

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité : Informatique

École Doctorale Informatique, Télécommunications et Électronique (Paris)

Présentée par :

Audrey Bornancin Plantier

Pour obtenir le grade de :

DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Conception de modèles de prévision des crues éclair par apprentissage artificiel

Directeurs de thèse : Madame Anne Johannet et Monsieur Gérard Dreyfus

Encadrement : Monsieur Pierre Roussel

Soutenue le 25 février 2013

Devant le jury composé de :

Monsieur Denis Dartus, *rapporteur*

Monsieur Claude Touzet, *rapporteur*

Madame Caroline Wittwer, *examinatrice*

Monsieur Pierre Ribstein, *examineur*

Monsieur Georges-Marie Saulnier, *examineur*

Madame Anne Johannet, *co-directrice de thèse*

Monsieur Gérard Dreyfus, *co-directeur de thèse*

Monsieur Pierre Roussel, *encadrant*

Remerciements

Cette thèse s'est déroulée au Centre des Matériaux des Mines d'Alès, je remercie Yannick Vimont et José-Marie Lopez-Cuesta, successivement directeurs de ce laboratoire de m'avoir accueillie. Je remercie aussi Eric Garcia-Diaz, responsable du pôle Matériaux et Génie Civil, pour son soutien.

Je tiens à remercier l'Agence Nationale de la Recherche pour le financement de cette thèse dans le cadre du projet FLASH.

Je tiens à adresser de vifs remerciements à mes co-directeurs et mon encadrant : Anne Johannet pour sa disponibilité et le suivi régulier de ce travail, Gérard Dreyfus pour son expertise sur l'apprentissage statistique et pour sa confiance, et Pierre Roussel pour ses conseils précieux et sa réactivité.

Je remercie Denis Dartus et Claude Touzet d'avoir accepté d'être les rapporteurs de cette thèse. Je remercie également Caroline Wittwer, Georges-Marie Saulnier et Pierre Ribstein pour avoir accepté de faire partie de mon jury de thèse en tant qu'examineurs.

Merci à Vazken Andréassian et à Bruce Denby d'avoir accepté d'être membres de mon jury pour la soutenance de mi-parcours et pour les conseils qu'ils m'ont donnés à cette occasion.

Je remercie le SCHAPI, Météo France, le SPC Grand Delta, l'OHMCV et Olivier Laganier pour les données hydrométéorologiques utilisées dans cette thèse.

Je souhaite exprimer ma reconnaissance à tous les participants du projet FLASH pour leurs conseils et les discussions intéressantes lors des différentes réunions. En plus de ceux déjà cités : merci à Pierre-Alain Ayrat pour ses visites guidées instructives, sa connaissance du terrain ainsi que son expertise pour le traitement des fichiers de données radar, à Bernard Vayssade pour ses remarques constructives, à Brunot Janet et Arthur Marchandise pour leurs informations sur le SCHAPI, à William Castaings pour ses explications sur TOPMODEL, à Yann Laborda du SPC Grand Delta, et aussi à tous les stagiaires, thésards et post-doctorants qui ont collaboré à ce projet.

Je tiens à exprimer mes remerciements à Marc Vinches pour m'avoir encouragée à me lancer dans ce projet ainsi que pour son soutien et ses conseils.

Je remercie tous les doctorants et post-doctorants de l'équipe Structures et Hydrosystèmes, en particulier : Guillaume Artigue pour la création des cartes présentées dans ce travail, sa connaissance du terrain et des prévisions météorologiques, Line Kong A Siou, Cécile Baudry, Hang Tran, Virgile Taver et Thomas Darras.

Je n'oublie pas les élèves de l'Ecole des Mines d'Alès qui par le biais de différents projets ont contribué à ce travail. Merci également aux stagiaires ayant intégré ponctuellement l'équipe, parmi lesquels je tiens à remercier plus spécialement Yann Visserot pour sa motivation et les résultats sur les pluies obtenues à partir des mesures des radars.

Mes remerciements d'adressent à Sylvie Cruvellier et Danièle Larroze pour leur aide précieuse concernant les démarches administratives. Merci aux membres du service informatique, à l'équipe de la bibliothèque et à l'équipe du Garage.

Je remercie Dominique Bertin pour la création du logiciel RNFpro et les améliorations régulières qu'il y apporte suite aux demandes des membres de l'équipe, merci également pour la réalisation du simulateur pour le projet FLASH.

Enfin, je tiens à remercier tout particulièrement ma mère et ma grand-mère pour leur soutien durant toutes mes études et leurs encouragements pendant la thèse.

Résumé

Le Sud de la France subit régulièrement des inondations dévastatrices et meurtrières résultant d'épisodes pluvieux très intenses et localisés. La prévision de ces événements rapides et complexes est très difficile.

Créé dans ce contexte, le projet FLASH (Flood forecasting with machine Learning, data Assimilation and Semi-pHysical modeling) regroupe plusieurs laboratoires partenaires qui ont pour objectif de fournir au SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations) des modèles de prévision des crues afin d'alimenter la carte de vigilance des crues disponible sur internet.

La zone d'étude principale est le Gardon d'Anduze. Des modèles à réseaux de neurones avec deux types d'architecture sont réalisés pour prévoir la hauteur du cours d'eau à partir des pluies et des hauteurs passées. La sélection du nombre de neurones cachés, du nombre de variables, de certains paramètres de l'algorithme d'apprentissage, ainsi que l'initialisation des paramètres des réseaux, déterminante pour l'estimation des performances des modèles, est effectuée par validation croisée. Les prévisions obtenues en test permettent une anticipation utile jusqu'à un horizon de prévision de 2h voire 3h suivant l'événement en test. La mise en œuvre d'un apprentissage adaptatif est décevante.

L'utilisation de hauteurs de pluies issues des mesures radar, plutôt que celles provenant des pluviomètres, a conduit à des premiers résultats équivalents.

Enfin, la méthodologie établie a été appliquée à la conception de modèles pour la prévision des crues sur le Gardon à Remoulins, bassin versant qui inclut celui précédemment étudié ; ces modèles donnent des résultats satisfaisants pour un horizon de prévision de 7h voire 9h.

Mots-clés : Crue éclair, apprentissage artificiel, réseaux de neurones, modélisation, prévision, Gardon, sélection, généralisation

Abstract

Machine Learning Models Design for Flash Flood Forecasting

The South of France is often subject to dramatic floods, which cause casualties and damages. Very intense, localized rainfalls generate fast, complex flash floods that are very difficult to forecast.

The FLASH project (Flood forecasting with machine Learning, data Assimilation and Semi-pHysical modeling) was created in this context. It brings together several laboratories from different scientific fields, whose purpose is to provide the French Flood Surveillance Service (SCHAPI), with a model of flood forecasting. These forecasts will feed the real-time flood vigilance map that is available on the Internet.

The main watershed under investigation here is the *Gardon d'Anduze*. Two types of neural networks are designed and trained to forecast the water level at *Anduze* from the past water levels and rainfalls. The selection of the number of hidden neurons, of the number of inputs, of some parameters of the training algorithm, and of the initialization of the networks parameters, which is crucial for estimating the generalization capability of the models, is performed by cross validation. The forecasts on the test events are satisfactory for 2 to 3 hour-ahead predictions, depending on the test event. An attempt at on-line training for model adaptation was unconvincing.

Encouraging preliminary results are obtained by using rainfall estimates from radar images instead of rain gauge measurements.

Finally, the methodology is applied to design predictive models of the water level of the *Gardon* at *Remoulins*, a watershed that includes the *Gardon d'Anduze* catchment. The level forecasts at *Remoulins* are satisfactory up to a prediction horizon of 7 to 9 hours.

Keywords: Flash Flood, machine learning, neural networks, modeling, *Gardon*, selection, generalization

Table des matières

REMERCIEMENTS	1
RÉSUMÉ	3
ABSTRACT	4
TABLE DES MATIÈRES	5
LISTE DES FIGURES	8
LISTE DES TABLEAUX	10
INTRODUCTION GÉNÉRALE	13
1. L'APPRENTISSAGE STATISTIQUE AVEC RÉSEAUX DE NEURONES	15
1.1. LE DILEMME BIAIS-VARIANCE.....	15
1.2. LES RÉSEAUX DE NEURONES	16
1.2.1. <i>Le neurone formel</i>	16
1.2.2. <i>Définition d'un réseau de neurones</i>	17
1.2.3. <i>Réseaux de neurones statiques</i>	17
1.2.4. <i>Réseaux de neurones dynamiques</i>	18
1.2.4.1. Définition d'un réseau dynamique.....	18
1.2.4.2. Différentes représentations d'un système dynamique.....	19
1.3. LE PERCEPTRON MULTICOUCHE ET SES PROPRIÉTÉS FONDAMENTALES	21
1.3.1. <i>Le perceptron multicouche</i>	21
1.3.2. <i>L'approximation universelle</i>	22
1.3.3. <i>La parcimonie</i>	23
1.4. APPRENTISSAGE STATISTIQUE.....	23
1.4.1. <i>Principe de l'apprentissage statistique</i>	23
1.4.2. <i>Optimisation de la fonction de coût</i>	24
1.4.2.1. Initialisation des paramètres	24
1.4.2.2. Pas d'apprentissage	25
1.4.3. <i>Apprentissage pour les réseaux de neurones dynamiques</i>	25
1.5. PROBLÈME DU SURAJUSTEMENT	25
1.6. MÉTHODES DE RÉGULARISATION	26
1.6.1. <i>L'arrêt précoce</i>	26
1.6.2. <i>La modération des poids</i>	27
1.7. SÉLECTION DES VARIABLES	27
1.7.1. <i>Principe</i>	27
1.7.2. <i>La méthode de la variable sonde</i>	28
1.8. SÉLECTION DE MODÈLE PAR LA VALIDATION CROISÉE	28
1.9. LA MODÉLISATION PAR RÉSEAUX DE NEURONES POUR LA PRÉVISION	29
1.9.1. <i>Simulation et prévision</i>	29
1.9.2. <i>Réseaux de neurones pour la prévision en hydrologie</i>	30
1.9.3. <i>Méthodologie pour la conception d'un modèle de prévision avec réseau de neurones</i>	31
1.10. CONCLUSION	31
2. CONTEXTE ET SITE D'ÉTUDE PRINCIPAL : LE GARDON D'ANDUZE	33
2.1. LES CRUES ÉCLAIR	33
2.2. LA ZONE D'ÉTUDE.....	35
2.3. ERREURS ET INCERTITUDES DES DONNÉES MESURÉES.....	36
2.3.1. <i>Précipitation</i>	36
2.3.2. <i>Hauteur/débit</i>	37
2.3.3. <i>Échantillonnage</i>	37
2.4. MODÉLISATION DES CRUES SUR LE BASSIN VERSANT DU GARDON D'ANDUZE	38
2.4.1. <i>Modélisation en hydrologie</i>	38
2.4.1.1. Les différents types de modèles.....	38
2.4.1.2. Principes des modèles hydrologiques.....	38
2.4.2. <i>Différents modèles utilisés sur le Gardon d'Anduze</i>	40
2.4.2.1. MARINE	40
2.4.2.2. TOPMODEL	40

2.4.2.3.	SCS.....	40
2.4.3.	<i>Critères de performance.....</i>	41
2.4.3.1.	RMSE.....	41
2.4.3.2.	Critère de Nash.....	42
2.4.3.3.	Critère de persistance	43
2.4.3.4.	PWRMSE.....	43
2.4.3.5.	Critères d'estimation de la qualité du pic de crue.....	44
2.5.	LE PROJET FLASH.....	44
2.5.1.	<i>Objectifs</i>	44
2.5.2.	<i>Les partenaires.....</i>	45
2.5.2.1.	L'EMA	45
2.5.2.2.	L'ESPCI ParisTech	46
2.5.2.3.	L'EDYTEM	46
2.5.2.4.	Le SCHAPI	46
2.6.	LA BASE DE DONNÉES	46
2.7.	CONCLUSION	48
3.	PRÉVISION DES CRUES ÉCLAIR SUR LE BASSIN VERSANT DU GARDON D'ANDUZE	51
3.1.	INTRODUCTION	51
3.2.	UTILISATION DES RÉSEAUX DE NEURONES POUR LA PRÉVISION DE LA HAUTEUR D'EAU À ANDUZE.....	51
3.2.1.	<i>Type de modèle utilisé.....</i>	51
3.2.2.	<i>Amélioration du modèle de base</i>	53
3.3.	SÉLECTION DE MODÈLES	54
3.3.1.	<i>Sélection du nombre de neurones cachés.....</i>	54
3.3.1.1.	Méthode de sélection.....	54
3.3.1.2.	Sensibilité du score de validation croisée partielle au nombre de neurones cachés	57
3.3.2.	<i>Sélection de la fenêtre temporelle des pluies.....</i>	59
3.3.2.1.	Méthode de sélection.....	59
3.3.2.2.	Sensibilité du score de validation croisée partielle à la dimension de la fenêtre temporelle des pluies.....	62
3.3.3.	<i>Sélection de l'initialisation des paramètres du réseau et ajustement de l'algorithme d'apprentissage</i>	63
3.3.3.1.	Sélection de l'initialisation du pas d'apprentissage et de son facteur de décroissance	63
3.3.3.2.	Sensibilité du score de validation croisée partielle à l'initialisation du pas d'apprentissage et à son facteur de décroissance.....	66
3.3.3.3.	Sélection de modèles issus de différentes initialisations des paramètres du réseau	67
3.3.4.	<i>Conclusion sur la sélection de modèles.....</i>	68
3.4.	PERFORMANCES DES MODÈLES SÉLECTIONNÉS SUR LES ÉVÉNEMENTS EN TEST	68
3.4.1.	<i>Caractéristiques des événements en test</i>	68
3.4.2.	<i>Prévisions des 20 modèles sélectionnés</i>	69
3.4.2.1.	Événement 19.....	69
3.4.2.2.	Événement 26.....	71
3.4.2.3.	Événement 27.....	74
3.4.3.	<i>Médiane des prévisions.....</i>	76
3.4.3.1.	Événement 19.....	77
3.4.3.2.	Événement 26.....	78
3.4.3.3.	Événement 27.....	81
3.4.4.	<i>Conclusion sur les performances des modèles à perceptron multicouche</i>	83
3.5.	MODIFICATION DU MODÈLE AVEC AJOUT D'UNE PARTIE LINÉAIRE	83
3.5.1.	<i>Modèle modifié.....</i>	83
3.5.2.	<i>Choix des hyperparamètres.....</i>	84
3.5.3.	<i>Prévisions des 20 modèles sélectionnés</i>	84
3.5.3.1.	Évènement 19.....	85
3.5.3.2.	Évènement 26.....	87
3.5.3.3.	Évènement 27.....	89
3.5.4.	<i>Médiane des sorties.....</i>	90
3.5.4.1.	Évènement 19.....	91
3.5.4.2.	Évènement 26.....	93
3.5.4.3.	Évènement 27.....	95
3.5.5.	<i>Conclusion sur les modèles avec une partie linéaire</i>	96
3.6.	CONCLUSION SUR LES MODÈLES POUR LA PRÉVISION DES CRUES À ANDUZE	97
4.	INTRODUCTION DE L'ADAPTATIVITÉ DANS LES MODÈLES DE PRÉVISION DES CRUES SUR LE BASSIN VERSANT DU GARDON D'ANDUZE.....	99

4.1.	INTRODUCTION	99
4.2.	PRINCIPE DE L'ADAPTATIVITÉ.....	99
4.3.	SÉLECTION DE LA FENÊTRE D'ADAPTATIVITÉ	101
4.3.1.	Évènement 19.....	101
4.3.2.	Évènement 26.....	104
4.3.3.	Évènement 27.....	107
4.4.	CONCLUSION SUR L'ADAPTATIVITÉ	109
5.	PRÉVISION DES CRUES SUR LE GARDON D'ANDUZE AVEC L'UTILISATION DES DONNÉES DE PLUIE FOURNIES PAR LES MESURES DES RADARS	111
5.1.	INTRODUCTION	111
5.2.	BASE DE DONNÉES	111
5.2.1.	Les données radar.....	111
5.2.2.	Traitements appliqués aux données.....	111
5.2.3.	Les événements.....	112
5.3.	COMPARAISON ENTRE LES PLUIES MESURÉES AU NIVEAU DES PLUVIOMÈTRES ET CELLES OBTENUES À PARTIR DES MESURES DU RADAR.....	113
5.4.	MODÈLES UTILISÉS ET RÉPARTITION DES ÉVÉNEMENTS DANS LES DIFFÉRENTS ENSEMBLES	115
5.4.1.	Sélection des modèles.....	116
5.4.2.	Résultats obtenus pour les différents horizons de prévision.....	117
5.4.2.1.	Évènement 13.....	117
5.4.2.2.	Évènement 19.....	119
5.5.	CONCLUSION SUR L'UTILISATION DES DONNÉES DE PLUIES FOURNIE PAR LES MESURES DU RADAR POUR LA PRÉVISION DES CRUES À ANDUZE.....	120
6.	PRÉVISION DES CRUES SUR LE BASSIN VERSANT DU GARDON À REMOULINS	123
6.1.	INTRODUCTION	123
6.2.	SITE D'ÉTUDE	123
6.3.	SÉLECTION ET CONSTITUTION DE LA BASE DE DONNÉES	124
6.4.	TYPES DE MODÈLES UTILISÉS	126
6.5.	CHOIX DES HYPERPARAMÈTRES	127
6.6.	RÉSULTATS	131
6.6.1.	Évènement 26.....	131
6.6.2.	Évènement 27.....	134
6.7.	CONCLUSION SUR LA PRÉVISION DES CRUES À REMOULINS	136
	CONCLUSION GÉNÉRALE	137
	BIBLIOGRAPHIE.....	139
	ANNEXES	147

Liste des figures

Figure 1 : Schéma d'un neurone formel avec ses variables x , ses paramètres w , sa fonction d'activation f et sa sortie y	16
Figure 2 : Schéma d'un réseau de neurones statique.....	18
Figure 3 : Schéma d'un réseau de neurones dynamique.....	18
Figure 4 : Modèle postulé à hypothèse bruit d'état.....	20
Figure 5 : Modèle postulé à hypothèse bruit de sortie.....	20
Figure 6 : Prédicteur neuronal associé à l'hypothèse bruit d'état.....	20
Figure 7 : Modèle neuronal associé à l'hypothèse bruit de sortie.....	21
Figure 8 : Schéma d'un perceptron avec une couche de neurones cachés.....	22
Figure 9 : Complexité du modèle et capacité de généralisation.....	26
Figure 10 : Schéma de la méthode de validation croisée.....	29
Figure 11 : Carte "Vigicrues" nationale du 7 Septembre 2010 avec vigilance orange pour les Gardons, la Cèze et l'Ardèche.....	34
Figure 12 : Carte "Vigicrues" du SPC Grand Delta le 3 Novembre 2011.....	35
Figure 13 : Carte du bassin versant du Gardon d'Anduze (conception G. Artigue).....	36
Figure 14 : Représentation systémique de la chaîne hydrométéorologique.....	45
Figure 15 : Différence entre modèles utilisés en prévision et en simulation avec k l'instant présent, f_T la fenêtre temporelle des pluies, r la fenêtre temporelle des hauteurs et h_p l'horizon de prévision.....	52
Figure 16 : Schéma du type de réseau utilisé.....	53
Figure 17 : Pour chaque nombre de neurones cachés* : moyenne, médiane et minimum sur les 200 scores de validation croisée partielle et moyenne des 20 meilleurs scores de validation croisée partielle, chaque graphique correspond à un horizon de prévision.....	55
Figure 18 : Méthode de sélection du nombre de neurones cachés.....	56
Figure 19: Pour chaque dimension de fenêtre temporelle : moyenne, médiane et minimum sur les 200 scores de validation croisée partielle et moyenne des 20 meilleurs scores de validation croisée partielle : chaque graphique correspond à un horizon de prévision.....	60
Figure 20: Schéma de la méthode de sélection de la fenêtre temporelle.....	61
Figure 21: Pour chaque combinaison de $(\mu_{mi}; \mu_{dec})$: moyenne, médiane et minimum sur les 200 scores de validation croisée partielle et moyenne des 20 plus petits scores de validation croisée partielle ; chaque graphique correspond à un horizon de prévision.....	64
Figure 22: Schéma de la méthode de sélection du pas d'apprentissage et de son facteur de décroissance.....	65
Figure 23: Schéma de la méthode de sélection de l'initialisation des paramètres du réseau.....	68
Figure 24: Événement 19, les 20 hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision.....	70
Figure 25: Valeurs des critères C_P , C_N , RMSE et PWRMSE pour le test de l'événement 19 sur les 20 modèles pour chaque horizon de prévision. L'axe des abscisses correspond au rang du modèle : le rang 1 correspond au modèle ayant le meilleur C_P et le 20 au plus mauvais C_P	71
Figure 26: Événement 26, les 20 hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision.....	72
Figure 27: Valeurs des critères C_P , C_N , RMSE et PWRMSE pour le test de l'événement 26 sur les 20 modèles pour chaque horizon de prévision. L'axe des abscisses correspond au rang du modèle par C_P décroissants.....	74
Figure 28: Événement 27, les 20 hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision.....	75
Figure 29: Valeurs des critères C_P , C_N , RMSE et PWRMSE pour le test de l'événement 27 sur les 20 modèles pour chaque horizon de prévision. L'axe des abscisses correspond au rang du modèle par C_P décroissants.....	76
Figure 30: Événement 19, hauteur médiane calculée sur les hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision.....	77
Figure 31: Événement 26, hauteur médiane calculée sur les hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision.....	80
Figure 32: Événement 27, hauteur médiane calculée sur les hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision.....	82
Figure 33 : Modèle modifié avec introduction d'une partie linéaire.....	84
Figure 34: Événement 19, les 20 hauteurs prévues par les 20 modèles, avec partie linéaire, sélectionnés pour chaque horizon de prévision.....	85
Figure 35: Événement 26, les 20 hauteurs prévues par les 20 modèles, avec partie linéaire, sélectionnés pour chaque horizon de prévision.....	87
Figure 36: Événement 27, les 20 hauteurs prévues par les 20 modèles, avec partie linéaire, sélectionnés pour chaque horizon de prévision.....	89

Figure 37: Événement 19, hauteurs médianes sur les hauteurs prévues par les 20 modèles avec partie linéaire et les 20 modèles à perceptron multicouche sélectionnés pour chaque horizon de prévision	91
Figure 38: Événement 26, hauteurs médianes sur les hauteurs prévues par les 20 modèles, avec partie linéaire, et les 20 modèles à perceptron multicouche sélectionnés pour chaque horizon de prévision	93
Figure 39: Événement 27, hauteurs médianes sur les hauteurs prévues par les 20 modèles avec partie linéaire et les 20 modèles à perceptron multicouche sélectionnés pour chaque horizon de prévision	95
Figure 40 : Schéma des données nécessaires pour réaliser les modèles adaptatifs, f_T étant la fenêtre temporelle des pluies et F_a la fenêtre d'adaptativité	100
Figure 41 : Schéma expliquant la modification des paramètres au cours de l'événement étudié.....	101
Figure 42 : Limnigrammes prévus à différents horizons de prévision pour l'événement 19 avec des modèles utilisant des fenêtres d'adaptativité différentes	102
Figure 43 : Limnigrammes prévus à différents horizons de prévision pour l'événement 26 avec des modèles utilisant des fenêtres d'adaptativité différentes	105
Figure 44 : Limnigrammes prévus à différents horizons de prévision pour l'événement 27 avec des modèles utilisant des fenêtres d'adaptativité différentes	107
Figure 45 : Découpage du bassin versant du Gardon d'Anduze [conception Y. Visserot], zone 1 = Saint Jean Amont, zone 2 = Mialet Rive Droite, zone 3 = Mialet Rive Gauche, zone 4 = Saint Jean Aval, zone 5 = Mialet aval, zone 6 = Salindrenque.....	112
Figure 46 : Événement 13, les hauteurs prévues par les modèles utilisant les données de pluviomètres et ceux utilisant les données radar pour chaque horizon de prévision	117
Figure 47 : Événement 19, les hauteurs prévues par les modèles utilisant les données de pluviomètres et ceux utilisant les données radar pour chaque horizon de prévision	119
Figure 48 : Bassin versant du Gardon (conception G. Artigue).....	124
Figure 49 : Modèle à perceptron multicouche avec h_p l'horizon de prévision, f_T , la fenêtre temporelle des pluies, f_1 à f_4 les fenêtres temporelles pour les hauteurs correspondantes et R_1 à R_5 les retards appliqués aux pluies et hauteurs correspondantes.....	126
Figure 50 : Modèle avec ajout d'une partie linéaire au perceptron multicouche avec h_p l'horizon de prévision, f_T , la fenêtre temporelle des pluies, f_1 à f_4 les fenêtres temporelles pour les hauteurs correspondantes et R_1 à R_5 les retards appliqués aux pluies et hauteurs correspondantes.....	127
Figure 51 : Moyenne sur tous les événements des corrélogrammes croisés avec pour série de référence la hauteur à Remoulins	129
Figure 52 : Pour chaque horizon de prévision, hauteurs médianes prévues pour l'événement 26 par les modèles à perceptron multicouche (PM) et avec une partie linéaire ajoutée (PM+PL)	133
Figure 53 : Pour chaque horizon de prévision, hauteurs médianes prévues pour l'événement 27 par les modèles à perceptron multicouche (PM) et avec une partie linéaire ajoutée (PM+PL)	135

Liste des tableaux

Tableau 1 : Les événements de la base de données.....	47
Tableau 2 : Caractérisation des événements	48
Tableau 3 : Dimension de la fenêtre temporelle utilisée pour la sélection de l'architecture des différents modèles	54
Tableau 4 : Nombre de neurones cachés en fonction de l'horizon de prévision	57
Tableau 5 : Pour chaque horizon de prévision, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de $M_{10\%}$	57
Tableau 6 : Pour chaque horizon de prévision et pour le nombre de neurones cachés déterminé, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de S_{vcp} sur 200 modèles obtenus avec 200 initialisations différentes des paramètres du réseau.....	58
Tableau 7 : Dimension de la fenêtre temporelle des pluies en fonction de l'horizon de prévision	61
Tableau 8 : Pour chaque horizon de prévision, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de $M_{10\%}$	62
Tableau 9 : Pour chaque horizon de prévision et pour la dimension de fenêtre temporelle déterminée, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de S_{vcp} sur 200 modèles obtenus avec 200 initialisations différentes des paramètres du réseau.....	62
Tableau 10 : Valeurs de l'initialisation du pas d'apprentissage et de son facteur de décroissance en fonction de l'horizon de prévision	65
Tableau 11 : Pour chaque horizon de prévision, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de $M_{10\%}$	66
Tableau 12 : Pour chaque horizon de prévision et pour la combinaison (μ_{ini} ; μ_{dec}) déterminée, moyenne, minimum, maximum, écart type, S_1 et S_2 des S_{vcp} sur 200 modèles obtenus avec 200 initialisations différentes des paramètres du réseau.....	66
Tableau 13 : Valeurs du C_p , C_N , PPD, SPPD et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 19.....	78
Tableau 14: Valeurs du C_p , C_N , PPD, SPPD et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 26.....	81
Tableau 15 : Valeurs du C_p , C_N , PPD, SPPD et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 27.....	83
Tableau 16 : Pour chaque horizon de prévision, sur les 20 prévisions de l'événement 19, calcul du minimum, du maximum, de la moyenne, de la médiane, de l'écart-type, des critères S_1 et S_2 des critères de persistance. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire (PM+PL).....	86
Tableau 17 : Pour chaque horizon de prévision, sur les 20 sorties de l'événement 26, calcul du minimum, du maximum, de la moyenne, de la médiane, de l'écart-type, des critères S_1 et S_2 des critères de persistance. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire (PM+PL).....	88
Tableau 18 : Pour chaque horizon de prévision, sur les 20 sorties de l'événement 27, calcul du minimum, du maximum, de la moyenne, de la médiane, de l'écart-type, des critères S_1 et S_2 des critères de persistance. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire ajoutée (PM+PL).....	90
Tableau 19 : Valeurs du C_p , C_N , PPD, SPPD et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 19. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire ajoutée (PM+PL).....	92
Tableau 20 : Valeurs du C_p , C_N , PPD, SPPD et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 26. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire ajoutée (PM+PL).....	94
Tableau 21 : Valeurs du C_p , C_N , PPD, SPPD et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 19. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire ajoutée (PM+PL).....	96
Tableau 22: Valeurs du C_p et du SPPD sur l'événement 19 pour les modèles ayant des fenêtres d'adaptativité différentes.....	103
Tableau 23 : Valeurs du C_p et du SPPD sur l'événement 26 pour les modèles ayant des fenêtres d'adaptativité différentes.....	106
Tableau 24 : Valeurs du C_p et du SPPD sur l'événement 27 pour les modèles ayant des fenêtres d'adaptativité différentes.....	108
Tableau 25 : Base de données comportant le numéro de l'événement, la date, la durée, le cumul de pluie des pluviomètres, le cumul de pluie du radar, la hauteur maximale à Anduze et le niveau de vigilance.....	113

Tableau 26 : Erreur relative (en %) entre les pluies des zones radar et les pluies des pluviomètres correspondantes.....	114
Tableau 27 : Coefficient de corrélation linéaire entre les pluies issues des pluviomètres et du radar.....	115
Tableau 28 : Pour chaque horizon de prévision, choix de la fenêtre temporelle et du nombre de neurones cachés pour les modèles avec les pluies des pluviomètres et avec les pluies des zones radar.....	116
Tableau 29 : Valeurs du C_p , C_N , PPD, SPPD et Retard H_{pic} suivant l'horizon de prévision pour les hauteurs prévues sur l'événement 13 par les modèles utilisant les pluies des pluviomètres et ceux utilisant les pluies du radar.	118
Tableau 30 : Valeurs du C_p , C_N , PPD, SPPD et Retard H_{pic} suivant l'horizon de prévision pour les hauteurs prévues sur l'événement 19 par les modèles utilisant les pluies des pluviomètres et ceux utilisant les pluies du radar.	120
Tableau 31: Base de données comportant le numéro de l'événement, la date, la durée, le cumul de pluie à Remoulins, la hauteur maximale à Remoulins et le niveau de vigilance.....	125
Tableau 32 : Caractérisation des événements	125
Tableau 33 : Décalage temporel par rapport à la hauteur à Remoulins pour une intercorrélacion maximale (selon la corrélation croisée).....	129
Tableau 34 : Pour chaque horizon de prévision, retard initial des différentes variables par rapport à l'instant de prévision.....	130
Tableau 35 : Pour chaque horizon de prévision, fenêtre temporelles des différentes variables d'entrée pour les deux types de modèles (PM = perceptron multicouche, $PM+PL$ = perceptron multicouche + partie linéaire) 131	131
Tableau 36 : Pour chaque horizon de prévision, critères de performance calculés sur les prévisions médianes de l'événement 26 obtenues par les modèles à perceptron multicouche (PM) et avec une partie linéaire ajoutée ($PM+PL$).....	132
Tableau 37 : Pour chaque horizon de prévision, critères de performance calculés sur les prévisions médianes de l'événement 27 obtenues par les modèles à perceptron multicouche (PM) et avec une partie linéaire ajoutée ($PM+PL$).....	134

Introduction générale

Les inondations constituent le risque naturel qui cause le plus de catastrophes en France. Le pourtour de l'arc méditerranéen est particulièrement touché, comme le montrent les événements marquants de ces vingt dernières années : Vaison-la-Romaine en 1992, l'Aude en 1999, le Gard en 2002 et le Var en 2010, qui ont eu pour conséquences un nombre important de victimes (dépassant la vingtaine dans chaque cas), et des dégâts matériels considérables qui ont atteint 1,2 milliards d'euros pour le Gard en 2002. Ces inondations dévastatrices sont causées par un phénomène appelé crue éclair, qui est difficilement prévisible et résulte de précipitations très intenses et localisées sur des bassins versants à réponse rapide.

Compte tenu des enjeux humains, la mise en place de moyens de prévention est indispensable. En 2003, le SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations) a été créé. Il est chargé de prévenir la population par une carte de vigilance disponible sur internet depuis 2006 : *vigicrues* (<http://www.vigicrues.gouv.fr/>). Les SPC (Services de Prévision des Crues) relaient les actions du SCHAPI au niveau local. Cependant, la complexité des processus hydrométéorologiques en jeu font de la modélisation des crues éclair une tâche ardue et pas encore complètement fiable.

Le projet FLASH (Flood forecasting with machine Learning data Assimilation and Semi-physical modeling) a été monté dans ce contexte, afin de réunir les compétences de plusieurs laboratoires et celles du SCHAPI pour se confronter au défi de la prévision des crues éclair. L'objectif final est de parvenir à la définition d'un prototype, utilisable en temps réel au SCHAPI pour alimenter la carte de *vigicrues* pour les cours d'eau étudiés. Les travaux contenus dans ce mémoire constituent une contribution à ce projet.

Dans une première partie, nous présenterons l'outil mathématique qui servira aux modélisations : les réseaux de neurones. Leurs propriétés fondamentales et leur fonctionnement seront détaillés, ainsi que les précautions à prendre lors de leur utilisation pour obtenir des résultats fiables. Les applications dans notre domaine d'intérêt, l'hydrologie, seront précisées.

La deuxième partie du mémoire commencera par une description de la zone principale d'étude, le bassin versant du Gardon d'Anduze, et une présentation de la disponibilité et de la qualité des mesures qui y ont été réalisées. Nous mentionnerons des exemples de modélisations pluie-débit sur ce bassin versant, issues de modèles faisant autorité en hydrologie. En alternative à ces modèles, nous proposerons l'approche du projet FLASH en mettant à profit les méthodes de l'apprentissage statistique. Enfin, les événements constitutifs de la base de données seront décrits.

La troisième partie sera centrée sur l'élaboration des modèles pour la prévision des crues sur le Gardon d'Anduze à Anduze, avec comme variables les pluies provenant des pluviomètres situés sur le bassin versant et des hauteurs précédentes à Anduze. Avec une méthode dérivant de la validation croisée, nous sélectionnerons différents hyperparamètres concernant l'architecture, le nombre de variables, l'ajustement de l'algorithme d'apprentissage et l'initialisation des paramètres du réseau. Nous estimerons empiriquement la sensibilité de la performance des modèles aux hyperparamètres puis nous évaluerons l'efficacité des modèles sur des événements qui leur sont inconnus. Deux types de

modèles seront utilisés : un perceptron multicouche (modèle non linéaire), et une combinaison linéaire d'un perceptron multicouche et d'un modèle linéaire.

Dans une quatrième partie, visant à améliorer les performances en généralisation, nous rendrons les modèles adaptatifs en ajustant les paramètres des réseaux en cours d'événement. L'essai sur trois événements-tests permettra de juger de la pertinence d'une telle forme d'adaptativité.

La cinquième partie verra l'introduction d'une nouvelle source de variables afin de remédier à l'éventuelle insuffisance de la seule estimation des précipitations par des mesures locales venant de pluviomètres. Nous appliquerons les pluies spatialisées issues de mesures radar que nous regrouperons par zones afin de pouvoir comparer les performances des modèles utilisant les données issues des pluviomètres à celles des modèles utilisant les données issues des radars.

Dans la sixième et dernière partie, la zone d'étude sera étendue, et notre travail sera consacré à la prévision des crues sur le bassin versant du Gardon à Remoulins. Ce bassin versant inclut le bassin versant d'Anduze précédemment étudié. Les surfaces et topographies de ces deux bassins n'étant pas comparables, un modèle légèrement différent, mettant à profit la connaissance que l'on a du bassin, sera proposé afin de prévoir au mieux la hauteur à Remoulins. Néanmoins, la méthodologie précédemment proposée au Chapitre 3 sera appliquée, et nous verrons qu'elle permet de prévoir les crues rapides à la station de Remoulins avec une qualité et une anticipation tout à fait satisfaisantes.

1. L'Apprentissage statistique avec réseaux de neurones

Quel que soit le type de modèle utilisant l'apprentissage statistique (réseau de neurones, machine à vecteurs de support, modèle de Markov caché, mélange de gaussiennes, arbre de régression,...), l'objectif est d'obtenir un modèle qui donne une réponse correcte à une situation qui n'a jamais été utilisée pendant la phase de conception du modèle. Cette propriété est appelée la *généralisation*. Le problème auquel nous devons faire face afin de parvenir à une bonne capacité de *généralisation* est de trouver la complexité optimale du modèle qui résoudra le *dilemme biais-variance* détaillé dans la première partie de ce chapitre.

Dans ce mémoire, nous avons choisi d'utiliser les réseaux de neurones formels pour la modélisation hydrologique. Le terme de « neurone » issu du vocabulaire de la biologie provient de l'origine historique de cet opérateur mathématique : les neurones formels ont été proposés par analogie aux neurones biologiques par le neurophysiologue Mc Culloch et le mathématicien W. Pitts en 1943 [McCulloch 1943], la fonction d'activation correspond à la relation entre le potentiel électrique de membrane et la fréquence des potentiels d'action, les entrées aux dendrites et la sortie à l'axone. Cependant cette analogie s'arrête là : les réseaux de neurones formels que nous utilisons dans ce mémoire se situent exclusivement dans le domaine de l'apprentissage statistique.

Les réseaux de neurones formels, que nous appellerons « réseaux de neurones », sont des modèles non linéaires paramétrés ; ils sont utilisés ici pour réaliser des fonctions. Contrairement aux modèles fondés sur des lois physiques ou empiriques provenant de l'étude des processus, les réseaux de neurones sont des modèles de type « boîte noire ». Dans cette étude, le processus ainsi modélisé est le déclenchement d'une crue éclair suite à des précipitations intenses.

La présentation des réseaux de neurones sera suivie par un énoncé de leurs propriétés fondamentales : l'approximation universelle et la parcimonie. Le problème principal de l'apprentissage statistique, la *généralisation*, sera traité à l'aide de méthodes de régularisation et de sélection. Enfin nous présenterons la modélisation des crues éclair avec les réseaux de neurones.

1.1. Le dilemme biais-variance

On suppose généralement que l'on peut modéliser l'observation y comme une réalisation d'une variable aléatoire qui est la somme d'une variable aléatoire *certaine* $\varphi(\mathbf{x})$, fonction des variables \mathbf{x} , appelée *fonction de régression*, et d'une variable aléatoire b , d'espérance mathématique nulle et de variance σ^2 , qui modélisent les bruits et perturbations non mesurées qui affectent le processus étudié :

$$y(\mathbf{x}) = \varphi(\mathbf{x}) + b$$

L'objectif de la conception d'un modèle par apprentissage est de trouver un modèle qui soit aussi proche que possible de la fonction de régression. Selon [Geman 1992], l'erreur de prédiction théorique, dans la limite d'un nombre d'exemples infini, est la somme de trois termes. Le premier terme ne provient que des données : il s'agit de la variance σ de la variable b . Considérant que les paramètres d'un modèle dépendent de l'ensemble d'apprentissage

choisi, on peut considérer les paramètres d'un modèle comme des réalisations de variables aléatoires. Dans cette approche, le modèle paramétré est donc lui-même une fonction de variables aléatoires : on peut alors définir sa variance, qui décrit la sensibilité du modèle au choix de l'ensemble d'apprentissage, et son biais, qui est l'espérance mathématique de la distance entre le modèle et la fonction de régression.

Comme tous ces termes sont positifs, l'erreur de généralisation d'un modèle obtenu par apprentissage ne peut pas être inférieure à la variance du bruit. Par ailleurs, le biais et la variance évoluent en sens inverse quand la complexité du modèle augmente. Plus un modèle est complexe, plus ses estimées seront proches des observations, mais le modèle est également plus sensible aux perturbations comprises dans l'ensemble d'apprentissage, ce qui fera augmenter la variance. Il convient ainsi de trouver la complexité du modèle pour laquelle la somme du biais et de la variance est minimum.

Nous devons évidemment tenir compte du dilemme-biais variance lors de la conception de nos modèles avec des réseaux de neurones.

1.2. Les réseaux de neurones

1.2.1. Le neurone formel

Un neurone formel (Figure 1) est un opérateur mathématique non linéaire qui est caractérisé par des paramètres, généralement appelés poids ou coefficients synaptiques, et par une fonction dite d'activation. Un neurone peut avoir une ou plusieurs variables et une ou plusieurs sorties.

Nous noterons :

- $\mathbf{x} = \{x_j, 0 \leq j \leq n\}$ le vecteur des variables
- y la sortie scalaire (nous n'utiliserons que des neurones à sortie unique)
- $\mathbf{w} = \{w_j, 0 \leq j \leq n\}$ le vecteur des paramètres
- f la fonction d'activation

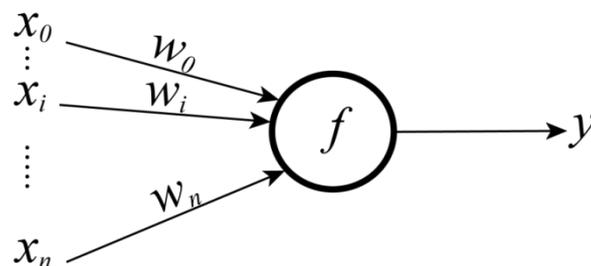


Figure 1 : Schéma d'un neurone formel avec ses variables \mathbf{x} , ses paramètres \mathbf{w} , sa fonction d'activation f et sa sortie y

La transformation des variables d'entrée par le neurone se fait en deux étapes. Le neurone calcule d'abord son potentiel, égal à la combinaison linéaire des variables dont les

coefficients sont les paramètres du neurone. Ensuite la fonction d'activation est appliquée au potentiel pour obtenir la sortie :

$$y = f\left(\sum_{i=0}^n w_i x_i\right)$$

Dans sa forme la plus simple, le neurone se réduit à un opérateur linéaire : lorsque la fonction d'activation est la fonction identité, la sortie est égale à la combinaison linéaire des entrées.

Les fonctions d'activation le plus souvent utilisées par des réseaux de neurones sont des fonctions sigmoïdes, par exemple la tangente hyperbolique. Ces fonctions d'activation sont bien adaptées aux tâches de classification pour lesquelles une décision binaire doit être prise, mais elles conviennent également comme fonctions élémentaires pour l'identification des paramètres d'un réseau modélisant d'un processus.

Dans ce travail, nous choisirons la tangente hyperbolique, qui est la plus couramment utilisée :

$$f(v) = \frac{2}{1 + e^{-2v}} - 1$$

1.2.2. Définition d'un réseau de neurones

Un réseau de neurones est une fonction réalisée par la composition de fonctions neurones élémentaires décrites dans la section précédente. Les neurones sont de deux sortes : les neurones cachés et les neurones de sortie. Les sorties des neurones dits « cachés » ne sont pas des sorties du réseau, tandis que les sorties des autres neurones sont celles du réseau. Cette distinction est importante, car, lorsque l'apprentissage est supervisé, ce qui est le cas dans ce travail, on dispose de valeurs mesurées pour les sorties du réseau, mais pas pour celles des neurones cachés.

On peut classer les réseaux de neurones en deux catégories, les réseaux de neurones statiques et les réseaux de neurones dynamiques.

1.2.3. Réseaux de neurones statiques

Les réseaux statiques sont des réseaux dont les sorties dépendent uniquement des valeurs des variables à l'instant considéré. Les sorties sont ainsi des fonctions à n variables où le temps ne joue pas de rôle fonctionnel ; si les variables ne varient pas dans le temps alors les sortie du réseau ne varient pas non plus. Ces réseaux sont qualifiés de « non bouclés » (Figure 2).

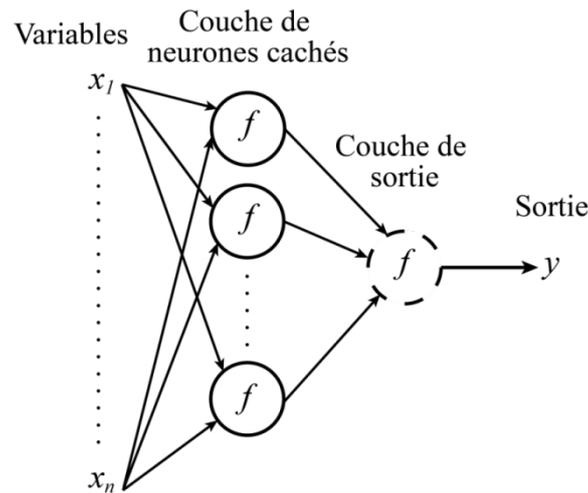


Figure 2 : Schéma d'un réseau de neurones statique

Lorsque les variables sont des signaux échantillonnés, le réseau est un filtre numérique transverse non linéaire et non récurrent :

$$y(k) = \varphi_{RN}(x_1(k), \dots, x_n(k), \mathbf{w})$$

où :

- $y(k)$ est la sortie du filtre à l'instant discret k
- $\mathbf{x}(k)$ est le vecteur des variables à l'instant k
- φ_{RN} est la transformation non linéaire effectuée par le filtre transverse
- \mathbf{w} est la matrice des paramètres

1.2.4. Réseaux de neurones dynamiques

1.2.4.1. Définition d'un réseau dynamique

Dans le cas des réseaux dynamiques, les sorties du réseau à un instant donné sont fonction des n variables (variables exogènes) et des valeurs passées des sorties du réseau (variables récurrentes). Le temps joue un rôle fonctionnel car même si les valeurs des variables exogènes ne varient pas dans le temps, les sorties du réseau peuvent varier. Le réseau est donc dynamique.

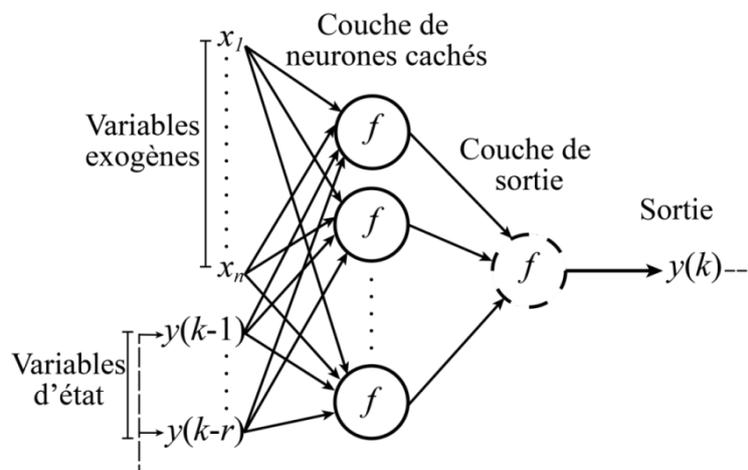


Figure 3 : Schéma d'un réseau de neurones dynamique

La sortie du réseau est reliée par au moins un retard à l'entrée. Les réseaux de neurones dynamiques sont des réseaux bouclés (Figure 3).

Un réseau bouclé dont les entrées sont les variables exogènes et des valeurs passées de la sortie, est un filtre récursif non linéaire.

1.2.4.2. Différentes représentations d'un système dynamique

Un système physique dynamique peut être décrit par un modèle reliant ses entrées et ses sorties discrétisées. Si le système physique est déterministe, la sortie du modèle est donnée par l'équation récurrente suivante :

$$y(k+1) = \varphi(y(k), \dots, y(k-r+1), u(k), \dots, u(k-m+1))$$

avec :

- k l'instant présent discrétisé
- y la sortie du modèle
- φ une fonction
- \mathbf{u} le vecteur des variables exogènes
- r et m des entiers positifs

Si le système dynamique est représenté par un modèle non déterministe, il peut être décrit mathématiquement différemment selon les processus internes considérés. Si l'on s'intéresse au rôle des perturbations, dont nous verrons dans la suite de ce mémoire qu'elles sont d'une importance majeure pour la prévision des crues rapides, on peut considérer deux types de prises en compte, selon que ces dernières opèrent principalement sur l'état du système ou sur sa sortie [Dreyfus, 2008]. Chacune de ces hypothèses conduit à un modèle postulé différent. Nous présentons dans la suite quels sont les prédicteurs adaptés à chacun des modèles ainsi postulés.

Supposons dans un premier temps que les perturbations peuvent être modélisées comme un bruit additif opérant sur l'état du système (Figure 4). Cette hypothèse est classiquement appelée *hypothèse bruit d'état*. On a alors :

$$y^p(k+1) = \varphi(y^p(k), \dots, y^p(k-r+1), u(k), \dots, u(k-m+1)) + b(k+1)$$

avec :

- y^p la sortie du processus observée
- \mathbf{u} le vecteur des variables exogènes
- b modélise les perturbations

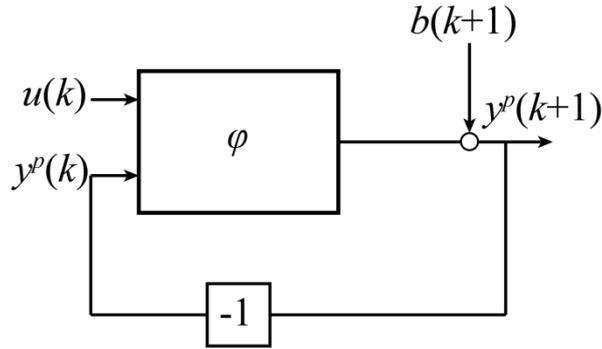


Figure 4 : Modèle postulé à hypothèse bruit d'état

Supposons à présent que les perturbations agissent sur la sortie (Figure 5) ; il s'agit typiquement d'un bruit de mesure supposé additif. Dans ce cas :

$$y^p(k+1) = \varphi(y^s(k), \dots, y^s(k-r+1), u(k), \dots, u(k-m+1)) + b(k+1)$$

avec :

- y^p la sortie du processus observée
- y^s la sortie d'état du processus représentée par le modèle entrée-sortie
- \mathbf{u} le vecteur des variables exogènes
- b représente les perturbations, dans ce cas c'est un bruit de mesure.

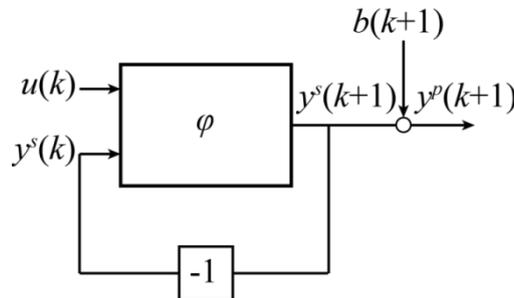


Figure 5 : Modèle postulé à hypothèse bruit de sortie

À chacun de ces modèles postulés correspond un modèle prédictif spécifique. Le premier type de modèle est un modèle de type NARX (Non linéaire Auto Régressif à entrées eXogènes) optimal pour l'hypothèse « bruit d'état » [Dreyfus, 2008]. Le modèle neuronal correspondant réalise la fonction :

$$y(k+1) = \varphi_{RN}(y^p(k), \dots, y^p(k-r+1), u(k), \dots, u(k-m+1), \mathbf{w})$$

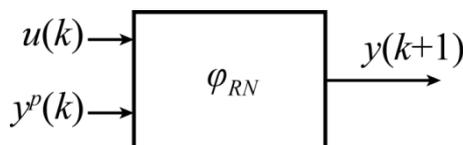


Figure 6 : Prédicteur neuronal associé à l'hypothèse bruit d'état

Ce modèle n'est pas récurrent : les entrées du réseau de neurones sont les variables exogènes et les sorties observées du processus (Figure 6). Comme on doit fournir au modèle les valeurs observées de la grandeur à modéliser, ce modèle peut être utilisé uniquement pour une prévision à un pas de temps. Si l'on souhaite obtenir une prédiction plus lointaine, on peut substituer aux grandeurs mesurées les sorties calculées aux pas de temps précédents, mais les performances sont alors détériorées.

Dans l'hypothèse « bruit de sortie », le modèle prédictif optimal est récurrent [Dreyfus, 2008]. Sa sortie est calculée de la façon suivante :

$$y(k+1) = \varphi_{RN}(y(k), \dots, y(k-r+1), u(k), \dots, u(k-m+1), \mathbf{w})$$

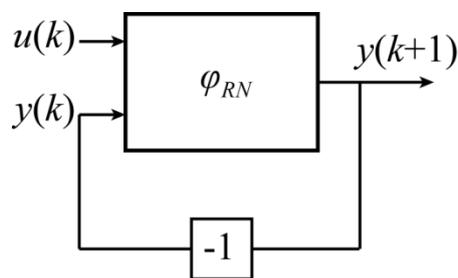


Figure 7 : Modèle neuronal associé à l'hypothèse bruit de sortie

Ce modèle est récurrent (Figure 7), et n'utilise pas les valeurs observées de la sortie du processus. Ce prédicteur peut fonctionner pour tout horizon de prédiction si l'on dispose des valeurs futures des variables exogènes, ce qui n'est pas toujours le cas pour les variables naturelles.

1.3. Le perceptron multicouche et ses propriétés fondamentales

1.3.1. Le perceptron multicouche

Un type de réseaux de neurones souvent utilisé est le perceptron multicouche. Les neurones sont organisés en couches, les sorties des neurones cachés sont les entrées des neurones de la couche suivante, les neurones cachés de la première couche ayant pour entrées les variables du réseau. Les neurones d'une même couche ne sont pas reliés entre eux. Les neurones des couches cachées ont des fonctions d'activation non linéaires et ceux de la couche de sortie ont des fonctions d'activation non linéaires ou linéaires. Le perceptron multicouche possède deux propriétés fondamentales intéressantes : l'approximation universelle et la parcimonie.

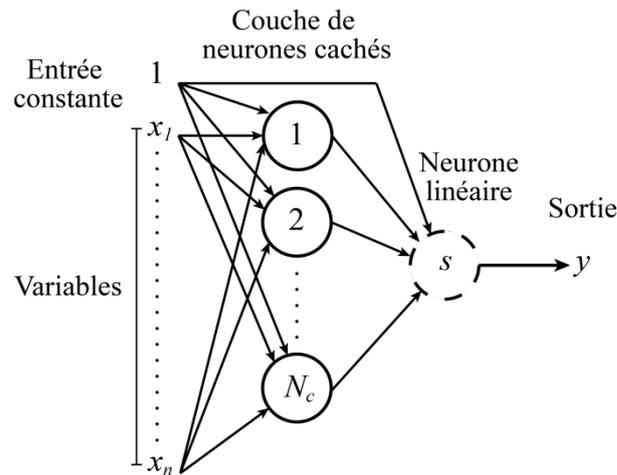


Figure 8 : Schéma d'un perceptron avec une couche de neurones cachés

Dans le cas d'un perceptron composé d'une seule couche de neurones cachés et d'une sortie unique comme présenté sur la Figure 8, la sortie est calculée selon l'équation suivante :

$$y = w_{s,0} + \sum_{j=1}^{N_c} w_{s,j} f \left(w_{j,0} + \sum_{i=1}^n w_{j,i} x_i \right)$$

avec :

- x_i la variable d'indice i
- N_c le nombre de neurones cachés
- $w_{s,0}$ le paramètre du réseau reliant l'entrée constante au neurone de sortie
- $w_{s,j}$ le paramètre reliant le neurone caché d'indice j au neurone de sortie
- $w_{j,i}$ le paramètre reliant la variable d'indice i au neurone caché d'indice j

Une entrée constante, ou biais, a été ajoutée aux variables. Elle permet, d'une part, de décaler horizontalement les sigmoïdes des neurones cachés par l'intermédiaire des poids $\{w_{j,0} \mid j = 1 \dots N_c\}$, et, d'autre part, d'ajouter un terme constant à la valeur de la sortie, si nécessaire, par l'intermédiaire du poids $w_{s,0}$.

1.3.2. L'approximation universelle

Un réseau de neurones construit avec une couche cachée, comportant un nombre fini de neurones ayant tous la même fonction d'activation et un neurone de sortie linéaire, possède la propriété d'approximation universelle : ce modèle peut approcher toute fonction bornée suffisamment régulière avec une précision arbitraire dans un domaine fini de l'espace de ses variables [Hornik 1989].

La propriété d'approximation universelle est un théorème d'existence, elle ne permet pas de calculer les paramètres du réseau la réalisant. Néanmoins, dans le domaine de l'identification des paramètres de modèles non linéaires, cette propriété est importante et justifie le choix de cette architecture dans notre travail.

1.3.3. *La parcimonie*

Un approximateur non linéaire en ses paramètres nécessite généralement moins de paramètres, toutes choses égales par ailleurs, qu'un approximateur linéaire en ses paramètres.

Ainsi pour un modèle linéaire en ses paramètres, le nombre de paramètres augmente exponentiellement lorsque le nombre de variables croît, alors que dans le cas d'un modèle non linéaire en ses paramètres comme les réseaux de neurones, le nombre de paramètres augmente linéairement [Barron 1993]. Cela confère aux réseaux de neurones la propriété de parcimonie. On peut augmenter le nombre de variables sans accroître considérablement la complexité du modèle.

1.4. Apprentissage statistique

L'étape d'apprentissage du réseau de neurones consiste à ajuster les paramètres du réseau de sorte que les valeurs calculées par le modèle soient aussi proches que possible des observations de la grandeur étudiée. Nous présentons dans ce qui suit comment l'apprentissage statistique exprime ce problème sous la forme d'une régression.

1.4.1. *Principe de l'apprentissage statistique*

L'apprentissage est l'estimation des paramètres du modèle ; il est réalisé à partir d'une base de données, appelée ensemble d'apprentissage. Chaque exemple de l'ensemble d'apprentissage est composé d'un couple comprenant les valeurs des variables du modèle et la valeur correspondante de la sortie du processus : $(\mathbf{x}_i ; y^p_i)$. Pour chaque valeur \mathbf{x}_i des variables, le réseau calcule la valeur correspondante y_i de sa sortie. Les paramètres du réseau sont alors modifiés afin de minimiser une fonction des écarts entre les valeurs des sorties du processus et celles qui sont calculées par le modèle.

La fonction généralement utilisée est la fonction de coût des moindres carrés :

$$J(\mathbf{w}) = \sum_{k=1}^N (y_k^p - y(\mathbf{x}_k, \mathbf{w}))^2$$

avec :

- N la dimension de l'ensemble d'apprentissage
- \mathbf{w} la matrice des paramètres du réseau
- \mathbf{x}_k le vecteur des variables du réseau pour l'exemple k de l'ensemble d'apprentissage
- y_k^p la valeur mesurée de la sortie du processus pour l'exemple k
- $y(\mathbf{x}_k, \mathbf{w})$ la valeur de la sortie du réseau pour l'exemple k

Le but de l'apprentissage est de minimiser cette fonction de coût. Comme le modèle est non linéaire en ses paramètres, la fonction de coût n'est pas quadratique en ses paramètres. La méthode des moindres carrés n'est donc pas applicable, et l'on a recours à des méthodes itératives d'optimisation.

1.4.2. Optimisation de la fonction de coût

Il existe plusieurs méthodes d'optimisation itératives. Elles sont fondées sur des algorithmes du type :

$$\mathbf{w}^i = \mathbf{w}^{i-1} - \mu \mathbf{M}^i$$

avec :

- \mathbf{w}^i les grandeurs par rapport auxquelles on effectue l'optimisation, à l'itération i (les paramètres du modèle dans le cas de l'apprentissage de celui-ci)
- μ un hyperparamètre de la méthode d'optimisation
- \mathbf{M}^i dépend uniquement du gradient de la fonction de coût pour les méthodes du premier ordre ; elle dépend également de la courbure de la fonction de coût pour les méthodes du second ordre

Dans notre travail nous avons essentiellement eu recours à une méthode du deuxième ordre: la méthode de Levenberg Marquardt [Marquardt 1963]. Cette méthode utilise les valeurs des dérivées premières du réseau, calculées à l'aide de l'algorithme de rétropropagation [Rumelhart 1986], pour estimer les dérivées secondes de la sortie par rapport aux paramètres. L'algorithme de Levenberg Marquardt (présentée plus amplement en Annexe 1) fait appel à des hyperparamètres μ_{ini} et μ_{dec} qui seront ajustés de manière heuristique durant cette étude. Les enjeux de cet ajustement et la méthode utilisée seront présentés dans le Chapitre 3.

Actuellement la méthode de Levenberg-Marquardt est reconnue comme une des méthodes les plus efficaces parmi les méthodes de gradient pour l'optimisation en variables continues.

1.4.2.1. Initialisation des paramètres

L'initialisation des paramètres du réseau joue un rôle important sur le déroulement de l'apprentissage et sur la performance du modèle en fin d'apprentissage. Le but recherché est d'atteindre le minimum de la fonction de coût. Cependant la fonction de coût, qui dépend non-linéairement des paramètres, possède généralement des minima locaux et différentes valeurs initiales des paramètres peuvent conduire à différents minima locaux de la fonction de coût. Par conséquent le choix de l'initialisation peut se révéler être déterminant pour l'efficacité de l'apprentissage.

Les enjeux de l'initialisation des paramètres sont liés à la vitesse de convergence lors de la phase d'apprentissage et à la capacité de généralisation du modèle une fois l'apprentissage réalisé. Comme nous l'avons dit précédemment, les modifications des paramètres reposent sur le gradient de la fonction de coût et donc sur la dérivée de la fonction d'activation par rapport à ces paramètres. Comme la fonction d'activation que nous utilisons est une fonction tangente hyperbolique, il est nécessaire qu'à l'initialisation, les fonctions d'activation ne soient pas saturées pour que leur dérivée ne soit pas pratiquement nulle. Si elles sont saturées, alors les modifications des paramètres sont trop faibles pour permettre une convergence de l'algorithme.

Pour éviter ce phénomène, les méthodes d'initialisation affectent une valeur aléatoire aux paramètres tirée dans une distribution satisfaisant à des contraintes sur sa variance. G. Thimm [Thimm 1997], propose une présentation des différentes méthodes d'initialisation qui conduisent en général à choisir les paramètres de manière aléatoire entre les deux valeurs $\{-a/\sqrt{n}; a/\sqrt{n}\}$, où n est le nombre de variables d'entrée des neurones. L'hyperparamètre a peut être choisi en fonction de la variance des paramètres initiaux et de la pente de la tangente hyperbolique, il varie en fonction des méthodes entre 0,77 et 3. Il apparait cependant que le principal facteur influençant le choix de la meilleure méthode d'initialisation est la base de données elle-même. Dans ce travail, nous adopterons une plage de variation de l'initialisation des paramètres aux valeurs $\{-1/\sqrt{n}; 1/\sqrt{n}\}$.

L'initialisation des paramètres peut également être déterminante pour la capacité de généralisation du modèle. En effet, pour un modèle non linéaire en ses paramètres, la fonction de coût n'est pas quadratique et peut présenter des minima locaux. En fonction de l'initialisation des paramètres, le processus d'optimisation va donc converger vers le minimum local le plus proche et ne pas en sortir. La procédure d'optimisation est donc répétée un grand nombre de fois afin d'explorer de manière la plus exhaustive possible l'ensemble des minima de la fonction de coût.

1.4.2.2. Pas d'apprentissage

Si le pas d'apprentissage, μ , des règles de premier ordre est trop petit, la vitesse de convergence est faible et le nombre d'itérations nécessaire pour atteindre un minimum peut devenir prohibitif. Si le pas est trop grand, il y a risque d'oscillations autour du minimum et même de divergence du processus d'optimisation. Un compromis peut consister à modifier la valeur du pas au cours de l'apprentissage.

1.4.3. Apprentissage pour les réseaux de neurones dynamiques

Selon [Nerrand 1993], tout réseau de neurones dynamique à sortie scalaire peut être mis sous une forme canonique composée d'un réseau de neurones non bouclé recevant en entrée des valeurs retardées de la sortie. L'apprentissage se déroule alors comme pour un réseau non bouclé en utilisant autant de « copies » du réseau non bouclé qu'il y a d'échantillons dans la séquence d'apprentissage, la ou les sorties de la copie correspondant à l'instant k faisant parties des variables de la copie $k+1$ (« dépliement temporel »). Nous ne rentrerons pas plus dans le détail de l'apprentissage pour les réseaux de neurones dynamiques, car ce type de réseaux ne sera pas utilisé dans nos travaux.

1.5. Problème du surajustement

Les résultats obtenus en apprentissage peuvent, au vu de l'approximation universelle, être aussi proches des sorties désirées qu'on le souhaite, si la complexité du modèle est suffisante. La complexité d'un modèle est définie par sa dimension de Vapnik-Cervonenkis [Vapnik 1998], qui, le plus souvent, et notamment dans le cas des réseaux de neurones, est une fonction croissante du nombre de paramètres du modèle : plus le nombre de paramètres du modèle est grand, plus sa complexité est grande, et plus il a de chance d'apprendre très précisément les exemples de l'ensemble d'apprentissage.

