

"Models are lies that help us see the truth."

Chaitin, after Picasso

2. Données et chaînes de modélisation hydrologique

Dans ce chapitre, nous présentons les bases de données utilisées : les prévisions météorologiques, les jeux de bassins versants qui ont servi de base à cette étude et leurs séries historiques de données observées. Ces données servent comme données d'entrée pour la modélisation hydrologique, effectuée par deux modèles pluie-débit globaux : le modèle GRP et le modèle MORDOR. Les fonctionnalités de ces modèles de prévision hydrologique sont également exposées ici.

2.1 Introduction

Avant de commencer toute opération liée aux prévisions, il est essentiel de bien connaître les bases de données et les outils de modélisation disponibles. Les travaux de cette thèse utilisent :

- les prévisions météorologiques d'ensemble du Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT ou ECMWF en anglais) ;
- les prévisions météorologiques d'ensemble de la base TIGGE, mise en place dans le cadre du programme THORPEX de l'Organisation Mondiale de la Météorologie ;
- les prévisions par analogie, issues de la méthode de recherche de situations analogues utilisée dans les centres de prévision opérationnelle d'EDF.

L'application de ces prévisions est effectuée sur deux jeux de bassins versants :

- un échantillon de 74 grands bassins versants en France,
- un échantillon de 11 bassins versants français de fort intérêt pour la production hydroélectrique.

Sur l'ensemble des bassins versants étudiés, nous disposons de données historiques de précipitation et température moyennes de bassin issues de la ré-analyse SAFRAN de Météo-France et des observations de débits journaliers issues de la banque HYDRO du Ministère en charge de l'environnement.

Deux modèles de prévision hydrologique sont ici utilisés au pas de temps journalier :

- le modèle *GRP*, développé par Irstea depuis une dizaine d'années et actuellement utilisé opérationnellement en France par 14 Services de Prévision de Crues (SPC) (Figure 4, à gauche),
- le modèle *MORDOR* développé à EDF et également utilisé en prévision opérationnelle dans ses centres de prévision (Figure 4, à droite).

Dans les paragraphes qui suivent, nous décrivons plus en détail les prévisions météorologiques, les bassins versants et les modèles hydrologiques qui composent la base des travaux de cette thèse.

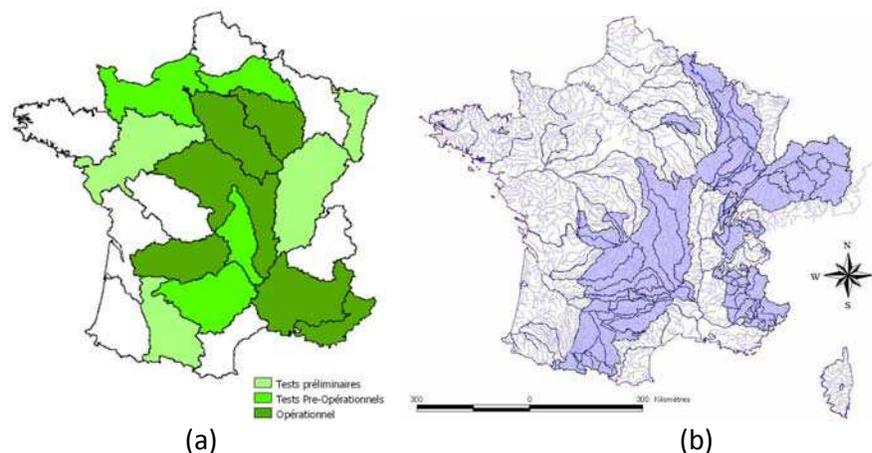


Figure 4 : Carte des régions où : (a) le modèle GRP est utilisé en opérationnel/ pré-opérationnel ou des tests préliminaires sont effectués par les services de prévision de crue SPC (source : SCHAPI) et (b) le modèle MORDOR est utilisé en opérationnel (source : EDF-DTG).

2.2 Prévisions météorologiques

2.2.1 Prévisions d'ensemble du CEPMMT

Le CEPMMT est une organisation intergouvernementale soutenue par 34 états membres (18 états européens) ayant pour objectif principal l'émission de prévisions météorologiques à moyen terme sur l'Europe et l'avancement de la recherche scientifique dans le domaine. Les prévisions sont faites par des modèles numériques à partir de données tirées des observations atmosphériques ou au sol. Les principaux types de prévisions émises sont :

- les prévisions de temps à moyen terme jusqu'à 15 jours en avance ;
- les prévisions mensuelles, émises toutes les semaines, allant jusqu'à 30 jours d'échéance ;
- les prévisions saisonnières avec un horizon de 12 mois.

Ces produits alimentent ensuite les services météorologiques nationaux, où ils servent principalement comme un complément aux prévisions à plus court terme.

Le modèle global du CEPMMT émet une prévision déterministe et des prévisions probabilistes d'ensemble. Les prévisions d'ensemble émises par le système de prévision d'ensemble (EPS, *Ensemble Prediction System*) sont constituées d'une prévision dite "*contrôle*", obtenue à partir de l'état initial issu de l'assimilation des données (condition initiale non-perturbée), et de membres qui correspondent à des trajectoires obtenues à partir des perturbations de cet état initial.

Les perturbations sont construites à partir de la méthode des vecteurs singuliers et choisies afin d'optimiser la croissance d'énergie totale sur un intervalle de 48h (Buizza et al., 2005). Plus récemment, des méthodes d'assimilation d'ensemble ont également été introduites (Buizza, 2010). Ces perturbations sont ensuite introduites dans le modèle numérique de temps, où certains processus physiques sont paramétrés. Les incertitudes liées au modèle numérique sont prises en compte en ajoutant des perturbations stochastiques dans les schémas de paramétrisation physique du modèle. Les trajectoires de l'évolution de l'atmosphère obtenues forment ainsi la prévision d'ensemble EPS.

La prévision d'ensemble du CEPMMT est émise deux fois par jour (à 00h et à 12h UTC) et fait partie de son système opérationnel de prévision depuis 1992. Au fil des années plusieurs modifications ont été apportées au système afin notamment d'améliorer les résolutions horizontales et verticales des prévisions et mieux capturer les phénomènes météorologiques à prévoir. Au début, chaque ensemble était construit avec 32 membres sur un horizon maximal de 10 jours et une résolution spatiale qui s'est améliorée au fil des années (Palmer *et al.*, 1993 ; Molteni *et al.*, 1996). En 2005, les prévisions d'ensemble étaient constituées de 51 membres et émises sur une grille spatiale d'environ 80 [km] x 80 [km]. A partir de septembre 2006, les 51 prévisions d'ensemble sont émises avec une résolution spatiale variable : 32 [km] x 32 [km] jusqu'à 10 jours et 64 [km] x 64 [km] au-delà (Buizza *et al.*, 2007a ; Leutbecher et Palmer, 2007). Par ailleurs, depuis mars 2008 les prévisions d'ensemble à moyen terme sont fusionnées avec les prévisions mensuelles : une fois par semaine, les jeudis, les prévisions d'ensemble émises sont ainsi prolongées jusqu'à 32 jours.

Dans notre étude, nous avons utilisé une archive de prévisions météorologiques du CEPMMT qui s'étend sur la période 2005-2008. Les 50 membres perturbés des prévisions journalières des précipitations totales et de températures émises à 12h ont été utilisés. Les données nous ont été fournies à la résolution spatiale de 0.5° lon x 0.5° lat, jusqu'à un horizon de 10-15 jours selon la période.

