Pour l'étude, la série temporelle disponible se termine en 2019. Sa méthode de génération induit une contrainte au niveau du nombre de variables disponibles. Seulement des variables considérées proches de la surface terrestre sont disponibles, sans toutefois tout ce qui est lié aux étendues d'eau. Les variables utilisées pour l'étude sont la température de l'air au sol, la température du point de rosée, la vitesse du vent, l'évaporation, l'évaporation potentielle, les précipitations totales, la pression atmosphérique, la radiation solaire nette ainsi que la radiation solaire extraterrestre (seulement pour ERA5 dans ce dernier cas).

Tableau 2.1 Caractéristiques principales de ERA5 et ERA5-Land

	ERA5	ERA5-Land
Résolution spatiale	0,25° par 0,25°	0,1° par 0,1°
Domaine spatial	Global (terre + eau)	Global (au-dessus de la terre, excluant les grandes étendues d'eau)
Résolution temporelle	Horaire	Horaire
Domaine temporel (Début)	1950	1950
Domaine temporel (Fin)	Temps quasi-réel (délais de 5 jours)	En continu, délai de 2-3 mois

2.3 Secteur à l'étude

Le secteur à l'étude comprend l'ensemble de l'Amérique du Nord avec des bassins versants situés au Canada, aux États-Unis et au Mexique (Figure 2.1). Dans le but d'augmenter la transposabilité de l'analyse, un total de 1169 bassins a été sélectionné. Ils sont à leur tour éparpillés parmi 23 zones climatiques selon les régions de Bukovsky (Bukovsky, 2011). L'ampleur du territoire observé permet ainsi l'inclusion d'une panoplie de conditions physiques, météorologiques et hydrologiques variées.

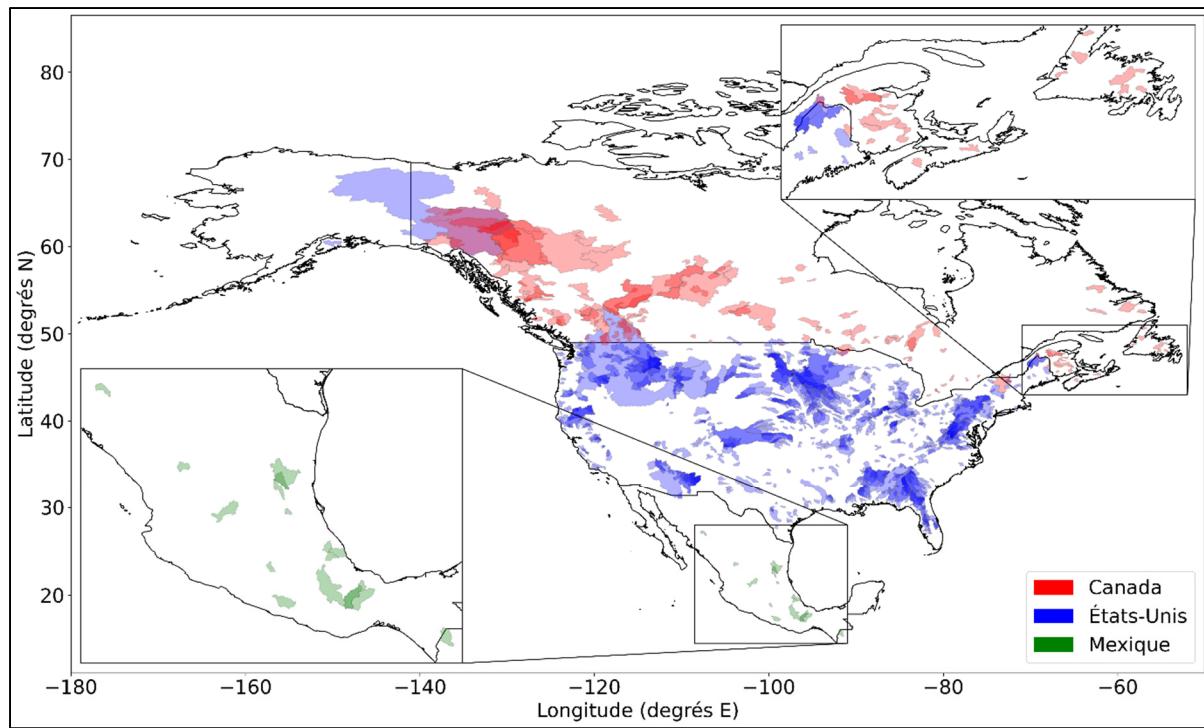


Figure 2.1 Localisation des bassins versants en Amérique du Nord

2.3.1 Zones climatiques

Les régions climatiques de Bukovsky résultent d'une régionalisation du territoire nord-américain, plus spécifiquement du Canada et des États-Unis. Elles sont une contribution faite au programme international « North American Regional Climate Change Assessment Program » (NARCCAP). Cette sous-régionalisation du nord de l'Amérique permet d'aider à l'analyse des simulations climatiques régionales que crée le NARCCAP (Bukovsky, 2011). Bukovsky (2011) décrit que ce produit découle d'une simplification et combinaison des travaux de Ricketts et al. (1999) et de Kampe et al. (2010) qui avait précédemment fait des séparations du territoire afin de pouvoir capturer la variation du climat à un niveau plus régional. La Figure 2.2 montre cette régionalisation accompagnée des centroïdes des bassins versants à l'étude (représentés par les points noirs).

Il y a au total 29 régions qui peuvent aussi être groupées en 13 plus grandes surfaces, combinant des régions présentant des caractéristiques similaires (Bukovsky, 2011). Aux fins de la

présente étude, l'option de base à 29 régions est utilisée. La délimitation des régions est basée sur une grille de 0.5° par 0.5° en longitude et latitude, générant une séparation inexacte pour les régions côtières. Certaines côtes ont un débordement d'une surface d'eau à l'intérieur de la frontière, alors que d'autres ont un débordement terrestre en dehors de la frontière (Bukovsky, 2011). Une modification manuelle a été apportée pour corriger cinq bassins dont la localisation de leur centroïde par rapport aux régions climatiques de Bukovsky indiquait qu'ils étaient dans une étendue d'eau. À cette modification manuelle s'ajoutent aussi 13 autres bassins qui se retrouvaient dans deux régions en même temps. Dans ce cas, la région ayant la plus grande superficie du bassin est conservée. Le tableau 2.2 présente les proportions de bassins versants présents dans chaque région. Comme ces régions ne couvrent pas l'entièreté du Mexique, tous les bassins mexicains non catégorisés sont classés dans la région -1.

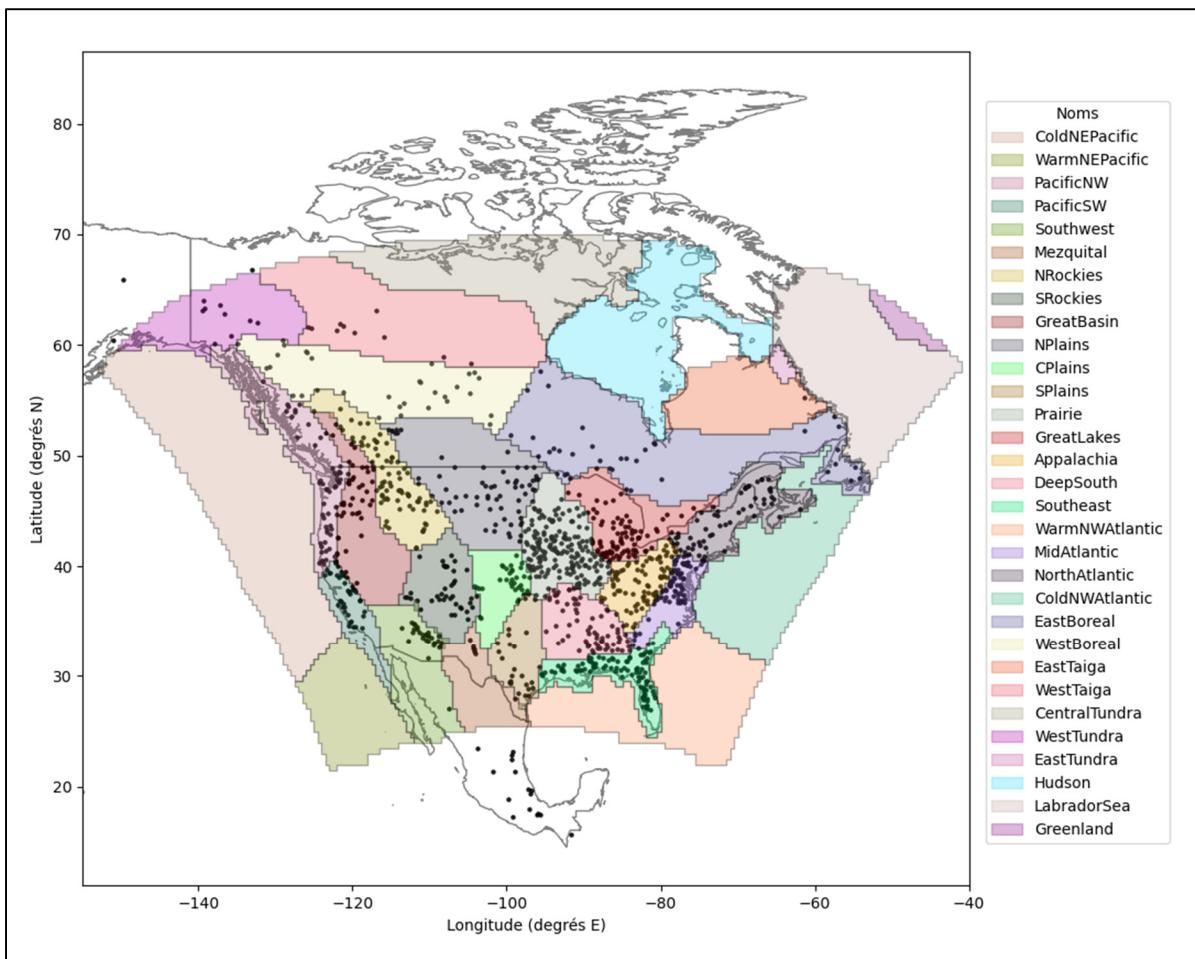


Figure 2.2 Localisation des bassins versants par rapport aux zones climatiques de Bukovsky (2011). Le code de couleur, sur la figure, réfère aux différentes régions climatiques.

Tableau 2.2 Distribution des bassins versants à travers des zones climatiques de Bukovsky.

Nom	Proportion de basins		Nom	Proportion de basins	
	Quantité	%		Quantité	%
PacificNW	61	5,22	Appalachia	86	7,36
PacificSW	40	3,42	DeepSouth	69	5,9
Southwest	30	2,57	Southeast	98	8,38
Mezquital	7	0,6	MidAtlantic	77	6,59
NRockies	77	6,59	NorthAtlantic	56	4,79
SRockies	39	3,34	EBoreal	35	2,99
GreatBasin	40	3,42	WBoreal	31	2,65
NPlains	62	5,30	EastTaiga	1	0,09
CPlains	35	2,99	WestTaiga	10	0,86
SPlains	23	1,97	WestTundra	12	1,03
Prairie	175	14,97	Mexique	16	1,37
GreatLakes	89	7,61			

2.3.2 Domaine de l'étude

Les 1169 bassins ont été sélectionnés à partir de la base de données HYSETS. Précisément, les bassins versants sont distribués sur le territoire nord-américain le plus également possible en fonction des critères de sélection présentés dans le prochain chapitre. Par exemple, des critères de sélection importants sont une aire minimale de 500 km², une série de données de débits la plus complète possible ou encore une répartition la plus égale possible dans chaque zone climatique.

Au final, il y a 173 bassins au Canada, 979 aux États-Unis et 17 au Mexique, soit respectivement une proportion équivalente de 14,8 %, 83,7 % et 1,5 %. Les caractéristiques, autant hydrométéorologiques que physiographiques, sont très variées. Entre autres, les aires varient entre 500,84 km² et 613 824 km² et les localisations vont de 15,69°N à 66,79°N en latitude et -151,08°E à -54,28°E en longitude. L'ensemble des distributions des caractéristiques

du domaine d'étude est présenté dans le tableau 2.3 qui se retrouve à la fin de ce chapitre. À noter que l'occupation du sol est présentée en pourcentage de la superficie totale d'un bassin versant et que les distributions de température, précipitation et débits sont toutes issues des données d'observation.

2.3.2.1 Caractéristiques physiographiques

En termes de caractéristiques physiographiques, les États-Unis se démarquent des deux autres pays par le grand nombre de bassins qui s'y retrouvent (figure 2.1). Cet effet est en partie dû au critère de sélection vis-à-vis des zones climatiques. Les États-Unis comptent plus de zones climatiques que le Canada et elles sont aussi plus petites, affectant directement la sélection des bassins versants. Quant au Mexique, le petit nombre de bassins est issu de deux facteurs : (1) la qualité des données est moindre, excluant dès le départ beaucoup de bassins et (2) le nombre total de bassins mexicains présent dans HYSETS est de seulement 46 alors qu'il y en a 2 375 pour le Canada et 12 004 pour les États-Unis.

La grandeur du territoire affecte directement l'étendue des tailles des bassins versants. Le Canada et les États-Unis offrent une étendue plus large comparativement au Mexique qui n'a aucun bassin plus grand que 12 000 km² dans la base de données HYSETS. Du côté des deux autres pays, le Canada compte 30 bassins plus grands que cette aire et les États-Unis 84. En moyenne, la distribution montre que les bassins versants canadiens sont plus grands que ceux des États-Unis. En effet, les valeurs des 25^e, 50^e et 75^e percentiles sont toutes supérieures à celles des États-Unis. La plus forte asymétrie positive pour ce pays peut probablement être attribuée au nombre de bassins américains qui est beaucoup plus vaste, surtout si l'on tient compte du fait que dans la base de données HYSETS, plusieurs bassins versants sont en fait des sous-bassins de bassins plus grands.

La variabilité de l'occupation du sol est aussi très grande. En considérant un regroupement des catégories d'occupation du sol en végétation (forêt, arbustes, prairies), zones urbaines, zones agricoles et en eau (étendues d'eau, neige et glace), on obtient les distributions présentées à la

figure 2.3. On remarque bien l'ampleur de la végétation qui prend en moyenne approximativement 65% de la superficie des bassins, et ce, à travers l'Amérique du Nord au complet. Toutefois, chaque pays offre un aspect différent permettant une comparaison des résultats en différents milieux. Le Mexique se distingue par ses zones agricoles, le Canada se démarque par l'ampleur de ses étendues d'eau, de neige et de glace; ce qui est normal en raison des bassins en zones nordiques et des bassins dans les rocheuses alors que grâce au grand nombre de bassins aux États-Unis, une nette supériorité de zones urbaines est présente.

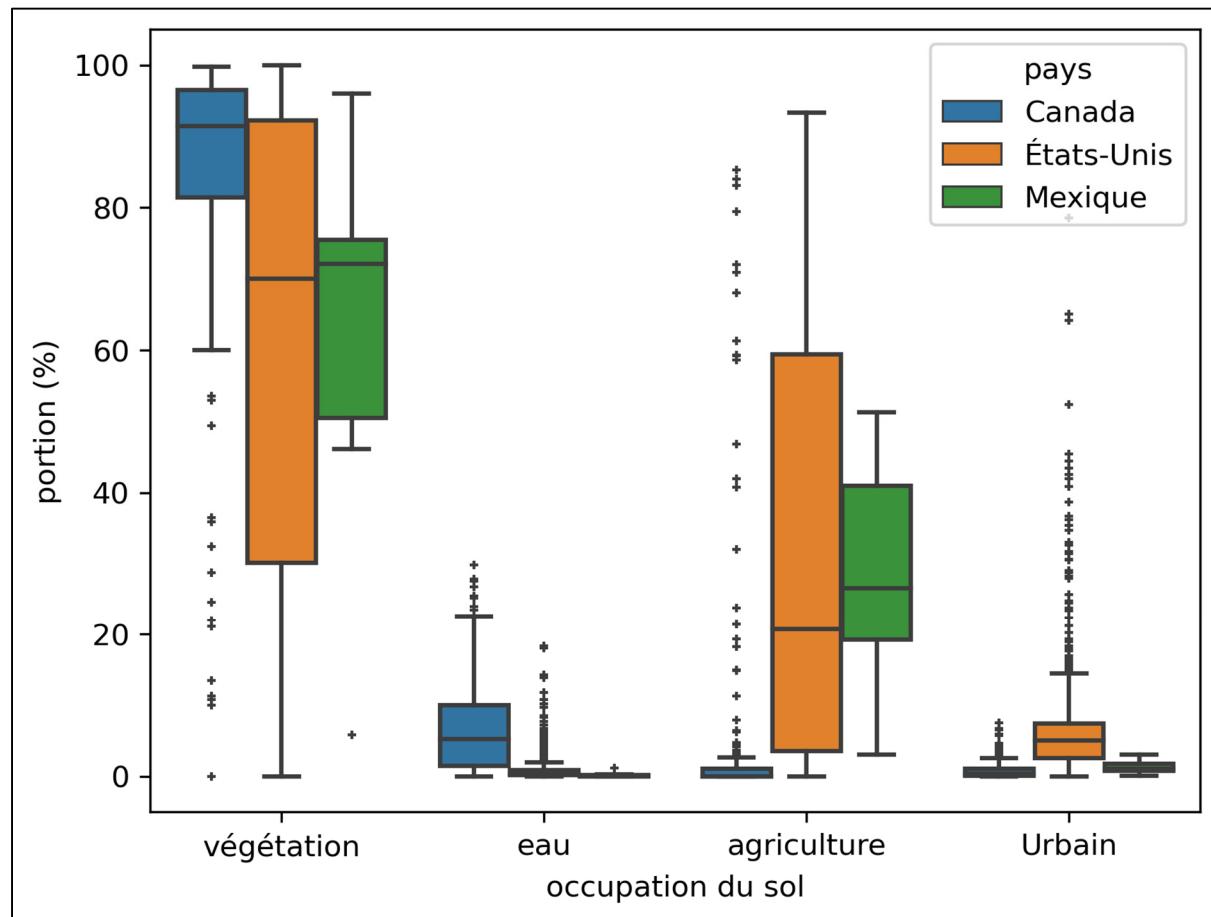


Figure 2.3 Distributions des occupations de sols par pays. La végétation inclut la forêt, la prairie, les arbustes ainsi que les zones humides alors que l'eau comprend aussi la neige et la glace.

2.3.2.2 Caractéristiques météorologiques

Les températures moyennes annuelles pour chacun des pays sont très différentes. Le Mexique ressent très peu de dispersion de température avec un écart de 15,6°C, dont le minimum est de 10,4°C. C'est d'ailleurs l'opposé au Canada, avec des températures plus froides. Au minimum, les températures moyennes sont de -9,1 °C, au maximum de 10,4°C et en général elles naviguent aux alentours de 0,0°C avec une médiane de 1,2°C. Ce sont les États-Unis qui offrent la plus grande dispersion de températures parmi les trois pays. En effet, elles s'étendent sur 27,6°C, débutant sous le seuil de congélation à -2,9°C. L'ensemble des températures varie toutefois peu puisque l'écart interquartile entre le 25^e quartile et le 75^e est de seulement 6,0°C. La moitié des températures se situent donc entre 7,2°C et 13,3°C. De par les distributions de températures associées à chaque pays, il est possible d'en déduire que les précipitations seront plutôt solides pour le Canada, mélangées pour les États-Unis et uniquement liquides pour le Mexique. (Tableau 2.4)

Les précipitations annuelles moyennes en provenance des observations (valeurs totales, en mm équivalents d'eau) se ressemblent au niveau des États-Unis et du Mexique. Les distributions de quartiles sont presque les mêmes avec des différences entre les valeurs de seulement quelques centaines de millimètres de différences, à l'exception de l'extrême maximal qui a une différence d'environ 0,5 m. Les plus grosses différences sont avec le Canada. Comme illustré à la figure 2.4 par les précipitations annuelles moyennes issues des données d'observations, les États-Unis et le Mexique reçoivent en moyenne 1,5 fois plus de précipitation annuellement, et ce, parmi la totalité des distributions. La figure 2.4 montre aussi bien la différence de précipitations reçues dans les zones plus humides, soit les zones côtières ainsi que le sud du Mexique qui est plus humide et moins aride que le Nord.

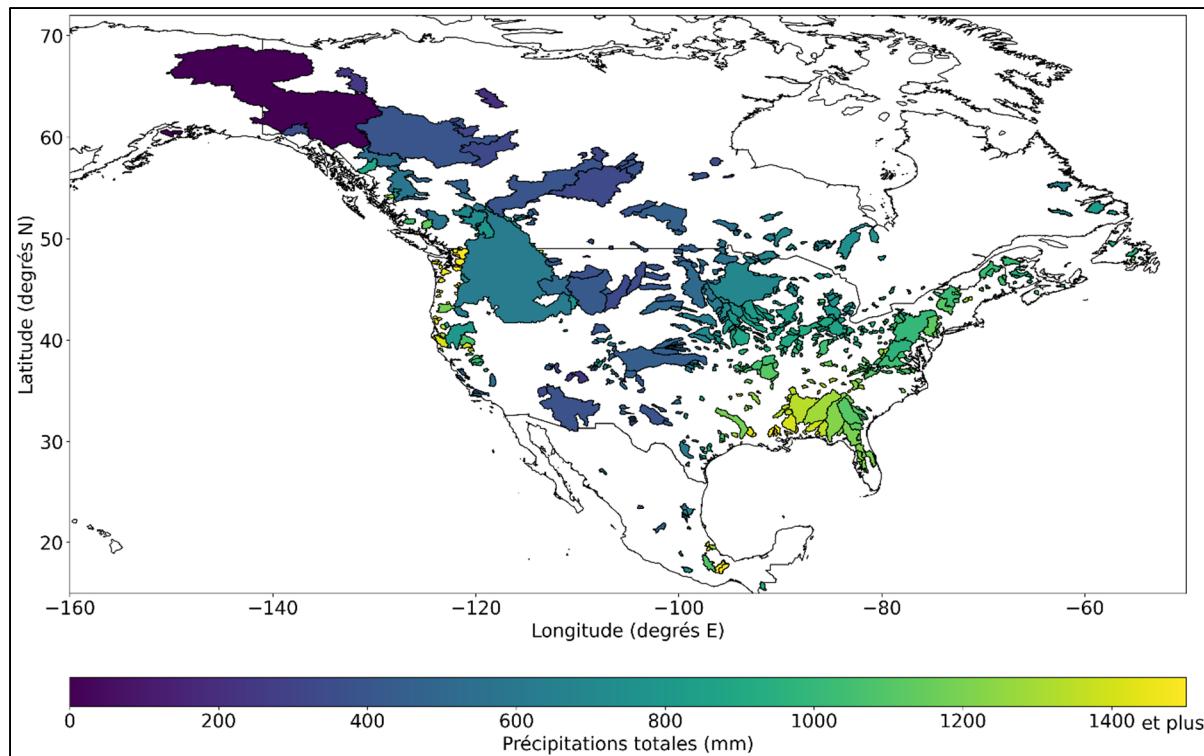


Figure 2.4 Moyennes des précipitations totales annuelles.

L'évapotranspiration annuelle moyenne est quant à elle distincte pour chacun des pays, ce à quoi l'on devrait s'attendre considérant que les données utilisées ont été générées à partir de la formulation de Oudin (Oudin et al., 2005). La formule utilise comme données d'entrée la radiation solaire extraterrestre et la température moyenne journalières donc l'estimation des valeurs se doit d'être supérieure plus au sud qu'au nord considérant plus d'heures de soleil annuellement et des températures moyennes plus élevées. La figure 2.5 présente la distribution d'évapotranspiration potentielle annuelle moyenne selon la formulation d'Oudin (2005). On peut d'ailleurs y voir le net impact de la radiation solaire sur l'évapotranspiration qui est liée à la localisation en latitude des bassins versants.

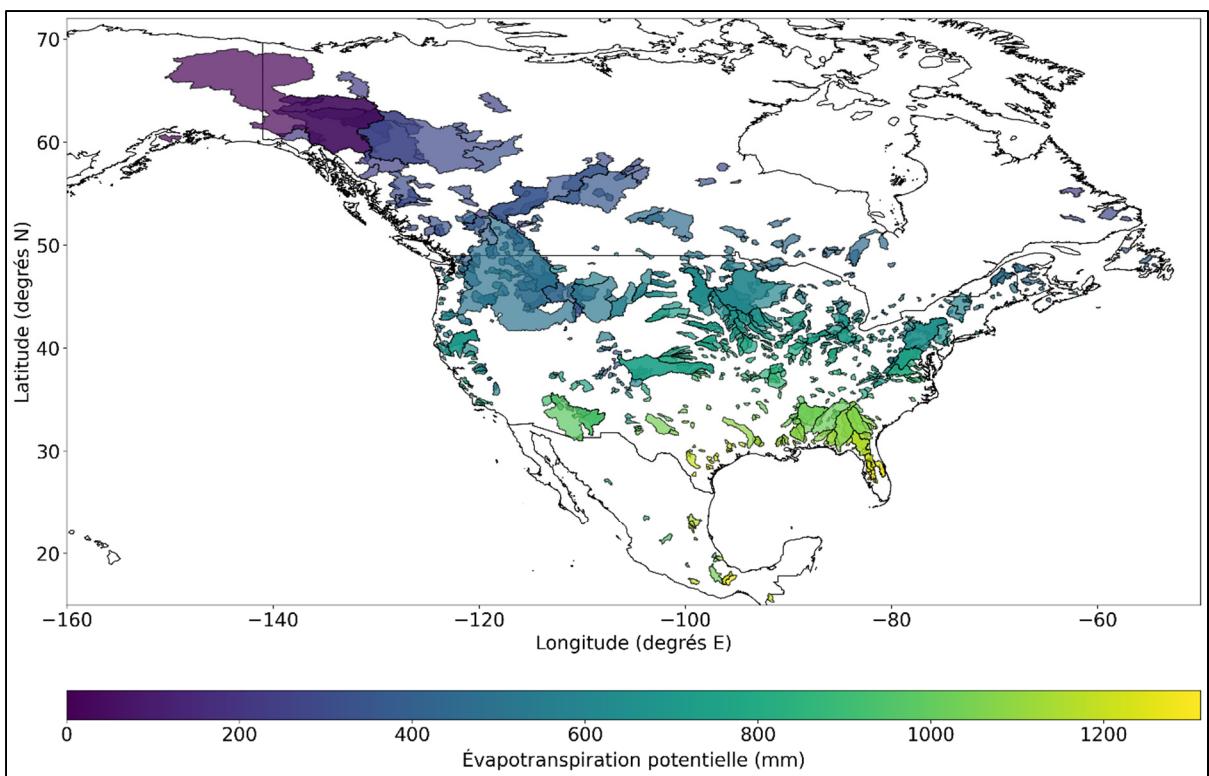


Figure 2.5 Évapotranspiration annuelle moyenne. Les valeurs sont estimées avec la formule de Oudin (2005)

La disponibilité en eau localement pour les bassins versants joue un rôle sur l'équilibre eau-énergie dans l'ET. Selon l'hypothèse de Budyko (Sposito, 2017), il est possible de mettre en relation les précipitations et l'évapotranspiration afin de classifier l'équilibre eau-énergie d'un bassin. Plusieurs adaptations ont été apportées à la classification initiale qui ne fait qu'une distinction entre des conditions mouillées et des conditions sèches (Sposito, 2017). Ici, les classes apportées par T. McVicar et al. (2012) sont utilisés. Le principe repose sur le calcul de l'index de sécheresse, c'est-à-dire le ratio à long terme de l'évapotranspiration annuelle moyenne par rapport aux précipitations annuelles moyennes. La classification des bassins est montrée à la figure 2.6. Les différentes classes sont : limité en énergie (conditions mouillées), équitant (conditions équilibrées) et limité en eau (conditions sèches). Les précipitations jouent un rôle important dans cette classification. Le tout est bien montré par le classement des bassins situés plus loin des zones côtières et plus au Sud comme étant limités par l'eau (sec) ou équitant

alors que les bassins sur les côtes et plus au Nord (moins de radiation) sont classés comme équitant ou limités par l'énergie (mouillé).