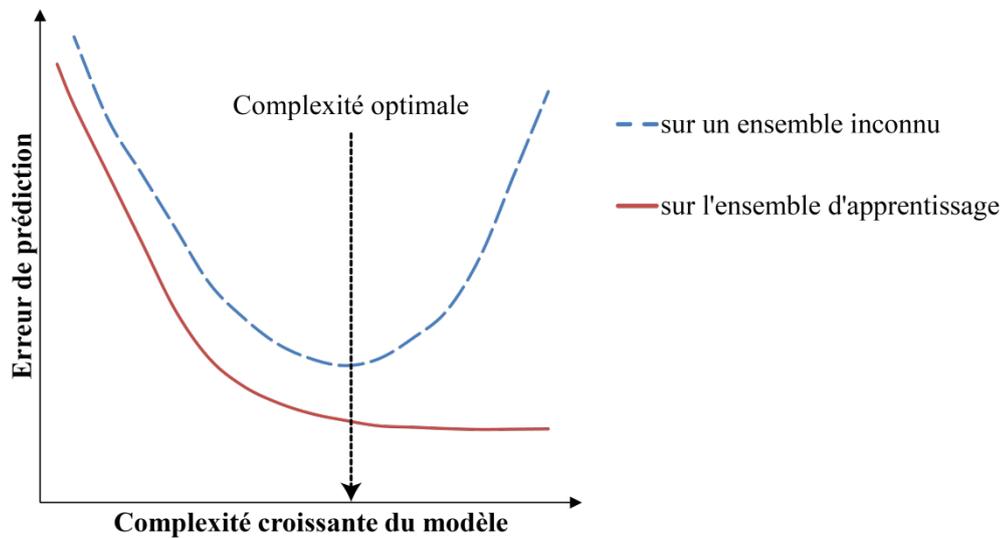


Figure 9 : Complexité du modèle et capacité de généralisation

Si un modèle est trop complexe, on observe le phénomène de surajustement : le modèle reproduit très bien les exemples d'apprentissage mais est incapable de généraliser comme l'illustre la Figure 9. Or, la capacité de *généralisation* étant primordiale, il va falloir résoudre au mieux le *dilemme biais-variance* en trouvant une complexité conduisant à minimiser la somme du biais et de la variance. La réalisation de ce compromis est difficile à obtenir lorsque l'ensemble d'apprentissage est restreint [Gallinari 1999], comme ce sera le cas dans nos travaux. Nous utiliserons des méthodes de régularisation pour y parvenir.

1.6. Méthodes de régularisation

Le surajustement peut être évité grâce à l'utilisation de méthodes de régularisation ; on en distingue deux types :

- Les méthodes qui diminuent la variance en opérant durant l'apprentissage, par exemple en limitant l'amplitude des paramètres du réseau au cours de l'apprentissage (modération des poids), ou en contrôlant le nombre d'itérations lors de l'apprentissage (arrêt précoce). Ces méthodes sont parfois qualifiées d'*actives*.

- Les méthodes qui diminuent la variance en éliminant les modèles surajustés une fois l'apprentissage réalisé, comme la validation croisée. Ces méthodes sont parfois qualifiées de *passives*.

1.6.1. L'arrêt précoce

Afin de limiter l'erreur de généralisation, l'arrêt précoce utilise un sous-ensemble de la base de données que l'on appelle l'*ensemble d'arrêt précoce*. À chaque itération du processus d'optimisation de la fonction de coût, on mesure la performance du réseau sur cet ensemble, c'est à dire la somme des carrés des écarts entre la sortie mesurée et la sortie calculée sur cet ensemble. Au fil des itérations, dans un premier temps, cette erreur diminue: le modèle devient de plus en plus performant à la fois sur l'ensemble d'apprentissage et sur l'ensemble d'arrêt précoce. On observe ensuite une augmentation de l'erreur quadratique calculée sur

l'ensemble d'arrêt: le modèle subit un surajustement sur l'ensemble d'apprentissage et se trouve incapable de généraliser correctement à l'ensemble d'arrêt. Si cette erreur quadratique ne diminue pas lors des itérations suivantes, on décide alors de stopper l'apprentissage à l'itération correspondant au minimum de l'erreur sur l'ensemble d'arrêt.

Selon [Sjöberg 1992], l'utilisation de l'arrêt précoce revient à l'ajout dans la fonction de coût d'un terme limitant l'amplitude des paramètres. Le fait de limiter l'amplitude des paramètres implique que certains d'entre eux restent très faibles et n'ont aucun rôle dans le calcul effectué par le réseau. Ainsi la complexité effective du modèle est réduite. La régularisation par limitation de l'amplitude des paramètres est décrite plus en détail dans la section suivante.

1.6.2. La modération des poids

Au cours de l'apprentissage, la valeur des paramètres du réseau augmente, ainsi des paramètres ayant une valeur importante peuvent être signe de surajustement.

D'après [Bartlett 1997], l'aptitude d'un réseau à généraliser dépend de la norme des poids plutôt que de leur nombre. Ainsi en empêchant les paramètres d'atteindre des valeurs trop importantes, on pourrait éviter le surajustement, une des alternatives de l'arrêt précoce revient à définir une nouvelle fonction de coût en y introduisant un terme de régularisation :

$$J' = J + \frac{\alpha}{2} \|\mathbf{w}\|^2$$

avec

- J la fonction de coût des moindres carrés,
- α un hyperparamètre qui détermine l'importance relative de la modération des poids par rapport à J .
- \mathbf{w} la matrice des paramètres du réseau

Plusieurs variantes existent, notamment avec un α différent selon la couche à laquelle appartiennent les paramètres.

Le calcul de ces hyperparamètres peut être réalisé par des méthodes statistiques [McKay 1992] ou heuristiques [Dreyfus 2008].

1.7. Sélection des variables

1.7.1. Principe

Comme pour une modélisation statistique classique, le choix des variables est important. En effet, si les variables sont très corrélées entre elles, elles apportent des informations équivalentes tout en augmentant la complexité, ce qui d'une part complique inutilement le système, mais également rend la matrice hessienne plus difficilement inversible. Ainsi l'apprentissage est plus lent et l'optimisation de la fonction de coûts sera moins réussie. Pour cela de nombreuses méthodes de sélection de variables ont été proposées.

La sélection de variables nécessite de mettre en œuvre :

- un critère quantitatif de performance qui sert à classer les variables selon leur pertinence,
- un seuil à partir duquel les variables seront jugées non pertinentes ; on peut par exemple rejeter les variables qui influencent moins la sortie du modèle qu'une variable dont les valeurs sont tirées aléatoirement [Stoppiglia 2003],

Dans ce qui suit nous allons présenter une méthode de sélection de variables : la méthode de la variable sonde. La méthode de validation croisée peut également être utilisée dans ce but, c'est cette dernière que nous choisirons au cours de l'élaboration de notre modèle.

1.7.2. La méthode de la variable sonde

La méthode de la variable sonde [Stoppiglia 2003], [Dreyfus 2006] consiste à :

- classer les variables candidates par ordre de pertinence décroissance, l'indice de pertinence d'une variable étant fondé sur la corrélation entre cette variable et la grandeur à modéliser,
- engendrer, et classer avec les autres variables, des « variables sondes » qui sont des réalisations de variables aléatoires,
- pour chaque variable candidate i , estimer la probabilité p_i pour qu'une variable sonde soit mieux classée qu'elle,
- fixer le risque δ de conserver une variable candidate alors qu'elle n'est pas pertinente ; si les exemples sont peu nombreux, il faut être très sélectif sur les variables, donc adopter une valeur faible pour δ ; si on dispose de nombreuses données, on peut être moins sélectif, donc admettre un risque δ plus grand,
- éliminer les variables candidates telles que $p_i > \delta$.

1.8. Sélection de modèle par la validation croisée

La validation croisée a été proposée par [Stone 1974] afin d'estimer l'erreur de généralisation en utilisant l'ensemble de la base de données disponible. Elle se déroule de la manière suivante (Figure 10) :

- on répartit les données en D sous-ensembles disjoints et de même cardinal,
- on effectue l'apprentissage du modèle sur $D-1$ sous-ensembles et on utilise le dernier sous-ensemble comme ensemble de validation,
- on calcule l'erreur de prédiction (l'erreur quadratique moyenne) sur le sous-ensemble mis en validation,
- on recommence D fois de sorte que tous les sous-ensembles soient tous passés une fois en validation,
- afin de quantifier la performance du modèle, on calcule un score de validation qui est la moyenne de l'erreur de prédiction évaluée sur chacun des sous-ensembles :

$$J_i = \frac{1}{N} \sum_{m=1}^N \left(y^m - y(x^m, \mathbf{w}) \right)^2, \text{ pour le sous ensemble } i \text{ qui comprend } N \text{ exemples,}$$

$$S = \sqrt{\frac{1}{D} \sum_{i=1}^D J_i}$$
 est le score de validation croisée.

Le score de validation croisée est une estimation de la capacité de généralisation du modèle considéré. Il peut servir à comparer les capacités de généralisation de différents modèles construits à partir des mêmes données, donc à sélectionner, parmi des modèles candidats, le « meilleur » modèle compte tenu des données disponibles. On n'utilise donc pas l'erreur d'apprentissage qui, comme nous l'avons indiqué précédemment, n'est pas un bon estimateur de l'erreur de généralisation.

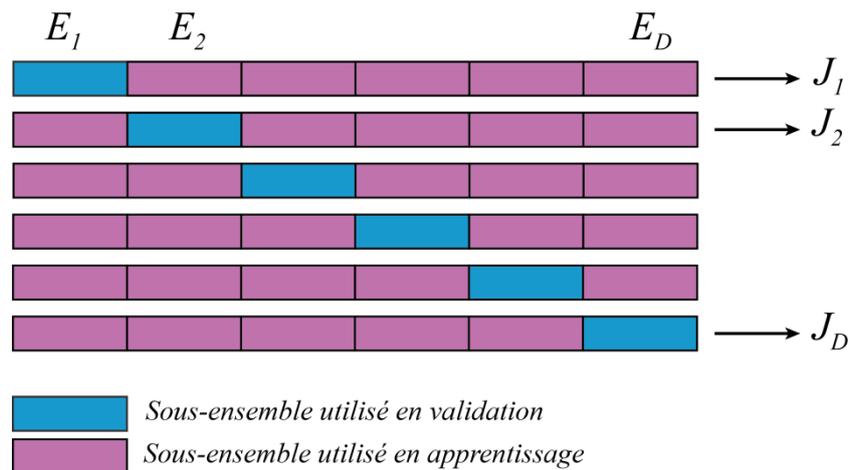


Figure 10 : Schéma de la méthode de validation croisée

Un cas particulier de la validation croisée est le leave-one-out où chaque sous-ensemble n'est composé que d'un seul exemple. Le score de leave-one-out est un estimateur non biaisé de l'erreur de généralisation. Il est coûteux en temps de calcul, si les données sont nombreuses, mais il est très utile lorsque les données sont peu nombreuses.

La validation croisée sera utilisée dans notre travail pour sélectionner les variables ainsi que la complexité du modèle et l'initialisation des paramètres.

1.9. La modélisation par réseaux de neurones pour la prévision

1.9.1. Simulation et prévision

Lorsque l'on s'intéresse à la modélisation d'un processus physique, il convient de différencier le mode de simulation et celui de prévision. La simulation (à distinguer de la simulation numérique) reproduit le fonctionnement du processus physique tel qu'observé, elle peut être réalisée par des modèles physiques, ou des modèles statistiques. Ainsi, si l'on considère l'instant k , la simulation d'un processus consiste à estimer la sortie de ce processus à l'instant k à partir des valeurs des variables d'entrée observées jusqu'à l'instant k .

De manière différente, la prévision des sorties d'un processus physique délivre ces sorties pour des instants postérieurs à k : à l'instant $k + h_p$, où h_p est l'horizon de prévision, à partir de variables physiques observées jusqu'à l'instant k . Le modèle de prévision ne dispose donc pas de l'intégralité des informations dont dispose le modèle de simulation ; le modèle de

prévision doit donc prévoir la sortie du processus au moyen de prévisions implicites des variables qui ne lui sont pas fournies à l'instant de prévision k . Le modèle de prévision, parce qu'il effectue potentiellement la prévision des variables exogènes ne peut donc représenter seulement le fonctionnement du système physique.

Simulation et prévision ont ainsi des visées différentes, la simulation permet une meilleure connaissance du processus alors que la prévision en fournit une anticipation. Néanmoins un modèle de simulation peut être utilisé en prévision si l'on dispose de séquences de prévisions des variables exogènes nécessaires.

1.9.2. Réseaux de neurones pour la prévision en hydrologie

Les réseaux de neurones sont utilisés dans des domaines très divers : ingénierie, physique, chimie, biologie, médecine, sciences humaines, économie, finance... Dans ce travail nous nous intéresserons au domaine de l'hydrologie.

De très nombreux travaux sont effectués pour appliquer les réseaux de neurones en hydrologie ; ils seront présentés de manière plus approfondie dans le chapitre suivant. Néanmoins on peut présenter synthétiquement quelques éléments.

Compte tenu de la complexité des processus intervenant à l'échelle d'un bassin versant, la propriété principale qui justifie l'utilisation des réseaux de neurones est la propriété d'approximation universelle. La modélisation statistique, ou boîte noire, a donc l'avantage de remédier au manque de connaissances dont on dispose sur les hydrosystèmes. Dans le même ordre d'idée, la spécificité de la prévision, différente de la modélisation physique, justifie plus encore l'utilisation de boîtes noires lorsque les prévisions des précipitations sont indisponibles en temps réel.

Considérant les modèles de réseaux de neurones utilisés, le perceptron multicouche est le plus utilisé et ses propriétés et avantages en hydrologie ont été mis en valeur par plusieurs articles de synthèse [Maier 2000], [Govindaraju 2000]. Concernant la prise en compte de l'aspect dynamique du processus physique, le modèle dirigé qui prend en compte les variables de débit ou hauteur précédentes afin de disposer d'une information sur l'état du système est le plus utilisé car le plus performant en général. Seules quelques approches utilisant un modèle récurrent ont été présentées [Johannet 2008] [Kong A Siou 2011] [Artigue 2012.b] [Roussel 2001] elles visent à remédier au défaut de mesure du débit du système pendant l'échéance de prévision, soit parce que les variables ne sont pas mesurées (cours d'eau non instrumenté), soit parce que ce sont des grandeurs physiques futures.

Si la modélisation boîte noire présente des intérêts pour le domaine de l'hydrologie, à l'inverse l'utilisation de modèles statistiques à des fins de meilleure connaissance des aquifères commence à être envisagée et est un défi scientifique actuel. Nous pouvons à ce propos citer le travail de L. Kong A Siou [Kong A Siou 2011], [Kong A Siou 2012.b], qui a développé la méthode KnoX (Knowledge eXtraction) afin d'extraire une certaine connaissance du modèle neuronal après l'apprentissage.

1.9.3. *Methodologie pour la conception d'un modèle de prévision avec réseau de neurones*

[Dreyfus 2008] détaille les différentes étapes de la création d'un prédicteur neuronal. Pour résumer, il faut :

- prétraiter les données en les normalisant notamment,
- sélectionner les variables pertinentes,
- réaliser l'apprentissage sur des modèles d'architectures différentes (faire varier le nombre de neurones cachés),
- sélectionner les modèles les plus efficaces,
- évaluer la performance de ces modèles sur un ensemble de test.

Nous nous inspirerons de cette démarche pour optimiser les modèles réalisés en vue d'effectuer la prévision des crues éclair du Gardon d'Anduze.

1.10. Conclusion

Nous avons présenté dans ce chapitre comment l'apprentissage statistique pouvait présenter une alternative à la modélisation physique des processus. Nous avons insisté sur le fait que la qualité des modèles utilisant l'apprentissage statistique dépend de leur capacité de *généralisation*. Il faut ainsi lors de leur conception parvenir à une complexité optimale qui satisfait le *dilemme biais-variance*.

L'utilisation des réseaux de neurones est intéressante pour la représentation de processus non linéaires statiques ou dynamiques. En particulier, le perceptron multicouche avec sa propriété d'approximation universelle est le type de réseau plébiscité pour l'identification ou la prévision en hydrologie.

Les paramètres des réseaux de neurones sont ajustés au cours de l'apprentissage statistique, fondé sur une base de données de couples variables mesurées/sortie mesurée. Un algorithme d'optimisation minimise la fonction de coût qui traduit l'erreur de modélisation à partir de la différence entre la sortie calculée par le modèle et la sortie mesurée sur le processus. Si l'optimisation peut être extrêmement efficace pour diminuer l'erreur de modélisation sur l'ensemble d'apprentissage, nous avons vu qu'en revanche elle ne garantit pas de bonnes propriétés de *généralisation* sur un ensemble inconnu. Qui plus est, l'erreur d'apprentissage et l'erreur de validation évoluent en sens opposés lorsque la complexité optimale du modèle est dépassée.

Ce *dilemme biais-variance* contraint le concepteur de modèle à sélectionner les variables et la complexité du modèle de manière méthodique et rigoureuse. À cette fin nous avons présenté plusieurs méthodes de sélection de variables et de complexité, ainsi que des méthodes de régularisation comme l'arrêt précoce ou la modération des poids.

Finalement nous avons pu aborder la fonction de prédiction qui peut être réalisée de manière immédiate en utilisant un modèle statistique mais ne peut prétendre modéliser seulement le phénomène physique en question ; il convient donc de différencier les modes de simulation et de prévision. Les réseaux de neurones sont utilisés dans de nombreux domaines dont l'hydrologie, notamment pour la prévision des crues éclair. Ils représentent une

alternative aux modèles basés sur des lois physiques largement employés en hydrologie. Cette propriété pourra être mise à profit pour l'anticipation des crues éclair du Gardon, crues que nous allons présenter plus amplement dans le prochain chapitre.

2. Contexte et site d'étude principal : le Gardon d'Anduze

Les inondations sont le phénomène naturel qui cause le plus de dommages en Europe ; améliorer les systèmes d'alerte à différentes échelles est donc une nécessité [Alfieri 2012]. Plus particulièrement, le pourtour méditerranéen se distingue par l'importante occurrence des crues éclair [Gaume 2009], crues catastrophiques du fait de leur soudaineté et des volumes d'eau engagés. Ainsi, les derniers épisodes d'inondation du Gardon en 2002 et du Var en 2010 ont causé en cumulé près de 50 victimes et plus de 2 milliards d'euros de dégâts ; c'est cette zone à risque qui est la motivation et le lieu de notre étude.

Le Gardon, affluent du Rhône, est un cours d'eau emblématique des crues éclair. Ce phénomène de montée très rapide du niveau d'eau est dû à des pluies intenses et localisées ; il prend localement le nom de *Gardonnades*, dues à des *épisodes cévenols* ou à des épisodes méditerranéens, tristement célèbres pour leur occurrence remarquable en automne. La première partie de ce chapitre sera consacrée à la présentation de ces crues éclair qui peuvent gonfler les cours d'eau d'un facteur 100 en moins d'une journée. Nous étudierons en particulier, le Gardon d'Anduze, partie amont du Gardon qui a son exutoire à Anduze.

Dans la seconde partie nous présenterons les données nécessaires à la modélisation : les pluies issues des pluviomètres et les hauteurs d'eau mesurées par les limnimètres. Nous nous pencherons sur les erreurs et incertitudes des mesures récoltées dont nous savons qu'elles sont particulièrement critiques pour un système conçu par apprentissage statistique. Nous pourrions ainsi mettre en évidence que, autant les incertitudes que l'indisponibilité des informations de pluie future causent une réelle difficulté pour un système statistique de prévision du débit. Nous examinerons également les avantages et inconvénients liés à la prévision de l'une ou l'autre des deux grandeurs de sortie possibles : la hauteur ou le débit.

Compte tenu des enjeux sociétaux s'y rapportant, le bassin versant du Gardon d'Anduze est l'objet de plusieurs programmes de recherches et de nombreuses modélisations hydrologiques ; nous présenterons dans la troisième partie les plus abouties d'entre elles, ainsi que leurs limites. Le projet FLASH, dans le cadre duquel s'est déroulée cette thèse, a pour but de proposer une alternative aux modèles hydrologiques fondés sur des approches physiques ou conceptuelles. Nous détaillerons les objectifs et les partenaires du projet.

Enfin nous présenterons la base de données dont nous disposons. Elle est composée de 17 événements de crues éclair ayant différentes formes de limnigrammes et de distribution des pluies.

2.1. Les crues éclair

Les crues éclair surviennent suite à des précipitations intenses et localisées sur des bassins versants de taille moyenne ou faible et à pente élevée, induisant ainsi une réponse rapide. Ces épisodes pluvieux sont difficilement prévisibles tant au niveau des cumuls de précipitation que de leur localisation. Les réponses des bassins à ces pluies sont très rapides avec des débits spécifiques (le débit spécifique est le débit rapporté à la surface du bassin versant) importants ; ainsi la hauteur des cours d'eau peut augmenter de plusieurs mètres en seulement quelques heures. Ces crues, du fait de leur soudaineté et de leur ampleur, peuvent causer de graves dommages en termes de vies humaines et de dégâts matériels. On peut citer

ainsi les crues de Nîmes (1988), Vaison-la-Romaine (1992), de l'Aude (1999), du Gard (2002 et 2003) ou encore du Var (2010).

Afin de limiter les conséquences dramatiques des inondations et de ces événements en particulier, le ministère en charge de l'Écologie a créé en 2003 le SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations) qui met en place des outils pour prévoir les crues et diffuser une carte de vigilance en libre accès sur internet : la carte *vigicrues* (<http://www.vigicrues.gouv.fr/>) (Figure 11, Figure 12). Au niveau régional, les SPC (Service de Prévision des Crues) remplacent les SAC (Service d'Annonce des Crues) depuis 2006 ; ils sont chargés de la surveillance, de la prévision et de la transmission de l'information des crues. Plus spécifiquement, le Gardon d'Anduze est sous la responsabilité du SPC Grand Delta.

La carte *vigicrues*, renseigne en temps réel le public sur les niveaux de vigilance pour les crues. A chaque cours d'eau et aux différentes sections instrumentées de ce cours d'eau est affecté un niveau de vigilance auquel correspond une couleur [vigicrues] :

- rouge = « *risque de crue majeure. Menace directe et généralisée de la sécurité des personnes et des biens* »,
- orange = « *risque de crue génératrice de débordements importants susceptibles d'avoir un impact significatif sur la vie collective et la sécurité des biens et des personnes* »,
- jaune = « *risque de crue ou de montée rapide des eaux n'entraînant pas de dommages significatifs, mais nécessitant une vigilance particulière dans le cas d'activités saisonnières et/ou exposées* »,
- vert = « *pas de vigilance particulière requise* ».

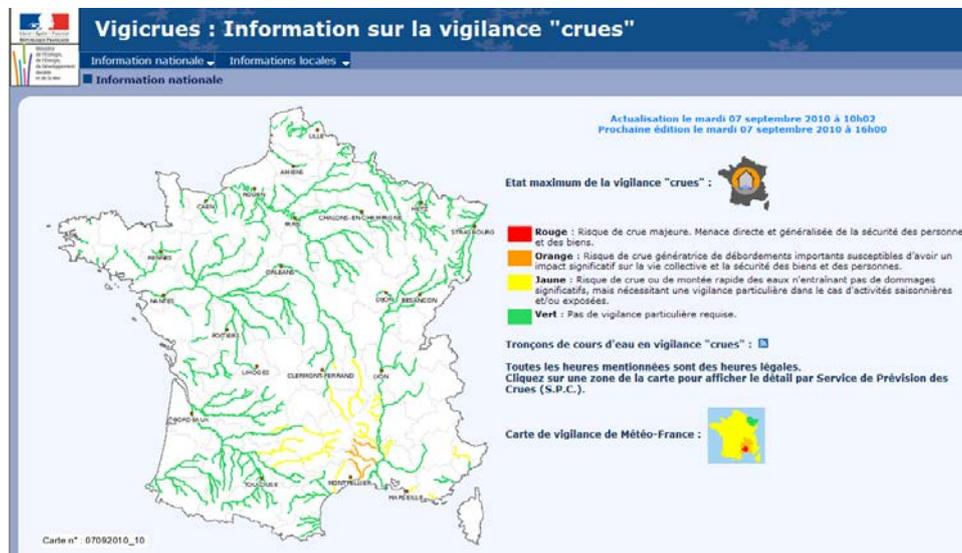


Figure 11 : Carte "Vigicrues" nationale du 7 Septembre 2010 avec vigilance orange pour les Gardons, la Cèze et l'Ardèche

En plus de la carte nationale (Figure 11), les informations locales sur les stations sont accessibles en cliquant sur la zone désirée (Figure 12).

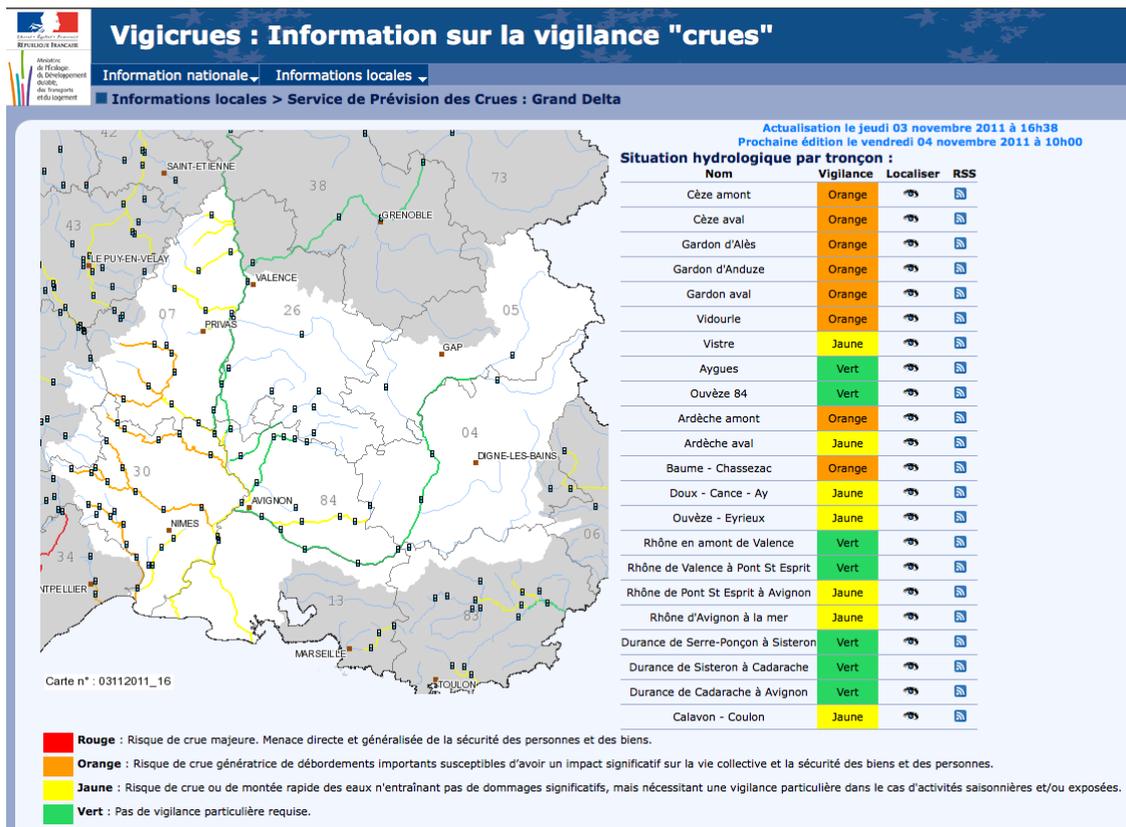


Figure 12 : Carte "Vigicrues" du SPC Grand Delta le 3 Novembre 2011

2.2. La zone d'étude

Le bassin versant du Gardon d'Anduze situé dans le Gard (Figure 13) a une superficie de 545 km². Du point de vue géologique, la partie amont du bassin est composée de schistes, micaschistes et gneiss jusqu'à Mialet ; puis suit une zone granitique autour de Soudorgues et on trouve des calcaires, dolomies et grès à l'aval. L'altitude maximale est de 1060m et celle de l'exutoire vaut 130m [Ayrat 2005] ; ce gradient important, associé à une pente moyenne de l'ordre de 10%, contribue au temps de réponse très court du bassin estimé entre 3 et 8 heures [Moussa 1991], ou environ 4h selon [Marchandise 2007].

Sur ce petit bassin, selon les informations fournies par Météo France, on mesure un grand nombre d'épisodes de pluies intenses. Dans un rayon de 25 km autour de Saint-Jean-du-Gard (situé au cœur du bassin versant d'Anduze), Météo France recense 53 événements dont la pluviométrie a dépassé 200 mm entre 6 h du matin d'un jour J et 6 h du jour $J+1$. Le nombre réel d'événements dont la pluviométrie a dépassé ce seuil sur une fenêtre de 24 heures glissante est donc nécessairement supérieur [Artigue 2012.a]. La valeur de 200 mm peut être mise en relation avec la pluviométrie moyenne à Paris qui s'élève à environ 600 mm : ces épisodes pluvieux sont donc réellement intenses.

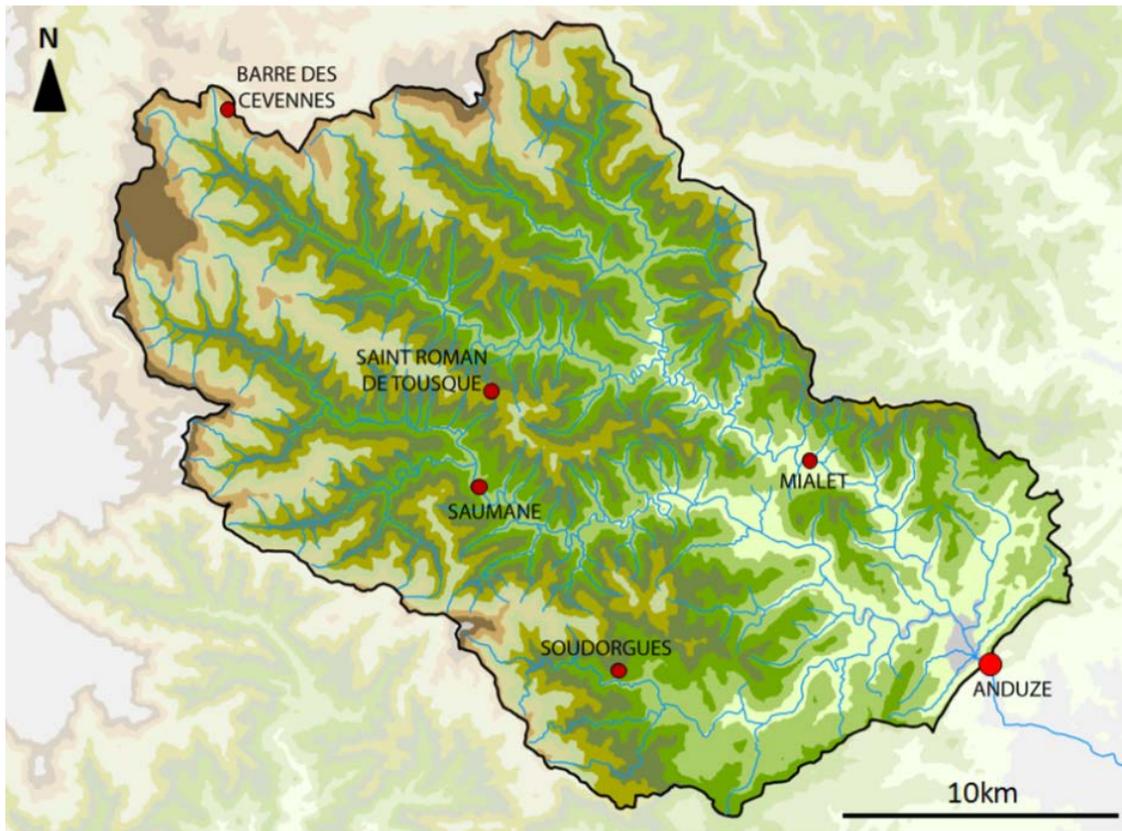


Figure 13 : Carte du bassin versant du Gardon d'Anduze (conception G. Artigue)

2.3. Erreurs et incertitudes des données mesurées

2.3.1. Précipitation

Les précipitations responsables des crues éclair sur le bassin versant du Gardon d'Anduze présentent une forte variabilité spatiale. Les cumuls de pluie relevés aux pluviomètres au cours d'un événement peuvent être très différents : c'est le cas par exemple pour l'événement d'octobre 2008 où le pluviomètre de Mialet a enregistré un cumul de 479 mm alors que celui de Barre-des-Cévennes, situé à moins de 30km, a eu un cumul de 9 mm. Selon [Obled 2009], 6 pluviomètres devraient suffire à représenter correctement la variabilité spatiale sur ce bassin versant. Il ne faut cependant pas omettre la possibilité que la majorité de la pluie se concentre sur une zone située entre les pluviomètres. Par ailleurs, la précision de la mesure des pluviomètres est elle-même sujette à des incertitudes qui vont de 10% à 20% d'après [Marchandise 2007].

Afin d'obtenir une meilleure résolution spatiale de la pluie, les lames d'eau (volumes de précipitations par unité de surface, équivalentes à des hauteurs de précipitations) fournies par les radars météorologiques sont de plus en plus utilisées. Deux radars couvrent le bassin versant du Gardon d'Anduze : celui de Nîmes et celui de Bollène ; les informations de pluies sont connues sur des mailles carrées d'une superficie de 1km². Les radars sont cependant sources d'autres erreurs [Jordan 2003] relatives aux taux de masquage (dus au reliefs), aux échos fixes (réflexion sur un obstacle) et à la forte non linéarité de la relation entre la réflectivité et la lame d'eau, relation qui peut beaucoup évoluer en fonction des conditions

d'utilisation. De plus, on notera un biais important pour certaines lames d'eau radar lorsque celles-ci sont calibrées à partir des données fournies par les pluviomètres. Qui plus est, l'emploi des données radar en modélisation hydrologique est relativement récent, et la technologie évoluant, les traitements effectués évoluent dans le temps et rendent donc l'utilisation de ces mesures difficile au sein d'une même base d'apprentissage. Nous utiliserons donc dans cette étude les pluies enregistrées par les pluviomètres, sauf dans le Chapitre 5 de ce mémoire, où nous évoquerons l'utilisation des données radar. Dans ce dernier cas, nous mettrons à profit la nouvelle analyse de ces données effectuée par Météo France afin d'homogénéiser les traitements que subissent les données brutes qui restaient jusqu'à présent très hétérogènes (réanalyse Comephore).

2.3.2. Hauteur/débit

La grandeur physique représentant l'état du bassin versant est le débit. On préférera cependant utiliser la hauteur du cours d'eau plutôt que le débit afin de s'affranchir des erreurs causées par la courbe de tarage (courbe indiquant la relation entre le débit et la hauteur d'un cours d'eau) qui n'est pas fiable pour les grandes valeurs de hauteur, les mesures de débit étant difficilement réalisables en cas de crue majeure.

Néanmoins, la mesure de la hauteur d'eau est également soumise à des sources d'incertitudes. Le lit de la rivière est susceptible de changer de forme suite à des crues où le débit a été particulièrement intense, ou bien à cause de travaux d'aménagement. En outre, il arrive que la station de mesure soit endommagée par la crue et que l'on soit contraint d'estimer la hauteur a posteriori, ce qui a été le cas pour l'événement de Septembre 2002 [SIEE 2004], l'erreur est encore accentuée. De manière schématique on peut estimer le cumul de toutes ces erreurs à une valeur de 30% [Marchandise 2007].

2.3.3. Échantillonnage

Suivant les contraintes opérationnelles du SCHAPI et de la dynamique du phénomène, la période d'échantillonnage a été fixée à ½ heure dans une précédente étude [Toukourou 2009.a]. Pour cette période, il a été montré que l'on pouvait perdre près de 10% de l'amplitude du pic sur l'événement le plus intense de septembre 2002. La même étude montre qu'un échantillonnage à 1h conduit à une sous-estimation de ce pic de 17% et à un retard du pic pouvant atteindre 1h. De manière générale, l'échantillonnage à 1h engendre des erreurs pouvant atteindre 30% sur l'estimation du pic de hauteur [Marchandise 2007].

Nous pensons que le choix de la demi-heure est un bon compromis entre la précision du signal et la complexité concernant le nombre de variables d'entrée qu'un échantillonnage plus fin imposerait au réseau de neurones. Nous verrons en effet que le modèle doit prendre en compte un historique des précipitations et que plus l'échantillonnage est fin, plus cela augmente le nombre de variables du modèle, donc le nombre des paramètres à estimer. Or, du fait du compromis biais-variance, le modèle doit être aussi parcimonieux que possible en termes de nombre de paramètres. Ce raisonnement est valable que ce soit pour les débits ou pour les hauteurs, même si, compte tenu de la non-linéarité de la courbe de tarage, sous-estimer le pic de hauteur de l'événement de 2002 de 10% conduit à sous-estimer le pic de débit de 20%. Au vu de ces éléments, il faut noter que les données antérieures à 2002 ont dû être ré-échantillonnées car elles ont été archivées avec un pas d'échantillonnage de 1h.

2.4. Modélisation des crues sur le bassin versant du Gardon d'Anduze

2.4.1. *Modélisation en hydrologie*

2.4.1.1. **Les différents types de modèles**

Les modèles hydrologiques sont classés suivant différents critères [Estupina Borrel 2004], [Mathevet 2005] :

- modèle déterministe/stochastique : les modèles stochastiques introduisent une part d'aléatoire alors que les modèles déterministes sont constitués d'équations portant sur des variables certaines,
- le domaine temporel : événementiel/continu : les données peuvent soit constituer des séries temporelles continues soit être composées d'événements ponctuels comme ce sera notre cas pour les crues,
- le fondement théorique : les modèles à base physique sont contrôlés par des lois physiques connues, les modèles conceptuels implémentent un concept global qui paraît pertinent comme celui des réservoirs, les modèles empiriques sont uniquement basés sur les observations ; les réseaux de neurones appartiennent à cette dernière catégorie,
- le niveau de spatialisation : les modèles globaux considèrent certaines caractéristiques du système comme homogènes sur sa totalité alors que les modèles spatialisés tiennent compte de l'hétérogénéité spatiale, des précipitations, ou du bassin versant (pente, géologie, ...),
- le mode d'affectation des valeurs des paramètres : ils peuvent être estimés par mesure sur le terrain ou par calage à partir de séries de données,
- la période d'échantillonnage : 30min, horaire, journalier, mensuel, annuel... Dans le cas des crues éclair, le pas de temps est généralement de 30min ou une heure,
- l'application : un modèle utilisé en simulation permet de mieux comprendre les processus en jeu dans le bassin, tandis qu'un modèle de prévision a pour objectif d'anticiper. Dans cette étude, nous serons conduits à développer des modèles de prévision qui estiment la hauteur d'eau dans le futur en fonction des seules données (pluies et hauteur) disponibles à l'instant de la prédiction. Ainsi, la fonction qui est demandée au modèle n'est plus de représenter le fonctionnement réel du bassin versant mais d'intégrer dans le modèle de prévision également une anticipation implicite sur les précipitations.

2.4.1.2. **Principes des modèles hydrologiques**

Les modèles hydrologiques cherchent à établir la relation entre la pluie tombée sur le bassin versant et le débit à son exutoire. Or, les processus hydrométéorologiques et hydrologiques responsables des crues éclair sont très complexes et seulement partiellement connus [Artigue 2012.b].

Concernant les pluies, plusieurs phénomènes peuvent être considérés :

- une partie de la pluie contribue directement à la crue via le ruissellement,
- une autre partie de la pluie s'infiltré dans le sol : elle participe donc à l'humidification du bassin, et une partie rejoint le cours d'eau,

- enfin une partie des précipitations n'a aucun impact sur la crue à cause des phénomènes d'évapotranspiration et d'interception par la végétation. Cette dernière proportion de pluie ne sera pas déterminante dans les crues éclair compte tenu des volumes d'eau mis en jeu et de la rapidité du phénomène.

L'hétérogénéité spatiale des précipitations est prise en compte par de nombreux modèles, par les *champs de pluie* qui décrivent cette variabilité spatiale. Ces champs de pluie peuvent être obtenus par l'image radar ou grâce à la mesure issue de plusieurs pluviomètres et spatialisée par des méthodes de krigeage.

Dans le cadre de la construction d'un prédicteur, les entrées de pluie posent deux problèmes principaux : d'une part, comme nous l'avons indiqué précédemment, la mesure de la pluie est entachée d'incertitudes et d'erreurs, et, d'autre part, il n'existe pas de prévision de pluie aux échelles spatiales (quelques dizaines de km²) et temporelles (moins d'une dizaine d'heures) pertinentes pour les crues éclair. Or pour calculer le débit, les modèles hydrologiques physiques et conceptuels ont besoin de l'information de pluie. Il est donc nécessaire, pour les utiliser, de formuler des hypothèses d'évolution de la pluie future.

Classiquement, trois types d'hypothèses sont utilisés : les pluies futures nulles, les pluies futures constantes (égales à la dernière mesure de pluie) et les pluies dites « parfaites » c'est-à-dire qui correspondent aux pluies observées. Bien entendu le dernier mode de *pluies parfaites* n'est pas utilisable pour effectuer des prévisions en temps réel. Les deux premiers modes ne peuvent également donner satisfaction car le débit prévu est soit fortement sous-estimé soit fortement surestimé suivant le cours de l'événement ; tout au plus ces hypothèses peuvent-elles apporter un encadrement des valeurs du débit qui se produira réellement. En revanche, les modèles empiriques comme les réseaux de neurones peuvent apprendre une relation, qui, comme nous l'avons indiqué précédemment, ne permet pas d'expliquer le fonctionnement du bassin versant, mais qui permet de réaliser une prévision.

Les modèles hydrologiques physiques et conceptuels calculent les estimations de débit à l'aide de fonctions dites de production et de transfert. La fonction de transfert traduit la propagation de l'eau dans la rivière. La fonction de production utilise la pluie et les caractéristiques du bassin versant afin de déterminer la quantité d'eau rejoignant le cours d'eau. Deux hypothèses sur le ruissellement apparaissent généralement dans les modèles à base physique :

- Le ruissellement hortonien [Horton 1933] apparaît lorsque l'intensité des pluies devient supérieure à la capacité d'infiltration du sol. Ce ruissellement peut se ré-infiltrer ou rejoindre le cours d'eau.

- Le ruissellement hewlettien [Hewlett 1961] confère aux sols une grande capacité d'infiltration jusqu'à leur saturation, à partir de laquelle le terrain est considéré comme imperméable. Le ruissellement dépend donc de la surface saturée, qui est influencée par la nature des sols et la topographie.

Plusieurs modèles hydrologiques ont été testés sur le bassin versant du Gardon d'Anduze afin de simuler et/ou prévoir les crues éclair. Nous allons brièvement présenter trois d'entre eux, qui ont participé au projet BVNE (Bassins Versants Numériques Expérimentaux) initié

par le SCHAPI et qui ont présenté les résultats les plus significatifs. Deux d'entre eux (MARINE et SCS) sont en cours d'implémentation dans la chaîne de prévision en temps réel du SCHAPI, en parallèle du modèle EMEAU avec réseaux de neurones sur lequel nous avons travaillé.

2.4.2. Différents modèles utilisés sur le Gardon d'Anduze

2.4.2.1. MARINE

MARINE (Modélisation de l'Anticipation du Ruissellement et des Inondations pour des événements Extrêmes) s'intéresse aux phénomènes physiques à différentes échelles : le MNT (Modèle Numérique de Terrain), l'occupation des sols (forêts, urbanisation, champs...), la texture du sol dont dépend la capacité d'infiltration, et les caractéristiques du chenal du cours d'eau (géométrie et rugosité) [Estupina Borrel 2004]. Le processus essentiel est le ruissellement. L'utilisation de MARINE sur le Gardon d'Anduze a donné des résultats très satisfaisants pour simuler le débit. En particulier [Braud 2010] a montré que la distribution spatiale des pluies a un rôle déterminant sur la génération des crues. L'avantage de ce type de modèles à base physique, entre autres informations sur le fonctionnement du bassin, est qu'il peut produire une estimation du débit en tout point du réseau hydrographique.

2.4.2.2. TOPMODEL

TOPMODEL (TOPography based hydrological MODEL) qui a été initialement développé par Beven et Kirkby [Beven 1979], est un modèle à base physique qui repose sur l'hypothèse des aires contributives saturées : l'eau ruisselle sur ces zones par incapacité d'infiltration sur un sol ayant dépassé le seuil de saturation [Dunne 1983]. Le ruissellement par aires contributives dépend essentiellement de la topographie. Ce modèle est spatialisé : il prend en considération la variabilité spatiale des pluies et des caractéristiques du bassin versant. Le bassin versant est ainsi maillé. [Le Lay 2007] et [Saulnier 2009] en appliquant TOPMODEL aux bassins cévenols (Gardon, Cèze, Ardèche) ont montré que l'information à prendre en compte est principalement la distribution spatiale des pluies. En second niveau apparaît l'humidité du sol, qui conditionne la saturation du sol et donc la réponse hydrologique. Les simulations sur l'événement très intense de septembre 2002 sont convaincantes sur la capacité du modèle à reproduire le fonctionnement des bassins.

2.4.2.3. SCS

SCS (Soil Conservation Services) [Bouvier 2006] est un modèle conceptuel où le bassin versant est considéré comme un réservoir que les précipitations remplissent. Le débit à l'exutoire correspond à la vidange du réservoir et dépend de la hauteur d'eau dans ce dernier. Seuls deux paramètres sont calculés : la transmissivité et la capacité du réservoir, qui sont les mêmes sur l'ensemble du bassin versant. Les pluies sont spatialisées. Ce modèle donne des résultats intéressants en simulation au vu de sa simplicité. [Coustau 2011] a appliqué ce modèle au Gardon d'Anduze et [Bouvier 2010] à un de ses tout petits bassins amont ; ils mettent en évidence le rôle majeur de l'initialisation de l'humidité des sols qui, pour SCS traduit le remplissage initial du réservoir « sol ».

En synthèse des travaux visant à appliquer ces trois modèles à base physique ou conceptuelle sur les bassins cévenols tels que celui du Gardon d'Anduze, on peut noter que ces modèles ne sont pas utilisés pour la prévision des débits ; ils n'ont pas été conçus pour cela, et, à notre connaissance, il n'y a pas de publication présentant des résultats de prévision opérationnelle (qui n'utilisent pas les pluies parfaites), avec ces modèles.

Il convient de noter également que ces trois modèles présentent des capacités de représentation des événements de crues satisfaisantes, alors qu'il sont fondés sur des hypothèses de fonctionnement hydrologiques très différentes, voire contradictoires. Ceci n'ôte rien à leur intérêt pour simuler le débit en tout point du réseau hydrographique, mais suggère que le fonctionnement du bassin qui conduit aux crues éclair n'est pas encore bien connu et que les modèles soumis à une procédure de calage automatique performante vont bien s'adapter aux hydrogrammes mesurés, sans pour autant permettre de valider le fondement physique sous-jacent. Ce phénomène est appréhendé en hydrogéologie par des auteurs travaillant sur l'eau souterraine, qui cherchent pour cela à ajouter des contraintes supplémentaires au modèle durant son calage, par exemple des contraintes relatives à la chimie des eaux (qui est liée à son temps de transfert en souterrain) [Charlier 2012]. On peut donc penser que l'approche de prévision des crues par apprentissage statistique est une voie digne d'intérêt car elle ne demande pas d'hypothèse sur la fonction à implémenter et qu'elle permet de réaliser des prévisions de hauteur ou de débit en l'absence de prévision de pluies [Toukourou 2010]. En revanche les études hydrologiques nous ont renseignés sur les variables pertinentes à prendre en compte : les précipitations spatialisées et le taux d'humidité du sol avant et pendant le déroulement de l'événement (mesure physique difficile à obtenir en temps réel et avec une bonne représentativité spatiale compte tenu de l'hétérogénéité des sols).

Les prévisions de pluie sont nécessaires lorsque l'horizon de prévision augmente et pas seulement en deçà du temps de réponse du bassin, puisqu'à un instant donné, les pluies tombées depuis un temps plus court que le temps de réponse n'ont pas encore d'impact sur le débit. [Dolciné 2001]. [Alfieri 2012] souligne l'intérêt du couplage entre modèles météorologiques et hydrologiques dans le cadre du système d'alerte aux crues Européen (EFAS). Ces couplages sont, à n'en pas douter, les pistes les plus prometteuses pour l'amélioration de la prévision hydrologique.

2.4.3. Critères de performance

Afin d'estimer la qualité des résultats d'un modèle et de pouvoir comparer ses performances avec celles d'autres modèles, on utilise des critères de performance. Parmi tous les critères existants nous en avons sélectionné quatre qui paraissent appropriés pour évaluer les crues et les performances en prévision.

2.4.3.1. RMSE

Le RMSE ou Root Mean Squared Error, est la racine carrée de l'erreur quadratique moyenne. Ce critère a la même dimension que la grandeur étudiée, ce qui donne un ordre de grandeur de l'erreur commise.

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k^p - y_k)^2}$$

avec :

- y_k^p sortie mesurée à l'instant k
- y_k sortie calculée à l'instant k

Cependant, en hydrologie, ce critère ne permet pas de comparer la performance du modèle entre plusieurs événements. En effet un même RMSE pour des événements ayant des amplitudes de hauteur différentes, n'est pas interprété de la même façon : un même RSME sur un événement avec une hauteur de pic importante correspond à un meilleur résultat que sur un événement de faible hauteur de pic, l'erreur relative étant plus petite avec des valeurs de hauteurs plus grandes. C'est pourquoi il est intéressant de calculer également des critères normés comme le critère de Nash ou la Persistance.

2.4.3.2. Critère de Nash

Le R^2 , ou critère de Nash [Nash 1970] rapporte l'erreur quadratique moyenne à la variance des mesures.

$$C_N = 1 - \frac{\sum_{k=1}^n (y_k^p - y_k)^2}{\sum_{k=1}^n (y_k^p - \overline{y^p})^2}$$

avec :

- y_k^p sortie mesurée à l'instant k
- y_k sortie calculée à l'instant k
- $\overline{y^p}$ moyenne de la sortie mesurée sur l'ensemble considéré

Les valeurs du critère de Nash sont comprises entre $-\infty$ et 1. Le critère de Nash vaut 1 quand la sortie calculée est égale à la sortie mesurée, donc quand le modèle est parfait. Si la valeur du critère de Nash est nulle, cela signifie que la sortie calculée ne donne pas de meilleure performance en termes d'erreur quadratique que la moyenne des mesures. En hydrologie, le critère de Nash est considéré comme bon lorsqu'il a une valeur supérieure à 0,8.

Ce critère est intéressant car il permet de comparer des événements pluvieux de caractéristiques différentes. Toutefois, dans le cas des crues, il faut prendre un certain recul sur l'interprétation de la valeur de ce critère. En effet, il tient compte de toutes les valeurs de hauteur (ou de débit) au cours de l'événement : on peut donc obtenir un bon critère de Nash si le modèle estime correctement les hauteurs juste avant le pic et lors de la décroissance, tout en ayant une mauvaise prévision du pic. Or la qualité du pic est primordiale pour un modèle de prévision de crues : il faut donc utiliser d'autres critères pour compléter l'appréciation conférée par le critère de Nash. De plus, le critère de Nash ne prend pas en compte le fait que

hauteur obtenue par le modèle est une prévision ; c'est pourquoi on utilise également le critère de persistance.

2.4.3.3. Critère de persistance

Le critère de persistance [Kitadinis 1980] permet de comparer la prévision du modèle à la prévision naïve, c'est-à-dire la prévision qui donne à la grandeur concernée une valeur future égale à la valeur actuelle.

$$C_p = 1 - \frac{\sum_{k=1}^n \left(y_{k+h_p}^p - y_{k+h_p} \right)^2}{\sum_{k=1}^n \left(y_{k+h_p}^p - y_k^p \right)^2}$$

avec :

- $y_{k+h_p}^p$ sortie mesurée à l'instant $k+h_p$ où h_p est l'horizon de prévision
- y_{k+h_p} sortie calculée à l'instant $k+h_p$
- y_k^p sortie mesurée à l'instant k

Le critère de persistance peut prendre des valeurs comprises entre $-\infty$ et 1. Lorsque le critère de persistance est égal à 1, cela signifie que les valeurs calculées par le modèle sont égales aux valeurs mesurées ; le modèle est le meilleur possible. Si le critère de persistance est nul, alors les valeurs calculées par le modèle sont de qualité équivalente aux valeurs de la prévision naïve qui consiste à prédire que la grandeur à l'instant $k+1$ aura la même valeur que la grandeur mesurée à l'instant k . Si le critère de persistance est négatif, le modèle donne de moins bons résultats que la prévision naïve : il présente donc peu d'intérêt.

2.4.3.4. PWRMSE

Le critère Peak-Weighted Root Mean Squared Error ou PWRMSE [Amengual 2007], pondère l'erreur quadratique par un coefficient qui augmente avec la valeur de la grandeur concernée. Ce coefficient est donc plus grand pour les valeurs du pic de crue, ce qui donne plus d'importance à l'erreur quadratique au niveau du pic, et par conséquent reflète mieux la performance du modèle au niveau du pic, qui est la zone la plus intéressante.

$$PWRMSE = \left[\frac{1}{2n} \sum_{k=1}^n \left[\left(y_k^p - y_k \right)^2 \left(1 + \frac{y_k^p}{y^p} \right) \right] \right]^{\frac{1}{2}}$$

avec :

- y_k^p sortie mesurée à l'instant k
- y_k sortie calculée à l'instant k
- $\overline{y^p}$ moyenne de la sortie mesurée

2.4.3.5. Critères d'estimation de la qualité du pic de crue

Pour estimer la qualité des prévisions au niveau du pic de crue, nous utiliserons trois critères qui évalueront l'amplitude du pic prévu et sa synchronisation par rapport au pic observé :

- le Percentage Peak Discharge, *PPD*, qui correspond au pourcentage de la hauteur maximale calculée du pic par rapport à la hauteur maximale du pic observée,
- le Synchronous Percentage Peak Discharge, *SPPD*, qui est le pourcentage de la hauteur calculée à l'instant de la hauteur maximale du pic observée par rapport à la hauteur maximale du pic observée,
- le décalage temporel entre la hauteur maximale du pic prévu et la hauteur maximale du pic observé, que l'on appellera Retard H_{pic} .

Ces critères seront utilisés dans ce travail pour effectuer la sélection du modèle et/ou pour comparer les différentes prévisions effectuées.

2.5. Le projet FLASH

L'importance des enjeux d'une meilleure compréhension, modélisation et prévision des crues éclair est telle que de nombreux projets de recherche se concentrent sur cet objectif. De manière non exhaustive, citons les projets ANR Floodscale qui étudie le phénomène de crue à diverses échelles spatiales et temporelles, PreDiFlood (<http://heberge.lcpc.fr/prediflood/>) qui s'intéresse à la prévision des submersions de routes dans le Gard, et ExtraFlo (<https://extraflo.cemagref.fr/>) qui vise à mieux appréhender la répétition des événements extrêmes. Citons également les projets européens PREVIEW (<http://www.preview-risk.com>), qui avait pour objet l'acquisition d'informations géographiques pour mieux prévoir les risques environnementaux dont le risque crue, IMPRINTS (<http://www.imprints-fp7.eu/>) dont l'objectif est de diminuer le coût en vie humaine et les dommages créés par les crues et les laves torrentielles, grâce à la définition de procédures et d'outils de gestion de crise et l'alerte précoce, et HyMex (<http://www.hymex.org/>) qui vise à approfondir la compréhension du cycle de l'eau en Méditerranée afin d'améliorer la prévision des risques hydrométéorologiques dont les crues éclair.

À ces projets s'ajoute le projet FLASH grâce auquel ce travail a pu se dérouler.

2.5.1. Objectifs

Le projet FLASH est un projet financé par l'agence Nationale de la Recherche, pour les années 2009-2013. Il part du constat d'un manque d'efficacité opérationnelle des modèles utilisés pour prévoir les crues éclair, comme nous l'avons souligné en section 2.4.2. Le projet FLASH va tenter de dépasser ces limites en mettant en œuvre un nouveau paradigme de construction de modèle, l'apprentissage statistique.

Comme l'apprentissage statistique ne requiert pas d'hypothèse ou de modèle physique pour construire une relation, on pourra représenter la chaîne hydrométéorologique, dans son ensemble, sous forme systémique et étudier séparément ou simultanément chacune des boîtes qui la composent. La chaîne hydrométéorologique comporte en entrée les différentes sources

d'acquisition de la mesure de précipitation : la pluie issue des pluviomètres, la pluie radar, la réflectivité radar. Sa sortie peut être le débit, variable décrivant l'état du système hydrologique, ou la hauteur d'eau, qui est en général la variable mesurée. On peut donc trouver autant de relations à identifier que de couples entrée-sortie. On peut également envisager à l'intérieur de la chaîne, des relations liant la hauteur au débit (courbe de tarage), et la réflectivité radar aux lames d'eau (pluviomètre ou radar) ; ce schéma est représenté en Figure 14.

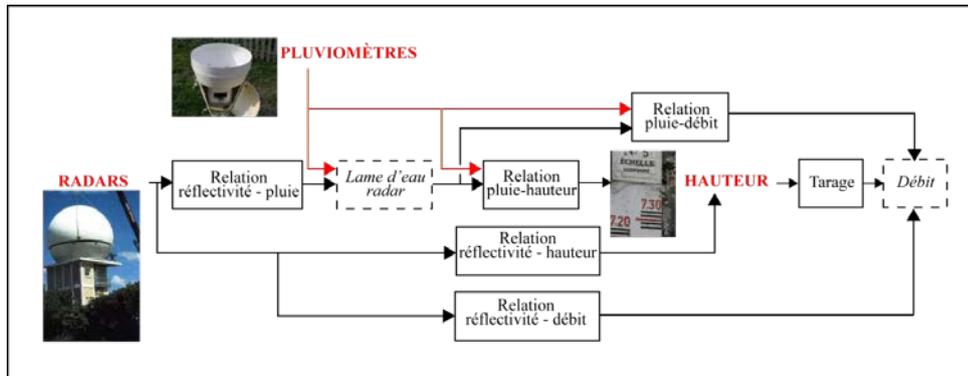


Figure 14 : Représentation systémique de la chaîne hydrométéorologique

L'objectif de notre travail est la représentation des fonctions de certaines des relations précédemment évoquées grâce aux réseaux de neurones formels. D'autres partenaires du projet s'intéressent à la réalisation de ces relations par une autre méthode d'apprentissage statistique : la régression par machines à vecteurs supports (SVR pour Support Vector Regression). De plus, une approche originale liant les modèles statistiques et hydrologiques sera également recherchée. Les bassins visés sont Le Gard et ses Gardons, la Cèze, l'Ardèche pour les crues rapides, et le bassin de la Somme pour les crues de nappe. Ces dernières sont intégrées au projet afin de valider les changements d'échelles tant spatiales que temporelles et pour appliquer une modélisation semi-physique (physique et statistique) au fonctionnement en prévision de la nappe souterraine alimentant la Somme.

2.5.2. Les partenaires

2.5.2.1. L'EMA

Deux laboratoires de l'EMA (Ecole des Mines d'Alès) participent au projet : le LGEI (Laboratoire de Génie de l'Environnement Industriel et des risques industriels et naturels) et le C2MA (Centre des Matériaux des Mines d'Alès). M^{me} Anne Johannet du C2MA coordonne le projet et participe à l'encadrement de cette thèse, en vue de la réalisation de modèles effectuant la prévision des crues éclair en utilisant des réseaux de neurones. Une autre thèse réalisée au C2MA par M. Guillaume Artigue s'intéresse à la généralisation des modèles neuronaux réalisés sur les Gardons à des bassins différents : ceux de la Cèze et de l'Ardèche. M. Bernard Vayssade est chargé de la prévision pour le bassin de la Somme. Enfin, M. Pierre-Alain Ayral du LGEI s'attache à l'étude sur le terrain de petits bassins versant expérimentaux situés dans le bassin versant du Gardon d'Anduze.

2.5.2.2. L'ESPCI ParisTech

L'axe de recherche du laboratoire SIGMA (SIGnaux, Modèles, Apprentissage statistique) de l'ESPCI ParisTech (Ecole Supérieure de Physique et de Chimie Industrielles de la ville de Paris) est l'apprentissage artificiel pour la modélisation des systèmes complexes, naturels ou artificiels. Outre la co-direction de cette thèse avec M. Gérard Dreyfus et M. Pierre Roussel, l'ESPCI participe au projet par la mise en place sur les Gardons d'un modèle semi-physique à base de machines à vecteurs supports *via* le post-doctorat de M. Khaled Boukharouba.

2.5.2.3. L'EDYTEM

L'EDYTEM (Environnements DYnamiques et TERRitoires de Montagne) est un laboratoire interdisciplinaire dépendant du CNRS et de l'Université de Savoie, qui travaille sur des problématiques liées à la géologie, l'hydrogéologie, la géographie et l'économie dans le cadre des milieux et territoires de montagne. L'EDYTEM, qui a déjà appliqué le modèle TOPMODEL aux crues rapides (section 4.2.4), participe au projet en s'intéressant à la modélisation semi-physique et à l'assimilation de données avec les travaux de M. Georges-Marie Saulnier et M. William Castaings.

2.5.2.4. Le SCHAPI

Concernant le projet FLASH, le SCHAPI est l'utilisateur final de la recherche ; il est chargé de la collecte des données utilisées pour les modélisations et réalisera l'outil informatique qui mettra en œuvre les méthodes développées dans le projet. Cet outil servira à alimenter la carte « *Vigicrues* ». M^{me} Caroline Wittwer, M. Arthur Marchandise et M. Bruno Janet suivent les avancées du projet.

2.6. La base de données

L'instrumentation du bassin versant du Gardon d'Anduze est composée de 6 pluviomètres (Figure 13) : Barre-des-Cévennes, Saint-Roman-de-Tousque, Saumane, Soudorgues, Mialet et Anduze, et de 3 stations de mesure limnimétrique à Saumane, Mialet et Anduze. Une station a été installée à Saint-Jean-du-Gard depuis 2003, mais comme des événements antérieurs sont pris en compte, on ne pourra pas l'utiliser dans une base de données commune.

N°	Date	Durée (h)	Pluie moyenne cumulée (mm)	Hauteur maximum (m)	Niveau de vigilance [SPCGD 2010]
1	22-24 Septembre 1994	35	189	3,71	vert
3	3-6 Octobre 1995	53	319	5,34	jaune
4	13-15 Octobre 1995	38	194	5	jaune
8	10-13 Novembre 1996	81	157	2,71	vert
107	4-7 Novembre 1997	89	273	4,20	jaune
108	24-28 Novembre 1997	81	107	2,58	vert
109	16-20 Décembre 1997	103	376	5,37	jaune
117	20-21 Octobre 1999	33	117	3,64	vert
13	28-30 Septembre 2000	45	184	4,80	jaune
120	12-15 Novembre 2000	70	125	2,77	vert
19	8-9 Septembre 2002	35	406	9,71	rouge
22	24-25 Septembre 2006	37	61	2,24	vert
23	18-20 Octobre 2006	54	198	6,61	orange
24	16-18 Novembre 2006	33	59	2,75	vert
25	20-23 Novembre 2007	70	124	2,69	vert
26	21-23 Octobre 2008	51	173	5,57	jaune
27	31 Octobre – 3 Novembre 2008	81	272	5,53	jaune

Tableau 1 : Les événements de la base de données

La base de données (Tableau 1) contient 17 événements de crues éclair ; la numérotation a été établie par [Ayrat 2005] jusqu'au numéro 19. Les crues de moindres importances ajoutées chronologiquement entre les événements précédents par [Toukourou 2009.b] sont numérotées à partir de 100. Ces événements composent des configurations variées : distribution des pluies, durée, amplitude de hauteurs différentes et la réponse à la pluie peut donner un événement mono-pic ou un événement multi-pics, les limnigrammes et hyétogrammes se trouvent en Annexe 2.