La Figure 5 présente un exemple des données CEPMMT utilisées dans cette étude. Les prévisions émises sur la France par les 50 membres de la prévision d'ensemble et la prévision contrôle (la dernière figure à droite) pour le 30 mai 2008, à 7 jours d'avance (c'est-à-dire, prévisions émises le 23 mai 2008), sont représentées. Ce jour là, un événement pluvieux sévère a touché le sud-est de la France. Les bassins versants de la Durance à Serre Ponçon et la Drôme à Saillans ont enregistré respectivement 35,7 [mm] et 21,1 [mm] de cumul de pluie journalier en moyenne sur la surface du bassin. Dans la figure, nous observons que la plupart des membres ont bien localisé la région affectée avec 7 jours d'avance. De plus, un tiers des membres ont bien prévu l'intensité de l'événement.

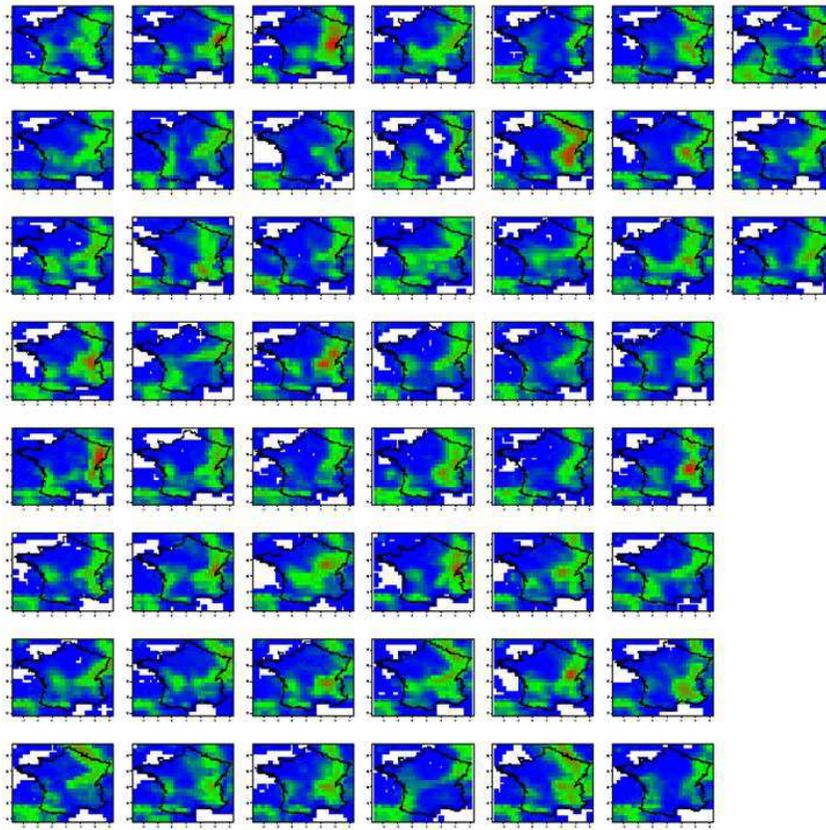


Figure 5 : Cartes de France illustrant les précipitations prévues par les 51 membres de la prévision d'ensemble du centre météorologique européen (CEPMET) pour le 30 mai 2008 et un horizon de prévision de 7 jours (prévisions émises le 23 mai 2008).

2.2.2 Prévisions d'ensemble de la base TIGGE

La base de prévisions TIGGE fait partie du programme THORPEX de l'Organisation Météorologique Mondiale ("*The Observing System Research and Predictability Experiment*"). Ce programme vise à mieux prévoir le temps et amenuiser ses effets négatifs sur la population. La base TIGGE est constituée de prévisions d'ensemble émises à partir d'octobre 2006, provenant de différents centres météorologiques (Richardson, 2005 ; Park *et al.*, 2008 ; Bougneault *et al.*, 2010). En effet, plusieurs études (ex., Pappenberger *et al.*, 2008, Roulin, 2007, Goswami *et al.*, 2007) ont montré que les prévisions d'ensemble provenant d'un centre météorologique ne prennent en compte que partiellement les incertitudes des conditions initiales. L'idée de base derrière l'initiative TIGGE est que l'utilisation combinée des prévisions fournies par les différents centres, avec leurs différentes méthodes de perturbation et paramétrisation stochastique, permettrait une meilleure description de l'incertitude prédictive (Park *et al.*, 2008).

L'archive mise en place au début des travaux de cette thèse est composée des prévisions de précipitation de 8 centres météorologiques internationaux, couvrant la période d'octobre 2006 jusqu'à septembre 2008 (Figure 6). Les centres météorologiques qui composent notre base de données sont : le centre européen (ci-après *ECMF*), britannique (*EGRR*), japonais (*RJTD*), américain (*KWBC*), chinois (*BABJ*), australien (*AMMC*), canadien (*CWAO*), brésilien (*SBSJ*). Nous notons que, même si nous disposons de données sur environ 2 ans, la longueur de la période pendant laquelle les prévisions d'ensemble sont disponibles varie pour chaque centre météorologique (indiquée sur la Figure 6). Nous observons, par exemple, que le centre brésilien (*SBSJ*), le dernier à rejoindre la base TIGGE, n'y contribue qu'avec 242 jours.

Le Tableau 1 résume les principales caractéristiques des prévisions TIGGE utilisées dans cette étude. On observe que les prévisions d'ensemble de tous les centres météorologiques combinées composent un grand ensemble d'un total maximum de 232 membres d'ensemble par bassin versant et par jour de prévision. L'horizon maximal des prévisions est de 15 jours.

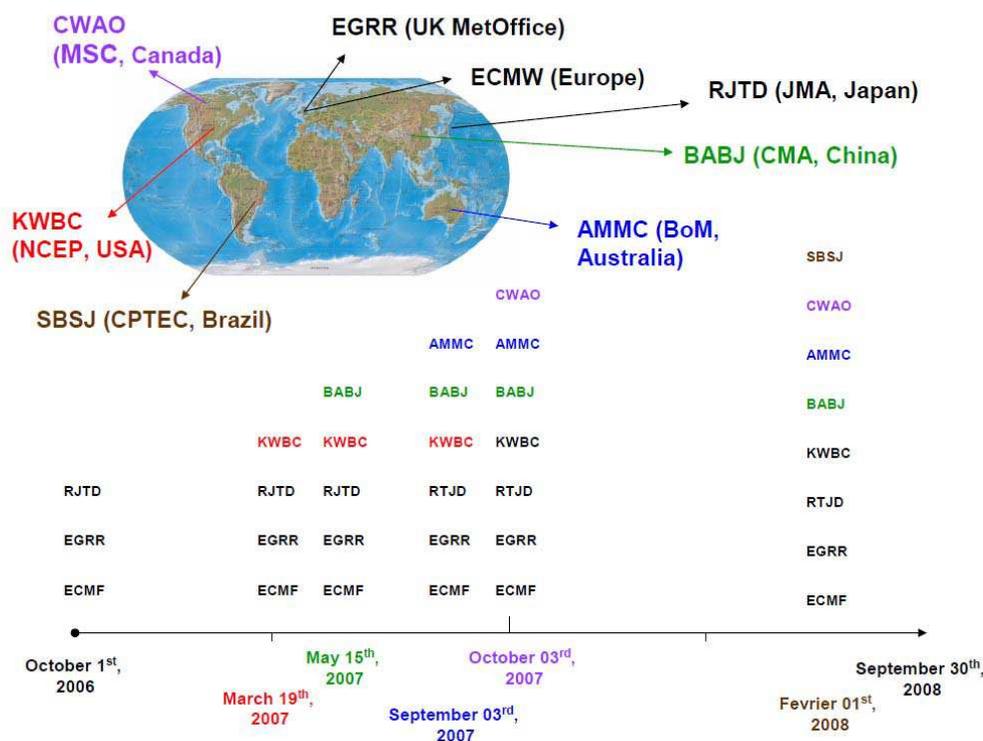


Figure 6 : Centres météorologiques contribuant à la base de prévisions de précipitation TIGGE utilisée dans cette étude, avec indication de la date de début de participation de chaque centre.