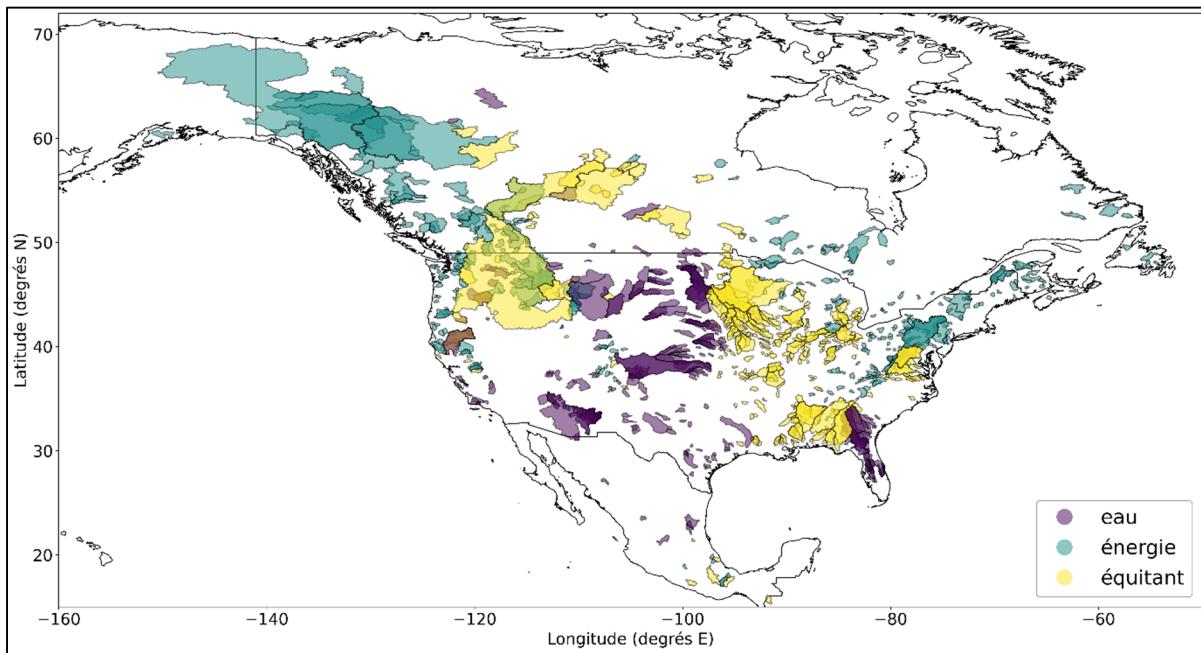


Figure 2.6 Carte l'équilibre eau-énergie dans l'évapotranspiration pour chaque bassin versant.

2.3.2.3 Caractéristiques hydrologiques

Les distributions des débits annuels sont assez variées. Alors que les maximums sont tous différents, et ce, avec plusieurs milliers de mètres cubes par seconde les séparant, il est possible de remarquer des valeurs plutôt semblables entre les États-Unis et le Canada. La variance dans les débits est difficilement explicable en raison des nombreuses caractéristiques différenciant les bassins versants, entre autres l'occupation du sol ou encore leur superficie.

Au niveau du régime hydrique, on peut voir plusieurs ressemblances entre les trois pays. Dans l'est du Canada et des États-Unis, surtout plus au Nord, il est régulier d'observer des régimes qui sont plus constants en raison de précipitations plus régulières annuellement, autant solides que liquides. Au contraire, en zone plus centrale des États-Unis ainsi qu'au Mexique, les

régimes hydriques sont beaucoup moins constants, même inexistant, par endroit pendant la majorité de l'année. Par exemple, le centre sud des États-Unis ainsi que le Mexique central et nordique sont des régions très arides, souvent en temps de sécheresse et donc les cours d'eau disparaissent entièrement jusqu'à ce qu'il y ait de nouvelles précipitations.

Les conditions météorologiques affectent directement la constance du régime hydrique, mais aussi son origine. La figure 2.7 montre une carte de la dominance du régime hydrique par bassin versant. La dominance du régime hydrique a été déterminée en se basant sur les graphiques de coefficients mensuels, issus des données d'observation, présentés dans les travaux de Zeiringer, Seliger, Greimel, et Schmutz (2018). Comme attendu, les régions au sud ont un régime dominé par la pluie puisqu'il ne fait pas assez froid pour recevoir des précipitations solides, alors que plus au nord, les régimes peuvent être dominés par un mélange de pluie et de neige ou encore seulement par la neige. À noter que certains bassins versants semblent être situés dans des régions chaudes et plus arides et montrent un régime dominé par la neige. Ces cas sont des bassins versants situés en zones montagneuses recevant de la neige en hiver en haute élévation. C'est d'ailleurs le cas pour les régions telles que le Colorado, le Nouveau-Mexique ou encore la Californie. Le régime hydrique dans ces cas se voit donc dominé par la période de fonte de neige en élévation.

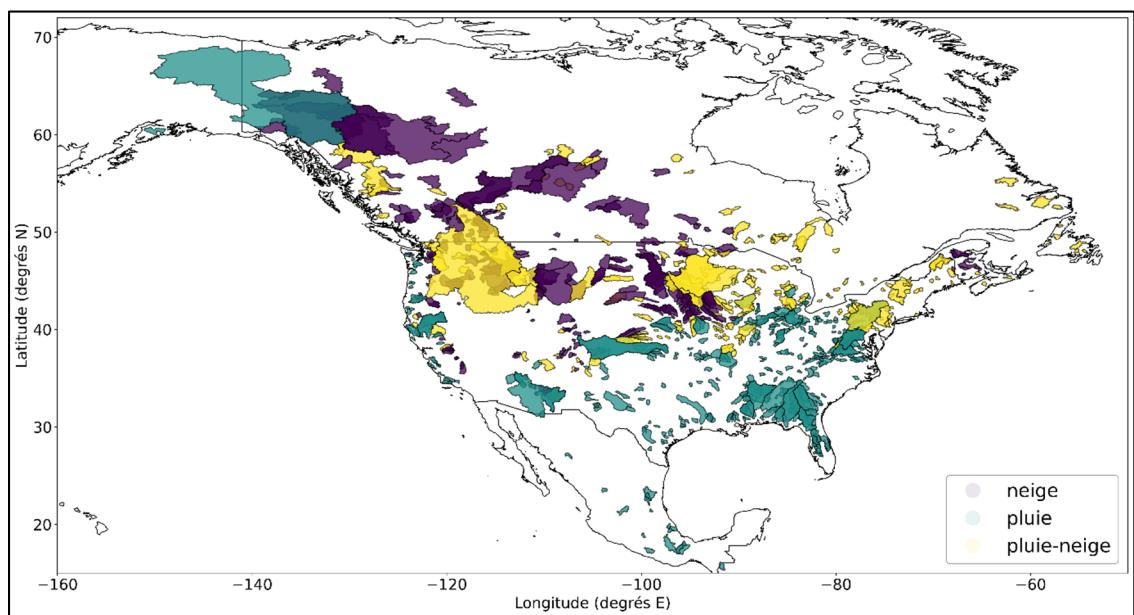


Figure 2.7 Carte de la dominance du régime hydrologique

Tableau 2.3 Distributions des principales caractéristiques des bassins versants

	Aire (km ²)	Latitude (°N)	Longitude e (°E)	Q (m ³ /s)	P (mm)	T (°C)	Occupation du sol (%)					
							F	G	ZH	W	U	S
Canada												
Minimum	501	42,83	-139,50	0	0	-9,1	0,1	0	0	0,08	0	0,01
25 ^e percentile	1242	49,24	-121,61	16	422,89	-1,3	46,92	2,08	0,08	0,97	0,08	3,80
Médiane	2847	51,85	-114,34	38	637,9	1,2	64,45	8,77	0,44	2,69	0,39	7,32
75 ^e percentile	8574	55,65	-86,30	104	935,9	2,8	75,77	21,50	3,19	7,43	1,11	11,78
Maximum	274 164	66,79	-54,28	3463	3053,5	10,4	96,56	75,98	39,58	27,86	7,47	47,41
États-Unis												
Minimum	502	26,93	-151,08	0	0	-2,9	0,01	0,01	0	0	0	0
25 ^e percentile	895	36,29	-104,71	6	621,6	7,2	8,55	0,92	0	0,15	2,55	0,07
Médiane	1637	39,95	-91,98	14	937,1	10,1	35,61	3,14	0	0,39	5,04	2,19
75 ^e percentile	3662	42,82	-84,09	33	1207,6	13,3	60,30	9,60	0,02	0,87	7,37	15,02
Maximum	613 824	65,86	-67,94	7574	5263,3	24,7	95,48	88,70	0,25	18	78,52	87,1
Mexique												
Minimum	661	15,69	-107,46	0	0	10,4	2,93	1,24	0	0	0,11	0,16
25 ^e percentile	1894	17,63	-99,41	12	593,5	16,5	28,32	4,07	0,27	0,2	0,78	3
Médiane	2579	19,63	-98,99	49	927,7	20,5	42,43	11,86	1,1	0,09	1,08	4,61
75 ^e percentile	5089	22,47	-96,74	108	1527,5	21,5	55,57	26,76	4,87	0,17	1,73	5,79
Maximum	11 982	27,05	-91,71	3079	4739,3	26,0	85,45	48,24	62,54	1,19	3,05	20,14
												51,27
												0

* Dans le tableau, Q représente le débit volumique annuel moyen, P les précipitations totales annuelles moyennes et T les températures annuelles moyennes. Pour l'occupation du sol, F représente la forêt, G l'herbe, ZH les zones humides, W l'eau, U les zones urbaines, S les arbustes, C les zones agricoles et SI la neige et la glace.

CHAPITRE 3

MÉTHODOLOGIE

Le suivant chapitre présente les méthodes et outils utilisés afin de réaliser l'étude. Une présentation générale des grandes étapes pour obtenir les débits simulés est d'abord faite, suivie par une plus ample explication des parties intrinsèques de ces dernières.

3.1 Étapes générales de l'étude

L'étude a nécessité la réalisation de plusieurs étapes secondaires avant de s'attaquer à l'objectif principal. Il faut préparer tous les intrants nécessaires aux modèles hydrologiques et puis les calibrer. La figure 3.1 représente un organigramme de toutes les étapes nécessaires à la compléTION de l'étude.

Les intrants comprennent l'ensemble de toutes les données qui sont fournies à la chaîne de modélisation hydrologique, soient les données hydrométéorologiques et les données physiographiques. La majeure partie du temps, elles ne peuvent être directement utilisées pour la modélisation et il est donc nécessaire d'y appliquer un traitement. Le cas échéant, l'utilisation de HYSETS a simplifié la manipulation de données puisque la totalité des données physiographiques ainsi qu'une partie des données hydrométéorologiques sont déjà traitées. La majorité du traitement de données est fait pour les données de réanalyses de ERA5 et de ERA5-Land.

Une fois les intrants prêts, le processus de modélisation peut être effectué. À cette étape, certains choix sont à faire quant aux modèles, aux algorithmes de calibration et aux critères jugeant la performance des modèles calibrés. Puisque l'étude repose sur l'exploration de différents types de données et de différentes formulations d'estimation de l'évapotranspiration,

12 différentes formules sont utilisées, auxquelles s'ajoutent deux sources de données fournissant des données directes d'ET, donc ne nécessitant aucune formule. Pour réduire le bruit dans les résultats, deux modèles hydrologiques différents sont utilisés avec un algorithme de calibration et un critère de performance sur les débits simulés. Cette chaîne de modélisation est aussi répétée une seconde fois pour assurer une uniformité des résultats. En totalité, plus de 180 000 calibrations sont faites si l'on considère toutes les combinaisons possibles de données, de formulations d'évapotranspiration et de modèles.

La dernière grande étape concerne les extrants. Ces derniers sont les débits simulés issus de la meilleure des deux calibrations, pour chaque bassin. En raison du nombre de calibrations, et par le fait même du temps de calcul nécessaire pour compléter le tout, les débits simulés lors de la modélisation ayant donné les meilleurs paramètres sont utilisés pour le restant de l'étude (voir la section 3.4).

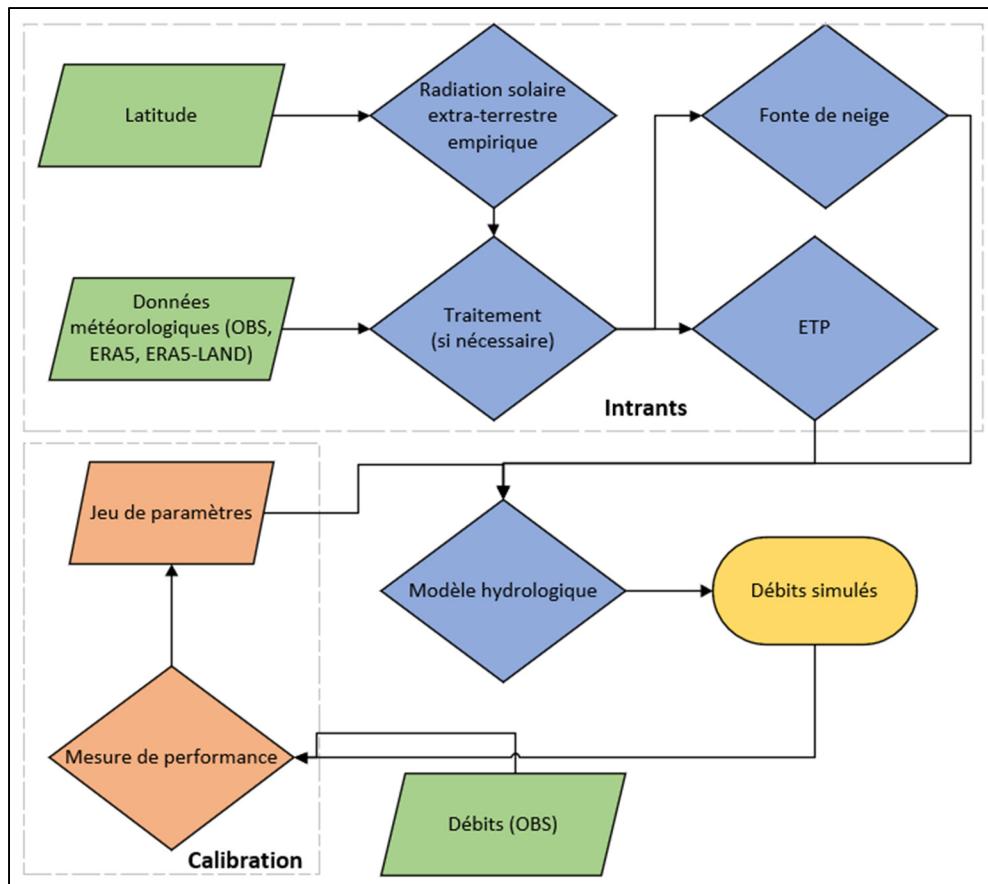


Figure 3.1 Organigramme des étapes générales de l'étude

3.2 Intrants

Pour cette étude, les intrants sont nombreux et importants. Les bassins versants ont d'abord été choisis afin de délimiter l'ampleur du projet. À partir des bassins versants, il est ensuite possible d'extraire les données hydrométéorologiques qui y sont associées et de les traiter pour qu'elles soient utilisables en modélisation hydrologique. Les prochaines sous-sections décrivent plus amplement ces processus.

3.2.1 Choix des bassins versants

Il a été décidé d'utiliser la base de données HYSETS (Arsenault et al., 2020). Elle est relativement nouvelle et inclut beaucoup d'éléments désirés pour cette étude, notamment la

disponibilité des données des réanalyses ERA5 et ERA5-Land ainsi qu'une couverture spatiale nord-américaine. Elle est donc le point de référence de toutes les données du projet et donc le choix des bassins versants en est issu.

HYSETS comprenant 14 425 bassins, il fallait donc en couper la majorité par souci du temps de calcul. En premier lieu, un filtre au niveau de l'aire des bassins et l'exhaustivité des séries temporelles de données sont appliqués. Le filtre pour l'aire est fixé à une aire minimale 500 km². Ce critère permet une meilleure homogénéité des caractéristiques physiques et hydrologiques, mais surtout d'assurer que le temps de concentration de chaque bassin est supérieur au pas de temps journalier des données. Le deuxième critère sur les données à arbitrairement fixé à un maximum de 40% de données manquantes. Ce filtre est seulement pour assurer une majorité de données présentes. De ça, tous les bassins mexicains ont été sélectionnés puisqu'il n'en restait que seulement 17. La deuxième partie de la sélection des bassins ne s'est faite que sur les bassins canadiens et américains. Une nouvelle coupure est faite au niveau du nombre de données manquante dans les séries temporelles. Seulement les bassins ne manquant pas de données ont été conservés

Au final, l'application de ces filtres, a résulté en un nombre de 1169 bassins versants. Après vérification manuelle, la sélection de bassins à ce point assurait aussi une distribution acceptable équitable de bassins dans chaque zone climatique, sauf pour un débordement dans les prairies.

3.2.2 Données générées

Certaines données doivent parfois être calculées puisqu'elles sont indisponibles, non recensées ou qu'il n'est pas nécessaire de les comptabiliser. C'est le cas pour certaines données solaires. Dans la présente étude, les données non recensées sont l'angle du coucher du soleil et le nombre d'heures journalières d'ensoleillement alors que les données non disponibles sont la radiation solaire extra-terrestre pour les données d'observations ainsi que pour ERA5-Land.

Pour la génération de ces données, deux techniques ont été utilisées en fonction du type original de données.

Le premier élément à calculer est l'angle du coucher du soleil ω , plus exactement le nombre de degrés de rotation de la Terre dans l'intervalle de temps entre midi et l'heure du coucher du soleil. Cette variable est nécessaire pour déterminer le nombre d'heures journalier d'ensoleillement ainsi que pour calculer la radiation solaire extra-terrestre, des variables nécessaires pour les formules d'ETP utilisant la radiation solaire. Cette valeur dépend de la latitude ainsi que de la déclinaison solaire. L'équation 3.1 est une version journalière de l'équation présentée par Morton (1983) permettant le calcul de cet angle en radians :

$$\omega = \arccos\left(-\tan\left(\frac{L \cdot \pi}{180}\right) \cdot \tan(\delta_j)\right) \quad (3.1)$$

où L est la latitude du centroïde des bassins et δ_j la déclinaison solaire. Quant à elle, la déclinaison solaire se calcule par l'équation 3.2 ci-dessous (Fortin et Turcotte, 2007) :

$$\delta_j = 0.41 \cdot \sin\left(\frac{j - 80}{365} \cdot 2\pi\right) \quad (3.2)$$

où j est le jour julien. Le nombre d'heures journalier d'ensoleillement DL s'obtient par l'utilisation de l'équation 3.3 (Fortin et Turcotte, 2007) :

$$DL \approx \frac{24}{\pi} \cdot \omega \quad (3.3)$$

Comme ce calcul dépend de la déclinaison solaire et par le fait même de la latitude et du jour julien, il est possible d'en conclure qu'annuellement, les valeurs sont les mêmes pour une même latitude. Il en est d'ailleurs de même pour la radiation solaire extra-terrestre G_E (en MJ) qui se calcule avec l'équation 3.7 (Morton, 1983). Les équations 3.4 à 3.6 sont des angles et des fonctions nécessaires à la simplification et au calcul de l'équation 3.7. À noter aussi que la

version originale dans Morton (1983) est établie pour des estimations mensuelles en Wm^{-2} . Ici les formules sont adaptées afin de leur donner un pas de temps journalier et un résultat en MJ :

$$\cos Z = \max_{0,001 \leq \cos Z} \cos\left(\frac{L}{57,3} - \theta\right) \quad (3.4)$$

$$\cos z = \cos Z + \cos\left(\frac{L}{57,3}\right) \cdot \cos \theta \cdot \frac{\sin \omega}{\omega - 1} \quad (3.5)$$

$$\varphi = 1 + \frac{\cos\left(\frac{L}{58,1}\right)}{30} \quad (3.6)$$

$$G_E = 446 - \theta \cdot \varphi \cdot \cos z \cdot 0,0864 \quad (3.7)$$

où Z et z sont respectivement les distances angulaires entre le zénith et le soleil à midi et en moyenne; L est la latitude du centroïde du bassin; θ est la déclinaison solaire, δ est donné par l'équation 3.2; ω est l'angle du coucher de soleil (équation 3.1) et φ le vecteur de diamètre du soleil.

3.2.3 Extraction des données

En ce qui concerne les données d'observation, elles sont tirées directement de HYSETS puisqu'elles sont déjà dans un format fait pour la modélisation hydrologique et disponible par bassin. Pour toutes les autres données, soit les données en provenance de ERA5, ERA5-Land et les données concernant le soleil (angle du coucher, heure d'ensoleillement et radiation extraterrestre), elles sont toutes extraites à partir de grilles de points couvrant l'ensemble de l'Amérique du Nord. Pour extraire les données, les géométries des bassins versants sont utilisées afin de déterminer les points des grilles qui sont à l'intérieur des contours de ces derniers.

3.2.4 Préparation des données

Pour alimenter les modèles hydrologiques, les séries temporelles doivent avoir le format nécessaire. Pour cette étude, le format de base est une série unique par bassin, pour chaque variable hydrométéorologique, sous un pas de temps journalier. Cette section explique le traitement apporté aux données afin de les mettre sous le format standard de l'étude. Selon le type et la source des différentes données (observations ou réanalyses), les techniques de traitement varient. Dans le cas les données d'observations météorologiques, elles n'ont eu besoin d'aucun traitement étant donné qu'elles proviennent directement de HYSETS.

3.2.4.1 Données d'observations hydriques

Ces données proviennent de HYSETS. Pour celles-ci, le seul traitement nécessaire est un changement d'unité. À l'origine, les séries temporelles sont fournies en m^3/s alors que les modèles utilisés ne tiennent pas compte de l'aire des bassins versants, les débits simulés sortant des modèles sont une lame d'eau en mm. Par défaut, puisque le pas de temps est journalier, la lame d'eau est donc en mm/j pour les modèles hydrologiques utilisés dans l'étude. Pour faire le changement d'unités, les séries de données ont été divisées par l'aire respective de chaque bassin puis multipliées par 86 400 secondes.

3.2.4.2 Données météorologiques de réanalyses

Les données de réanalyses ont subi deux traitements distincts, soit une agglomération temporelle suivie d'une agglomération spatiale. L'agglomération temporelle est réalisée de deux façons différentes selon si les variables sont de type instantané ou accumulé. Les variables instantanées sont toutes celles qui sont prises à un temps précis, par exemple la température à 12h, alors que l'on mesure une accumulation pendant une période de temps pour celles accumulées, par exemple les précipitations totales de 12h à 13h. Le tableau 3.1 présente la liste des variables ainsi que leur type de recensement.

Tableau 3.1 Liste des variables par type de recensement

Variables avec mesures instantanées	Variables avec mesures accumulées
Température de l'air à 2 m	Évaporation
Température du point de rosée à 2 m	Évaporation potentielle
Composante V de la vitesse du vent à 10 m	Précipitation totale
Composante U de la vitesse du vent à 10 m	Radiation solaire nette à la surface
Pression à la surface	Radiation solaire extraterrestre

Tel que mentionné à la section 2.2.2, les données de réanalyses sont fournies au temps d'UTC. Donc l'ensemble des points de grilles ont été décalés dans le temps afin de rendre les données à l'heure locale. Le décalage horaire de chaque point a été déterminé au préalable afin d'assurer que chaque point reçoive le décalage approprié. L'application du décalage permet l'agglomération temporelle adéquate. Pour les variables instantanées, le passage du pas de temps horaire au pas de temps journalier nécessite d'effectuer la moyenne journalière des 24 valeurs horaires. Dans le cas de la température de l'air, les minimums et maximums journaliers sont aussi enregistrés. En ce qui concerne les variables accumulées, il y a une distinction à faire entre ERA5 et ERA5-Land. Pour ERA5, l'accumulation est horaire, c'est-à-dire que les données horaires fournies ne comprennent que l'accumulation pour la période de 1h. Ces données nécessitent donc une simple addition des 24 valeurs journalières afin d'obtenir l'accumulation totale de chaque jour. Du côté de ERA5-Land, les données sont cumulatives de façon journalière, avec une mesure prise chaque heure. Cela veut dire qu'il faut faire la différence entre les valeurs horaires consécutives dans le but de décumuler les données, puis additionner le tout journalièrement après l'application du décalage horaire. Pour l'agglomération spatiale, une moyenne est effectuée, chaque jour, sur l'ensemble des points de grilles à l'intérieur des limites des bassins pour ne conserver qu'une seule série temporelle par bassin.