Événement	intense	long	Multi-pic
1			X
3	X	X	X
4	X		
8		X	X
107		X	X
108		X	X
109	X	X	
117			
13	X		
120		X	X
19	X		
22			
23	X	X	
24			
25		X	X
26	X		
27	X	X	X

Tableau 2 : Caractérisation des événements

Afin de comparer les événements, on considère en particulier leurs durées et leurs intensités (Tableau 2). Les hauteurs maximales des événements sont comprises entre 2,71 et 9,71m ; on note que le deuxième événement le plus intense a une hauteur maximale de 6,61 m, qui correspond à une vigilance orange, n’atteignant que 68% de celle de l’événement le plus intense. On considère qu’un événement est intense si la hauteur maximale est plus grande que la moitié de la hauteur maximale de l’événement le plus intense : l’événement 19 est le plus intense et seul à correspondre à une vigilance rouge, 8 événements sont donc considérés comme intenses.

Par ailleurs, la durée des événements s’étend de 33 à 103 h. Un événement sera qualifié de long si sa durée est supérieure à la moitié de celle de l’événement le plus long. L’événement le plus long est le 109, 9 événements sont ainsi considérés comme longs.

Enfin un événement peut posséder un seul pic ou plusieurs pics.

2.7. Conclusion

Après avoir présenté, dans le Chapitre 1, l’apprentissage statistique et sa mise en œuvre pour la modélisation de systèmes dynamiques non linéaires, nous nous sommes penchés dans ce chapitre sur le processus qui nous a intéressé dans ce travail : la création d’une crue rapide à partir de pluies intenses. Nous avons tout d’abord présenté les crues rapides, puis le théâtre de leurs débordements : les Gardons Cévenols, rivière éponyme des célèbres *Gardonnades*.

Nous nous sommes ensuite attachés à présenter les spécificités de ces bassins rapides caractérisés par des orages très violents et des chutes de pluies très hétérogènes tant

spatialement que temporellement. De ce fait, l'acquisition des mesures doit être réalisée de manière suffisamment fréquente et rapprochée, ce qui malheureusement n'a pas pu toujours être le cas dans le passé, même si ce défaut tend à être corrigé par l'introduction de nouvelles stations de mesures et par la mise en œuvre de la mesure de champs de pluie par les radars météorologiques.

Malgré les efforts consacrés à l'acquisition des mesures, nous avons pu constater que les erreurs et incertitudes restaient considérables et soulevaient un problème majeur pour appliquer l'apprentissage statistique. Nous avons en effet rappelé dans le Chapitre 1 que la précision des modèles conçus par apprentissage statistique est limitée par la présence de bruit de mesure et de perturbations non mesurées, ainsi que par la taille de la base de données disponible.

Nous avons ensuite présenté les modèles hydrologiques qui visent à simuler le fonctionnement des crues éclair, ainsi que les travaux menés en ce sens sur les Gardons. Ces travaux seront très utiles pour définir les variables des modèles que nous mettrons en œuvre et pour mieux comprendre leurs performances. Enfin nous avons présenté le projet qui nous a permis de mener à bien ce travail : le projet ANR FLASH, ainsi que la base de données des événements disponibles. Nous envisagerons dans le prochain chapitre la mise en œuvre de l'apprentissage statistique sur une telle base de données, afin de réaliser des modèles de prévision dédiés au Gardon d'Anduze.

3. Prévision des crues éclair sur le bassin versant du Gardon d'Anduze

3.1. Introduction

Dans le premier chapitre, la modélisation par réseaux de neurones a été présentée ; nous avons insisté notamment sur le problème – général pour tout modèle conçu par apprentissage statistique – du dilemme biais-variance, et sur les propriétés d'approximation universelle et de parcimonie que possèdent les réseaux de neurones. Le deuxième chapitre, consacré aux crues éclair sur le Gardon d'Anduze, a montré les difficultés rencontrées lors de la modélisation de ces phénomènes en raison de leur complexité et des incertitudes sur les données.

Ce chapitre est consacré à la prévision des crues éclairs du bassin versant du Gardon d'Anduze avec des réseaux de neurones. Tout d'abord, nous présenterons le type de modèle choisi et les améliorations que l'on souhaite y apporter.

Ensuite, à l'aide de la méthode de validation croisée présentée au Chapitre 1, nous sélectionnerons les différents hyperparamètres, tant relatifs à l'architecture qu'à l'algorithme d'apprentissage du réseau, afin d'optimiser les modèles. Nous évaluerons la pertinence des modèles ainsi définis, pour différents horizons de prévision, sur trois événements majeurs en test (événements qui n'ont été utilisés ni pour l'apprentissage ni pour la sélection de modèles ou d'hyperparamètres). La qualité des prévisions des modèles se révélant très sensible à l'initialisation des paramètres, nous montrerons que l'intégration d'une partie linéaire au perceptron multicouche permet de diminuer la variabilité par rapport à l'initialisation des paramètres.

Enfin, dans l'objectif de l'étude systémique de la chaîne hydrométéorologique, nous introduirons les données radar sous forme de lame d'eau, comme variable des modèles. Nous comparerons les résultats obtenus avec les mesures de pluie provenant des pluviomètres et celles obtenues avec le radar.

3.2. Utilisation des réseaux de neurones pour la prévision de la hauteur d'eau à Anduze

3.2.1. Type de modèle utilisé

Rappelons que pour effectuer la prévision des hauteurs d'eau, on se place à l'instant k , et l'on souhaite prévoir la hauteur à Anduze à l'instant $k+h_p$, h_p étant l'horizon de prévision. On dispose des valeurs des pluies observées aux 6 pluviomètres présents sur le bassin versant (cf. Chapitre 2) jusqu'au temps k , ainsi que de la hauteur d'eau mesurée au limnimètre de la station d'Anduze jusqu'au temps k . On réalisera des modèles établissant des prévisions et non des simulations, la différence entre ces deux familles de modèles est illustrée sur la Figure 15.

Si l'on désire prévoir les crues à Anduze à l'aide du modèle de simulation du bassin versant, on devra disposer des précipitations futures (par rapport à l'instant d'émission de la prévision : l'instant k , et utiliser le modèle de simulation fonctionnant dans le futur pour estimer le débit à $k+h_p$, connaissant les pluies prévues jusqu'à $k+h_p$, et les hauteurs d'eau estimées par le modèles jusqu'à $k+h_p-1$. Si l'on réalise cette même prévision avec le modèle

de prévision, ce dernier estime la hauteur d'eau à $k+h_p$ en fonction des pluies mesurées jusqu'à k et des hauteurs d'eau mesurées ou estimées jusqu'à k . Il manque donc de l'information au modèle entre les instant k et $k+h_p-1$ et l'on peut penser que lors de son apprentissage le modèle aura appris à remédier à ce manque d'information en réalisant de manière implicite une anticipation sur les pluies.

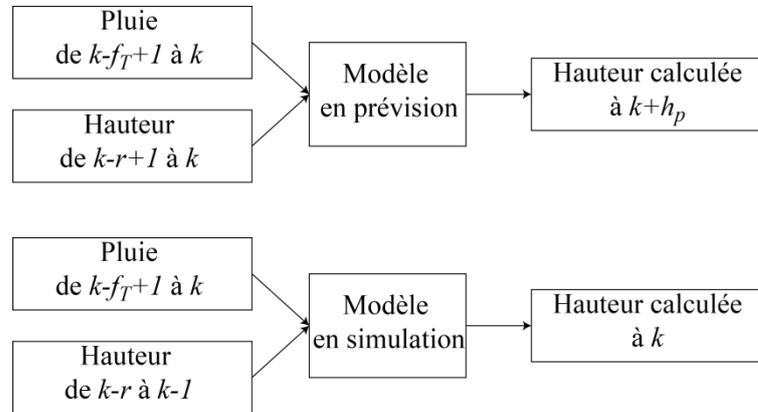


Figure 15 : Différence entre modèles utilisés en prévision et en simulation avec k l'instant présent, f_T la fenêtre temporelle des pluies, r la fenêtre temporelle des hauteurs et h_p l'horizon de prévision.

Le prédicteur retenu (cf. Chapitre 1) est celui qui est associé à l'hypothèse de bruit d'état. Ce modèle utilise les valeurs précédentes observées de la sortie qu'il doit estimer : l'apprentissage du réseau est dirigé. Cela évite les divergences que pourrait subir un modèle associé à l'hypothèse de bruit de sortie qui calcule sa sortie à partir des valeurs précédentes de celle-ci. Ces valeurs calculées étant soumises à des erreurs, les suivantes risquent de s'éloigner encore plus des sorties observées.

Le modèle réalise la fonction suivante :

$$y(k+h_p) = \varphi_{RN}(y^p(k), \dots, y^p(k-r+1), \mathbf{u}(k), \dots, \mathbf{u}(k-f_T+1), \mathbf{w})$$

avec :

- $y(k+h_p)$ la sortie calculée par le modèle : la hauteur prévue à Anduze pour l'instant $k+h_p$
- h_p l'horizon de prévision
- $y^p(k)$ la sortie observée (la hauteur d'eau à Anduze) à l'instant k
- f_T la fenêtre temporelle des pluies
- r la fenêtre temporelle des hauteurs
- \mathbf{u} le vecteur les variables exogènes : les pluies mesurées aux 6 pluviomètres

Le réseau de neurones utilisé est un perceptron multicouche comportant une couche de neurones cachés avec des tangentes hyperboliques comme fonction d'activation et un neurone de sortie linéaire. On reprend le type de modèle mis en œuvre par M. Toukourou dans sa thèse (Figure 16).

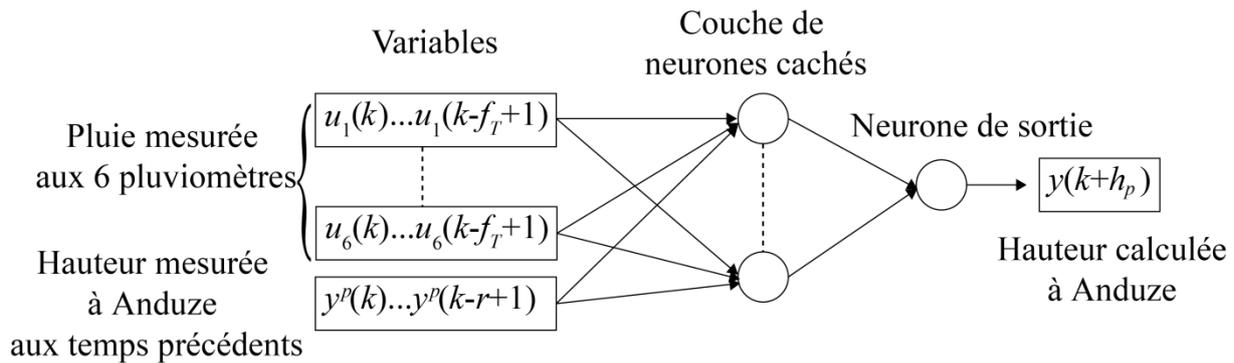


Figure 16 : Schéma du type de réseau utilisé

Il y a un réseau de neurones différent pour chaque horizon de prévision pour tenir compte du fait que l'anticipation demandée est différente pour chaque horizon. On réalisera ainsi 6 réseaux pour les horizons de prévision suivants : 30min, 1h, 2h, 3h, 4h, et 5h.

3.2.2. Amélioration du modèle de base

Nous avons pour but d'améliorer le modèle de base (Figure 16). Nous n'introduisons pas de nouvelles variables : selon [Toukourou 2009.a] l'utilisation du cumul des précipitations n'apporte pas d'information pertinente au modèle ; pour ce qui est de l'humidité des sols, la période d'échantillonnage de 24h des données ISBA, Interaction Sol Biosphère Atmosphère (développé par le Centre National de Recherches Météorologiques), n'est pas compatible avec la période de 30 min utilisée pour les mesures des autres grandeurs.

Comme nous l'avons vu dans le Chapitre 1, la complexité du modèle joue un rôle important dans le dilemme biais-variance : un modèle de faible complexité donne de mauvais résultats en apprentissage et par conséquent en généralisation, alors qu'un modèle trop complexe est trop ajusté à l'ensemble d'apprentissage et ne peut pas généraliser de manière satisfaisante. Il convient de trouver un compromis : nous allons donc chercher à optimiser la complexité du modèle, qui, pour les réseaux de neurones, est liée essentiellement au nombre de neurones cachés et au nombre de variables. Nous chercherons donc à optimiser ces deux facteurs. Puis nous étudierons l'influence de deux hyperparamètres de l'algorithme d'apprentissage sur la capacité de généralisation : l'initialisation du pas d'apprentissage de l'algorithme de Levenberg-Marquardt et le facteur décroissance de ce pas au fil des itérations. Enfin nous nous efforcerons de trouver la meilleure initialisation possible des paramètres du réseau. Nous comparerons également la sensibilité des modèles à ces hyperparamètres.

Pour réaliser cela, nous sélectionnerons pour chaque horizon les modèles les plus performants à l'aide de la méthode de validation croisée partielle expliquée au Chapitre 1. Les événements de la base de données sont répartis comme suit dans les différents ensembles :

- l'ensemble d'apprentissage comprend tous les événements de la base excepté les événements 13, 19, 26 et 27,
- l'événement 13 est choisi pour l'arrêt précoce car c'est l'événement qui donne les meilleurs résultats lorsqu'il est utilisé en test. Il est donc bien représentatif de l'ensemble de la base d'apprentissage et il y a une cohérence certaine à le choisir pour arrêter l'apprentissage,

- les événements 19, 26 et 27 sont réservés pour le test car ce sont des événements intenses. D’ailleurs, les événements 26 et 27 ont été choisis comme événements tests par le SCHAPI dans le projet BVNE (Chapitre 2,2.4). Nous avons également retenu l’événement 19 car il s’agit de l’événement le plus intense de la base de données. Cependant les mesures ont été en grande partie reconstruites *a posteriori* : les erreurs sur les valeurs de hauteur sont donc probablement très importantes,

- les événements servant d’ensemble de « validation » pour la validation croisée partielle sont les événements 3, 4, 109 et 23, événements parmi les plus intenses de la base d’apprentissage puisque que l’on cherche à sélectionner les modèles qui prévoient le mieux les événements les plus intenses. Nous n’effectuerons ainsi pas une validation croisée complète : nous calculerons le score de validation seulement sur ces événements intenses. Cette procédure est appelée *validation croisée partielle*.

3.3. Sélection de modèles

Comme indiqué ci-dessus nous allons sélectionner les modèles en nous intéressant successivement au nombre de neurones cachés, à la fenêtre temporelle des pluies, au pas d’apprentissage et à son facteur de décroissance, puis à initialisation des paramètres du réseau.

3.3.1. Sélection du nombre de neurones cachés

3.3.1.1. Méthode de sélection

Afin d’obtenir la complexité optimale pour le réseau de neurones, nous allons faire varier le nombre de neurones cachés pour chaque horizon de prévision. Le nombre de neurones cachés sera choisi grâce à la méthode de validation croisée partielle.

Pour chaque horizon de prévision, on fait varier le nombre de neurones cachés, N_c , de 1 à 7 tout en fixant la fenêtre temporelle des pluies et les valeurs des hyperparamètres de l’algorithme d’apprentissage avec des valeurs établies lors de précédentes simulations [Bornancin Plantier 2011]. La fenêtre temporelle dépend de l’horizon de prévision comme indiqué dans le Tableau 3, le pas d’apprentissage est initialisé à 10^{-4} , son facteur de décroissance de 0,9.

Horizon de prévision	30 min	1h	2h	3h	4h	5h
Fenêtre temporelle	2h30min	3h	3h	2h	30min	30min

Tableau 3 : Dimension de la fenêtre temporelle utilisée pour la sélection de l’architecture des différents modèles

Pour chaque valeur de N_c , on calcule le score de validation croisée partielle, S_{vcp} , (cf. Chapitre 1, 1.8) pour 200 modèles obtenus pour 200 différentes initialisations des paramètres du réseau. On obtient ainsi 200 S_{vcp} dont on calcule la moyenne, la médiane, le minimum, le maximum, l’écart-type (les valeurs du maximum et de l’écart type sont consignées en Annexe

3), ainsi que la moyenne des 20 meilleurs S_{vcp} sur les 200 disponibles, $M_{10\%}$, soit la moyenne sur le premier décile des S_{vcp} .

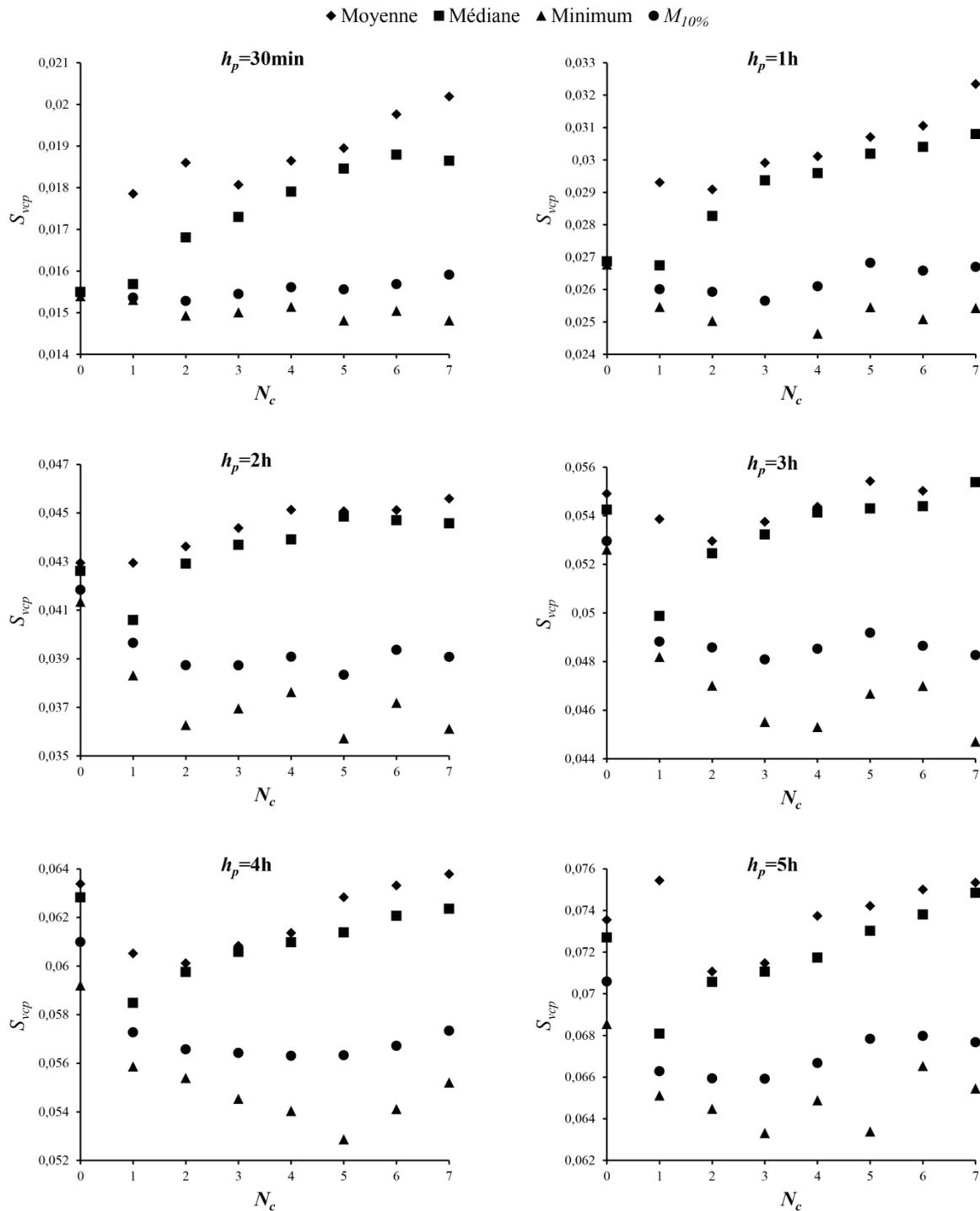


Figure 17 : Pour chaque nombre de neurones cachés* : moyenne, médiane et minimum sur les 200 scores de validation croisée partielle et moyenne des 20 meilleurs scores de validation croisée partielle, chaque graphique correspond à un horizon de prévision

* $N_c=0$ correspond à un réseau ayant un seul neurone caché linéaire, si les S_{vcp} n'ont pas la même valeur pour un même horizon de prévision, cela est dû à l'arrêt précoce qui interrompt l'apprentissage avant la convergence vers la solution unique du modèle linéaire

Sur les graphiques de la Figure 17, on remarque que les différentes grandeurs calculées ne présentent pas le même minimum et ne conduisent donc pas à la sélection d'un même nombre de neurones cachés: il faut donc choisir l'une d'entre elles pour pouvoir construire un modèle.

La moyenne est très influencée par le maximum des S_{vcp} qui peut atteindre une valeur dix fois supérieure à cette dernière : la variance des scores est telle que la différence entre les scores pour les différentes complexités est peu significative. La médiane des scores, quant à elle désigne, pour chaque horizon de prévision, les modèles ayant un seul neurone caché : elle favorise les réseaux à faible complexité. Ces réseaux ayant un plus petit nombre de paramètres, on peut supposer que l'ajustement de ces paramètres est moins complexe, ce qui les conduit à des valeurs finales plus proches et a pour conséquence des valeurs de médianes de S_{vcp} plus petites. Enfin, fonder le choix du modèle sur la comparaison entre les scores de validation croisée minimums, obtenus pour chaque complexité, favorise les réseaux de plus grande complexité, néanmoins, compte tenu de la variance des scores obtenus, une telle démarche peut être hasardeuse. Nous avons donc opté pour un compromis et nous avons effectué la sélection du nombre de neurones cachés à partir de la moyenne des 20 plus petits S_{vcp} , $M_{10\%}$, favorisant ainsi les meilleurs résultats tout en assurant une certaine robustesse.

On note que les réseaux de neurones ne possédant pas de neurones avec fonction d'activation non linéaire ($N_c=0$), ne sont jamais sélectionnés, et la différence de S_{vcp} par rapport aux modèles avec neurones cachés à fonction sigmoïde devient de plus en plus marquée lorsque l'horizon de prévision augmente. Les modèles linéaires semblent donc moins bien adaptés à la tâche que nous cherchons à réaliser, compte tenu des données dont nous disposons, que les modèles non linéaires.

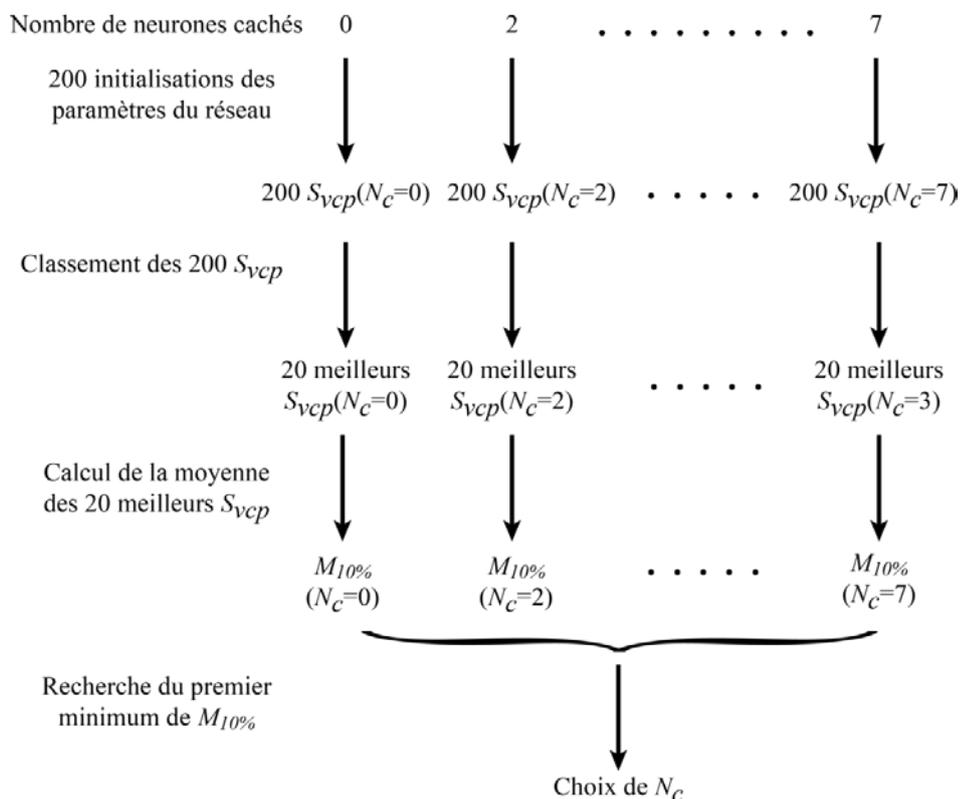


Figure 18 : Méthode de sélection du nombre de neurones cachés

Cette méthode, schématisée Figure 18, conduit aux résultats présentés dans le Tableau 4.

Horizon de prévision	30 min	1h	2h	3h	4h	5h
Nombre de neurones cachés	2	3	2	3	4	3

Tableau 4 : Nombre de neurones cachés en fonction de l'horizon de prévision

Le nombre de neurones cachés n'est donc pas corrélé avec l'horizon de prévision. La difficulté de modélisation augmente logiquement lorsque la prévision devient plus lointaine, on pourrait s'attendre à ce qu'en conséquence les réseaux soient plus complexes pour les horizons lointains. Cependant, comme les réseaux ne disposent pas de toutes les informations nécessaires à une bonne qualité de ces prévisions lointaines, la complexité des réseaux n'évolue pas nécessairement de la même façon que la complexité des processus à modéliser.

3.3.1.2. Sensibilité du score de validation croisée partielle au nombre de neurones cachés

Nous allons maintenant déterminer l'influence relative de la variation du nombre de neurones cachés sur la capacité de généralisation du modèle. Pour effectuer cette analyse de sensibilité, nous introduisons deux critères, S_1 et S_2 , qui quantifient la dispersion des résultats obtenus pour les différentes complexités :

$$S_1 = \frac{\text{Maximum}(S_{vcp}) - \text{Minimum}(S_{vcp})}{\text{Moyenne}(S_{vcp})}$$

$$S_2 = \frac{\text{Ecart-type}(S_{vcp})}{\text{Moyenne}(S_{vcp})}$$

Pour chaque horizon de prévision, on étudie la dispersion de la moyenne des 20 meilleurs S_{vcp} obtenus, $M_{10\%}$, qui est le critère utilisé pour la sélection de la complexité du modèle. On calcule la moyenne, le minimum, le maximum, l'écart-type et les critères S_1 et S_2 ; des 8 valeurs de $M_{10\%}$ (correspondant aux 8 valeurs du nombre de neurones cachés). Ces valeurs sont reportées dans le Tableau 5.

$M_{10\%}$	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
Moyenne	$1,56.10^{-2}$	$2,63.10^{-2}$	$3,90.10^{-2}$	$4,86.10^{-2}$	$5,67.10^{-2}$	$6,69.10^{-2}$
Minimum	$1,53.10^{-2}$	$2,57.10^{-2}$	$3,83.10^{-2}$	$4,81.10^{-2}$	$5,63.10^{-2}$	$6,59.10^{-2}$
Maximum	$1,59.10^{-2}$	$2,68.10^{-2}$	$3,96.10^{-2}$	$4,92.10^{-2}$	$5,73.10^{-2}$	$6,80.10^{-2}$
Écart-type	$2,13.10^{-4}$	$4,44.10^{-4}$	$4,39.10^{-4}$	$3,60.10^{-4}$	$4,32.10^{-4}$	$9,06.10^{-4}$
S_1	$4,1.10^{-2}$	$4,5.10^{-2}$	$3,4.10^{-2}$	$2,3.10^{-2}$	$1,8.10^{-2}$	$3,1.10^{-2}$
S_2	$1,4.10^{-2}$	$1,7.10^{-2}$	$1,1.10^{-2}$	$7,4.10^{-3}$	$7,6.10^{-3}$	$1,4.10^{-2}$

Tableau 5 : Pour chaque horizon de prévision, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de $M_{10\%}$.

Le Tableau 5 montre que les valeurs de $M_{10\%}$ varient peu en fonction du nombre de neurones cachés, comme on pouvait l'observer sur la Figure 17, et donc que le minimum est peu marqué. On en déduit que les scores de validation croisée partielle, et par leur estimation, la capacité de généralisation du modèle, sont peu sensibles au nombre de neurones cachés. Cela peut en partie s'expliquer par l'utilisation de l'arrêt précoce comme méthode de régularisation qui réduit efficacement les effets du surajustement [Kong A Siou 2012.a].

Les valeurs de moyenne, de minimum et de maximum augmentent avec l'horizon de prévision ce qui confirme que la difficulté de modélisation devient plus grande lorsque la prévision est plus éloignée et que l'on ne dispose plus de données de pluie et de hauteur temporellement proches de la hauteur à prévoir.

Les critères de dispersion S_1 et S_2 ainsi que l'écart-type ne suivent pas de tendance par rapport à l'horizon de prévision : la sensibilité du S_{vcp} au nombre de neurones cachés ne semble pas liée à la difficulté de modélisation.

Il est intéressant de comparer les variations de $M_{10\%}$ en fonction de N_c à celles des S_{vcp} obtenus pour les 200 initialisations des paramètres du réseau utilisées. Pour le nombre de neurones cachés choisi au cours de l'étape précédente, on étudie l'influence de l'initialisation des paramètres du réseau sur la dispersion des S_{vcp} obtenus.

S_{vcp}	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
N_c	2	3	2	3	4	3
Moyenne	$1,86.10^{-2}$	$2,99.10^{-2}$	$4,36.10^{-2}$	$5,37.10^{-2}$	$6,14.10^{-2}$	$7,15.10^{-2}$
Minimum	$1,49.10^{-2}$	$2,37.10^{-2}$	$3,63.10^{-2}$	$4,55.10^{-2}$	$5,40.10^{-2}$	$6,33.10^{-2}$
Maximum	$1,45.10^{-1}$	$4,58.10^{-2}$	$5,55.10^{-2}$	$7,52.10^{-2}$	$8,65.10^{-2}$	$8,38.10^{-2}$
Écart-type	$1,01.10^{-2}$	$3,42.10^{-3}$	$3,61.10^{-3}$	$4,00.10^{-3}$	$3,85.10^{-3}$	$3,64.10^{-3}$
S_1	7,0	$7,4.10^{-1}$	$4,4.10^{-1}$	$5,5.10^{-1}$	$5,3.10^{-1}$	$2,9.10^{-1}$
S_2	$5,4.10^{-1}$	$1,1.10^{-1}$	$8,3.10^{-2}$	$7,4.10^{-2}$	$6,3.10^{-2}$	$5,1.10^{-2}$

Tableau 6 : Pour chaque horizon de prévision et pour le nombre de neurones cachés déterminé, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de S_{vcp} sur 200 modèles obtenus avec 200 initialisations différentes des paramètres du réseau.

Le Tableau 6 illustre la sensibilité du score de validation croisée partielle à l'initialisation des paramètres du réseau et permet de la comparer à la sensibilité au nombre de neurones cachés. Il faut cependant noter que la comparaison avec le Tableau 5 n'est pas entièrement rigoureuse car les descripteurs statistiques ne sont pas calculés sur une population de même taille dans les deux tableaux, cependant on peut en tirer quelques remarques.

La dispersion des valeurs de S_{vcp} est importante : le maximum peut atteindre dix fois la valeur de la moyenne et les critères S_1 et S_2 ont des valeurs de dix à cent fois supérieures à celles du Tableau 5. Le score de validation croisée partielle, et par conséquent la capacité de généralisation du modèle, sont ainsi très sensibles à l'initialisation des paramètres du réseau.

Par ailleurs, les critères de dispersion tendent à décroître lorsque l'horizon de prévision augmente. La variabilité de l'estimation de la capacité de généralisation par validation croisée décroît lorsque l'horizon de prédiction augmente.

3.3.2. *Sélection de la fenêtre temporelle des pluies*

3.3.2.1. *Méthode de sélection*

La complexité du réseau étant déterminée, on étudie le nombre de variables, c'est-à-dire la dimension de la fenêtre temporelle des pluies. Il aurait été intéressant d'estimer une fenêtre temporelle propre à chaque pluviomètre, car on pourrait supposer que les pluviomètres plus éloignés de l'exutoire ne fournissent pas l'information la plus pertinente au même instant que ceux qui en sont situés plus proches. Cependant, l'importante variabilité spatiale et temporelle des pluies d'un événement à l'autre ne permet pas de tirer de conclusions pertinentes sur les fenêtres respectives des 6 pluviomètres, comme le confirment quelques essais de sélections de fenêtres temporelles différentes. On se contentera ici d'imposer la même fenêtre temporelle pour tous les pluviomètres.

Pour chaque horizon de prévision, on fait varier cette fenêtre temporelle, f_T , de 30min à 5h30min, de demi-heure en demi-heure. Le nombre de neurones cachés est celui figurant dans le Tableau 6, le pas d'apprentissage est initialisé à 10^{-4} et le facteur de décroissance à 0,9.

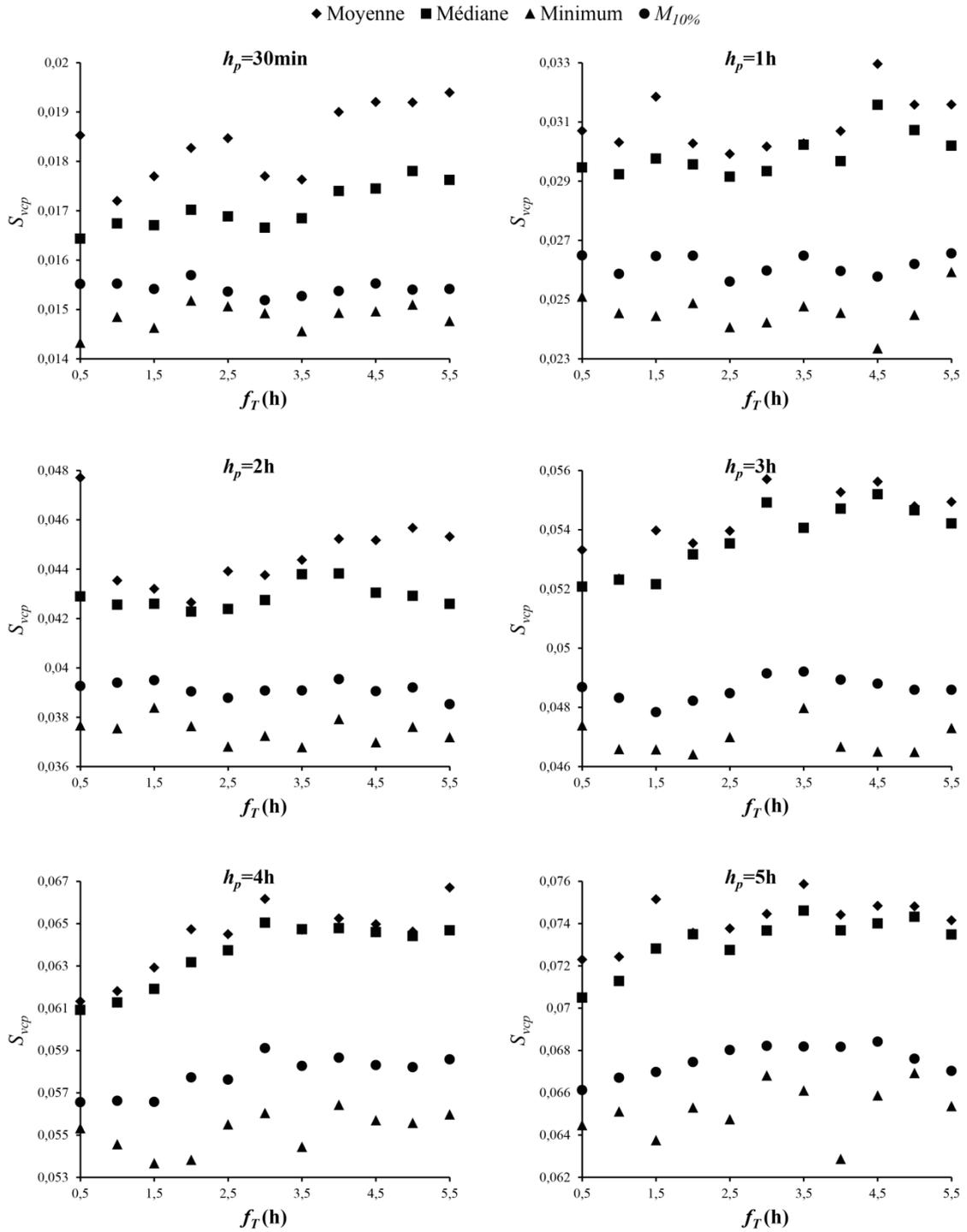


Figure 19: Pour chaque dimension de fenêtre temporelle : moyenne, médiane et minimum sur les 200 scores de validation croisée partielle et moyenne des 20 meilleurs scores de validation croisée partielle : chaque graphique correspond à un horizon de prévision

Comme à l'étape précédente, pour chaque valeur de f_T , on réalise l'apprentissage de 200 modèles avec 200 initialisations des paramètres du réseau différentes. Les score de validation croisée partielle, S_{vcp} , sont calculés. Sur chaque série de 200 scores (Figure 20), on calcule la moyenne, la médiane, le minimum, le maximum, l'écart-type (les valeurs du maximum et de l'écart type sont consignées en Annexe 3), mais on utilisera $M_{10\%}$, la moyenne des 20 plus

petits scores de validation croisée pour chaque f_T , (Figure 19) pour choisir la fenêtre temporelle la plus judicieuse, comme précédemment.

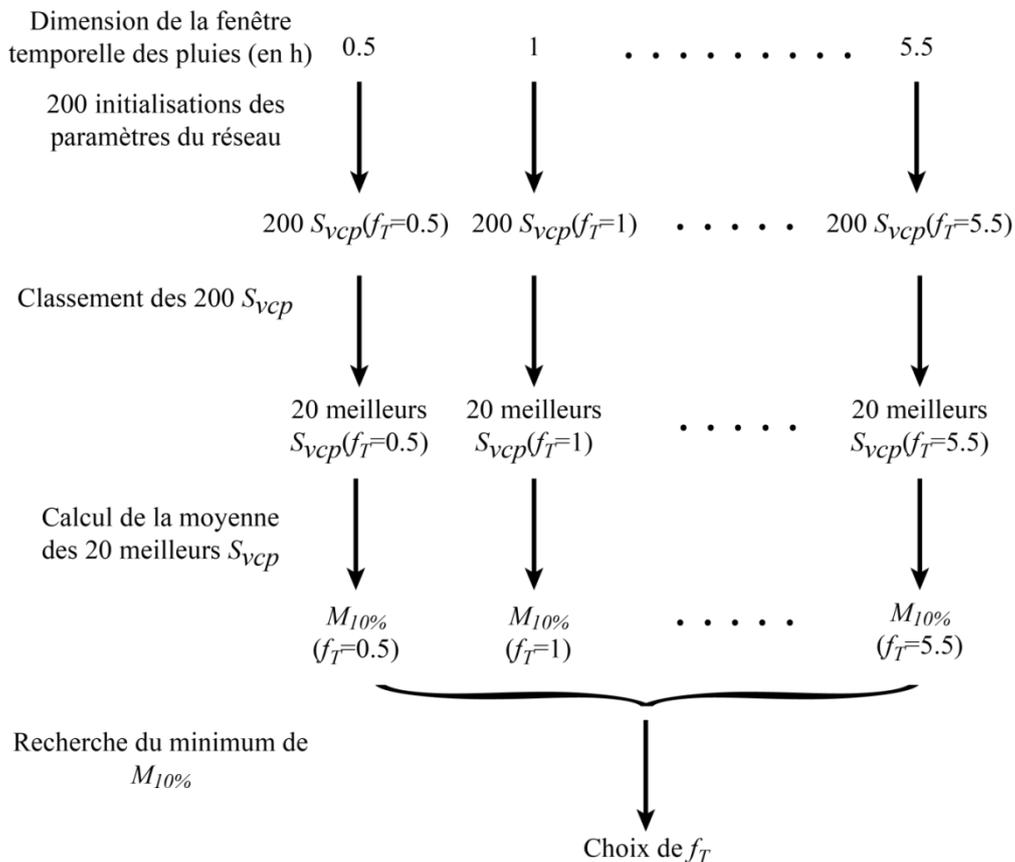


Figure 20: Schéma de la méthode de sélection de la fenêtre temporelle

Les valeurs des fenêtres temporelles, estimées selon la méthode précédente, sont consignées dans le Tableau 7.

Horizon de prévision	30 min	1h	2h	3h	4h	5h
Fenêtre temporelle des pluies	3h	2h30min	2h30min	1h30min	30min	30min

Tableau 7 : Dimension de la fenêtre temporelle des pluies en fonction de l'horizon de prévision

Contrairement au nombre de neurones cachés, la dimension des fenêtres temporelles des pluies sélectionnées est corrélée à l'horizon de prévision : plus l'horizon de prévision est grand, plus la fenêtre temporelle est réduite. Cette constatation est cohérente avec la physique du phénomène. En effet, lorsque l'horizon de prédiction augmente, les données de pluie proches de l'instant de prévision ne sont pas disponibles. Elles sont remplacées en entrée par des pluies de plus en plus anciennes par rapport à la prévision et elles deviennent donc non pertinentes.

3.3.2.2. Sensibilité du score de validation croisée partielle à la dimension de la fenêtre temporelle des pluies

Comme pour le nombre de neurones cachés, nous allons déterminer l'influence de la variation de la fenêtre temporelle des pluies en se basant sur la dispersion de $M_{10\%}$, la moyenne des 20 plus petits S_{vcp} , pour les 10 valeurs de f_T .

$M_{10\%}$	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
Moyenne	$1,54.10^{-2}$	$2,62.10^{-2}$	$3,91.10^{-2}$	$4,86.10^{-2}$	$5,78.10^{-2}$	$6,75.10^{-2}$
Minimum	$1,52.10^{-2}$	$2,56.10^{-2}$	$3,85.10^{-2}$	$4,78.10^{-2}$	$5,66.10^{-2}$	$6,61.10^{-2}$
Maximum	$1,57.10^{-2}$	$2,66.10^{-2}$	$3,95.10^{-2}$	$4,92.10^{-2}$	$5,91.10^{-2}$	$6,84.10^{-2}$
Écart-type	$1,38.10^{-4}$	$3,44.10^{-4}$	$2,99.10^{-4}$	$4,04.10^{-4}$	$9,04.10^{-4}$	$7,44.10^{-4}$
S_1	$3,3.10^{-2}$	$3,6.10^{-2}$	$2,6.10^{-2}$	$2,8.10^{-2}$	$4,4.10^{-2}$	$3,4.10^{-2}$
S_2	$8,9.10^{-3}$	$1,3.10^{-2}$	$7,7.10^{-3}$	$8,3.10^{-3}$	$1,6.10^{-2}$	$1,1.10^{-2}$

Tableau 8 : Pour chaque horizon de prévision, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de $M_{10\%}$.

Si l'on compare le Tableau 8 et le Tableau 9, on trouve pour la fenêtre temporelle des valeurs d'écart type, de S_1 et S_2 semblables à celles obtenues pour différents nombres de neurones cachés. Les scores de validation croisée partielle des modèles présentent une sensibilité équivalente et faible à ces deux hyperparamètres. Le choix de la fenêtre temporelle a, tout comme le choix du nombre de neurones cachés, peu d'influence sur l'estimation de la capacité de généralisation des modèles.

Les valeurs des écarts types de S_1 et S_2 ne sont pas, non plus, corrélés avec l'horizon de prévision, donc la sensibilité de l'estimation, par validation croisée partielle, de la qualité de prévision des modèles à la dimension de la fenêtre temporelle des pluies n'a pas un lien direct avec la complexité de la prévision.

Pour les dimensions de fenêtre temporelles retenues, nous regardons l'influence de l'initialisation des paramètres du réseau sur la capacité de généralisation des modèles.

S_{vcp}	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
f_t	3	2,5	2,5	1,5	0,5	0,5
Moyenne	$1,77.10^{-2}$	$2,99.10^{-2}$	$4,39.10^{-2}$	$5,57.10^{-2}$	$6,13.10^{-2}$	$7,23.10^{-2}$
Minimum	$1,49.10^{-2}$	$2,41.10^{-2}$	$3,68.10^{-2}$	$4,60.10^{-2}$	$5,53.10^{-2}$	$6,45.10^{-2}$
Maximum	$7,18.10^{-2}$	$7,26.10^{-2}$	$1,47.10^{-1}$	$9,76.10^{-2}$	$7,52.10^{-2}$	$1,69.10^{-1}$
Écart-type	$4,72.10^{-3}$	$4,47.10^{-3}$	$8,93.10^{-3}$	$5,52.10^{-3}$	$3,53.10^{-3}$	$1,09.10^{-2}$
S_1	3,2	1,6	2,5	$9,3.10^{-1}$	$3,2.10^{-1}$	1,4
S_2	$2,7.10^{-1}$	$1,5.10^{-1}$	$2,0.10^{-1}$	$9,9.10^{-2}$	$5,8.10^{-2}$	$1,5.10^{-1}$

Tableau 9 : Pour chaque horizon de prévision et pour la dimension de fenêtre temporelle déterminée, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de S_{vcp} sur 200 modèles obtenus avec 200 initialisations différentes des paramètres du réseau.

Le Tableau 9 pour la dimension de la fenêtre temporelle donne des conclusions semblables à celles du Tableau 6 concernant le nombre de neurones cachés. Les critères de dispersion S_1 et S_2 sur l'initialisation des paramètres sont de dix à cent fois supérieurs à ceux

sur la dimension de la fenêtre temporelle : pour une dimension donnée de la fenêtre temporelle. La variabilité des scores de validation croisée partielle à l'initialisation des paramètres des réseaux est plus grande que la variabilité que l'on observe en faisant varier la dimension de la fenêtre temporelle.

3.3.3. Sélection de l'initialisation des paramètres du réseau et ajustement de l'algorithme d'apprentissage

3.3.3.1. Sélection de l'initialisation du pas d'apprentissage et de son facteur de décroissance

L'architecture du réseau et le nombre de variables étant établis, nous passons à l'ajustement de l'initialisation du pas d'apprentissage de l'algorithme de Levenberg-Marquardt, μ_{ini} , et de son facteur de décroissance, μ_{dec} . Ces deux hyperparamètres semblant liés, nous avons décidé de les choisir simultanément.

Pour chaque horizon de prévision, nous faisons varier μ_{ini} de 10^{-4} à 10 par puissance de 10, et μ_{dec} de 0,1 à 0,9 par pas de 0,1, soit 54 combinaisons différentes. Le nombre de neurones cachés et la dimension de la fenêtre temporelle sont fixés selon les résultats des deux paragraphes précédents.

Comme auparavant, on effectue 200 initialisations des paramètres du réseau différentes pour chaque combinaison (μ_{ini} ; μ_{dec}). On calcule sur les 200 S_{vcp} de chaque série, la moyenne, la médiane, le minimum, le maximum, l'écart-type (les valeurs du maximum et de l'écart type sont consignées en Annexe 3) ainsi que la moyenne des 20 plus petits scores de validation croisée partielle (Figure 21). On détermine la combinaison que l'on va conserver grâce à la validation croisée partielle, en utilisant la moyenne des 20 plus petits S_{vcp} dans chaque série de 200 initialisations (Figure 22).

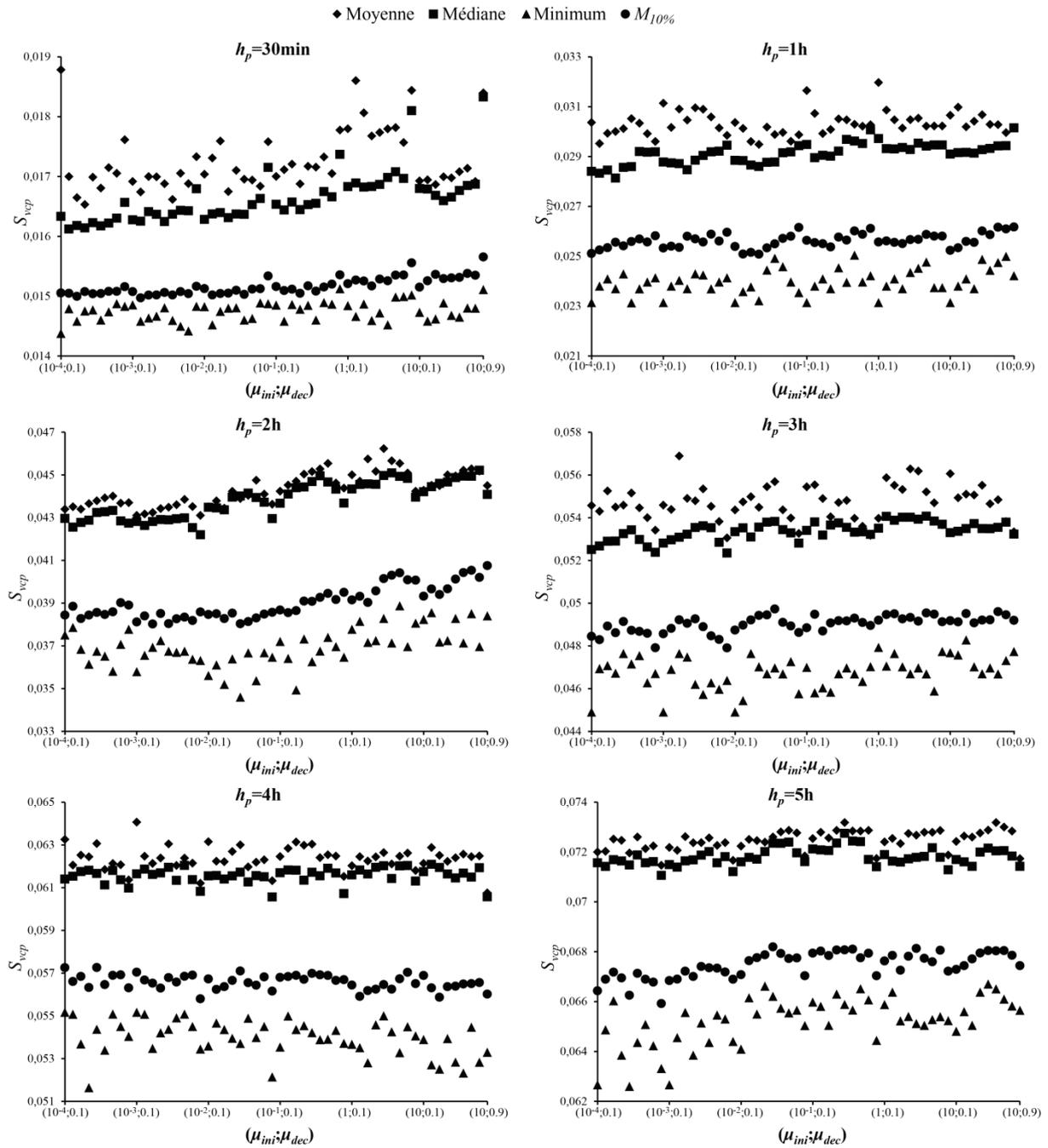


Figure 21: Pour chaque combinaison de $(\mu_{ini}; \mu_{dec})$: moyenne, médiane et minimum sur les 200 scores de validation croisée partielle et moyenne des 20 plus petits scores de validation croisée partielle ; chaque graphique correspond à un horizon de prévision

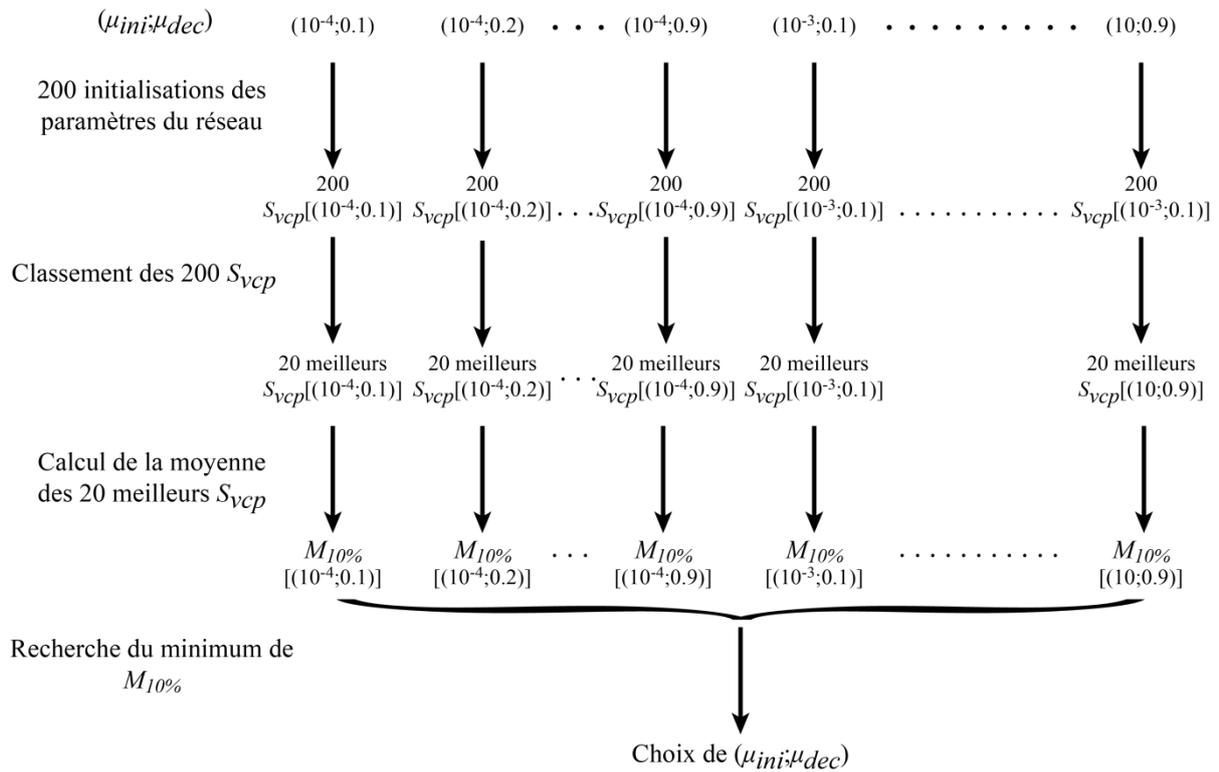


Figure 22: Schéma de la méthode de sélection du pas d'apprentissage et de son facteur de décroissance

Les résultats de la sélection de la combinaison $(\mu_{ini} ; \mu_{dec})$ en fonction de l'horizon de prévision sont répertoriés dans le Tableau 10.

Horizon de prévision	30 min	1h	2h	3h	4h	5h
$(\mu_{ini} ; \mu_{dec})$	$(10^{-3} ; 0,2)$	$(10^{-2} ; 0,4)$	$(10^{-3} ; 0,3)$	$(10^{-3} ; 0,9)$	$(10^{-3} ; 0,9)$	$(10^{-4} ; 0,9)$

Tableau 10 : Valeurs de l'initialisation du pas d'apprentissage et de son facteur de décroissance en fonction de l'horizon de prévision

On peut noter que les valeurs de μ_{ini} retenues sont parmi les plus faibles proposées. Comme indiqué dans l'Annexe 1, ceci signifie que la solution trouvée est meilleure si la méthode d'optimisation prend en compte, dès le départ, les dérivées secondes. En d'autres termes, et sous toutes réserves (notamment concernant le rôle effectif de l'arrêt précoce), on peut penser que la fonction comporte un très grand nombre de minima locaux autour du point d'initialisation qui peuvent servir de piège à l'optimisation avec une règle de gradient simple (μ_{ini} grand). Par ailleurs, l'optimisation en utilisant dès le départ une forme proche des règles de deuxième ordre (μ_{ini} petit) atteindrait, grâce à des mouvements de plus grande amplitude, des minima locaux situés globalement à des niveaux de coût meilleurs. On remarque également que pour les petits horizons de prévision, la méthode privilégie des valeurs faibles de μ_{dec} , alors que pour les horizons de prévision plus importants c'est la valeur maximale de

μ_{dec} qui est choisie. L'algorithme d'apprentissage nécessite donc une variation du pas d'apprentissage plus fine dans le cas des horizons de prévision les plus courts, qui sont également ceux pour lesquels l'information apportée par les variables exogènes est la plus récente.

3.3.3.2. Sensibilité du score de validation croisée partielle à l'initialisation du pas d'apprentissage et à son facteur de décroissance

Dans ce paragraphe nous étudions l'influence de la variation de $(\mu_{ini}; \mu_{dec})$ en étudiant la dispersion de la moyenne des 20 plus petits S_{vcp} , $M_{10\%}$ pour les 54 combinaisons $(\mu_{ini}; \mu_{dec})$, on calcule la moyenne, le minimum, le maximum, l'écart-type, S_1 et S_2 .

$M_{10\%}$	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
Moyenne	$1,52.10^{-2}$	$2,56.10^{-2}$	$3,90.10^{-2}$	$4,90.10^{-2}$	$5,66.10^{-2}$	$6,75.10^{-2}$
Minimum	$1,50.10^{-2}$	$2,51.10^{-2}$	$3,80.10^{-2}$	$4,79.10^{-2}$	$5,58.10^{-2}$	$6,59.10^{-2}$
Maximum	$1,57.10^{-2}$	$2,62.10^{-2}$	$4,08.10^{-2}$	$4,97.10^{-2}$	$5,73.10^{-2}$	$6,82.10^{-2}$
Écart-type	$1,45.10^{-4}$	$2,84.10^{-4}$	$7,64.10^{-4}$	$4,07.10^{-4}$	$3,37.10^{-4}$	$5,23.10^{-4}$
S_1	$4,5.10^{-2}$	$4,3.10^{-2}$	$7,0.10^{-2}$	$3,7.10^{-2}$	$2,6.10^{-2}$	$3,4.10^{-2}$
S_2	$9,6.10^{-3}$	$1,1.10^{-2}$	$2,0.10^{-2}$	$8,3.10^{-3}$	$6,0.10^{-3}$	$7,8.10^{-3}$

Tableau 11 : Pour chaque horizon de prévision, moyenne, minimum, maximum, écart-type, S_1 et S_2 des valeurs de $M_{10\%}$.

Les valeurs des critères S_1 , S_2 et de l'écart-type (Tableau 11) montrent que le score de validation croisée partielle est peu sensible aux variations de $(\mu_{ini}; \mu_{dec})$. Cette faible sensibilité est par ailleurs comparable à la sensibilité du S_{vcp} au nombre de neurones cachés et à la dimension de la fenêtre temporelle. Le fait que le choix de $(\mu_{ini}; \mu_{dec})$, dont les valeurs sont souvent prises par défaut, soit aussi influent sur la qualité de la solution que le choix de l'architecture et des variables est très surprenant. Peut-être que le nombre d'événements de l'ensemble d'apprentissage est trop réduit pour ajuster efficacement des paramètres ce qui donne, en contrepartie, plus d'influence aux hyperparamètres de l'algorithme d'apprentissage.

Par ailleurs, aucune tendance ne se dégage de l'évolution des critères de dispersion en fonction de l'horizon de prévision.

Pour les combinaisons $(\mu_{ini}; \mu_{dec})$ sélectionnées, on évalue la sensibilité des modèles à l'initialisation des paramètres du réseau.

S_{vcp}	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
$(\mu_{ini}; \mu_{dec})$	$(10^{-3}; 0,2)$	$(10^{-2}; 0,4)$	$(10^{-3}; 0,3)$	$(10^{-3}; 0,9)$	$(10^{-3}; 0,9)$	$(10^{-4}; 0,9)$
Moyenne	$1,67.10^{-2}$	$2,95.10^{-2}$	$4,32.10^{-2}$	$5,31.10^{-2}$	$6,12.10^{-2}$	$7,15.10^{-2}$
Minimum	$1,46.10^{-2}$	$2,32.10^{-2}$	$3,69.10^{-2}$	$4,64.10^{-2}$	$5,34.10^{-2}$	$6,33.10^{-2}$
Maximum	$5,13.10^{-2}$	$6,21.10^{-2}$	$6,68.10^{-2}$	$7,71.10^{-2}$	$7,79.10^{-2}$	$8,38.10^{-2}$
Écart-type	$2,95.10^{-3}$	$4,46.10^{-3}$	$4,05.10^{-3}$	$4,13.10^{-3}$	$3,66.10^{-3}$	$3,64.10^{-3}$
S_1	2,2	1,3	$6,9.10^{-1}$	$5,8.10^{-1}$	$4,0.10^{-1}$	$2, .10^{-1}$
S_2	$1,8.10^{-1}$	$1,5.10^{-1}$	$9,4.10^{-2}$	$7,8.10^{-2}$	$6,0.10^{-2}$	$5,1.10^{-2}$

Tableau 12 : Pour chaque horizon de prévision et pour la combinaison $(\mu_{ini}; \mu_{dec})$ déterminée, moyenne, minimum, maximum, écart type, S_1 et S_2 des S_{vcp} sur 200 modèles obtenus avec 200 initialisations différentes des paramètres du réseau

Dans le Tableau 12, on note que la sensibilité du score de validation croisée partielle à l'initialisation des paramètres du réseau est beaucoup plus marquée que la sensibilité de ce score aux hyperparamètres $(\mu_{ini} ; \mu_{dec})$: S_1 et S_2 sont multipliés par un facteur allant de 10 à 100, l'écart type d'un facteur 10.

Au regard de l'évolution de ces critères en fonction de l'horizon de prévision, la sensibilité du S_{vcp} à l'initialisation des paramètres du réseau ne suit pas de tendance particulière.

3.3.3.3. Sélection de modèles issus de différentes initialisations des paramètres du réseau

Les trois paragraphes précédents ont révélé la sensibilité importante de la capacité de généralisation des modèles à l'initialisation des paramètres du réseau. Compte tenu de cette situation, on va retenir non pas un seul, mais 20 modèles, chacun issu d'une initialisation des paramètres différentes afin d'obtenir un ensemble de prévisions et de pouvoir analyser les différences de qualité des limnigrammes. On considère les simulations effectuées au paragraphe précédent sous un nouveau point de vue : pour chacune des 200 initialisations, on considère les résultats obtenus pour les 54 combinaisons de $(\mu_{ini} ; \mu_{dec})$. Le nombre de neurones cachés et la dimension de la fenêtre temporelle des pluies sont fixés : $(\mu_{ini} ; \mu_{dec})$ sont les seuls hyperparamètres sur lesquels on peut jouer afin d'évaluer la pertinence d'une même initialisation, car modifier le nombre de neurones cachés ou la dimension de la fenêtre temporelle des pluies change le nombre de paramètres du réseau, les initialisations ne sont par conséquent plus comparables.

On obtient ainsi 54 S_{vcp} correspondant à 54 modèles issus d'une même initialisation, puis on calcule le minimum et l'écart-type de ces 54 scores. Dans un premier temps, on sélectionne les 50 initialisations donnant les modèles avec les plus petits minimums de S_{vcp} afin de retenir les initialisations à l'origine des meilleurs modèles. Puis, parmi les 50, on choisit les 20 initialisations donnant les modèles avec les plus faibles écarts-types de S_{vcp} pour conserver les initialisations conduisant à des modèles pour lesquels les estimations de la capacité de généralisation sont les moins dispersées. La Figure 23 illustre la méthode employée.