Tableau 1 : Caractéristiques principales des centres météorologiques qui forment la base de prévisions de précipitation TIGGE utilisée dans cette étude (d'après : <http://tigge.CEPMMT.int/models.html>).

	Pays	Nombre de membres	Horizon maximal (jours)	Résolution Spatiale (degrés °)	Méthode de perturbation utilisée	Nombre de jours	Référence
ECMF	Organisation avec 34 Etats Membres (ECMWF)	50	10-15	0.5° x 0.5°	Vecteurs Singulieus et Filtre d'ensemble Kalman	731	www.CEPMMT.int/
EGRR	Royaume Uni	23	15	0.8333° x 0.5555°	Filtre d'Ensemble Kalman Transformé	731	www.metoffice.gov.uk/
RJTD	Japon	50	9	0.5625° x 0.5625°	Méthode des erreurs croissantes BGM	731	www.jma.go.jp/jma/indexe.html
KWBC	Etats-Unis	20	15	1.0° x 1.0°	Méthode des erreurs croissantes BGM	561	www.ncep.noaa.gov/
BABJ	Chine	14	10	1.0° x 1.0°	Méthode des erreurs croissantes BGM	504	www.cma.gov.cn/cma_new/
AMMC	Australie	32	10	1.5 x 1.5°	Vecteurs Singulieus	393	www.bom.gov.au/
CWAO	Canada	20	15	~ 1.0° x 1.0°	Filtres d'ensemble Kalman	363	www.weatheroffice.gc.ca/canada_e
SBSJ	Brésil	14	15	0.9375 °x0.9375 °	Fonction orthogonale empirique	242	www.cptec.inpe.br/

2.2.3 Prévisions probabilistes par analogie

Au sein de la Division Technique Générale d'EDF, la notion de prévision probabiliste est présente depuis les années 1950, sous le terme " prévisions probabilistes expertisées " (Garçon *et al.*, 2009). Ce terme signalait la prise en compte de l'expertise humaine pendant le processus de prévision. La première approche probabiliste était basée sur des simples analyses statistiques de pluie et de débit prévus en ajoutant des intervalles de confiance de régression autour de l'estimation initiale. A peu près en même temps (~1953) commence l'effort d'un archivage journalier des champs de pression observés sur chaque bassin versant d'intérêt, conjointement à l'archivage déjà existant des précipitations et températures observées. Dans les années 1970, cette archive comptait déjà vingt ans de données et commençait à devenir opérationnelle. Duband (1970, 1980) a alors formalisé sous forme d'algorithmes la démarche de prévision par analogie déjà utilisée par les prévisionnistes.

La prévision par analogie consiste à mieux comprendre les situations futures à l'aide des situations synoptiques observées dans le passé et mises en relation avec les conditions locales, afin d'obtenir des prévisions de précipitation à partir des prévisions des champs synoptiques. L'hypothèse derrière cette approche est basée sur les travaux de Lorenz (1969), selon lesquels deux situations synoptiques suffisamment similaires doivent conduire à des effets locaux comparables. Nous parlons de prévision probabiliste par analogie, car il ne s'agit pas seulement de trouver la situation la plus ressemblante, mais de construire un sous-ensemble de situations similaires permettant d'obtenir une distribution des scénarios futurs possibles (Zorita et von Storch, 1999 ; Obled *et al.* 2002). Le prévisionniste cherche alors à lier la situation météorologique prévue aux situations analogues observées dans le passé pour expertiser sa prévision.

L'archivage de plusieurs variables comme le champ de pression, l'humidité, etc. est proposé dans la littérature concernant les méthodes par analogie (Zorita et von Storch, 1999 ; Obled *et al.*, 2002 ; Bontron, 2004 ; Ben Daoud, 2009). Dans cette étude, seulement les champs de pression observés à 0h00 et 24h00, à 700hPa et 1000hPa sont utilisés. Les champs prévus proviennent des analyses du modèle *ARPEGE* de Météo-France (Geleyn *et al.*, 1994) pour les champs à 00 hTU, ensuite des prévisions *ARPEGE* (champs à 24, 48 et 72 hTU) et enfin des prévisions du modèle du Centre Européen pour les Prévisions Météorologiques à Moyen Terme CEPMMT (Simmons *et al.*, 1989) (pour les échéances de 72 à 168 hTU). Il faut noter que les approches par analogie sont également une forme de descente d'échelle, car elles ajoutent des informations locales, à résolution fine, à partir des sorties des modèles numériques de temps. Ces derniers tournent sur l'ensemble du globe ou sur des fenêtres régionales larges, avec un maillage qui ne leur permet pas de bien capturer la topographie ou les phénomènes locaux.

Le critère utilisé pour trouver le sous-ensemble de situations analogues au champ de géopotentiel est proposé par Teweless et Wobus (1954). Ce critère est calculé entre chacune des situations de l'archive météorologique et la situation de la journée cible sur la base des gradients horizontaux d'altitude pour plusieurs niveaux de pression atmosphérique. Les 50 situations les plus ressemblantes à la journée cible sont identifiées dans l'archive météorologique. Une fois ces 50 situations définies les précipitations associées aux bassins versants de l'étude sont retrouvées. Par conséquent, pour chaque champ de géopotentiel prévu à une échéance donnée, 50 scénarios (valeurs) de précipitation observée par bassin versant, qui forment la prévision probabiliste par analogie pour la journée cible et le bassin versant en question, sont obtenus. La Figure 7 présente schématiquement l'approche de l'analogie pour l'association des valeurs observées du géopotentiel à des situations similaires au passé et leur transformation en précipitation.

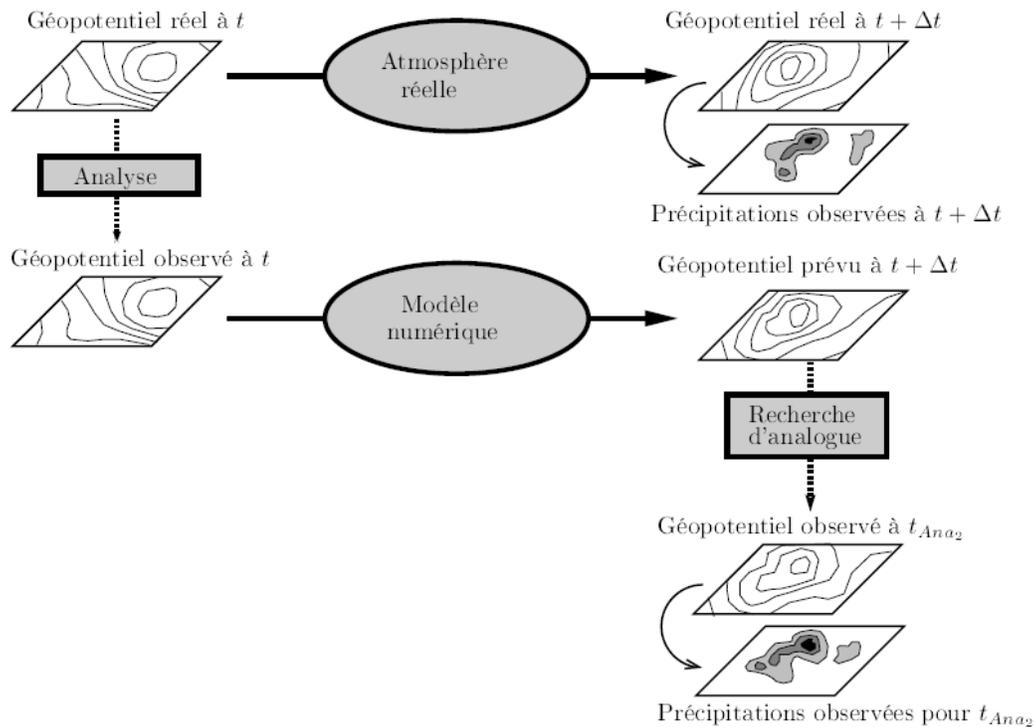


Figure 7 : Description schématique de l'approche de prévisions par analogie (source : Bontron, 2004).