Le dernier traitement se fait sur les données de vent. Le vent suit généralement un profil logarithmique en fonction de la hauteur et de l'occupation du sol. Cela sous-entend que le vent a une vitesse plus grande plus l'altitude est haute puisqu'il y a moins de friction. La formulation

d'ETP modifiée de la FAO nécessite que les mesures de vent soient à une hauteur équivalente aux données de température, dans le présent cas à une hauteur de 2m. Par contre, les données de ERA5 et ERA5-Land sont données à une hauteur de 10m et en deux composantes. La composante u représentant la vitesse des vents en direction de l'est (d'ouest en est) et la composante v est pour les vents en direction nord (du sud au nord). Il est donc nécessaire d'ajuster la vitesse des vents et de combiner les composantes pour obtenir un seul vecteur. L'ajustement de la vitesse est fait avec un facteur de conversion proposé par la FAO. Il tient compte de la hauteur au-dessus d'une surface d'herbe courte. La valeur du vent ajustée se détermine ainsi en multipliant le facteur de conversion à la valeur de vitesse tel qu'indiqué par l'équation 3.8 :

$$u_2 = u_z \frac{4,87}{\ln(67,8z - 5,42)} \quad (3.8)$$

où z est la hauteur au-dessus de la surface et u_z est la valeur du vent à la hauteur z .

Une fois l'ajustement fait, la vitesse du vent s'obtient en calculant la norme du vecteur formé par les composantes u et v .

3.2.5 Données d'évapotranspiration potentielle

Les données d'ETP pour cette étude sont obtenues de deux manières différentes, la première étant directement des variables en provenance des réanalyses et l'autre par méthode traditionnelle d'estimation avec des formules empiriques. Un des points intéressants des réanalyses ERA5 et ERA5-Land est la disponibilité des variables d'ETP mais aussi d'ETR. Ces deux variables, pour chaque réanalyse, sont directement utilisées comme intrants dans les modèles suivant leur traitement, expliqué à la section précédente. Le reste des données d'ETP sont issues de formulations pour son estimation.

Un total de 12 différentes formules d'ETP sont sélectionnées, ces dernières se classant parmi les types généralement utilisés, soit : (1) à base de température, (2) à base de radiation et (3)

combinatoire. Puisque l'étude relève des travaux de Seiller et Anctil (2016) et de Dallaire et al. (2021), les formules prises pour l'étude sont les mêmes que celles dans Dallaire et al. (2021) auxquelles s'ajoute une formule combinatoire se retrouvant dans Seiller et Anctil (2016). Un but de l'étude est de pouvoir comparer différents types de formules, donc le choix des formules spécifiquement était moins important que d'avoir des formules distribuées dans différentes classes. À noter que pour l'étude, les types de données, c'est-à-dire observations ou réanalyse, ne sont pas mélangés ensemble. Pour cette raison, la rareté des données d'observations pour l'ensemble des bassins comme le vent restreint l'utilisation de la formulation FAO-PM pour laquelle il y a des résultats seulement avec les données de réanalyses. Les formules ainsi que leur classification sont présentées dans le tableau 3.2 ci-dessous.

Tableau 3.2 Classemement et description des équations d'estimation d'ETP

Nom	Équation	Référence
Formules à base de température		
Blaney-Criddle (BC)	$ETP = 0,825 \left(100 \frac{DL}{365 * 12} \right) (0,46T_m + 8,13)$	Blaney & Criddle (1950)
Hamon (H)	$ETP = \left(\frac{DL}{12} \right)^2 \exp \left(\frac{T_m}{16} \right)$	Hamon (1960)
Hydro-Québec (HQ)	$ETP = 0,029718(T_{max} - T_{min}) \exp \left(0,019 \left(\frac{9}{5} (T_{max} - T_{min}) + 64 \right) \right)$	Fortin (2000)
Kharrufa (K)	$ETP = 0,34 \left(100 \frac{DL}{365 * 12} \right) T_m^{1,3}$	Kharrufa (1985)
MOHYSE (M)	$ETP = \frac{4,088}{\pi} \omega * \exp \left(\frac{17,3 * T_m}{238 + T_m} \right)$	Fortin & Turcotte (2006)
Thornthwaite (T)	$ETP = 16 \left(\frac{DL}{360} \right) \left(\frac{10T_m}{I} \right)^K$	Thornthwaite (1948)

	$I = \sum_{n=1}^{12} \left(\frac{T_{mn}}{5} \right)^{1,514}$ $K = 0,49239 + 1,792 * I * 10^{-2} - 0,771 * I^2 * 10^{-4}$	
Formules à base de radiation		
Baier- Robertson (BR)	$ETP = 0,157T_{max} + 0,158(T_{max} - T_{min}) + 0,109R_e - 5,39$	Baier & Robertson (1965)
Hargreaves- Samani (HS)	$ETP = 1000 * 0,0023 \frac{R_e}{\lambda\rho} \sqrt{(T_{max} - T_{min})} (T_m + 17,8)$	Hargreaves & Samani (1985)
Jensen-Haise (JH)	$ETP = 1000 \frac{R_e}{\lambda\rho} \frac{T_m}{40}$	Jensen & Haise (1963)
McGuinness- Bordne (MB)	$ETP = \frac{R_e}{\lambda\rho} \frac{T_m + 5}{68}$	McGuinness & Bordne (1972)
Oudin (O)	$ETP = \frac{R_e}{\lambda\rho} \frac{T_m + 5}{100}$	Oudin et al. (2005a, 2005b)
Formule combinatoire		
FAO Penman- Monteith (FAO-PM)	$ETP = \frac{0,408\Delta R_n + \gamma \left(\frac{900}{T_m + 273} \right) * U_2(e_s - e_a)}{\Delta + \gamma(1 + 0,34U_2)}$	Allen et al. (2005)

Dans le tableau 3.2 ci-dessus, les variables se définissent ainsi :

- DL : nombre d'heures d'ensoleillement par jour (h/j);
- ω : angle du coucher du soleil (rad);
- T_m : Température moyenne journalière ($^{\circ}\text{C}$);
- T_{max} : Température maximale journalière ($^{\circ}\text{C}$);
- T_{min} : Température minimale journalière ($^{\circ}\text{C}$);

- T_{mm} : Température mensuelle moyenne ($^{\circ}\text{C}$);
- R_e : Radiation solaire extra-terrestre ($\text{MJ/m}^2\text{j}$);
- R_n : Radiation solaire nette ($\text{MJ/m}^2\text{j}$);
- U_2 : Vitesse du vent à 2m de la surface (m/s);
- e_s : Tension vapeur à saturation (kPa) (utilisant la pression atmosphérique)
- e_a : Tension vapeur actuelle (kPa) (utilisant la pression atmosphérique)
- Δ : pente de la courbe de tension vapeur (kPa/ $^{\circ}\text{C}$);
- λ : Énergie latente de vaporisation (MJ/kg);
- ρ : Masse volumique de l'eau (kg/m^3);
- γ : Constante psychométrique (MJ/kg);

L'obtention des valeurs dérivées du soleil et des valeurs ajustées de vent sont décrites aux sections 3.2.2 et 3.2.4 respectivement.

À noter que certaines modifications sont prises en compte dans les formules. La formule de BC a normalement un facteur « k » qui se calibre en fonction de la saison et de la végétation. Pour cette étude, la valeur utilisée de 0,825 dans Dallaire et al. (2021) est conservée. Elle équivaut approximativement à la valeur moyenne de « k » selon l'intervalle de possibilités donné dans Seiller et Anctil (2016). L'autre modification est l'omission du flux thermique du sol dans la formule de FAO-PM. Normalement, cette valeur doit être soustraite à l'énergie de radiation solaire nette. Par contre, à un pas de temps journalier, la valeur de ce flux tend vers zéro et est très petite comparativement à la radiation solaire nette. Il peut donc être négligé (R. G. Allen, Pereira, Raes, & Smith, 1998).

3.3 Modélisation hydrologique

Cette section aborde le processus de modélisation hydrologique. L'ensemble du processus s'est fait à l'aide du « HydrOIlogical Prediction LABaratory » (HOOPLA) (Thiboult et al., 2017). Cette boîte à outils favorise la simplicité de la modélisation hydrologique de façon modulaire.

Les différents outils de modélisation utilisés dans HOOPLA sont expliqués dans les prochaines sections.

3.3.1 HOOPLA

HOOPLA est une boîte à outils offrant les possibilités de modélisation hydrologique modulaire créée par Thiboult et al. (2020). À la base, l'application comprend 20 modèles hydrologiques, un modèle de fonte de neige, trois formules d'estimation de l'ETP, deux algorithmes pour la calibration automatique, deux routines d'assimilation de données et 17 fonctions-objectifs de performance. La diversité dans l'offre explique la caractéristique de modularité. L'environnement principal de HOOPLA permet de sélectionner multiples combinaisons de modèles (hydrologique, neige, ETP), d'algorithmes de calibration, et de critères de performance pour ainsi procéder facilement et rapidement à de multiples études. Considérant le nombre de calibrations à faire pour cette étude (plus de 180 000), cet outil permet de sauver beaucoup de temps pour cette situation de recherche.

Afin de rendre l'outil le plus polyvalent possible, certaines modifications ont été apportées aux modèles hydrologiques intégrés afin de tous les rendre globaux et conceptuels et fonctionnels dans le cadre de HOOPLA, le degré de modification variant pour chacun des modèles. Ici, GR4J n'a subi aucune modification alors que MOHYSE est qualifié de « similaire », signifiant que très peu de modifications ont été apportées. Les modifications seront abordées dans la section présentant MOHYSE. (Thiboult et al., 2020)

3.3.2 Modèles hydrologiques

La modélisation hydrologique est un outil utilisé dans le but de caractériser le cycle de l'eau sur une certaine région. Le processus se fait généralement en plusieurs parties, notamment la détermination de l'eau disponible pour l'écoulement (phase de production), suivi de l'évaluation des transferts latéraux de cette dernière dans les sols et les rivières (phase de transfert), puis enveloppant ces calculs, le calage des paramètres libres des modèles. Cet ensemble donne comme résultats des débits simulés qui s'obtiennent donc par l'utilisation de

la suite de formules mathématiques définissant les processus du cycle de l'eau. Cette définition s'avère complexe et souvent plus globale que locale. Les formules internes des modèles sont donc accompagnées de coefficients, appelés paramètres de calage, qui permettent d'optimiser l'application d'un modèle à un site, soit un bassin versant, en particulier.

Pour cette étude, le grand nombre de bassins et la disparité dans la disponibilité des données ont réduit la sélection à des modèles globaux et conceptuels. Les modèles sélectionnés sont MOHYSE et GR4J. Restreindre le choix à une seule classe de modèle empêche malheureusement dans ce cas de tester la transférabilité à d'autres types de modèles. Les prochaines sections expliquent les modèles utilisés. Ces deux modèles sont choisis pour leur simple utilisation et leur efficacité en modélisation conceptuelle globale.

3.3.2.1 MOHYSE

MOHYSE, qui signifie « MOdèle HYdrologique Simplifié à l'Extrême » (MOHYSE) est un modèle global et conceptuel à 10 paramètres (Fortin & Turcotte, 2007). Comme plusieurs modèles hydrologiques, MOHYSE est basé sur un schéma de production qui, bien que similaire à celui d'autres modèles hydrologiques, repose sur une simplification de ses parties intrinsèques (Fortin & Turcotte, 2007). Le schéma comprend deux processus complémentaires.

Le premier processus représente l'échange vertical d'eau. MOHYSE définit cet échange entre trois compartiments, imaginés comme des réservoirs dans lesquels de l'eau s'ajoute et desquels elle s'écoule. Le cheminement vertical correspond donc à toute l'eau de précipitation et de neige qui atteint la surface. Cette eau sera retenue en surface, ruissellera ou s'infiltrera. L'eau infiltrée passe au deuxième compartiment qui représente la zone vadose du sol (zone non saturée). Une fois dans cette zone, une partie de l'eau retourne vers l'atmosphère par évapotranspiration, une partie s'écoulera horizontalement vers l'exutoire et une dernière portion s'écoule vers l'aquifère de façon gravitaire. L'aquifère représente le troisième et dernier compartiment de l'échange vertical. Il est considéré dans MOHYSE que la totalité de

l'eau dans l'aquifère ne peut que s'écouler vers l'exutoire via le processus de transfert horizontal (décrit ci-dessous).

Le deuxième processus définit le transfert horizontal de l'eau, c'est-à-dire la transformation de la lame d'eau (issue des échanges verticaux) qui s'écoule vers l'exutoire. Ce processus est géré par l'utilisation d'un hydrogramme unitaire basé sur une fonction de densité de probabilité gamma. L'hydrogramme est ajustable en forme et en échelle par l'entremise de deux paramètres afin de bien caractériser le laminage de chaque bassin. La figure 3.2 ci-dessous représente le schéma de production de MOHYSE. (Fortin & Turcotte, 2007).

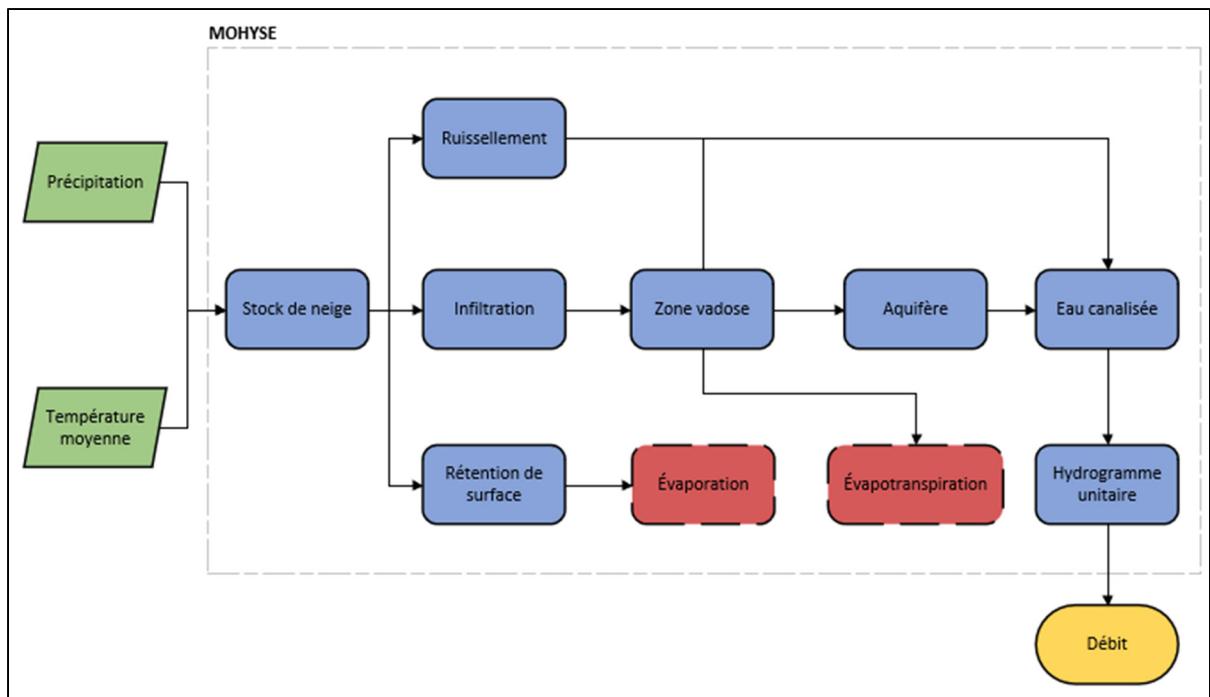


Figure 3.2 Schéma original de production de MOHYSE

À la base, MOHYSE contient dix paramètres, soit huit pour l'ajustement de l'échange d'eau vertical et deux pour l'ajustement de la forme et de l'échelle de l'hydrogramme unitaire (échange horizontal). Pour l'intégration de ce modèle dans HOOPLA, quelques modifications sont nécessaires dans le processus d'échange vertical, plus précisément dans le premier réservoir et deuxième réservoir. HOOPLA étant modulaire dans le choix de ses sous-modèles,

des formules d'ETP et de fonte de neige, le modèle MOHYSE est modifié afin de permettre le forçage de ces valeurs ainsi omettre leur calcul à l'interne même du modèle. Cette action retire donc aussi des paramètres de calibration et réduit le nombre total de paramètres de dix à sept. Les trois paramètres retirés sont : le coefficient d'ajustement de l'évapotranspiration, le taux de fonte de neige et la température seuil pour la fonte de la neige.

L'ETP est aussi une valeur qui peut être forcée dans le modèle. L'estimation de ces valeurs s'ajoute donc comme intrant supplémentaire aux précipitations et à la température. Tout comme l'ETP, l'évolution du couvert de neige est gérée à l'extérieur du modèle, par un autre modèle spécifiquement fait pour ça. Le modèle utilisé est CemaNeige (Valéry, 2010). La figure 3.3 présente le schéma de production modifié de MOHYSE dans HOOPLA. La liste des paramètres peut aussi être consultée à l'annexe I.

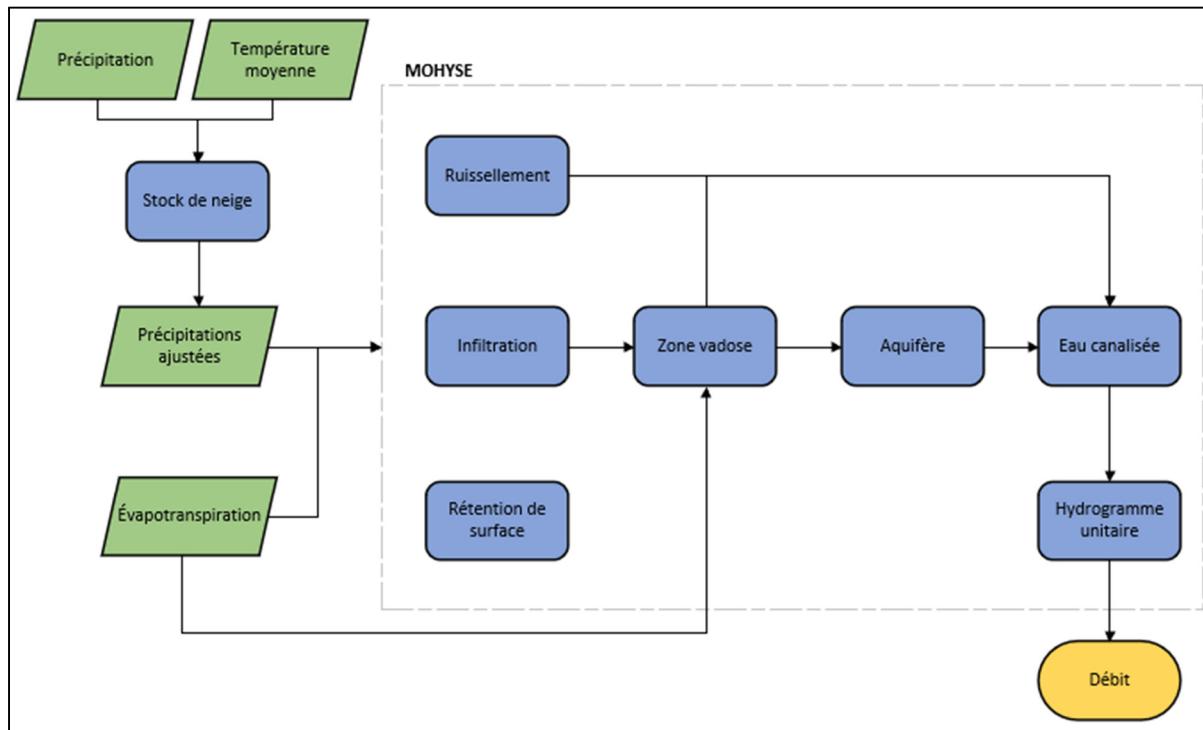


Figure 3.3 Schéma de production du modèle MOHYSE dans HOOPLA

3.3.2.2 GR4J

Le nom du modèle signifie « modèle du Génie Rural à 4 paramètres Journaliers » (GR4J) (Perrin, Michel, & Andréassian, 2003). Tout comme le modèle MOHYSE, GR4J se base sur la simplification des échanges entre des réservoirs. Le modèle comporte quatre paramètres lui permettant son ajustement à différents bassins versants et compte deux intrants. Le premier intrant est les valeurs d'ETP alors que le deuxième est la lame d'eau générée par les précipitations. GR4J ne s'occupe pas de l'évolution du couvert de neige à la base, il est donc utilisé conjointement avec le modèle de neige CemaNeige (Valéry, 2010). L'intrant de précipitation fourni au modèle consiste donc en la précipitation sous forme de pluie et l'eau de fonte consiste en l'extract du modèle de neige, qui est également fourni à GR4J.

La première étape du modèle est l'interception des précipitations et de l'évapotranspiration brutes, dans un réservoir ne considérant aucune capacité de stockage. Les niveaux de précipitations et d'évapotranspiration nets résultant de la neutralisation dépendent simplement des quantités brutes de chaque intrant. Le résultat de l'interception est ensuite neutralisé. Si l'interception génère une précipitation nette, une partie de celle-ci ruisselle vers l'exutoire alors que l'autre partie s'ajoute au réservoir de production. Dans le cas où l'interception nette est de l'évapotranspiration, une portion du réservoir de production se voit vidé. L'échange avec le réservoir de production engendre une percolation d'eau qui contribuera au débit vers l'exutoire, avec la portion de précipitation qui ruissellera. Le routage est basé sur l'utilisation de deux hydrogrammes unitaires. Le premier hydrogramme est utilisé afin d'alimenter le dernier réservoir, celui de routage alors que le deuxième génère un débit qui est routé directement vers l'exutoire. Le débit final est la combinaison du débit extract du réservoir de routage et de celui acheminé par le deuxième hydrogramme, après avoir subi des échanges souterrains. Le modèle n'a pas nécessité aucune modification pour son inclusion dans HOOPLA et est donc utilisé dans sa forme originale. Le schéma de production est présenté à la figure 3.4.

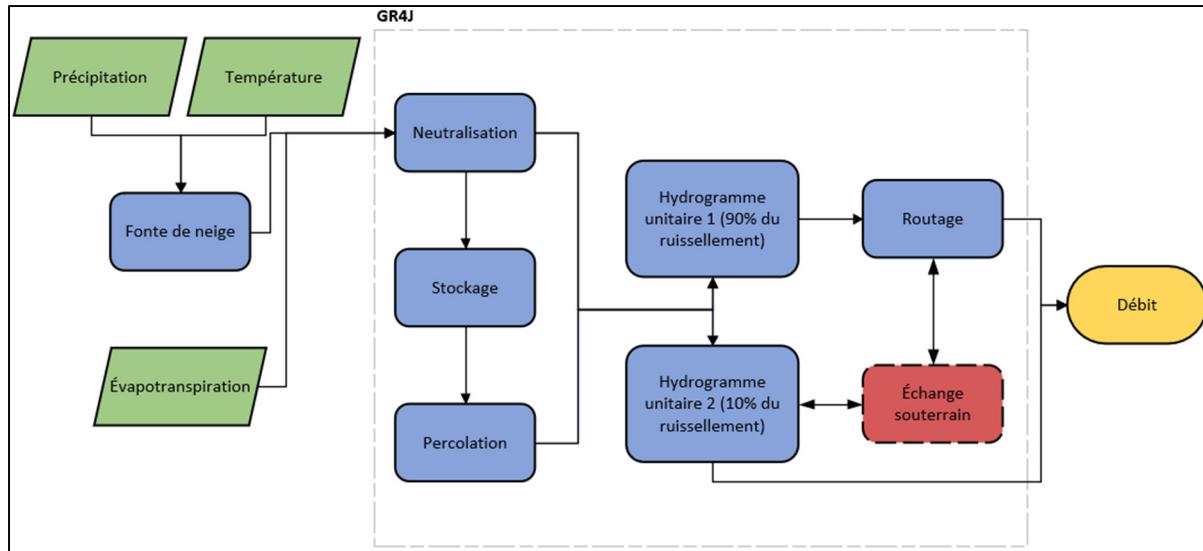


Figure 3.4 Schéma de production de GR4J

3.3.3 Modélisation du couvert neigeux

À l'origine, CemaNeige (Valery, 2010) est un modèle semi-distribué à deux paramètres de calibration, développé sous une base journalière et prenant comme intrants les précipitations totales et les températures moyennes. Le modèle est dit semi-distribué puisque son utilisation se fait à différents niveaux d'élévation sur le bassin versant. Il faut diviser le bassin en cinq bandes d'élévation pour lesquelles CemaNeige est appliqué séparément. La modification des précipitations et des températures à l'élévation moyenne de la bande en question se fait à l'aide de gradients de température et de précipitation. La paire de paramètres doit donc être calibrée distinctement pour chaque bande d'élévation. L'utilisation de cette version du modèle apporte toutefois son lot de désavantages, surtout lorsque la routine de neige s'intègre à un modèle hydrologique conceptuel.

Dans la présente étude, une version conceptualisée de CemaNeige est utilisée. Cette version conceptuelle est originale adaptée au Québec, elle ne considère donc aucune bande d'élévation pour l'application de la séparation des précipitations. La séparation est faite, peu importe l'élévation avec les températures seuils prédéterminées à 0°C pour les précipitations

solides et mixtes ainsi que 3°C pour les précipitations mixtes et liquides. Ces seuils sont ceux ressortant de Harpold et al. (2017).

La routine comprend au total cinq étapes. La première étape calcule la séparation des précipitations solides et liquides avec les seuils de température mentionnés ci-haut. La portion liquide s'ajoute à la lame d'eau de précipitation qui sera fournie au modèle hydrologique alors que la portion solide s'ajoute au manteau neigeux. Suite à l'accumulation de neige, le déficit énergétique pour la fonte (glace et neige) est mis à jour à l'aide d'une fonction basée sur la température. Il y a fonte de neige seulement lorsque la température de la neige est de 0°C. Cette fonction comprend le premier paramètre qui est un facteur du déficit énergétique. Le niveau de ce déficit précède le calcul de la fonte de neige potentielle qui se fait à l'aide d'un facteur de fonte, le deuxième paramètre du modèle. La valeur réelle de fonte qui est ajoutée à la lame d'eau de précipitation liquide est évaluée empiriquement à 90% de la fonte potentielle. La figure 3.5 montre le schéma de production de la lame de précipitation liquide finale. Cette lame d'eau devient l'intrant de précipitation fourni à GR4J et MOHYSE.