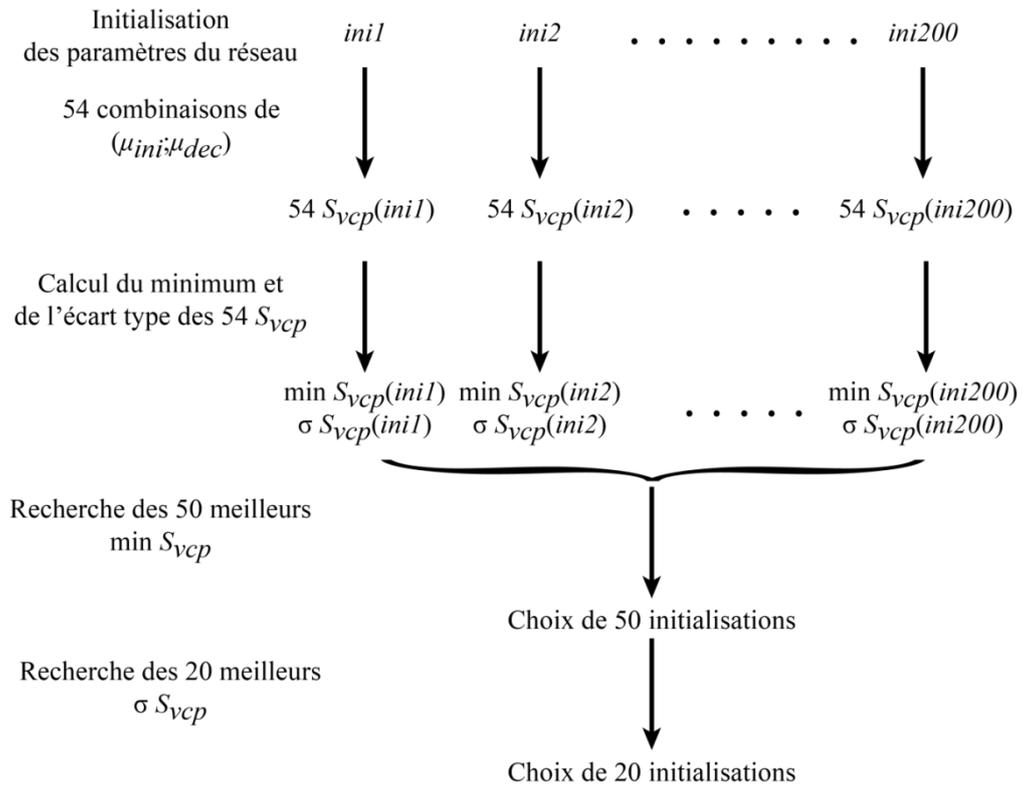


Figure 23: Schéma de la méthode de sélection de l'initialisation des paramètres du réseau

3.3.4. Conclusion sur la sélection de modèles

Les différents hyperparamètres des modèles : nombre de neurones cachés, dimension de la fenêtre temporelle des pluies, initialisation du pas d'apprentissage, facteur de décroissance du pas d'apprentissage et initialisation des paramètres du réseau sont maintenant optimisés pour la prévision de la hauteur du Gardon d'Anduze. L'apprentissage par l'algorithme de Levenberg Marquardt régularisé par l'arrêt précoce sur l'événement 13, est réalisé sur tous les événements de la base excepté les événements 13, 19, 26 et 27. On va tester les performances des modèles sur les événements 19, 26 et 27 qui sont inconnus des modèles.

3.4. Performances des modèles sélectionnés sur les événements en test

3.4.1. Caractéristiques des événements en test

Les trois événements présentés *successivement* en test sont le 19 (septembre 2002), le 26 (octobre 2008) et le 27 (novembre 2008) : on ne présente pas aux réseaux un unique ensemble de test comprenant les trois événements, mais chaque événement est testé séparément.

Comme indiqué dans le Chapitre 2 (2.6), l'événement 19 a été retenu car il est le plus intense de la base de données. On doit noter que l'événement le plus intense de la base d'apprentissage a un pic de hauteur plus faible de 30%. Les hauteurs d'eau étant également des variables des modèles, l'événement 19 permettra donc de tester les prévisions en dehors du domaine de variation des variables dans l'ensemble d'apprentissage. Sachant, comme nous l'avons présenté au Chapitre 2, que les processus physiques intervenant lors d'une crue rapide ne sont pas complètement connus, il est difficile de préciser si le réseau généralise à des

grandeurs supérieures à celles de son ensemble d'apprentissage ou s'il extrapole à de nouveaux processus.

Les événements 26 et 27 ont des hauteurs maximales semblables à celle de l'événement le plus intense de l'ensemble d'apprentissage : l'évaluation des modèles se centre ainsi sur la qualité de prévision des événements les plus intenses, comme souhaité par le SCHAPI.

Ces trois événements présentent des caractéristiques très différentes. Il sera ainsi possible d'apprécier les performances des modèles sur des configurations variées. Les événements 19 et le 26 présentent un seul pic, alors que le 27 en possède plusieurs. Les 19 et 26 sont considérés comme courts, alors que le 27 est environ deux fois plus long que les deux autres. L'état d'humidité des sols et la distribution spatiale des précipitations sont également très différents pour ces trois événements.

Pour chaque horizon de prévision, on dispose de 20 modèles différents provenant de 20 initialisations des paramètres du réseau différentes. On va observer les résultats des modèles sur les 3 événements en test.

3.4.2. Prévisions des 20 modèles sélectionnés

Par souci de lisibilité des limnigrammes, cette partie est consacrée essentiellement à la comparaison des différents modèles pour un même horizon de prévision ; l'étude de la qualité des modèles en fonction de l'horizon de prévision se fera par la suite.

3.4.2.1. Événement 19

Pour chaque horizon de prévision, on dispose de 20 hauteurs calculées par les 20 modèles retenus ; les limnigrammes ainsi obtenus pour l'événement 19 sont représentés Figure 24.

On remarque immédiatement que pour un même horizon de prévision, les modèles testés donnent des prévisions très différentes. Certaines hauteurs sont même aberrantes avec des valeurs jusqu'à 60m de hauteur pour $h_p=1h$, d'autres sont négatives pour les horizons de prévision 4h et 5h. Jusqu'à $h_p=2h$, dans chaque cas deux ou trois modèles présentent des hauteurs prévues de qualité intéressante. Ainsi on peut souligner deux constatations :

- certains limnigrammes sont aberrants et complètement inexploitable,
- ces limnigrammes aberrants en test sont obtenus avec des modèles ayant pourtant été sélectionnés par la validation croisée.

Cet échec de la sélection par la validation croisée est explicable par le fait que les événements intenses sont très peu nombreux dans la base de données et donc tous différents les uns des autres. En tout état de cause, il apparaît à nouveau que le choix de l'initialisation des paramètres du réseau a une influence déterminante sur les performances des modèles en test.

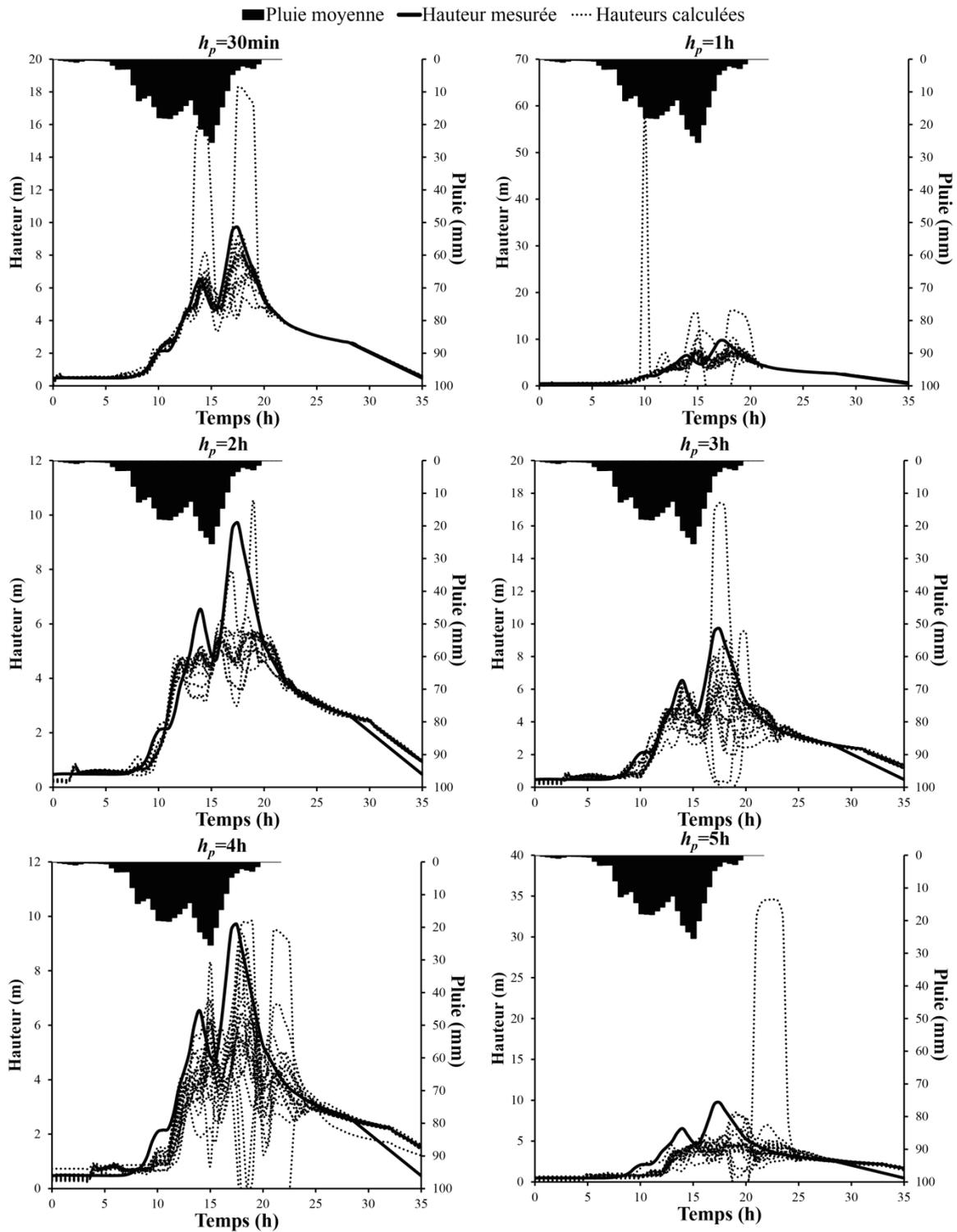


Figure 24: Événement 19, les 20 hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision

La Figure 25 illustre bien la dispersion des critères sur les 20 modèles ; pour ne pas alourdir la présentation des résultats, les tableaux de valeurs se trouvent en Annexe 4, les modèles sont classés du meilleur au moins bon selon le critère de persistance. On peut noter que ce classement est le même si on trie les modèles selon les autres critères. Pour $h_p=30\text{min}$, les valeurs du critère de persistance s'échelonnent de 0,54 à -37 (ce dernier point n'est pas

représenté sur le graphique afin de ne pas « écraser » l'échelle des autres résultats). Cela conforte les observations faites à partir des limnigrammes, pour chaque horizon de prévision la dispersion des performances des modèles est semblable : la qualité se dégrade régulièrement jusqu'aux 3 à 5 derniers modèles du classement qui présentent des valeurs de critères très mauvaises correspondant aux hauteurs aberrantes.

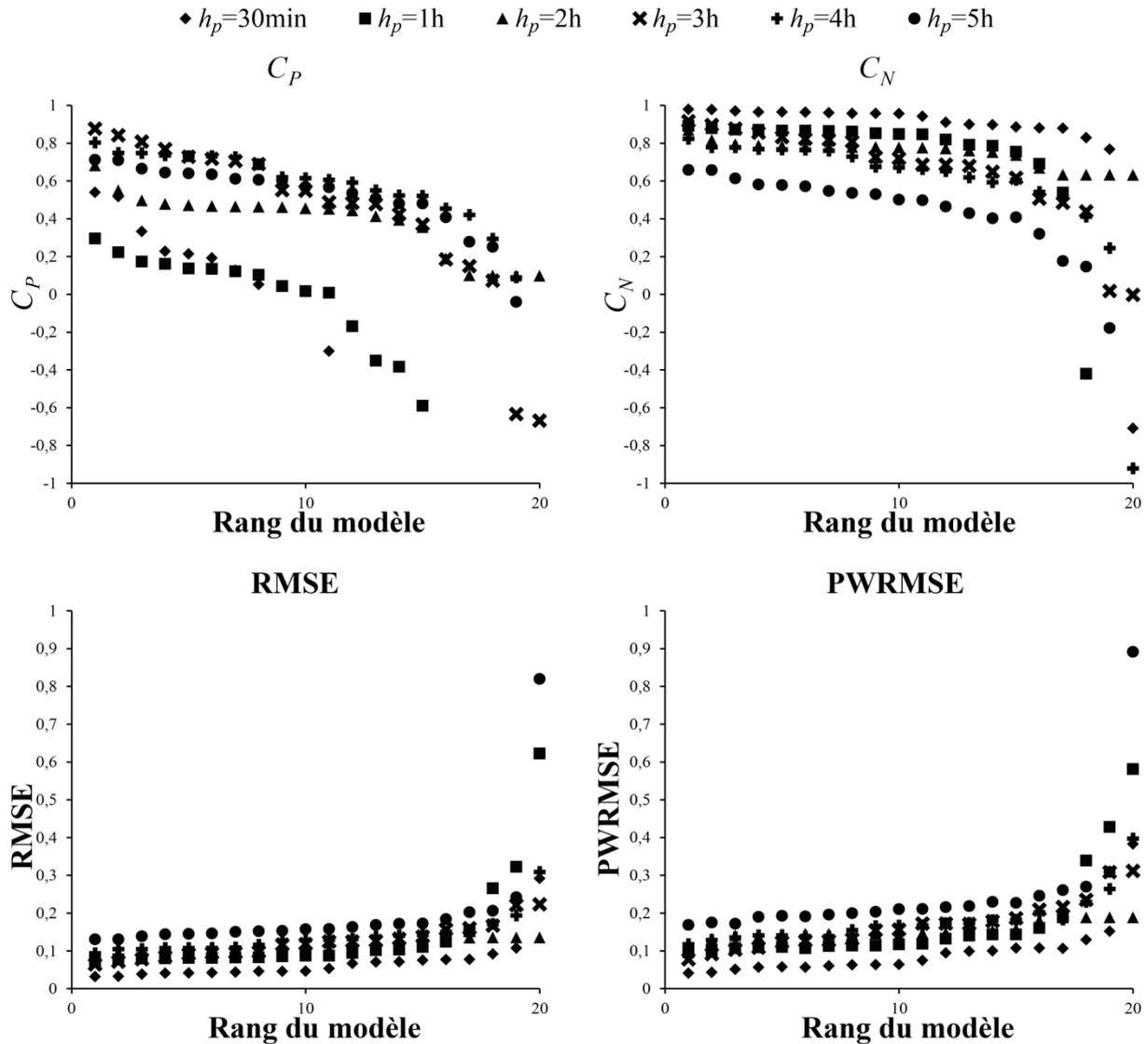


Figure 25: Valeurs des critères C_P , C_N , RMSE et PWRMSE pour le test de l'événement 19 sur les 20 modèles pour chaque horizon de prévision. L'axe des abscisses correspond au rang du modèle : le rang 1 correspond au modèle ayant le meilleur C_P et le 20 au plus mauvais C_P .

3.4.2.2. Événement 26

Les prévisions relatives à l'événement 26 sont globalement plus satisfaisantes que celles présentées ci-dessus pour l'événement 19, comme on peut le constater sur Figure 26. On note une dispersion des hauteurs prévues par les différents modèles, cependant cette dispersion est bien moindre que pour l'événement 19. Les performances des modèles sur un événement dont la hauteur maximale est présente dans l'ensemble d'apprentissage, contrairement au 19, sont plus cohérentes : il n'y a pas de hauteurs aberrantes comme lorsque les modèles doivent extrapoler sur les hauteurs particulièrement grandes du 19.

Tous les modèles prévoient un faux pic, d'amplitude plus ou moins importante, en réaction au pic de pluie très intense à Mialet en début d'événement : il est tombé, sur un sol sec, 91mm de pluie en 1h30 dont 62mm en 30min. Par ailleurs, toutes les hauteurs prévues présentent des oscillations dont l'amplitude augmente avec l'horizon de prévision ; ces oscillations sont apparemment liées aux impulsions de pluie, nombreuses pour cet événement.

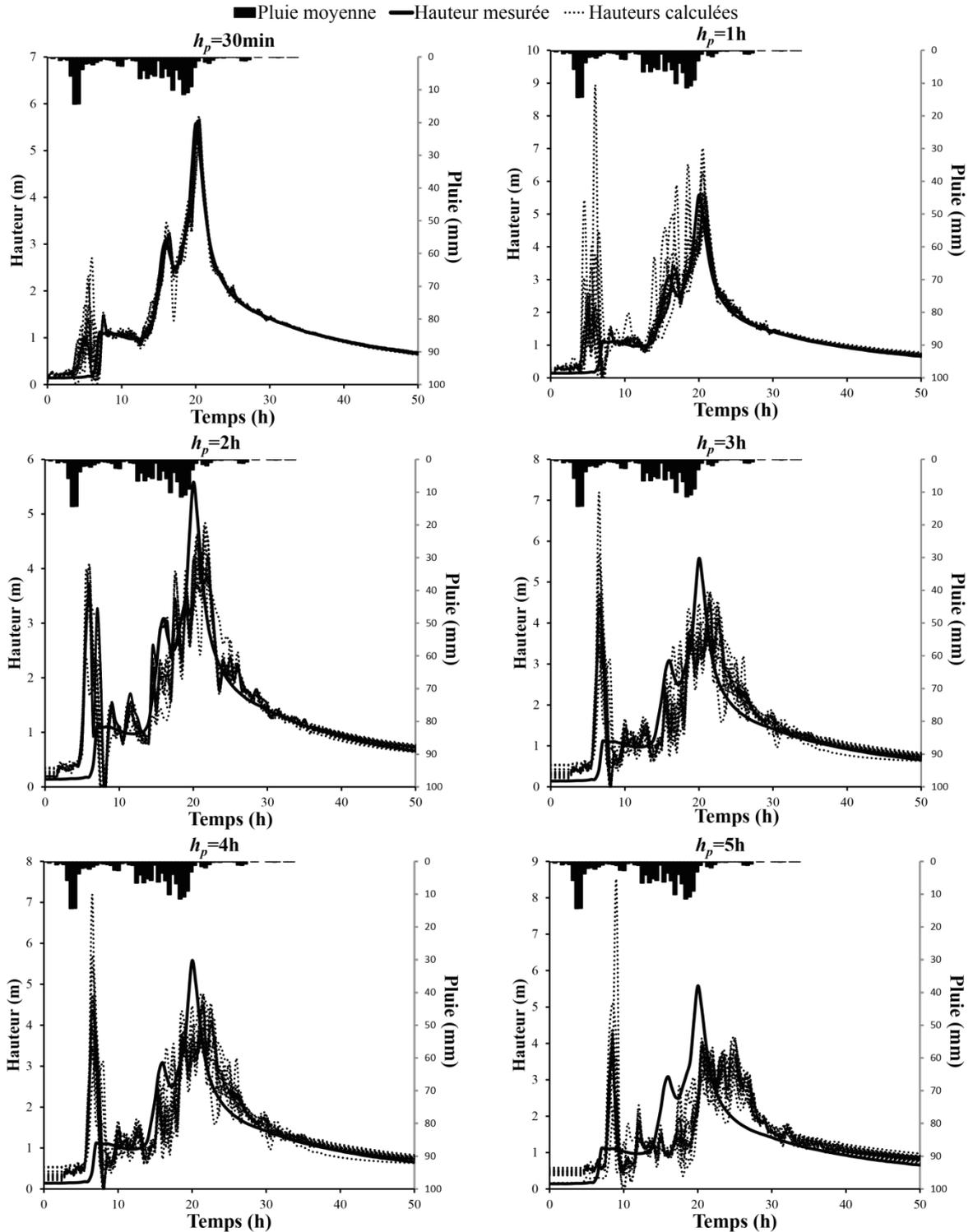


Figure 26: Événement 26, les 20 hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision

La Figure 27 présente les critères de performance pour les 20 modèles, par horizon de prévision, les tableaux de valeurs se trouvent Annexe 4. Les modèles sont classés du meilleur au moins bon selon le critère de persistance ; ce classement est le même si on trie les modèles selon les autres critères exceptés quelques petites différences pour PWRMSE. Le critère de persistance rend le mieux compte de la dispersion des résultats en particulier pour $h_p=30\text{min}$ et $h_p=1\text{h}$ où l'écart entre le meilleur et le plus mauvais modèle est très marqué. Cependant, comme pour les limnigrammes, la dispersion est moins importante que pour l'événement 19.

On peut noter sur la Figure 27, de manière plus marquée que sur la Figure 25, que ce sont les modèles de l'horizon 30min qui sont les plus « performants » pour le critère de Nash (C_N), tandis que, ce sont les modèles relatifs à l'horizon 4h qui semblent les meilleurs pour le critère de persistance, contredisant ainsi l'appréciation de la forme des limnigrammes de la Figure 26. Pour expliquer cela, il faut revenir à la présentation faite des critères dans le Chapitre 2 : pour un horizon de prévision très court, le limnigramme qui recopie en prévision la mesure observée au temps précédent a un excellent score de Nash alors qu'il n'apporte aucune anticipation. En revanche le critère de persistance qui prend en compte la comparaison de la prévision réalisée avec la prévision naïve apporte une meilleure appréciation sur la valeur ajoutée du modèle.

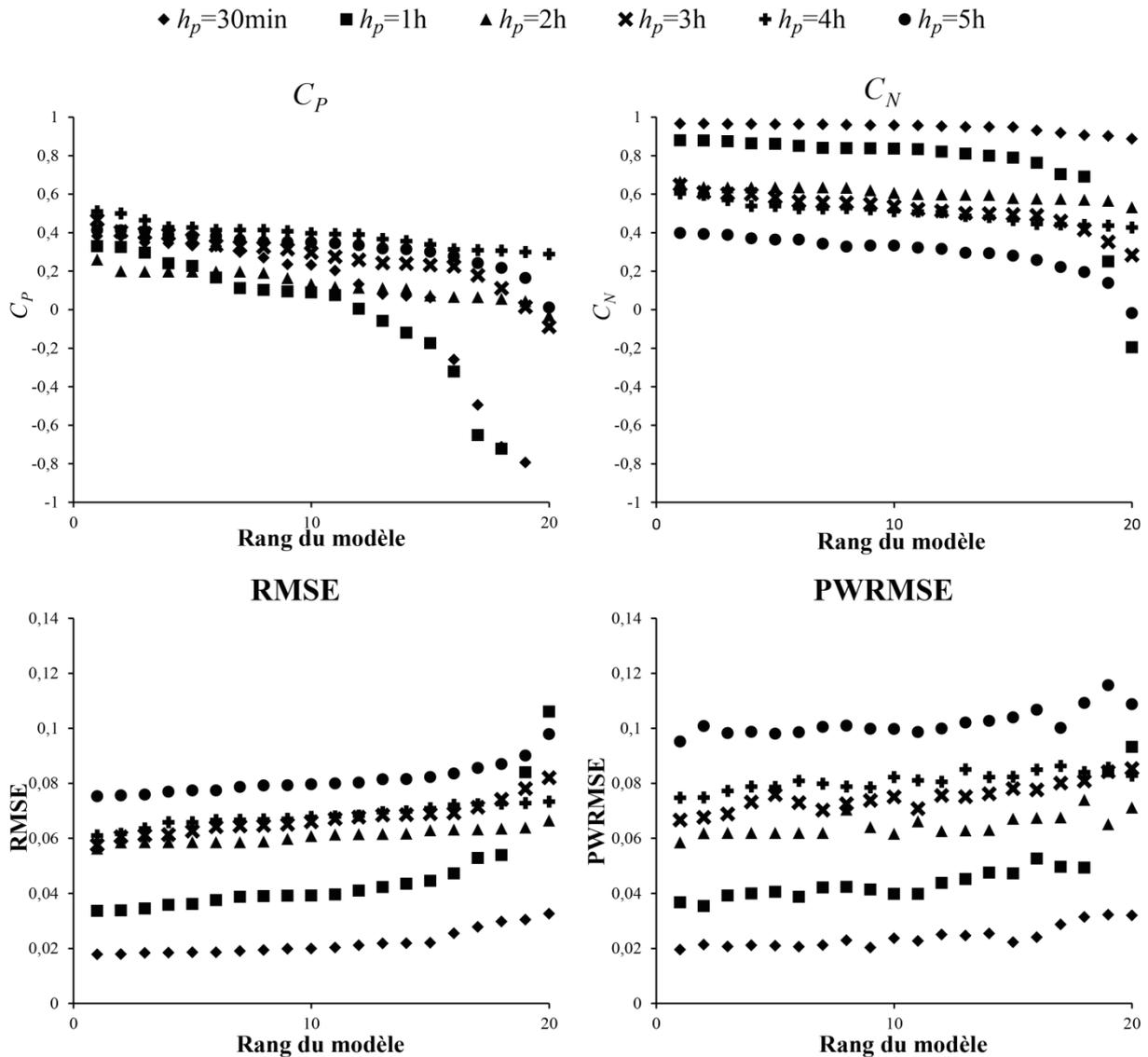


Figure 27: Valeurs des critères C_P , C_N , RMSE et PWRMSE pour le test de l'événement 26 sur les 20 modèles pour chaque horizon de prévision. L'axe des abscisses correspond au rang du modèle par C_P décroissants.

3.4.2.3. Événement 27

Les résultats obtenus pour l'événement 27 sont présentés sur la Figure 28. La dispersion est faible pour cet événement, seules les hauteurs les plus éloignées du limnigramme observé se distinguent les unes des autres. Aucune hauteur n'est aberrante comme pour le 19. Des fluctuations rapides apparaissent comme pour l'événement 26, elles sont là-aussi liées à la forme du hyétogramme. Comme pour les autres événements, les prévisions se dégradent régulièrement lorsque l'horizon de prévision augmente. Ceci est compréhensible puisque les modèles ne disposent pas de prévision des précipitations ni d'information sur l'état futur du système par des mesures de hauteur.

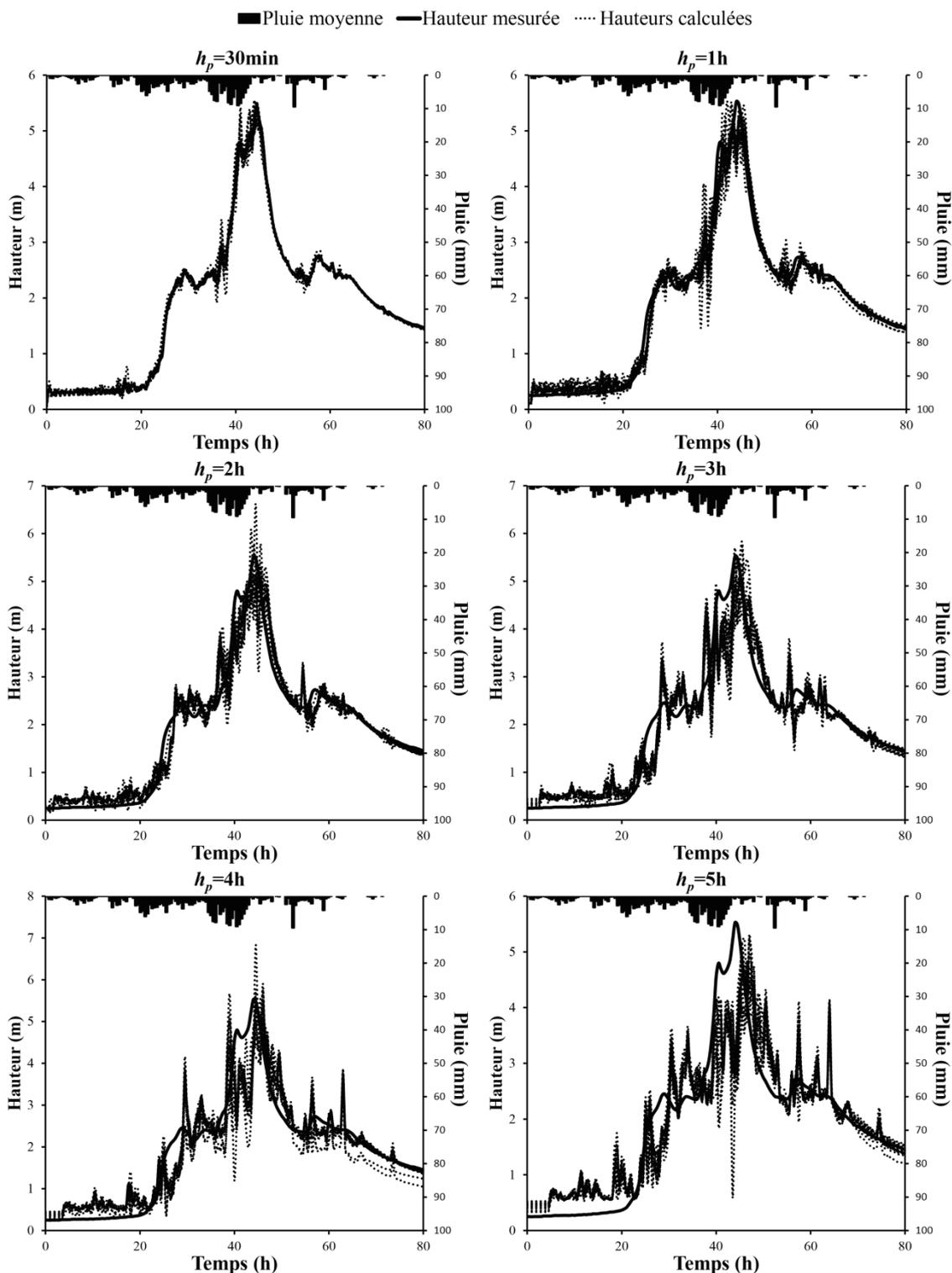


Figure 28: Événement 27, les 20 hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision

La Figure 29 présente les critères de performance pour les 20 modèles par horizon de prévision, les tableaux de valeurs se trouvent Annexe 4. Les modèles sont classés du meilleur au moins bon selon le critère de persistance.

Seul le critère de persistance présente une dispersion importante des performances des différents modèles notamment pour $h_p=30\text{min}$ et $h_p=1\text{h}$ où le C_P prend des valeurs négatives.

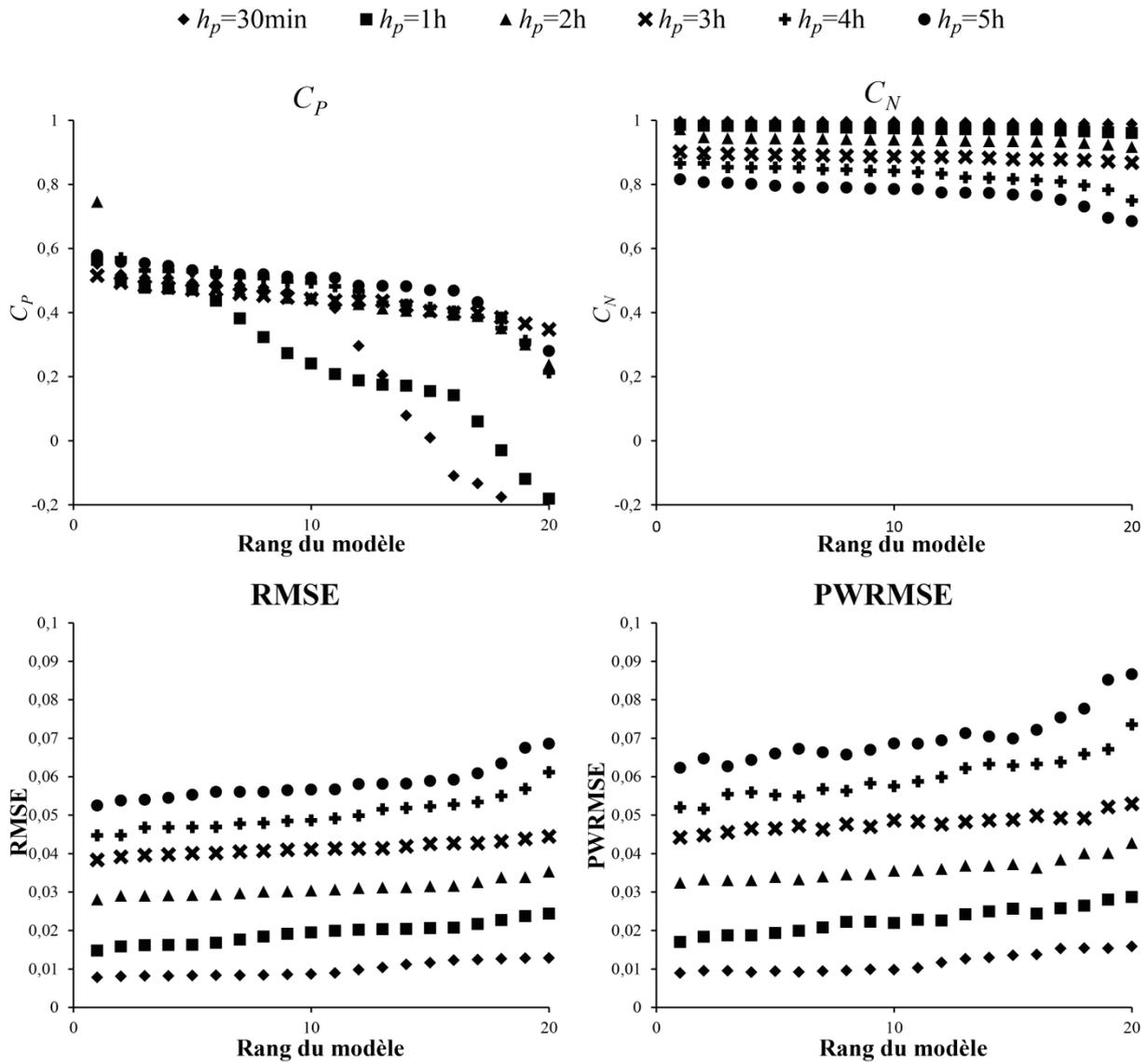


Figure 29: Valeurs des critères C_P , C_N , RMSE et PWRMSE pour le test de l'événement 27 sur les 20 modèles pour chaque horizon de prévision. L'axe des abscisses correspond au rang du modèle par C_P décroissants.

3.4.3. Médiane des prévisions

Afin de diminuer la variabilité des sorties obtenues avec les 20 meilleurs modèles sélectionnés par la validation croisée, et vu que l'on ne peut pas déterminer à l'avance quel modèle sera le plus pertinent, on se propose de procéder de façon analogue à la méthode des ensembles [Diettrich 2000], en calculant à chaque pas de temps la médiane des 20 prévisions obtenues par les différents modèles, et représentées dans le paragraphe précédent. La médiane est préférée à la moyenne pour ses propriétés de robustesse vis-à-vis des valeurs aberrantes.

Pour juger de la qualité de cette prédiction, nous calculons le critère de persistance, le critère de Nash, à comparer au critère de Nash de la prédiction naïve, le retard entre le sommet du pic de la médiane et celui de la mesure, ainsi que les *PPD* et *SPPD* précédemment présentés au Chapitre 2. Tous ces indicateurs permettent d'évaluer la qualité des limnigrammes.

3.4.3.1.

Événement 19

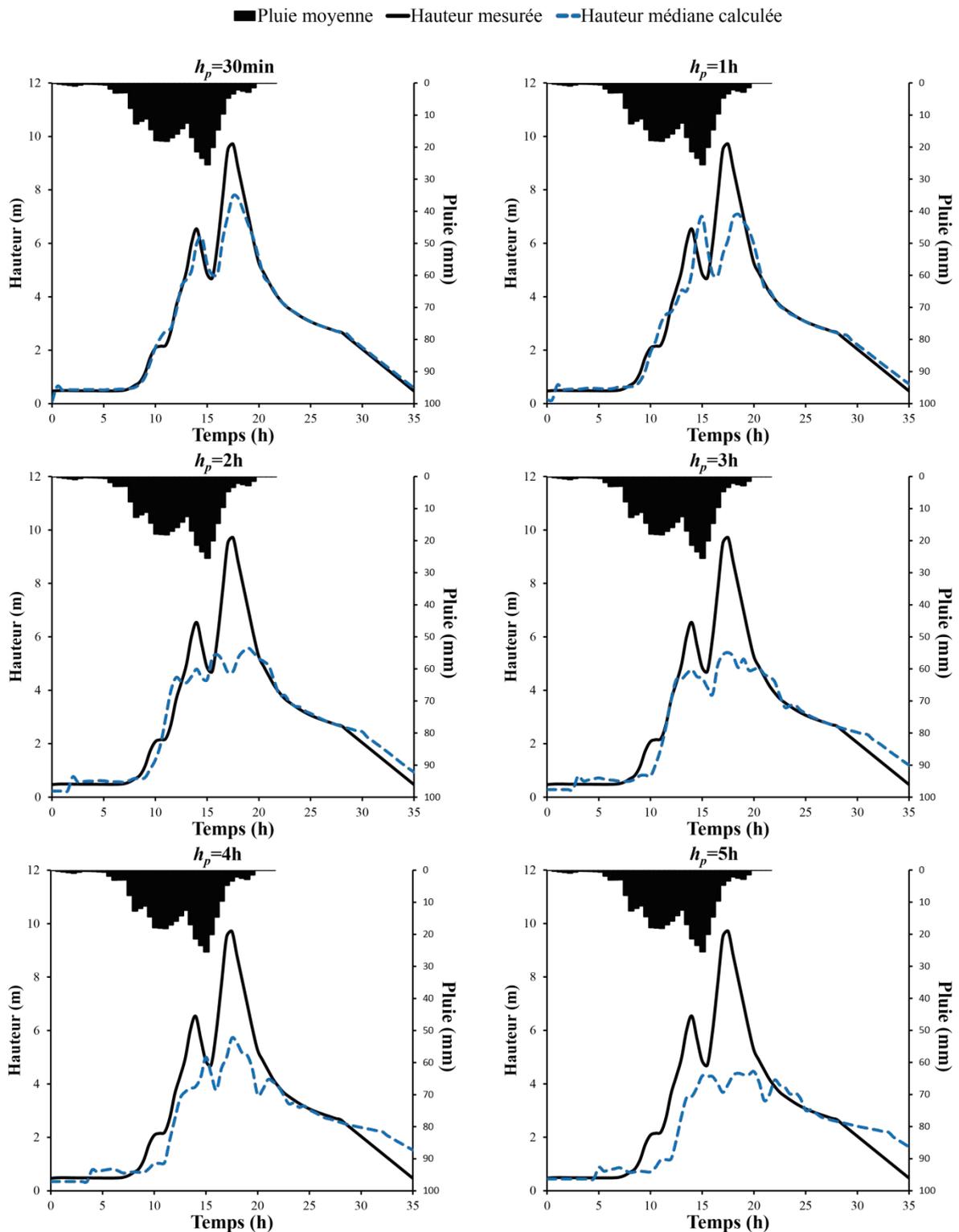


Figure 30: Événement 19, hauteur médiane calculée sur les hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision

	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
C_P	~ 0	0,08	0,46	0,64	0,69	0,62
C_N	0,96	0,86	0,78	0,78	0,73	0,56
C_N naïve	0,95	0,84	0,57	0,34	0,07	-0,33
PPD (%)	80	73	57	56	59	46
SPPD (%)	80	63	48	56	59	41
Retard H_{pic} (h)	0	1	1,5	0	0	2,5

Tableau 13 : Valeurs du C_P , C_N , PPD , $SPPD$ et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 19

La Figure 30 présente les prévisions de hauteur de l'événement 19 pour les différents horizons et le Tableau 13 les critères correspondants. La forme du limnigramme prévu est la même que celle du limnigramme mesuré pour $h_p=30\text{min}$ et $h_p=1h$.

Pour $h_p=30\text{min}$, la prévision est correcte : le modèle donne un pic synchrone avec une amplitude de 80% par rapport à celui observé, même si sur l'ensemble de l'événement la persistance nulle indique la qualité du modèle est équivalente à celle de la prévision naïve. La prévision naïve donne bien évidemment une hauteur maximale parfaite, mais avec un retard égal à l'horizon de prévision.

Pour $h_p=1h$ et $h_p=2h$, le pic calculé subit un retard de l'ordre de l'horizon de prévision et la hauteur prévue au moment du pic diminue à 63% et 48% de la valeur du pic. Au regard des critères, la prévision est meilleure que la prévision naïve et les critères de Nash restent satisfaisants, mais cette appréciation cache une anticipation assez mauvaise, tant pour l'instant du maximum du pic que pour son amplitude.

Pour $h_p=3h$ et $h_p=4h$, le pic prévu est à nouveau synchrone et son amplitude représente 56% et 59% par rapport à celui observé. Pour cet événement, ces horizons sont cependant trop grands pour que la prévision soit correcte car le modèle ne dispose pas de l'information sur le pic de pluie qui arrive environ 2,5h avant le pic de hauteur.

Pour $h_p=5h$, la qualité de la prévision est très insuffisante, une hauteur prévue au moment du pic égale à seulement 41% de la hauteur observée n'a aucun intérêt ; les informations de pluie semblent donc indispensables pour modéliser cet événement à un horizon supérieur à 2 heures.

3.4.3.2. Événement 26

De même que pour l'événement 19, la médiane est calculée à partir des 20 meilleurs modèles sélectionnés par la validation croisée. La Figure 31 rassemble les prévisions de hauteur de l'événement 26 pour les différents horizons, et le Tableau 14 les critères correspondants. Comme on pouvait le voir sur les prévisions des modèles, à partir de $h_p=2h$, le modèle répond à l'important pic de pluie à Mialet, par un faux pic de hauteur en début

d'événement avec un retard de l'ordre de l'horizon de prévision par rapport au pic de pluie, et, d'une amplitude de l'ordre de 4m qui dépasse, pour certains horizons de prévision, celle du « vrai » pic. En situation opérationnelle, ce pic pourrait déclencher une fausse alerte (qui serait cependant détectée par le prévisionniste comme telle compte tenu de la situation). De plus la hauteur prévue est très sensible aux variations d'intensité de pluie ce qui engendre de nombreuses fluctuations importantes d'amplitude.

La forme du limnigramme prévu est la même que celle du limnigramme observé pour $h_p=30\text{min}$ et $h_p=1\text{h}$. Pour $h_p=30\text{min}$, la prévision est excellente, la hauteur prévue au moment du pic observé est de 93%, la prévision du modèle est meilleure que celle de la prévision naïve malgré un décalage du pic égal à l'horizon de prévision. Pour $h_p=1\text{h}$, la prévision est satisfaisante avec *SPPD* de 72%.

Malgré une dégradation de la qualité de la montée du pic, les limnigrammes prévus pour $h_p=2\text{h}$, $h_p=3\text{h}$ et $h_p=4\text{h}$ sont du même ordre d'amplitude avec un *SPPD* de 63%, ce qui est en deçà de la tolérance admise sur l'estimation des débits des crues rapides (30% selon [Marchandise 2007]). Pour $h_p=5\text{h}$, même si le pic prévu est quasiment synchrone avec le pic observé, la montée prévue est tellement éloignée de la montée observée que la qualité de la prévision en est nettement détériorée.

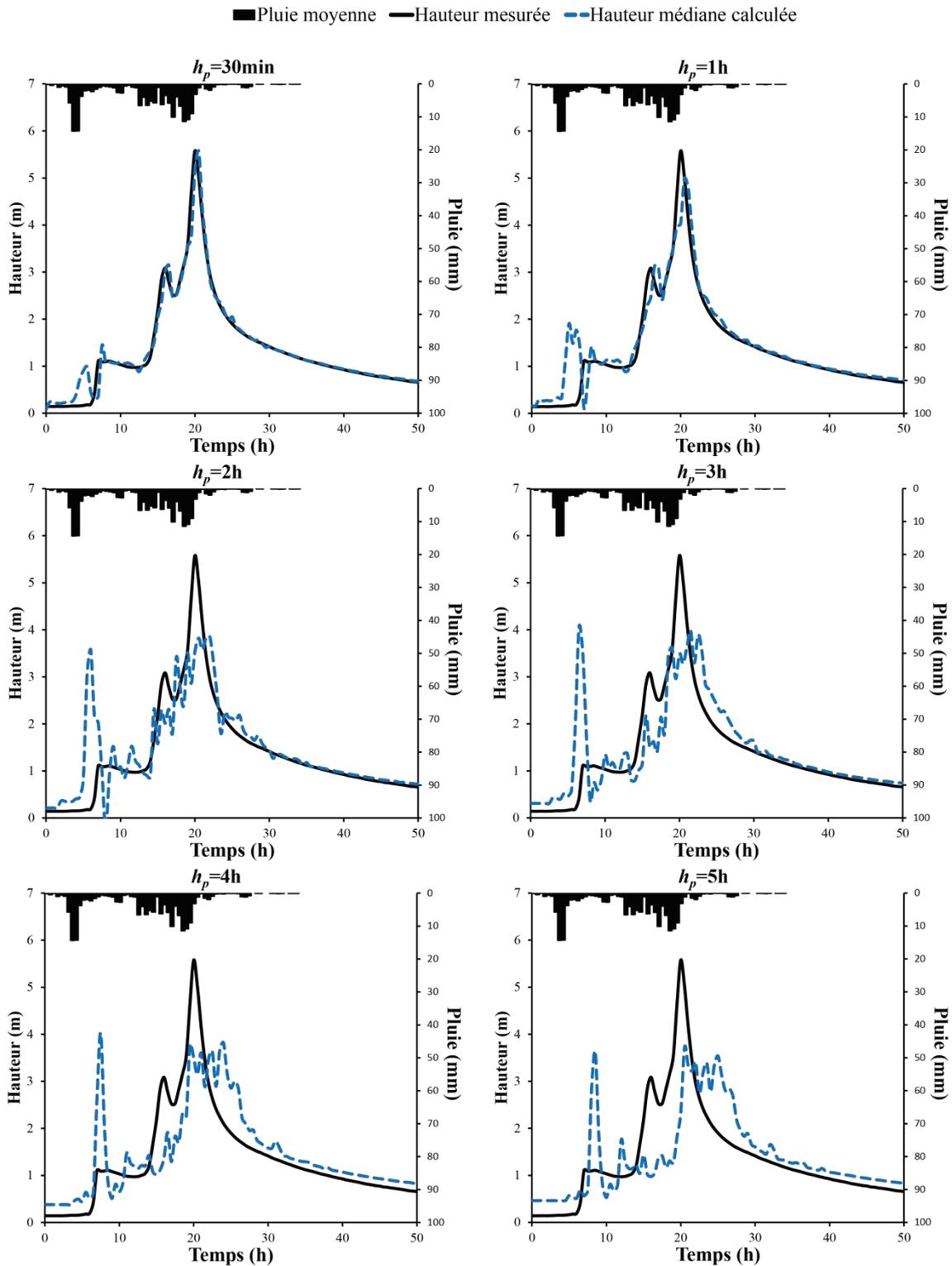


Figure 31: Événement 26, hauteur médiane calculée sur les hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision

	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
C_P	0,31	0,23	0,17	0,33	0,44	0,36
C_N	0,96	0,86	0,62	0,56	0,55	0,35
C_N naïve	0,94	0,82	0,52	0,29	0,10	-0,18
PPD (%)	~100	89	69	71	69	67
SPPD (%)	93	72	63	63	63	42
Retard H_{pic} (h)	0,5	0,5	2	1,5	4	0,5

Tableau 14: Valeurs du C_P , C_N , PPD , $SPPD$ et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 26

3.4.3.3. Événement 27

De même que pour les événements 19 et 26, la médiane est calculée à partir des 20 meilleurs modèles sélectionnés par la validation croisée partielle. La Figure 32 illustre les prévisions de hauteur de l'événement 27 pour les différents horizons et le Tableau 15 les critères correspondants. Comme pour l'événement 26, le hyétogramme très irrégulier se répercute sur la hauteur prévue sous formes d'oscillations plus ou moins marquées suivant l'horizon de prévision. La forme du limnigramme prévu est semblable à celle du limnigramme observé jusqu'à $h_p=2h$.

Pour $h_p=30min$, la prévision est excellente avec un critère de Nash approchant 1 et un $SPPD$ de 95 %. Pour $h_p=1h$ et $h_p=2h$, la prévision reste très bonne avec un pic prévu presque synchrone et un $SPPD$ de 85%.

Pour $h_p=3h$ et $h_p=4h$, l'amplitude des oscillations parasites augmente, mais les prévisions sont toujours intéressantes avec peu de retard du pic et un PPD autour 90%.

Pour $h_p=5h$, la prévision n'est pas satisfaisante ; elle présente un PPD correct mais souffre d'un retard dans le pic de 3,5h qui limite grandement son intérêt.

Les valeurs élevées de PPD et $SPPD$ pour cet événement, par rapport aux deux précédents, s'explique par sa durée : le 27 est de l'ordre de 2 fois plus lent que le 19 et le 26 ; d'une part, le manque de prévision de pluie se fait moins ressentir, et d'autre part, la hauteur variant moins vite, la dernière valeur observée est plus proche de la valeur attendue.

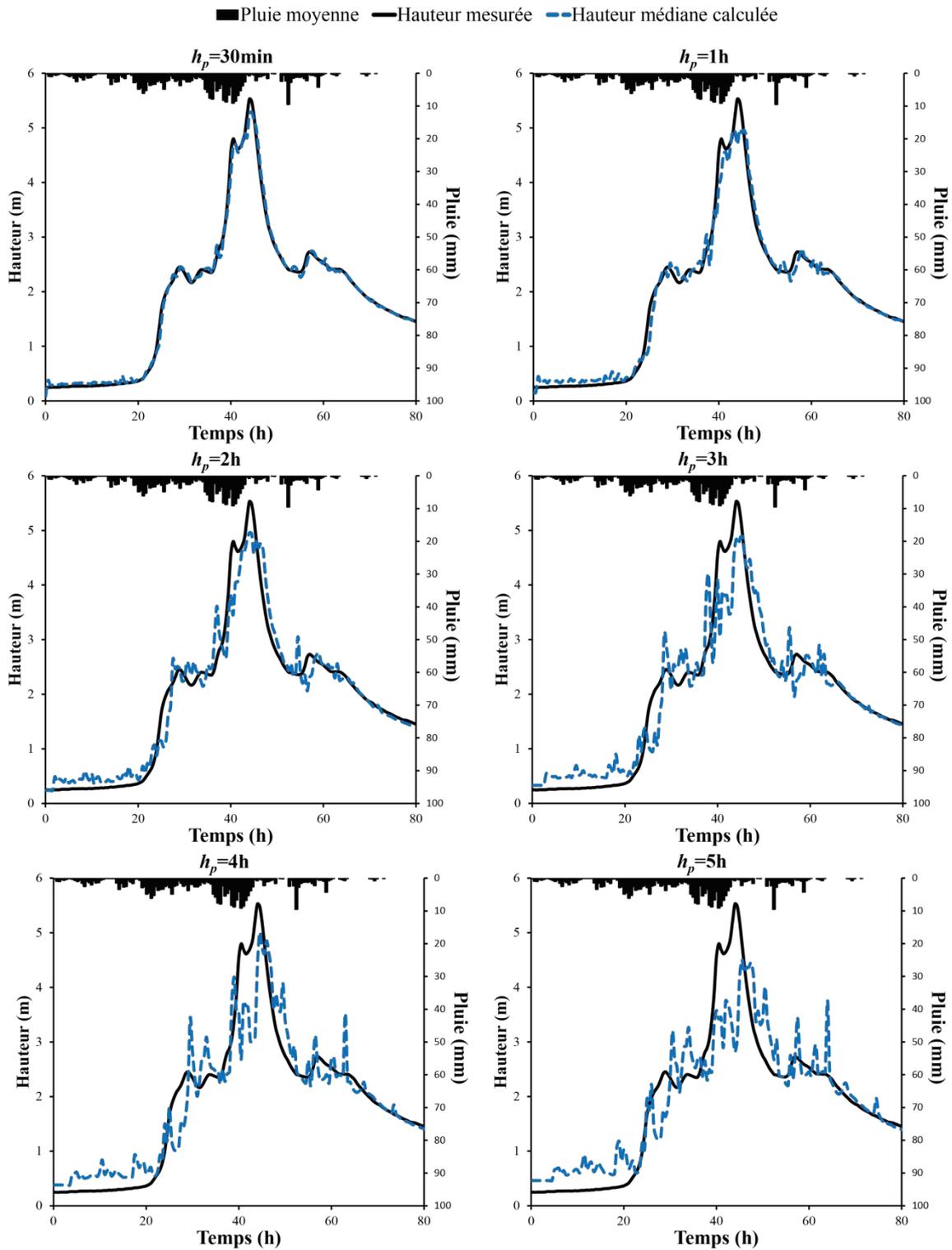


Figure 32: Événement 27, hauteur médiane calculée sur les hauteurs prévues par les 20 modèles sélectionnés pour chaque horizon de prévision

	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
C_P	0,53	0,48	0,52	0,49	0,52	0,54
C_N	~1	0,98	0,95	0,90	0,85	0,80
$C_{N\text{ naïve}}$	0,99	0,97	0,89	0,78	0,66	0,51
PPD (%)	96	90	90	88	91	81
$SPPD$ (%)	95	85	85	87	79	51
Retard H_{pic} (h)	0,5	0,5	0,5	1	0,5	3,5

Tableau 15 : Valeurs du C_P , C_N , PPD , $SPPD$ et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 27

3.4.4. Conclusion sur les performances des modèles à perceptron multicouche

Nous avons constaté qu'avec le type de modèle choisi, les performances en test peuvent être très variables à cause, notamment, de la sensibilité de l'estimation de la capacité de généralisation des modèles à l'initialisation des paramètres du réseau qui se révèle bien plus importante que le choix des variables et de la complexité du modèle. Cette variabilité a pour conséquence que les meilleurs modèles sélectionnés par la validation croisée partielle ne sont pas nécessairement les meilleurs modèles en test. Une méthode autre que la validation croisée pour la sélection de modèle ne pourrait résoudre le problème puisqu'il est dû à la grande hétérogénéité des événements ; c'est pourquoi nous allons choisir une autre approche.

3.5. Modification du modèle avec ajout d'une partie linéaire

3.5.1. Modèle modifié

La variabilité des performances en test a montré la difficulté des modèles à perceptron multicouche à généraliser. Afin d'améliorer les prévisions, nous avons ajouté aux modèles des termes linéaires, encore appelés *termes directs* [Dreyfus 2008]. On suppose que les modèles précédemment utilisés possèdent une composante linéaire, l'ajout des termes directs permet ainsi de mieux apprendre cette linéarité.

Le réseau de neurones proposé, présenté Figure 33, est composé d'un perceptron multicouche standard auquel des connexions linéaires entre les variables et le neurone de sortie (dont la fonction d'activation est l'identité) ont été ajoutées, ce qui conduit à une relation linéaire.

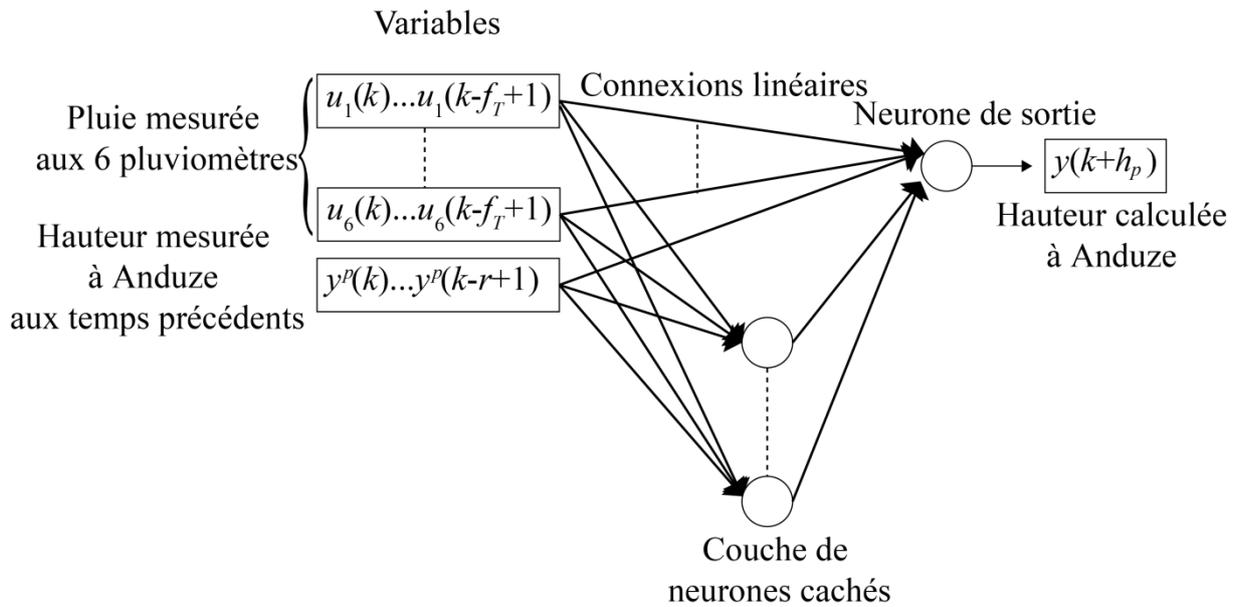


Figure 33 : Modèle modifié avec introduction d'une partie linéaire

Afin de pouvoir établir des comparaisons avec les résultats présentés précédemment, la base de données et la répartition des différents événements dans les ensembles d'apprentissage, d'arrêt et de test restent les mêmes.

3.5.2. *Choix des hyperparamètres*

Pour être totalement rigoureux, il aurait été nécessaire de réaliser toutes les étapes de sélection de modèles mises en œuvre pour l'architecture précédente : sélection du nombre de neurones cachés, de la dimension de la fenêtre temporelle, de l'initialisation du pas d'apprentissage et du facteur de décroissance du pas d'apprentissage. Cependant, la durée limitée de la thèse n'a pas permis de réaliser toutes ces étapes car ce travail a été effectué tardivement. De plus, du fait de la faible sensibilité de la capacité de généralisation des modèles à ces hyperparamètres, il semble raisonnable de reprendre les valeurs établies précédemment pour le perceptron multicouche et de ne s'intéresser qu'à l'initialisation des paramètres du réseau. Ainsi, pour chaque horizon de prévision, sur 200 modèles issus de 200 initialisations, on choisira les 20 modèles qui ont les plus petits scores de validation croisée partielle.

3.5.3. *Prévisions des 20 modèles sélectionnés*

Dans ce paragraphe, on s'intéresse à la dispersion des résultats obtenus sur les prévisions des 20 modèles sélectionnés par la validation croisée partielle, pour chaque horizon de prévision. Le critère de performance retenu est la persistance, car on a pu constater dans la section précédente que c'est le critère qui, d'une part, est spécifique à la prévision, et d'autre part, est le plus discriminant car le plus variable.

3.5.3.1. Évènement 19

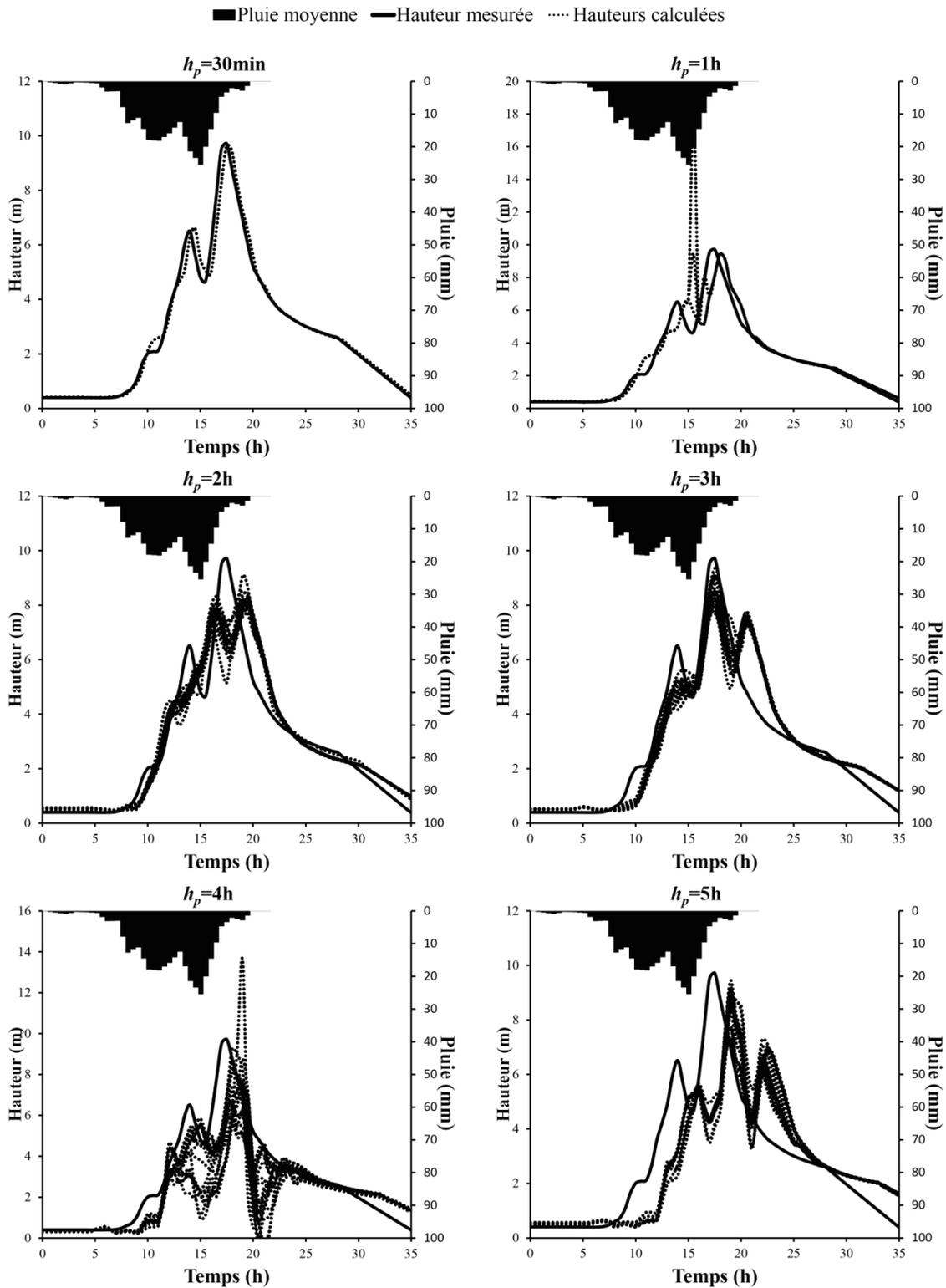


Figure 34: Évènement 19, les 20 hauteurs prévues par les 20 modèles, avec partie linéaire, sélectionnés pour chaque horizon de prévision

La Figure 34 présente les différents limnigrammes obtenus pour l'évènement 19. Sur l'ensemble des limnigrammes, on remarque que, contrairement aux modèles composés

uniquement d'un perceptron multicouche, il y a moins de hauteurs incohérentes : 18 m pour $h_p=1h$ et deux hauteurs négatives pour $h_p=4h$; même si elles sont tout autant gênantes, leur ordre de grandeur est moins aberrant comparé aux 60 m prévus avec les modèles sans partie linéaire.

Par ailleurs, les 20 prévisions proposées sont beaucoup moins dispersées avec les modèles disposant d'une partie linéaire. Pour $h_p=30min$ elles sont parfaitement superposées, pour $h_p=1h$ seules deux hauteurs se distinguent des autres, il n'y a que pour $h_p=4h$ que la dispersion est marquée.

Le Tableau 16 quantifie les remarques précédentes sur les limnigrammes avec l'étude des critères de persistance pour les deux types de modèles. Les critères de dispersion (écart-type, S_1 et S_2), gagnent six ordres de grandeur pour $h_p=30min$ et $h_p=1h$ lorsque l'on ajoute les termes directs, et un à deux ordres de grandeur pour les autres horizons de prévision. La partie linéaire réduit donc très efficacement l'influence de l'initialisation des paramètres du modèle pour les deux horizons de prévision les plus proches, et de manière significative pour les autres.

Ainsi les minima des critères de persistance pour les modèles avec partie linéaire ne sont jamais négatifs, les modèles donnent de meilleures prévisions que la prévision naïve, alors que les modèles à perceptron multicouche simple peuvent être beaucoup plus mauvais que le prédicteur naïf. Cependant pour les horizons les plus lointains, les meilleures prévisions sont fournies par les modèles sans partie linéaire ; l'arrêt précoce s'effectuant plus rapidement avec la partie linéaire, il ne permet pas à la fonction résultant du perceptron multicouche d'atteindre son optimum.