Actuellement l'archive des analogues est bien plus robuste et compte presque 60 ans de données. Dans cette thèse, nous avons utilisé cette archive pour former notre base de prévisions par analogie. Pour chaque échéance et un horizon maximal de 7 jours, les prévisions d'ensemble du CEPMMT ont été utilisées dans la recherche des situations analogues. Cette information sera utilisée pour la correction statistique des biais des prévisions EPS du CEPMMT. La méthodologie adoptée est décrite en détail dans le chapitre 3.

2.3 Bassins versants étudiés

Cette thèse s'appuie sur la base de plus de 3000 bassins versants mise en place dans l'équipe hydrologique d'Irstea à Antony dans le cadre des travaux de thèse passés (Mathevet, 2005 ; Le Moine, 2008), qui était disponible au début de nos travaux. Aujourd'hui cette base, qui regroupe des données météorologiques et hydrologiques, présente déjà près de 3500 bassins versants français. Il s'agit de bassins versants pour lesquels nous disposons de données journalières de pluie et de débit, et qui sont à signification hydrologique et sans influence majeure d'ouvrages hydrauliques selon la Banque de données hydrologiques HYDRO.

Sur chaque bassin versant, les précipitations journalières de bassin sont calculées à partir des données de la ré-analyse SAFRAN de Météo-France (Quintana-Ségui *et al.*, 2008 ; Vidal *et al.*, 2009), disponibles sur la période 1958-2009. Les débits observés ont été extraits de la banque nationale HYDRO. Ils y sont également au pas de temps journalier, avec une longueur des séries historiques variable selon le bassin versant. Enfin, l'information sur l'évapotranspiration potentielle (ETP) est également fournie dans cette base de données. Cette donnée est calculée à partir de la formule d'Oudin *et al.* (2005), qui utilise la température de l'air, le rayonnement extraterrestre, la chaleur latente de vaporisation de l'eau et sa masse volumique dans le calcul de l'évapotranspiration.

Dans notre étude, nous avons cherché à construire deux échantillons de bassins versants, chacun destiné à des buts spécifiques de notre recherche :

- un échantillon de 74 grands bassins versants (surface supérieure à 1000 km²), et
- un échantillon de 11 bassins versants à fort intérêt pour la production hydroélectrique (bassins plutôt de montagne).

Les deux jeux de bassins versants utilisés dans cette étude sont présentés ci-après.

2.3.1 74 grands bassins versants en France

Ce jeu de bassins versants a été mis en place pour répondre à nos questions concernant l'utilisation de l'approche multi-modèle en prévision hydrologique d'ensemble, à partir des prévisions de précipitation de la base TIGGE (Chapitre 5). La sélection de cet échantillon a été alors faite sur la base de trois critères :

- le premier critère a été la surface : les bassins versants devaient avoir une surface supérieure à 1000 km², choisie en fonction de la résolution spatiale des prévisions TIGGE (archive fournie avec une résolution spatiale d'environ 45 [km] x 45 [km] sur la France), une limite de surface plus petite aurait amené à des bassins le plus souvent couverts par une seule maille de la grille des prévisions. D'autre part, une limite de surface plus grande aurait amené à un nombre trop petit de bassins versants, étant donné les échelles d'espace sur le territoire français,

- le deuxième critère a été la disponibilité de données observées : étant donné la courte période dont nous disposons des prévisions TIGGE (2006 à 2008), afin de préserver la robustesse statistique de nos résultats, seuls les bassins avec au moins 95% de données observées disponibles sur cette période ont été retenus. De plus, nous avons également veillé à ce qu'une période antérieure d'au moins 10 années soit également disponible pour le calage du modèle hydrologique utilisé,
- le troisième critère a été la localisation des bassins versants : nous avons cherché à avoir une certaine diversité des bassins versants en termes de caractéristiques physiographiques, climatiques et hydrologiques, et avons évité, dans la mesure du possible, d'avoir beaucoup de bassins versants sur une même rivière ou un tributaire proche.

La localisation des 74 bassins versants sélectionnés est présentée à la Figure 8. La surface moyenne de l'échantillon de bassins versants est d'environ 3 400 km², avec une surface minimale égale à 1 070 km² et maximale égale à 9 387 km² (à l'exception du bassin versant de la Seine à Paris qui ressort de l'échantillon, avec 43 800 km²). Dans l'échantillon, la longueur des séries temporelles des débits varie entre 16 et 51 années, avec une majorité de bassins versants (65 %) disposant d'une période de 40 à 44 années de données complètes.

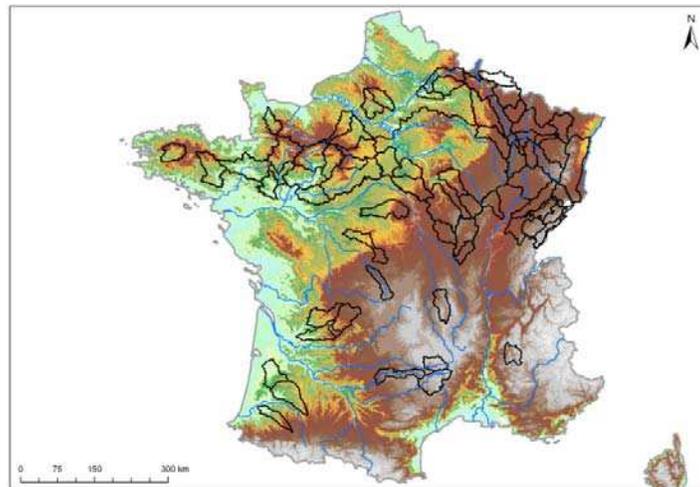


Figure 8 : Localisation du jeu de 74 larges bassins versants français.

Le Tableau 2 présente les principales caractéristiques du jeu de 74 bassins versants étudiés. Ces caractéristiques sont représentées par leur valeur moyenne avec, entre crochets, les valeurs minimales et maximales.

Tableau 2 : Principales caractéristiques des 74 bassins versants étudiés en termes de valeurs moyennes (entre les crochets sont affichées les valeurs minimales et maximales).

Altitude moyenne (m)	Moyenne interannuelle des précipitations (mm/an)	Fraction solide des précipitations (%)	Moyenne interannuelle d'ETP - Oudin (mm/an)	Moyenne interannuelle des débits (mm/an)
313	962	5.10	677	389
[74, 1046]	[638, 1565]	[0.6, 20.4]	[590, 793]	[99, 1304]

Nous remarquons que l'échantillon des bassins sélectionnés inclut surtout des bassins de plaine, peu influencés par la neige. La plupart des bassins présentent également une alimentation pluviale dominante avec une hauteur des précipitations par année assez élevée. Par conséquent, l'influence de la température sur la formation des débits des fontes nivales, ainsi que l'influence de la topographie sont modérées pour la majorité des bassins versants étudiés dans cet échantillon.