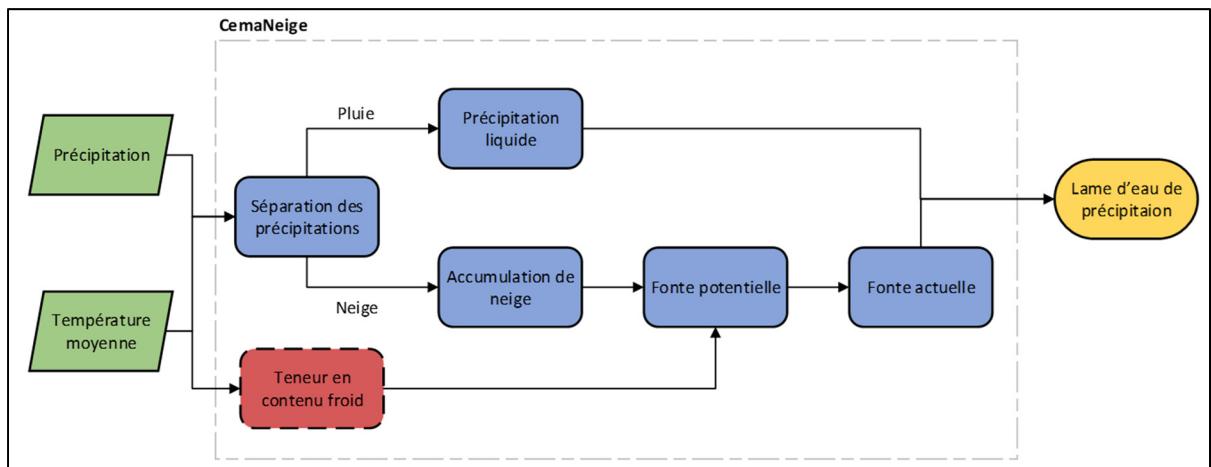


Figure 3.5 Schéma de production de CemaNeige

3.3.4 Calibration des modèles hydrologique

Le processus de calibration des modèles hydrologiques permet l'ajustement du modèle au bassin versant sur lequel le modèle est utilisé. Généralement, la calibration se fait par processus itératif automatique. Cette méthode offre la meilleure robustesse et la plus grande dissociation des biais pouvant être introduits par l'utilisateur en comparaison avec une calibration manuelle par exemple. Peu importe l'algorithme de calibration utilisé, un nombre d'itérations sera réalisé. Cette limite est généralement fixée par l'utilisateur. Une itération de calage comprend les étapes suivantes, un visuel du processus est montré à la figure 3.6 :

1. Simuler les débits à l'aide d'un premier jeu de paramètres donné en intrant;
2. Vérifier la performance du modèle par l'adéquation de la simulation par rapport à des données connues en utilisant une fonction-objectif quelconque;
3. Modifier le jeu de paramètres et répéter si la performance obtenue à l'étape précédente n'est pas adéquate.

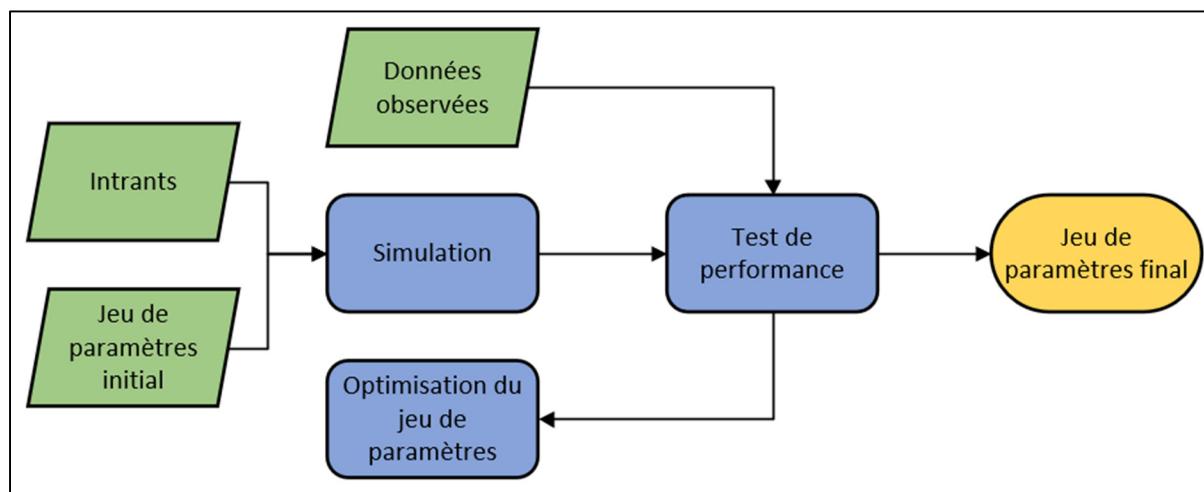


Figure 3.6 Schéma du processus de calibration d'un modèle hydrologique

Les modèles hydrologiques ont dû être calibrés pour chaque combinaison d'intrants différente. Pour cette étude, une variation dans les intrants est notable pour chaque bassin versant (1169 bassins) sur lesquels différents jeux de données de précipitation et température sont utilisés (3 jeux : observations, ERA5, ERA5-Land) en combinaison avec différentes méthodes pour

estimer l'ETP (12 formules et les estimations d'ERA5 et d'ERA5-Land). L'ensemble des combinaisons (bassins, modèles, données, ETP) s'élève donc ici à un total de 91 182. L'utilisation de modèles à plusieurs paramètres augmente toutefois le risque d'obtenir un jeu de paramètres se trouvant dans une zone de mauvaise qualité de l'espace de paramètres. Pour minimalement contrer ce phénomène, le processus de calibration est répété une seconde fois, totalisant 182 364 calibrations. La prochaine section décrit plus en profondeur la méthode de calage, l'algorithme d'optimisation utilisé pour la calibration ainsi que la fonction-objectif.

3.3.4.1 Méthode de calibration

Habituellement, un processus de calibration régulier se fait en deux parties : (1) la calibration et (2) la validation. La méthode la plus utilisée est de scinder les données en deux parties égales (ici en nombre d'années) dans le but d'associer une partie des données à la calibration alors que la seconde partie sert à vérifier la performance de la calibration. Cette méthode permettrait la meilleure évaluation de la performance de la calibration puisque le modèle calibré est utilisé avec des données jamais vues auparavant. Les processus hydrologiques sont toutefois très complexes et l'approche par séparation a son lot d'inconvénients. Entre autres, un événement hydrologique spécial peut se retrouver dans la période de validation et représente donc un manque d'information en période de calibration. Arsenault, Brissette, et Martel (2018) proposent une technique de calibration qui utilise la totalité des données. Leur étude a démontré que par cette méthode, la calibration produit un jeu de paramètres beaucoup plus robuste et un modèle plus performant.

Pour cette étude, l'utilisation de différentes données et formules demande davantage de robustesse dans la représentation des processus sous-jacents, alors que l'exactitude des résultats simulés est moins importante que la robustesse de ceux-ci. La méthode proposée par Arsenault et al. (2018) est utilisée et la totalité des données est fournie au processus de calibration.

3.3.4.2 Algorithme d'optimisation

La méthode de calibration choisie pour l'étude est par calage itératif automatique. L'algorithme choisi est le « Dynamically Dimensioned Search » (DDS) (Tolson & Shoemaker, 2007) qui est un algorithme efficient lorsque le nombre de paramètres à calibrer est grand. Cet algorithme est aussi avantageux puisqu'il y a une méthode interne afin d'éviter les optimums locaux ainsi qu'en raison de sa capacité à rapidement réduire l'espace de recherche des jeux de paramètres possibles. (Arsenault et al., 2014)

Malgré que certains optimiseurs performent mieux que DDS lorsqu'ils sont utilisés avec des modèles avec un bas nombre de paramètres (10 et moins), la force de DDS survient dans sa rapidité d'exécution et sa plus faible demande en temps de calculs (Arsenault et al., 2014; Tolson & Shoemaker, 2007). Étant donné le nombre de calibrations à faire, le temps de calcul total est le facteur le plus important et c'est pour cela que l'optimiseur choisi est DDS.

Le temps de calcul total est fonction de la complexité du modèle hydrologique, de la grandeur de l'espace de paramètres, mais aussi du critère d'arrêt de la calibration. Ce critère se définit comme un nombre maximal d'itérations pour DDS. En raison de sa rapide convergence, DDS est habituellement en mesure de trouver un bon jeu de paramètres avec aussi peu que 250 itérations (Huot et al., 2019). Par contre, conservant l'idée qu'après la convergence initiale, DDS peut être dans ou près d'un optimum local qui peut être amélioré, le critère d'arrêt est fixé à un nombre supérieur d'itérations, c'est-à-dire 2000 itérations pour cette étude. Étant environ huit fois plus grand que le nombre moyen d'itérations nécessaires pour converger, il y a une plus grande chance que DDS offre le jeu de paramètres le plus optimal possible.

Pour ce qui est de l'espace de paramètres, il est limité par l'utilisation de bornes minimales et maximales. Pour chaque modèle, les bornes utilisées sont les mêmes que celles dans (Dallaire, 2019), qui relève des bornes conseillées par Turcotte & Fortin (2007) pour MOHYSE et par Perrin et al. (2003) pour GR4J. De ces bornes, le jeu de paramètres initial fourni à DDS

représente les valeurs moyennes des bornes pour chaque paramètre respectivement. Les bornes sont présentées à l'annexe I.

3.3.4.3 Fonction-objectif de performance

Le jeu de paramètre défini en processus de calibration est celui qui permet de fournir les meilleurs débits simulés, c'est-à-dire ceux qui se rapprochent le plus de la réalité. Pour juger cette performance, une fonction d'efficience est utilisée, la fonction-objectif de performance. Pour cette étude, le critère de performance est le KGE modifié (KGEm) (Kling et al., 2012). Cette version du KGE original utilise le coefficient de variation au lieu de l'écart type dans le calcul de la variabilité. Cela évite la corrélation croisée possible entre les éléments de variabilité et de biais lorsqu'un biais est présent dans les données d'entrées (ex. : précipitation) (Gupta et al., 2009). L'équation 3.9 présente le calcul du KGEm alors que les équations 3.10 et 3.11 décrivent respectivement le calcul des composantes de biais (β) et de variabilité (γ).

$$KGEm = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (3.9)$$

$$\beta = \frac{\mu_s}{\mu_o} \quad (3.10)$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \quad (3.11)$$

Dans les équations ci-dessus, r représente de coefficient de corrélation, β le ratio de biais, γ le ratio de variabilité, μ la moyenne, CV le coefficient de corrélation et σ l'écart-type. Les indices s et o signifient *simulé* et *observé* respectivement. Le KGEm peut prendre une valeur allant de $-\infty$ à 1. Un résultat qui tend vers $-\infty$ est considéré mauvais alors qu'un résultat tendant vers 1 est considéré bon. Dans l'univers de la modélisation hydrologique, une valeur de KGEm se retrouvant entre 0.6 et 0.9 est généralement considérée bonne. Il est à noter que des valeurs au-delà de 0.9 sont excellentes, mais souvent rares et qu'une valeur exacte de 1 est impossible à atteindre.

3.4 Analyse des résultats

Afin de tirer des conclusions spécifiques aux caractéristiques des jeux de données ainsi que sur leur utilisation en modélisation hydrologique, les résultats sont scindés en trois grandes sections. Les trois jeux de données sont d'abord comparés les uns aux autres, suivis d'une comparaison entre les valeurs d'ETP que ces derniers génèrent par l'utilisation de différentes formules présentées au tableau 3.2. La présentation des résultats se termine par une comparaison de leur performance en modélisation hydrologique, c'est-à-dire de leur pouvoir à simuler des débits à l'exutoire d'un bassin versant. Cette dernière comparaison inclut l'utilisation directe des données d'ETP et d'ETR provenant des réanalyses.

Pour l'ensemble des résultats présentés, lorsqu'une comparaison entre des jeux de données est réalisée, le jeu de référence peut varier. En effet, si une analyse est faite contre les observations (OBS vs ERA5 et OBS vs ERA5-Land), ce sont ces dernières qui sont considérées comme la référence. Dans le cas des comparaisons entre les réanalyses, ce sont les données de ERA5 qui sont la référence (ERA5 vs ERA5-Land).

3.4.1 Comparaison des jeux de données

Pour la comparaison entre jeux de données, les comparaisons sont effectuées via l'analyse des coefficients de corrélation de Pearson (r) (équation 3.12) ainsi que de la différence moyenne entre les valeurs (équation 3.13). Le but est de voir dans quelle mesure les données sont semblables ou différentes entre elles.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}} \quad (3.12)$$

où r est le coefficient de corrélation, x représente les observations ou ERA5 et y les données de comparaison (ERA5 ou ERA5-Land).

$$\overline{diff} = \frac{\sum(x - y)}{n} \quad (3.13)$$

où \overline{diff} est la différence moyenne, x représente les observations ou ERA5, y les données de comparaison (ERA5 ou ERA5-Land) et n la taille de l'échantillon.

Dans le cas des valeurs générées pour les variables liées au soleil (radiation solaire extra-terrestre par exemple), la référence est considérée comme étant les valeurs calculées par la formulation empirique de base (équation 3.7).

3.4.2 Valeurs d'ETP

Dans la structure de HOOPLA, les données d'ETP sont fournies en entrée aux modèles hydrologiques et devaient donc être calculées préalablement. Les valeurs de ces résultats sont présentées et comparées via l'utilisation de graphiques sous la forme de boîtes à moustaches. L'utilisation de ces graphiques permet d'établir la distribution des données facilement et offre une comparaison visuelle et rapide entre les distributions. Afin d'en tirer le plus d'information possible, les résultats sont présentés en considérant différents regroupements. Par exemple, les jeux de données sont comparés entre eux en fonction des formulations d'ETP utilisées.

3.4.3 Performance en modélisation hydrologique

Les résultats de performance en modélisation hydrologique sont présentés en fonction des valeurs de KGE obtenues en calibration. Étant donné que la calibration est réalisée sur la totalité de la période de disponibilité de données dans chacun des bassins versants, les débits simulés obtenus en calibration sont conservés, ce qui évite de devoir refaire une simulation par bassin avec les paramètres issus de la calibration. Les jeux de paramètres et débits issus correspondant au meilleur KGE obtenus en calibration sont conservés. Pour s'aligner avec les objectifs du projet, les résultats en modélisation hydrologique sont principalement présentés suivant 2 axes, soit (1) en comparant les résultats issus des observations contre ceux issus des réanalyses; et (2) en comparant les formulations d'ETP ainsi que leur type. Comme pour les résultats des valeurs d'ETP, les résultats sont présentés sous forme de boîtes à moustaches afin de permettre d'établir les distributions des résultats.

CHAPITRE 4

RÉSULTATS

Le présent chapitre présente l'ensemble des résultats obtenus dans le cadre de l'étude. Une comparaison des données est premièrement faite, suivie d'une comparaison des valeurs d'ETP. Ces valeurs sont celles directement calculées à partir des données et servent à faire le parallèle entre les données utilisées pour les calculs et les valeurs d'ETP. Le chapitre se termine par la présentation des résultats de modélisation, soit les débits simulés.

Au fil du chapitre, les résultats sont souvent liés à des formules d'ETP ou à des zones climatiques. Le tableau 4.1 ci-dessous offre un rappel des acronymes pour les formules ETP ainsi que ceux utilisés pour les variables de réanalyses. Le tableau 4.2 quant à lui, définit la numérotation des zones climatiques.

Tableau 4.1 Acronymes associés aux formulations d'ET

Acronyme	Formulation	Acronyme	Formulation
BC	Blaney-Criddle	HS	Hargreaves-Samani
BR	Baier-Robertson	HQ	Hydro-Québec
K	Kharrufa	JH	Jensen-Haise
FAO-PM	FAO Penman-Monteith	MB	McGuinness-Bordne
H	Hamon	M	MOHYSE
E5R	Groupe d'ERA5 et ERA5-Land ETR	O	Oudin
E5E	Groupe d'ERA5 et ERA5-Land ETP	T	Thornthwaite

Tableau 4.2 Numérotation des zones climatiques

Numéro	Région climatique	Numéro	Région climatique
1	ColdNEPacific	17	Southeast
2	WarmNEPacific	18	WarmNWAtlantic
3	PacificNW	19	MidAtlantic
4	PacificSW	20	NorthAtlantic
5	Southwest	21	ColdNWAtlantic
6	Mezquital	22	EastBoreal
7	NRockies	23	WestBoreal
8	SRockies	24	EastTaiga
9	GreatBasin	25	WestTaiga
10	NPlains	26	CentralTundra
11	CPlains	27	WestTundra
12	SPlains	28	EastTundra
13	Prairie	29	Hudson
14	GreatLakes	30	LabradorSea
15	Appalachia	31	Greenland
16	DeepSouth	-1	Mexique

4.1 Analyse et comparaison des jeux des données

L'analyse a été réalisée en regroupant les données de différentes façons. Principalement, les données sont comparées sur l'ensemble de leur disponibilité, en regardant les moyennes saisonnières ou encore en regroupant les données par régions climatiques. Les résultats des groupements par région climatique ne permettaient pas de déceler des tendances différentes ou de montrer des différences par rapport aux groupements présentés ci-après, ils sont donc omis. Des liens avec la géolocalisation, et donc par défaut les zones climatiques, sont toutefois présentés.

4.1.1 Variables disponibles dans les trois jeux de données

Les disponibilités des variables d'observations sont très limitées. L'intersection des variables disponibles ne comprend que les températures et les précipitations. De plus, la radiation solaire extra-terrestre n'est disponible que pour ERA5. Pour les observations (OBS) et ERA5-Land, leurs valeurs proviennent d'un calcul empirique (équation 3.7). Le tableau 4.3 ci-dessous présente les valeurs moyennes des r et des différences moyennes entre les réanalyses et les observations ainsi qu'entre les réanalyses.

Au niveau des températures, les résultats sont très probants, aucun coefficient de corrélation moyen ne se trouvant sous 0,96 et les différences moyennes sont au maximum de 2°C. Ces moyennes proviennent des valeurs obtenues par bassin au pas de temps journalier. Les distributions des coefficients de corrélation et des différences moyennes sont globalement petites. La moins bonne corrélation parmi tous les bassins est de 0,71 et l'erreur maximale est de 9,16°C, et ce pour la comparaison entre les observations et ERA5-Land. Similairement, les résultats pour la radiation solaire extra-terrestre sont bons. La relation entre OBS et ERA5-Land est attendue étant donné que pour ces deux jeux de données, cette variable n'était pas disponible et a donc été calculée de la même façon, c'est d'ailleurs pour cela aussi que les valeurs entre OBS et ERA5 sont les mêmes. Il est donc plus important de regarder la force d'ERA5 seulement pour cette variable. Au niveau local, le plus bas coefficient de corrélation est de 0,99.

Tableau 4.3 Valeurs moyennes, pour les 1169 bassins, des coefficients de corrélation et des différences moyennes

Variable	OBS vs ERA5		OBS vs ERA5-Land		ERA5 vs ERA5-Land	
	r	Déférence moyenne	r	Déférence moyenne	r	Déférence moyenne
Tmax (°C)	0,96	0,6	0,96	1,2	1	0,62
Tmoy (°C)	0,98	-0,51	0,98	-0,3	1	0,23
Tmin (°C)	0,97	-1,9	0,97	-2,1	0,99	-0,18

P (mm/j)	0,73	-0,02	0,74	0,09	0,98	0,1
Exrad (kJ/j)	1	0,57	1	0,03	1	-0,54

Dans le tableau 4.3 ci-dessus, Tmax représente la température maximale, Tmoy la température moyenne, Tmin la température minimale, P les précipitations et Exrad la radiation solaire extra-terrestre.

C'est surtout au niveau des précipitations que ERA5 et ERA5-Land diffèrent des observations. La variabilité dans les précipitations est difficilement captable, un problème connu des précipitations de ERA5 (M. Tarek et al., 2020). Cette faiblesse est toutefois localisée. Les valeurs du tableau 4.3 seraient donc différentes si les valeurs par bassins étaient en sous-groupes régionaux. Comme le montre la figure 4.1, les coefficients de corrélation sur l'ensemble de l'Amérique du Nord sont plus faibles pour presque l'ensemble de la zone étudiée par rapport à l'ouest des États-Unis. Il est à noter que ces résultats contredisent un peu ceux retrouvés généralement dans la littérature. La plus grosse différence ici est que les coefficients de corrélation moyens montrés à la figure 4.1 sont issus de la moyenne entre la corrélation des données correspondantes en date. Si les corrélations entre les moyennes annuelles ou saisonnières étaient plutôt utilisées, les corrélations seraient plus fortes puisque l'utilisation de ces moyennes réduirait l'impact des valeurs journalières. Par exemple, les coefficients de corrélation en utilisant les moyennes annuelles de précipitation pour OBS versus ERA5 et OBS versus ERA5-Land sont respectivement 0,87 et 0,88. Une augmentation de plus de 0,1 dans les deux cas. Pour la précipitation, le pire coefficient de corrélation s'obtient entre les observations et ERA5 avec une valeur de 0,31.

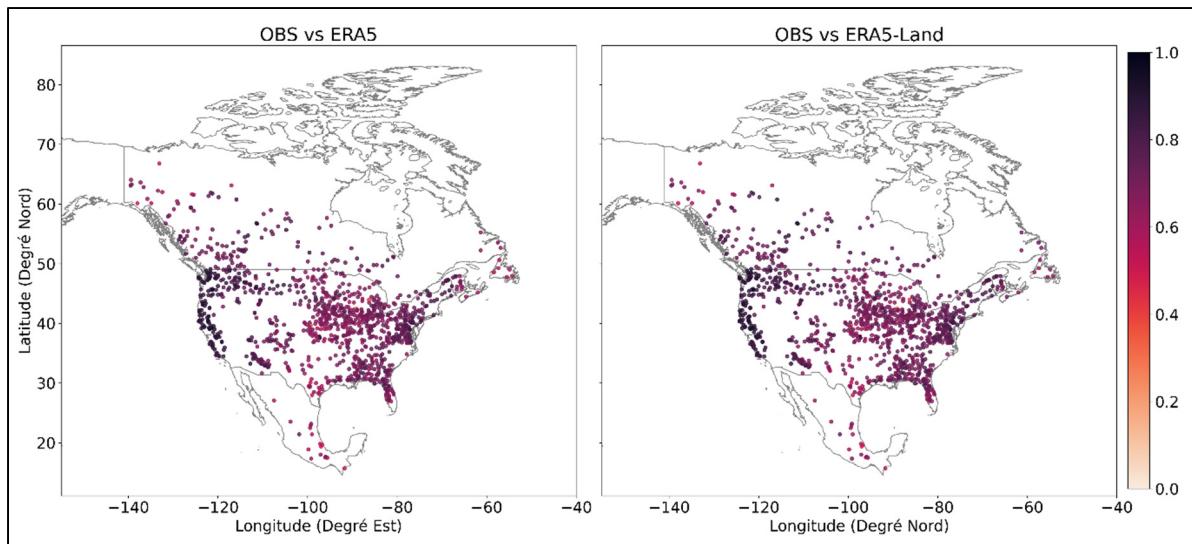


Figure 4.1 Distribution spatiale des coefficients de corrélation pour la précipitation

4.1.2 Variables disponibles que pour les réanalyses

Un grand avantage des réanalyses est l'exhaustivité de la liste des variables disponibles. Pour ERA5 et ERA5-Land et dans le cadre de l'étude, les variables suivantes s'ajoutent aux variables présentées à la section précédente : la température du point de rosée, la vitesse du vent, la pression atmosphérique, la radiation solaire nette, l'ETP et l'ETR.

À l'exception de la pression atmosphérique (analysée plus en détail ci-après), les corrélations sont fortes entre les deux réanalyses comme le montre le tableau 4.4. Ce sont des résultats attendus puisque ERA5-Land a été créée dans le but d'être représentatif de ERA5, mais à une résolution spatiale plus fine. L'ETP voit son coefficient de corrélation un peu plus bas que celui des autres variables (excluant la pression atmosphérique). La différence moyenne témoigne bien de la disparité entre les résultats. En effet, tel que le montre la figure 4.2, en général, les valeurs d'ETP sont bien corrélées. La diminution du coefficient moyen est causée par les bassins environnant les Grands Lacs et dans les régions des maritimes pour lesquels des coefficients de corrélation négatifs atteignent des valeurs jusqu'à -0,48. De plus, malgré que le coefficient de corrélation moyen soit considérablement bon, il est montré par la différence moyenne qu'une différence non négligeable est présente dans les données. Au maximum la

différence atteint -22,03 mm. Le signe négatif de la différence maximale et de la moyenne du Tableau 4.4 montre que ERA5-Land sous-estime les valeurs comparativement à ERA5. Ceci est développé davantage dans la discussion du Chapitre 5.

Tableau 4.4 Coefficients de corrélation moyens et différences moyennes

Variable	r	Déférence moyenne
Tdew (°C)	1	0,15
w (m/s)	0,98	0,33
pr (kPa)	0,32	-0,09
netrad (kJ/j)	1	0,33
ETP (mm/j)	0,88	-7,64
ETR (mm/j)	0,96	0,01

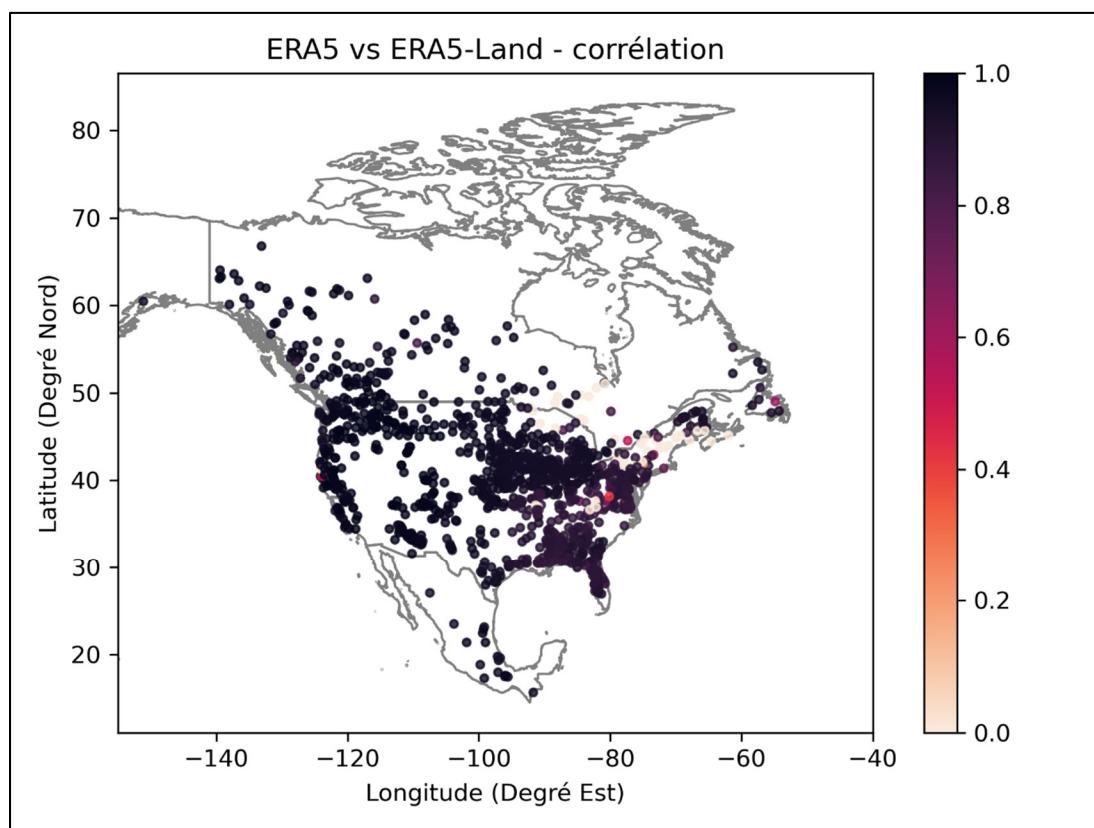


Figure 4.2 Distribution spatiale des coefficients de corrélation pour l'ETP journalière entre les réanalyses ERA5 et ERA5-Land.