C_p		$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
Minimum	PM+PL	0,54	0,48	0,52	0,72	0,03	0,49
	PM	-37,51	-49,58	0,10	-0,67	-1,34	-11,12
Maximum	PM+PL	0,54	0,48	0,77	0,83	0,76	0,67
	PM	0,54	0,30	0,68	0,88	0,80	0,71
Moyenne	PM+PL	0,54	0,48	0,68	0,78	0,56	0,61
	PM	-2,56	-3,67	0,38	0,43	0,49	-0,06
Médiane	PM+PL	0,54	0,48	0,69	0,79	0,67	0,63
	PM	-0,14	0,01	0,45	0,52	0,61	0,57
Écart-type	PM+PL	$6,4 \cdot 10^{-6}$	$1,3 \cdot 10^{-6}$	$6,4 \cdot 10^{-2}$	$2,5 \cdot 10^{-2}$	$2,2 \cdot 10^{-1}$	$5,1 \cdot 10^{-2}$
	PM	8,3	$1,1 \cdot 10^1$	$1,7 \cdot 10^{-1}$	$4,3 \cdot 10^{-1}$	$4,7 \cdot 10^{-1}$	2,6
S_1	PM+PL	$4,2 \cdot 10^{-5}$	$1,3 \cdot 10^{-5}$	$3,8 \cdot 10^{-1}$	$1,3 \cdot 10^{-1}$	1,3	$3,0 \cdot 10^{-1}$
	PM	$-1,5 \cdot 10^1$	$-1,4 \cdot 10^1$	1,5	3,6	4,4	$-1,9 \cdot 10^2$
S_2	PM+PL	$1,2 \cdot 10^{-5}$	$2,7 \cdot 10^{-6}$	$9,5 \cdot 10^{-2}$	$3,2 \cdot 10^{-2}$	$3,9 \cdot 10^{-1}$	$8,3 \cdot 10^{-2}$
	PM	-3,3	-3,1	$4,5 \cdot 10^{-1}$	1,0	$9,5 \cdot 10^{-1}$	$-4,1 \cdot 10^1$

Tableau 16 : Pour chaque horizon de prévision, sur les 20 prévisions de l'événement 19, calcul du minimum, du maximum, de la moyenne, de la médiane, de l'écart-type, des critères S_1 et S_2 des critères de persistance. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire (PM+PL).

3.5.3.2. Événement 26

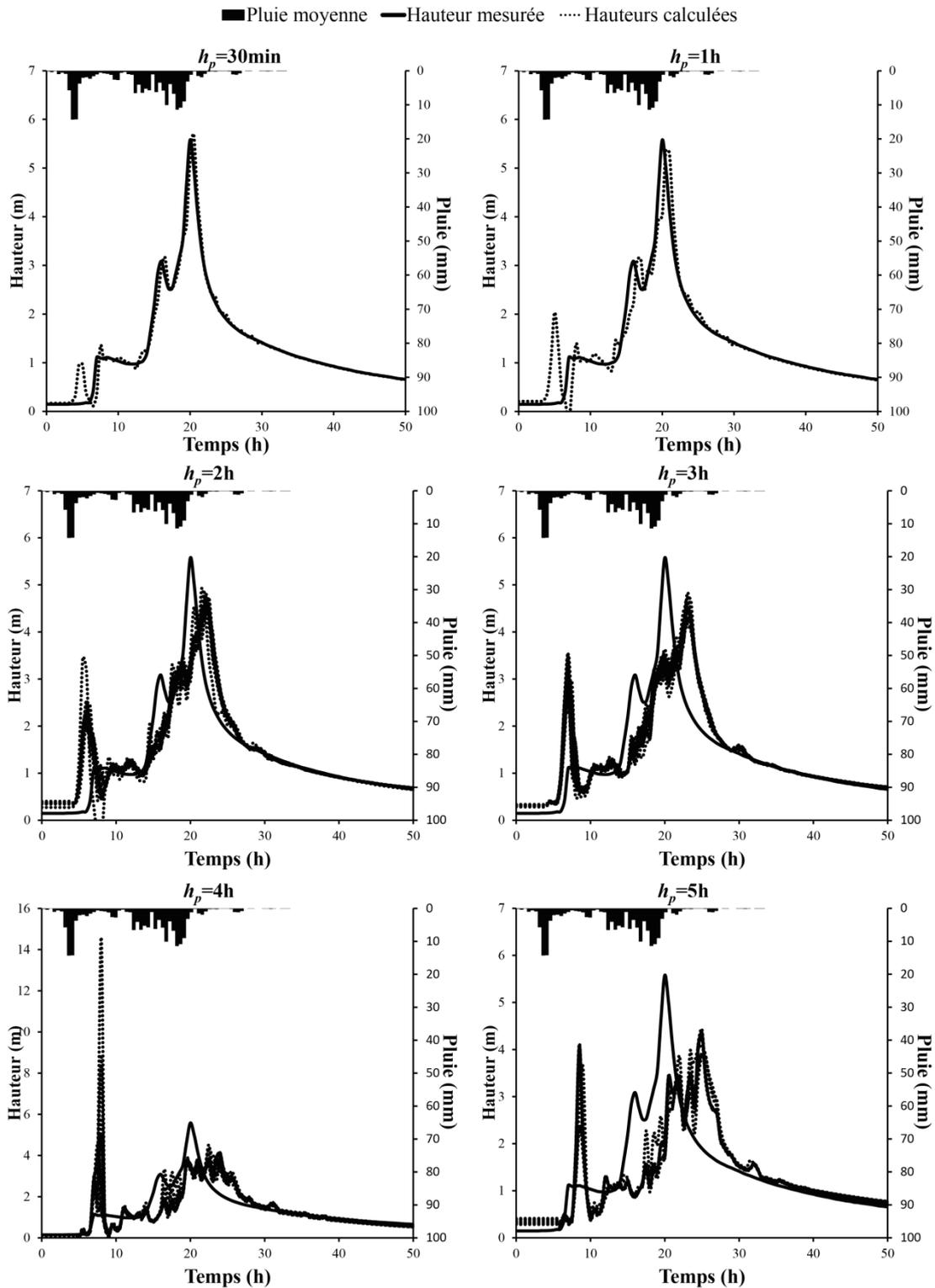


Figure 35: Événement 26, les 20 hauteurs prévues par les 20 modèles, avec partie linéaire, sélectionnés pour chaque horizon de prévision

Les différents limnigrammes prévus par les 20 modèles pour les différents horizons de prévision sont compilés dans la Figure 35. La dispersion des 20 hauteurs prévues est très faible : elles sont même superposées pour $h_p=30\text{min}$ et $h_p=1\text{h}$.

Les valeurs des critères de persistance répertoriées dans le Tableau 17 confirment la dispersion très réduite des hauteurs prévues par les modèles avec partie linéaire en comparaison des modèles à perceptron multicouche seul. Pour $h_p=30\text{min}$ et $h_p=1\text{h}$ les valeurs de l'écart type, de S_1 et S_2 sont jusqu'à un million de fois plus petits dans le cas des modèles avec partie linéaire, ce facteur varie autour de 2 et 10 pour les autres horizons de prévision, excepté pour $h_p=4\text{h}$ où la dispersion est moins importante pour les modèles sans partie linéaire.

Hormis pour $h_p=4\text{h}$, toutes les prévisions fournies par les modèles avec partie linéaire sont meilleures que les prévisions naïves, contrairement aux prévisions provenant des modèles sans partie linéaire. En contrepartie, les meilleures prévisions des modèles sans partie linéaire sont de qualité supérieure à celles des modèles avec partie linéaire sauf pour $h_p=2\text{h}$.

C_p		$h_p=0,5\text{h}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Minimum	PM+PL	0,32	0,23	0,07	0,30	-1,50	0,27
	PM	-1,06	-5,65	-0,03	-0,09	0,29	0,01
Maximum	PM+PL	0,32	0,23	0,38	0,46	0,50	0,32
	PM	0,38	0,33	0,26	0,47	0,51	0,41
Moyenne	PM+PL	0,32	0,23	0,29	0,38	0,26	0,30
	PM	0,02	-0,44	0,13	0,27	0,39	0,32
Médiane	PM+PL	0,32	0,23	0,31	0,39	0,39	0,30
	PM	0,22	0,08	0,13	0,29	0,40	0,35
Écart-type	PM+PL	$2,8 \cdot 10^{-6}$	$7,6 \cdot 10^{-7}$	$8,6 \cdot 10^{-2}$	$4,3 \cdot 10^{-2}$	$4,5 \cdot 10^{-1}$	$1,1 \cdot 10^{-2}$
	PM	$4,4 \cdot 10^{-1}$	1,4	$7,3 \cdot 10^{-2}$	$1,3 \cdot 10^{-1}$	$6,5 \cdot 10^{-2}$	$9,8 \cdot 10^{-2}$
S_1	PM+PL	$3,5 \cdot 10^{-5}$	$1,5 \cdot 10^{-5}$	1,1	$4,3 \cdot 10^{-1}$	7,7	$1,7 \cdot 10^{-1}$
	PM	$7,4 \cdot 10^1$	$-1,4 \cdot 10^1$	2,2	2,1	$5,7 \cdot 10^{-1}$	1,3
S_2	PM+PL	$8,8 \cdot 10^{-6}$	$3,3 \cdot 10^{-6}$	$3,0 \cdot 10^{-1}$	$1,1 \cdot 10^{-1}$	1,7	$3,7 \cdot 10^{-2}$
	PM	$2,3 \cdot 10^1$	-3,3	$5,5 \cdot 10^{-1}$	$5,1 \cdot 10^{-1}$	$1,7 \cdot 10^{-1}$	$3,1 \cdot 10^{-1}$

Tableau 17 : Pour chaque horizon de prévision, sur les 20 sorties de l'événement 26, calcul du minimum, du maximum, de la moyenne, de la médiane, de l'écart-type, des critères S_1 et S_2 des critères de persistance. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire (PM+PL).

3.5.3.3. Événement 27

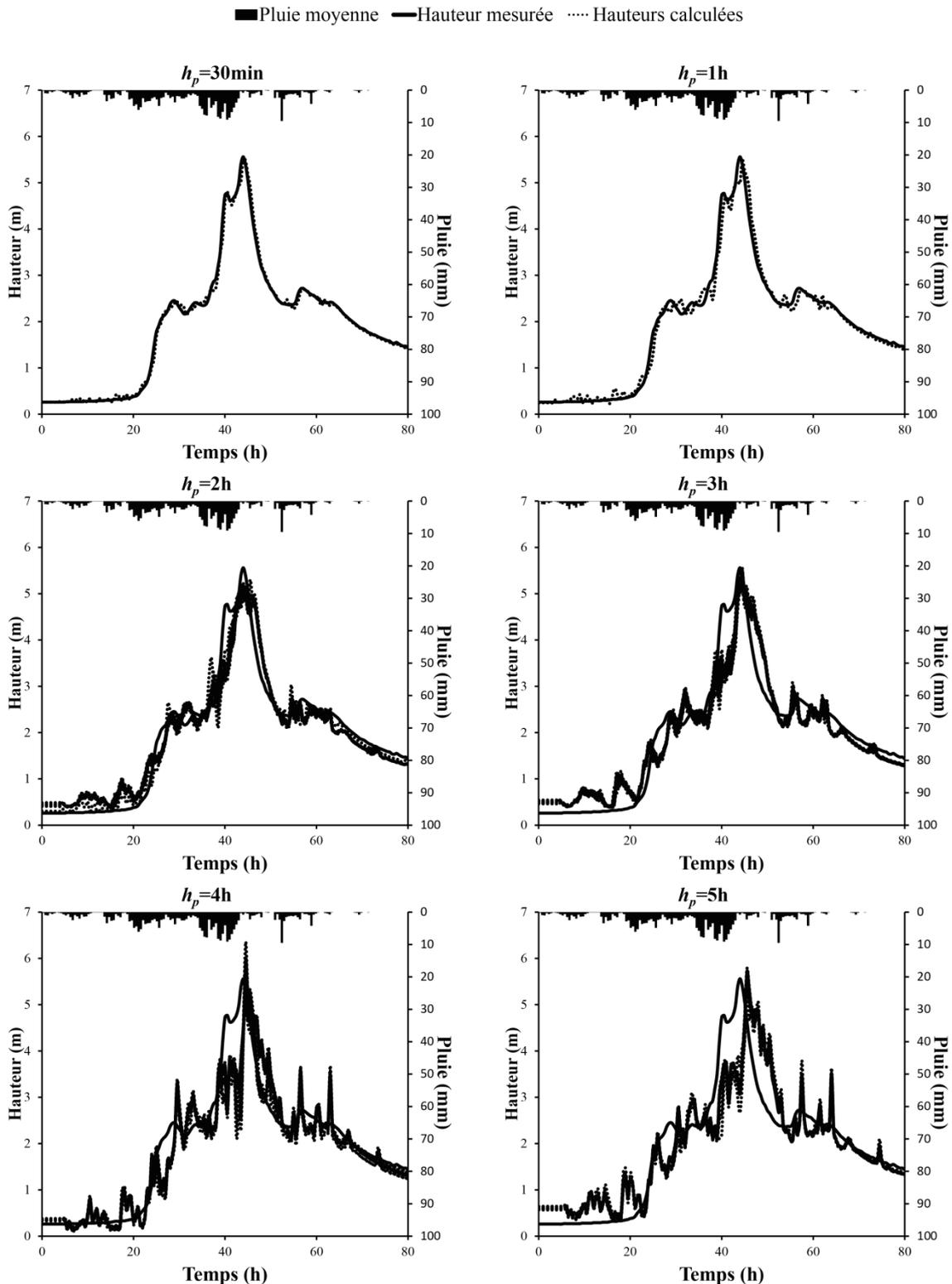


Figure 36: Événement 27, les 20 hauteurs prévues par les 20 modèles, avec partie linéaire, sélectionnés pour chaque horizon de prévision

On trouve sur la Figure 36 les 20 prévisions pour l'événement 27 suivant les horizons de prévision. Comme pour les deux autres événements en test, ces prévisions sont beaucoup moins dispersées que celles issues de modèles sans partie linéaire. Pour l'événement 27, les

20 hauteurs prévues sont très proches quel que soit l'horizon de prévision, et aucune hauteur aberrante n'apparaît.

Concernant les valeurs du critère de persistance dans le Tableau 18, la différence de dispersion des sorties des modèles avec et sans partie linéaire est très marquée pour $h_p=30\text{min}$ et $h_p=1\text{h}$ avec un rapport de 10^5 sur l'écart-type, S_1 et S_2 . Pour les autres horizons de prévision, ce facteur varie entre 2 et 10.

On note à nouveau que les valeurs du minima des critères de persistance sont moins mauvaises pour les prévisions des modèles avec partie linéaire, alors que les meilleures prévisions issues des modèles sans partie linéaire sont en général de qualité plus grande que celles des modèles avec partie linéaire.

C_p		$h_p=0,5\text{h}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Minimum	PM+PL	0,58	0,52	0,38	0,41	0,43	0,36
	PM	-0,22	-0,18	0,24	0,35	0,21	0,28
Maximum	PM+PL	0,58	0,52	0,45	0,52	0,60	0,52
	PM	0,55	0,56	0,75	0,52	0,57	0,58
Moyenne	PM+PL	0,58	0,52	0,41	0,47	0,54	0,48
	PM	0,26	0,25	0,44	0,44	0,46	0,48
Médiane	PM+PL	0,58	0,52	0,40	0,48	0,55	0,49
	PM	0,43	0,22	0,44	0,44	0,49	0,51
Ecart type	PM+PL	$9,7 \cdot 10^{-6}$	$2,3 \cdot 10^{-6}$	$2,1 \cdot 10^{-2}$	$2, \cdot 10^{-2}$	$5,3 \cdot 10^{-2}$	$4,3 \cdot 10^{-2}$
	PM	$2,9 \cdot 10^{-1}$	$2,1 \cdot 10^{-1}$	$9,9 \cdot 10^{-2}$	$4,4 \cdot 10^{-2}$	$9,1 \cdot 10^{-2}$	$7,9 \cdot 10^{-2}$
S_1	PM+PL	$6,4 \cdot 10^{-5}$	$2,2 \cdot 10^{-5}$	$1,8 \cdot 10^{-1}$	$2,4 \cdot 10^{-1}$	$3,2 \cdot 10^{-1}$	$3,5 \cdot 10^{-1}$
	PM	3,0	3,0	1,2	$3,8 \cdot 10^{-1}$	$7,8 \cdot 10^{-1}$	$6,2 \cdot 10^{-1}$
S_2	PM+PL	$1,7 \cdot 10^{-5}$	$4,3 \cdot 10^{-6}$	$5,1 \cdot 10^{-2}$	$5,7 \cdot 10^{-2}$	$9,9 \cdot 10^{-2}$	$9,1 \cdot 10^{-2}$
	PM	1,1	$8,6 \cdot 10^{-1}$	$2,2 \cdot 10^{-1}$	$1,0 \cdot 10^{-1}$	$2,0 \cdot 10^{-1}$	$1,6 \cdot 10^{-1}$

Tableau 18 : Pour chaque horizon de prévision, sur les 20 sorties de l'événement 27, calcul du minimum, du maximum, de la moyenne, de la médiane, de l'écart-type, des critères S_1 et S_2 des critères de persistance. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire ajoutée (PM+PL).

3.5.4. Médiane des sorties

De la même manière que précédemment, nous allons évaluer la qualité des prévisions obtenues en choisissant à chaque pas de temps la médiane des 20 hauteurs proposées à chaque horizon de prévision.

On comparera les résultats provenant des modèles avec partie linéaire à ceux issus des modèles à perceptron multicouche sans termes directs.

3.5.4.1. Évènement 19

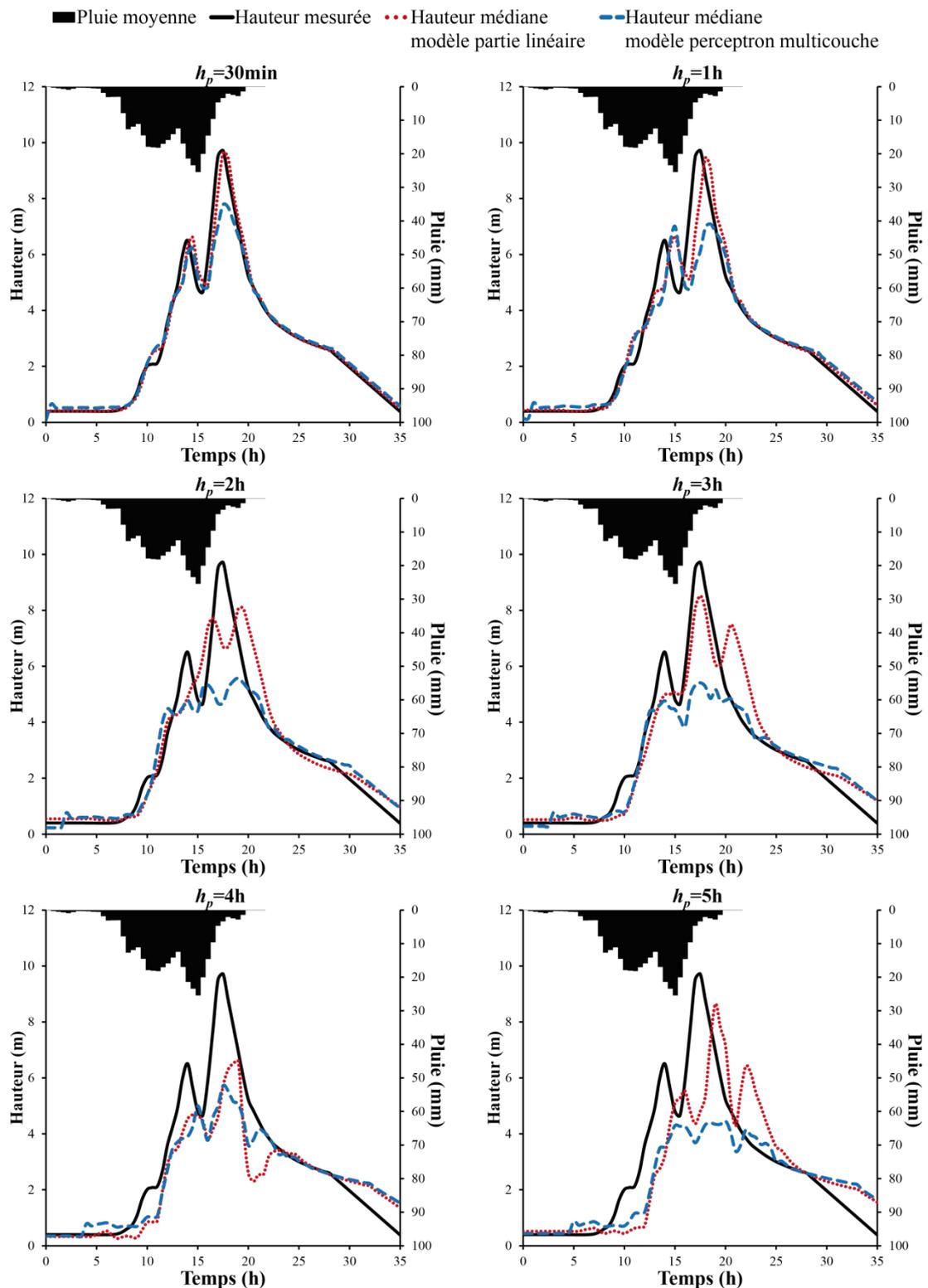


Figure 37: Évènement 19, hauteurs médianes sur les hauteurs prévues par les 20 modèles avec partie linéaire et les 20 modèles à perceptron multicouche sélectionnés pour chaque horizon de prévision

Si l'on compare sur la Figure 37 les hauteurs prévues par les deux types de modèles, avec et sans partie linéaire, on constate que les modèles à partie linéaire prévoient beaucoup mieux la hauteur d'eau maximale. L'ajout de la partie linéaire permet de mieux modéliser la

composante linéaire du processus importante pour cet événement où les hauteurs en jeu sont beaucoup plus grandes que celles de la base d'apprentissage.

Concernant les hauteurs prévues par les modèles avec partie linéaire, pour $h_p=30\text{min}$, la prévision est excellente, le pic prévu est synchrone avec celui observé avec un *PPD* de 99%, comme reporté dans le Tableau 19. Pour $h_p=1\text{h}$, le pic prévu a un retard de 30min par rapport à celui observé, mais le *SPPD* de 83% donne une information intéressante sur la valeur de la hauteur. Pour $h_p=2\text{h}$ et $h_p=3\text{h}$, la forme du limnigramme prévu est différente de celle du limnigramme observé, ce qui n'empêche pas d'avoir une prévision du pic convenable avec des valeurs de *SPPD* appréciables notamment de 88% pour $h_p=3\text{h}$.

Pour $h_p=4\text{h}$, le choix du modèle avec partie linéaire est discutable : le *PPD* est meilleur, mais le pic prévu arrive avec 1h30min de retard alors qu'il était synchrone pour le modèle à perceptron multicouche. Pour $h_p=5\text{h}$, la hauteur prévue par le modèle à partie linéaire redevient plus pertinente car à la fois on gagne en amplitude et, comparé à l'autre modèle, on diminue le retard d'une heure par rapport au pic observé. Ce retard étant constant sur l'ensemble du limnigramme et largement plus petit que l'horizon de prévision, ce résultat peut être exploitable par les prévisionnistes.

Médianes		$h_p=0,5\text{h}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
C_P	PM+PL	0,54	0,48	0,70	0,81	0,68	0,65
	PM	0,00	0,08	0,46	0,64	0,69	0,62
C_N	PM+PL	0,98	0,92	0,88	0,88	0,73	0,59
	PM	0,96	0,86	0,78	0,78	0,73	0,56
<i>PPD</i> (%)	PM+PL	99	97	83	88	68	89
	PM	80	73	57	56	59	46
<i>SPPD</i> (%)	PM+PL	99	83	69	88	59	49
	PM	80	63	48	56	59	41
Retard H_{pic} (h)	PM+PL	0	0,5	2	0	1,5	1,5
	PM	0	1	1,5	0	0	2,5

Tableau 19 : Valeurs du C_P , C_N , *PPD*, *SPPD* et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 19. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire ajoutée (PM+PL).

3.5.4.2. Événement 26

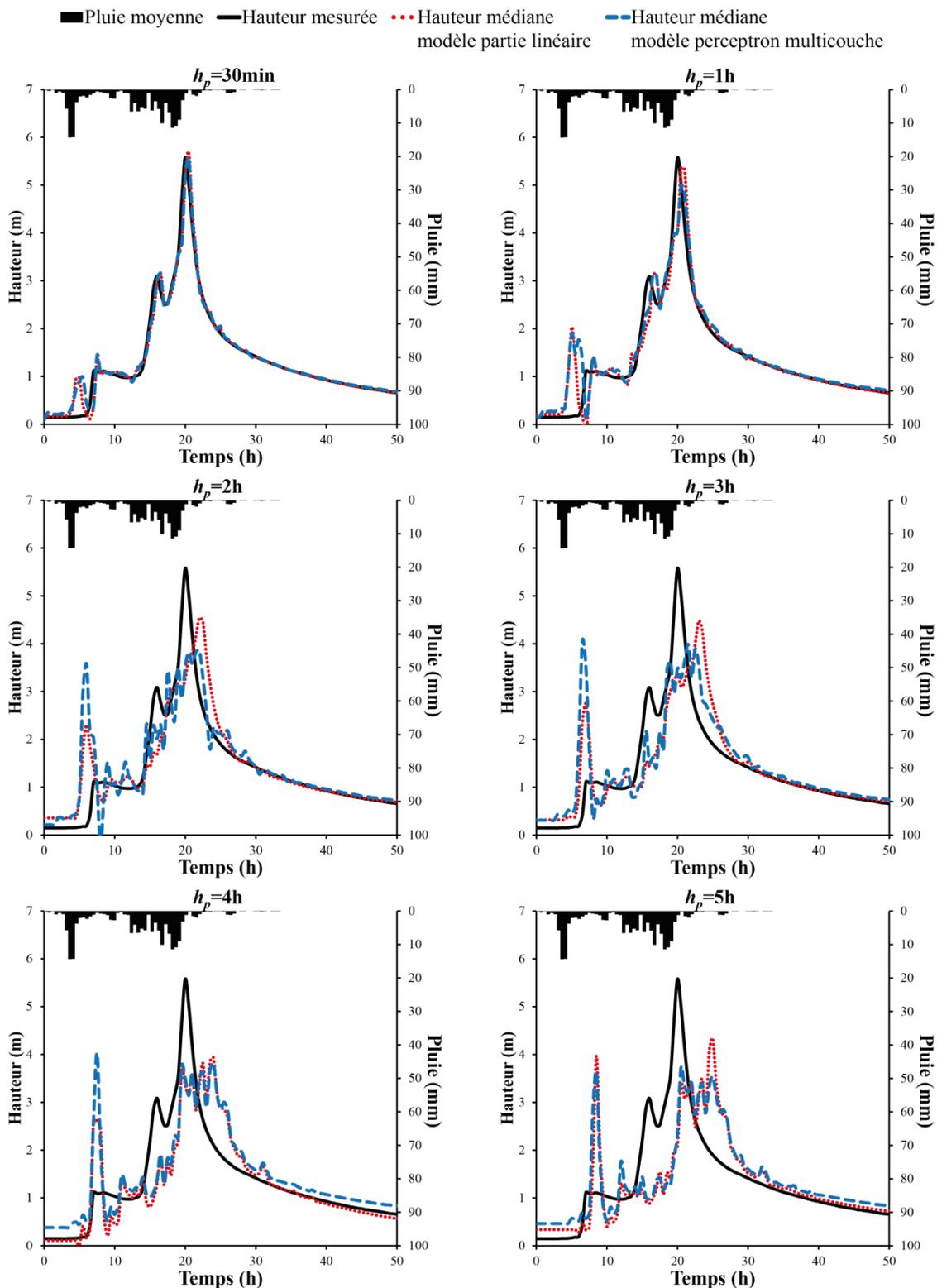


Figure 38: Événement 26, hauteurs médianes sur les hauteurs prévues par les 20 modèles, avec partie linéaire, et les 20 modèles à perceptron multicouche sélectionnés pour chaque horizon de prévision

La Figure 38 montre les hauteurs prévues à partir des deux types de modèles pour les différents horizons de prévision, et le Tableau 20 renseigne sur les valeurs des critères de

performance. On voit sur les limnigrammes que les modèles avec partie linéaire donnent une meilleure amplitude au niveau du pic, les oscillations en réponses aux brusques variations d'intensité de pluie sont moins importantes, en particulier celle causant un faux pic en début d'événement, qui perd jusqu'à 1m par rapport aux résultats des modèles sans partie linéaire.

Pour l'événement 26, la différence entre les résultats des deux types de modèles est bien moins marquée que pour l'événement 19. Pour $h_p=30\text{min}$ et $h_p=1\text{h}$, les deux prévisions de qualité très satisfaisante sont très proches avec toutefois un meilleur *PPD* pour les modèles avec partie linéaire ; la même remarque s'applique pour $h_p=2\text{h}$ et $h_p=4\text{h}$ avec des prévisions de qualité insuffisante mais donnant des indications au prévisionniste. Pour $h_p=3\text{h}$ et $h_p=5\text{h}$, la préférence d'un type de modèle sur l'autre n'est pas évidente, car les modèles avec partie linéaire donnent un meilleur *PPD*, alors que les modèles sans partie linéaire présentent un retard du pic prévu par rapport à au pic observé moins important avec un meilleur *SPPD*. Quoiqu'il en soit, pour $h_p=5\text{h}$, la montée du pic prévu est trop éloignée de celle du pic observé pour que les prévisions soient exploitables.

Médianes		$h_p=0,5\text{h}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
C_P	PM+PL	0,32	0,23	0,32	0,39	0,46	0,30
	PM	0,31	0,23	0,17	0,33	0,44	0,36
C_N	PM+PL	0,96	0,86	0,69	0,60	0,57	0,28
	PM	0,96	0,86	0,62	0,56	0,55	0,35
<i>PPD</i> (%)	PM+PL	102	96	81	80	70	78
	PM	~100	89	69	71	69	67
<i>SPPD</i> (%)	PM+PL	93	72	59	61	60	37
	PM	93	72	63	63	63	42
Retard	PM+PL	0,5	0,5	2	3	4	5
H_{pic} (h)	PM	0,5	0,5	2	1,5	4	0,5

Tableau 20 : Valeurs du C_P , C_N , *PPD*, *SPPD* et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 26. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire ajoutée (PM+PL).

3.5.4.3. Événement 27

Pluie moyenne
 Hauteur mesurée
 Hauteur médiane modèle partie linéaire
 Hauteur médiane modèle perceptron multicouche

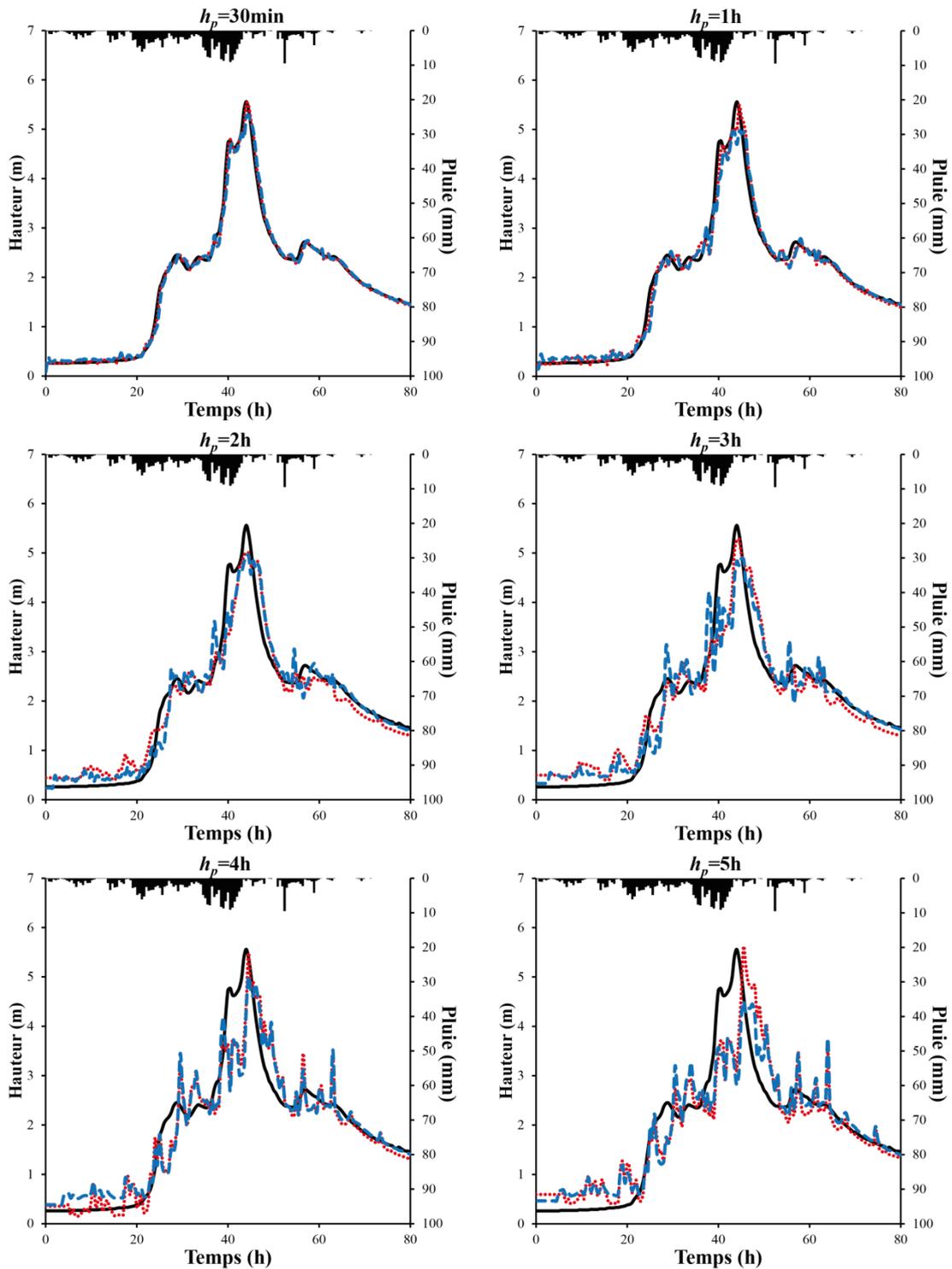


Figure 39: Événement 27, hauteurs médianes sur les hauteurs prévues par les 20 modèles avec partie linéaire et les 20 modèles à perceptron multicouche sélectionnés pour chaque horizon de prévision

Les hauteurs prévues pour l'événement 27 à partir des deux types de modèles et les critères de performance correspondants se trouvent Figure 39 et Tableau 21. La différence de résultat entre les deux hauteurs proposées à chaque horizon de prévision est peu visible sur les limnigrammes ; on constate une légère diminution des oscillations causées par les variations du hyétogramme, mais ce lissage est moindre que celui obtenu pour l'événement 26. On note que le *PPD* et le *SPPD* sont meilleurs pour les modèles avec partie linéaire de même que la synchronisation entre le pic prévu et le pic observé, même si les critères de Nash et de persistance sont plus élevés pour les modèles sans partie linéaire pour $h_p=2h$ et $h_p=5h$.

On retient que les prévisions sont excellentes pour $h_p=30min$ et $h_p=1h$, satisfaisantes pour $h_p=2h$ et $h_p=3h$ et le retard dans la montée du pic prévu par rapport au pic observé les rend plus difficilement utilisables pour $h_p=4h$ et surtout $h_p=5h$ malgré un *PPD* de l'ordre de 100%.

Médianes		$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
C_P	PM+PL	0,58	0,52	0,42	0,49	0,58	0,48
	PM	0,53	0,48	0,52	0,49	0,52	0,54
C_N	PM+PL	~1	0,98	0,94	0,89	0,87	0,77
	PM	~1	0,98	0,95	0,90	0,85	0,80
<i>PPD</i> (%)	PM+PL	~100	99	90	~100	99	101
	PM	96	90	90	88	91	81
<i>SPPD</i> (%)	PM+PL	~100	90	90	95	80	55
	PM	95	85	85	87	79	51
Retard	PM+PL	0	0,5	0,5	0,5	0,5	1,5
H_{pic} (h)	PM	0,5	0,5	0,5	1	0,5	3,5

Tableau 21 : Valeurs du C_P , C_N , *PPD*, *SPPD* et Retard H_{pic} suivant l'horizon de prévision pour la hauteur prévue médiane des sorties sur l'événement 19. Comparaison entre les modèles à perceptron multicouche uniquement (PM) et ceux avec une partie linéaire ajoutée (PM+PL).

3.5.5. Conclusion sur les modèles avec une partie linéaire

Dans cette partie, nous avons comparé les résultats obtenus en test par les deux types de modèles utilisés : un perceptron multicouche standard et un perceptron multicouche muni de termes directs.

La différence entre les prévisions obtenues par ces deux modèles est surtout remarquable pour l'événement 19 où la qualité est significativement augmentée jusqu'à l'horizon de prévision de 3h avec l'ajout de la partie linéaire. Cette amélioration est liée à une meilleure amplitude du pic prévu. La partie linéaire a permis au modèle de répondre de manière très intense à l'occurrence d'une impulsion très forte de précipitation, pour cet événement dont la hauteur maximale dépasse notablement celles de la base d'apprentissage (30% de plus que la plus grande hauteur en apprentissage). En revanche, la partie non linéaire est nécessaire pour une bonne montée de crue lorsque les précipitations ne sont pas trop intenses et aussi pour assurer une bonne décrue.

Pour les événements 26 et 27, la préférence d'un modèle sur l'autre est moins évidente à déterminer. Les différences entre les deux types de modèles sont relativement faibles. Pour ces deux événements, les modèles avec partie linéaire donnent des prévisions avec une

amplitude de pic très proche du pic observé. Pour les horizons de prévisions les plus lointains, la synchronisation du pic prévu par rapport au pic observé est meilleure avec les modèles sans partie linéaire pour le 26 et pour les modèles avec partie linéaire dans le cas du 27.

On proposera ainsi de privilégier les modèles avec partie linéaire pour les horizons de prévision jusqu'à 2h et d'étudier les hauteurs prévues par les deux types de modèles au-delà.

3.6. Conclusion sur les modèles pour la prévision des crues à Anduze

Dans ce chapitre, nous nous sommes attachés à appliquer la méthodologie de conception d'un modèle de prévision neuronal présentée au Chapitre 1 à la prévision des crues éclair décrite au Chapitre 2. À cette fin, nous avons élaboré des modèles à base de perceptrons multicouche, qui calculent la prévision de la hauteur à Anduze pour différents horizons de prévision. Contre toute attente, et vraisemblablement du fait de l'utilisation rigoureuse de méthodes de régularisation, ce ne sont ni l'architecture du réseau ni le nombre de variables d'entrées qui ont l'influence la plus déterminante sur l'estimation de la capacité de généralisation des modèles, mais l'initialisation des paramètres du réseau. Cette estimation joue un rôle fondamental sur les résultats en test. Ainsi il est apparu que si de bons prédicteurs existaient et pouvaient être obtenus par apprentissage, ils n'étaient pas sélectionnés par la méthode de validation croisée partielle. Nous avons choisi de travailler avec les 20 meilleurs modèles issus de la validation croisée et de retenir la médiane des sorties des modèles comme prévision à chaque pas de temps.

L'analyse de la dispersion des estimations de la capacité de généralisation des modèles en fonction de l'initialisation de leurs paramètres nous a conduits à proposer l'ajout d'une partie linéaire au perceptron multicouche. Ceci a permis de limiter considérablement la dispersion des prévisions en test en fonction de l'initialisation des paramètres lors de l'apprentissage, en particulier pour les petits horizons de prévision. Cependant ce type de modèle n'améliore pas toujours la synchronisation entre pic prévu et observé pour les horizons de prévision les plus lointains. Nous proposons donc de conserver les prévisions des deux types de modèles. Une piste de recherche intéressante consistera à étudier la sensibilité de cette dispersion à l'absence de disponibilité de pluies futures, à l'aide des « pluies parfaites » (pluies futures observées mises à disposition des modèles) par exemple ; en effet cette lacune importante d'information pourrait être la source de la dispersion des performances des modèles en test.

Les différents modèles réalisés dans ce chapitre permettent de fournir des prévisions de la hauteur à Anduze de bonne qualité pour $h_p=30\text{min}$ et $h_p=1\text{h}$, utilisables avec certaines précautions pour $h_p=2\text{h}$ et $h_p=3\text{h}$ et qui peuvent servir d'indications pour $h_p=4\text{h}$ et $h_p=5\text{h}$. Il faut cependant tenir compte que les performances des modèles sont dépendantes de l'intensité et la rapidité des événements : ainsi les prévisions de l'événement 27 qui est intense mais assez lent sont bien meilleures que celles sur l'événement 19 qui est beaucoup plus intense et plus rapide.

4. Introduction de l'adaptativité dans les modèles de prévision des crues sur le bassin versant du Gardon d'Anduze

4.1. Introduction

Dans le chapitre précédent, différents modèles pour la prévision des crues sur le bassin versant du Gardon d'Anduze à Anduze ont été présentés. On a pu constater que les modèles issus d'une sélection par validation croisée partielle ont des performances de qualité très variable en test, dépendant en particulier de l'initialisation des paramètres du réseau avant la phase d'apprentissage. Afin de réduire l'écart de performance entre les résultats sur l'ensemble d'apprentissage et ceux obtenus en test, on va dans ce chapitre doter les modèles déjà créés d'une capacité d'adaptation : les paramètres des réseaux seront modifiés au cours de l'événement testé, on adaptera ainsi le modèle en « temps réel ».

4.2. Principe de l'adaptativité

Les paramètres des réseaux utilisés dans le Chapitre 3 ont été ajustés de sorte à approcher au mieux les données de l'ensemble d'apprentissage tout en satisfaisant les conditions de généralisation imposées par l'arrêt précoce. De même, la sélection des modèles à partir de la validation croisée a mis à contribution l'ensemble d'apprentissage et celui d'arrêt précoce. Cependant, les résultats obtenus en test montrent que les choix des hyperparamètres et en particulier de l'initialisation des paramètres du réseau ne conduisent pas à des prévisions satisfaisantes pour les événements en test.

En d'autres termes, des modèles sélectionnés avec des scores de validation croisée très proches, conduisent à des performances en test très inégales [Bornancin-Plantier 2011]. Or, si l'on recherche quel est le meilleur résultat en test possible (en effectuant la sélection de modèles sur la performance en test), il apparaît qu'il peut exister des modèles plus performants (Annexe 5), ce qui est particulièrement notable pour l'événement 19, mais moins apparent pour les événements 26 et 27.

On peut comprendre que si ces modèles sont optimaux pour un événement spécifique, et lui seul, ils conduisent effectivement à des scores assez médiocres en validation croisée, tout en étant d'un grand intérêt potentiel pour l'alerte en cours d'événement pluvieux. En particulier ceci est possible si chaque événement correspond à un fonctionnement particulier du bassin versant qui ne se retrouve pas dans l'ensemble de la base d'apprentissage.

Il vient alors naturellement à l'idée d'étudier la possibilité de modifier, en temps réel, les paramètres du modèle sélectionné à l'issue de la validation croisée, afin de le spécialiser sur l'événement en cours de déroulement au moyen d'un apprentissage adaptatif. Notons que l'adaptation des paramètres du réseau en fonction de mesures effectuées en temps réel s'appelle en météorologie et hydrologie l'*assimilation de données*. L'assimilation de données est en fait plus large que l'adaptativité car elle permet de modifier à la fois les paramètres, les initialisations des modèles (pour les modèles dynamiques, ce qui n'est pas le cas dans le présent travail) ainsi que leurs variables. En ce qui concerne ce chapitre, nous allons restreindre notre étude à la modification des paramètres du réseau au cours de l'événement en test, pour spécialiser le réseau sur cet événement et ainsi réaliser de meilleures prévisions.

Si l'on se place dans des conditions d'utilisation en temps réel, à l'instant k , on cherche à prévoir la hauteur à l'instant $k+h_p$. On dispose des données de pluie et de hauteur jusqu'à cet instant k inclus. On va utiliser ces données sur une période F_a , fenêtre d'adaptativité, pour modifier les paramètres du réseau.

Considérons que l'on cherche à adapter, par apprentissage, un modèle existant de prévision à l'horizon h_p . Pour effectuer l'apprentissage, il convient donc de disposer des valeurs observées de hauteur à prévoir, c'est-à-dire avec une anticipation de h_p par rapport à l'instant courant. Ceci implique que si l'on se situe à l'instant k , on ne peut adapter le modèle qu'en utilisant des données observées entre $k - h_p$ et $k - h_p - F_a$. Un laps de temps de $2 h_p$ sépare l'instant où les mesures de précipitations et de hauteur peuvent être utilisées, de celles auxquelles la prévision devra être confrontée. Nous verrons dans la suite de ce chapitre que cet écart de temps va considérablement handicaper les capacités d'adaptation des modèles. La Figure 40 présente les différentes fenêtres considérées durant l'apprentissage adaptatif.

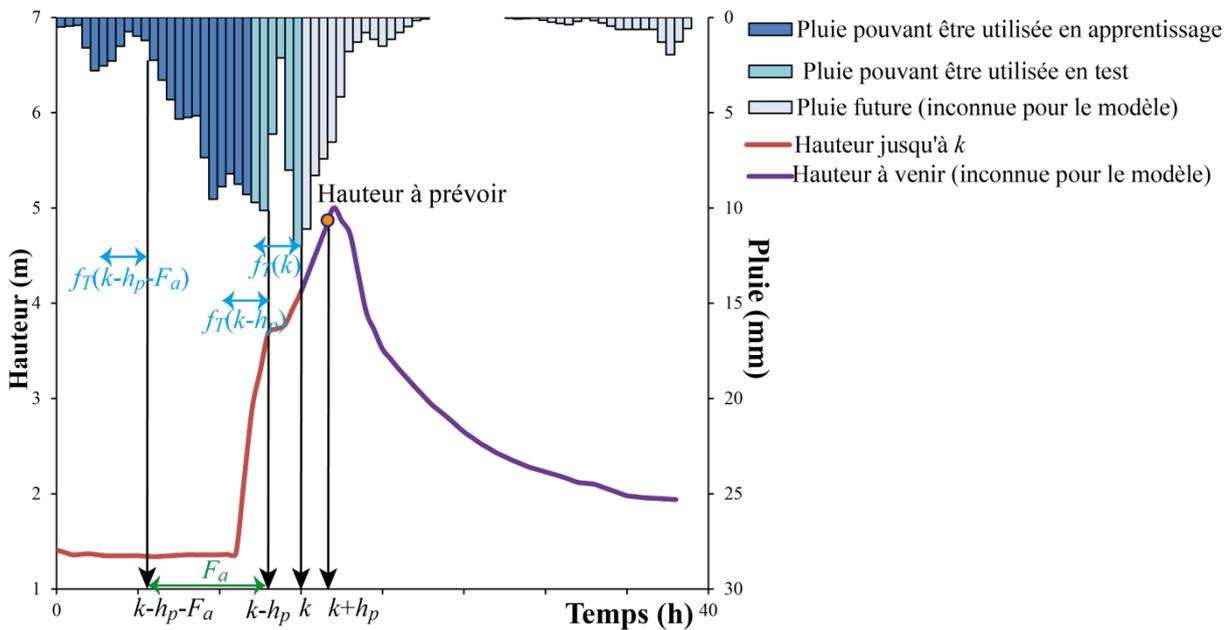


Figure 40 : Schéma des données nécessaires pour réaliser les modèles adaptatifs, f_T étant la fenêtre temporelle des pluies et F_a la fenêtre d'adaptativité

Pour chaque horizon de prévision, on part d'un modèle avec l'architecture d'un perceptron multicouche, sélectionné comme nous l'avons décrit dans la section 3 du chapitre 3. On choisira comme modèle initial le modèle qui avait obtenu le meilleur score de validation croisée partielle. On modifie les paramètres du réseau à chaque pas de temps comme illustré par la Figure 41, en réalisant un nouvel apprentissage sur la fenêtre d'adaptativité de taille F_a .

L'algorithme utilisé pour l'étape d'apprentissage adaptatif est celui de Levenberg-Marquardt et sans méthode de régularisation. On constate que le nombre d'itérations nécessaires à la minimisation de la fonction de coût est faible (<10), ce qui est soit la conséquence du nombre réduit d'exemples présentés en apprentissage adaptatif : l'ajustement

des paramètres n'est pas assez contraint, soit le modèle existant est très proche du modèle modifié.

L'apprentissage, à chaque présentation d'un nouvel exemple au modèle, en situation de temps réel simulé, est réalisé sur une fenêtre glissante d'exemples (les valeurs précédentes de précipitations et de débits observés) de longueur à préciser : il s'agit de la fenêtre que nous avons appelée fenêtre d'adaptativité. Il convient donc de déterminer la fenêtre d'adaptativité la plus pertinente en appliquant les modèles adaptatifs aux événements 19, 26 et 27. La procédure suivie est présentée en Figure 41.

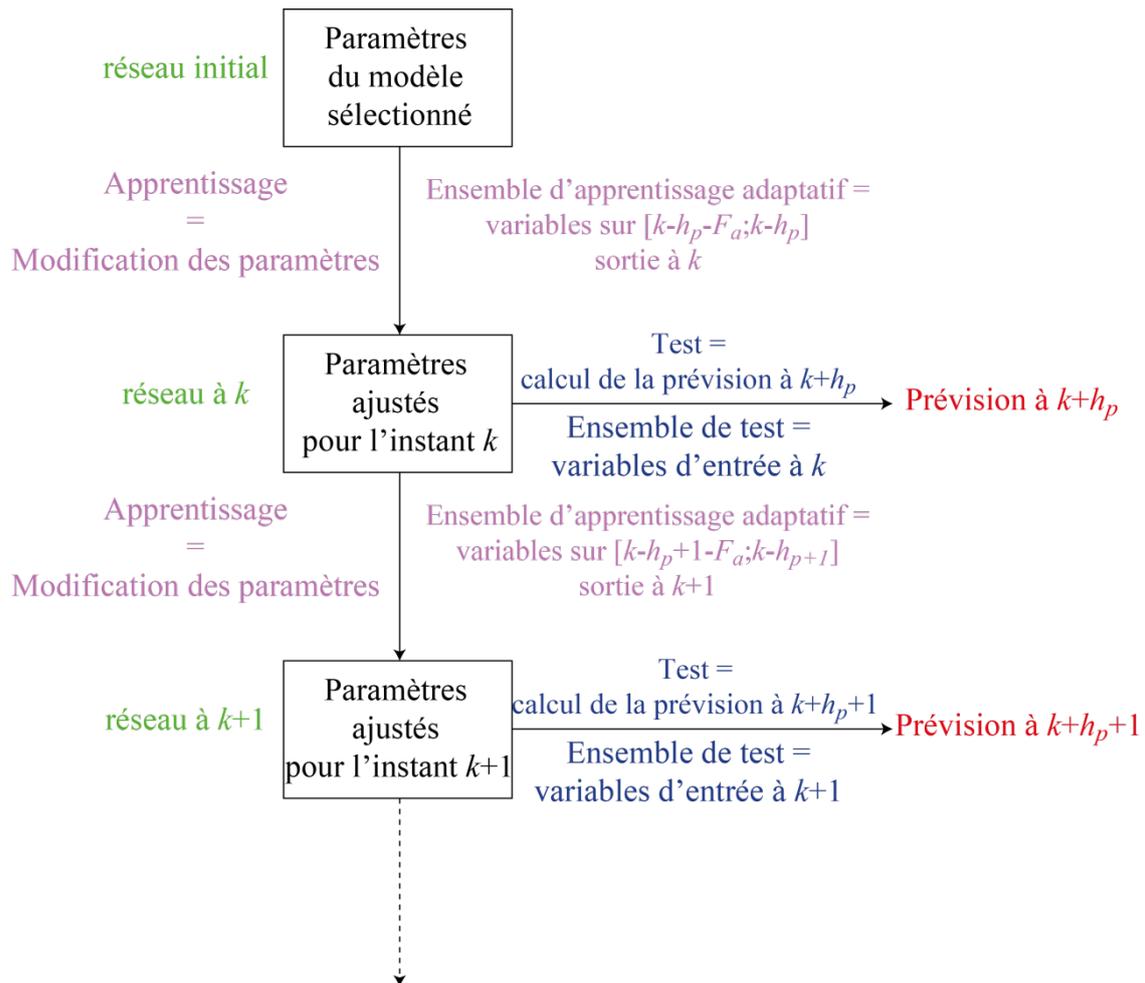


Figure 41 : Schéma expliquant la modification des paramètres au cours de l'événement étudié

4.3. Sélection de la fenêtre d'adaptativité

Comme nous avons supposé que le fonctionnement du bassin versant est différent pour chaque événement, l'application de l'adaptativité sera étudiée pour chacun des trois événements considérés en test : les événements 19, 26 et 27.

4.3.1. Évènement 19

Considérons tout d'abord l'évènement 19 ; pour chaque horizon de prévision, on réalise des prévisions pour des modèles modifiés suivant différentes fenêtres d'adaptativité : 30 min (1 exemple en apprentissage), 1h (2 exemples en apprentissage), 2h (4 exemples en

apprentissage), 3h (6 exemples en apprentissage) et 4h (8 exemples en apprentissage). La Figure 42, représente les limnigrammes des prévisions pour l'événement 19 et le Tableau 22, les critères de persistance et les *SPDD* correspondants, les valeurs des autres critères sont précisées en Annexe 6.

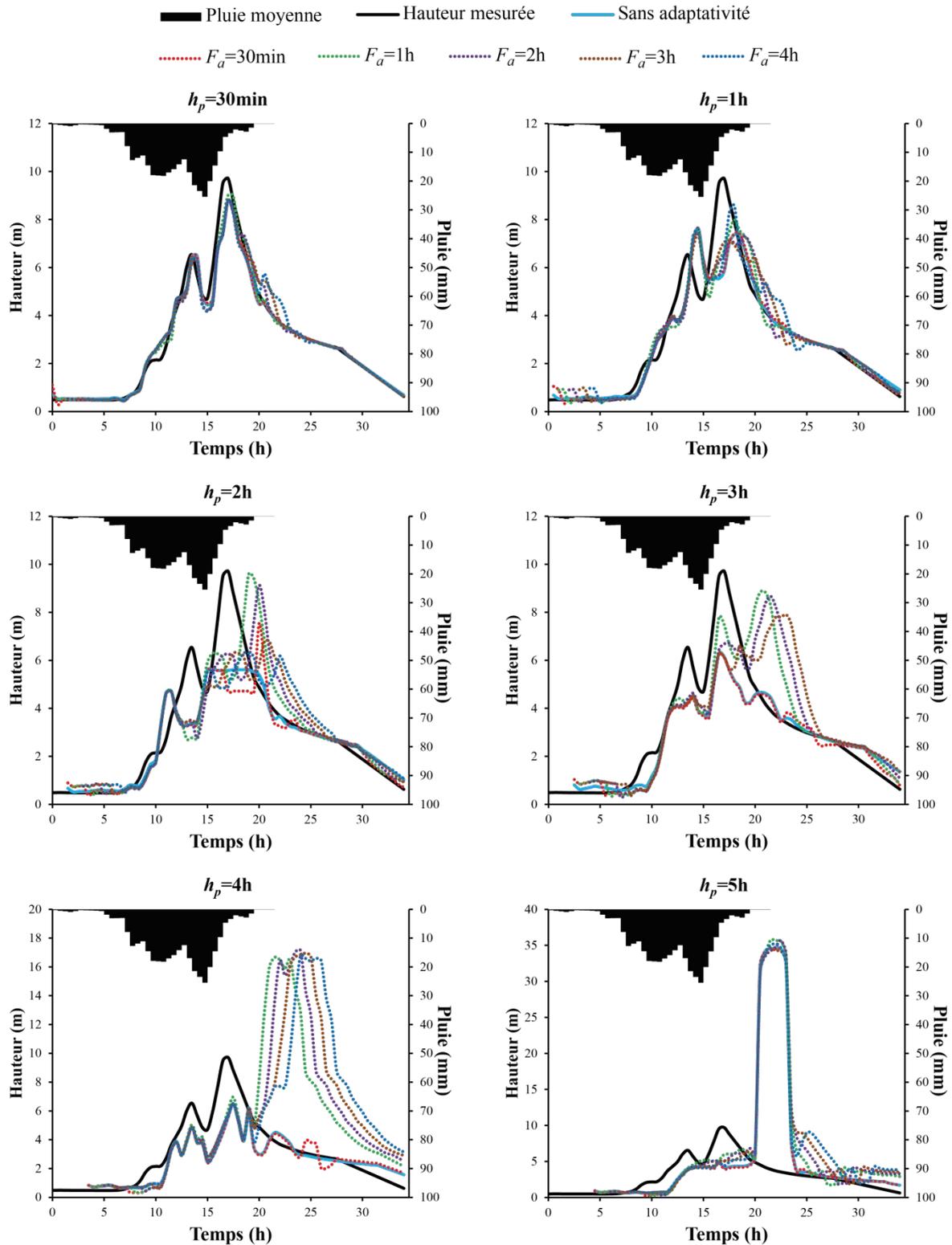


Figure 42 : Limnigrammes prévus à différents horizons de prévision pour l'événement 19 avec des modèles utilisant des fenêtres d'adaptativité différentes

		Sans adaptativité	$F_a=30\text{min}$	$F_a=1\text{h}$	$F_a=2\text{h}$	$F_a=3\text{h}$	$F_a=4\text{h}$
$h_p=0,5\text{h}$	C_p	0,34	0,31	0,49	0,23	0,14	0,06
	SPPD (%)	90	90	93	91	91	91
$h_p=1\text{h}$	C_p	0,02	0,04	0,12	0,07	0,21	0,09
	SPPD (%)	60	61	69	67	73	67
$h_p=2\text{h}$	C_p	0,46	0,25	0,11	0,23	0,33	0,23
	SPPD (%)	57	48	54	64	60	50
$h_p=3\text{h}$	C_p	0,66	0,64	0,44	0,32	0,24	-0,11
	SPPD (%)	63	63	77	69	63	32
$h_p=4\text{h}$	C_p	0,61	0,59	-3,0	-3,4	-3,7	-3,9
	SPPD (%)	57	57	62	59	59	59
$h_p=5\text{h}$	C_p	-11,1	-11,2	-11,6	-11,5	-11,3	-11,5
	SPPD (%)	42	41	61	52	52	45

Tableau 22: Valeurs du C_p et du SPPD sur l'événement 19 pour les modèles ayant des fenêtres d'adaptativité différentes

On constate, sur les limnigrammes présentés, que les modèles ayant une fenêtre d'adaptativité de 30 min calculent une hauteur presque identique à celle des modèles sans adaptativité : un seul exemple dans l'ensemble d'apprentissage n'est pas suffisant pour modifier sensiblement les paramètres des modèles.

On note également que l'adaptativité ne permet pas de modifier suffisamment les paramètres pour corriger les modèles donnant des résultats aberrants comme c'est le cas du modèle pour $h_p=5\text{h}$, où l'initialisation des paramètres correspondant au meilleur score de validation croisée conduit à un réseau qui prévoit une hauteur de 35m, totalement irréaliste, pour l'événement 19.

Deux constatations principales ressortent de ces limnigrammes. La première est que les modèles adaptatifs arrivent à augmenter notablement l'amplitude du pic, jusqu'à 88% du pic mesuré avec $F_a=4\text{h}$ pour $h_p=1\text{h}$ au lieu de 76% sans adaptativité, jusqu'à 99% du pic mesuré avec $F_a=1\text{h}$ pour $h_p=2\text{h}$ contre 58% sans adaptativité et 91% du pic mesuré avec $F_a=1\text{h}$ pour $h_p=3\text{h}$ par rapport à 65% sans adaptativité. Par contre pour $h_p=4\text{h}$, l'amplitude du pic est surestimée de plus de 70%, les hauteurs prévues n'ont ainsi aucune validité. Cette augmentation de l'amplitude du pic peut être due à une meilleure prise en compte des précipitations très intenses et/ou à un ajustement important du modèle aux hauteurs passées.

La seconde constatation est que le retard du pic prévu par rapport au pic observé est de plus en plus grand lorsque l'horizon de prévision augmente. D'une part ce retard a tendance à augmenter lorsque la fenêtre d'adaptativité est plus longue, et d'autre part la différence de retard entre les modèles est d'autant plus marquée que l'horizon de prévision est lointain. Ainsi pour $h_p=3\text{h}$, le pic du modèle avec $F_a=1\text{h}$ a un retard de 3,5h et celui avec $F_a=4\text{h}$ un retard de 6h ; quant aux pics prévus pour $h_p=4\text{h}$, celui du modèle avec $F_a=1\text{h}$ a un retard de 4,5h et celui avec $F_a=4\text{h}$ un retard de 7h. Ces retards très importants allant de 1 fois à 2 fois

la valeur de l'horizon de prévision s'expliquent par le fait que l'apprentissage prend en compte des variables mesurées à des temps éloignés de celui auquel est émise la prévision (le temps k), par un intervalle de temps égal à deux fois l'horizon de prévision (cf. Figure 40). Les retards avec lesquels le pic principal est prévu rendent évidemment les prévisions inutilisables.

Néanmoins, on peut retenir une légère amélioration des prévisions par rapport au modèle sans adaptativité pour $h_p = 30\text{min}$; avec $F_a = 1\text{h}$ on passe d'un *SPPD* de 90% à 93% et d'un critère de persistance de 0,34 à 0,49. De même pour $h_p = 1\text{h}$, avec $F_a = 3\text{h}$ le *SPPD* passe de 60% à 73% et la persistance de 0,02 à 0,21.

Dans le cas de l'événement 19, l'adaptativité apporte une légère amélioration des prévisions uniquement pour $h_p = 30\text{min}$ et $h_p = 1\text{h}$, par rapport au modèle sans contribution linéaire, mais on ne peut pas vraiment privilégier une fenêtre d'adaptativité. La proposition consistant à prendre comme prévision la médiane des 20 prévisions du modèle mixte linéaire et non linéaire présente donc de meilleurs résultats.

4.3.2. Évènement 26

De même que pour l'événement 19, pour chaque horizon de prévision, nous avons effectué des prévisions de l'événement 26 avec des modèles utilisant différentes fenêtres d'adaptativité. La Figure 43 rassemble les limnigrammes des prévisions et le Tableau 23, les critères de persistance et les *SPDD* correspondants, les autres critères se trouvent en Annexe 6.

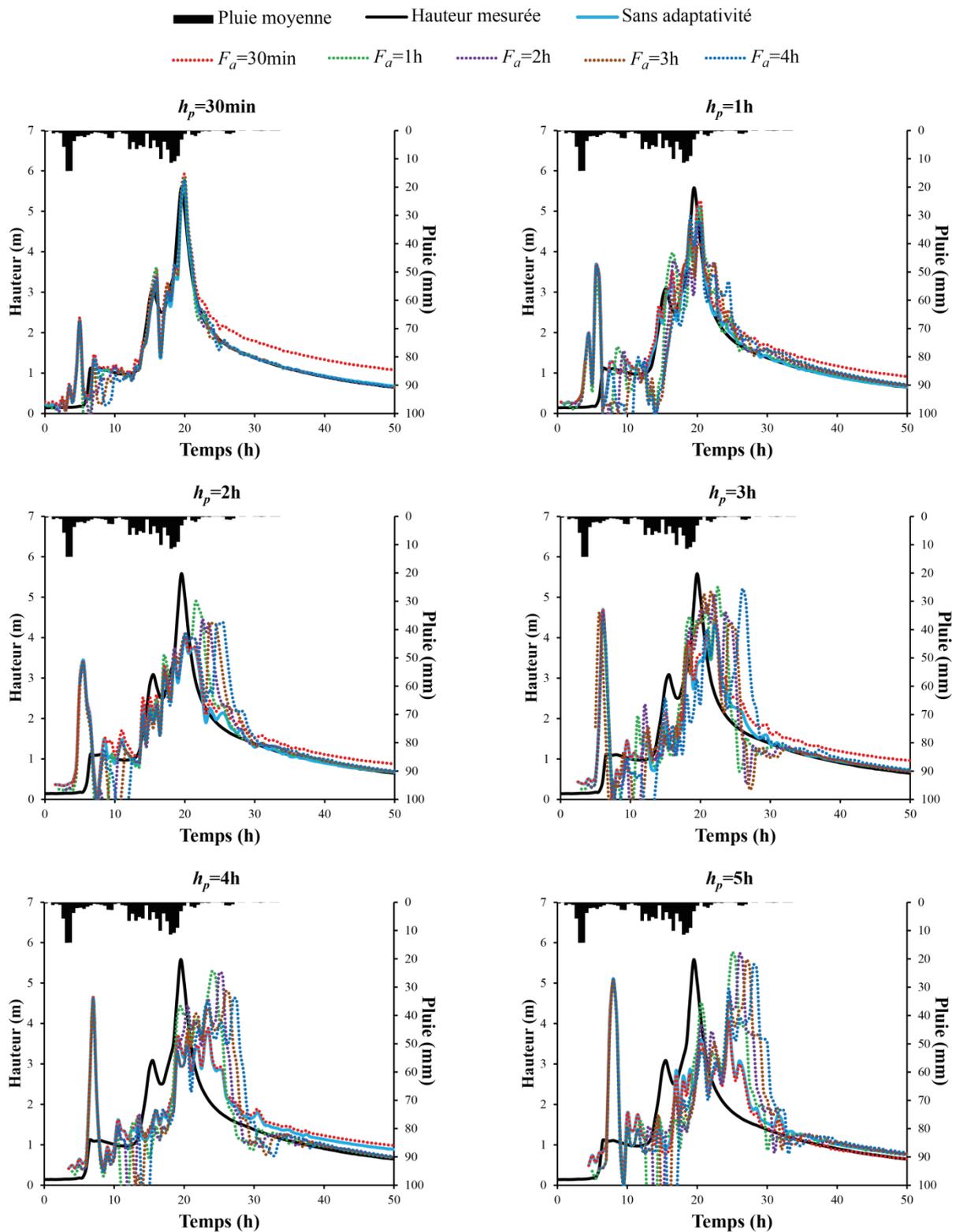


Figure 43 : Limnigrammes prévus à différents horizons de prévision pour l'événement 26 avec des modèles utilisant des fenêtres d'adaptivité différentes

Les remarques concernant les effets de l'adaptativité sur les prévisions de l'événement 26 rejoignent celles faites pour l'événement 19.