Ces conclusions sont confirmées par la classification proposée par Sauquet (2006) concernant les régimes hydrologiques des bassins versants français. Selon cette classification, la grande majorité des bassins versants sélectionnés dans notre échantillon font partie des groupes caractérisés par un régime hydrologique pluvial, fortement influencé par la pluie d'hiver et d'automne, et avec des écoulements faibles pendant les mois d'été. Seul un petit nombre de nos bassins (3 parmi les 74) est dominé par un régime hydrologique plutôt identifié à un régime méditerranéen : avec des événements de précipitations intenses surtout en automne et des étés particulièrement secs, comme il en est pour le Luy à Saint-Pandelon et l'Aveyron à Laguëpie. Enfin, nous retrouvons également des régimes hydrologiques de transition, où les variations saisonnières de l'écoulement sont affectées aussi bien par les précipitations que par la température et l'influence topographique, dans un nombre limité de nos bassins versants (6 parmi les 74), comme il en est pour le Lot à Basannac et la Drôme à Saillans, où les influences nivale et pluviale coexistent.

2.3.2 11 bassins versants de fort intérêt pour la production hydro-électrique

Ce deuxième échantillon de bassins versants a été construit dans le but d'étudier la qualité et la valeur des prévisions hydrologiques d'ensemble dans le cas des bassins versants de fort intérêt économique pour la production hydroélectrique en France (Chapitres 6 et 7). Le choix a été basé sur les critères de sélection suivants :

- des bassins situés en amont des retenues d'eau, dont les apports alimentent les lacs-réservoirs des barrages gérés par EDF, ou font figure de témoins importants pour la génération d'électricité,
- des bassins versants de superficies et de régimes hydrologiques variés,
- des bassins versants distribués dans les principaux massifs français : les Alpes, le Massif Central, les Cévennes et le Jura.

La localisation des bassins versants sélectionnés pour cet échantillon est présentée dans la Figure 9.

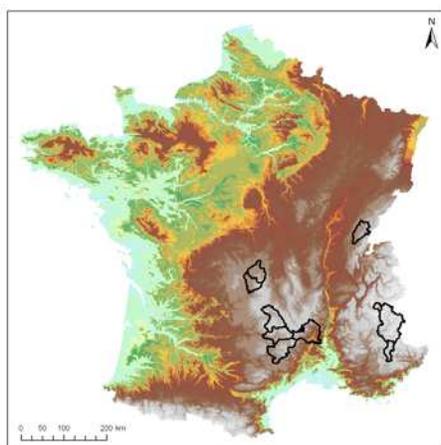


Figure 9 : Localisation du jeu de 11 bassins versants de fort intérêt en hydroélectricité de cette étude.

La plupart des bassins versants examinés ici sont des bassins de montagne avec pour la plupart des fortes influences nivales, surtout pour les bassins alpins. En suivant la classification des régimes hydrologiques de Sauquet (2006), nous notons que ces bassins versants appartiennent aux groupes caractérisés par des écoulements forts liés à la fonte de neige, avec un maximum de débit au printemps et en été. Les bassins versants moins influencés par la neige sont ceux situés dans le Massif Central et le Jura, citons les exemples du Tarn à Millau, de l'Ardèche à Sauze, de la Sioule aux Fades et de la Truyère à Grandval. Ces bassins versants sont caractérisés par des écoulements forts dès la fin d'automne, jusqu' au printemps, ainsi que par des étés particulièrement secs.

Le Tableau 3 présente les principales caractéristiques du jeu de 11 bassins versants étudiés. Les caractéristiques sont représentées par leur valeur moyenne avec, entre crochets, leurs valeurs minimales et maximales. Le Tableau 4 résume également les principales caractéristiques des retenues liées aux 11 bassins versants étudiés.

Tableau 3 : Principales caractéristiques des 11 bassins versants étudiés en termes de valeurs moyennes (entre crochets sont affichées les valeurs minimales et maximales).

Altitude moyenne (m)	Moyenne interannuelle des précipitations (mm/an)	Fraction solide des précipitations (%)	Moyenne interannuelle d'ETP - Oudin (mm/an)	Moyenne interannuelle des débits (mm/an)
1170	1257	19.9	673	775
[490, 2362]	[955, 1838]	[5, 61]	[570, 826]	[430, 1352]

Tableau 4 : Principales caractéristiques des retenues liées aux 11 bassins versants de fort intérêt en hydroélectricité de cette étude.

Bassin Versant	Surface (km²)	Localisation dans les massifs montagneux français	Volume utile de la retenue (hm³)	Vocation	Puissance hydroélectrique installée (MW)
Durance à Serre Ponçon	3580	Alpes	1030	Hydroélectricité et Irrigation	380
Borne au Pont Nicoulaud	63	Cévennes	35	Hydroélectricité	100
Verdon à Castillon	656	Alpes	113	Hydroélectricité et Irrigation	50
Ain à Vouglans	1195	Jura	419	Hydroélectricité	262
Truyère à Grandval	1795	Massif Central	225	Hydroélectricité	68
Dordogne à Bort	1005	Massif Central	408	Hydroélectricité et Tourisme	240
Sioule aux Fades	1300	Massif Central	56	Hydroélectricité	29
Tarn à Millau	2143	Massif Central	4	Hydroélectricité	40
Lot à Castelnau	1158	Massif Central	20	Hydroélectricité	41
Ardèche à Sauze	2261	Massif Central	-	-	-
Romanche aux Chambons	223	Alpes	50,8	Hydroélectricité	116

2.4 Modélisation hydrologique

Une fois définies les prévisions météorologiques qui seront utilisées dans le système de prévision hydrologique d'ensemble, les données disponibles (précipitations, débits et températures observés) et les bassins versants qui seront étudiés, il reste à choisir le modèle hydrologique qui permettra de transformer les précipitations en débit et effectuer le calage de ses paramètres libres.

Dans cette étude, deux modèles hydrologiques globaux sont utilisés : le modèle *GRP* dans sa version sans module de prise en compte de stockage et fonte de la neige (Tangara, 2005) et avec module de prise en compte de la neige (Valery, 2010 ; Nicolle *et al.*, 2011), et le modèle *MORDOR* (Garçon, 1999 et Paquet, 2004). D'une part, le modèle *GRP* sans module neige est utilisé pour la modélisation du jeu de 74 grands bassins versants français peu influencés par la neige. D'autre part, le modèle *GRP* avec module neige et le modèle *MORDOR* sont utilisés pour la modélisation du jeu de 11 bassins versants situés dans les principaux massifs montagneux français.

Dans un premier temps, nous présentons les principales caractéristiques de chaque modèle. Ensuite, nous présentons la procédure et les résultats du calage effectué sur les deux jeux de bassins versants de notre étude.

2.4.1 Modèle hydrologique *GRP*

Le modèle de prévision *GRP*, développé par l'équipe Hydrologie d'Irstea à Antony, est un modèle global à 3 paramètres, dérivé du modèle *GR4J* (Perrin, 2000) et adapté pour la prévision des crues (Tangara, 2005). Dans cette étude, le modèle *GRP* est utilisé au pas de temps journalier, dans sa version adaptée aux prévisions d'ensemble (Ramos *et al.*, 2008). Pendant son développement, le modèle a été testé sur plus d'un millier de bassins versants français, ce qui lui permet de s'adapter à une grande variété de conditions hydroclimatiques, géologiques et d'occupation du sol.

GRP modélise le bassin versant comme un assemblage de réservoirs représentant les stocks d'eau dans les sols et les nappes et leur transfert vers l'exutoire du bassin. La structure du modèle repose sur un réservoir de production qui permet d'assurer un suivi des conditions d'humidité du bassin, une fonction d'ajustement de la pluie efficace, un hydrogramme unitaire et un réservoir de routage quadratique (Figure 10).