La pression atmosphérique montre une faible corrélation alors que la différence r moyenne est basse. Cela est expliqué par une corruption dans les données de pression d'ERA5-Land qui s'est introduite lors du téléchargement de celles-ci. Il s'agit en fait d'une valeur constante de 76,89 kPa journalièrement pour la 2^e moitié de l'année 2006. La corruption est très isolée et c'est pourquoi l'erreur moyenne est si faible, il n'en demeure pas moins que l'impact est plus significatif sur le coefficient de corrélation qui lui regarde les valeurs journalières. En considérant les pressions annuelles moyennes dans le but d'atténuer l'impact de la corruption, le coefficient de corrélation monte à 0,97. La figure 4.3 ci-dessous montre l'étendue spatiale et l'impact de la corruption. Les bassins versants situés en altitude (par exemple dans les rocheuses) voient leurs coefficients de corrélation à atteindre de meilleures valeurs puisque dans ces cas la pression est généralement plus basse en raison de l'altitude et donc la corruption n'affecte pas autant les séries temporelles de données.

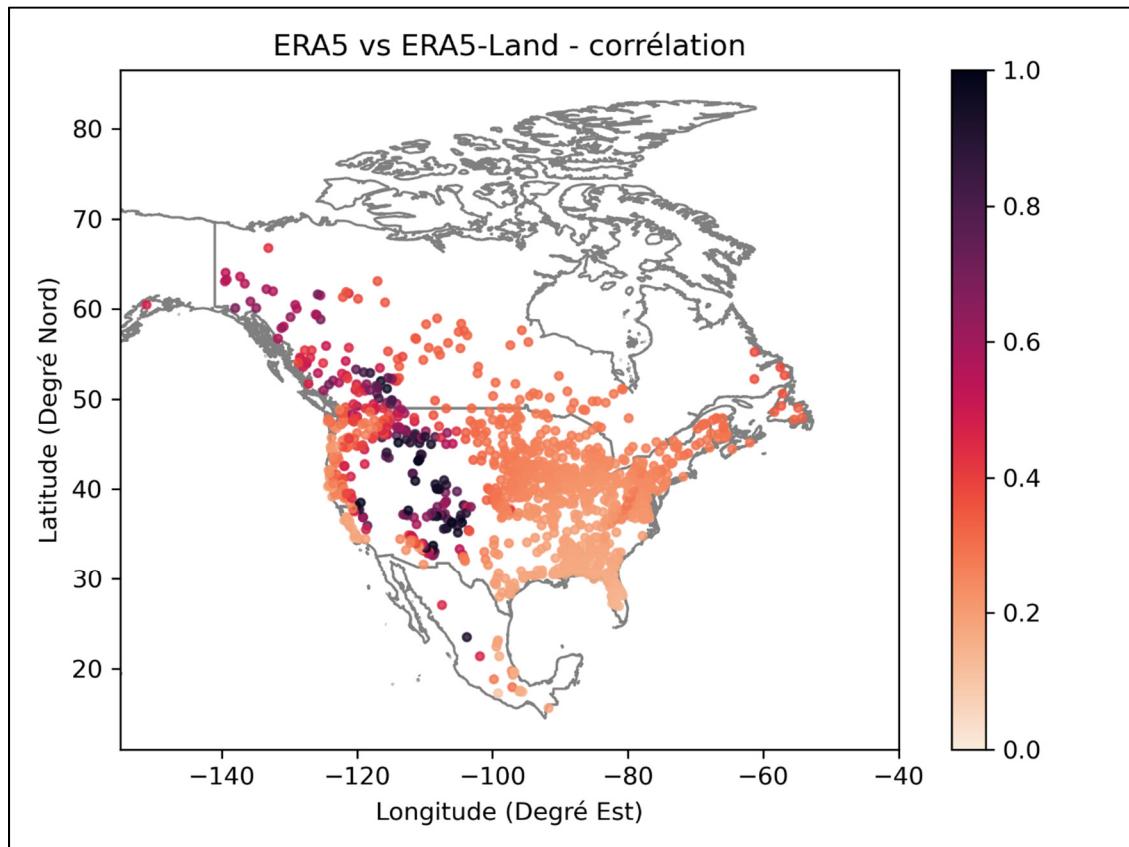


Figure 4.3 Distribution spatiale des coefficients de corrélation pour la pression atmosphérique journalière.

4.2 Estimation de l'ETP

Cette sous-section analyse les valeurs des estimations d'ETP. La figure 4.4 ci-dessous montre les valeurs annuelles d'ETP regroupées par formulation (les acronymes peuvent être consultés au tableau 4.1). Deux autres regroupements ont été réalisés, l'un en fonction des saisons et l'autre par région climatique. La différence n'étant pas significative dans les résultats, ces groupements sont omis au profit de seulement la figure 4.4.

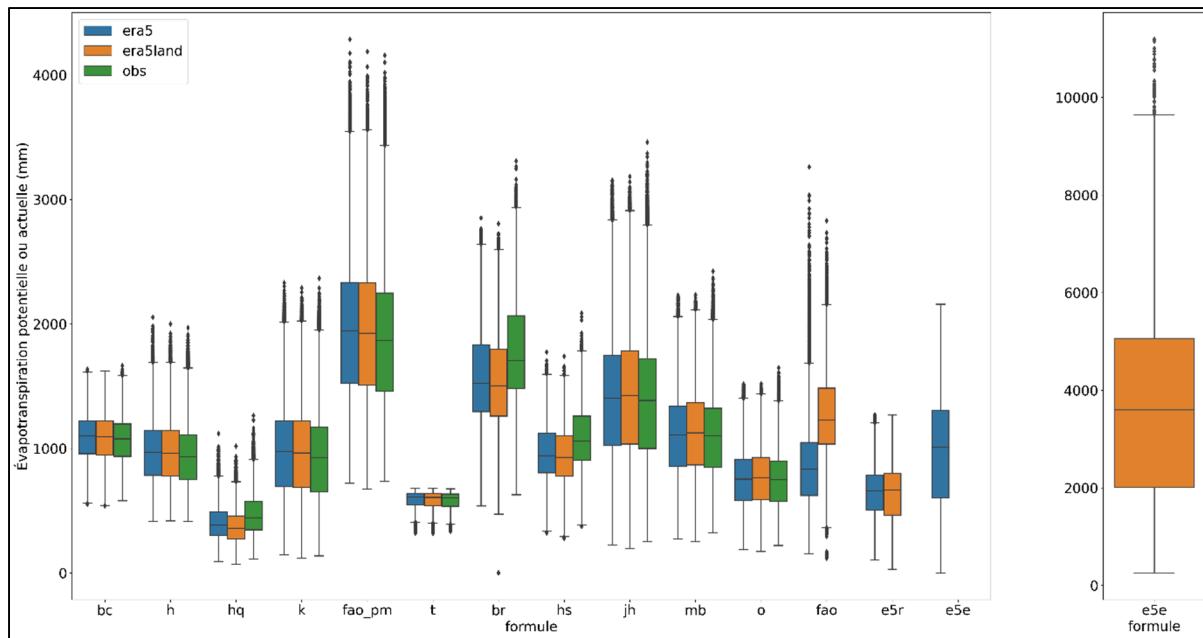


Figure 4.4 Boîtes à moustaches des valeurs d'ETP annuelle par formulation d'ETP pour tous les bassins. Ici, la boîte séparée à droite est isolé pour éviter d'affecter l'échelle des autres boîtes.

4.2.1 Observations et réanalyses

Comme le présente la figure 4.4, les résultats d'estimation d'ETP sont assez constants entre les différents jeux de données en considérant les résultats avec OBS contre ceux de réanalyses pour une même formule. Ce sont des résultats auxquels il est possible de s'attendre en considérant la ressemblance entre les données montrées à la section 4.1.1. Les résultats en sont de même, peu importe le groupement, c'est-à-dire par saison ou par région climatique. Les

seules exceptions surviennent pour la formule d'estimation de FAO-PM, pour la variable E5L-ETP ainsi que pour les OBS dans la région climatique -1. L'analyse concernant E5L-ETP, ce sera présenté à la sous-section suivante. La sensibilité de FAO-PM aux données d'entrée est bien montrée par les résultats d'ETP. Malgré que les données d'entrées se ressemblent beaucoup, une différence non significative peut mener à une grande différence dans les estimations d'ETP avec FAO-PM. Le groupement par région climatique souligne bien cet aspect. En effet, pour la totalité des régions, les valeurs obtenues avec ERA5-Land sont supérieures à celles obtenues avec ERA5 (annexe II).

Quant à la région climatique -1, c'est elle qui englobe tous les bassins n'étant pas couverts par les séparations des régions climatiques de Bukovsky. Ceux-ci sont localisés au centre et au sud du Mexique. Comme le montre la figure 4.5, peu importe la formule utilisée, les résultats avec les OBS contre ceux avec les données de réanalyses sont montrent des différences importantes. La distribution est beaucoup plus large et le grand écart interquartile suggère une grande dispersion des résultats. La qualité des données au Mexique est plus incertaine et le potentiel d'opter pour des réanalyses dans ces cas est bien soutenu.

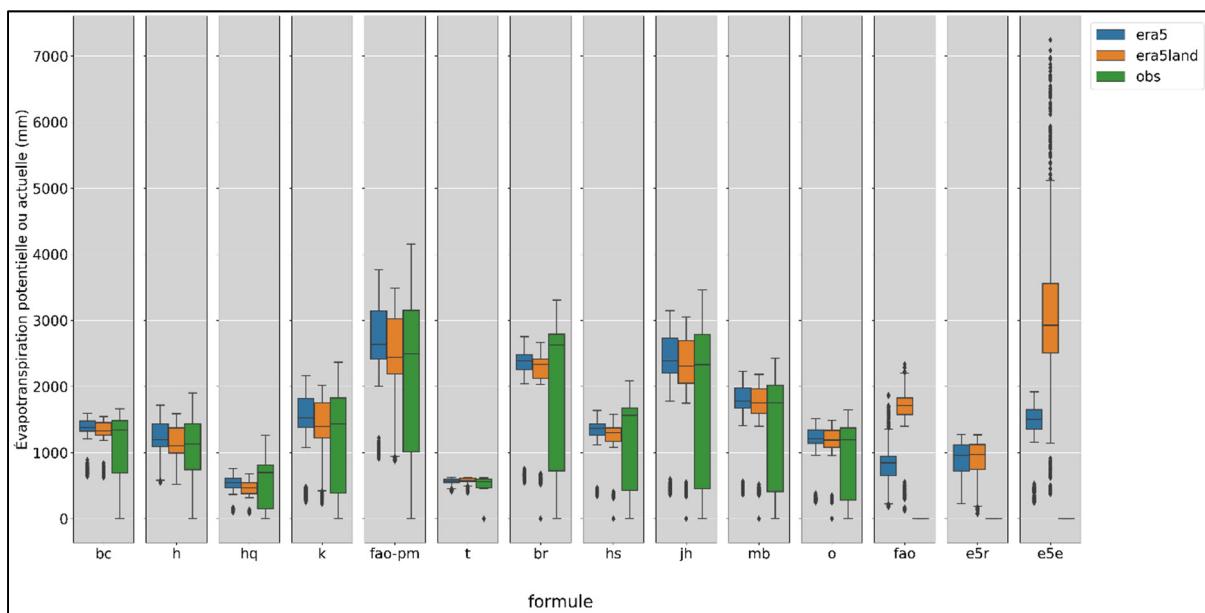


Figure 4.5 Boîtes à moustaches des valeurs d'ET pour la région climatique -1 (centre et sud du Mexique) toute formules d'ET confondues

4.2.2 Formules d'ETP

La qualité et la pertinence des formules en tant que telle sont difficilement jugeables. Les résultats montrent qu'autant des formules à base de température (HQ, M et T) que celles à base de radiation (BR et JH) peuvent mener à des valeurs plus extrêmes et à beaucoup de dispersion dans les résultats produits vis-à-vis des autres formulations dans leur même famille (figure 4.4). Comme mentionné plus haut, c'est similairement le cas en regroupant les valeurs par saison ou régions climatiques.

En suivant la littérature et en prenant la formulation de FAO-PM comme référence, les formulations à base de température sont celles qui s'éloignent le plus des valeurs de FAO-PM. Certaines formules, comme BC, H ou K, donnent des résultats satisfaisants et similaires à FAO-PM et à plusieurs formules à base de radiation (HS, MB et O). En soi, les formules à base de température peuvent offrir une bonne performance, elles ne considèrent toutefois qu'une seule variable parmi les quatre principales dirigeant la demande évaporative atmosphérique, et donc par défaut offrent une moins bonne représentativité physique.

4.2.3 Variables d'ETP et d'ETR des réanalyses

Il y a une grande différence entre les différentes variables d'ETP et d'ETR des réanalyses, tel que montré à la figure 4.5. Évidemment, il est normal de voir l'ETR donner des valeurs plus basses, surtout au niveau de la différence entre ERA5 et ERA5-Land. Pour l'ETR, les valeurs sont beaucoup plus similaires que pour l'ETP, toutefois ERA5-Land montre une distribution plus large que ERA5 et donc par le fait même une plus grande dispersion dans ses résultats. Leur différence ne semble toutefois pas significative et probablement qu'en modélisation hydrologique, la différence entre les deux réanalyses pour cette même variable ne sera pas non plus considérable. En comparaison avec les autres formulations d'ETP, l'ETR donne des résultats qui semblent réalistes. En effet, à l'exception d'HQ et de T, les valeurs d'ETR pour chacune des réanalyses sont plus basses que le reste des valeurs d'ETP estimées avec les autres formulations. Ces constats concordent aussi en regroupant les estimations d'ETP par saison alors que la différence entre les deux variables se fait sentir un peu plus par région climatique.

En effet, ERA5-Land retourne des valeurs considérablement plus élevées que celles d'ERA5 dans la région des Grands Lacs et à l'est des États-Unis alors que c'est l'opposé pour les rocheuses et l'ensemble de la côte ouest-canadienne et américaine. Ce phénomène concorde avec la méthode de calcul de l'ETP d'ERA5-Land qui est un calcul d'évaporation de bac comparativement à la méthode d'ERA5 qui emploie plutôt un calcul d'évaporation pour agriculture avec un sol saturé en eau.

La variable d'E5-ETP montre du potentiel. Les valeurs sont généralement comparables aux autres formulations utilisées. Malgré une distribution plus étendue, les valeurs de l'E5-ETP sont très similaires à celles de FAO-PM en considérant cette formule comme étant la référence telle que le prescrit la littérature. En regroupant les valeurs par saison, la dispersion des valeurs d'E5-ETP peut d'ailleurs être associée aux valeurs automnales et estivales qui sont sensiblement plus élevées comparativement à celles de FAO-PM (figure 4.6). Pour ce qui est de l'E5L-ETP, les valeurs montrent une imprécision et des valeurs impossibles. En effet, les valeurs médianes environnent les 4000 mm d'ETP annuellement et des valeurs extrêmes allant jusqu'à 10 000 mm sont décelées. Ces valeurs équivalent à plus que le double des valeurs de n'importe quelle autre formulation utilisée pour l'étude. Ces résultats sont d'ailleurs bien démontrés à la figure 4.4 plus haut.

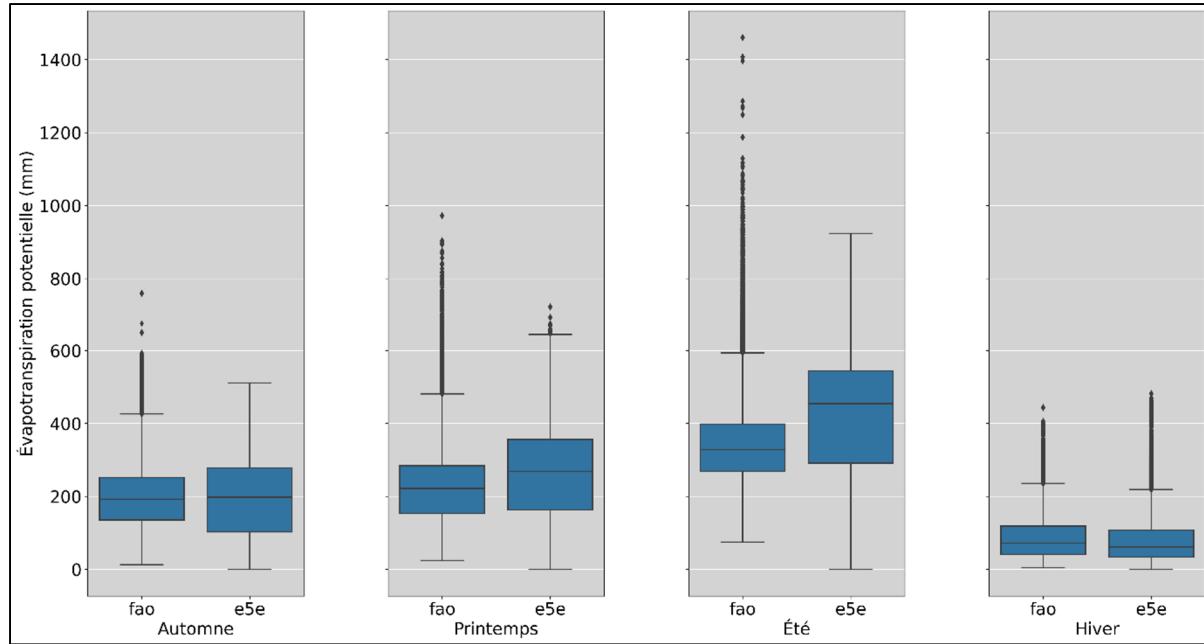


Figure 4.6 Comparaison des valeurs d'ETP saisonnières totales entre FAO-PM et E5-ETP provenant d'ERA5

4.3 Résultats de modélisation hydrologique

Cette sous-section présente les résultats en modélisation hydrologique. Tel que montre la figure 4.7 ci-dessous, les KGE ont globalement de bonnes valeurs en modélisation hydrologique et ce peu importe le type de données ou la formule d'estimation de l'ETP. Il est important de souligner que comme la littérature le soutient, les OBS restent en tête en termes de performance générale en calibration des modèles hydrologiques. Cela n'enlève toutefois rien aux réanalyses qui donnent de très bonnes performances aussi. Ces résultats sont considérés généralisés puisque plusieurs regroupements montrent les mêmes tendances. Qu'ils soient séparés seulement par le jeu de données ou encore par pays et jeux de données, les OBS tendent à offrir de meilleurs résultats que les réanalyses sans que les réanalyses ne soient nécessairement loin derrière. Les réanalyses se distinguent du lot surtout en considérant les caractéristiques physiographiques des bassins. La section 4.3.1 aborde ce point plus en détail.

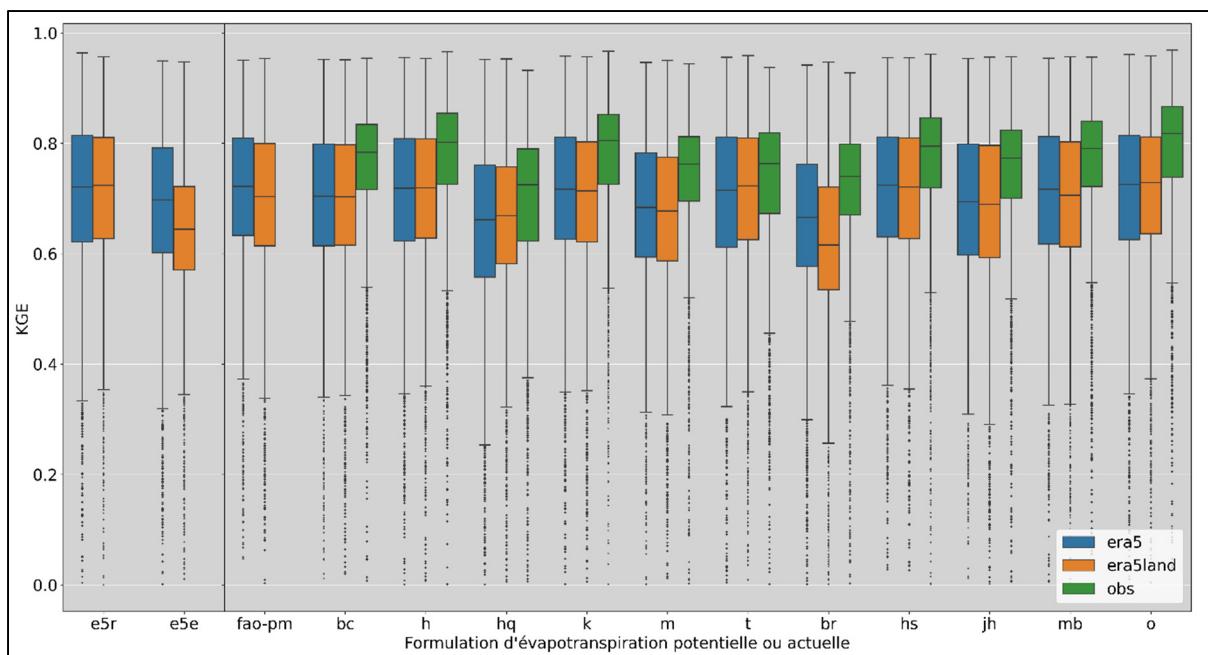


Figure 4.7 Distribution des KGE en calibration par jeux de données et par formulation d’ET

4.3.1 Observations et réanalyses

Comme le montre la figure 4.7, peu importe le jeu de données utilisé, il est possible d’atteindre des valeurs satisfaisantes en modélisation hydrologique. La figure 4.8 quant à elle permet de renforcer les résultats montrés précédemment. Elle permet de montrer les résultats sur le plan spatial pour l’ensemble des résultats par bassin, toutes formules confondues. Cela signifie dans les deux cas qu’une valeur négative sous-entend une meilleure performance pour le jeu de données de référence (OBS) alors qu’une valeur positive équivaut à une performance supérieure en modélisation pour le jeu de donnée en comparaison (ERA5). Il est aussi important de souligner que la comparaison directe entre les OBS et ERA5-Land n’est pas faite. Il en est ainsi puisque la performance de ERA5-Land par rapport aux OBS peut en partie être déduite de la comparaison avec ERA5 puisqu’ERA5 est comparée aux OBS. Il est donc considéré, par ce que montre la figure 4.7, qu’en général ERA5 est plus performant que ERA5-Land.

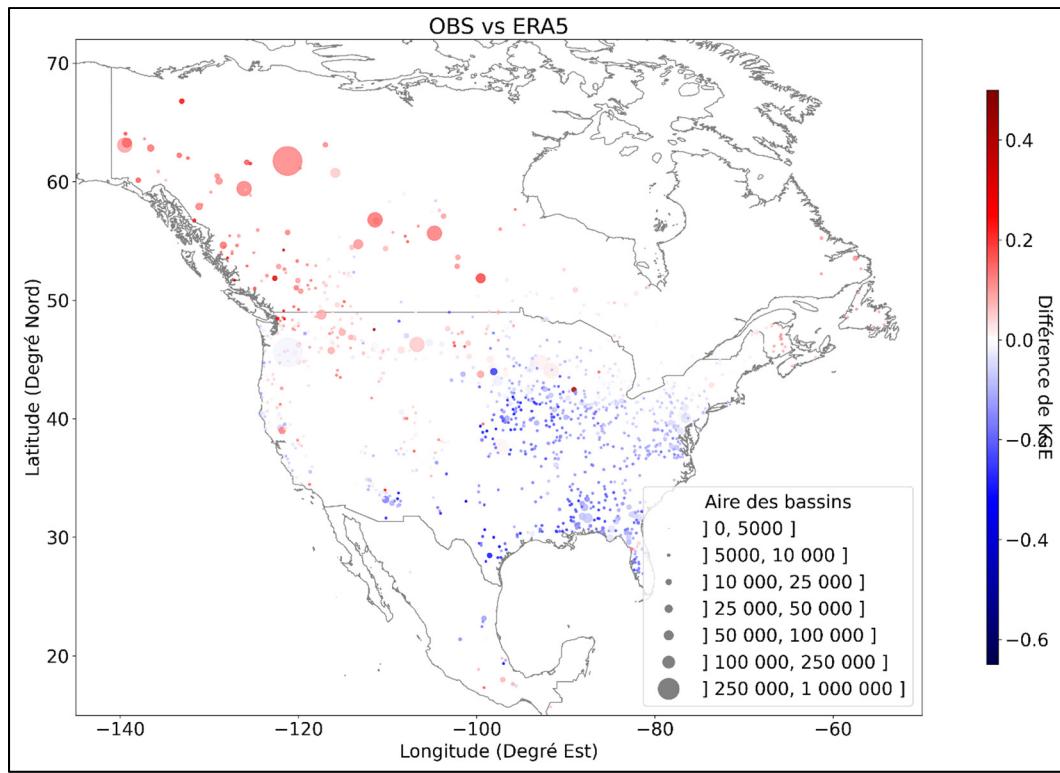


Figure 4.8 Différence moyenne entre les valeurs de KGE en fonction du jeu de données et du bassin versant. Ici, la couleur rouge sous-entend une meilleure performance de ERA5 et une couleur bleue sous-entend une meilleure performance des OBS.