Les prévisions réalisées avec une fenêtre d'adaptativité de 30 min sont semblables à celles des modèles sans adaptativité

Cependant, l'amplitude des pics prévus est meilleure avec l'adaptativité et ce, quel que soit l'horizon de prévision, ce qui n'était pas le cas pour l'événement 19. Par exemple pour $h_p=2h$, le pic prévu avec $F_a=1h$ fait 88% du pic observé contre 72% sans adaptativité et pour $h_p=5h$, le pic prévu avec $F_a=3h$ a la même amplitude que celle du pic observé alors qu'il n'en faisait que 74% sans adaptativité.

Mais à côté de ces gains d'amplitude, les retards du pic prévu par rapport au pic observé sont beaucoup plus importants dans le cas de l'adaptativité et augmentent avec la largeur de la fenêtre d'adaptativité et l'éloignement de l'horizon de prévision. Ainsi pour $h_p=2h$, le retard du pic prévu par le modèle, avec $F_a=1h$ est de 2h ; celui avec $F_a=4h$, de 6h ; et pour $h_p=5h$, le pic prévu par le modèle avec $F_a=1h$ a un retard de 5,5h ; et celui avec $F_a=4h$ un retard de 8,5h.

On constate également que le phénomène d'oscillations de la hauteur prévue en réponses aux impulsions de pluie, est plus marqué lorsqu'il y a adaptativité, ce qui semble indiquer que les pluies sont affectées d'un poids plus important. Ces oscillations entraînent des hauteurs négatives qui nuisent à la qualité et à la crédibilité des prévisions. On remarque aussi que le faux pic en début d'événement n'est ni évité ni diminué.

Aucune prévision de l'événement 26 provenant des modèles avec adaptativité n'apparaît meilleure que celles des modèles sans adaptativité. Le gain en *SPPD* obtenu grâce à l'adaptativité est contrebalancé par la perte de persistance qui croît avec l'horizon de prévision.

		Sans adaptativité	$F_a=30min$	$F_a=1h$	$F_a=2h$	$F_a=3h$	$F_a=4h$
$h_p=0,5h$	C_p	-0,69	-2,1	-0,72	-0,66	-0,89	-1,3
	<i>SPPD</i> (%)	94	99	97	96	96	96
$h_p=1h$	C_p	-0,65	-0,91	-1,6	-2,3	-2,1	-2,9
	<i>SPPD</i> (%)	68	72	63	53	60	74
$h_p=2h$	C_p	0,20	0,16	-0,14	-0,26	-0,38	-0,48
	<i>SPPD</i> (%)	62	64	67	67	67	62
$h_p=3h$	C_p	0,17	0,15	-0,09	-0,20	-0,20	-0,58
	<i>SPPD</i> (%)	59	62	80	74	77	53
$h_p=4h$	C_p	0,43	0,41	0,06	0,00	-0,08	-0,18
	<i>SPPD</i> (%)	61	62	79	64	53	0,53
$h_p=5h$	C_p	0,26	0,24	-0,40	-0,72	-0,88	-0,74
	<i>SPPD</i> (%)	41	39	54	38	37	44

Tableau 23 : Valeurs du C_p et du *SPPD* sur l'événement 26 pour les modèles ayant des fenêtres d'adaptativité différentes

4.3.3. Évènement 27

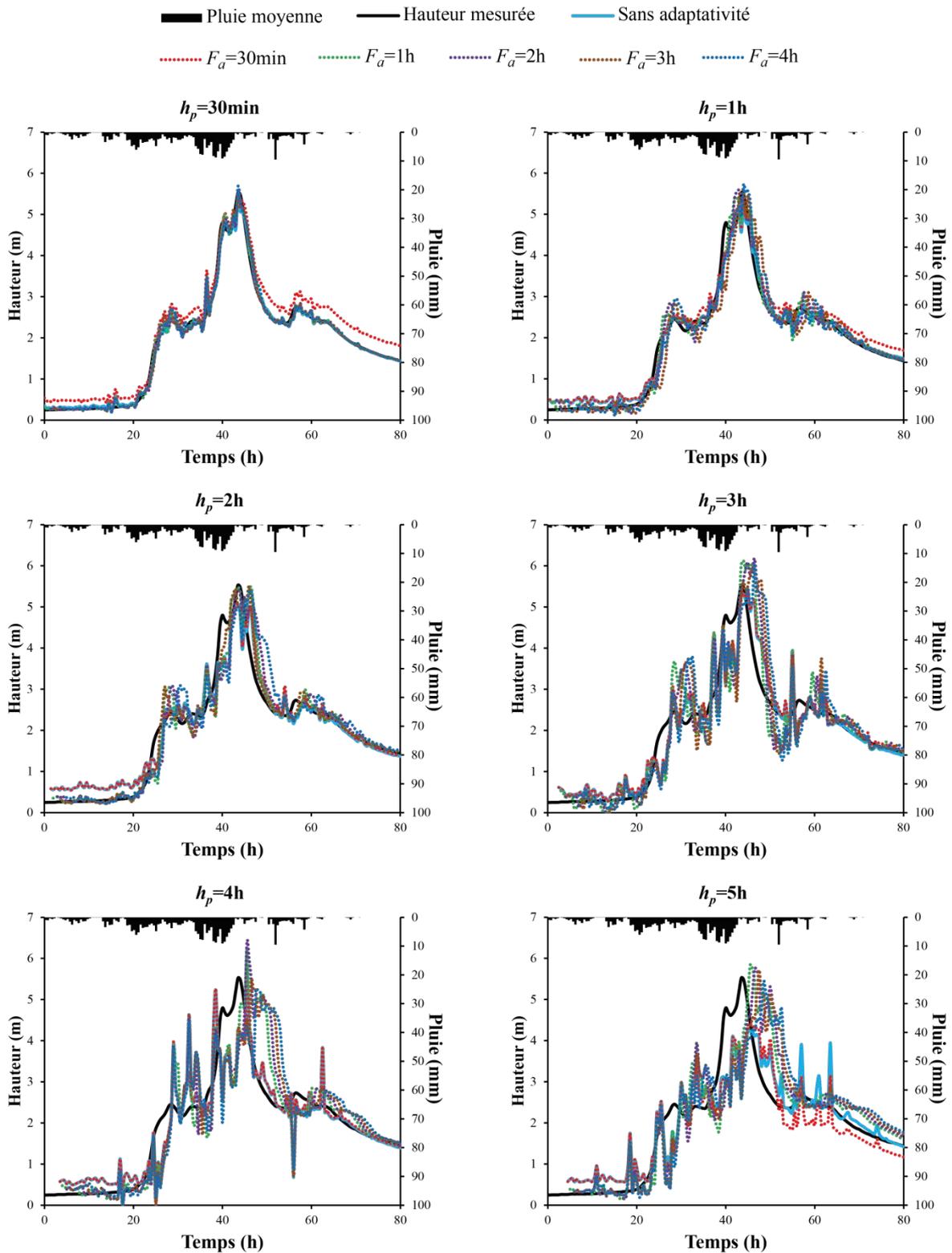


Figure 44 : Limnigrammes prévus à différents horizons de prévision pour l'évènement 27 avec des modèles utilisant des fenêtres d'adaptativité différentes

La Figure 44, regroupe les limnigrammes des prévisions pour l'événement 27 et le Tableau 24, les critères de persistance et les *SPPD* correspondants, les autres critères se trouvent en Annexe 6.

Pour un même horizon de prévision, les hauteurs prévues pour l'événement 27 avec les différentes fenêtres d'adaptativité sont plus semblables que dans les deux cas précédents. Cela est dû au fait que l'événement 27 est plus lent donc les variations de hauteurs entre deux pas de temps sont plus faibles et, par ailleurs, les pluies sont plus homogènes temporellement que pour les événements 19 et 26.

En conséquence, le gain en amplitude du pic prévu est moins important. Pour $h_p = 2h$, le pic prévu avec $F_a = 2h$ a la même amplitude que l'observé contre 92% sans adaptativité ; et pour $h_p = 5h$, le pic prévu avec $F_a = 2h$ est légèrement plus grand que le pic observé alors qu'il n'en faisait que 77% sans adaptativité.

On peut également remarquer que le retard des pics prévus avec adaptativité par rapport aux pics observés est moins grand que pour les événements 19 et 26, tout en restant très handicapant. Ce retard est de l'ordre de l'horizon de prévision jusqu'à $h_p = 3h$. Concernant l'écart de retard entre les modèles avec différentes fenêtres d'adaptativité, il est surtout visible pour $h_p = 4h$ et $5h$: pour $h_p = 4h$, le retard du pic prévu par le modèle avec $F_a = 1h$ est de 2h et de 4h avec $F_a = 4h$, pour $h_p = 5h$, le pic prévu par le modèle avec $F_a = 1h$ a un retard de 2h alors que celui avec $F_a = 4h$ un retard de 5.5h.

Les oscillations de hauteur dues aux impulsions de pluie ont une amplitude semblable pour les modèles avec et sans adaptativité. Enfin, il est difficile de privilégier une fenêtre d'adaptativité : l'adaptativité donne en général une information meilleure sur la hauteur au niveau du pic, mais une moins bonne persistance due à une moins bonne synchronisation.

		Sans adaptativité	$F_a=30min$	$F_a=1h$	$F_a=2h$	$F_a=3h$	$F_a=4h$
$h_p=0,5h$	C_p	0,09	-5,0	0,14	-0,05	-0,07	-0,06
	<i>SPPD</i> (%)	98	102	100	100	99	103
$h_p=1h$	C_p	0,21	-0,17	0,15	-0,10	-1,4	-0,15
	<i>SPPD</i> (%)	83	84	86	90	100	90
$h_p=2h$	C_p	0,30	0,24	0,12	-0,02	0,46	-0,33
	<i>SPPD</i> (%)	92	92	99	98	91	91
$h_p=3h$	C_p	0,40	0,39	-0,06	-0,20	-0,36	-0,31
	<i>SPPD</i> (%)	95	95	109	99	93	88
$h_p=4h$	C_p	0,41	0,40	-0,04	-0,21	-0,27	-0,39
	<i>SPPD</i> (%)	77	77	67	66	78	72
$h_p=5h$	C_p	0,52	0,50	0,22	0,08	0,00	-0,02
	<i>SPPD</i> (%)	54	57	62	56	47	47

Tableau 24 : Valeurs du C_p et du *SPPD* sur l'événement 27 pour les modèles ayant des fenêtres d'adaptativité différentes

4.4. Conclusion sur l'adaptativité

L'intérêt de disposer d'un modèle adaptatif s'est fait ressentir au vu de la difficulté de généralisation des modèles sur les événements en test. Cependant, les résultats obtenus en utilisant un apprentissage adaptatif ne sont pas à la hauteur de nos attentes. L'adaptativité permet aux réseaux de mieux prendre en compte les données de pluie et d'ajuster les modèles aux hauteurs spécifiques à l'événement en cours, ce qui contribue à une augmentation de l'amplitude du pic prévu. Cependant, l'écart temporel entre l'instant où l'on effectue l'apprentissage adaptatif et celui de la prévision souhaitée, est égal à deux fois l'horizon de prévision. Cela entraîne un retard considérable du pic prévu par rapport au pic observé en particulier pour les événements rapides comme les événements 19 et 26, ce qui donne une prévision de très mauvaise qualité. Ainsi, hormis pour les petits horizons de prévision dans le cas de l'événement 19, les modèles sans adaptativité obtiennent de meilleures prévisions.

Si l'on souhaite rendre les modèles adaptatifs plus performants, il faut trouver une autre voie que celle qui consiste à faire subir aux réseaux de neurones un nouvel apprentissage. Ainsi, dans [Dimopoulos 1996], un filtre de Kalman est ajouté après la sortie du réseau de neurones ; ce filtre sert à modéliser l'erreur réalisée par le réseau. Les paramètres du filtre sont ensuite modifiés en temps réel pour rendre le modèle adaptatif. Une légère amélioration par rapport au seul réseau de neurones a été constatée pour des prévisions de débit à pas de temps journalier ; cependant la pertinence de ce type de modèle dans le contexte des crues éclair n'est pas prouvée. En particulier l'intervalle de 2 fois l'horizon de prévision séparant l'obtention des données mesurées et l'échéance de prévision est un obstacle qui n'est pas traité par ce type de méthode.

5. Prédiction des crues sur le Gardon d'Anduze avec l'utilisation des données de pluie fournies par les mesures des radars

5.1. Introduction

La très forte hétérogénéité spatiale des précipitations constitue une des sources d'incertitude majeure que nous avons soulignées dans la présentation du site d'étude et des données associées. Afin de tenir compte de cette variabilité spatiale, nous avons utilisé l'information de précipitation issue de six pluviomètres répartis sur le bassin. Même s'il a été montré [Obled 2009] que ce nombre est suffisant pour bien représenter cette variabilité, nous avons souhaité évaluer l'intérêt d'utiliser des précipitations spatialisées au moyen des mesures effectuées par un radar météorologique. Le travail présenté dans cette section s'intègre donc dans la modélisation de la chaîne hydrométéorologique (Figure 14). Ce travail a été réalisé grâce à la contribution de Monsieur Yann Visserot, au cours de son stage de Master 2 « Gestion des Catastrophes et des Risques Naturels » de l'Université Montpellier 3 [Visserot 2012].

5.2. Base de données

5.2.1. Les données radar

Les données radar dont nous disposons sur la zone d'étude proviennent de deux radars : celui de Nîmes et celui de Bollène. On utilisera uniquement le radar de Nîmes car les échos fixes sur le bassin versant du Gardon d'Anduze sont plus nombreux avec le radar de Bollène, sur la période étudiée, avec en moyenne 60km² d'échos fixes pour le radar de Nîmes contre 235km² pour celui de Bollène.

Les mesures sont des niveaux radar que l'on transforme à l'aide de tables fournies par Météo France en réflectivité et en intensité de pluie. La résolution est de 1 km² et le pas de temps est de 5 min.

Outre ces données brutes acquises dans le cadre du projet FLASH, le SCHAPI et Météo France ont fourni les lames d'eau radar, au pas de temps horaire, issues de la réanalyse COMEPHORE [Tabary 2012], ainsi que les métadonnées associées suivantes : les taux de masquage et d'échos fixes pour chaque année, et les facteurs correctifs quotidiens établis à partir des pluviomètres. Les métadonnées du radar de Nîmes n'étant actuellement disponibles que sur la période 2000-2006, la base de données utilisée pour l'étude présentée dans cette section sera restreinte à cette période.

5.2.2. Traitements appliqués aux données

La résolution étant de 1 km², les fichiers de données contiennent 512 valeurs de niveau radar, une pour chaque pixel représentant le bassin versant. Mettre en entrée des réseaux de neurones un vecteur de variables de dimension 512, multiplié par l'historique des précipitations à considérer, n'est pas raisonnablement envisageable, du fait du dilemme biais-variance, pour une base encore plus réduite que celle sur laquelle nous avons travaillé dans le Chapitre 3. Nous avons préféré regrouper les pixels en six zones correspondant à des sous-bassins, tels que présentés Figure 45. Nous avons pris soin de choisir autant de zones que de pluviomètres utilisés dans la première partie de ce chapitre pour deux raisons :

- la possibilité d'une comparaison entre les résultats issus des deux bases de données,
- le fait que, d'après [Obled 2009], le champ de pluie présent sur le Gardon d'Anduze peut être efficacement représenté par un vecteur de six composantes.

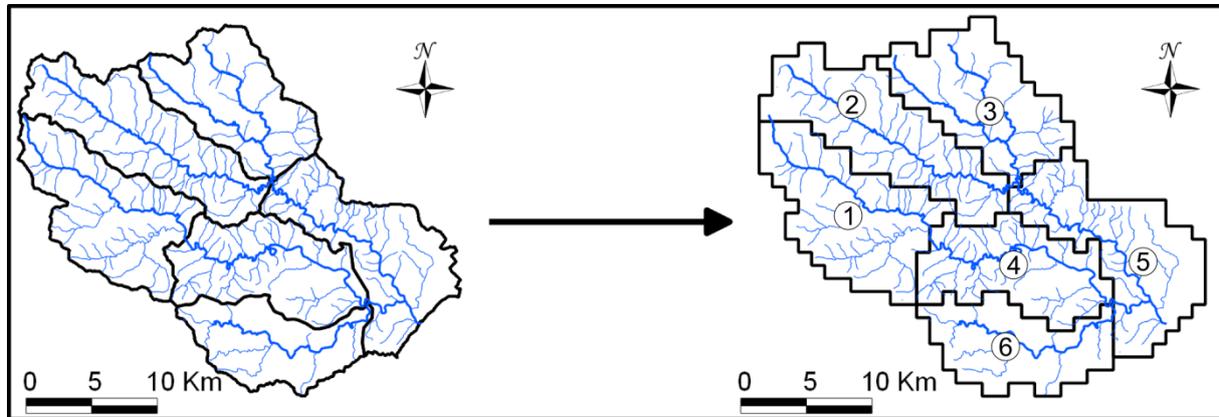


Figure 45 : Découpage du bassin versant du Gardon d'Anduze [conception Y. Visserot], zone 1 = Saint Jean Amont, zone 2 = Mialet Rive Droite, zone 3 = Mialet Rive Gauche, zone 4 = Saint Jean Aval, zone 5 = Mialet aval, zone 6 = Salindrenque

Après le découpage en zones, les niveaux radar sont transformés en intensités puis en cumuls de pluie. Ensuite, on supprime les échos fixes et on corrige les valeurs de cumuls à l'aide du taux de masquage. On passe ensuite du pas de temps 5 min à 30 min en réalisant la somme des six valeurs de cumuls concernées, et enfin on effectue la moyenne des cumuls sur chacune des 6 zones, ce qui donne 6 valeurs de pluie sur le bassin à chaque pas de temps. Compte tenu de la non linéarité de la relation Réflectivité-Pluie (ou Z-R, dans la littérature), nous avons préféré réaliser les ré-échantillonnages tant spatial que temporel après la conversion Z-R.

5.2.3. Les événements

La base de données constituée à partir des données radar s'étend sur la période 2000-2006 et comporte 11 événements dont 6 en commun avec la base de données utilisée précédemment ; les événements ajoutés ont été mis à notre disposition en cours de thèse. Par rapport à la base de données précédente, de petites différences apparaissent dans les caractéristiques de ces événements : en effet, ceux-ci ont été sélectionnés à partir des données brutes qui ne proviennent pas toujours des mêmes sources et le traitement des données n'a pas été exactement le même. C'est pourquoi nous présentons les caractéristiques des événements utilisés dans cette section dans le Tableau 25.

N°	Date	Durée (h)	Pluie moyenne cumulée pluviomètres (mm)	Pluie moyenne cumulée radar (mm)	Hauteur maximum (m)	Niveau de vigilance [SPCGD 2010]
12	19-20 Septembre 2000	39	114	87	2,72	vert
13	28-29 Septembre 2000	51	207	127	4,80	jaune
120	12-14 Novembre 2000	76	128	42	2,77	vert
15	14-16 Mars 2002	55	67	42	3,67	vert
19	8-9 Septembre 2002	41	351	209	10,43	rouge
200	15-18 Novembre 2003	57	150	75	3,81	jaune
201	21-25 Novembre 2005	90	182	109	5,06	jaune
21	5-10 Septembre 2005	123	262	184	2,06	vert
22	24-25 Septembre 2006	33	72	33	2,27	vert
23	19-20 Octobre 2006	66	237	103	6,54	orange
24	17-18 Novembre 2006	55	84	58	2,72	vert

Tableau 25 : Base de données comportant le numéro de l'événement, la date, la durée, le cumul de pluie des pluviomètres, le cumul de pluie du radar, la hauteur maximale à Anduze et le niveau de vigilance

Les limnigrammes et hyétogrammes de ces différents événements sont présentés en Annexe 7.

5.3. Comparaison entre les pluies mesurées au niveau des pluviomètres et celles obtenues à partir des mesures du radar

Afin d'alléger la rédaction, nous appellerons désormais la pluie mesurée au niveau des pluviomètres, « pluie des pluviomètres », et la pluie obtenue à partir des mesures radar, « pluie du radar ». Avant de commencer la modélisation, on a souhaité comparer les données obtenues avec les pluviomètres et celles issues du radar. Pour ce faire, nous proposons de calculer l'erreur relative entre le cumul de pluie d'un événement mesuré par un pluviomètre et celui de la zone radar dans laquelle se trouve ce pluviomètre.

$$\varepsilon = \frac{P_1 - P_2}{P_2} \times 100$$

avec :

- P₁ le cumul de pluie provenant d'une zone radar
- P₂ le cumul de pluie provenant du pluviomètre correspondant à cette zone

Bien entendu, la valeur ponctuelle du pluviomètre et la moyenne des valeurs spatialisées du radar correspondant ne sont pas rigoureusement comparables, mais cela donne néanmoins une idée de la différence entre les variables des modèles. Le pluviomètre de Saumane est associé aux zones 1 et 4 car il est situé à la limite entre les deux zones, celui de Barre-des-Cévennes et celui de Saint-Roman à la zone 2, celui de Mialet et celui d'Anduze à la zone 5, celui de Soudorgues à la zone 6 ; la zone 3 n'est pas représentée car aucun pluviomètre ne s'y trouve. Le Tableau 26 présente les erreurs relatives calculées pour chaque couple {pixel de zone – pluviomètre}.

événement	Zone 1-Saumane	Zone 2-Barre des Cévennes	Zone 2-St Roman	Zone 4-Saumane	Zone 5-Anduze	Zone 5-Mialet	Zone 6-Soudorgues	Moyenne
12	-52	25	-48	-30	-22	-19	-34	-26
13	-51	-39	-52	-49	-18	-42	-27	-40
120	-69	-63	-77	-51	-38	-76	-66	-63
15	-65	2650*	-58	53	-43	-34	19	-21
19	-59	44	-67	-32	-48	0,3	-17	-26
200	-51	-40	-52	-53	-34	-58	-58	-49
201	-33	-66	-38	-39	4	-26	-29	-32
21	-16	-5	-18	-10	-45	-48	-31	-25
22	-56	-57	-55	-65	-19	-64	-60	-54
23	-49	-62	-52	-57	4	-64	-76	-51
24	-30	-21	-31	-7	-15	-35	-49	-27
Moyenne	-48	-29	-50	-31	-25	-42	-39	-

Tableau 26 : Erreur relative (en %) entre les pluies des zones radar et les pluies des pluviomètres correspondantes

* Pour l'événement 15, le cumul de pluie au pluviomètre de Barre-des-Cévennes est proche de zéro, ce qui donne cette valeur très élevée de l'erreur relative, cette valeur n'a pas été retenue pour calculer les moyennes.

D'après le Tableau 26, la pluie du radar pour une zone est dans la grande majorité des cas sous-estimée par rapport à la pluie des pluviomètres. Sur un même événement, la pluie du radar sur une zone est inférieure de 25 à 63% à celle observée au pluviomètre correspondant : l'écart entre les deux types de données varie fortement d'un événement à l'autre. L'écart entre la pluie observée sur un pluviomètre donné et la pluie radar associée va de 25% pour Anduze à 50% pour Saint Roman.

La sous-estimation de la pluie par les données radar par rapport aux pluviomètres est importante, il convient maintenant de savoir si les formes des hyétogrammes sont semblables pour les deux types de signaux. On calcule ainsi le coefficient de corrélation linéaire, r :

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

avec :

- x_i la pluie du pluviomètre au pas de temps i
- \bar{x} la pluie moyenne du pluviomètre

- y_i la pluie du radar au pas de temps i
- \bar{y} la pluie moyenne du radar

Plus ce coefficient est proche de 1 en valeur absolue, meilleure est la corrélation entre les deux séries de pluie.

événement	Zone 1-Saumane	Zone 2-Barre des Cévennes	Zone 2-St Roman	Zone 4-Saumane	Zone 5-Anduze	Zone 5-Mialet	Zone 6-Soudorgues	Moyenne
12	0,24	0,67	0,86	0,12	0,84	0,26	0,84	0,55
13	0,88	0,88	0,91	0,73	0,79	0,87	0,89	0,85
120	0,70	0,68	0,76	0,53	0,61	0,79	0,68	0,68
15	0,75	-0,04*	0,47	0,24	0,88	0,83	0,50	0,61
19	0,86	0,45	0,87	0,87	0,86	0,79	0,91	0,80
200	0,86	0,90	0,90	0,88	0,79	0,78	0,83	0,85
201	0,63	0,75	0,67	0,68	0,65	0,63	0,85	0,70
21	0,73	0,77	0,80	0,72	0,84	0,86	0,85	0,79
22	0,85	0,74	0,83	0,69	0,66	0,97	0,91	0,81
23	0,75	0,76	0,84	0,57	0,63	0,86	0,80	0,74
24	0,07	0,84	0,21	0,13	0,49	0,66	0,66	0,44
Moyenne	0,66	0,74	0,74	0,56	0,73	0,75	0,79	-

Tableau 27 : Coefficient de corrélation linéaire entre les pluies issues des pluviomètres et du radar

* Pour l'événement 15, le cumul de pluie au pluviomètre de Barre des Cévennes est proche de zéro, ce qui donne cette valeur presque nulle du r , cette valeur n'a pas été retenue pour calculer les moyennes.

Pour un même événement, le coefficient de corrélation prend des valeurs allant de 0,44 à 0,85 avec une moyenne de 0,71, Tableau 27. La corrélation entre les pluies des pluviomètres et celles des zones radar est très variable suivant les événements mais une analogie entre les signaux existe bien. Le pluviomètre de Saumane est moins bien corrélé avec ses zones associées que les autres pluviomètres qui présentent tous des coefficients de corrélation supérieurs à 0,7 ; ceci peut être expliqué par le fait que le pluviomètre de Saumane est à la frontière entre les zones 1 et 4.

Les hyétogrammes de la pluie provenant des zones radar et des pluviomètres ont des amplitudes différentes avec des formes ressemblantes mais non identiques : ainsi les modèles réalisés à partir de ces deux types de données n'auront *a priori* pas la même architecture.

5.4. Modèles utilisés et répartition des événements dans les différents ensembles

Ce travail ayant eu lieu avant la proposition d'ajout des termes linéaires aux modèles, on avait choisi de prendre des modèles avec perceptron multicouche vus au Chapitre 3, Figure 16. Pour chaque horizon de prévision, on réalisera deux types de modèles : l'un avec les pluies des 6 pluviomètres et l'autre avec les pluies des 6 zones radar. La conception du modèle utilisant les pluies des pluviomètres doit être entièrement refaite, car la base de

donnée est, dans cette partie, réduite aux événements communs entre les deux sources de donnée : radar et pluviomètres. On sélectionne dans chaque cas la valeur de la fenêtre temporelle des pluies, le nombre de neurones cachés et l'initialisation des paramètres du réseau à l'aide de la validation croisée.

Les 11 événements de la base de données seront répartis comme suit :

- les événements 13 et 19 seront utilisés successivement comme test. Le 19 est l'événement le plus intense de la base, et le 13 est également un événement intense, à un seul pic,
- l'événement 13 étant utilisé en test, ce sera l'événement 200 qui servira pour l'arrêt précoce car c'est celui qui obtenait les meilleurs résultats en test et est donc le plus représentatif de la base d'apprentissage,
- les autres événements composeront l'ensemble d'apprentissage. Lorsque l'événement 19 sera testé, l'événement 13 sera dans l'ensemble d'apprentissage et inversement.

5.4.1. Sélection des modèles

En utilisant une procédure de sélection de modèle analogue à celle présentée précédemment, nous avons abouti au dimensionnement des réseaux indiqué dans le Tableau 28.

		$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
f_T (h)	Pluviomètres	0,5	5	4,5	4,5	5	4,5
	Zones radar	0,5	0,5	0,5	0,5	0,5	4,5
N_c	Pluviomètres	2	2	2	2	1	1
	Zones radar	1	3	4	3	4	1

Tableau 28 : Pour chaque horizon de prévision, choix de la fenêtre temporelle et du nombre de neurones cachés pour les modèles avec les pluies des pluviomètres et avec les pluies des zones radar.

Comme le montre le Tableau 28, les modèles utilisant les pluies des pluviomètres et celles des zones radar présentent des différences notables au niveau du nombre de variables et du nombre de neurones cachés. On constate également que les modèles alimentés par les pluies observées aux pluviomètres ne sont pas identiques à ceux sélectionnés avec la base de données utilisée précédemment. La faible sensibilité des modèles à la fenêtre temporelle des pluies et au nombre de neurones cachés explique autant de différences entre des réseaux modélisant un même processus et dont les ensembles d'apprentissage comprenaient des événements en commun.

Compte tenu des valeurs de f_T , les modèles qui utilisent les pluies des pluviomètres semblent avoir besoin d'une fenêtre de pluie plus large que les modèles avec les pluies des zones radar. La complexité de ces derniers modèles réside dans l'architecture : ils nécessitent (à l'exception des horizons 5h et 30min) plus de neurones cachés que les modèles avec les pluies des pluviomètres.

Pour chaque horizon de prévision et pour les deux types de pluie en entrée, on réalise 20 modèles correspondant à 20 initialisations différentes des paramètres du réseau. Les 20 initialisations n'ont pas été sélectionnées par validation croisée partielle car on a montré en Annexe 8 que la médiane des prévisions des 20 modèles ainsi sélectionnés était très semblable à la médiane des prévisions de 20 modèles issus de 20 initialisations des paramètres non

sélectionnées par validation croisée partielle. Afin d'évaluer les performances de ces modèles respectifs, on les teste successivement avec les événements 19 et 13.

5.4.2. Résultats obtenus pour les différents horizons de prévision

Les hauteurs prévues représentées sur les limnigrammes correspondent à la, médiane à chaque pas de temps, des 20 prévisions obtenues par les 20 modèles correspondants.

5.4.2.1. Événement 13

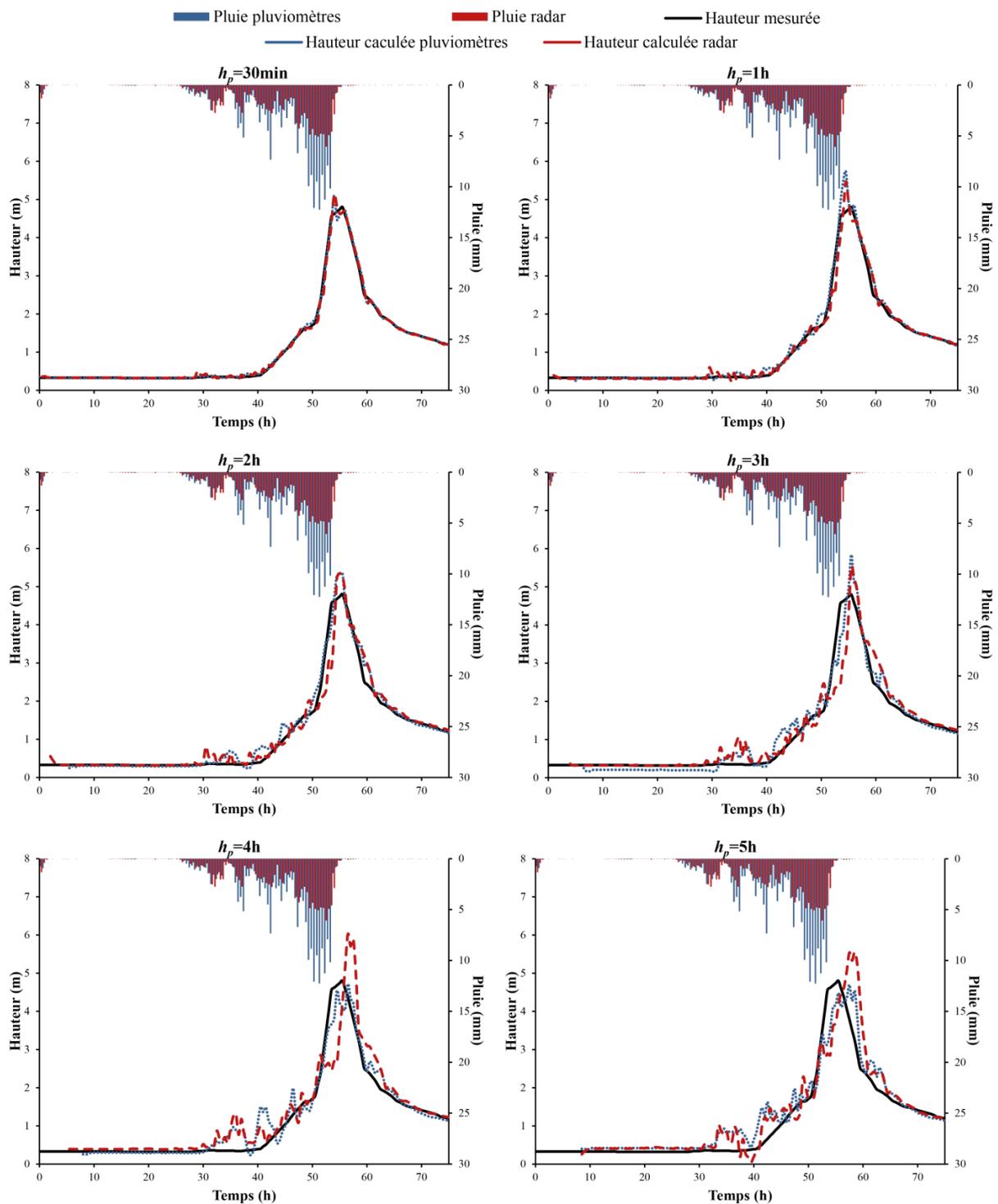


Figure 46 : Événement 13, les hauteurs prévues par les modèles utilisant les données de pluviomètres et ceux utilisant les données radar pour chaque horizon de prévision

Les hauteurs prévues par les modèles pour l'événement 13 sont illustrées en Figure 46 et les valeurs des critères sont consignées Tableau 29. Jusqu'à $h_p=3h$, l'amplitude du pic prévu est équivalente pour les deux types de modèles avec une surestimation de l'ordre de 15%, qui reste acceptable car elle ne correspond pas à un changement de niveau de vigilance. Pour $h_p=4h$ et $h_p=5h$, la surestimation du pic des prévisions des modèles à base de pluies du radar devient gênante ; pour les modèles utilisant les pluies des pluviomètres, la hauteur au niveau du pic est bonne pour $h_p=4h$ et sous-estimée pour $h_p=5h$.

Les limnigrammes montrent que les modèles utilisant les données radar sont systématiquement en retard lors de la montée du pic par rapport aux modèles avec les pluies des pluviomètres qui proposent une montée satisfaisante jusqu'à $h_p=4h$. Ce retard peut être dû en partie au manque d'information sur l'historique des pluies car les modèles avec les pluies du radar ne prennent en considération qu'une seule valeur de pluie pour la majorité des horizons de prévision.

Les résultats obtenus avec l'événement 13 en test poussent à privilégier les modèles utilisant les pluies des pluviomètres, ces modèles donnent des prévisions très intéressantes jusqu'à $h_p=4h$ et utilisables comme indications pour $h_p=5h$.

		$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
C_P	Pluviomètres	0,74	0,62	0,85	0,84	0,88	0,85
	Zones radar	0,61	0,61	0,66	0,64	0,59	0,69
C_N	Pluviomètres	1,00	0,98	0,98	0,95	0,94	0,90
	Zones radar	1,00	0,98	0,95	0,89	0,80	0,79
PPD (%)	Pluviomètres	106	119	112	122	98	94
	Zones radar	107	114	111	115	125	116
$SPPD$ (%)	Pluviomètres	97	98	112	122	86	93
	Zones radar	97	93	110	115	87	87
Retard H_{pic} (h)	Pluviomètres	-1,5	-1	0	0	1	3
	Zones radar	-1,5	-1	-0,5	0	1	3

Tableau 29 : Valeurs du C_P , C_N , PPD , $SPPD$ et Retard H_{pic} suivant l'horizon de prévision pour les hauteurs prévues sur l'événement 13 par les modèles utilisant les pluies des pluviomètres et ceux utilisant les pluies du radar.

5.4.2.2. Événement 19

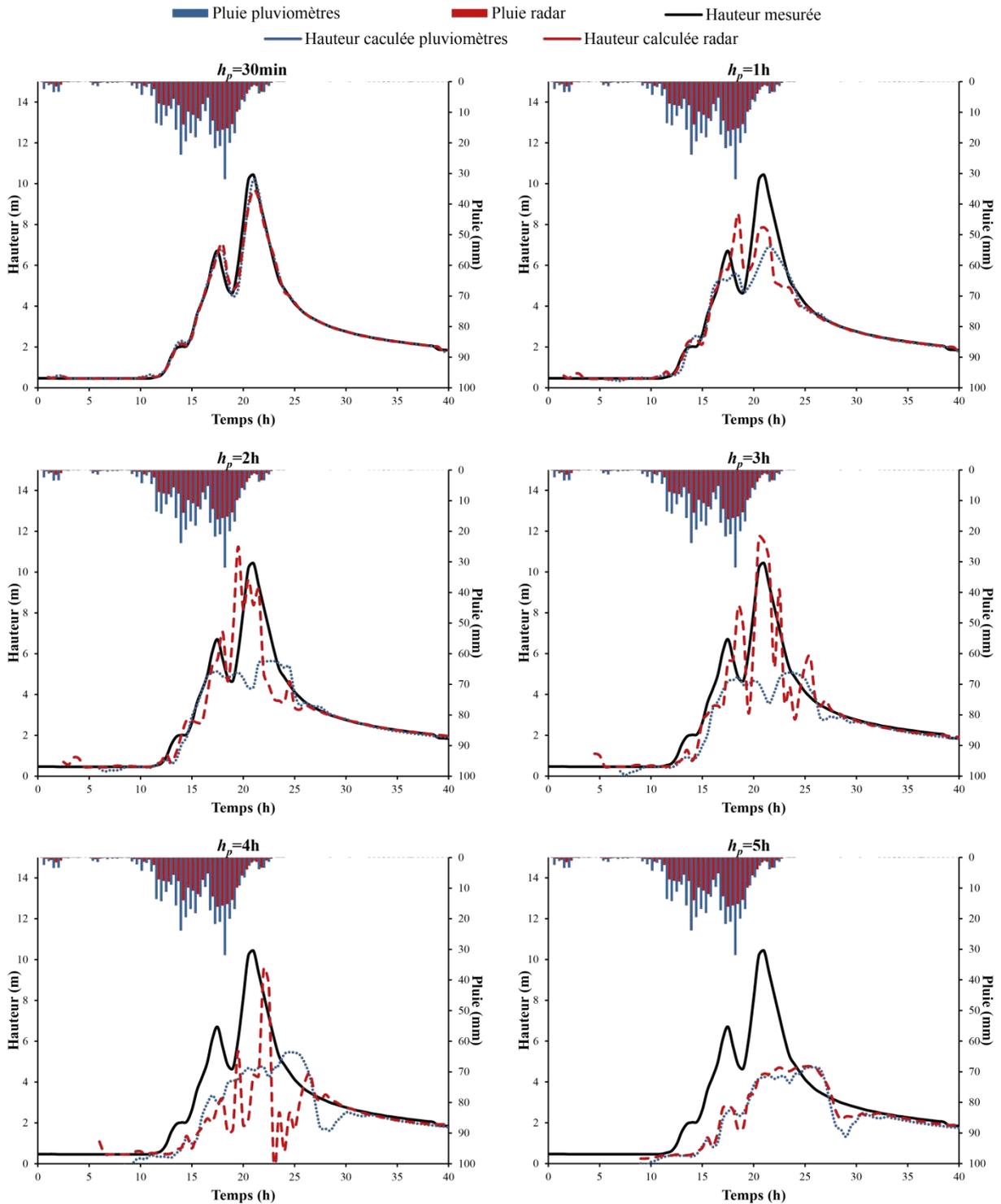


Figure 47 : Événement 19, les hauteurs prévues par les modèles utilisant les données de pluviomètres et ceux utilisant les données radar pour chaque horizon de prévision

Les prévisions réalisées par les deux types de modèles sur l'événement 19, événement plus intense que les exemples vus dans la base d'apprentissage, se trouvent Figure 47 avec les valeurs des critères de performance Tableau 30.

Tout d'abord, on peut noter que les prévisions sur l'événement 19 sont du même ordre de qualité que celles obtenues dans le Chapitre 3 avec une base de données plus étoffée. Des différences plus marquées que dans le cas de l'événement 13 apparaissent entre les hauteurs prévues par les deux types de modèles. Les modèles utilisant les pluies des pluviomètres sous-estiment fortement la hauteur du pic dès $h_p=2h$ avec un *PPD* de 50%, alors que concernant les modèles avec les pluies du radar, le *PPD* reste intéressant jusqu'à $h_p=4h$.

Pour $h_p=30\text{min}$ et $h_p=1h$, malgré des hauteurs de pic prévues très satisfaisantes, les modèles avec les pluies des pluviomètres accusent un retard non négligeable lors de la montée du pic, ce qui réduit d'autant l'anticipation en temps réel. Ce retard est moindre pour les modèles avec les pluies du radar.

Dans le cas de l'événement 19, les prévisions effectuées par les modèles utilisant les pluies du radar sont de meilleure qualité : elles sont utilisables pour $h_p=30\text{min}$ et $h_p=1h$, peuvent donner des indications intéressantes aux prévisionnistes pour $h_p=2h$ et $h_p=3h$, mais ne sont pas exploitables au-delà.

		$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
C_P	Pluviomètres	0,73	0,30	0,43	0,42	0,49	0,47
	Zones radar	0,70	0,31	0,63	0,72	0,13	0,51
C_N	Pluviomètres	0,99	0,88	0,74	0,60	0,52	0,24
	Zones radar	0,98	0,88	0,83	0,81	0,17	0,29
<i>PPD</i> (%)	Pluviomètres	98	64	54	49	52	45
	Zones radar	93	75	108	112	92	46
<i>SPPD</i> (%)	Pluviomètres	98	63	42	39	44	41
	Zones radar	93	75	80	110	42	42
Retard H_{pic} (h)	Pluviomètres	0	1	2	2,5	3,5	4,5
	Zones radar	0	0	-1,5	-0,5	1	4,5

Tableau 30 : Valeurs du C_P , C_N , *PPD*, *SPPD* et Retard H_{pic} suivant l'horizon de prévision pour les hauteurs prévues sur l'événement 19 par les modèles utilisant les pluies des pluviomètres et ceux utilisant les pluies du radar.

5.5. Conclusion sur l'utilisation des données de pluies fournie par les mesures du radar pour la prévision des crues à Anduze

La sous-estimation des pluies des zones radar par rapport aux pluies provenant des pluviomètres correspondant, n'a pas d'impact sur l'amplitude des hauteurs prévues au niveau du pic ; au contraire les modèles avec les données radar ont tendance à prévoir des pics d'amplitude plus grande. Les réseaux de neurones ont su s'ajuster à l'échelle des valeurs de la base d'apprentissage. On remarque que les modèles avec les pluies du radar utilisent un historique des pluies plus restreint que ceux avec les pluies des pluviomètres.

Les deux événements proposés en test ne permettent pas de conclure sur la prédominance des performances d'un type de modèle par rapport à l'autre. Pour l'événement 13, ce sont les modèles alimentés par les pluies des pluviomètres qui sont meilleurs, alors que pour l'événement 19 c'est le contraire. L'apprentissage sur une base de données plus étoffée et le test sur d'autres événements pourraient contribuer à une estimation plus pertinente des

capacités respectives des deux types de modèles. Un autre traitement des données radar avec de meilleurs facteurs correctifs et un zonage différent sont également à envisager.

L'utilisation des données de pluies provenant de radar comme variables d'entrée a contribué à modéliser une autre partie de la chaîne hydrométéorologique. Ces modèles seront utiles en cas de panne des pluviomètres ou des télétransmissions en situation opérationnelle.

6. Prévision des crues sur le bassin versant du Gardon à Remoulins

6.1. Introduction

Au-delà de la prévision des crues rapides sur les amonts du Gardon, comme à Anduze, le projet FLASH prévoyait d'étudier la généralisation des méthodes développées à des bassins plus grands, ayant une dynamique de crues plus diversifiée encore. À cette fin, nous nous sommes intéressés à l'ensemble du bassin versant du Gardon, c'est-à-dire au bassin versant du Gardon à Remoulins. En tenant compte des remarques faites sur le bassin versant du Gardon d'Anduze, on créera des modèles à différents horizons de prévision qui prévoiront la hauteur du Gardon à Remoulins.

6.2. Site d'étude

Le bassin versant du Gardon à Remoulins, représenté Figure 48, a une superficie de 1848 km². Le Gardon d'Anduze et le Gardon d'Alès se rejoignent un peu avant Ners pour former le Gardon. La topographie du bassin versant du Gardon peut être décrite en quatre parties. De l'amont vers l'aval :

- les Cévennes avec de fortes pentes et des vallées encaissées,
- le piémont cévenol avec des reliefs moins prononcés et des plaines assez étroites,
- la Gardonnenque, qui commence peu après Ners, est une large plaine constituée d'alluvions du tertiaire et du quaternaire,
- les gorges calcaires qui s'étendent sur environ 25 km et qui se terminent quelques kilomètres avant Remoulins ; le lit du Gardon y est très encaissé.

Ces variations importantes de topographie ont pour conséquences des régimes hydrauliques très différents qui doivent être pris en compte dans les modèles hydrologiques comme c'est le cas pour n-TOPMODELS dans [Bonnifait 2009].

Les stations de mesures que nous avons utilisées sont indiquées sur la Figure 48. La prévision de la hauteur à Remoulins est réalisée grâce aux hauteurs d'eau amont mesurées à Ners, Alès et Anduze, ainsi qu'aux précipitations reçues par l'aval du bassin à Ners et à Remoulins. De cette manière la conversion des pluies intense reçues par les amonts en hauteur à Anduze et Alès n'est plus à réaliser par le modèle, et, comme nous le verrons dans la section 6.6, le temps de propagation de la crue depuis les stations d'Alès et d'Anduze confère une capacité d'anticipation suffisante au modèle.



Figure 48 : Bassin versant du Gardon (conception G. Artigue)

6.3. Sélection et constitution de la base de données

Pour alimenter les modèles, on a utilisé, au pas de temps 30 min, les hauteurs observées aux stations limnimétriques d'Anduze, d'Alès, de Ners et de Remoulins, et les pluies mesurées aux pluviomètres de Ners et de Remoulins. La station limnimétrique de Russan a été exclue car elle n'a été créée qu'en 2003. Les hauteurs ont été privilégiées aux débits car les courbes de tarages fournies par le SPC Grand-Delta ne couvraient pas l'intégralité de la période d'étude.

On note que suite à des travaux près de la station limnimétrique de Remoulins en 2006, une différence importante de la hauteur initiale a été constatée entre les événements antérieurs à 2006 et les autres. Afin d'homogénéiser les données en vue de leur utilisation par l'apprentissage statistique, on a décidé d'ajouter 0,93m aux hauteurs à partir de 2006, ce qui correspond à la différence entre la moyenne des hauteurs avant 2006 et la moyenne des « vraies » hauteurs après 2006.

On a repris les événements qui constituaient la base de données pour le travail réalisé sur le Gardon d'Anduze. Cependant l'événement 1 (septembre 1994) a été supprimé car il manquait 12 heures de données de hauteur à Remoulins au moment supposé du pic. L'événement 19 (septembre 2002) a également dû être écarté car des données de pluie et de hauteur étaient manquantes à Ners, la hauteur à Alès était vraisemblablement erronée avec un maximum de 2,6m (en contradiction avec tous les relevés visuels effectués) et enfin la hauteur à Remoulins présentait une allure très irrégulière. On a gardé l'événement 8 (novembre 1996), malgré une hauteur constante à Alès, afin de ne pas trop réduire le nombre d'événements. La base de données contient donc les 15 événements répertoriés dans le Tableau 31. Les limnigrammes et hyétogrammes de la base de données sont représentés en Annexe 9.

N°	Date	Durée (h)	Cumul de pluie à Remoulins (mm)	Hauteur maximum (m)	Niveau de vigilance [SPCGD 2010]
3	3-6 Octobre 1995	66	67	3,80	jaune
4	13-15 Octobre 1995	54	36	3,80	jaune
8	10-14 Novembre 1996	105	1	3,00	jaune
107	4-8 Novembre 1997	101	36	2,70	vert
108	24-27 Novembre 1997	72	4	1,87	vert
109	15-21 Décembre 1997	142	58	4,74	jaune
117	20-22 Octobre 1999	59	107	1,26	vert
13	28-30 Septembre 2000	60	52	2,89	vert
120	12-15 Novembre 2000	80	10	2,10	vert
22	23-26 Septembre 2006	70	63	1,55	vert
23	18-21 Octobre 2006	64	39	4,10	jaune
24	16-18 Novembre 2006	47	49	2,79	vert
25	20-24 Novembre 2007	84	85	2,37	vert
26	21-23 Octobre 2008	53	10	3,93	jaune
27	31 Octobre – 4 Novembre 2008	95	36	4,14	jaune

Tableau 31: Base de données comportant le numéro de l'événement, la date, la durée, le cumul de pluie à Remoulins, la hauteur maximale à Remoulins et le niveau de vigilance

Événement	3	4	8	107	108	109	117	13	120	22	23	24	25	26	27
Intense	X	X	X	X		X		X			X	X		X	X
Long			X	X	X	X			X				X		X
Multi-pic	X		X	X	X				X				X		

Tableau 32 : Caractérisation des événements

Pour comparer les caractéristiques des événements, on les classe dans le Tableau 32 suivant leur intensité (hauteur maximale), leur durée et l'allure du limnigramme. Les hauteurs maximales sont comprises entre 1,26 et 4,74m. Un événement est considéré comme intense si

sa hauteur maximale est supérieure à la moitié de celle de l'événement le plus intense : 4,74m pour l'événement 109. Les durées des événements, quant à elles, vont de 47h à 142h. Un événement est classé comme long si sa durée est supérieure à la moitié de celle de l'événement le plus long : 142 heures pour l'événement 109. La base de données comprend donc 10 événements intenses, 7 événements longs et 6 événements présentent plusieurs pics.

6.4. Types de modèles utilisés

On reprend les deux types de modèles utilisés dans le Chapitre 3 pour la prévision des crues du Gardon d'Anduze : un premier type de modèles à perceptron multicouche, représenté en Figure 49, et un second type de modèle à perceptron multicouche avec ajout d'une partie linéaire, représenté en Figure 50.

Si l'on se place à l'instant k , les modèles calculent la hauteur prévue à Remoulins au temps $k+h_p$, h_p étant l'horizon de prévision. La nature des variables exogènes diffère de celles utilisées pour les modèles du Gardon d'Anduze. En effet, on n'utilisera pas uniquement des données de pluie, mais également des données de hauteurs issues de stations amont. Les variables exogènes sont ainsi la pluie mesurée approximativement au centre du bassin (Ners), en aval du bassin (Remoulins), et les hauteurs à Anduze, Alès et Ners. Les modèles subissent un apprentissage dirigé de par la mise à disposition de la hauteur mesurée à Remoulins connue jusqu'à l'instant auquel est effectuée la prévision : k .

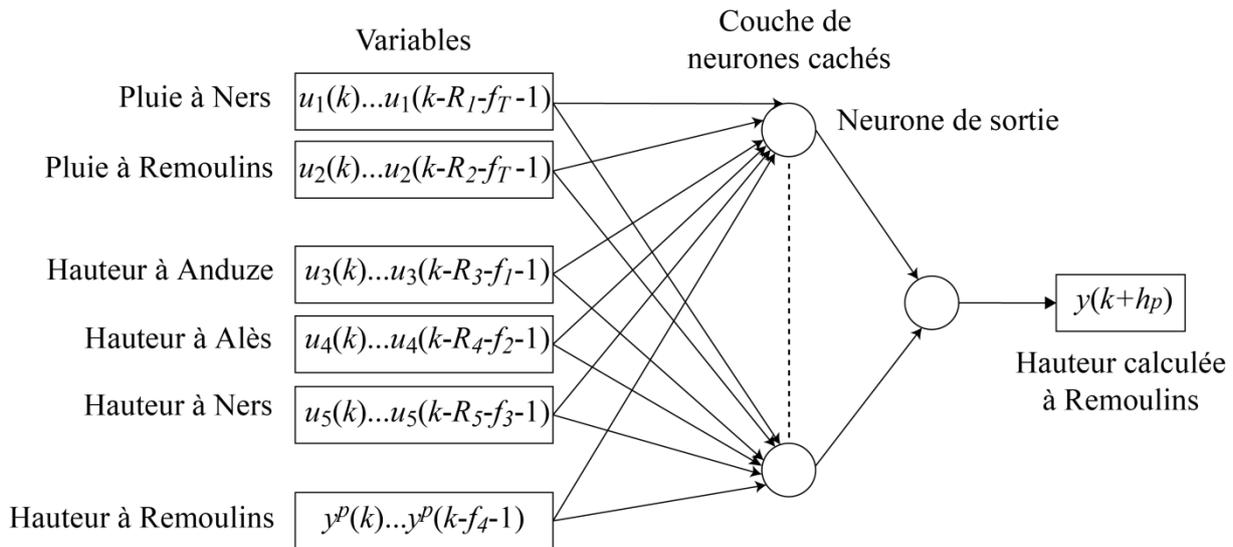


Figure 49 : Modèle à perceptron multicouche avec h_p l'horizon de prévision, f_T la fenêtre temporelle des pluies, f_1 à f_4 les fenêtres temporelles pour les hauteurs correspondantes et R_1 à R_5 les retards appliqués aux pluies et hauteurs correspondantes

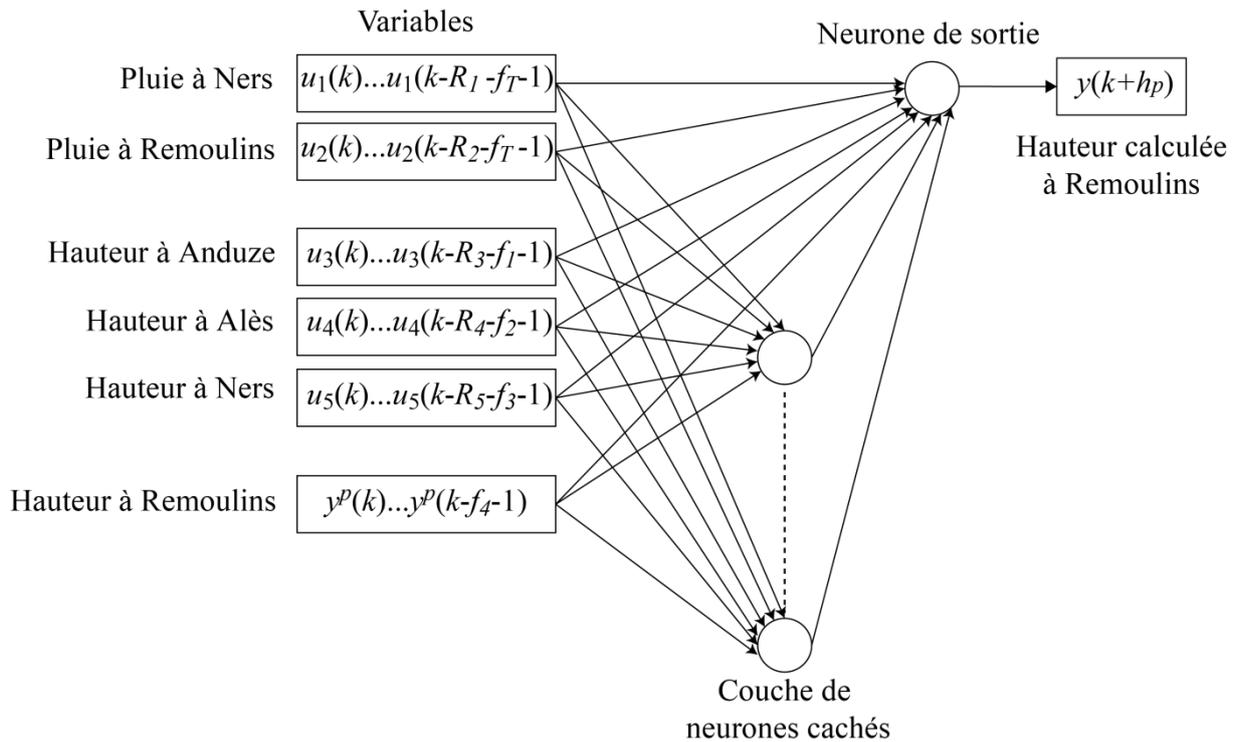


Figure 50 : Modèle avec ajout d'une partie linéaire au perceptron multicouche avec h_p l'horizon de prévision, f_T , la fenêtre temporelle des pluies, f_1 à f_4 les fenêtres temporelles pour les hauteurs correspondantes et R_1 à R_5 les retards appliqués aux pluies et hauteurs correspondantes

Pour les deux types de réseaux, on réalise des modèles pour chaque horizon de prévision : 1h, 3h, 5h, 7h, 9h et 11h. Dans chaque cas, on va choisir les hyperparamètres présentés au Chapitre 3.

6.5. Choix des hyperparamètres

Compte tenu des constatations faites au Chapitre 3 sur la faible sensibilité des modèles aux différents hyperparamètres, on décide de fixer le nombre de neurones cachés à 3 (valeur moyenne observée au Chapitre 3), la valeur initiale du pas d'apprentissage et son facteur de décroissance seront celles choisies par défaut par le logiciel RNFpro ($\mu_{ini}=0,001$ et $\mu_{dec}=0,1$), et la fenêtre temporelle des pluies sera fixée à 1,5h (valeur intermédiaire observée précédemment). En revanche, il est nécessaire de sélectionner la fenêtre temporelle des différentes hauteurs mesurées aux stations en amont présentées en entrée.

Comme la superficie du bassin versant du Gardon à Remoulins est environ 3,4 fois supérieure à celle du bassin versant du Gardon d'Anduze à Anduze, et que la pente moyenne du cours d'eau est plus faible, la dynamique des processus et le temps de réponse sont donc plus lents sur le bassin versant du Gardon à Remoulins. Le temps de réponse comporte le temps de concentration des précipitations (réalisé par le processus de production de la crue), et le temps de propagation et d'étalement de cette crue de l'amont vers l'aval ; s'ajoute aussi la production d'autres apports dus aux pluies tombant sur l'aval du bassin et aux affluents non jaugés (sans mesure) rejoignant le Gardon le long de son trajet aval. Ces différents temps de propagation vont être pris en compte dans le modèle en introduisant des retards (tant que

l'horizon de prévision le permet) sur les variables par rapport à la hauteur à Remoulins. Afin de diminuer le nombre de plans d'expériences, et en supposant que ces retards jouent principalement le rôle de retards purs, on détermine la valeur de ces retards à l'aide de corrélations croisées.

Les calculs de corrélation linéaire, utilisés en hydrologie notamment par [Mangin 1975], permettent pour la corrélation simple de calculer la corrélation linéaire d'une série chronologique par rapport à elle-même avec un décalage temporel (il s'agit alors d'autocorrélation ou corrélation simple). La corrélation croisée ou intercorrélacion, sert quant à elle, à établir la corrélation linéaire entre deux séries chronologiques en fonction du décalage de ces deux séries dans le temps. La corrélation croisée r_k , entre deux séries temporelles x et y , pour un décalage temporel k , se calcule comme ci-dessous :

$$r_k = \frac{\frac{1}{n} \sum_{i=1}^{n-k} (x_i - \bar{x}) \cdot (y_{i+k} - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

où :

- x la série temporelle étudiée, ici la hauteur à Remoulins
- y sera successivement la hauteur à Remoulins, la hauteur à Anduze, la hauteur à Alès, la hauteur à Ners, la pluie à Ners et la pluie à Remoulins
- k le décalage en unité 30 min
- n le nombre de pas de temps correspondant à la durée de l'événement

On calcule, pour chaque événement de la base de données, les corrélations croisées entre les grandeurs en entrée du modèle et la hauteur à Remoulins pour un décalage maximum de 20h. Puis on calcule la moyenne de ces corrélations croisées sur les 15 événements, Figure 51, afin de déterminer le retard à appliquer aux variables d'entrée du modèle. L'ensemble des corrélogrammes pourra être consulté en Annexe 10.

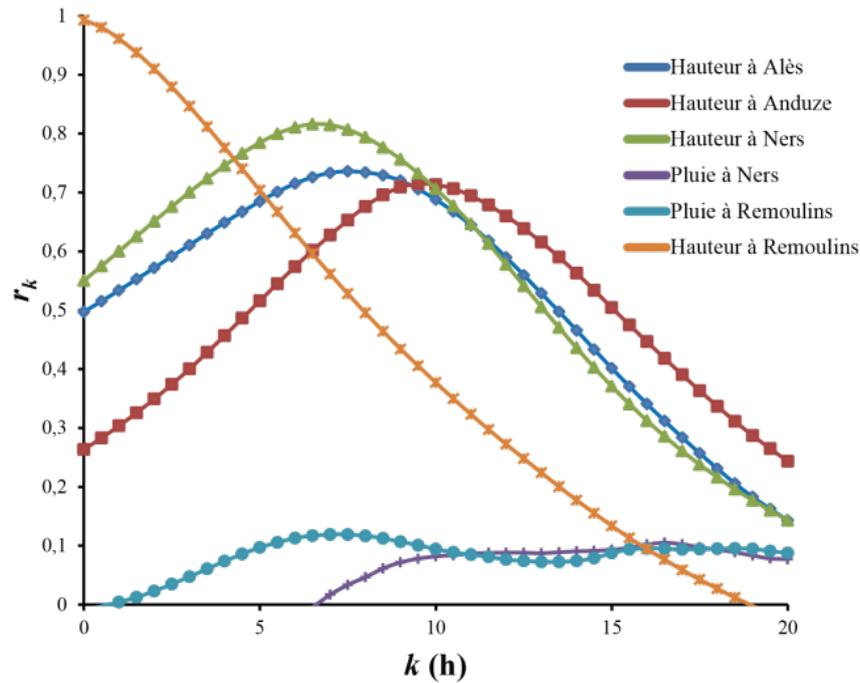


Figure 51 : Moyenne sur tous les événements des corrélogrammes croisés avec pour série de référence la hauteur à Remoulins

D'après [Mangin 1981], la corrélation seuil est de 0,2 : en deçà de cette valeur, on ne devrait pas considérer que les deux séries sont corrélées. Si l'on regarde les corrélogrammes reportés en Figure 51, on ne devrait pas tenir compte de la pluie. Le corrélogramme de la pluie à Ners est même négatif en deçà de 7h, ce qui traduit la non-linéarité de la relation entre la pluie à Ners et la hauteur à Remoulins. On peut effectivement constater que pour certains événements les cumuls de pluie à Ners et à Remoulins sont relativement faibles comparés à la réponse du cours d'eau : la crue est alors causée uniquement par la pluie à l'amont du bassin (par exemple événement 26 cf. Tableau 31) .Ainsi une analyse plus fine considérant les corrélogrammes, événement par événement, voir Annexe 10, montre que pour 11 événements sur 15 au moins, une des deux séries de pluie présente des valeurs de corrélation croisée supérieures à 0,2, mais dans chaque cas pour des k différents, ce qui conduit à cette valeur très faible valeur de la moyenne. Considérant que la moyenne masque l'importance des pluies, on gardera donc la pluie à Ners et à Remoulins comme variables.