Le premier paramètre du modèle représente la capacité à un jour du réservoir de routage ($X1$, en mm), le deuxième représente le facteur d'ajustement de la pluie ($X2$, paramètre adimensionnel supérieur ou inférieur à 1) et le troisième correspond au temps de base de l'hydrogramme unitaire ($X3$, en jours). La mise à jour du modèle est faite au niveau du réservoir de routage : son niveau est recalculé de manière à pouvoir générer un débit exactement égal au dernier débit observé à l'instant de la prévision (Tangara, 2005). Berthet (2010) a comparé la performance de l'approche de mise à jour à pas de temps horaire utilisée avec d'autres méthodes existantes (filtre de Kalman, réseaux neurones, etc) dans la littérature. Son étude a montré que :

- il y a un intérêt à appliquer la mise à jour au niveau du réservoir de routage, car le modèle est peu sensible aux mises à jour des autres états ;
- la mise à jour ajustant le niveau du réservoir de routage avec le dernier débit observé présente une meilleure performance par rapport aux autres approches examinées, notamment pour les plus longues échéances.

Les données d'entrée du modèle *GRP* sont les précipitations et l'évapotranspiration potentielle (ETP), tandis que la mise à jour a besoin du dernier débit observé pour être effective.

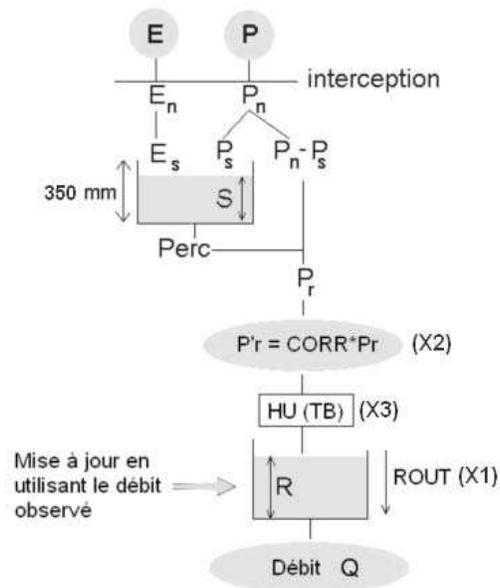


Figure 10 : Schématisation de la structure du modèle GRP sans module neige (Tangara, 2005)

Récemment un module de stockage et fonte de la neige a été développé pour la famille de modèles *GR*. Ce module neige, appelé *Cemaneige* a été développé par Valéry (2010), sur la base du modèle *GR4J* et intégré au modèle *GRP* par Nicolle (2010, 2012). L'intégration d'un module neige apporte deux paramètres supplémentaires au modèle : le facteur de fonte (K_f) et le coefficient de pondération de l'état thermique du manteau (CT_G).

La Figure 11 présente une schématique de l'intégration du module neige *Cemaneige* dans la structure du modèle *GRP*. Le module est inséré en amont du modèle hydrologique. La séparation de la pluie en liquide ou solide est basée sur les températures minimales et maximales, ou, à défaut, sur la température moyenne journalière. Le modèle a ainsi besoin, en plus des précipitations et de l'évapotranspiration potentielle, de la température de l'air en entrée. La neige est stockée au sein d'un réservoir conceptuel qui représente le manteau neigeux. Ce réservoir est alimenté uniquement par la fraction solide des précipitations. Le manteau neigeux présente un état thermique qui permet de retarder ou déclencher la fonte. Une fonction de transfert permet ensuite de calculer la lame de fonte écoulee, issue du stock de neige, qui va être intégrée au modèle hydrologique, en plus de la fraction liquide des précipitations (Nicolle, 2010).

Le logiciel *GRP* est constitué de deux utilitaires :

- l'utilitaire de calage, qui permet, à partir des chroniques des données historiques, de déterminer ses paramètres et d'évaluer les performances du modèle ;
- l'utilitaire de prévision, permettant de calculer les prévisions de débit.

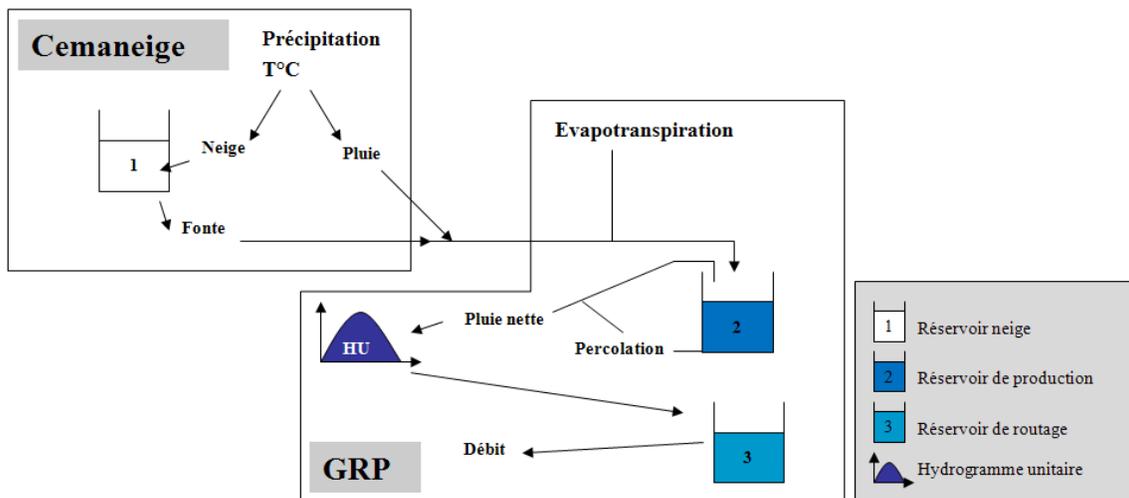


Figure 11 : Schématique de l'intégration du module neige *Cemaneige* dans la structure du modèle *GRP* (Valéry, 2010 ; Nicolle et al., 2012).

Les détails du fonctionnement des modèles *GRP* sans et avec module neige est présentée dans l'ANNEXE A.

La version du modèle MORDOR utilisée dans cette thèse présente au total 10 paramètres libres à caler : le coefficient de correction de pluie (X1), les 3 capacités de réservoirs (X8, X9, X10), les 4 paramètres pour la vidange des réservoirs L et N (respectivement : X2, X3 et X4, X5), ainsi que les 2 paramètres pour fixer la forme de l'hydrogramme unitaire (X6, X7). A noter que la correction de la précipitation en entrée du modèle permet un premier débiaisage des pluies. La capacité du réservoir profond N est considérée constante.

Le modèle MORDOR a également un module neige intégré dans sa structure. Ce module fonctionne autour du réservoir "NEIGE", qui permet la gestion du stock et de la fonte de neige. Dans ce module, une partie de la lame de fonte est directement dirigée vers le ruissellement tandis qu'une autre partie est ajoutée soit au réservoir de surface *U*, soit au réservoir profond *N*. La répartition se fait à l'aide d'un paramètre à optimiser lors de la phase de calage.

Les détails du fonctionnement du modèle MORDOR, ainsi que de son module neige sont présentés dans l'ANNEXE A.

2.4.3 Calage des modèles hydrologiques

Pour chaque bassin versant, les modèles hydrologiques ont été calés à partir des données observées disponibles sur la période du 01 août 1958 au 10 mars 2005. Nous utilisons l'ensemble de données disponibles avant le début de la période de prévision qui servira de validation des approches étudiées en prévision d'ensemble. Nous présentons ici les performances en calage. La performance des modèles en validation, c'est-à-dire, pendant la période de prévision est au cœur de cette thèse et sera présentée dans les chapitres 5 et 6, où la qualité des prévisions émises par les modèles est examinée.

L'objectif d'une procédure de calage d'un modèle hydrologique consiste à optimiser les paramètres libres du modèle en s'appuyant sur des données connues, comme le débit à l'exutoire. La fonction objectif utilisée pour l'optimisation des paramètres libres vise à minimiser la distance entre l'observation (valeur connue) et la simulation (valeur calculée). En résultat, il s'agit d'identifier le jeu de paramètres qui minimise l'écart entre l'observation et la simulation en optimisant la fonction objectif.