Tableau 4.5 Résumé des statistiques sur les différences entre les KGE issus des calibrations avec OBS versus les KGE avec ERA5. (La dominance définit le jeu de données dominant pour la région, la portion représente le pourcentage de bassins pour lesquels il y a la dominance)

	Maximum en faveur des OBS	Maximum en faveur d'ERA5	Moyenne	Dominance	Portion
Canada	-0,17	0,37	0,07	ERA5	82%
États-Unis	-0,45	0,48	-0,09	OBS	84%
Mexique	-0,31	0,21	-0,03	OBS	56%

Tel que le présentent les résultats généraux du tableau 4.5, il n'y a dominance en performance pour ERA5 qu'au Canada. Cette dominance est assez marquée par sa grande portion de bassins ayant une dominance d'ERA5, mais aussi par les différences maximales montrant une

distribution en faveur d'ERA5. En comparaison, lorsque la performance avec ERA5 est meilleure qu'avec les OBS, elle l'est significativement plus que lorsque les OBS offrent une meilleure performance. Aux États-Unis, il y a aussi une très grande dominance, mais dans ce cas ce sont les OBS. La dominance semble toutefois plus étendue. En regard des valeurs maximales du tableau 4.5, chacun des jeux de données performe similairement mieux que l'autre lorsque la différence est grande. L'étendue de cette distribution est assez distinctive. En effet, ERA5 performe généralement mieux dans le centre Nord et le nord-ouest (figure 4.8). Cela démontre bien la force d'ERA5 dans les régions montagneuses et concorde avec les valeurs au Canada montrant une nette supériorité de ERA5 dans les régions en montagne et les régions plus éloignées. Plus au centre, centre sud, au sud-est et l'est des États-Unis, la différence est flagrante en faveur des OBS, surtout pour les régions humides plus au sud. Les régions centrales plus arides et limitées en eau montrent toutefois une différence moins marquée. Au Mexique, les résultats sont assez répartis. Il y a une majorité des bassins offrant de meilleures performances par l'utilisation des OBS, mais elle est très faible avec seulement 56% des bassins montrant une supériorité en performance. De plus, 50% des valeurs de différences se situent approximativement entre -0,13 et 0,08. La performance entre les OBS et ERA5 est donc très semblable au Mexique.

Les statistiques pour les différences entre ERA5 et ERA5-Land ne sont pas incluses dans le tableau 4.5 étant donné le niveau de ressemblance entre la comparaison entre les OBS et ERA5. En fait les différences moyennes pour le Canada, les États-Unis et le Mexique sont respectivement de -0,01, -0,02 et -0,003. Il est jugé que ces faibles différences permettent d'inférer que les statistiques entre les OBS et ERA5-Land seraient très similaires et qu'elles ne changeraient pas les résultats. D'autant plus, les trois moyennes indiquent une dominance d'ERA5 sur l'entièreté de l'Amérique du Nord. C'est-à-dire que ERA5 performe en général mieux qu'ERA5-Land et donc l'analyse d'ERA5-Land contre les OBS serait plutôt accessoire.

En regard des caractéristiques physiographiques, les réanalyses se démarquent au niveau de l'élévation. En effet, les différences en performance sont généralement en grande faveur des réanalyses pour les régions des rocheuses et de la côte ouest pour le Canada et pour les régions

montagneuses des États-Unis, plus au nord et un peu sur la côte ouest. En regard de la figure 2.6, la différence de performance des OBS ou des réanalyses est plus importante pour les bassins versants ayant une limitation en énergie ou en eau alors que les performances sont généralement plus similaires pour les régions équilibrées. Pour ce qui est de la superficie des bassins versants, il ne semble pas y avoir de jeu de données qui se démarque. Autant des grands que des petits bassins montrent de meilleures performances avec les OBS ou ERA5. La figure 4.8 le montre bien.

En général, la figure 4.8 montre aussi une bonne corrélation spatiale entre la performance des réanalyses avec l'étude et la densité spatiales des stations météorologiques, tel que montré également dans Essou et al. (2017). C'est-à-dire que plus la densité de stations est faible, meilleures sont les réanalyses en comparaison avec les OBS. Encore une fois, les différences entre ERA5-Land et les OBS sont négligées en raison de la ressemblance entre les résultats de ERA5 et ceux de ERA5-Land. La figure montrant la différence de KGE entre ERA5 et ERA5-Land peut toutefois être consultée à l'annexe III.

4.3.2 Formulation d'estimation de l'ET

Les résultats suivant l'axe des formulations d'ETP représentent la moyenne des KGE obtenus pour chaque jeu de données. La section 4.3.1 montre bien que chaque jeu de données montre du potentiel et, qu'à l'exception des valeurs plus extrêmes, les performances entre les trois types de données sont très semblables. Cet axe tient donc plus des formules que des jeux de données. Les KGE sont moyennés par formule en combinant les jeux de données ensemble.

La figure 4.9 ci-dessous montre les résultats pour la formule de FAO-PM, qui selon la littérature représente l' ET_0 . Ici, seulement les résultats pour FAO-PM sont montrés puisqu'ils représentent sensiblement l'ensemble des résultats pour les autres formules. Les tendances entre les formules sont assez similaires sauf à l'exception des formules exhibant des valeurs plus extrêmes (bas ou haut) d'ETP par rapport aux autres dans la figure 4.4. Dans ces cas, les valeurs de KGE sont généralement plus basses, sans toutefois être complètement à l'opposé.

Par souci d'éviter la répétition, les autres figures sont donc omises, mais peuvent être consultées dans l'annexe III. La figure 4.9 montre d'ailleurs des résultats qui concordent avec Dallaire et al. (2021). La variabilité dans les valeurs estimées d'ETP se traduit dans la performance en calibration des modèles. La présentation des résultats est donc généralisée pour l'ensemble des formules.

En majorité pour le Canada, les résultats de calibration sont satisfaisants. En effet, la performance atteinte en moyenne est de 0,799, et plus de 94% des valeurs de KGE sont supérieures ou égales à 0,6. Pour ce qui est des États-Unis, les plus faibles performances dans le centre et au sud-est font descendre la moyenne qui est de 0,675. Il y a toutefois au moins 78% des bassins qui atteignent un KGE minimal de 0,6, soit des performances en calibration jugées adéquates. Puis pour le Mexique, la moyenne est de 0,651 avec plus de 74% de valeurs au-delà de 0,6.

Au niveau des caractéristiques physiques des bassins, il semble que ni l'aire, ni l'élévation n'aient un impact en considération de la formule d'estimation d'ETP utilisée. Autant des petits que des gros bassins versants ont de bonnes performances et pour chaque formule, les performances sont semblables en comparant en zones montagneuses ou encore en zones dont le relief est plus plat. L'impact sur la performance en modélisation est plutôt lié à la géolocalisation du bassin versant qu'à la formule d'ETP. En effet, les résultats en calibration sont considérablement plus faibles dans les régions limitées en eau pour chaque formule (figure 2.6). La formule d'ETP ne fait qu'introduire plus de variabilité dans la performance d'un modèle sans nécessairement rendre la performance nulle.

Il est à souligner que la variable d'ETR en provenance des réanalyses permet d'obtenir de très bons résultats. Ils ne sont pas significativement différents des résultats obtenus par d'autres formulations. C'est aussi le cas pour la variable de l'ETP. Cette dernière montre toutefois des résultats inférieurs à l'ETR, mais tout aussi satisfaisants. L'utilisation de l'ETP des réanalyses donne de bons résultats et en donnerait probablement des meilleurs en ne considérant pas les erreurs dans les valeurs de E5L-ETP. Ces valeurs sont présentées dans la figure de l'annexe

III. La différence de performance pour l'ETP relève en partie du fait que HOOPLA ne contient aucun coefficient d'ajustement, ni pour MOHYSE, ni pour GR4J qui normalement en ont. L'ETP n'est donc pas ajusté à une valeur plus réelle équivalente à l'ETR.

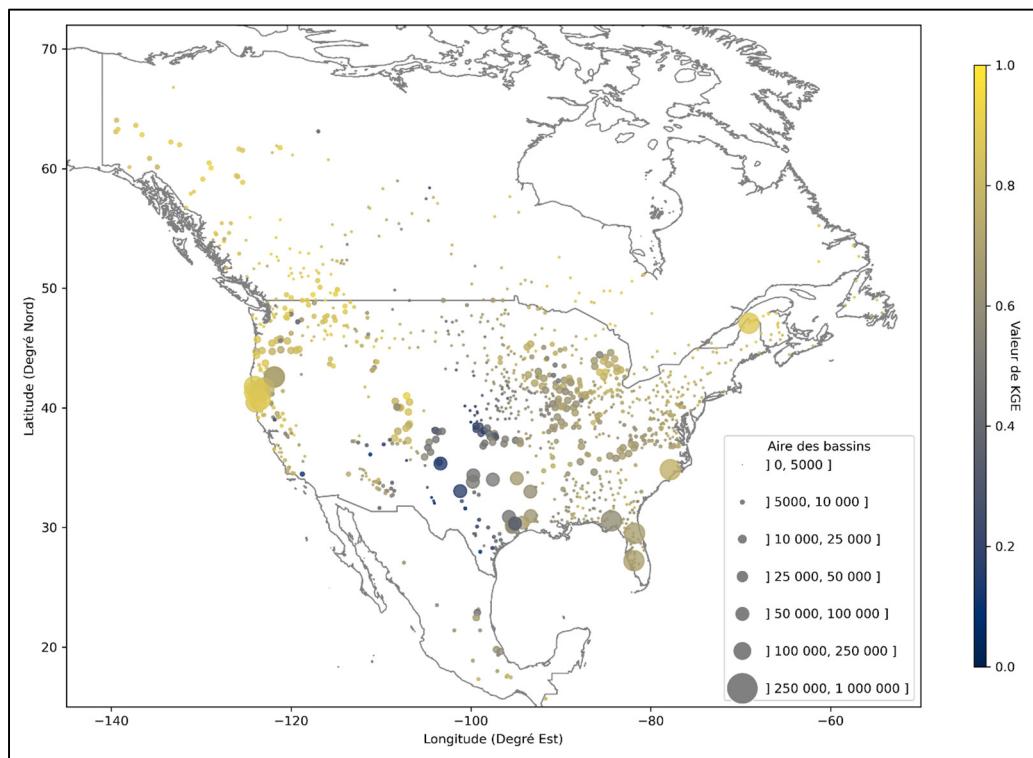


Figure 4.9 Distribution des valeurs de KGE en calibration pour la formule d'estimation de l'ETP FAO-PM

CHAPITRE 5

DISCUSSION

Ce chapitre offre une discussion sur les résultats. Les erreurs et incertitudes pouvant affecter les résultats sont d'abord abordées. Pour suivre les objectifs de l'étude, les résultats sont ensuite discutés en fonction de la représentativité physique des formules d'ETP et du potentiel des réanalyses en modélisation hydrologique. Le chapitre se termine par l'explication des limites de l'étude.

5.1 Corruption et exactitude des données de réanalyses

L'utilisation de sources de données provenant des réanalyses peut contenir son lot de problèmes et d'incertitudes. Certaines corruptions sont présentes dans les données de chacune des réanalyses. L'exactitude des valeurs est aussi un point important qui dépend grandement de la façon dont les valeurs sont calculées et à quel point les interactions avec le climat sont prises en compte. Cette section sert à expliquer ces corruptions et aborde l'exactitude des valeurs de certaines variables de ERA5 et ERA5-Land.

5.1.1 ERA5

Des deux réanalyses, ERA5 est la meilleure. Malgré qu'elle offre une résolution spatiale plus grossière, les valeurs sont plus certaines en omettant les biais connus (Hersbach et al., 2020; M. Tarek et al., 2020; Mostafa Tarek et al., 2020). L'exactitude des variables n'est donc pas remise en question dans ce cas-ci. ERA5 a été vérifiée maintes fois avant d'être publiée par ECMWF et il est possible d'utiliser les valeurs brutes. Par souci de précision, il est toujours possible de corriger le biais dans les données, qu'il s'agisse d'un biais systématique régional ou local qui peut exister dans un tel modèle. Une correction de biais peut mener à des valeurs

plus justes et peut par le fait même augmenter la performance en modélisation (Lompar, Lalić, Dekić, & Petrić, 2019; Ricard & Anctil, 2019). La littérature scientifique montre toutefois le potentiel d'ERA5 en modélisation hydrologique sans en avoir corrigé le biais (M. Tarek et al., 2020) dans les données. Étant donné le caractère exploratoire de la présente étude, le biais dans les données est donc considéré négligeable.

Aujourd’hui corrigées, les données d’ERA5 qui avaient été utilisées pour l’étude comprenaient une corruption (ECMWF, 2021). En soi, la corruption est très localisée et petite, elle correspond à la ligne grise près de la frontière entre le Canada et les États-Unis (figure 5.1). L’état de la corruption est analysé en regard du nombre de données affectées, en considérant un point de grille comme étant à une certaine longitude, latitude, date et heure. Pour chaque variable affectée, en termes de points de grilles, aucune corruption ne dépasse 1800 points alors que pour l’ensemble de l’Amérique du Nord, il y a environ 90 000 000 000 points. Les corruptions représentent donc une portion infime des points de grilles, c'est-à-dire 0,000002% des points inclus dans ce projet. La corruption est donc jugée non significative scientifiquement vis-à-vis de son impact sur les résultats, d’autant plus que les points de grilles sont moyennés à l’échelle du bassin versant avant de fournir les données aux modèles hydrologiques. Ce processus réduit davantage l’impact de la corruption. En raison de l’impact infime de la corruption et de la nature exploratoire de l’étude, il est jugé qu’il n’était pas nécessaire d’acquérir les nouvelles grilles de points corrigées. Le temps de calcul que cela aurait ajouté n’en valait pas la peine. La description exacte des variables affectées et de l’ampleur de la corruption peut être consultée à l’annexe IV.

ERA5 comprend aussi une corruption sur l’ensemble de l’Amérique du Nord pour les données de pression atmosphérique. Comme elle n’est pas listée dans les corruptions d’ERA5, une vérification auprès du support d’ECMWF a été faite (ECMWF, 2021). De leur côté, les valeurs sont justes et non corrompues. Il est donc évalué par l’auteur et le support d’ECMWF que la corruption est une erreur introduite lors du téléchargement des grilles de points. Elle correspond à une valeur fixe d’environ 76 893,98 Pa. Cette valeur est celle horaire pour la deuxième moitié de l’année 2006. Plus exactement, la corruption commence le 9 juin et se

termine le 31 décembre, représentant moins de 1,5 % des valeurs journalières par rapport à la longueur totale des séries temporelles (206 sur 14610 jours). L'effet de cette corruption est négligeable sur les estimations d'ETP avec FAO-PM. Pour la période touchée et pour l'ensemble des bassins, la différence de pression crée au maximum une différence absolue de 0,69 mm/j et en moyenne 0,36 mm/j. L'impact de la corruption est donc négligeable sur l'ensemble des résultats et sur les résultats en modélisation. Il a donc été décidé de ne pas corriger cette erreur et de seulement en tenir compte dans les résultats.

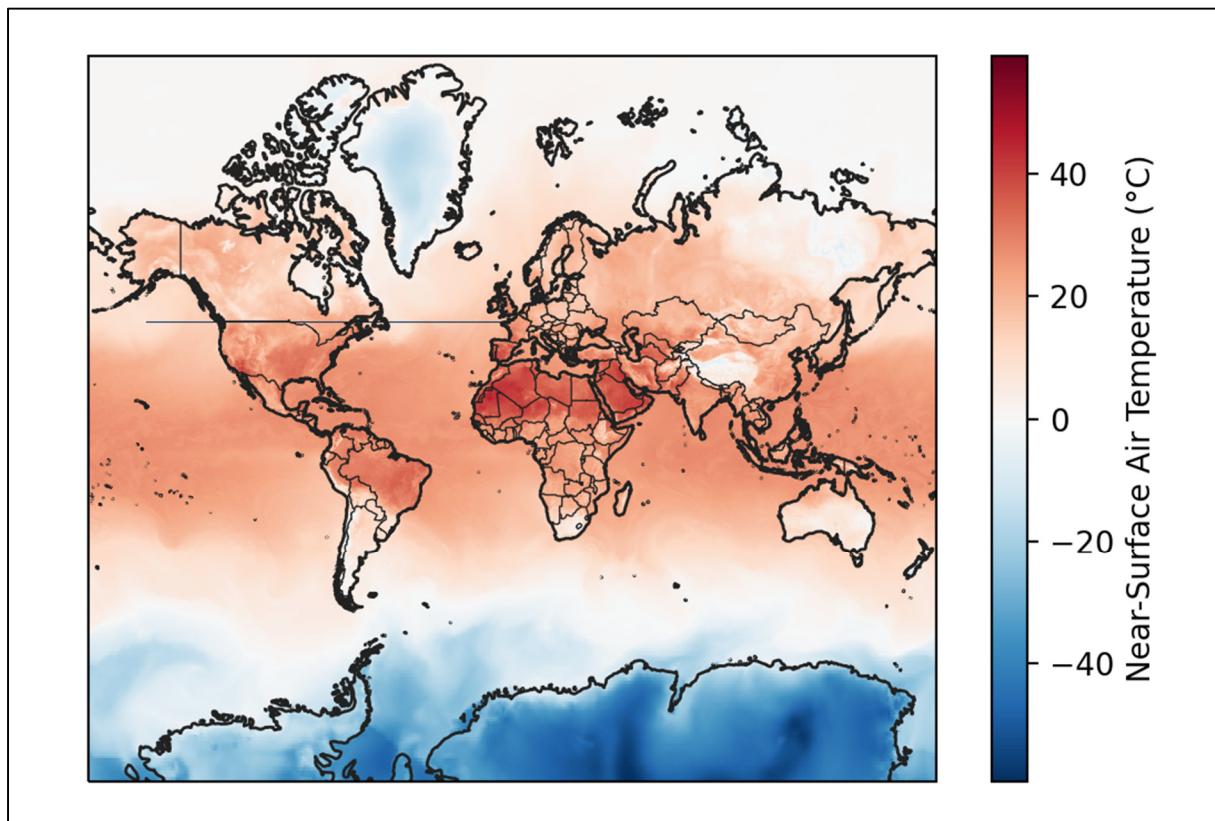


Figure 5.1 Exemple de la zone de la corruption (ligne grise) de ERA5 avec la température à 2m. Tiré de (ECMWF, 2021)

5.1.2 ERA5-Land

À ce jour, aucune corruption n'est connue pour ERA5-Land. Comme ils l'ont fait pour la corruption d'ERA5, ECMWF aurait publié un avis à ce sujet si une telle corruption dans ERA5-

Land était connue par l'organisation. L'exactitude des valeurs d'ETP est toutefois analysée. L'erreur moyenne montrée au tableau 4.4 entre les valeurs d'ERA5 et d'ERA5-Land souligne aussi l'impossibilité que les valeurs d'E5L-ETP représentent l'évaporation au niveau de la végétation. En effet, la moyenne des valeurs d'ETP pour tous les bassins et formulations combinés est de 3,11 mm/j. En considérant la différence moyenne dans le cas présent, les valeurs d'E5L-ETP seraient donc en moyenne de 11,05 mm/j. Cette valeur est normale en considérant la méthode de calcul employée, qui repose sur l'estimation de l'évaporation en bac, soit de l'évaporation de surface d'eau libre. Pour de l'ETP au niveau terrestre, c'est toutefois une valeur qui peut être considérée non adéquate. À noter que la moyenne de 3,11 mm/j contient aussi les valeurs erronées d'E5L-ETP. La moyenne, en ne considérant aucune valeur d'ETP en provenance de réanalyses, baisse la moyenne à 2,87 mm/j, accentuant l'écart entre les valeurs d'E5L-ETP et d'E5-ETP.

La grande différence notée ci-haut semble provenir de la méthode d'estimation de l'ETP pour E5L qui diffère d'E5. La méthode employée pour E5L, l'évaporation en bac tel que mentionné plus haut, est une méthode qui peut mener à de grandes incertitudes comme le décrit le premier chapitre. Les valeurs d'ETP obtenues à partir de cette technique d'estimation dépendent fortement de la DEA et peuvent donc mener à des surestimations ou sous-estimations dépendamment des autres variables entrant en jeu (radiation, vent et humidité). Comme le montre la figure 4.2, la totalité des valeurs inadéquates pour E5L-ETP sont situées dans des zones à risque près d'étendues d'eau et à l'est du continent où le biais de ERA5 et ERA5-Land est fort (Hersbach et al., 2020). Il a été essayé de retélécharger les grilles de points pour cette variable et d'analyser si cette erreur pouvait être dans la transformation des données, mais les erreurs persistent. Les erreurs sont ciblées dans une région spécifique, il serait donc surprenant que l'erreur provienne du traitement. Il est plus logique qu'elle soit à la source. Pour cette raison, il est recommandé d'utiliser ces données avec précautions.

5.2 Incertitude liée à la modélisation de l'ETP

La modélisation de l'ET peut constituer l'une des plus grandes sources d'incertitude dans la chaîne de modélisation hydrologique (Dallaire et al., 2021), particulièrement lorsqu'il est question de la modélisation des étiages. Bien que l'accent de la présente étude ne soit pas mis sur les étiages, plusieurs études montrent la dispersion des résultats d'une formule à l'autre dans des contextes de modélisation variés, tel qu'évoqué dans la revue de littérature du Chapitre 1. Le choix de la formule peut introduire une variabilité considérable dans les résultats en modélisation ou encore cette variabilité peut être introduite par les données d'entrée utilisées pour l'estimation de l'ETP.

C'est le cas pour la formule BC, entre autres. Elle comporte techniquement un paramètre libre de calage. Pour ce projet, l'effet de ce paramètre n'était pas pris en compte et donc la valeur arbitraire utilisée est la même que dans Dallaire et al. (2021) et Seiller et Anctil (2016). C'est une valeur qui permettait d'obtenir généralement de bons résultats en calibration pour ces deux études. Elle a donc été conservée pour la présente. La formulation de FAO-PM est aussi sujette à incertitude. Techniquement, cette formule a été développée pour de la modélisation hydrologique spécifiquement pour l'agriculture (R. G. Allen et al., 1998). Les estimations qu'elle produit représentent de l' ET_0 , à laquelle un coefficient de culture doit être appliqué dans le but d'obtenir les valeurs d'ETP. Pour l'étude, le but était de voir si l'utilisation d'une formule ayant des bases plus physiques permettait de générer des modèles bien calibrés. Il n'était pas question d'obtenir la meilleure représentation de la formulation de Penman-Monteith. Ainsi, les résultats d'ETP que cette formule donne sont utilisés dans un but exploratoire. L'incertitude reliée au coefficient de culture n'est pas considérée. Il est important de souligner toutefois que cette formulation, de base, donne des résultats satisfaisants en modélisation. Elle montre un bon potentiel et il est donc possible d'augmenter la performance des modèles avec cette formule en utilisant une forme plus adaptée de l'équation.

Certaines données pour le calcul de FAO-PM étaient manquantes ou devaient être transformées. En premier lieu, la variable de radiation solaire extra-terrestre pour ERA5-Land

n'était pas disponible. La comparaison des résultats d'ETP et en modélisation hydrologique n'est donc pas parfaite. La différence est toutefois négligée pour cette étude considérant la forte corrélation entre les valeurs d'ERA5 et celles calculées empiriquement pour la radiation solaire extra-terrestre (équation 3.7). En second lieu, pour chacune des réanalyses, les données de vitesse du vent ont dû être modifiées. FAO-PM considère des valeurs de vent à 2 m du sol, or les données de réanalyses n'offrent que les vents à 10 m. Ces valeurs sont réduites par la méthode proposée par la FAO (R. G. Allen et al., 1998). Elle utilise la hauteur de la végétation, qui dans la présente étude est considérée arbitrairement égale à la valeur proposée par la FAO soit de l'herbe courte. Ceci introduit une incertitude dans les données de vent et ultimement dans les performances et/ou l'adéquation de l'estimation de l'ETP. Considérant le caractère exploratoire de l'étude, cette incertitude est jugée comme n'ayant, à ce stade, pas d'impact significatif sur les résultats.

5.3 Taille des bassins versants

Dans l'étude, un filtre a été appliqué sur l'aire minimale des bassins versant afin d'assurer que ces derniers soient compatibles avec la modélisation au pas de temps journalier, mais aucun filtre sur la valeur maximale de superficie n'a été appliqué. Cela fait en sorte que l'échantillon comprend des bassins allant jusqu'à plusieurs centaines de milliers de kilomètres carrés (maximum de 613 000 km²), ce qui peut poser problème en modélisation hydrologique globale reposant sur une représentation de l'ensemble du bassin versant à l'aide d'une seule unité spatiale de simulation. Il est fort possible qu'il en résulte une perte d'information, comme par exemple dans le cas d'un événement pluvieux localisé. Malgré que négligé pour l'étude, l'inclusion de tels bassins peut être une source de bruit dans les résultats et insère de l'incertitude.

5.4 Représentation physique de l'ETP en modélisation hydrologique

La représentation des processus physiques sous-jacents à un modèle hydrologique est souvent mise de côté au profit de meilleures performances (Blöschl et al., 2019; Ricard & Anctil, 2019). Le degré d'adéquation d'un modèle avec les représentations plus physiques ne sous-entend

toutefois pas nécessairement de moins bons résultats en calibration. Il est vrai, d'un autre côté, que les changements climatiques demandent une meilleure compréhension des processus physiques gouvernant le cycle de l'eau. La littérature se multiplie sur le thème et, particulièrement pour l'ET, une augmentation en température ne signifie pas toujours une augmentation proportionnelle en ET. La présente étude montre que par l'utilisation de formules d'ETP à base de radiation et combinatoires, il est possible d'atteindre d'aussi bonnes, parfois meilleures, performances qu'avec des formules empiriques plus simples à base de température. L'utilisation de formules à base de radiation devrait en fait être le standard minimal pour l'estimation de l'ET. Il est possible d'utiliser ces formules avec souvent seulement la température et la géolocalisation d'un bassin versant. Il est donc possible d'augmenter la représentativité physique d'un modèle hydrologique sans demande en données supplémentaires, simplement en utilisant cette famille de formules.