L'abscisse qui correspond au maximum des corrélogrammes, est le décalage temporel entre la série de hauteur à Remoulins et celle de la grandeur en question, qui donne la meilleure corrélation entre les deux séries. C'est donc un retard de cet ordre, reporté Tableau 33, que l'on applique aux entrées du réseau par rapport à la hauteur à Remoulins.

	Hauteur à Alès	Hauteur à Anduze	Hauteur à Ners	Pluie à Ners	Pluie à Remoulins
k_{max} (h)	7,5	9,5	6,5	10	7

Tableau 33 : Décalage temporel par rapport à la hauteur à Remoulins pour une intercorrélacion maximale (selon la corrélation croisée)

Au vu de l'emplacement des stations (Figure 48), il est logique que la hauteur à Ners soit affectée d'un retard moins important que celles à Alès et Anduze. Les longueurs de drains et la distance des stations à Remoulins étant semblables, la différence de retard entre la hauteur à Alès et celle à Anduze peut s'expliquer par plusieurs points : la différence de pente moyenne : le Gardon d'Anduze à Anduze a une pente moyenne de 2,3% [Artigue 2012.a] et le Gardon d'Alès à Alès a une pente moyenne de 3,6% [Erlich 2012] ; des différences au niveau de la géologie et aussi la présence d'un endiguement sur près d'1km en aval de la station à Alès peuvent favoriser une propagation rapide.

Le retard à appliquer aux autres variables par rapport à la hauteur à Remoulins, ayant été estimé, on s'intéresse maintenant aux fenêtres temporelles des différentes hauteurs à appliquer au modèle. On va les faire varier, avec un pas de temps de 1h, d'une largeur de 30 min à 4h30 pour les hauteurs à Alès, Anduze et Ners et de 30 min à 3h30 pour la hauteur à Remoulins.

Pour les hauteurs à Alès, Anduze et Ners, on applique les retards initiaux mentionnés dans le Tableau 34. Ces retards ont été déterminés afin que, pour la fenêtre temporelle la plus large (4h30), le décalage estimé par corrélation croisée, k_{max} , se trouve au milieu de l'historique des pluies [$k_{max}-2h$; $k_{max}+2h$]. Cela tant que $k_{max}-2h$ est inférieur à h_p , le retard initial est nul sinon. Pour les pluies, la fenêtre temporelle, fixée à 1h30, correspond à [$k_{max}-2h$; k_{max}], tant que $k_{max}-2h$ est inférieur à h_p , le retard initial est nul sinon.

Retards initiaux (h)	$h_p=1h$	$h_p=3h$	$h_p=5h$	$h_p=7h$	$h_p=9h$	$h_p=11h$
Pluie à Ners	8,5	6,5	4,5	2,5	0,5	0
Pluie à Remoulins	5,5	3,5	1,5	0	0	0
Hauteur à Anduze	6,5	4,5	2,5	0,5	0	0
Hauteur à Alès	4,5	2,5	0,5	0	0	0
Hauteur à Ners	3,5	1,5	0	0	0	0
Hauteur à Remoulins	0	0	0	0	0	0

Tableau 34 : Pour chaque horizon de prévision, retard initial des différentes variables par rapport à l'instant de prévision

Pour chacune des 500 combinaisons de fenêtres temporelles des hauteurs à Anduze, Alès, Ners et Remoulins, nous avons effectué 20 initialisations différentes des paramètres du réseau, conduisant à 20 modèles. L'apprentissage est réalisé sur tous les événements de la base de données sauf le 13, le 26 et le 27. L'événement 13 sert d'ensemble d'arrêt précoce. La validation croisée n'a pas été utilisée ici car elle demande un temps considérable de calcul sans pour autant avoir donné des résultats probants pour le Gardon d'Anduze et la régularisation est assurée par l'arrêt précoce. On calcule la médiane des critères de persistance

sur les prévisions de l'ensemble d'apprentissage des 20 modèles ainsi obtenus. On choisit la combinaison de fenêtres temporelles pour laquelle la médiane des critères de persistance est la plus grande. Le Tableau 35 contient les fenêtres temporelles sélectionnées.

Fenêtres temporelles		$h_p=1h$	$h_p=3h$	$h_p=5h$	$h_p=7h$	$h_p=9h$	$h_p=11h$
Pluie à Ners	PM	1,5	1,5	1,5	1,5	1,5	1,5
	PM+PL	1,5	1,5	1,5	1,5	1,5	1,5
Pluie à Remoulins	PM	1,5	1,5	1,5	1,5	1,5	1,5
	PM+PL	1,5	1,5	1,5	1,5	1,5	1,5
Hauteur à Anduze	PM	4,5	2,5	2,5	3,5	3,5	4,5
	PM+PL	2,5	1,5	1,5	0,5	2,5	1,5
Hauteur à Alès	PM	2,5	1,5	4,5	4,5	1,5	2,5
	PM+PL	4,5	3,5	2,5	4,5	2,5	4,5
Hauteur à Ners	PM	4,5	3,5	4,5	4,5	1,5	3,5
	PM+PL	4,5	2,5	0,5	2,5	0,5	0,5
Hauteur à Remoulins	PM	3,5	3,5	2,5	2,5	3,5	1,5
	PM+PL	3,5	3,5	1,5	2,5	1,5	1,5

Tableau 35 : Pour chaque horizon de prévision, fenêtre temporelles des différentes variables d'entrée pour les deux types de modèles (PM = perceptron multicouche, PM+PL = perceptron multicouche + partie linéaire)

Comme pour les fenêtres de pluie dans le Chapitre 3, la largeur des fenêtres temporelles de hauteur n'évolue pas de manière déterminée en fonction de l'horizon de prévision. On remarque que le modèle à perceptron multicouche simple réclame des fenêtres plus larges pour la hauteur à Anduze et à Ners et plus faibles pour celle à Alès que le modèle avec partie linéaire. La largeur des fenêtres pour la hauteur à Remoulins est semblable dans les deux types de modèles.

6.6. Résultats

Les modèles élaborés au paragraphe précédent sont testés successivement sur deux événements intenses. L'un est court (événement 26), l'autre est long (événement 27). Pour chaque horizon de prévision et pour les deux types de modèles (perceptron multicouche et perceptron multicouche avec termes directs), les limnigrammes présentés sont la médiane à chaque pas de temps des prévisions des 20 modèles de l'étape de sélection.

6.6.1. Événement 26

Considérons tout d'abord la Figure 52, qui présente les limnigrammes prévus pour l'événement 26 aux différents horizons de prévision, et le Tableau 36, qui indique les critères de performances associés. La différence entre les deux types de modèles est peu marquée ; cependant les prévisions effectuées par les modèles à perceptron multicouche sont légèrement meilleures pour $h_p=1h$ et $h_p=3h$, alors que les modèles avec partie linéaires sont meilleurs pour les autres horizons de prévision.

Jusqu'à $h_p=9h$, les valeurs du *SPPD* évoluent de 79% à 96 % et celles du retard du pic prévu par rapport au pic observé, entre 30min et 1h, sont très satisfaisantes : les modèles nous renseignent donc bien sur la hauteur au moment du pic.

En revanche, la montée du pic est, hormis pour $h_p=1h$, retardée de plusieurs heures, ce qui gêne la qualité de la prévision. Les modèles ont tendance à prévoir plusieurs pics au lieu d'un seul, en particulier à $h_p=9h$ et $h_p=11h$. Le premier petit pic en début d'événement correspond à un petit pic semblable sur la hauteur à Alès et à Ners ; ce pic est visible sur les limnigrammes présentés en Annexe 9.

Les valeurs des critères de persistance sont très bonnes à cause des valeurs importantes des horizons de prévisions qui rendent la prévision naïve peu pertinente.

		$h_p=1h$	$h_p=3h$	$h_p=5h$	$h_p=7h$	$h_p=9h$	$h_p=11h$
C_P	PM	0,81	0,85	0,87	0,89	0,91	0,87
	PM+PL	0,77	0,80	0,89	0,92	0,93	0,87
C_N	PM	0,99	0,93	0,87	0,82	0,80	0,66
	PM+PL	0,99	0,91	0,89	0,86	0,84	0,66
PPD (%)	PM	101	94	87	83	105	105
	PM+PL	106	87	96	80	95	118
$SPPD$ (%)	PM	100	86	65	82	88	42
	PM+PL	96	85	80	79	79	50
Retard H_{pic} (h)	PM	1	1	1	0,5	0,5	2,5
	PM+PL	0,5	0,5	0,5	0,5	0,5	2,5

Tableau 36 : Pour chaque horizon de prévision, critères de performance calculés sur les prévisions médianes de l'événement 26 obtenues par les modèles à perceptron multicouche (PM) et avec une partie linéaire ajoutée (PM+PL)

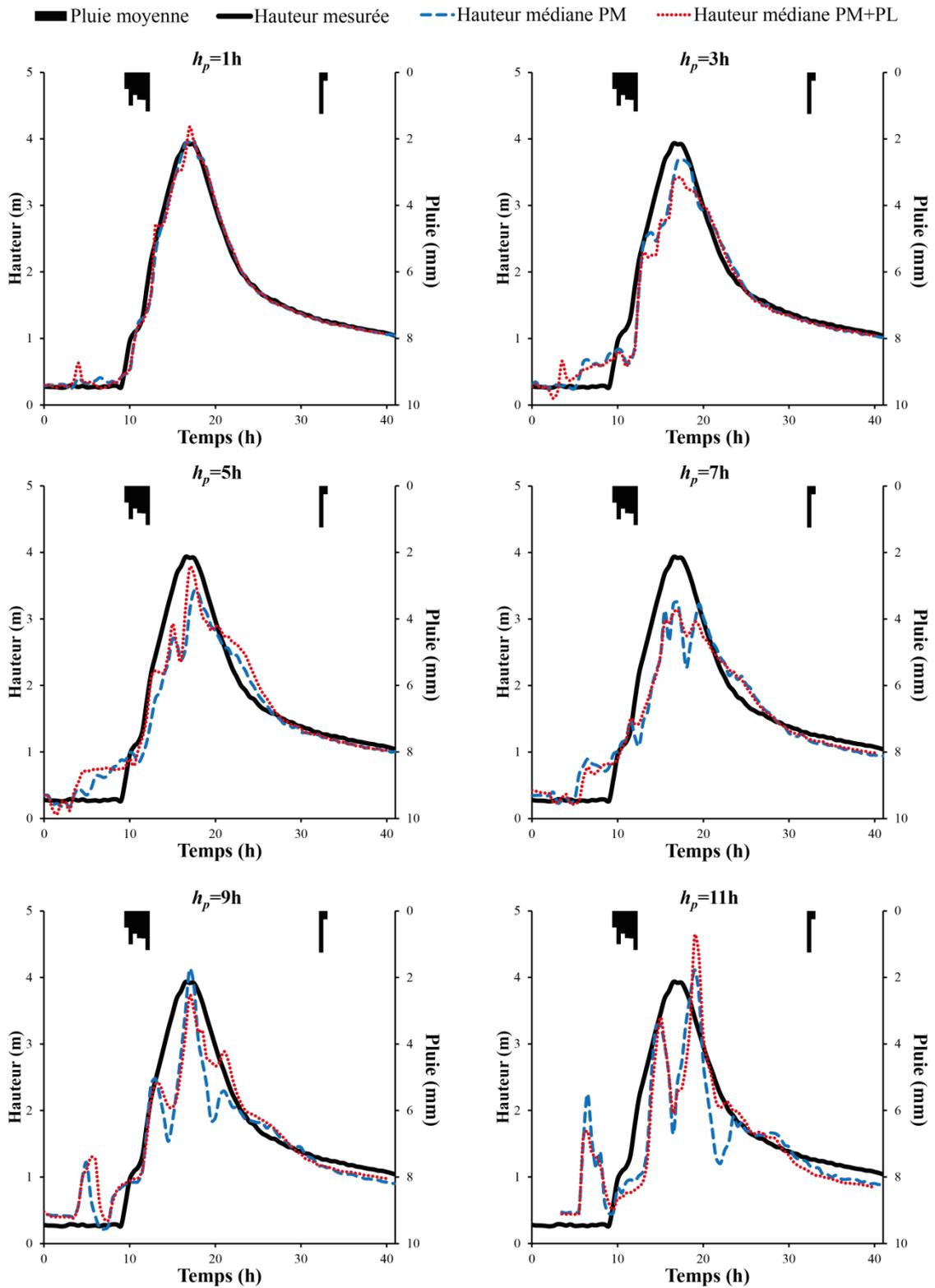


Figure 52 : Pour chaque horizon de prévision, hauteurs médianes prévues pour l'événement 26 par les modèles à perceptron multicouche (PM) et avec une partie linéaire ajoutée (PM+PL)

6.6.2. Événement 27

Les limnigrammes prévus pour l'événement 27 aux différents horizons de prévision sont regroupés sur la Figure 53, et les critères de performances correspondants dans le Tableau 37. Les prévisions réalisées par les deux types de modèles sont très semblables, plus encore que pour l'événement 26.

		$h_p=1h$	$h_p=3h$	$h_p=5h$	$h_p=7h$	$h_p=9h$	$h_p=11h$
C_P	PM	0,74	0,86	0,92	0,94	0,96	0,93
	PM+PL	0,67	0,83	0,89	0,96	0,96	0,90
C_N	PM	0,99	0,98	0,97	0,97	0,97	0,93
	PM+PL	0,99	0,97	0,96	0,98	0,97	0,90
PPD (%)	PM	100	99	96	95	96	103
	PM+PL	102	98	99	91	95	108
$SPPD$ (%)	PM	99	94	88	81	89	102
	PM+PL	98	92	84	89	89	102
Retard H_{pic} (h)	PM	1	1,5	2	1,5	-2	1
	PM+PL	1	1,5	1,5	2,5	-0,5	1

Tableau 37 : Pour chaque horizon de prévision, critères de performance calculés sur les prévisions médianes de l'événement 27 obtenues par les modèles à perceptron multicouche (PM) et avec une partie linéaire ajoutée (PM+PL)

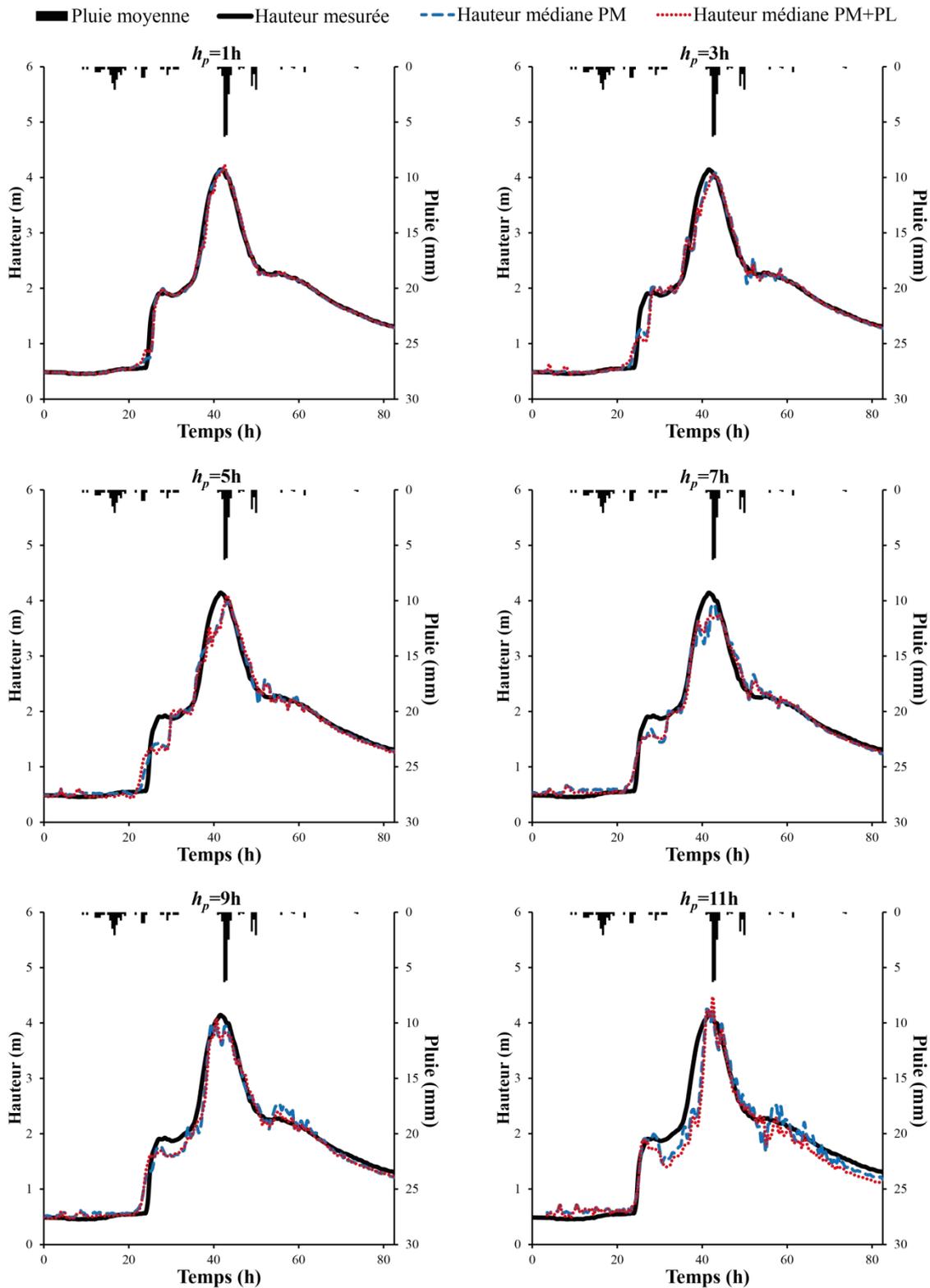


Figure 53 : Pour chaque horizon de prévision, hauteurs médianes prévues pour l'événement 27 par les modèles à perceptron multicouche (PM) et avec une partie linéaire ajoutée (PM+PL)

Les valeurs de *SPPD* comprises entre 84 et 102% sont bonnes. Le retard du pic prévu par rapport à l'observé reste en deçà de l'horizon de prévision, sauf pour $h_p = 1h$, et n'augmente pas pour les horizons les plus lointains ; le pic est même en avance pour $h_p = 9h$. La montée du

pic est correctement prévue par les modèles à l'exception de ceux pour $h_p=5h$ et $h_p=11h$, où la montée présente un retard de quelques heures.

6.7. Conclusion sur la prévision des crues à Remoulins

La superficie du bassin versant de Remoulins est plus de 3 fois supérieure à celle du bassin versant du Gardon d'Anduze, et la topographie est très différente entre l'amont et l'aval de ce bassin versant. Il n'est donc pas possible de généraliser directement les modèles construits pour la prévision des crues à Anduze en vue d'une utilisation à Remoulins. De plus, au niveau hydrologique, le phénomène de propagation de l'onde de crue apparaît sur l'aval alors qu'il est peu présent sur les amonts. Pour ces raisons, nous avons choisi de ne pas transposer les modèles à l'identique, mais d'adapter la méthode proposée au Chapitre 3 à cette configuration plus générale. Au lieu de n'utiliser que des données de pluie, on met à profit la connaissance des niveaux de crues sur les amonts acquise au moyen du réseau de mesure des hauteurs aux stations en amont de Remoulins : Anduze, Alès et Ners. Les données de pluies apportées au modèle sont celles du milieu du bassin versant à Ners et de l'aval à Remoulins. Il convient également de prendre en compte la propagation de la crue au moyen d'un retard pur, appliqué sur les fenêtres temporelles des hauteurs à Anduze, Alès et Ners, lorsque l'horizon de prévision est inférieur à ce retard dû à la propagation.

Avec les variables ainsi définies, les deux types de modèles élaborés au Chapitre 3 (perceptron multicouche et perceptron multicouche avec termes directs) ont été testés sur les événements 26 et 27, événements identifiés comme intenses mais moins que celui de septembre 2002 pour lequel nous ne disposons pas de séries de mesures exploitables. Ces deux types de modèles donnent des résultats semblables. L'événement 27, plus lent, est mieux prévu que le 26 où la hauteur au pic est relativement bien anticipée jusqu'à un horizon de 5h. Par contre la montée du pic est très retardée par rapport à celle du pic observé. Les prévisions obtenues pour l'événement 27 sont satisfaisantes jusqu'à $h_p=9h$.

Par rapport aux objectifs du projet FLASH qui devait étudier la généralisation des méthodes développées sur le site d'Anduze à des bassins de taille différente, il apparaît que, si les modèles ne sont pas directement transposables, en revanche, la méthodologie de conception est tout à fait adaptée et permet d'envisager la réalisation de systèmes d'alerte performants jusqu'à 7 heures à l'avance pour la station de Remoulins. Le remplacement des pluies mesurées à l'amont par des hauteurs pourrait être appliqué au bassin d'Anduze avec les stations amont de Mialet, de Saumane et de St Jean du Gard. Cette dernière, installée en 2003, n'a pour l'instant pas permis d'enregistrer suffisamment d'événements pour envisager un apprentissage statistique.

Conclusion générale

Dans un contexte où la prévision des crues éclair nécessite d'être améliorée afin de réduire le nombre de victimes des inondations dévastatrices qui en résultent régulièrement, les travaux réalisés au cours de cette thèse avaient pour objectif d'étudier la conception de modèles issus de l'apprentissage statistique afin de prévoir la hauteur des cours d'eau lors de tels phénomènes.

L'utilisation de l'apprentissage statistique est une alternative aux modélisations basées sur des équations décrivant des processus physiques complexes et mal connus. Les propriétés d'approximation universelle et de parcimonie font des réseaux de neurones, en particulier, un outil intéressant. Néanmoins, comme nous l'avons présenté dans le premier chapitre, leur utilisation doit se faire avec rigueur afin de préserver les modèles du surajustement et garantir ainsi des résultats fiables en généralisation.

Au cours du second chapitre nous avons pu présenter plusieurs obstacles à l'utilisation de l'apprentissage statistique. La base de données est très réduite et les événements la constituant présentent des caractéristiques variées tant pour la distribution spatiale et temporelle des précipitations que pour la forme de la réponse du cours d'eau. De nombreuses sources d'erreurs et d'incertitudes sur les mesures s'ajoutent à cela, entachant les données d'un bruit considérable. Néanmoins la capacité de l'apprentissage statistique de bâtir des modèles différents des modèles physiques permet de réaliser des prédicteurs de hauteur d'eau, même *en l'absence de prévision des pluies* au-delà de l'instant où est effectuée la prévision. Ce dernier avantage est un élément qu'il convient de souligner.

Ainsi, le Chapitre 3 montre comment, pour chaque horizon de prévision souhaité, en l'absence de prévision de pluies, nous avons construit des modèles composés d'un perceptron multicouche, qui calculent la hauteur d'eau prévue à Anduze à partir de valeurs passées de cette hauteur et de la pluie passée au niveau de six pluviomètres répartis sur le bassin versant. L'arrêt précoce sert de méthode de régularisation pour éviter le surajustement. Avec une méthode dérivée de la validation croisée, nous avons sélectionné les hyperparamètres des modèles : le nombre de neurones cachés, la largeur de la fenêtre temporelle des pluies, la valeur initiale du pas d'apprentissage de l'algorithme de Levenberg Marquardt et celle de son facteur de décroissance, et enfin, l'initialisation des paramètres du réseau. Alors que l'estimation des capacités de généralisation des modèles se révèle peu sensible à tous les autres hyperparamètres, le choix de l'initialisation des paramètres du modèle paraît déterminant. Dans le but d'améliorer les performances des modèles, des réseaux comportant une partie linéaire sont proposés et nous avons pu montrer que, outre une réduction efficace de la dispersion des prévisions provenant de modèles avec diverses initialisations des paramètres, les prévisions de l'événement le plus intense de la base sont nettement améliorées. En revanche, la synchronisation entre pic prévu et pic observé est moins bonne que pour les modèles sans partie linéaire. Ainsi, les prévisions pour l'événement le plus intense de la base, celui de septembre 2002, sont intéressantes jusqu'à un horizon de prévision de 3h et celles de deux autres événements intenses, ceux d'octobre et novembre 2008 sont satisfaisantes jusqu'à 2 ou 3h.

Dans le Chapitre 4, compte tenu de l'écart entre les performances des modèles sur l'ensemble d'apprentissage et leurs performances sur les événements en test, nous avons

souhaité rendre les modèles adaptatifs : nous avons ainsi modifié les paramètres en cours d'événement afin qu'ils s'ajustent à l'événement étudié. Les résultats ont été décevants : la prévision de l'amplitude du pic est améliorée grâce à une meilleure prise en compte des précipitations, mais le pic est prévu avec retard en raison du décalage temporel entre le moment où s'effectue l'adaptation du modèle et l'instant de la prévision, qui est égal à deux fois l'horizon de prévision.

Une autre piste d'amélioration de la qualité des prévisions a été abordée dans le Chapitre 5 : il s'agit d'améliorer la représentation du champ de pluies en appliquant au modèle les mesures issues de radar plutôt que les seules mesures ponctuelles issues des pluviomètres. Dans le cadre de l'étude systémique de la chaîne hydrométéorologique, cette approche est tout à fait complémentaire à la précédente ; après plusieurs traitements des données brutes correspondant à des images en niveaux de réflectivité, nous avons obtenu des images de lame d'eau dont six valeurs associées à un découpage du bassin en six sous-bassins ont été extraites. Nous avons constaté que les cumuls de pluie provenant du radar sont en moyenne de l'ordre de 40% inférieurs à ceux des cumuls de pluies fournis par les pluviomètres. La comparaison des prévisions sur deux événements en test avec les deux types de données ne permet pas de privilégier l'utilisation de l'un ou l'autre des types de données comme variables des modèles. L'utilisation d'une base de données plus étoffée, une meilleure connaissance des taux de masquages et un découpage en zones différent sont des pistes à suivre pour améliorer les résultats obtenus avec les données radar.

Pour terminer notre étude des modèles à apprentissage statistique pour la prévision des crues rapides, nous décrivons, dans le Chapitre 6, les modèles prédictifs que nous avons conçus pour la hauteur d'eau du Gardon à Remoulins, dont le bassin versant est environ trois fois plus grand que celui du Gardon d'Anduze, et inclut celui-ci. L'ensemble des variables des modèles a été étendu afin de prendre en compte la connaissance du bassin et l'offre plus importante de stations de mesures. Ces modèles sont alimentés par les valeurs passées des hauteurs amont et aval, ainsi que par les précipitations mesurées au milieu et à l'aval du bassin. La méthodologie proposée au Chapitre 3 a été appliquée à ce bassin ; les prévisions obtenues se sont révélées intéressantes jusqu'à un horizon de prévision de 7h ou 9h en fonction des événements considérés. Compte tenu de la diversité des phénomènes hydrologiques entrant en jeu du fait de la diversité de la topographie et de la géologie du bassin, on peut donc conclure que l'approche développée à partir de réseaux de neurones apporte une solution particulièrement simple à mettre en œuvre et peu gourmande en mesures de terrain pour effectuer la prévision de manière efficace en l'absence de prévision de pluies.

Enfin, les modèles réalisés sur le bassin versant du Gardon d'Anduze sont actuellement implémentés par la société Géonosis pour le SCHAPI. La première version est un simulateur déjà mis à disposition des partenaires du projet FLASH. Néanmoins, même si les résultats de cette étude sont encourageants et peuvent déjà être valorisés par le SCHAPI, il reste de nombreux travaux à effectuer, notamment la conception de modèles dynamiques et l'approfondissement des travaux sur l'adaptativité ou l'assimilation de données.

Bibliographie

[Alfieri 2012]

L. Alfieri, P. Salamon, F. Pappenberger, F. Wetterhall, J. Thielen, “*Operational early warning systems for water-related hazards in Europe*”, Environmental Science & Policy, p. 35-49, 2012

[Amengual 2007]

Amengual, R. Romero, M. Gomez, A. Martin, S. Alonso, “*A Hydrometeorological Modeling Study of a Flash-Flood Event over Catalonia*”, Spain, Journal of hydrometeorology, vol 8, p. 282-303, 2007

[Artigue 2012.a]

G. Artigue, “*Prévision des crues éclair par réseaux de neurones : généralisation aux bassins non jaugés*”, thèse de doctorat, Université Montpellier II, 2012

[Artigue 2012.b]

G. Artigue, A. Johannet, V. Borrell, S. Pistre, “*Flash flood forecasting in poorly gauged basins using neural networks: case study of the Gardon de Mialet basin (southern France)*”, NHESS in Press, 2012

[Ayrat 2005]

P.A. Ayrat, “*Contribution à la spatialisation du modèle opérationnel de prévision des crues éclair ALHTAIR ; approches spatiale et expérimentale ; application au bassin versant d’Anduze*”, thèse de doctorat, Université d’Aix-Marseille, 2005

[Barron 1993]

A.R. Barron, “*Universal Approximation Bounds for Superpositions of a Sigmoidal Function*”, IEEE Transactions on Information Theory, IT-39, p. 930-945, 1993

[Bartlett 1997]

P. L. Bartlett, “*For valid generalization, the size of the weights is more important than the size of the network*”, Neural Information Processing Systems, vol 9, Morgan Kaufmann, 1997

[Beven 1979]

K.J. Beven, M.J. Kirkby, “*A physical based, variable contributing areas model of basin hydrology*”, Hydrological Sciences Bulletin, 24, p. 43-69, 1979

[Bonnifait 2009]

L. Bonnifait, G. Delrieu, M. Le Lay, B. Boudevillain, A. Masson, P. Belleudy, E. Gaume, G.-M. Saulnier, “*Distributed hydrologic and hydraulic modelling with radar rainfall input: Reconstruction of the 8–9 September 2002 catastrophic flood event in the Gard region, France*”, Advances in Water Resources, 32, p 1077–1089, 2009

[Bornancin-Plantier 2011]

A. Bornancin-Plantier, A. Johannet, P. Roussel-Ragot, G. Dreyfus, “*Flash Flood Forecasting using Neural Networks without Rainfall Forecasts: Model Selection and Generalization capability*”, EGU 2011-1794, présentation orale, 2011

[Bouvier 2006]

C. Bouvier, P.A. Ayrat, P. Brunet, A. Marchandise, C. Martin, “*Recent advance in rainfall-runoff modelling: extrapolation to extreme floods in southern France*”, International Workshop on Hydrological Extreme, Rende, p. 3-4, 2006

[Bouvier 2010]

Y. Trambly, C. Bouvier, C. Martin, J.-F. Didon-Lescot, D. Todorovik, J.-M. Domergue, “*Assessment of initial soil moisture conditions for event-based rainfall-runoff modelling*”, Journal of Hydrology, 387, Issues 3–4, 15, p. 176-187, 2010

[Braud 2010]

I. Braud, H. Roux, S. Anquetin, M.M. Maubourguet, C. Manus, P. Viallet, D. Dartus, “*The use of distributed hydrological models for the Gard 2002 flash flood event: Analysis of associates hydrological processes*”, Journal of Hydrology, 394 1-2, p. 162-181, 2010

[Charlier 2012]

J.-B. Charlier, C. Bertrand, J. Mudry, “*Conceptual hydrogeological model of flow and transport of dissolved organic carbon in a small Jura karst system*”, Journal of Hydrology, 460–461, p. 52–64, 2012

[Coustau 2011]

M. Coustau, S. Ricci, A. Piacentini, O. Thual, A. Marchandise, C. Bouvier, V. Borrell-Estupina, “*Calibration of a rainfall-runoff hydrological model and flood simulation using data assimilation*”, Geophysical Research Abstracts, vol. 13, EGU2011-281-2, 2011

[Diettrich 2000]

T.G. Diettrich, “*Ensemble Methods in Machine Learning*”, in J. Kittler and F. Roli (Ed.), First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, p. 1-15, New York: Springer Verlag, 2000

[Dimopoulos 1996]

I. Dimopoulos, S. Lek, J. Lauga, “*Modélisation de la relation pluie-débit par les réseaux connexionnistes et le filtre de Kalman*”, Hydrological Sciences –Journal des Sciences Hydrologiques, 41(2), p. 179-193, 1996

[Dolciné 2001]

L. Dolciné, H. Andrieu, D. Sempere-Torres, D. Creutin, “*Flash Flood Forecasting with Coupled Precipitation Model in Mountainous Mediterranean Basin*”, Journal of Hydrology, 2001

[Dreyfus 2006]

G. Dreyfus, I. Guyon, Assessment methods, in *"Feature extraction, foundations and applications"*, edited by I. Guyon, S. Gunn, M. Nikravesh and L. Zadeh, Springer, 2006

[Dreyfus 2008]

G. Dreyfus, J.-M. Martinez, M. Samuelides, M.B. Gordon, F. Badran, S. Thiria, *"Apprentissage statistique"*, Eyrolles, Algorithmes, 417 p., 2008

[Dunne 1983]

T. Dunne, *"Relation of field studies and modeling in the prediction of storm runoff"*, Journal of Hydrology, 65, p. 25-48, 1983

[Erlich 2012]

M. Erlich, C. Girard, E. David, *"Intégration du modèle hydraulique 2D dans la chaîne de prévision des crues éclair en temps réel"*, symposium européen sur les problèmes actuels de la protection contre les inondations, mars 2012

[Estupina Borrel 2004]

V. Estupina Borrel, *"Vers une modélisation hydrologique adaptée à la prévision opérationnelle des crues éclair. Application à de petits bassins versants du sud de la France"*, thèse de doctorat, Institut National Polytechnique de Toulouse, 2004

[Gallinari 1999]

P. Gallinari, T. Cibas, *"Practical complexity control in multilayer perceptrons"*, Signal processing, vol 74, p. 29-46, 1999

[Gaume 2009]

E. Gaume, V. Bain, P. Bernardara, O. Newinger, M. Narbuc, A. Bateman, L. Blaškovičová, G. Blöschl, M. Borga, A. Dumitrescu, I. Daliakopoulos, J. Garcia, A. Irimescu, S. Kohnova, A. Koutroulis, L. Marchi, S. Matreata, V. Medina, E. Preciso, D. Sempere-Torres, G. Stancalie, J. Szolgay, I. Tsanis, D. Velasco, A. Viglione, *"A compilation of data on European flash floods"*, Journal of Hydrology, 367, p. 70-78, 2009

[Geman 1992]

S. Geman, E. Bienenstock, R. Doursat, *"Neural Networks and the Bias/Variance Dilemma"*, Neural Computation, vol 4, p. 1-58, 1992

[Govindaraju 2000]

R.S. Govindaraju, *"Artificial neural networks in hydrology II: Hydrologic application"*, Journal of Hydrologic Engineering, 5(2), p. 124-137, 2000

[Hewlett 1961]

J.D. Hewlett, “*Watershed management some ideas about storm runoff and baseflow*”, USDA Forest Service, 1961

[Hornik 1989]

K. Hornik, M. Stinchcombe, H. White, “*Multilayer feedforward networks are universal approximators*”, Neural Networks, vol 2, p. 359-366, 1989

[Horton 1933]

R.E. Horton, “*The role of infiltration in the hydrological cycle*”, Transaction American Geophysical Union, 14, p. 446-460, 1933

[Johannet 2008]

A. Johannet, A. Mangin, B. Vayssade, “*Modélisation d’un système karstique par réseaux de neurones. Simulation des débits du karst du Baget-France*”, Cahiers de Géographie-collection EDYTEM vol. 7, p. 51-62, 2008

[Jordan 2003]

P.W. Jordan, A.W. Seed, P.E. Weinmann, “*A stochastic model of radar measurement errors in rainfall accumulations at catchment scale*”, Journal of Hydrometeorology, 4, p.841-855, 2003

[Kitadinis 1980]

P.K. Kitadinis, R. Bras, “*Real-Time forecasting with a conceptual hydrologic model, applications and results*”, Water Resources Research, vol 16, p. 1034-1044, 1980

[Kong A Siou 2011]

L. Kong A Siou, K.Cros, A. Johannet, V. Borrel Estupina, S. Pistre, “*Modélisation hydrodynamique des karsts par réseaux de neurones: comment dépasser la boîte noire ?*”, Actes du 9^{ème} colloque d’hydrogéologie en pays calcaire, 1-3 septembre, Besançon, p 265-268, 2011

[Kong A Siou 2012.a]

L. Kong A Siou, A. Johannet, V. Borrel Estupina, S. Pistre, “*Optimization of the generalization capability for rainfall-runoff modeling by neural networks: the case of the Lez aquifer (southern France)*”, Environmental Earth Sciences, 65, p. 2365-2375, 2012

[Kong A Siou 2012.b]

L. Kong A Siou, P. Feury, A. Johannet, V. Borrell Estupina, N. Dörfliger, S. Pistre, “*How inputs of an hydrologic model have to be adjusted to its underlying physical hypothesis? Case study on the Lez hydrodynamic modeling (Southern France)*”, Assemblée Générale EGU, 2012

[Le Lay 2007]

M. Le Lay, G. M. Saulnier, “*Exploring the signature of climate and landscape spatial variabilities in flash flood events: Case of the 8-9 September 2002 Cévennes-Vivarais catastrophic event*”, *Geophysical Research Letters*, vol. 34, L13401, 5 p., 2007

[Maier 2000]

H.R. Maier, G.C. Dandy, “*Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications*”, *Environmental Modelling and Software* 15, p. 101-124, 2000

[Mangin 1975]

A. Mangin, “*Contribution à l'étude hydrodynamique des aquifères karstiques*”, Thèse Université de Dijon, Laboratoire souterrain du Centre national de la recherche scientifique, 1975

[Mangin 1981]

A. Mangin, “*Utilisation des analyses corrélatoire et spectrale dans l'approche des systèmes hydrologiques*”, *CR Acad. Sci. Paris*, 293, p.401-404, 1981

[Marchandise 2007]

A. Marchandise, “*Modélisation hydrologique distribuée sur le Gardon d'Anduze ; étude comparative de différents modèles pluie-débit, extrapolation de la normale à l'extrême et tests d'hypothèses sur les processus hydrologiques*”, thèse de doctorat, Montpellier II, 2007

[Marquardt 1963]

D.W. Marquardt, “*An algorithm for least-squares estimation of nonlinear parameters*”, *Journal of the Society of Industrial and Applied Mathematics*, vol 11, p 431-441, 1963

[Mathevet 2005]

T. Mathevet, “*Quels modèles pluie-débit globaux au pas de temps horaire ? Développements empiriques et comparaison de modèles sur un large échantillon de bassins versants*”, thèse de doctorat, Ecole Nationale du Génie Rural des Eaux et des Forêts, 2005

[McCulloch 1943]

W.S. McCullochs, W. Pitts, “*A logical calculus of the ideas immanent in nervous activity*”, *Bulletin of Mathematical Biophysics*, vol 5, p. 115-133, 1943

[McKay 1992]

D.J.C. McKay, “*A practical Bayesian framework for backpropagation networks*”, *Neural Computation*, vol 4, p. 448-472, 1992

[Moussa 1991]

R. Moussa, “*Variabilité spatio-temporelle et modélisation hydrologique. Application au bassin du gardon d'Anduze*”, thèse de doctorat, Montpellier II, 1991

[Nash 1970]

J.E. Nash, J.V. Sutcliffe, “*River Flow Forecasting through Conceptual Models. Part I – A Discussion of Principles*”, Journal of Hydrology, vol 10, p. 282-290, 1970

[Nerrand 1993]

O. Nerrand, P. Roussel-Ragot, L. Personnaz, G. Dreyfus, S. Marcos, “*Neural networks and non –linear adaptative filtering: unifying concepts and new algorithms*”, Neural computation, vol5, p. 165-197, 1993

[Obled 2009]

C. Obled, I. Zin, B. Hingray, “*Choix des pas de temps et d’espace pour des modélisations parcimonieuses en hydrologie des crues*”, la Houille Blanche, 5, p.81-87, 2009

[Roussel 2001]

P. Roussel, F. Moncet, B. Barrieu, A. Viola, “*Modélisation d’un processus dynamique à l’aide de réseaux de neurones bouclés. Application à la modélisation de la relation pluie-hauteur d’eau dans un réseau d’assainissement et à la détection de défaillance de capteur*”, Innovative Technologies in urban drainage, 1, p. 919-926, G.R.A.I.E., 2001

[Rumelhart 1986]

D.E. Rumelhart, G. E. Hinton, R. J. Williams , “*Learning Internal Representations by error backpropagation*”, Parallel Distributed Processing : Explorations in the Microstructure of Cognition, p. 318-362, MIT Press, 1986

[Saulnier 2009]

G.M. Saulnier, M. Le Lay, “*Sensitivity of flash flood simulations on the volume, the intensity and the localization in the Cévennes-Vivarais region (France)*”, Water Resources Research, 45, 2009

[SIEE 2004]

SIEE – Société d’Ingénierie Eau et Environnement, “*Validation des relevés hydrométriques de l’événement des 8 & 9 septembre 2002* ”, Rapport d’Etude, Direction Départementale de l’Equipement du Gard, 162 p., 2004

[Sjöberg 1992]

J. Sjöberg, L. Ljung, “*Overtraining regularization, and searching for minimum in neural networks*”, Preprint IFAC Symposium on Adaptive Systems in Control and Signal Processing, 1992

[SPCGD 2010]

Service de Prévision des Crues Grand Delta, “*Règlement de surveillance de prévision et de transmission de l’information sur les crues (RIC)*”, Préfecture du Gard-direction des Territoires et de la Mer, version modifiée en 2010

[Stone 1974]

M. Stone, “*Cross-validatory choice and assessment of statistical predictions*”, Journal of the Royal Statistical Society, B36, p. 111-147, 1974

[Stoppiglia 2003]

H. Stoppiglia, G. Dreyfus, R. Dubois, Y. Oussar, “*Ranking a Random Feature for Variable and Feature Selection, Journal of Machine Learning Research*”, p. 1399-1414

[Tabary 2012]

P. Tabary, P. Dupuy, G. L’Hénaff, O. Laurentin, C. Merlier, J.M. Soubeyrou, “*A 10-year (2000-2006) reanalysis of quantitative precipitation over France: methodology and first results*”, Hydrological Sciences Journal, in press, 11p., 2012

[Thimm 1997]

G. Thimm, E. Fiesler, “*High-Order and Multilayer Perceptron Initialization*”, IEEE Transactions on neural networks, vol8, 1997

[Toukourou 2009.a]

M.S. Toukourou, “*Application de l’apprentissage artificiel à la prévision des crues éclair*”, thèse de l’Ecole Nationale Supérieure des Mines de Paris, Spécialité Informatique Temps Réel, Robotique et Automatique, 175p, 2009

[Toukourou 2009.b]

M.S. Toukourou, A. Johannet, G Dreyfus, “*Flash Flood Forecasting by Statistical Learning in the Absence of Rainfall Forecast: A Case Study*”, Engineering Applications of Neural Networks, p. 98-107, 2009

[Toukourou 2010]

M.S. Toukourou, A. Johannet, G Dreyfus, P.A. Ayrat, “*Rainfall-runoff modeling of flash floods in the absence of rainfall forecasts: the case of Cevenol flash floods*” Journal of Applied Intelligence, 35 2, p. 178-189, 2010

[Vapnik 1998]

V. Vapnik, “*The nature of statistical learning theory*”, Springer, 1998

[Vigicrues]

<http://www.vigicrues.gouv.fr/>

[Visserrot 2012]

Y. Visserrot, “*Intégration de l’information hydrométéorologique spatialisée pour la modélisation et la prévision des crues éclair à l’aide des réseaux de neurones*”, Mémoire de Master Pro Gestion des Catastrophes et des Risques Naturels, Département de Géographie UFR III de l’université Paul Valérie, Montpellier III, 2012

Annexes

<i>Annexe 1 : La méthode de Levenberg-Marquardt (chapitre 1).....</i>	<i>148</i>
<i>Annexe 2 : Limnigrammes et hyétogrammes des événements compris dans la base de données pour le Gardon d'Anduze (chapitres 2, 3 et 4).....</i>	<i>149</i>
<i>Annexe 3 : Tableaux de l'écart-type et du maximum des 200 scores de validation croisée partielle (chapitre3).....</i>	<i>158</i>
<i>Annexe 4 : Tableaux de critères pour les sorties des 20 modèles à perceptron multicouche (chapitre 3)</i>	<i>164</i>
<i>Annexe 5 : Comparaison entre les sorties des modèles à perceptron multicouche sélectionnés par validation croisée partielle (médiane des 20 modèles, chapitre 3) et ceux sélectionnés sur l'événement en test (meilleurs critères en test).....</i>	<i>173</i>
<i>Annexe 6 : Tableaux des critères pour les différentes fenêtres d'adaptativité (chapitre 4).....</i>	<i>176</i>
<i>Annexe 7 : Limnigrammes et hyétogrammes (pluie issue des pluviomètres et pluie obtenue par le radar) des événements compris dans la base de données pour le Gardon d'Anduze (chapitres 5).....</i>	<i>180</i>
<i>Annexe 8 : Comparaison des hauteurs prévues selon le choix des modèles sélectionnés.....</i>	<i>186</i>
<i>Annexe 9 : Limnigrammes et hyétogrammes des événements compris dans la base de données pour le Gardon à Remoulins (chapitres 6)</i>	<i>188</i>
<i>Annexe 10 : Pour chaque événement, corrélogrammes croisés calculés par rapport à la hauteur à Remoulins (chapitre 6).....</i>	<i>196</i>
<i>Annexe 11 : Résumé de la présentation effectuée pendant la conférence European Geosciences Union 2011 à Vienne</i>	<i>199</i>
<i>Annexe 12 : Article pour Neurocomputing (en cours de finalisation, sera rajouté dans la version du mémoire remise au jury le jour de la soutenance).....</i>	<i>202</i>

Annexe 1 : La méthode de Levenberg-Marquardt (chapitre 1)

La méthode de Levenberg-Marquardt [Marquardt 1963] est une méthode d'optimisation de second ordre. La matrice hessienne, \mathbf{H} , est remplacée par \mathbf{H}_{lv} , où les composantes de \mathbf{H} ont subi l'approximation suivante : les dérivées secondes sont transformées en produit de dérivées premières; de plus un terme diagonal est ajouté, paramétré par le scalaire μ .

$$\mathbf{H}_{lv} = \mathbf{H} + \mu \mathbf{I}$$

Ainsi à l'itération i [Dreyfus 2008] :

$$\mathbf{w}^i = \mathbf{w}^{i-1} - \left[\mathbf{H}(\mathbf{w}^{i-1}) + \mu \mathbf{I} \right]^{-1} \nabla J(\mathbf{w}^{i-1})$$

avec \mathbf{w} le vecteur des paramètres du réseau

∇J : le vecteur des dérivées premières de la fonction de coût par rapport aux paramètres

Lorsque μ est petit par rapport aux coefficients diagonaux de \mathbf{H} alors $\mu \mathbf{I}$ est négligeable devant \mathbf{H} et on retrouve la méthode de Newton.

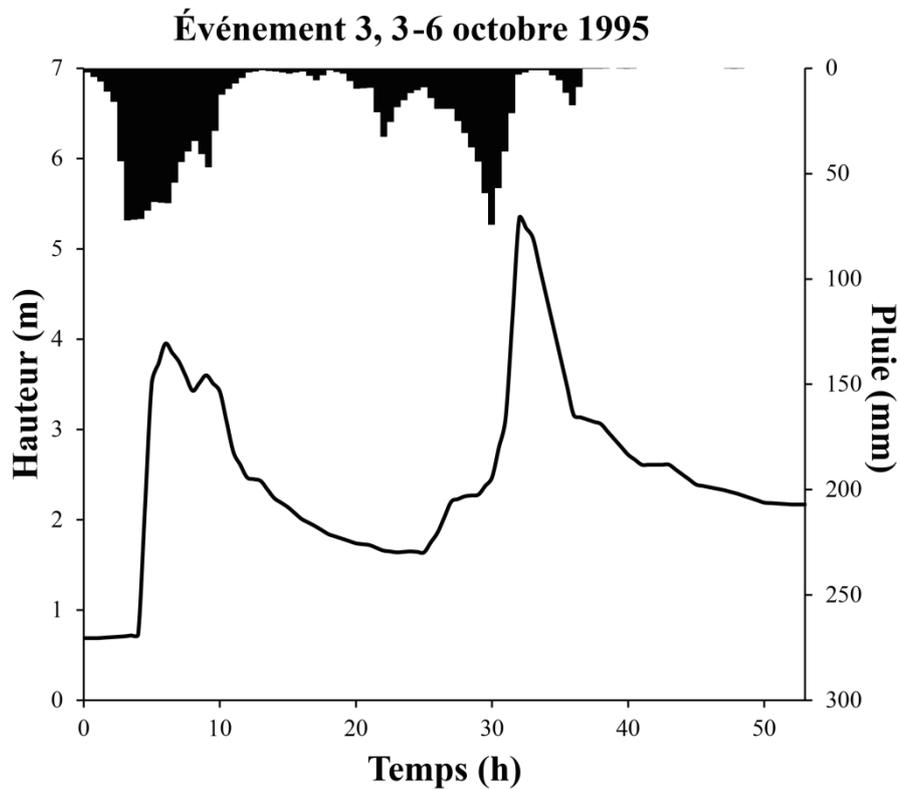
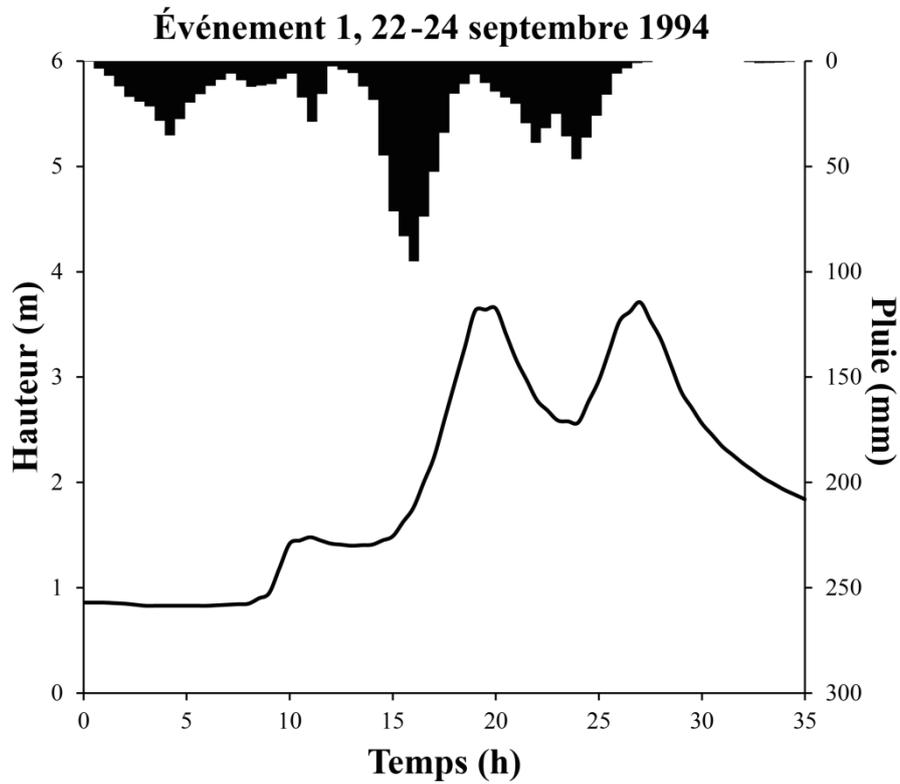
Si μ est grand par rapport aux coefficients diagonaux de \mathbf{H} , alors \mathbf{H} est négligeable devant $\mu \mathbf{I}$ et on retrouve une méthode de gradient simple avec un pas de $1/\mu$.

On choisit le pas d'apprentissage pour la première itération : μ_{ini} .

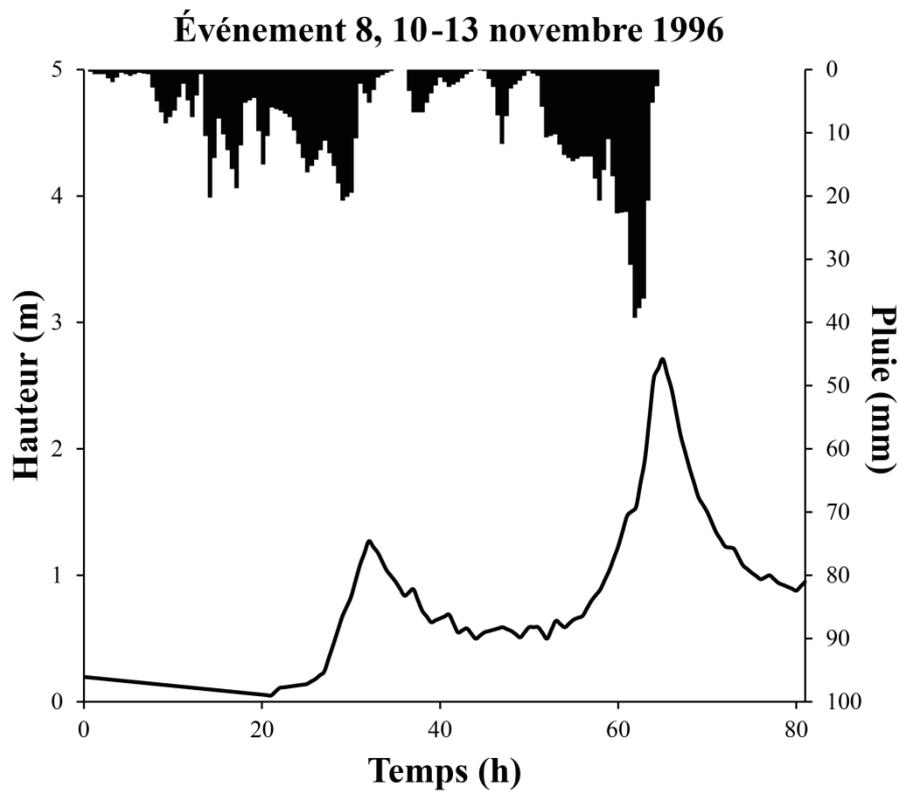
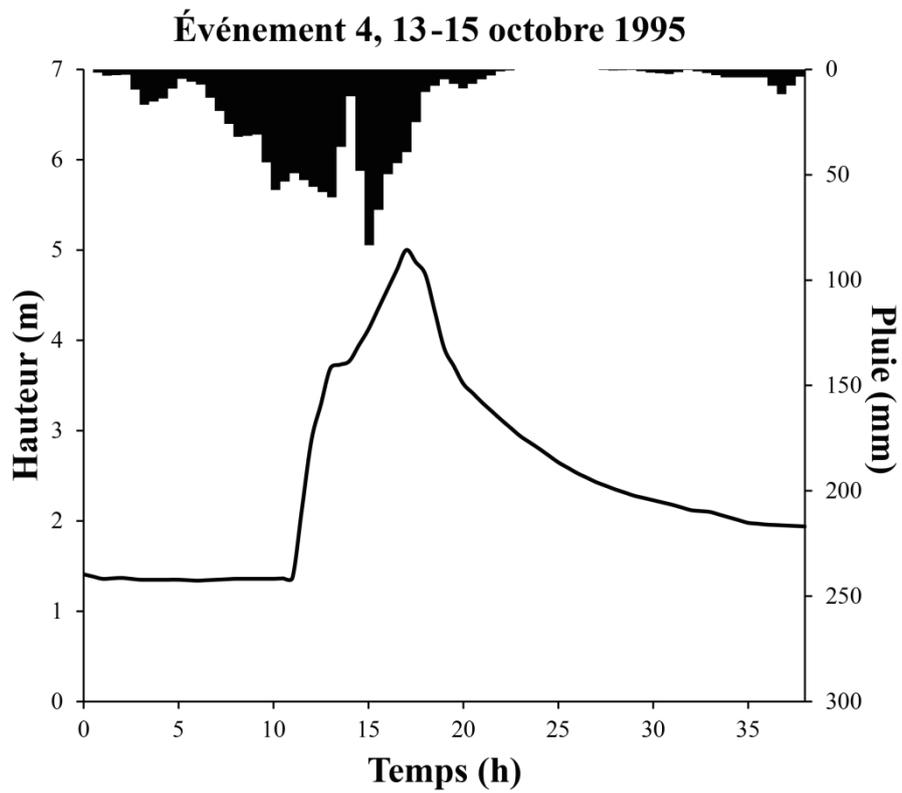
Si $J(\mathbf{w}^i) < J(\mathbf{w}^{i-1})$, alors on accepte le changement de paramètres et on diminue la valeur du pas par un facteur μ_{dec} . Sinon, on garde les valeurs des paramètres de l'itération précédente et on augmente la valeur du pas d'un facteur $1/\mu_{dec}$.