Les modèles hydrologiques *GRP* et *MORDOR* ont des fonctions objectifs déjà codées au préalable dans leur module de calage automatique. Nous avons utilisé ces fonctions pour caler les modèles. La performance en calage est ainsi évaluée par :

- le critère d'efficacité de Nash et Sutcliffe (1970) pour les deux modèles *GRP* et *MORDOR*,
- le critère de *C2MP*, introduit par Mathevet (2005) afin de donner au critère de persistance (Kitanidis et Bras, 1980) une borne négative à -1, pour le modèle *GRP*,
- le critère de *Kling-Gupta* (Gupta *et al.*, 2009) pour le modèle *MORDOR*.

Le critère d'efficacité de *Nash et Sutcliffe* (1970) est défini comme la part de variance des débits expliquée par le modèle. Il varie dans l'intervalle $[-\infty, 1]$ et estime sur N jours la performance du modèle par rapport à un modèle de référence qui donnerait comme débit simulé Q^{sim} le débit moyen observé \bar{Q}^{obs} :

$$NASH = 1 - \frac{\sum_{i=1}^N (Q_i^{sim} - Q_i^{obs})^2}{\sum_{i=1}^N (Q_i^{obs} - \bar{Q}^{obs})^2} \quad \text{Eq. 1}$$

Ainsi, si le critère de *Nash-Sutcliffe* est négatif, le modèle utilisé est plus mauvais que le modèle de référence ; si le critère est positif, il est meilleur.

Le critère de persistance (*PERS*) a été introduit pour permettre de comparer la performance du modèle à un modèle naïf qui donnerait comme débit simulé le dernier débit observé. Ce critère est aussi borné aux valeurs positives et varie dans l'intervalle $[-\infty, 1]$:

$$PERS = 1 - \frac{\sum_{i=1}^N (Q_i^{sim} - Q_i^{obs})^2}{\sum_{i=1}^N (Q_i^{obs} - Q_{i-1}^{obs})^2} \quad \text{Eq. 2}$$

Comme le critère de *Nash-Sutcliffe*, le critère de persistance est marqué par une absence de borne négative. Le critère de *C2MP* a été introduit par Mathevet (2005) afin de donner au critère de persistance une borne négative à -1. Le critère *C2MP* varie dans l'intervalle $[-1, 1]$, ou $[-100, 100]$ quand ceci est donné en pourcentage :

$$C2MP = \frac{PERS}{2 - PERS} \quad \text{Eq. 3}$$

Le critère de *Kling-Gupta* (Gupta *et al.*, 2009), quant à lui, est une modification du critère de *Nash-Sutcliffe* qui améliore la performance des modèles pluie-débit en simulation. Il s'agit d'une décomposition du critère de *Nash-Sutcliffe* en trois termes :

$$NASH = -\beta + \alpha^2 + 2r\alpha \quad \text{Eq. 4}$$

, ou :

$$\beta = \frac{(\bar{Q}^{obs} - \bar{Q}^{sim})^2}{(\sigma^{obs})^2} \quad \text{Eq. 5}$$

$$\alpha = \frac{\sigma^{sim}}{\sigma^{obs}} \quad \text{Eq. 6}$$

$$r = \frac{1}{n} \sum_{i=1}^N \frac{(Q_i^{obs} - \bar{Q}^{obs}) * (Q_i^{sim} - \bar{Q}^{sim})}{\sigma^{sim} \sigma^{obs}} \quad \text{Eq. 7}$$

La composante β est équivalente à l'erreur du bilan normée entre les moyennes des débits simulés et observés, la composante α représente le rapport des variances des débits simulés et observés et la composante r est le coefficient de corrélation entre débits simulés et observés.

Le critère de *Kling-Gupta* (*KGE*) est un critère de type distance euclidienne entre les valeurs prises par ces trois composantes et leurs valeurs optimales. Plus particulièrement le critère de *Kling-Gupta* est la différence entre 1 et cette distance euclidienne. En combinant ces termes, l'expression mathématique du critère de *Kling-Gupta* est :

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + \left(\frac{\sigma^{sim}}{\sigma^{obs}} - 1\right)^2} \quad \text{Eq. 8}$$

En calage, l'objectif étant d'atteindre les valeurs optimales des composantes, la distance euclidienne doit tendre vers zéro et maximiser ainsi la valeur du critère *KGE* (Eq. 8).

Nous rappelons que le modèle *GRP* a été calé sur les deux jeux de bassins versants de cette étude : les 74 grands bassins versants (modèle *GRP* sans module neige) et les 11 bassins versants à intérêt en hydroélectricité (modèle *GRP* avec module neige) et que le modèle *MORDOR* n'a été calé que sur ce dernier jeu de bassins versants.

Les Figure 13 et Figure 14 présentent les performances obtenues en calage selon le critère de *NASH* et *C2MP*, respectivement, pour le modèle *GRP*. La Figure 15 présente les performances obtenues en calage selon le critère de *NASH* et *Kling-Gupta* pour le modèle *MORDOR*.

Dans la Figure 13 nous observons que le modèle *GRP* montre de bonnes performances pour la majorité des bassins versants étudiés en termes de critère de *NASH*. Le modèle *GRP* obtient des performances plus faibles seulement sur quelques bassins versants situés en Alsace-Lorraine et dans les Cévennes. Cependant les valeurs du score de *NASH* sur ces bassins versants restent supérieures à 80%.

En ce qui concerne le score C2MP (Figure 14) les valeurs varient entre 15 et 70 pendant la période de calage. Rappelons que les valeurs les plus élevées sont souhaitées. Nous remarquons également que les bassins versants présentant les plus faibles performances ($C2MP < 30$) se situent dans le Nord du pays (bassins de la Seine aval et de l'Eure), ainsi que dans les Alpes et le Massif Central. Nous ne remarquons pas une tendance claire entre la surface du bassin versant et la valeur du critère C2MP.

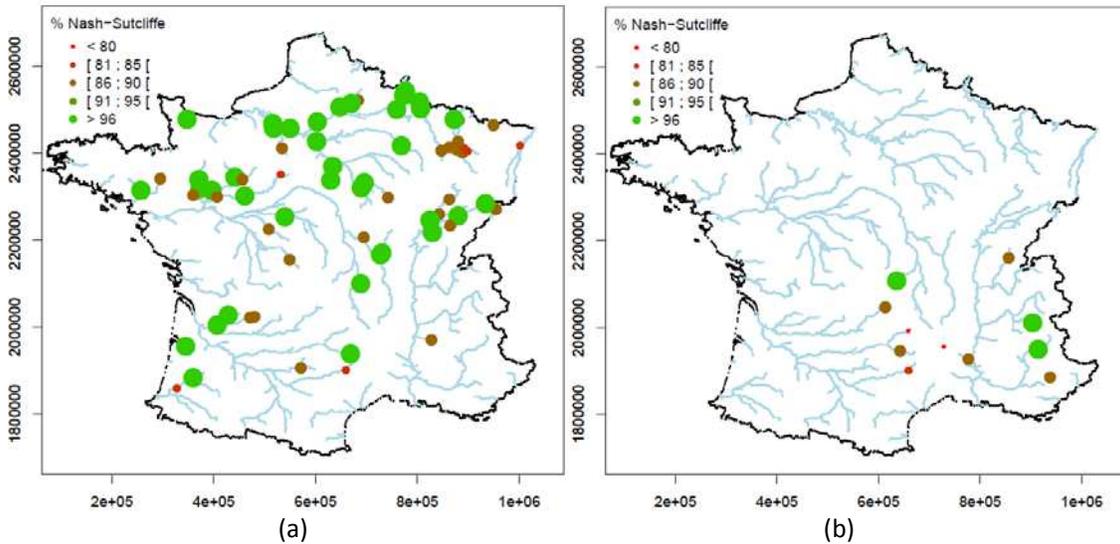


Figure 13 : Performance au calage du modèle GRP sans (à gauche) et avec (à droite) le module Cemaneige selon le critère de NASH et pour l'échantillon de : (a) 74 et (b) 11 bassins versants.