Cet aspect est d'autant plus vrai et important en considérant que la formule de FAO-PM alimentée avec les données de réanalyses permet d'obtenir de bonnes calibrations. Cette formule est régulièrement retenue comme la référence en matière d'estimation d'ET, mais aussi comme étant celle permettant la meilleure représentation physique de l'ET. Elle est toutefois souvent laissée de côté en raison de sa forte demande en données d'entrée et de sa sensibilité à l'exactitude de ces données. Ce sont deux points qui dans le passé étaient importants, mais qui perdent de leur sens avec la disponibilité de données comme celles d'ERA5 qui offrent un accès facile à toutes les variables nécessaires et qui permettent l'obtention de bonnes calibrations avec le maximum de représentation physique au niveau de l'ET dans les modèles hydrologiques. Son utilisation plus fréquente pourrait permettre de réduire l'incertitude des modélisations hydrologiques en changements climatiques. En effet, les formules combinatoires sont celles montrant le plus de stabilité dans les résultats en changements climatiques. C'est-à-dire le moins de variance en comparant les valeurs futures aux valeurs de références. (Seiller & Anctil, 2016)

Les variables d'ETP et d'ETR des réanalyses montrent aussi un bon potentiel pour effectuer de la modélisation hydrologique, tout en considérant la représentativité physique. Il semble

possible d'éviter l'incertitude liée au choix de la formule d'ETP en utilisant directement ces valeurs. Les résultats montrent que les valeurs sont valides et qu'elles permettent d'atteindre de bonnes performances en calibration, à l'exception d'E5L-ETP, pour les raisons évoquées précédemment. Étant générées par un modèle climatique, les données d'ETR et d'ETP des réanalyses offrent des valeurs qui sont basées sur l'interaction entre un ensemble de variables climatiques et donc assurent, à tout le moins, une cohérence physique entre les différentes variables climatiques et, possiblement, une représentativité physique minimale de l'ET. En termes de physique, cet aspect place l'ETR et l'ETP des réanalyses à un niveau supérieur aux formules à base de température et à un niveau égal ou même supérieur aux formules à base de radiation ou combinatoires. Cet argument n'est pas quantifié, mais ressort de la simple logique de la science entourant la génération des données d'ETR et d'ETP des réanalyses en comparaison avec celles produites par les formules empiriques de l'étude. Ceci dit, l'auteur ne peut se prononcer sur les données d'ETP et d'ETR provenant des réanalyses des générations précédentes, lesquelles n'ont pas été analysées dans le cadre de la présente étude. De ce point, il ressort que, comme pour les formules combinatoires, l'utilisation des variables d'ETR ou d'ETP des réanalyses ERA5 ou ERA5-Land pourrait stabiliser les résultats en changements climatiques (Seiller & Anctil, 2016).

Cela touche grandement aussi le paradoxe d'évapotranspiration. Il est généralement envisagé qu'une augmentation des variables climatiques comme la température aura un même impact sur l'ET. Le paradoxe soutient par contre que malgré cette augmentation, avec les changements climatiques, il y aura aussi une augmentation du couvert nuageux et en concentration d'aérosols dans l'atmosphère. Le couvert nuageux pourrait affecter négativement le niveau de radiation solaire alors qu'une concentration plus forte en aérosols dans l'air réduirait la capacité de l'air à emmagasiner de la vapeur d'eau. Il pourrait donc en résulter une stabilisation ou même une diminution de l'ET dans le futur en lieu d'une augmentation tel qu'il est généralement envisagé. Ce paradoxe met donc bien en avant le besoin de représentation physique dans les modèles, surtout en étude d'impact des changements climatiques. (Dallaire et al., 2021)

5.5 Potentiel des réanalyses en modélisation hydrologique

La littérature comprend de maintes études sur le potentiel en modélisation hydrologique des données de réanalyses, surtout sur l'ancêtre d'ERA5, ERA-Interim, mais plus récemment sur ERA5 aussi (M. Tarek et al., 2020; Mostafa Tarek et al., 2020). Le but de cette étude n'était pas tant de valider ce point, mais plutôt de confirmer le potentiel en considérant l'utilisation de nouvelles variables et en n'utilisant que des données de réanalyses. Les résultats montrent que les réanalyses peuvent atteindre d'aussi bons résultats que les observations et même se démarquer dans des zones montagneuses.

La confirmation du potentiel en modélisation hydrologique permet l'utilisation plus généralisée des variables de réanalyses. Des études ont montré que de bonnes performances sont atteintes autant en utilisant que des données de réanalyses brutes, des données de réanalyses corrigées ou encore en combinant les réanalyses aux OBS en remplaçant les données et variables manquantes par celle des réanalyses (Lompar et al., 2019; Ricard & Anctil, 2019; M. Tarek et al., 2020). Les arguments contre les observations comme l'indisponibilité de variables, du manque de densité en stations météorologiques dans les régions éloignées ou encore de l'incertitude dans les grilles d'observations sont donc renforcés par la littérature grandissante sur l'utilisation des données de réanalyses aux fins de modélisation hydrologique. Le grand éventail de variables disponibles permet aussi l'utilisation de formulations plus complexes pour la représentation des processus physiques sous-jacents à un modèle. La complexité a souvent été délaissée au profit de la simplicité en partie en raison de la faible disponibilité en variables, mais ERA5 permet de pallier ce problème et ouvre la porte sur l'utilisation de modèles plus physiques.

En soi, pour cette étude, ERA5 et ERA5-Land donnent des résultats semblables en modélisation hydrologique. La résolution plus fine des grilles d'ERA5-Land ne change pas significativement les résultats lorsqu'utilisée de concert avec de très grands bassins et des modèles conceptuels. La résolution plus fine d'ERA5-Land comporte tout de même son lot de défis. Les fichiers de données brutes sont considérablement plus gros, soit environ 8 à 10 fois

par rapport à ERA5. Cette caractéristique requiert donc de l'utilisateur beaucoup plus d'espace de stockage dans l'ordinateur utilisé. Cela affecte aussi le traitement de données qui prend beaucoup plus de temps et qui demande plus d'espace en mémoire vive. La demande supplémentaire en matériel informatique est jugée non avantageuse pour l'utilisateur en considérant qu'ERA5-Land ne performe pas significativement mieux qu'ERA5, dans le contexte de la présente étude. Pour ce genre d'étude, il est donc recommandé d'opter pour ERA5 au lieu d'ERA5-Land. Il n'est pas possible de se prononcer par contre pour des études différemment paramétrées, par exemple si les modèles utilisés sont distribués ou avec de bien plus petits bassins versants.

5.6 Limites de l'étude

La nature exploratoire de l'étude la rend très générale et ne permet pas une analyse en profondeur, elle comporte donc certaines limites. Le but ultime de l'étude est de déterminer si les données d'ERA5 et d'ERA5-Land peuvent être plus généralement utilisées en modélisation hydrologique, plus particulièrement des variables rarement considérées comme le vent et la radiation solaire, mais qui permettent l'utilisation de formules comportant un caractère plus physique et plus complexe. Par souci de temps, il était d'intérêt de réduire le temps de calcul au maximum étant donné l'ampleur du secteur à l'étude. Les limitations affectent donc autant les données d'entrées physiographiques et météorologiques que les choix des modèles et des méthodologies.

Au niveau des caractéristiques physiographiques, seulement les bassins versants de superficie supérieure à 500 km² ont été pris en compte, il n'est donc pas possible de généraliser les résultats pour des bassins sous ce seuil. Les résultats obtenus pour le Mexique sont jumelés à ceux des États-Unis et du Canada pour permettre leur meilleure couverture des conditions hydroclimatiques et transférabilité, toutefois il faut s'abstenir de considérer seulement les résultats au niveau du Mexique afin de généraliser les résultats à l'échelle du pays en raison du faible nombre de bassins utilisés pour l'étude.

Pour ce qui est des données hydrométéorologiques, il est important de souligner que le but n'était pas de juger de leur exactitude, mais plutôt d'évaluer leur potentiel. Pour l'ensemble des données (OBS, ERA5, ERA5-Land), elles sont donc utilisées à l'état brut sans y avoir appliqué aucune correction préalablement. Ce choix peut avoir introduit de l'incertitude et un biais dans les résultats, mais ces derniers sont jugés non significatifs en considérant que le point de comparaison de la performance n'était que le KGE. Pour ce qui est de l'ETR et de l'ETP des réanalyses, aucune analyse de tendances et variabilité n'est faite. La validité des valeurs n'est que confirmée par comparaison générale avec des valeurs d'ETP obtenues empiriquement. Le but était d'explorer leur potentiel en modélisation hydrologique seulement. Leur capacité à capter la variabilité et les tendances est donc négligée pour cette étude.

Au niveau de la méthodologie, il n'était pas question de trouver les meilleures données en fonction de l'étape de modélisation. Pour chacun de jeux de données, la totalité des données disponibles était donc utilisée et aucun jeu de données n'a été mélangé. Cela crée une disparité dans les dates de début et de fin, par exemple ERA5 commence en 1979 alors qu'ERA5-Land en 1981 et les OBS se terminent en 2015 alors que les réanalyses vont jusqu'en 2018. Il en est de la même lignée pour la radiation solaire extra-terrestre qui a été utilisée pour ERA5 alors que pour les deux autres (OBS et ERA5-Land), cette variable a été calculée empiriquement. Ces différences sont considérées négligeables pour cette étude puisqu'il était désiré de maximiser l'utilisation des données disponibles afin d'évaluer le potentiel en modélisation hydrologique. Puis toujours en termes de méthodologie, il y a une limite importante au niveau de la généralisation des résultats en fonction du type de modèle. Les modèles utilisés sont globaux et conceptuels, ne tenant donc pas compte de plusieurs caractéristiques physiographiques des bassins versants. Il ne faut donc pas généraliser les résultats obtenus dans cette étude à d'autres types de modèles.

CONCLUSION

En modélisation hydrologique, un important défi à ce jour reste l'équilibre entre la sélection d'un modèle offrant une bonne caractérisation du régime des rivières, mais qui permet un temps de calcul raisonnable selon les conditions de l'étude visée. Pour cette raison, les utilisateurs se rangent souvent derrière des modèles simples et empiriques. Le penchant pour des modèles conceptuels se fait surtout en raison de la disponibilité en données fiables. Cette voie comporte un grand désavantage puisque la représentativité physique du modèle se voit affectée. C'est d'ailleurs un point important soulevé par la communauté en hydrologie puisque la représentativité physique d'un modèle permet sa transférabilité autant sur l'échelle spatiale que temporelle, une préoccupation grandissante surtout pour les études en changements climatiques (Blöschl et al., 2019).

Dans cet esprit, le but de cette étude était d'explorer de nouvelles solutions afin d'ouvrir la porte sur l'augmentation de la représentativité des modèles hydrologiques. La réponse à ce but s'est faite selon deux axes, soit (1) par l'évaluation du potentiel en modélisation hydrologique de nouvelles sources de données offrant une plus grande exhaustivité en variables et (2) par l'utilisation de formules de l'estimation de l'ETP nécessitant plus de données d'entrées, mais avec une meilleure représentation du processus physique.

Cette étude se base sur la comparaison des résultats en modélisation hydrologique issus de différentes combinaisons. Celles-ci ont été générées à l'aide de trois sources de données hydroclimatiques (grilles observationnelles, ERA5, ERA5-Land), deux modèles hydrologiques conceptuels et globaux (MOHYSE, GR4J) calibrés en fonction du critère KGE et de 14 formulations d'estimation de l'ETP. Chacune de ces combinaisons a été appliquée sur 1169 bassins versants répartis sur l'ensemble de l'Amérique du Nord entre le Canada, les États-Unis et le Mexique. Il est à noter que pour réellement tester le potentiel des différentes sources de données, aucune combinaison ne comprend le mélange de sources de données. Les

calibrations des modèles se sont faites par l'entremise de 100% des données disponibles pour chacune des sources.

En modélisation, les observations restent le chef de file si l'on considère la généralisation et si l'on regarde seulement les valeurs de KGE obtenues en calibration. Il y a toutefois des exceptions à cette règle surtout dans les zones pour lesquelles les observations sont rares ou pour lesquelles la fiabilité des observations est discutable. Cela est principalement le cas pour les régions montagneuses comme les rocheuses au Canada, les régions plus au nord du Canada et pour le Mexique. Dans ces cas, en général, de meilleurs résultats en calibration avec les données de réanalyses (ERA5 et ERA5-Land) sont observables. Les observations offrent un grand pouvoir de modélisation aux États-Unis. Le pays comporte un large réseau de stations météorologiques et donc les observations sont très représentatives, surtout pour les régions humides du pays à l'est et au sud-est, où ERA5 et ERA5-Land ont un biais important dans les valeurs de précipitations. Néanmoins, la supériorité en calibration des observations pour ces régions ne sous-entend pas de mauvais résultats pour les données de réanalyses. Ces dernières permettent quand même d'atteindre de résultats satisfaisants, il faut seulement être conscient de leur faiblesse dans ces cas.

Le grand potentiel des données de réanalyses aux fins de modélisation hydrologique permet d'ouvrir la porte sur la représentativité physique des modèles puisque cet aspect est synonyme d'une nécessité accrue en données climatiques. Or les réanalyses offrent cette disponibilité. Les résultats en modélisation hydrologique pour les combinaisons utilisant l'ETP à base plus physique (FAO-PM, E5-ETR, E5-ETP, E5L-ETR) montrent qu'il est possible d'obtenir des modèles aussi bien calibrés que par l'entremise de formulations d'ETP plus simples et plus empiriques. L'utilisation de formulations plus physiques permet toutefois d'augmenter la représentativité physique des modèles en comparaison aux autres formules d'ETP plus empiriques. En général, certaines formulations plus simples génèrent de meilleurs résultats et il y a une difficulté notable en modélisation dans la région centrale des États-Unis. Cela n'enlève toutefois rien au potentiel en modélisation hydrologique utilisant des méthodes d'estimation de l'ETP plus physique.

Cette étude montre le potentiel en modélisation hydrologique favorable des réanalyses ainsi que le potentiel de ces données combinées à des méthodes à base plus physiques afin de déterminer l'ETP, un processus important de l'hydrologie des bassins versants. Elle se voit toutefois très limitée au niveau de l'échelle temporelle et au niveau des autres processus sous-jacents aux modèles. Il serait intéressant d'étendre l'analyse des données de réanalyses en étude de changements climatiques. Avec un grand potentiel, il serait aussi intéressant de tester si l'exhaustivité des données de réanalyses permet d'augmenter davantage, outre l'ETP, la représentativité d'autres processus hydrologiques. Il ne faut pas oublier que les modèles hydrologiques restent l'une des plus grandes sources d'incertitude dans les résultats obtenus (Dallaire et al., 2021).

RECOMMANDATIONS

L'étude a exploré des eaux peu naviguées, particulièrement l'utilisation de l'ETP et l'ETR de réanalyses. Sa nature exploratoire a donc nécessité d'apposer plusieurs limites au champ d'étude. C'est pourquoi les recommandations sont surtout axées sur les possibilités d'expansion de l'étude.

Les réanalyses offrent un grand potentiel de modélisation hydrologique. Cette étude est à un pas de temps journalier, mais les données de réanalyses sont fournies au pas de temps horaire. Il semble donc y avoir un potentiel en modélisation sous-journalière qui pourrait être étudié et qui serait d'intérêt pour les plus petits bassins versants (dont la superficie est inférieure à 500 km²).

La différence de résolution entre ERA5 et ERA5-Land n'a pas vraiment affecté les résultats dans la présente étude. Il pourrait donc être intéressant d'évaluer si la résolution plus fine d'ERA5-Land en vaut la peine sous différents angles, notamment en utilisant des modèles hydrologiques distribués ou en incluant dans le secteur étudié des bassins de petite taille (sous les 500 km²)

Finalement, les études les plus intéressantes seraient de tester potentiel des réanalyses ERA5 et ERA5-Land en étude d'impacts des changements climatiques en lieu des observations. C'est un terrain non exploré et avec leur potentiel en modélisation hydrologique, elles montrent du potentiel en changements climatiques. Ces études pourraient d'ailleurs être combinées à l'utilisation de FAO-PM ou des variables d'ETR et d'ETP comme estimateurs de l'ET.

ANNEXE I

PARAMÈTRES ET BORNES LIMITES DES MODÈLES HYDROLOGIQUES MOHYSE ET GR4J

Tableau-A I-1 Paramètres et bornes pour MOHYSE

Paramètre	Description	Borne inférieure	Borne supérieure
X(1)	Coefficient de transpiration	0	5
X(2)	Taux d'infiltration maximal	0,005	1000
X(3)	Coefficient de vidange (vadose – aquifère)	0	1
X(4)	Coefficient de vidange (vadose – rivière)	0	1
X(5)	Coefficient de vidange (aquifère – rivière)	0	1
X(6)	Paramètre de forme de l'hydrogramme unitaire	1	10
X(7)	Paramètre d'échelle de l'hydrogramme unitaire	0,1	20

Tableau-A I-2 Paramètres et bornes pour GR4J

Paramètre	Description	Borne inférieure	Borne supérieure
X(1)	Capacité maximale du réservoir de production	10	1000
X(2)	Coefficient d'échange sous-terrain	-5	3
X(3)	Capacité maximale du réservoir de routage un jour à l'avance	5	300
X(4)	Temps de base de l'hydrogramme unitaire	1	16

Tableau-A I-3 Paramètres et bornes pour CemaNeige

Paramètre	Description	Borne inférieure	Borne supérieure
X(1)	Coefficient thermique du manteau de neige	0	1
X(2)	Coefficient de fonte	0	20

ANNEXE II

ÉVAPOTRANSPIRATION POTENTIELLE ANNUELLE

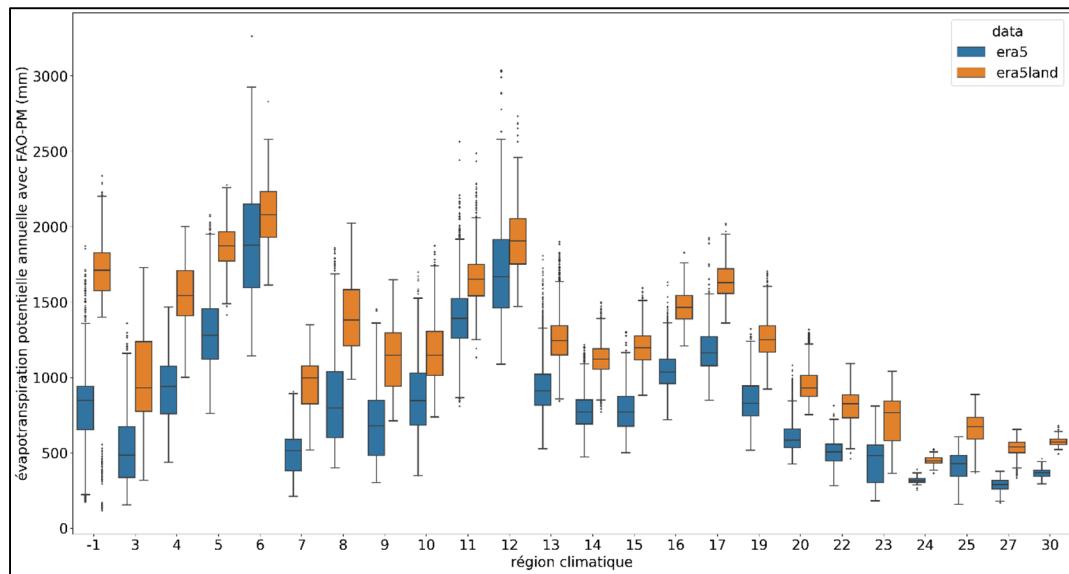


Figure-A II-1 Évapotranspiration potentielle annuelle moyenne obtenue par la formule de FAO-PM groupée par régions climatiques\

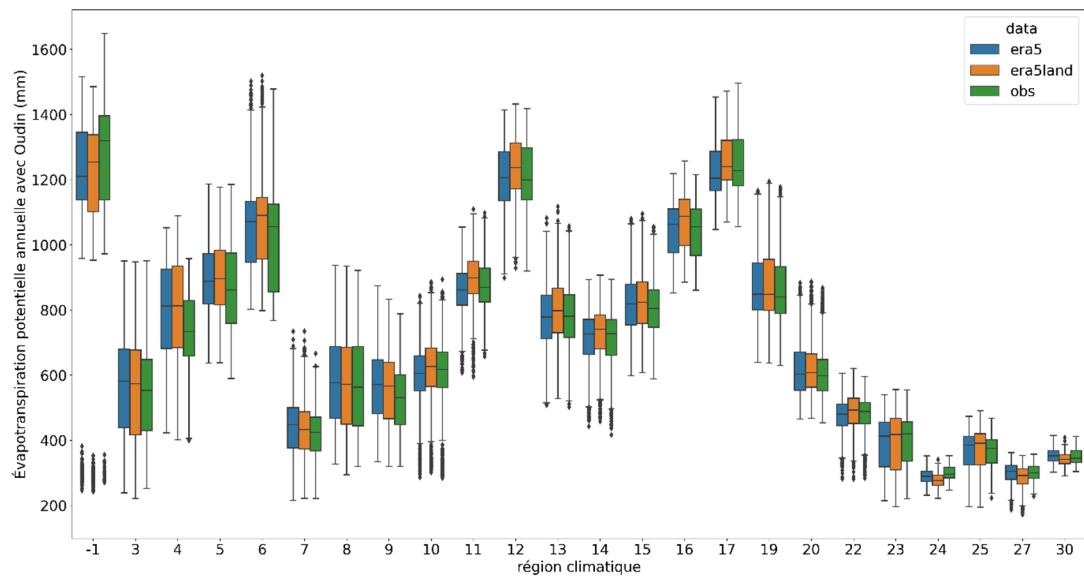


Figure-A II-2 Évapotranspiration potentielle annuelle moyenne obtenue par la formule de Oudin groupée par régions climatiques

ANNEXE III

RÉSULTATS DE PERFORMANCE EN CALIBRATION

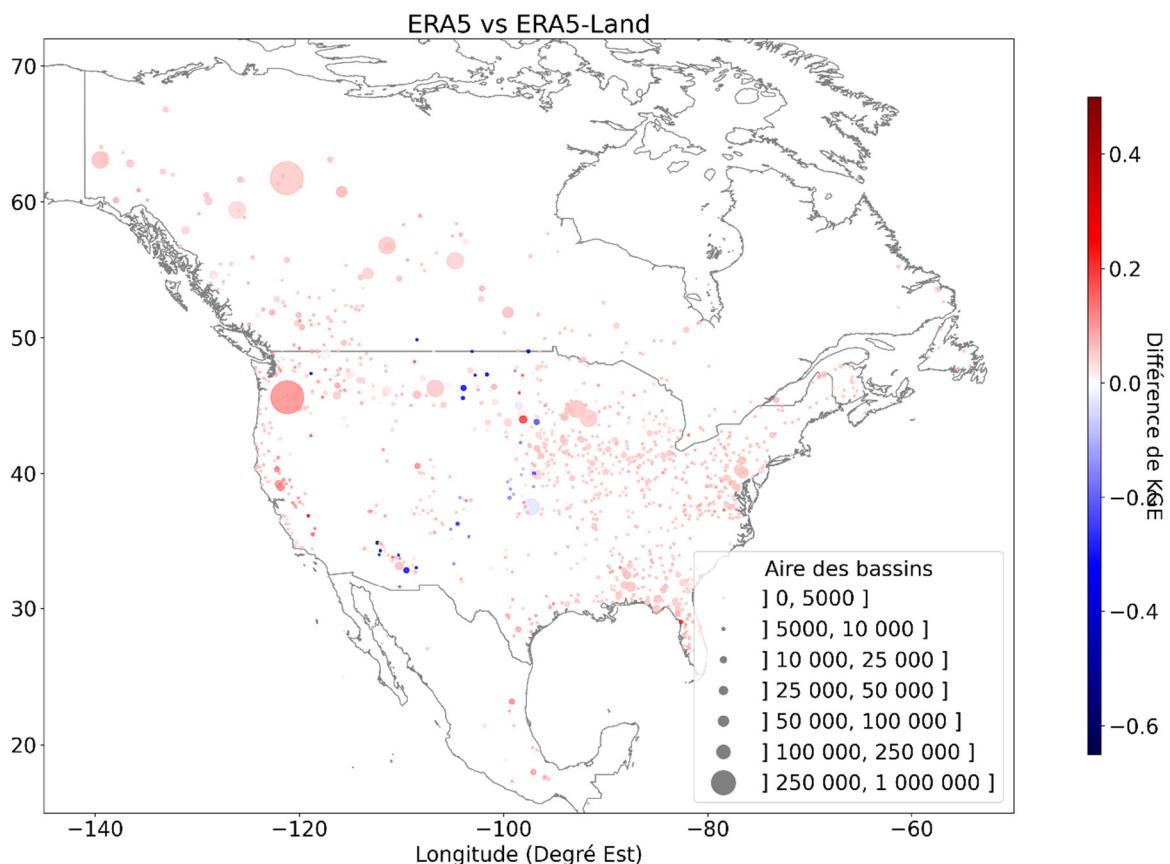


Figure-A III-1 Différence moyenne entre les valeurs de KGE en fonction du jeu de données et du bassin versant

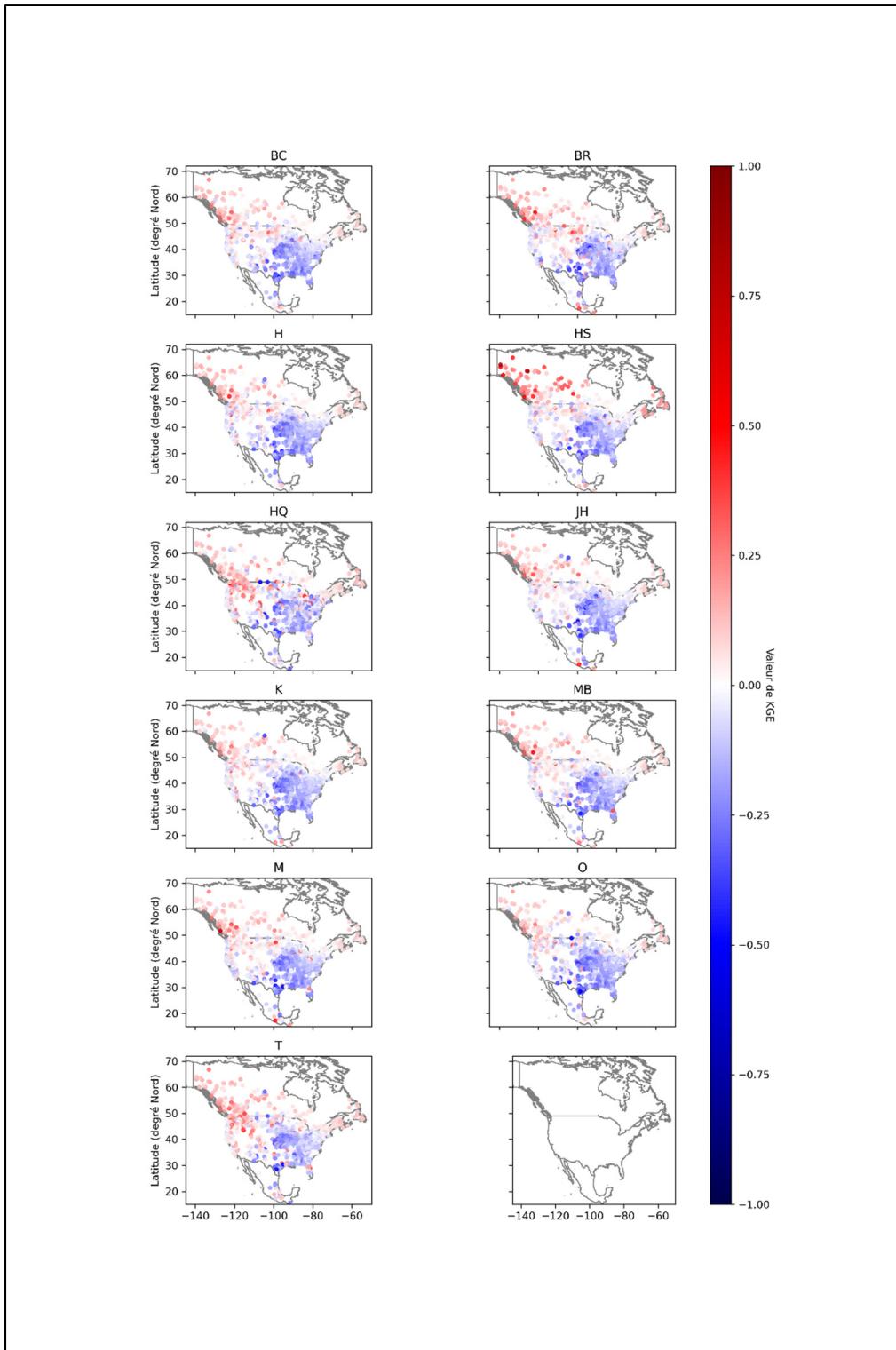


Figure-A-III-2 Différence moyenne entre les valeurs de KGE pour les jeux de données OBS et E5 par formule d'ET

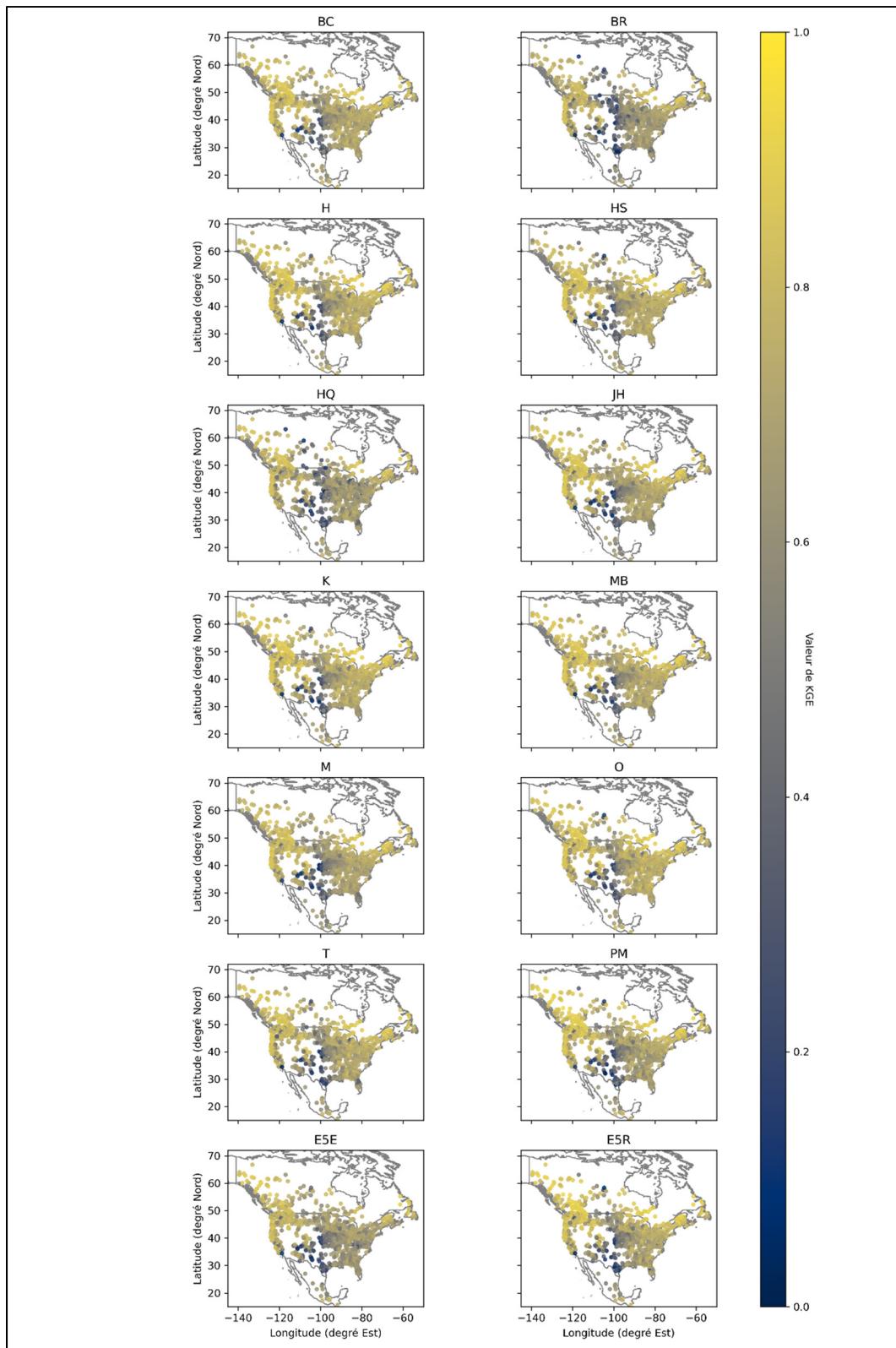


Figure-A III-3 Valeurs de KGE en fonction de la formule d'ETP et du bassin versant.

ANNEXE IV

CORRUPTION D'ERA5

Tableau-A IV-1 Description de l'étendue de la corruption des données d'ERA5.

Variable	Date	Heure	Nombre de points de grille affectés
Température de l'air à 2m	1988/07/26	8:00	1112
	1990/07/19	17:00	1751
	2014/05/07	2:00	151
Température du point de rosé à 2m	1980/03/03	23:00	710
	1996/08/17	16:00	1771
Vitesse du vent en direction est à 10m	1980/10/26	5:00	1471
	1983/09/10	9:00	356
Évaporation	1986/07/22	17:00	1957
	2013/06/08	21:00	1452
	2014/04/16	19:00	1710

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- Allen, M. R., & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419(6903), 228-232. doi: 10.1038/nature01092. Repéré à <https://doi.org/10.1038/nature01092>
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration. Guidelines for computing crop water requirements.
- Arsenault, R., Brissette, F., & Martel, J.-L. (2018). The hazards of split-sample validation in hydrological model calibration. *Journal of Hydrology*, 566, 346-362. doi: <https://doi.org/10.1016/j.jhydrol.2018.09.027>. Repéré à <https://www.sciencedirect.com/science/article/pii/S0022169418307145>
- Arsenault, R., Brissette, F., Martel, J.-L., Troin, M., Lévesque, G., Davidson-Chaput, J., . . . Poulin, A. (2020). A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds. *Scientific Data*, 7(1). doi: 10.1038/s41597-020-00583-2. Repéré à <https://dx.doi.org/10.1038/s41597-020-00583-2>
- Arsenault, R., Poulin, A., Côté, P., & Brissette, F. (2014). Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration. *Journal of Hydrologic Engineering*, 19(7), 1374-1384. doi: 10.1061/(ASCE)HE.1943-5584.0000938. Repéré à [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000938](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000938)
- Baier, W., & Robertson, G. W. (1965). Estimation of latent evaporation from simple weather observations. *Canadian journal of plant science*, 45(3), 276-284. doi: 10.1139/cjss-2017-0112
- Blaney, H. F., & Criddle, W. D. (1950). *Determining water requirements in irrigated areas from climatological and irrigation data*. Washington, DC: US Department of Agriculture, Soil Conservation Service.
- Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., . . . Zhang, Y. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrological Sciences Journal*, 64(10), 1141-1158. doi: 10.1080/02626667.2019.1620507. Repéré à <https://dx.doi.org/10.1080/02626667.2019.1620507>

- Bormann, H. (2011). Sensitivity analysis of 18 different potential evapotranspiration models to observed climatic change at German climate stations. *Climatic Change*, 104(3), 729-753. doi: 10.1007/s10584-010-9869-7. Repéré à <https://doi.org/10.1007/s10584-010-9869-7>
- Bosilovich, M. G. (2013). Regional Climate and Variability of NASA MERRA and Recent Reanalyses: U.S. Summertime Precipitation and Temperature. *Journal of Applied Meteorology and Climatology*, 52(8), 1939-1951. doi: 10.1175/jamc-d-12-0291.1. Repéré à <https://dx.doi.org/10.1175/jamc-d-12-0291.1>
- Breña-Naranjo, J. A., Laverde-Barajas, M. Á., & Pedrozo-Acuña, A. (2017). Changes in pan evaporation in Mexico from 1961 to 2010. *International Journal of Climatology*, 37(1), 204-213. doi: <https://doi.org/10.1002/joc.4698>. Repéré à <https://doi.org/10.1002/joc.4698>
- Bukovsky, M. S. (2011). Masks for the Bukovsky regionalization of North America. Consultée le 2020-09-22, à <http://www.narccap.ucar.edu/contrib/bukovsky/>
- Choi, W., Kim, S. J., Rasmussen, P. F., & Moore, A. R. (2009). Use of the North American Regional Reanalysis for Hydrological Modelling in Manitoba. *Canadian Water Resources Journal*, 34(1), 17-36. doi: 10.4296/cwrj3401017. Repéré à <https://dx.doi.org/10.4296/cwrj3401017>
- Craig, J. R., Brown, G., Chlumsky, R., Jenkinson, R. W., Jost, G., Lee, K., . . . Tolson, B. A. (2020). Flexible watershed simulation with the Raven hydrological modelling framework. *Environmental Modelling & Software*, 129, 104728. doi: <https://doi.org/10.1016/j.envsoft.2020.104728>. Repéré à <https://www.sciencedirect.com/science/article/pii/S136481521931165X>
- Dai, A. (2006). Recent Climatology, Variability, and Trends in Global Surface Humidity. *Journal of Climate*, 19(15), 3589-3606. Repéré à <https://www.proquest.com/scholarly-journals/recent-climatology-variability-trends-global/docview/222867054/se-2?accountid=27231>
- Dallaire, G. (2019). *Étude de l'incertitude liée à la modélisation de l'évapotranspiration potentielle dans les études d'impact des changements climatiques sur les ressources en eau au Canada, aux États-Unis et au Mexique* (Ecole de technologie supérieure, Montreal).
- Dallaire, G., Poulin, A., Arsenault, R., & Brissette, F. (2021). Uncertainty of potential evapotranspiration modelling in climate change impact studies on low flows in North America. *Hydrological Sciences Journal*, 66(4), 689-702. doi: 10.1080/02626667.2021.1888955. Repéré à <https://doi.org/10.1080/02626667.2021.1888955>

- Dallaire, G. (2019) *Étude de l'incertitude liée à la modélisation de l'évapotranspiration potentielle dans les études d'impact des changements climatiques sur les ressources en eau au Canada, aux États-Unis et au Mexique.* (Mémoire de maîtrise, Ecole de technologie supérieure, Montréal)
- Do, H. X., Gudmundsson, L., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. *Earth System Science Data*, 10(2), 765-785. doi: 10.5194/essd-10-765-2018. Repéré à <https://dx.doi.org/10.5194/essd-10-765-2018>
- Donohue, R. J., McVicar, T. R., & Roderick, M. L. (2010). Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate. *Journal of Hydrology*, 386(1), 186-197. doi: <https://doi.org/10.1016/j.jhydrol.2010.03.020>. Repéré à <https://www.sciencedirect.com/science/article/pii/S0022169410001460>
- ECMWF. (2021). ERA5 CDS: Data corruption. Repéré à <https://confluence.ecmwf.int/display/CKB/ERA5+CDS%3A+Data+corruption>
- Essou, G. R. C., Brissette, F., & Lucas-Picher, P. (2017). The Use of Reanalyses and Gridded Observations as Weather Input Data for a Hydrological Model: Comparison of Performances of Simulated River Flows Based on the Density of Weather Stations. *Journal of Hydrometeorology*, 18(2), 497-513. doi: 10.1175/jhm-d-16-0088.1. Repéré à <https://dx.doi.org/10.1175/jhm-d-16-0088.1>
- Essou, G. R. C., Sabarly, F., Lucas-Picher, P., Brissette, F., & Poulin, A. (2016). Can Precipitation and Temperature from Meteorological Reanalyses Be Used for Hydrological Modeling? *Journal of Hydrometeorology*, 17(7), 1929-1950. doi: 10.1175/jhm-d-15-0138.1. Repéré à <https://dx.doi.org/10.1175/jhm-d-15-0138.1>
- Essou, G. R. C., (2016) *Potentiel des données de précipitation et température des réanalyses atmosphériques en modélisation hydrologique.* (Mémoire de maîtrise, Ecole de technologie supérieure, Montréal)
- Feng, T., Su, T., Zhi, R., Tu, G., & Ji, F. (2019). Assessment of actual evapotranspiration variability over global land derived from seven reanalysis datasets. *International Journal of Climatology*, 39(6), 2919-2932. doi: 10.1002/joc.5992. Repéré à <https://dx.doi.org/10.1002/joc.5992>
- Fortin, V. (2000). *Le modèle météo-apport HSAMI: historique, théorie et application.* Varennes, QC : Institut de recherche d'Hydro-Québec.

Fortin, V., & Turcotte, R. (2006). *Le modèle hydrologique MOHYSE*. SCA7420 : Modèles hydrologique. Université du Québec à Montréal.

Gudmundsson, L., Do, H. X., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment. *Earth System Science Data*, 10(2), 787-804. doi: 10.5194/essd-10-787-2018. Repéré à <https://dx.doi.org/10.5194/essd-10-787-2018>

Gupta, H. V., 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(1), 80-91. doi: <https://doi.org/10.1016/j.jhydrol.2009.08.003>. Repéré à <https://www.sciencedirect.com/science/article/pii/S0022169409004843>

Hamon, W. R. (1960). *Estimating potential evapotranspiration*. (Thèse de doctorat, Massachusetts Institute of Technology, États-Unis).

Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied engineering in agriculture*, 1(2), 96-99. doi: 10.13031/2013.26773

Harpold, A. A., Kaplan, M. L., Klos, P. Z., Link, T., McNamara, J. P., Rajagopal, S., . . . Steele, C. M. (2017). Rain or snow: hydrologic processes, observations, prediction, and research needs. *Hydrol. Earth Syst. Sci.*, 21(1), 1-22. doi: 10.5194/hess-21-1-2017. Repéré à <https://hess.copernicus.org/articles/21/1/2017/>

Her, Y., & Chaubey, I. (2015). Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty. *Hydrological Processes*, 29(19), 4220-4237. doi: <https://doi.org/10.1002/hyp.10487>. Repéré à <https://doi.org/10.1002/hyp.10487>

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999-2049. doi: <https://doi.org/10.1002/qj.3803>. Repéré à <https://doi.org/10.1002/qj.3803>

Huntington, T. G. (2006). Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology*, 319(1), 83-95. doi: <https://doi.org/10.1016/j.jhydrol.2005.07.003>. Repéré à <https://www.sciencedirect.com/science/article/pii/S0022169405003215>

- Huot, P.-L., Poulin, A., Audet, C., & Alarie, S. (2019). A hybrid optimization approach for efficient calibration of computationally intensive hydrological models. *Hydrological Sciences Journal*, 64(10), 1204-1222. doi: 10.1080/02626667.2019.1624922. Repéré à <https://dx.doi.org/10.1080/02626667.2019.1624922>
- Huot, P.-L., (2014) *Évaluation de méthodes d'optimisation pour le calage efficace de modèles hydrologiques coûteux en temps de calcul.* (Mémoire de maîtrise, Ecole de technologie supérieure, Montreal)
- IPCC (2021). Climate change 2021 the Physical Science Basis. Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdoms and NewYork, NY, USA. doi: 10.1017/9781009157896
- Jensen, M. E., & Haise, H. R. (1963). Estimating evapotranspiration from solar radiation. *Proceedings of the American Society of Civil Engineers, Journal of the Irrigation and Drainage Division*, 89, 15-41.
- Kharrufa, N. S. (1985). Simplified equation for evapotranspiration in arid regions. *Beiträge zur Hydrologie*, 5(1), 39-47.
- Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424-425, 264-277. doi: <https://doi.org/10.1016/j.jhydrol.2012.01.011>. Repéré à <https://www.sciencedirect.com/science/article/pii/S0022169412000431>
- Knoben, W., Freer, J., & Woods, R. (2019). Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth System Sciences Discussions*, 1-7. doi: 10.5194/hess-2019-327
- Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., ... & Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the US, and Southern Canada 1950–2013. *Scientific data*, 2, 150042. doi:10.1038/sdata.2015.42
- Lompar, M., Lalić, B., Dekić, L., & Petrić, M. (2019). Filling Gaps in Hourly Air Temperature Data Using Debiased ERA5 Data. *Atmosphere*, 10(1), 13. Repéré à <https://www.mdpi.com/2073-4433/10/1/13>

- Manzanas, R., Amekudzi, L. K., Preko, K., Herrera, S., & Gutiérrez, J. M. (2014). Precipitation variability and trends in Ghana: An intercomparison of observational and reanalysis products. *Climatic Change*, 124(4), 805-819. doi: 10.1007/s10584-014-1100-9. Repéré à <https://dx.doi.org/10.1007/s10584-014-1100-9>
- Mao, Y., & Wang, K. (2017). Comparison of evapotranspiration estimates based on the surface water balance, modified Penman-Monteith model, and reanalysis data sets for continental China. *Journal of Geophysical Research: Atmospheres*, 122(6), 3228-3244. doi: 10.1002/2016jd026065. Repéré à <https://dx.doi.org/10.1002/2016jd026065>
- Martins, D. S., Paredes, P., Raziei, T., Pires, C., Cadima, J., & Pereira, L. S. (2017). Assessing reference evapotranspiration estimation from reanalysis weather products. An application to the Iberian Peninsula. *International Journal of Climatology*, 37(5), 2378-2397. doi: 10.1002/joc.4852. Repéré à <https://dx.doi.org/10.1002/joc.4852>
- McGuinness, J. L., & Bordne, E. F. (1972). *A comparison of lysimeter-derived potential evapotranspiration with computed values* (No. 1452). Washington, DC: US Dept. of Agriculture.
- McVicar, T., Roderick, M., Donohue, R., & Van Niel, T. (2012). Ecohydrology bearings—Invited commentary less bluster ahead? Ecohydrological implications of global trends of terrestrial near-surface wind speeds. *Ecohydrology*, 5. doi: 10.1002/eco.1298
- McVicar, T. R., Roderick, M. L., Donohue, R. J., Li, L. T., Van Niel, T. G., Thomas, A., . . . Dinpashoh, Y. (2012). Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *Journal of Hydrology*, 416-417, 182-205. doi: <https://doi.org/10.1016/j.jhydrol.2011.10.024>. Repéré à <https://www.sciencedirect.com/science/article/pii/S0022169411007487>
- Milly, P. C. D., & Dunne, K. A. (2011). On the Hydrologic Adjustment of Climate-Model Projections: The Potential Pitfall of Potential Evapotranspiration. *Earth Interactions*, 15(1), 1-14. doi: 10.1175/2010EI363.1. Repéré à <https://journals.ametsoc.org/view/journals/eint/15/1/2010ei363.1.xml>
- Morton, F. I. (1983). Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. *Journal of Hydrology*, 66(1), 1-76. doi: [https://doi.org/10.1016/0022-1694\(83\)90177-4](https://doi.org/10.1016/0022-1694(83)90177-4). Repéré à <https://www.sciencedirect.com/science/article/pii/0022169483901774>
- Ohmura, A., & Wild, M. (2002). Is the Hydrological Cycle Accelerating? *Science*, 298, 1345-1346.
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model?

Journal of Hydrology, 303(1-4), 290-306. doi: 10.1016/j.jhydrol.2004.08.026. Repéré à <https://dx.doi.org/10.1016/j.jhydrol.2004.08.026>

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005a). Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of hydrology*, 303(1-4), 290-306. doi: 10.1016/j.jhydrol.2004.08.026

Oudin, L., Michel, C., & Anctil, F. (2005b). Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 1—Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs?. *Journal of Hydrology*, 303(1-4), 275- 289. doi: 10.1016/j.jhydrol.2004.08.025

Pelosi, A., Terribile, F., D'Urso, G., & Chirico, G. B. (2020). Comparison of ERA5-Land and UERRA MESCAN-SURFEX Reanalysis Data with Spatially Interpolated Weather Observations for the Regional Assessment of Reference Evapotranspiration. *Water*, 12(6). doi: 10.3390/w12061669

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4), 275-289. doi: 10.1016/s0022-1694(03)00225-7. Repéré à [https://dx.doi.org/10.1016/s0022-1694\(03\)00225-7](https://dx.doi.org/10.1016/s0022-1694(03)00225-7)

Poulin, A., Brissette, F., Leconte, R., Arsenault, R., & Malo, J.-S. (2011). Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *Journal of Hydrology*, 409(3), 626-636. doi: <https://doi.org/10.1016/j.jhydrol.2011.08.057>. Repéré à <https://www.sciencedirect.com/science/article/pii/S0022169411006135>

Ramanathan, V., Crutzen, P. J., Kiehl, J. T., & Rosenfeld, D. (2001). Aerosols, Climate, and the Hydrological Cycle. *Science*, 294(5549), 2119-2124. doi: 10.1126/science.1064034. Repéré à <https://doi.org/10.1126/science.1064034>

Raza, A., Shoaib, M., Faiz, M. A., Baig, F., Khan, M. M., Ullah, M. K., & Zubair, M. (2020). Comparative Assessment of Reference Evapotranspiration Estimation Using Conventional Method and Machine Learning Algorithms in Four Climatic Regions. *Pure and Applied Geophysics*, 177(9), 4479-4508. doi: 10.1007/s00024-020-02473-5. Repéré à <https://dx.doi.org/10.1007/s00024-020-02473-5>

Ricard, S., & Anctil, F. (2019). Forcing the Penman-Montheith Formulation with Humidity, Radiation, and Wind Speed Taken from Reanalyses, for Hydrologic Modeling. *Water*, 11(6). doi: 10.3390/w11061214

Ruggieri, G., Allocca, V., Borfecchia, F., Cusano, D., Marsiglia, P., & De Vita, P. (2021). Testing Evapotranspiration Estimates Based on MODIS Satellite Data in the Assessment of the Groundwater Recharge of Karst Aquifers in Southern Italy. *Water*, 13(2), 118. doi: 10.3390/w13020118. Repéré à <https://dx.doi.org/10.3390/w13020118>

Seiller, G., & Anctil, F. (2016). How do potential evapotranspiration formulas influence hydrological projections? *Hydrological Sciences Journal*, 61(12), 2249-2266. doi: 10.1080/02626667.2015.1100302. Repéré à <https://doi.org/10.1080/02626667.2015.1100302>

Seiller, G., Anctil, F., & Perrin, C. (2012). Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. *Hydrol. Earth Syst. Sci.*, 16(4), 1171-1189. doi: 10.5194/hess-16-1171-2012. Repéré à <https://hess.copernicus.org/articles/16/1171/2012/>

Sposito, G. (2017). Understanding the Budyko Equation. *Water*, 9(4), 236. doi: 10.3390/w9040236. Repéré à <https://dx.doi.org/10.3390/w9040236>

Srivastava, P. K., Han, D., Ramirez, M. A. R., & Islam, T. (2013). Comparative assessment of Evapotranspiration derived from NCEP and ECMWF global datasets through Weather Research and Forecasting model. *Atmospheric Science Letters*, 14, 118-125.

Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrol. Earth Syst. Sci.*, 24(5), 2527-2544. doi: 10.5194/hess-24-2527-2020. Repéré à <https://hess.copernicus.org/articles/24/2527/2020/>

Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Large-Scale Analysis of Global Gridded Precipitation and Temperature Datasets for Climate Change Impact Studies. *Journal of Hydrometeorology*, 21(11), 2623-2640. doi: 10.1175/JHM-D-20-0100.1. Repéré à <https://journals.ametsoc.org/view/journals/hydr/21/11/JHM-D-20-0100.1.xml>

Thiboult, A., Seiller, G., Poncelet, C., & Anctil, F. (2020). The HOOPLA toolbox: a HydrOlOgical Prediction LABoratory to explore ensemble rainfall-runoff modeling. *Hydrol. Earth Syst. Sci. Discuss.*, 2020, 1-18. doi: 10.5194/hess-2020-6. Repéré à <https://hess.copernicus.org/preprints/hess-2020-6/>

Thorntwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical review*, 38(1), 55-94. doi : 10.2307/210739

- Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43(1), n/a-n/a. doi: 10.1029/2005wr004723. Repéré à <https://dx.doi.org/10.1029/2005wr004723>
- Valéry, A. (2010). *Modeling precipitation flow under snow influence: Elaboration of a snow module and evaluation on 380 catchment areas.* (Doctorat Hydrobiologie, Institut des Sciences et Industries du Vivant et de l'Environnement AgroParisTech). Repéré à <https://hal.inrae.fr/tel-02594605>
- Valéry, A., Andréassian, V., & Perrin, C. (2014). ‘As simple as possible but not simpler’: What is useful in a temperature-based snow-accounting routine? Part 1 – Comparison of six snow accounting routines on 380 catchments. *Journal of Hydrology*, 517, 1166-1175. doi: <https://doi.org/10.1016/j.jhydrol.2014.04.059>. Repéré à <https://www.sciencedirect.com/science/article/pii/S0022169414003333>
- Wild, M. (2009). Global dimming and brightening: A review. *Journal of Geophysical Research: Atmospheres*, 114(D10). doi: <https://doi.org/10.1029/2008JD011470>. Repéré à <https://doi.org/10.1029/2008JD011470>
- Zeiringer, B., Seliger, C., Greimel, F., & Schmutz, S. (2018). River Hydrology, Flow Alteration, and Environmental Flow. Dans (pp. 67-89). Springer International Publishing. doi: 10.1007/978-3-319-73250-3_4. Repéré à https://dx.doi.org/10.1007/978-3-319-73250-3_4

Chen, J., Brissette, F. P., Poulin, A., & Leconte, R. (2011). Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resources Research*, 47(12), n/a-n/a. doi: 10.1029/2011wr010602. Repéré à <https://dx.doi.org/10.1029/2011wr010602>

Smith, M. B., & Gupta, H. V. (2012). The Distributed Model Intercomparison Project (DMIP) – Phase 2 experiments in the Oklahoma Region, USA. *Journal of Hydrology*, 418-419, 1-2. doi: <https://doi.org/10.1016/j.jhydrol.2011.09.036>. Repéré à <https://www.sciencedirect.com/science/article/pii/S002216941100686X>