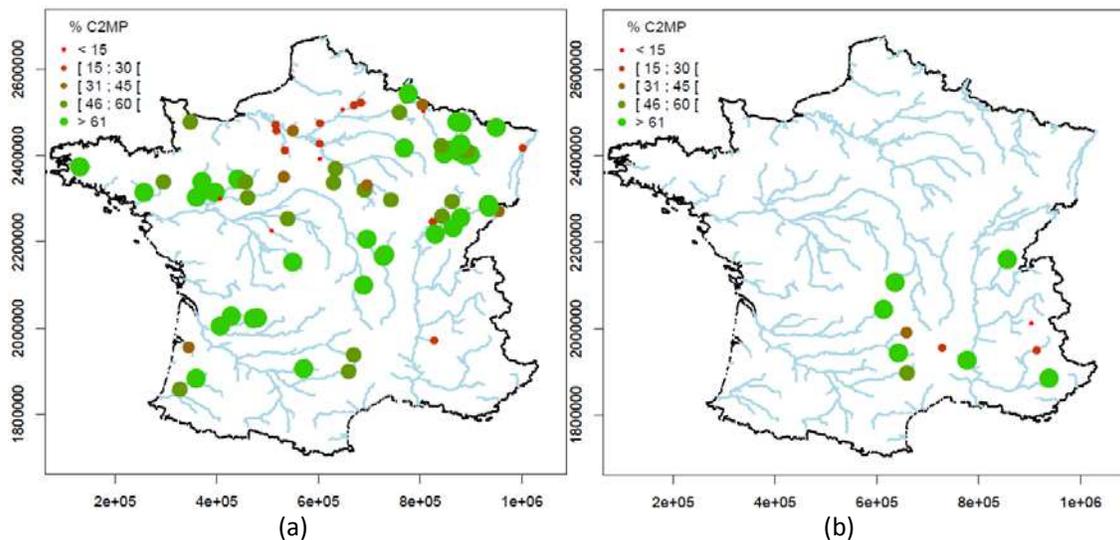


Figure 14 : Performance au calage du modèle GRP sans (à gauche) et avec (à droite) selon le critère de C2MP et pour l'échantillon de : (a) 74 et (b) 11 bassins versants.

Pour la performance en calage du modèle *MORDOR* (Figure 15), en ce qui concerne le critère de *NASH*, les valeurs obtenues varient entre 78 et 91. Deux bassins versants situés dans les Alpes et deux dans le Massif Central (Verdon à Castillon, Romanche aux Chambons, Sioule aux Fades et Truyère à Grandval) présentent les plus faibles performances (*NASH* < 85). Nous remarquons néanmoins que les valeurs du critère de *NASH* restent élevées et satisfaisantes pour tous les bassins versants étudiés. Quant au calage basé sur le critère *KGE*, nous remarquons que les bassins versants avec les plus faibles performances correspondent aux bassins versants de faible performance également avec le critère de *NASH*.

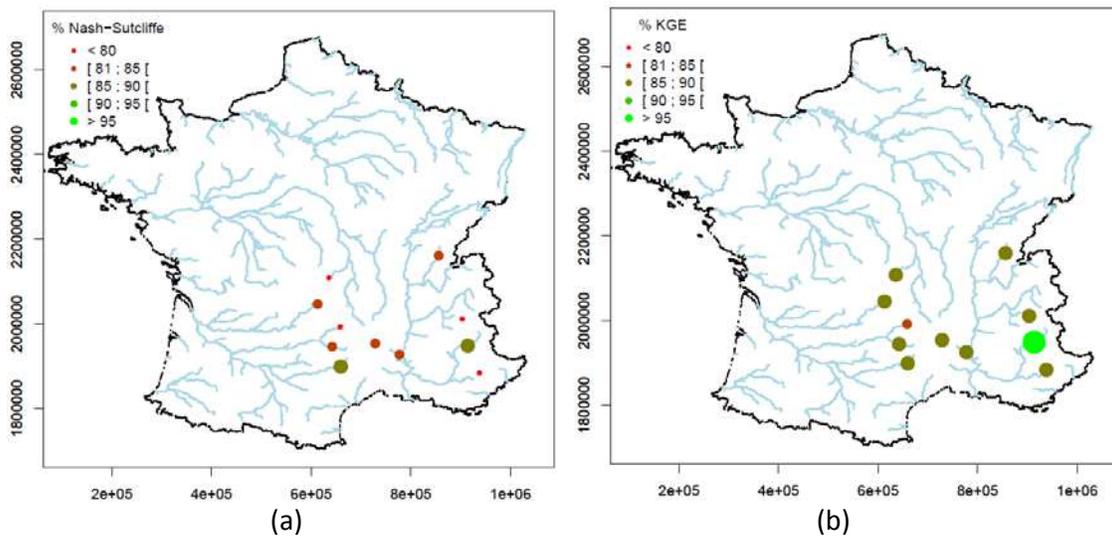


Figure 15 : Performance au calage du modèle *MORDOR* sur un échantillon de 11 bassins versants selon le critère de: (a) Nash-Sutcliffe (*NASH*), (b) Kling-Gupta (*KGE*).

2.5 Synthèse

Ce chapitre a présenté les données et les modèles qui ont été utilisés dans cette thèse. Pour répondre à nos objectifs, nous avons mis en place une large base de prévisions météorologiques (précipitation et températures) :

- les prévisions météorologiques d'ensemble du centre européen CEPMMT avec au total 51 membres sur un horizon de prévision de 10 à 15 jours et pour 4 années de prévisions disponibles (2005 à 2008),
- les prévisions d'ensemble de la base TIGGE avec 8 centres météorologiques disposant au total de 232 membres sur un horizon maximal de prévision de 15 jours et pour 2 années de prévisions (2006 à 2008),
- les prévisions par analogie avec 50 membres sur un horizon de prévision de 7 jours et pour 4 années de données (2005 à 2008).

Ces prévisions seront utilisées avec deux modèles hydrologiques : le modèle *GRP* et le modèle *MORDOR*, calés sur deux échantillons de bassins versants français : un jeu de 74 grands bassins versants et un jeu de 11 bassins versants à fort intérêt pour la production hydroélectrique. La constitution de ces échantillons de bassins versants s'est centrée sur deux objectifs principaux :

- d'un côté, avoir un échantillon suffisamment large afin de produire des conclusions d'une représentativité générale,
- de l'autre côté, avoir un échantillon suffisamment diversifié pour que les conclusions soient pertinentes et représentatives d'une large gamme de conditions climatiques et hydrologiques rencontrées en France.

L'utilisation des prévisions de plusieurs modèles météorologiques et l'application à deux modèles hydrologiques, qui ont des traits en commun (les deux sont des modèles globaux à vocation opérationnelle) mais se distinguent par des représentations différentes des processus physiques et des paramétrages différentes - avec un nombre de paramètres libres différent - d'un côté confère à cette étude une diversité de configurations à examiner lors de l'évaluation de la qualité et de la valeur des prévisions hydrologiques d'ensemble ; de l'autre côté fournit aux conclusions une généralité et une robustesse nécessaire. Ces deux aspects sont examinés dans les prochains chapitres.