Annexe 2 : Limnigrammes et hyétogrammes des événements compris dans la base de données pour le Gardon d'Anduze (chapitres 2, 3 et 4)

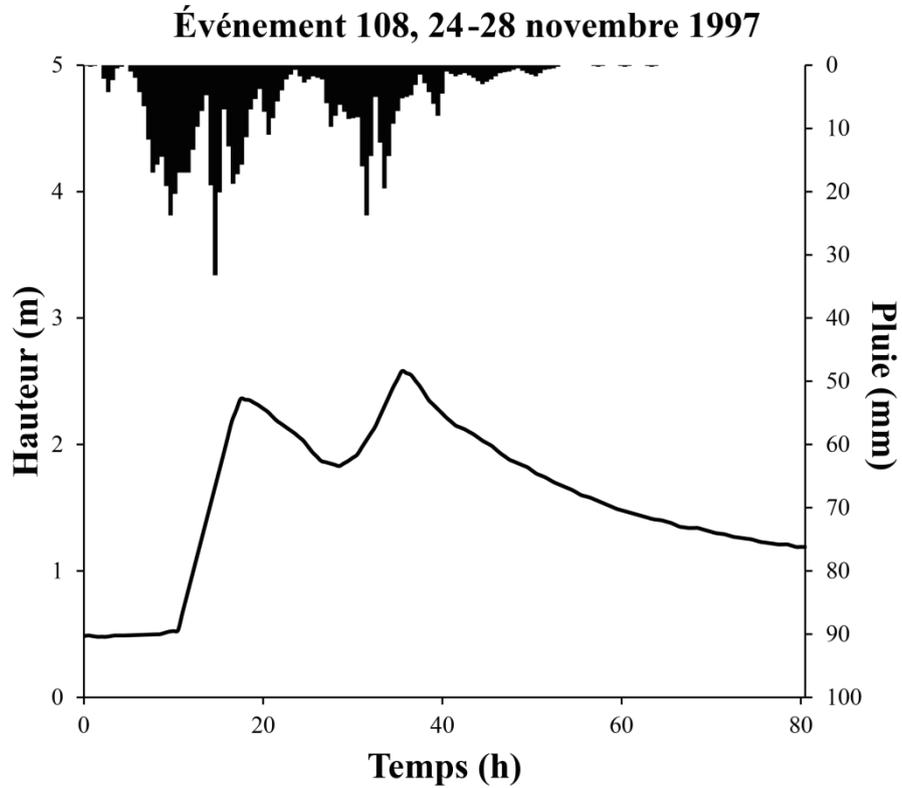
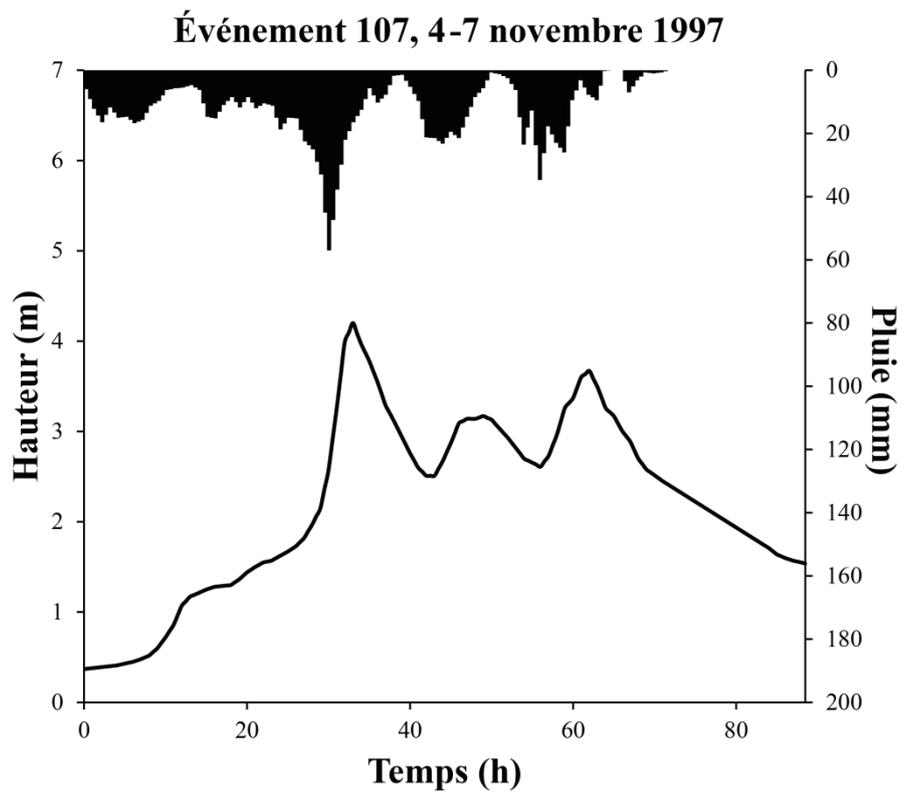
■ Cumul des pluies sur les 6 pluviomètres — Hauteur à Anduze



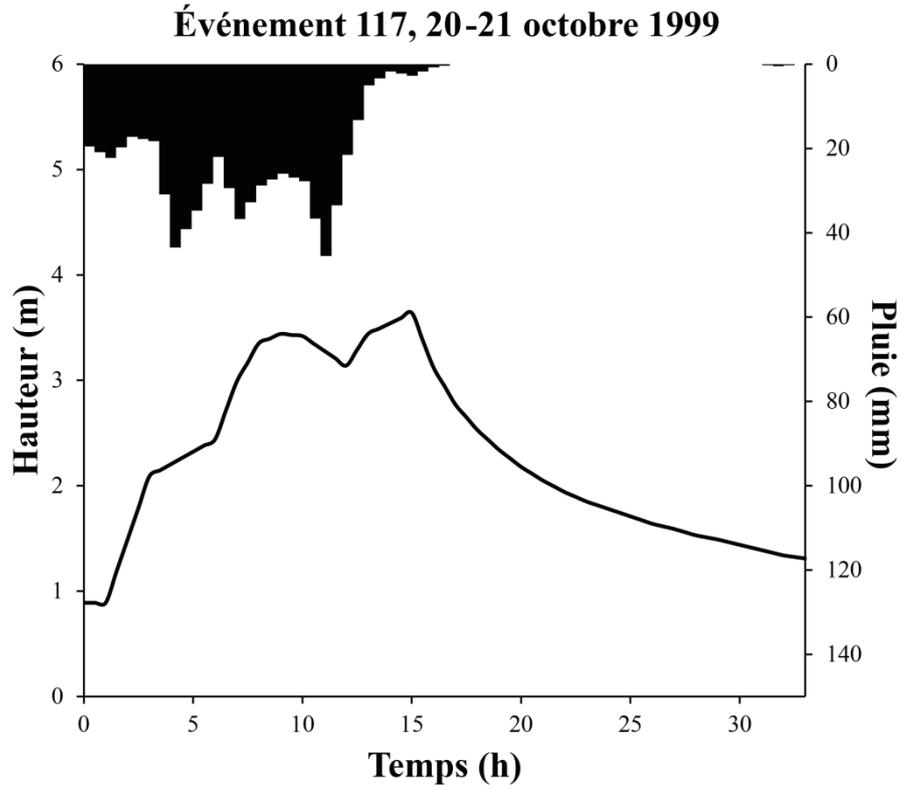
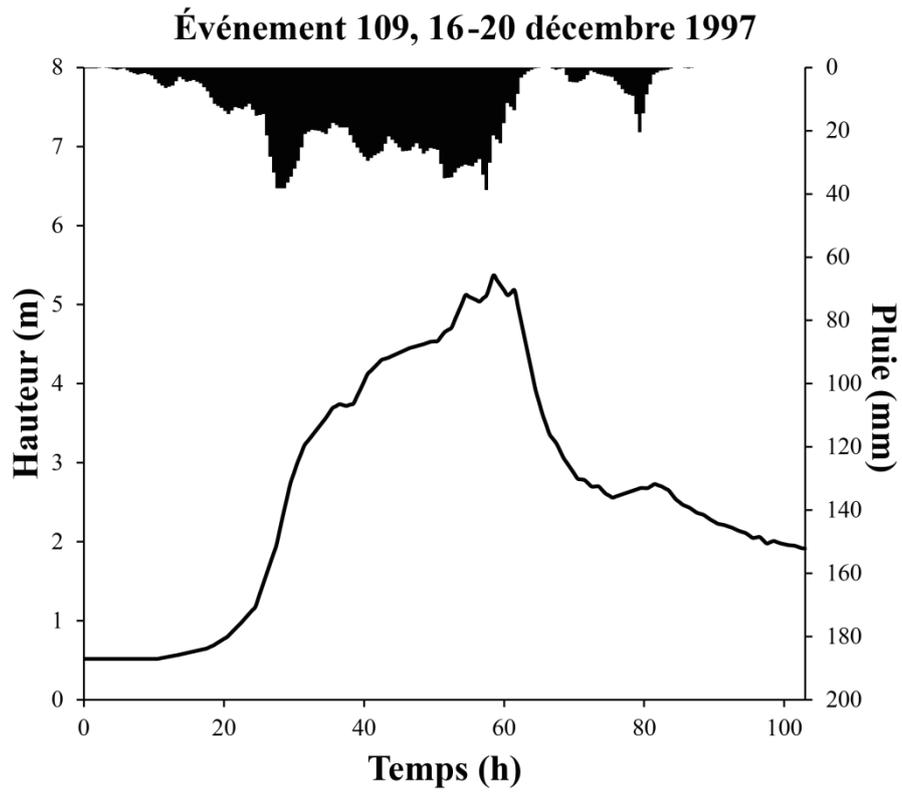
■ Cumul des pluies sur les 6 pluviomètres — Hauteur à Anduze



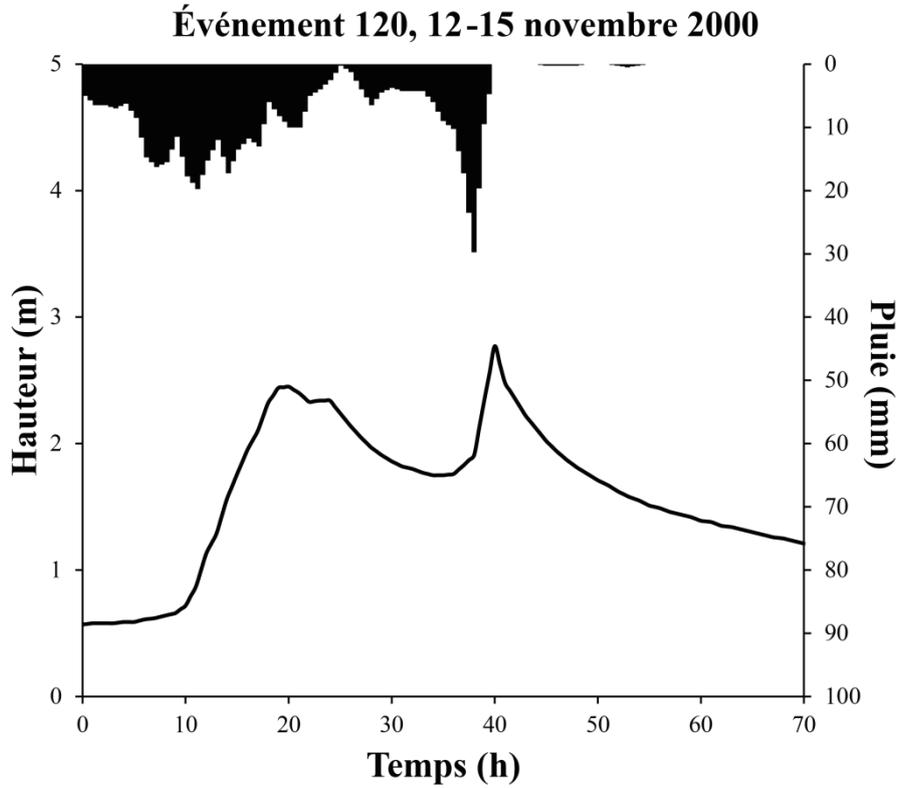
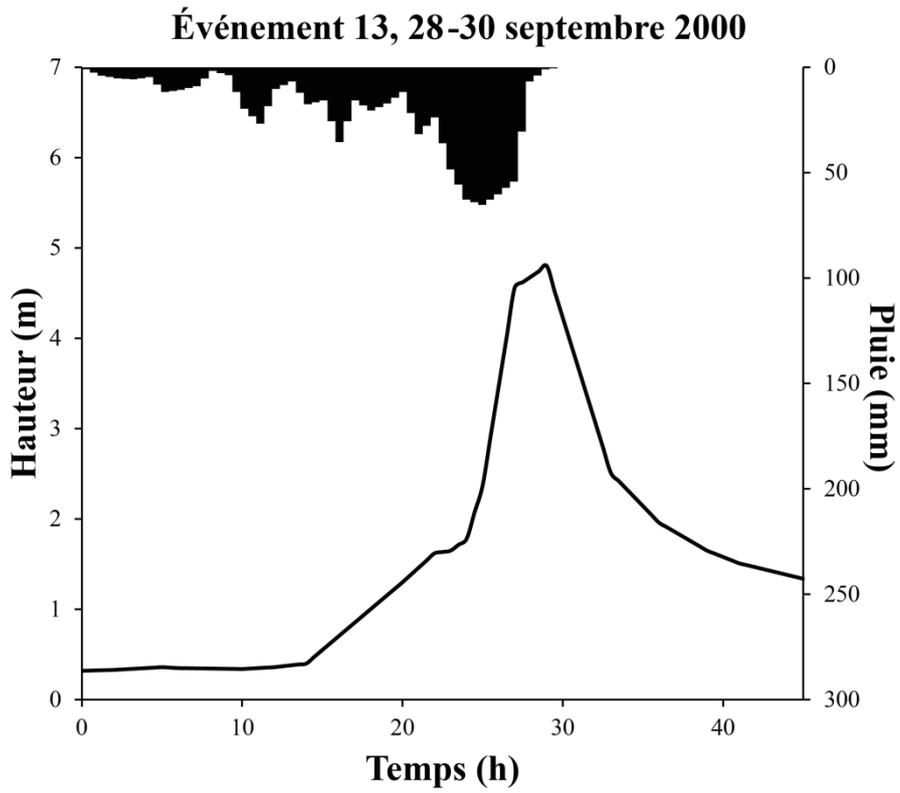
■ Cumul des pluies sur les 6 pluviomètres — Hauteur à Anduze



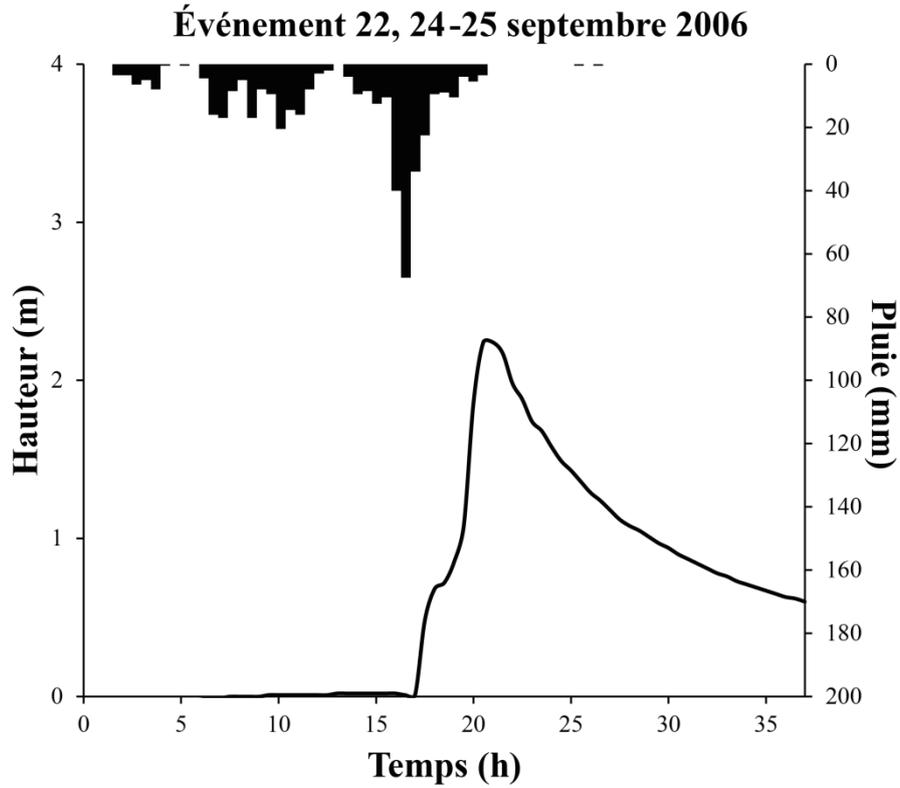
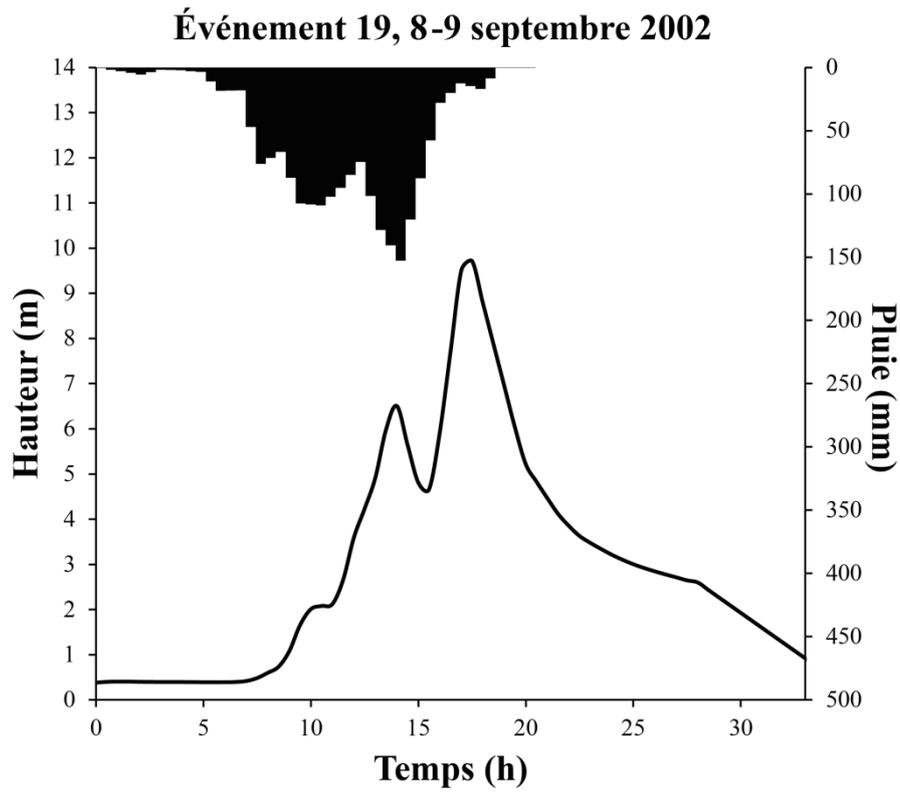
■ Cumul des pluies sur les 6 pluviomètres — Hauteur à Anduze



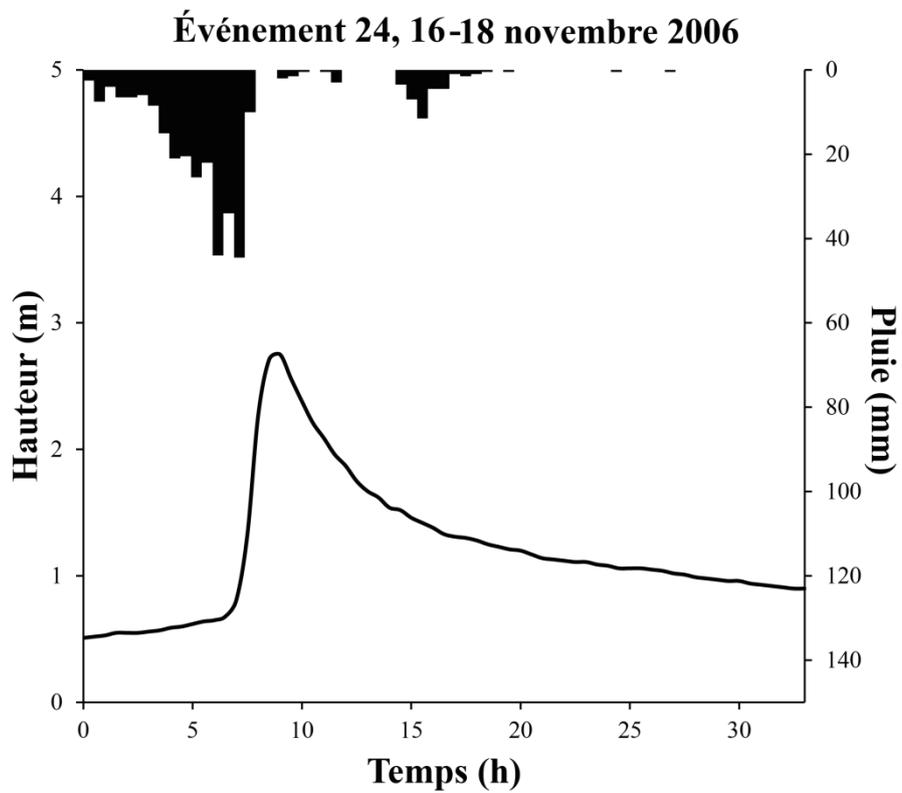
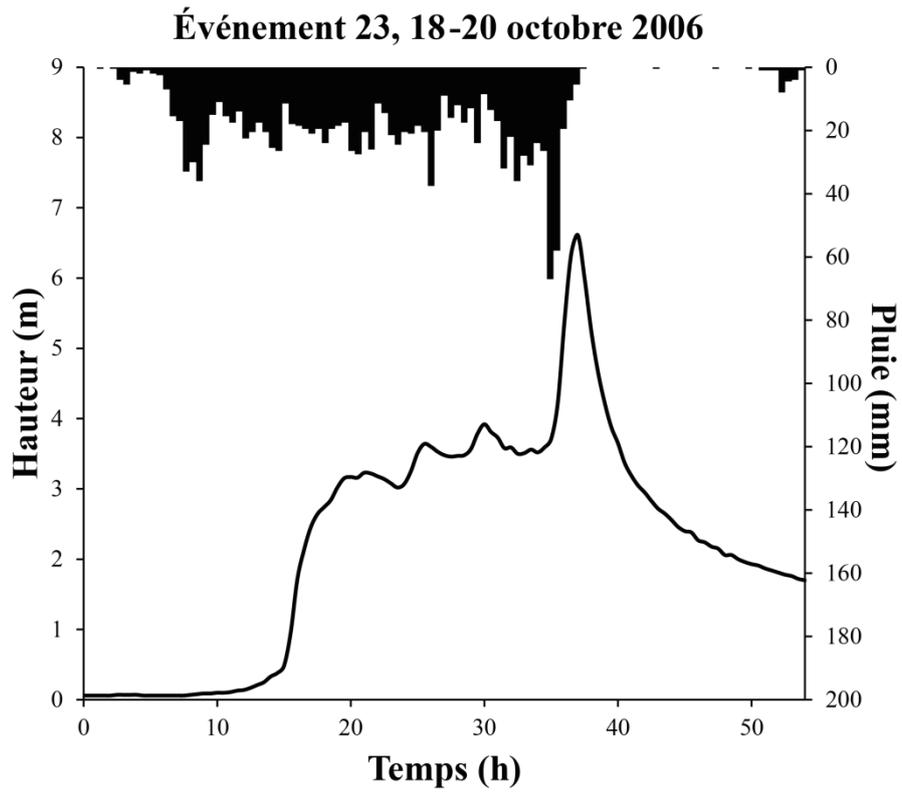
■ Cumul des pluies sur les 6 pluviomètres — Hauteur à Anduze



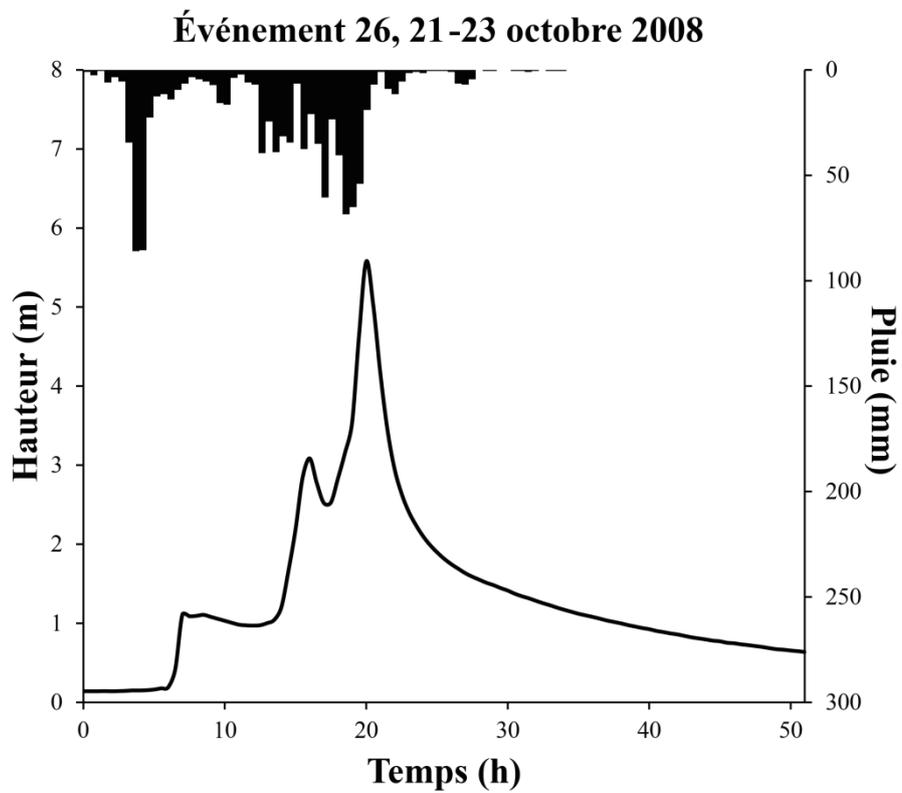
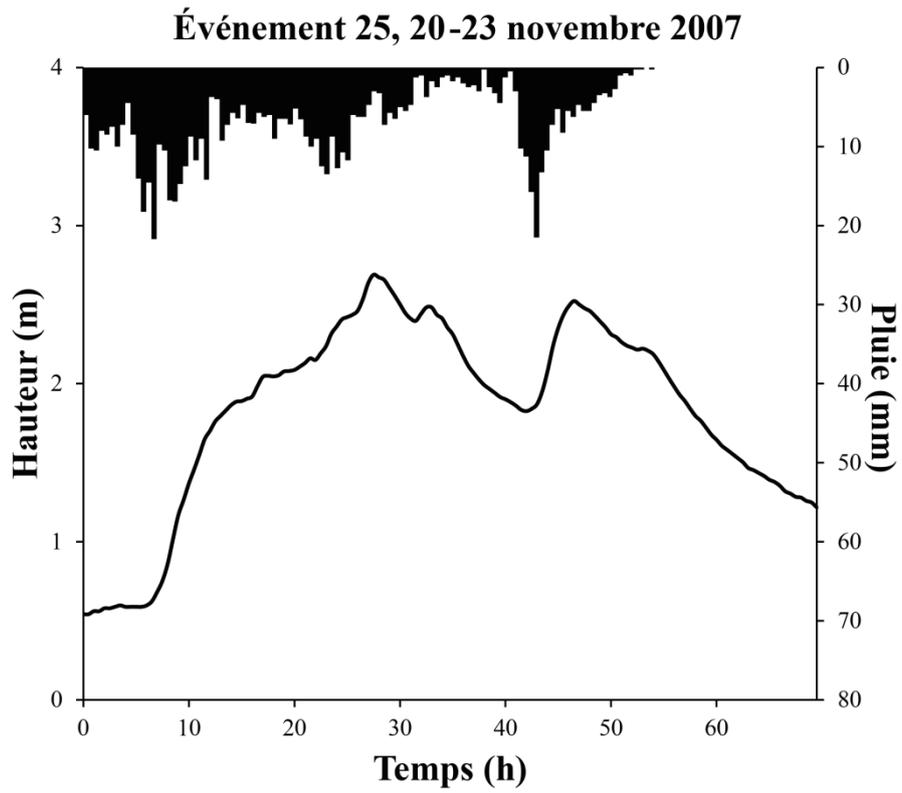
■ Cumul des pluies sur les 6 pluviomètres — Hauteur à Anduze



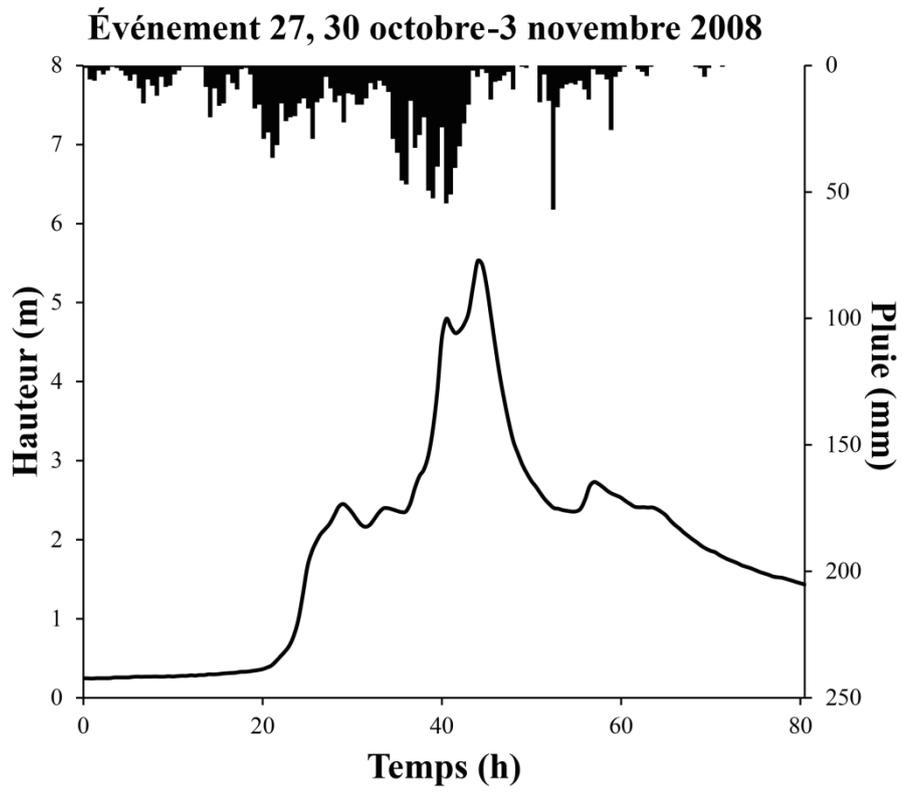
■ Cumul des pluies sur les 6 pluviomètres — Hauteur à Anduze



■ Cumul des pluies sur les 6 pluviomètres — Hauteur à Anduze



■ Cumul des pluies sur les 6 pluviomètres — Hauteur à Anduze



Annexe 3 : Tableaux de l'écart-type et du maximum des 200 scores de validation croisée partielle (chapitre3)

Pour le nombre de neurones cachés :

N_c	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
0	0,016	0,029	0,029	0,066	0,086	0,096
1	0,112	0,128	0,128	0,148	0,135	1,052
2	0,145	0,040	0,040	0,072	0,087	0,124
3	0,037	0,046	0,046	0,075	0,074	0,084
4	0,047	0,042	0,042	0,069	0,087	0,259
5	0,034	0,041	0,041	0,087	0,145	0,098
6	0,040	0,050	0,050	0,081	0,115	0,115
7	0,084	0,109	0,109	0,121	0,114	0,091

Maximum

N_c	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
0	0,0101	0,0129	0,0129	0,0146	0,0110	0,0703
1	0,0101	0,0026	0,0026	0,0034	0,0033	0,0053
2	0,0028	0,0034	0,0034	0,0040	0,0031	0,0036
3	0,0035	0,0030	0,0030	0,0038	0,0038	0,0149
4	0,0029	0,0028	0,0028	0,0055	0,0075	0,0052
5	0,0041	0,0037	0,0037	0,0051	0,0064	0,0058
6	0,0069	0,0072	0,0072	0,0069	0,0062	0,0052
7	0,0101	0,0129	0,0129	0,0146	0,0110	0,0703

Écart-type

Pour la dimension de la fenêtre temporelle des pluies :

f_T	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
30min	0,268	0,148	0,597	0,223	0,075	0,169
1h	0,036	0,088	0,077	0,067	0,082	0,170
1h30min	0,057	0,171	0,059	0,283	0,094	0,371
2h	0,092	0,049	0,067	0,072	0,161	0,085
2h30min	0,125	0,073	0,147	0,072	0,102	0,208
3h	0,072	0,113	0,073	0,098	0,148	0,088
3h30min	0,036	0,045	0,059	0,214	0,082	0,140

4h	0,134	0,054	0,116	0,069	0,122	0,137
4h30mi,	0,151	0,175	0,132	0,109	0,093	0,115
5h	0,052	0,055	0,101	0,083	0,078	0,122
5h30min	0,075	0,082	0,142	0,112	0,348	0,097

Maximum

f_T	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
30min	0,0185	0,0089	0,0429	0,0124	0,0035	0,0109
1h	0,0023	0,0059	0,0042	0,0028	0,0039	0,0079
1h30min	0,0043	0,0114	0,0032	0,0169	0,0051	0,0220
2h	0,0061	0,0032	0,0030	0,0038	0,0101	0,0037
2h30min	0,0086	0,0045	0,0089	0,0041	0,0054	0,0102
3h	0,0047	0,0067	0,0041	0,0055	0,0093	0,0044
3h30min	0,0026	0,0028	0,0040	0,0149	0,0042	0,0077
4h	0,0104	0,0041	0,0066	0,0043	0,0060	0,0059
4h30mi,	0,0101	0,0120	0,0090	0,0065	0,0046	0,0058
5h	0,0056	0,0043	0,0084	0,0045	0,0040	0,0062
5h30min	0,0064	0,0056	0,0111	0,0059	0,0207	0,0051

Écart-type

Pour les hyperparamètres de l'algorithme d'apprentissage :

$\mu_{ini}-\mu_{dec}$	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
0,0001-0,1	0,333	0,175	0,113	0,169	0,231	0,089
0,0001-0,2	0,070	0,151	0,113	0,168	0,085	0,088
0,0001-0,3	0,046	0,169	0,113	0,205	0,138	0,150
0,0001-0,4	0,025	0,162	0,114	0,172	0,111	0,143
0,0001-0,5	0,070	0,147	0,114	0,169	0,114	0,112
0,0001-0,6	0,070	0,148	0,115	0,172	0,083	0,106
0,0001-0,7	0,089	0,142	0,118	0,172	0,087	0,086
0,0001-0,8	0,070	0,140	0,122	0,174	0,086	0,091
0,0001-0,9	0,070	0,039	0,150	0,169	0,087	0,084
0,001-0,1	0,053	0,175	0,058	0,161	0,297	0,090
0,001-0,2	0,051	0,151	0,058	0,105	0,097	0,088
0,001-0,3	0,057	0,169	0,067	0,236	0,083	0,150
0,001-0,4	0,058	0,162	0,067	0,111	0,111	0,143

0,001-0,5	0,051	0,147	0,089	0,119	0,114	0,112
0,001-0,6	0,026	0,148	0,085	0,161	0,083	0,096
0,001-0,7	0,089	0,142	0,106	0,102	0,077	0,086
0,001-0,8	0,056	0,140	0,135	0,089	0,079	0,091
0,001-0,9	0,024	0,041	0,112	0,077	0,078	0,084
0,01-0,1	0,053	0,092	0,058	0,083	0,231	0,089
0,01-0,2	0,057	0,175	0,070	0,133	0,085	0,088
0,01-0,3	0,085	0,075	0,055	0,205	0,083	0,096
0,01-0,4	0,027	0,062	0,067	0,106	0,111	0,090
0,01-0,5	0,051	0,086	0,059	0,157	0,114	0,082
0,01-0,6	0,040	0,063	0,055	0,161	0,083	0,086
0,01-0,7	0,041	0,055	0,166	0,102	0,078	0,089
0,01-0,8	0,024	0,043	0,096	0,086	0,077	0,091
0,01-0,9	0,028	0,050	0,083	0,069	0,077	0,082
0,1-0,1	0,038	0,097	0,092	0,369	0,079	0,083
0,1-0,2	0,051	0,175	0,098	0,251	0,081	0,088
0,1-0,3	0,061	0,075	0,099	0,205	0,137	0,087
0,1-0,4	0,028	0,062	0,082	0,106	0,117	0,086
0,1-0,5	0,051	0,086	0,075	0,164	0,114	0,082
0,1-0,6	0,050	0,063	0,068	0,161	0,094	0,086
0,1-0,7	0,047	0,055	0,075	0,102	0,081	0,102
0,1-0,8	0,029	0,070	0,061	0,068	0,080	0,108
0,1-0,9	0,029	0,041	0,070	0,065	0,077	0,085
1-0,1	0,049	0,195	0,088	0,076	0,079	0,087
1-0,2	0,150	0,175	0,060	0,251	0,077	0,087
1-0,3	0,074	0,075	0,254	0,205	0,082	0,088
1-0,4	0,039	0,062	0,067	0,150	0,084	0,087
1-0,5	0,040	0,086	0,214	0,272	0,086	0,088
1-0,6	0,039	0,063	0,065	0,205	0,074	0,095
1-0,7	0,052	0,055	0,064	0,191	0,078	0,095
1-0,8	0,036	0,043	0,073	0,144	0,084	0,089
1-0,9	0,032	0,043	0,057	0,065	0,091	0,089
10-0,1	0,025	0,137	0,058	0,244	0,079	0,092

10-0,2	0,032	0,175	0,058	0,212	0,093	0,085
10-0,3	0,032	0,075	0,054	0,239	0,088	0,090
10-0,4	0,036	0,080	0,067	0,205	0,088	0,088
10-0,5	0,049	0,086	0,055	0,208	0,080	0,086
10-0,6	0,035	0,066	0,060	0,151	0,081	0,087
10-0,7	0,037	0,075	0,061	0,185	0,080	0,089
10-0,8	0,024	0,060	0,057	0,071	0,086	0,088
10-0,9	0,024	0,040	0,058	0,063	0,083	0,083

Maximum

$\mu_{ini}-\mu_{dec}$	$h_p=30min$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
0,0001-0,1	0,0229	0,0124	0,0058	0,0121	0,0149	0,0038
0,0001-0,2	0,0048	0,0092	0,0060	0,0099	0,0042	0,0036
0,0001-0,3	0,0026	0,0109	0,0059	0,0153	0,0066	0,0066
0,0001-0,4	0,0015	0,0105	0,0062	0,0100	0,0054	0,0068
0,0001-0,5	0,0047	0,0099	0,0060	0,0102	0,0067	0,0045
0,0001-0,6	0,0040	0,0097	0,0060	0,0126	0,0040	0,0047
0,0001-0,7	0,0065	0,0088	0,0063	0,0099	0,0038	0,0035
0,0001-0,8	0,0049	0,0083	0,0064	0,0096	0,0042	0,0039
0,0001-0,9	0,0044	0,0028	0,0082	0,0089	0,0038	0,0036
0,001-0,1	0,0035	0,0134	0,0033	0,0094	0,0207	0,0039
0,001-0,2	0,0029	0,0094	0,0037	0,0059	0,0049	0,0036
0,001-0,3	0,0038	0,0128	0,0041	0,0197	0,0039	0,0066
0,001-0,4	0,0035	0,0106	0,0038	0,0081	0,0053	0,0062
0,001-0,5	0,0030	0,0105	0,0049	0,0073	0,0063	0,0043
0,001-0,6	0,0014	0,0099	0,0047	0,0101	0,0041	0,0041
0,001-0,7	0,0053	0,0089	0,0057	0,0061	0,0037	0,0035
0,001-0,8	0,0031	0,0084	0,0073	0,0054	0,0041	0,0038
0,001-0,9	0,0019	0,0030	0,0058	0,0041	0,0037	0,0035
0,01-0,1	0,0035	0,0081	0,0032	0,0050	0,0128	0,0036
0,01-0,2	0,0046	0,0108	0,0040	0,0076	0,0045	0,0036
0,01-0,3	0,0070	0,0051	0,0032	0,0129	0,0043	0,0037
0,01-0,4	0,0017	0,0045	0,0042	0,0056	0,0058	0,0037

0,01-0,5	0,0037	0,0064	0,0036	0,0097	0,0058	0,0031
0,01-0,6	0,0024	0,0046	0,0036	0,0119	0,0039	0,0035
0,01-0,7	0,0023	0,0038	0,0094	0,0053	0,0040	0,0037
0,01-0,8	0,0014	0,0030	0,0050	0,0047	0,0043	0,0040
0,01-0,9	0,0020	0,0030	0,0044	0,0033	0,0037	0,0032
0,1-0,1	0,0025	0,0093	0,0049	0,0227	0,0041	0,0031
0,1-0,2	0,0034	0,0113	0,0054	0,0149	0,0047	0,0037
0,1-0,3	0,0041	0,0049	0,0054	0,0127	0,0072	0,0034
0,1-0,4	0,0018	0,0046	0,0049	0,0050	0,0077	0,0036
0,1-0,5	0,0034	0,0059	0,0047	0,0095	0,0061	0,0034
0,1-0,6	0,0032	0,0045	0,0039	0,0094	0,0046	0,0034
0,1-0,7	0,0033	0,0035	0,0051	0,0047	0,0042	0,0040
0,1-0,8	0,0019	0,0041	0,0035	0,0031	0,0045	0,0047
0,1-0,9	0,0023	0,0029	0,0040	0,0027	0,0038	0,0032
1-0,1	0,0039	0,0140	0,0053	0,0038	0,0036	0,0033
1-0,2	0,0107	0,0112	0,0037	0,0152	0,0046	0,0038
1-0,3	0,0057	0,0058	0,0153	0,0127	0,0044	0,0036
1-0,4	0,0035	0,0043	0,0043	0,0105	0,0044	0,0040
1-0,5	0,0037	0,0059	0,0125	0,0181	0,0044	0,0038
1-0,6	0,0033	0,0046	0,0041	0,0143	0,0037	0,0044
1-0,7	0,0037	0,0036	0,0040	0,0108	0,0039	0,0042
1-0,8	0,0026	0,0032	0,0037	0,0072	0,0043	0,0042
1-0,9	0,0025	0,0032	0,0030	0,0029	0,0041	0,0036
10-0,1	0,0013	0,0101	0,0032	0,0179	0,0036	0,0037
10-0,2	0,0015	0,0123	0,0033	0,0133	0,0050	0,0040
10-0,3	0,0014	0,0055	0,0030	0,0147	0,0049	0,0039
10-0,4	0,0024	0,0062	0,0034	0,0122	0,0045	0,0038
10-0,5	0,0028	0,0065	0,0029	0,0145	0,0045	0,0036
10-0,6	0,0023	0,0049	0,0031	0,0088	0,0046	0,0040
10-0,7	0,0023	0,0045	0,0033	0,0103	0,0045	0,0039
10-0,8	0,0011	0,0034	0,0031	0,0028	0,0046	0,0039
10-0,9	0,0017	0,0026	0,0029	0,0027	0,0034	0,0030

Écart-type

Annexe 4 : Tableaux de critères pour les sorties des 20 modèles à perceptron multicouche (chapitre 3)

Pour l'événement 19 :

C_p	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,54	0,30	0,68	0,88	0,80	0,71
Modèle 2	0,52	0,22	0,55	0,84	0,75	0,71
Modèle 3	0,33	0,17	0,50	0,81	0,75	0,67
Modèle 4	0,23	0,16	0,48	0,77	0,74	0,65
Modèle 5	0,21	0,14	0,47	0,73	0,73	0,64
Modèle 6	0,19	0,13	0,47	0,72	0,73	0,63
Modèle 7	0,13	0,12	0,47	0,71	0,73	0,61
Modèle 8	0,05	0,10	0,46	0,69	0,69	0,61
Modèle 9	0,05	0,04	0,46	0,55	0,62	0,60
Modèle 10	0,01	0,02	0,45	0,55	0,61	0,57
Modèle 11	-0,30	0,01	0,45	0,49	0,61	0,57
Modèle 12	-1,02	-0,17	0,44	0,48	0,59	0,54
Modèle 13	-1,25	-0,35	0,41	0,48	0,55	0,51
Modèle 14	-1,29	-0,38	0,39	0,42	0,52	0,48
Modèle 15	-1,58	-0,59	0,36	0,37	0,52	0,48
Modèle 16	-1,70	-1,01	0,19	0,18	0,45	0,41
Modèle 17	-1,72	-1,99	0,10	0,15	0,42	0,28
Modèle 18	-2,84	-8,22	0,10	0,07	0,29	0,25
Modèle 19	-4,23	-12,58	0,10	-0,63	0,09	-0,04
Modèle 20	-37,51	-49,58	0,10	-0,67	-1,34	-11,12

Critère de Persistance

C_N	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,98	0,89	0,87	0,92	0,82	0,66
Modèle 2	0,98	0,88	0,82	0,90	0,78	0,66
Modèle 3	0,97	0,87	0,79	0,88	0,78	0,61
Modèle 4	0,97	0,87	0,79	0,86	0,77	0,58
Modèle 5	0,97	0,87	0,78	0,83	0,77	0,58
Modèle 6	0,96	0,87	0,78	0,82	0,77	0,57
Modèle 7	0,96	0,86	0,78	0,82	0,76	0,55
Modèle 8	0,96	0,86	0,78	0,81	0,73	0,54

Modèle 9	0,96	0,85	0,78	0,73	0,67	0,53
Modèle 10	0,96	0,85	0,78	0,72	0,67	0,50
Modèle 11	0,94	0,85	0,78	0,69	0,66	0,50
Modèle 12	0,91	0,82	0,77	0,69	0,65	0,46
Modèle 13	0,90	0,79	0,76	0,68	0,62	0,43
Modèle 14	0,90	0,79	0,75	0,65	0,59	0,40
Modèle 15	0,89	0,75	0,74	0,62	0,61	0,41
Modèle 16	0,88	0,69	0,67	0,51	0,54	0,32
Modèle 17	0,88	0,54	0,63	0,48	0,51	0,18
Modèle 18	0,83	-0,42	0,63	0,44	0,41	0,15
Modèle 19	0,77	-1,09	0,63	0,02	0,25	-0,18
Modèle 20	-0,71	-6,79	0,63	0,00	-0,92	-12,49

Critère de Nash

RMSE	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,0319	0,0736	0,0806	0,0641	0,0936	0,1304
Modèle 2	0,0327	0,0773	0,0958	0,0718	0,1049	0,1306
Modèle 3	0,0384	0,0798	0,1011	0,0783	0,1049	0,1387
Modèle 4	0,0413	0,0803	0,1030	0,0847	0,1073	0,1443
Modèle 5	0,0417	0,0815	0,1036	0,0914	0,1078	0,1450
Modèle 6	0,0423	0,0815	0,1043	0,0941	0,1079	0,1460
Modèle 7	0,0439	0,0822	0,1045	0,0952	0,1092	0,1502
Modèle 8	0,0458	0,0830	0,1046	0,0975	0,1159	0,1519
Modèle 9	0,0459	0,0857	0,1047	0,1167	0,1273	0,1530
Modèle 10	0,0467	0,0869	0,1053	0,1181	0,1278	0,1575
Modèle 11	0,0536	0,0873	0,1056	0,1249	0,1295	0,1581
Modèle 12	0,0668	0,0947	0,1064	0,1250	0,1316	0,1633
Modèle 13	0,0705	0,1018	0,1092	0,1262	0,1376	0,1685
Modèle 14	0,0712	0,1031	0,1110	0,1322	0,1400	0,1717
Modèle 15	0,0755	0,1105	0,1144	0,1381	0,1422	0,1725
Modèle 16	0,0772	0,1241	0,1284	0,1567	0,1510	0,1840
Modèle 17	0,0775	0,1516	0,1351	0,1602	0,1556	0,2025
Modèle 18	0,0921	0,2659	0,1352	0,1671	0,1711	0,2061
Modèle 19	0,1075	0,3227	0,1352	0,2211	0,1939	0,2421
Modèle 20	0,2917	0,6227	0,1353	0,2234	0,3092	0,8196

RMSE

PWRMSE	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,0410	0,0965	0,1077	0,0774	0,1176	0,1686
Modèle 2	0,0434	0,1011	0,1321	0,0914	0,1289	0,1752
Modèle 3	0,0510	0,1069	0,1405	0,1033	0,1360	0,1719
Modèle 4	0,0567	0,1092	0,1378	0,1088	0,1414	0,1898
Modèle 5	0,0575	0,1100	0,1442	0,1230	0,1412	0,1928
Modèle 6	0,0569	0,1068	0,1450	0,1261	0,1386	0,1909
Modèle 7	0,0603	0,1128	0,1450	0,1271	0,1382	0,1959
Modèle 8	0,0633	0,1143	0,1447	0,1310	0,1553	0,1996
Modèle 9	0,0636	0,1142	0,1454	0,1540	0,1654	0,2035
Modèle 10	0,0644	0,1174	0,1405	0,1564	0,1669	0,2103
Modèle 11	0,0747	0,1187	0,1469	0,1724	0,1678	0,2108
Modèle 12	0,0946	0,1321	0,1479	0,1725	0,1735	0,2160
Modèle 13	0,0989	0,1401	0,1529	0,1714	0,1738	0,2183
Modèle 14	0,1003	0,1433	0,1533	0,1789	0,1814	0,2300
Modèle 15	0,1074	0,1433	0,1574	0,1862	0,1828	0,2272
Modèle 16	0,1076	0,1605	0,1805	0,2094	0,2023	0,2456
Modèle 17	0,1065	0,1913	0,1884	0,2162	0,1829	0,2609
Modèle 18	0,1298	0,3391	0,1885	0,2349	0,2300	0,2699
Modèle 19	0,1520	0,4281	0,1885	0,3085	0,2640	0,3091
Modèle 20	0,3832	0,5815	0,1886	0,3115	0,3975	0,8913

PWRMSE

Pour l'événement 26 :

C_P	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,38	0,33	0,26	0,47	0,51	0,41
Modèle 2	0,38	0,33	0,20	0,41	0,50	0,41
Modèle 3	0,35	0,30	0,20	0,40	0,46	0,41
Modèle 4	0,34	0,24	0,20	0,39	0,43	0,39
Modèle 5	0,33	0,23	0,20	0,37	0,43	0,38
Modèle 6	0,33	0,17	0,20	0,33	0,41	0,38
Modèle 7	0,30	0,11	0,20	0,33	0,41	0,36
Modèle 8	0,27	0,10	0,19	0,33	0,41	0,35
Modèle 9	0,24	0,09	0,16	0,31	0,41	0,35
Modèle 10	0,23	0,09	0,13	0,30	0,40	0,35
Modèle 11	0,20	0,08	0,12	0,27	0,39	0,34
Modèle 12	0,13	0,00	0,11	0,26	0,39	0,33
Modèle 13	0,08	-0,06	0,11	0,24	0,37	0,32
Modèle 14	0,07	-0,12	0,11	0,24	0,36	0,31
Modèle 15	0,06	-0,17	0,07	0,23	0,34	0,30
Modèle 16	-0,26	-0,32	0,07	0,23	0,31	0,28
Modèle 17	-0,49	-0,65	0,06	0,18	0,31	0,24
Modèle 18	-0,71	-0,72	0,06	0,11	0,31	0,22
Modèle 19	-0,79	-3,17	0,04	0,01	0,30	0,16
Modèle 20	-1,06	-5,65	-0,03	-0,09	0,29	0,01

Critère de Persistance

C_N	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,97	0,88	0,66	0,65	0,60	0,40
Modèle 2	0,97	0,88	0,64	0,61	0,60	0,39
Modèle 3	0,96	0,87	0,64	0,60	0,57	0,39
Modèle 4	0,96	0,86	0,64	0,60	0,54	0,37
Modèle 5	0,96	0,86	0,64	0,58	0,54	0,36
Modèle 6	0,96	0,85	0,64	0,56	0,53	0,36
Modèle 7	0,96	0,84	0,64	0,56	0,53	0,34
Modèle 8	0,96	0,84	0,63	0,56	0,53	0,33
Modèle 9	0,96	0,84	0,62	0,55	0,52	0,33
Modèle 10	0,96	0,84	0,61	0,54	0,51	0,33

Modèle 11	0,96	0,83	0,60	0,52	0,51	0,32
Modèle 12	0,95	0,82	0,60	0,51	0,51	0,32
Modèle 13	0,95	0,81	0,60	0,50	0,49	0,30
Modèle 14	0,95	0,80	0,60	0,50	0,48	0,29
Modèle 15	0,95	0,79	0,58	0,49	0,47	0,28
Modèle 16	0,93	0,76	0,58	0,49	0,45	0,26
Modèle 17	0,92	0,70	0,58	0,46	0,44	0,22
Modèle 18	0,91	0,69	0,57	0,41	0,44	0,20
Modèle 19	0,90	0,25	0,57	0,35	0,44	0,14
Modèle 20	0,89	-0,20	0,53	0,28	0,43	-0,02

Critère de Nash

RMSE	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,0179	0,0337	0,0562	0,0576	0,0610	0,0753
Modèle 2	0,0179	0,0338	0,0585	0,0606	0,0616	0,0756
Modèle 3	0,0183	0,0345	0,0586	0,0612	0,0636	0,0759
Modèle 4	0,0184	0,0358	0,0586	0,0614	0,0659	0,0770
Modèle 5	0,0186	0,0362	0,0586	0,0626	0,0659	0,0774
Modèle 6	0,0186	0,0376	0,0586	0,0642	0,0667	0,0774
Modèle 7	0,0190	0,0388	0,0586	0,0645	0,0668	0,0787
Modèle 8	0,0194	0,0390	0,0588	0,0647	0,0669	0,0792
Modèle 9	0,0199	0,0391	0,0598	0,0651	0,0671	0,0792
Modèle 10	0,0199	0,0392	0,0608	0,0660	0,0678	0,0796
Modèle 11	0,0203	0,0396	0,0614	0,0671	0,0680	0,0799
Modèle 12	0,0212	0,0410	0,0615	0,0677	0,0682	0,0803
Modèle 13	0,0218	0,0423	0,0616	0,0685	0,0695	0,0815
Modèle 14	0,0219	0,0435	0,0617	0,0686	0,0698	0,0816
Modèle 15	0,0220	0,0445	0,0629	0,0689	0,0710	0,0823
Modèle 16	0,0255	0,0473	0,0632	0,0692	0,0722	0,0836
Modèle 17	0,0278	0,0528	0,0632	0,0713	0,0724	0,0856
Modèle 18	0,0297	0,0540	0,0635	0,0742	0,0726	0,0870
Modèle 19	0,0304	0,0840	0,0639	0,0781	0,0728	0,0900
Modèle 20	0,0326	0,1060	0,0664	0,0821	0,0734	0,0979

RMSE

PWRMSE	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,0195	0,0367	0,0585	0,0667	0,0747	0,0952
Modèle 2	0,0214	0,0354	0,0619	0,0676	0,0748	0,1008
Modèle 3	0,0207	0,0392	0,0619	0,0688	0,0773	0,0983
Modèle 4	0,0211	0,0400	0,0619	0,0732	0,0788	0,0987
Modèle 5	0,0210	0,0406	0,0619	0,0759	0,0787	0,0981
Modèle 6	0,0206	0,0388	0,0619	0,0730	0,0809	0,0986
Modèle 7	0,0212	0,0422	0,0619	0,0703	0,0799	0,1005
Modèle 8	0,0229	0,0424	0,0704	0,0726	0,0788	0,1009
Modèle 9	0,0203	0,0414	0,0640	0,0738	0,0785	0,0998
Modèle 10	0,0237	0,0398	0,0617	0,0751	0,0823	0,0997
Modèle 11	0,0227	0,0398	0,0661	0,0708	0,0811	0,0986
Modèle 12	0,0250	0,0438	0,0626	0,0756	0,0805	0,0999
Modèle 13	0,0247	0,0452	0,0628	0,0752	0,0850	0,1021
Modèle 14	0,0254	0,0476	0,0630	0,0762	0,0823	0,1027
Modèle 15	0,0223	0,0473	0,0671	0,0780	0,0824	0,1039
Modèle 16	0,0240	0,0526	0,0674	0,0776	0,0850	0,1067
Modèle 17	0,0287	0,0497	0,0676	0,0801	0,0863	0,1001
Modèle 18	0,0314	0,0493	0,0739	0,0809	0,0841	0,1092
Modèle 19	0,0322	0,0843	0,0651	0,0842	0,0857	0,1156
Modèle 20	0,0320	0,0932	0,0711	0,0853	0,0828	0,1087

PWRMSE

Pour l'événement 27 :

C_P	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,55	0,56	0,75	0,52	0,57	0,58
Modèle 2	0,52	0,50	0,52	0,49	0,57	0,56
Modèle 3	0,51	0,48	0,49	0,48	0,53	0,55
Modèle 4	0,51	0,48	0,48	0,48	0,53	0,55
Modèle 5	0,49	0,47	0,48	0,47	0,53	0,53
Modèle 6	0,49	0,44	0,48	0,47	0,53	0,52
Modèle 7	0,49	0,38	0,47	0,46	0,51	0,52
Modèle 8	0,48	0,32	0,46	0,45	0,51	0,52
Modèle 9	0,46	0,27	0,45	0,45	0,50	0,51
Modèle 10	0,44	0,24	0,44	0,44	0,49	0,51
Modèle 11	0,41	0,21	0,44	0,44	0,48	0,51
Modèle 12	0,30	0,19	0,43	0,44	0,47	0,48
Modèle 13	0,20	0,17	0,41	0,44	0,43	0,48
Modèle 14	0,08	0,17	0,41	0,42	0,42	0,48
Modèle 15	0,01	0,15	0,40	0,40	0,41	0,47
Modèle 16	-0,11	0,14	0,39	0,40	0,40	0,47
Modèle 17	-0,13	0,06	0,39	0,40	0,39	0,43
Modèle 18	-0,18	-0,03	0,35	0,39	0,35	0,38
Modèle 19	-0,21	-0,12	0,30	0,37	0,31	0,30
Modèle 20	-0,22	-0,18	0,24	0,35	0,21	0,28

Critère de Persistance

C_N	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	1,00	0,99	0,97	0,90	0,87	0,82
Modèle 2	1,00	0,98	0,95	0,90	0,87	0,81
Modèle 3	1,00	0,98	0,94	0,90	0,85	0,80
Modèle 4	1,00	0,98	0,94	0,89	0,85	0,80
Modèle 5	1,00	0,98	0,94	0,89	0,85	0,80
Modèle 6	1,00	0,98	0,94	0,89	0,85	0,79
Modèle 7	1,00	0,98	0,94	0,89	0,85	0,79
Modèle 8	1,00	0,98	0,94	0,89	0,85	0,79
Modèle 9	1,00	0,98	0,94	0,89	0,84	0,79
Modèle 10	0,99	0,97	0,94	0,89	0,84	0,79

Modèle 11	0,99	0,97	0,94	0,89	0,84	0,79
Modèle 12	0,99	0,97	0,94	0,89	0,83	0,77
Modèle 13	0,99	0,97	0,94	0,89	0,82	0,77
Modèle 14	0,99	0,97	0,93	0,88	0,82	0,77
Modèle 15	0,99	0,97	0,93	0,88	0,82	0,77
Modèle 16	0,99	0,97	0,93	0,88	0,81	0,77
Modèle 17	0,99	0,97	0,93	0,88	0,81	0,75
Modèle 18	0,99	0,97	0,93	0,88	0,80	0,73
Modèle 19	0,99	0,96	0,92	0,87	0,78	0,70
Modèle 20	0,99	0,96	0,92	0,87	0,75	0,69

Critère de Nash

RMSE	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,0078	0,0148	0,0281	0,0383	0,0447	0,0525
Modèle 2	0,0081	0,0159	0,0290	0,0392	0,0448	0,0538
Modèle 3	0,0082	0,0162	0,0290	0,0396	0,0468	0,0540
Modèle 4	0,0082	0,0162	0,0291	0,0398	0,0468	0,0545
Modèle 5	0,0083	0,0163	0,0292	0,0401	0,0469	0,0553
Modèle 6	0,0083	0,0168	0,0294	0,0401	0,0469	0,0560
Modèle 7	0,0084	0,0176	0,0297	0,0405	0,0478	0,0560
Modèle 8	0,0084	0,0184	0,0301	0,0407	0,0479	0,0560
Modèle 9	0,0085	0,0191	0,0301	0,0410	0,0485	0,0564
Modèle 10	0,0087	0,0195	0,0304	0,0411	0,0486	0,0566
Modèle 11	0,0089	0,0199	0,0306	0,0413	0,0492	0,0567
Modèle 12	0,0098	0,0202	0,0310	0,0413	0,0499	0,0581
Modèle 13	0,0104	0,0203	0,0312	0,0413	0,0515	0,0581
Modèle 14	0,0112	0,0204	0,0312	0,0419	0,0518	0,0582
Modèle 15	0,0116	0,0206	0,0315	0,0425	0,0523	0,0589
Modèle 16	0,0123	0,0207	0,0316	0,0427	0,0528	0,0592
Modèle 17	0,0124	0,0217	0,0326	0,0427	0,0534	0,0609
Modèle 18	0,0126	0,0227	0,0338	0,0432	0,0550	0,0634
Modèle 19	0,0128	0,0237	0,0338	0,0438	0,0569	0,0675
Modèle 20	0,0129	0,0243	0,0353	0,0445	0,0612	0,0686

RMSE

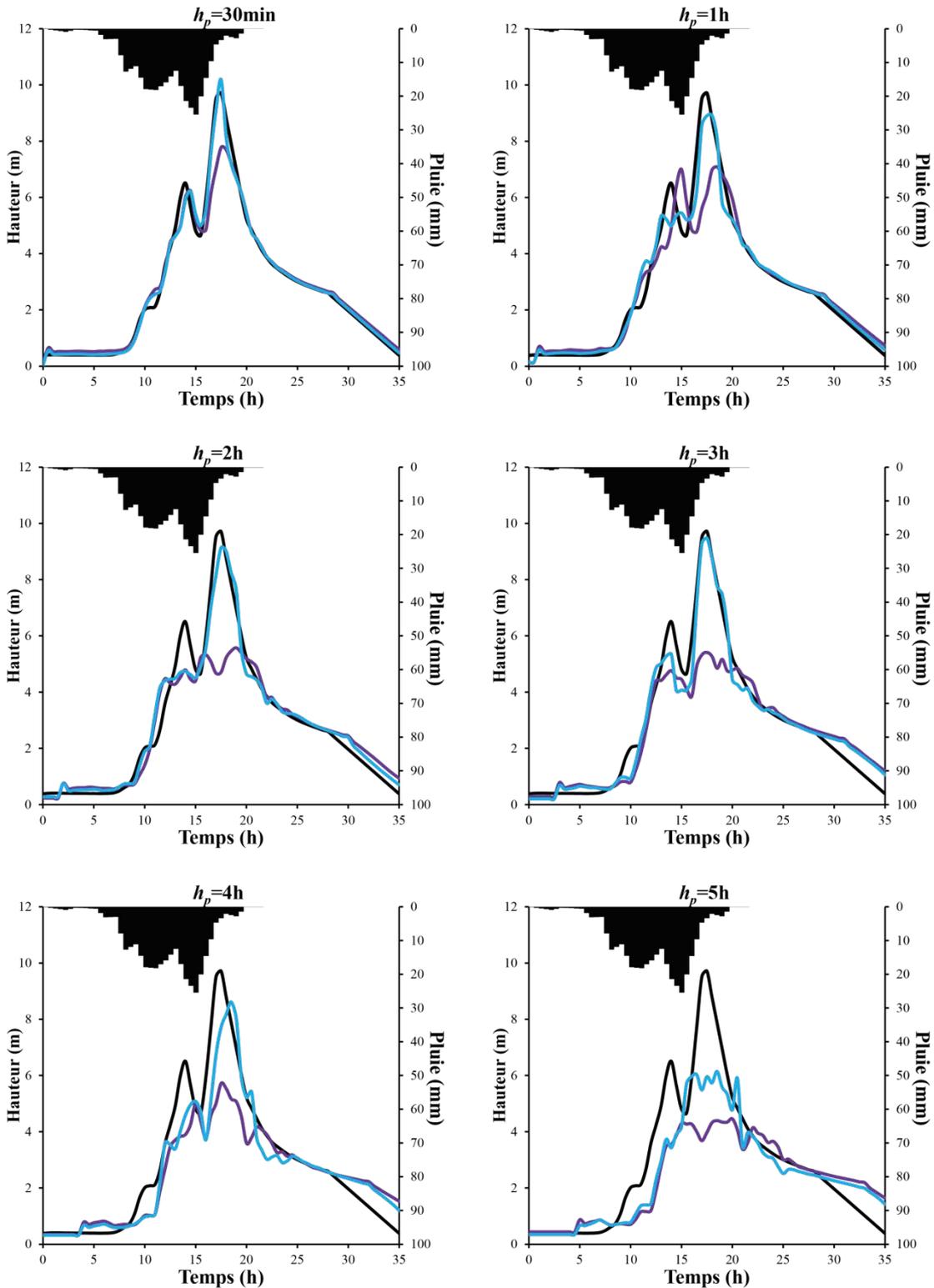
PWRMSE	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Modèle 1	0,0089	0,0171	0,0324	0,0442	0,0520	0,0623
Modèle 2	0,0095	0,0183	0,0332	0,0448	0,0516	0,0647
Modèle 3	0,0095	0,0188	0,0331	0,0455	0,0555	0,0627
Modèle 4	0,0091	0,0187	0,0331	0,0465	0,0559	0,0644
Modèle 5	0,0094	0,0193	0,0339	0,0465	0,0552	0,0660
Modèle 6	0,0092	0,0199	0,0333	0,0473	0,0549	0,0672
Modèle 7	0,0094	0,0208	0,0340	0,0463	0,0568	0,0663
Modèle 8	0,0096	0,0222	0,0345	0,0476	0,0563	0,0657
Modèle 9	0,0099	0,0223	0,0346	0,0470	0,0583	0,0670
Modèle 10	0,0098	0,0219	0,0355	0,0487	0,0575	0,0686
Modèle 11	0,0103	0,0227	0,0357	0,0483	0,0588	0,0686
Modèle 12	0,0117	0,0226	0,0360	0,0476	0,0599	0,0694
Modèle 13	0,0126	0,0242	0,0369	0,0482	0,0622	0,0713
Modèle 14	0,0130	0,0250	0,0368	0,0487	0,0633	0,0704
Modèle 15	0,0136	0,0256	0,0372	0,0488	0,0629	0,0699
Modèle 16	0,0138	0,0244	0,0363	0,0498	0,0633	0,0722
Modèle 17	0,0153	0,0257	0,0384	0,0492	0,0638	0,0754
Modèle 18	0,0154	0,0265	0,0401	0,0492	0,0659	0,0777
Modèle 19	0,0153	0,0280	0,0401	0,0521	0,0671	0,0852
Modèle 20	0,0158	0,0287	0,0427	0,0529	0,0736	0,0866

PWRMSE

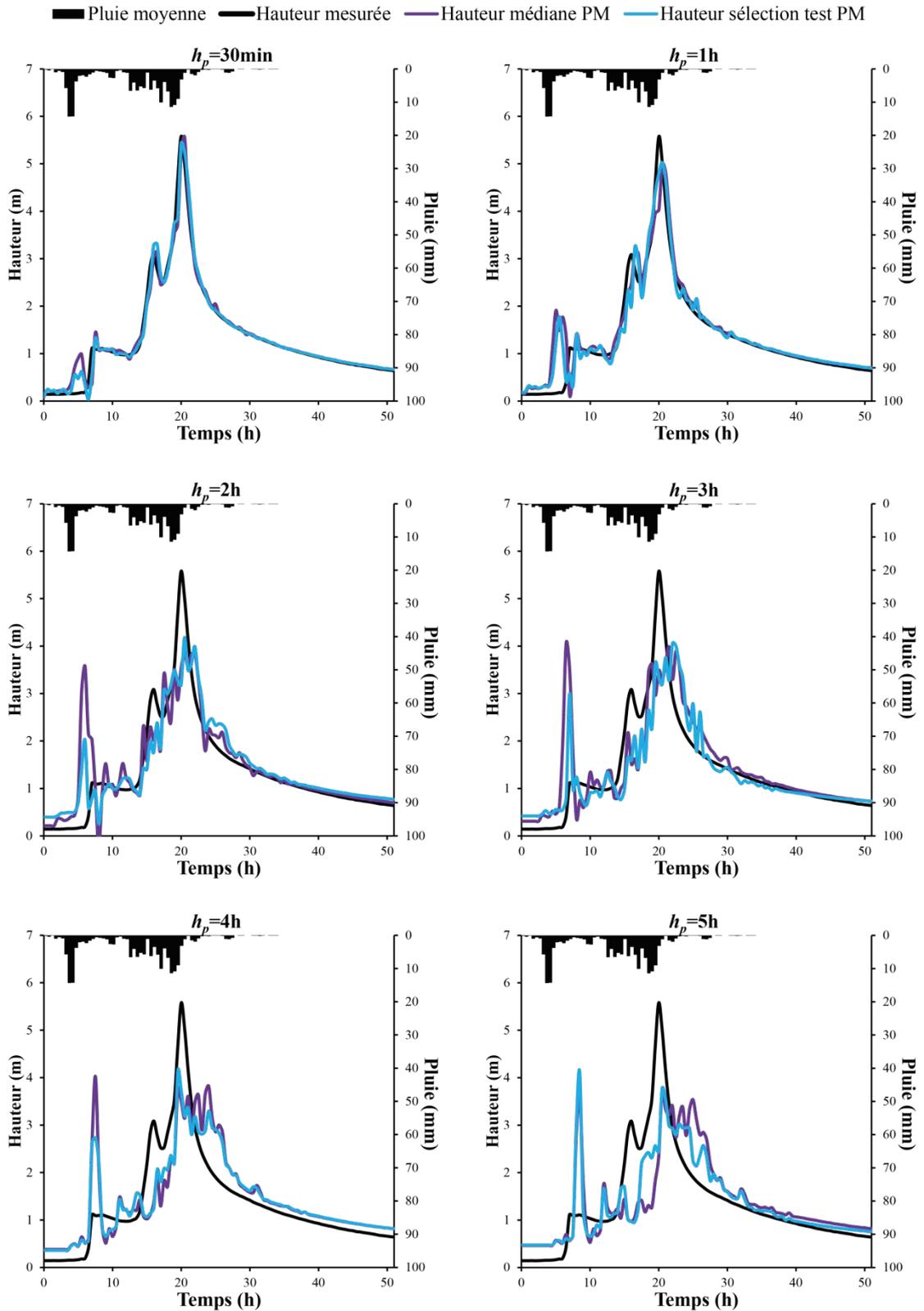
Annexe 5 : Comparaison entre les sorties des modèles à perceptron multicouche sélectionnés par validation croisée partielle (médiane des 20 modèles, chapitre 3) et ceux sélectionnés sur l'événement en test (meilleurs critères en test)

Pour l'événement 19 :

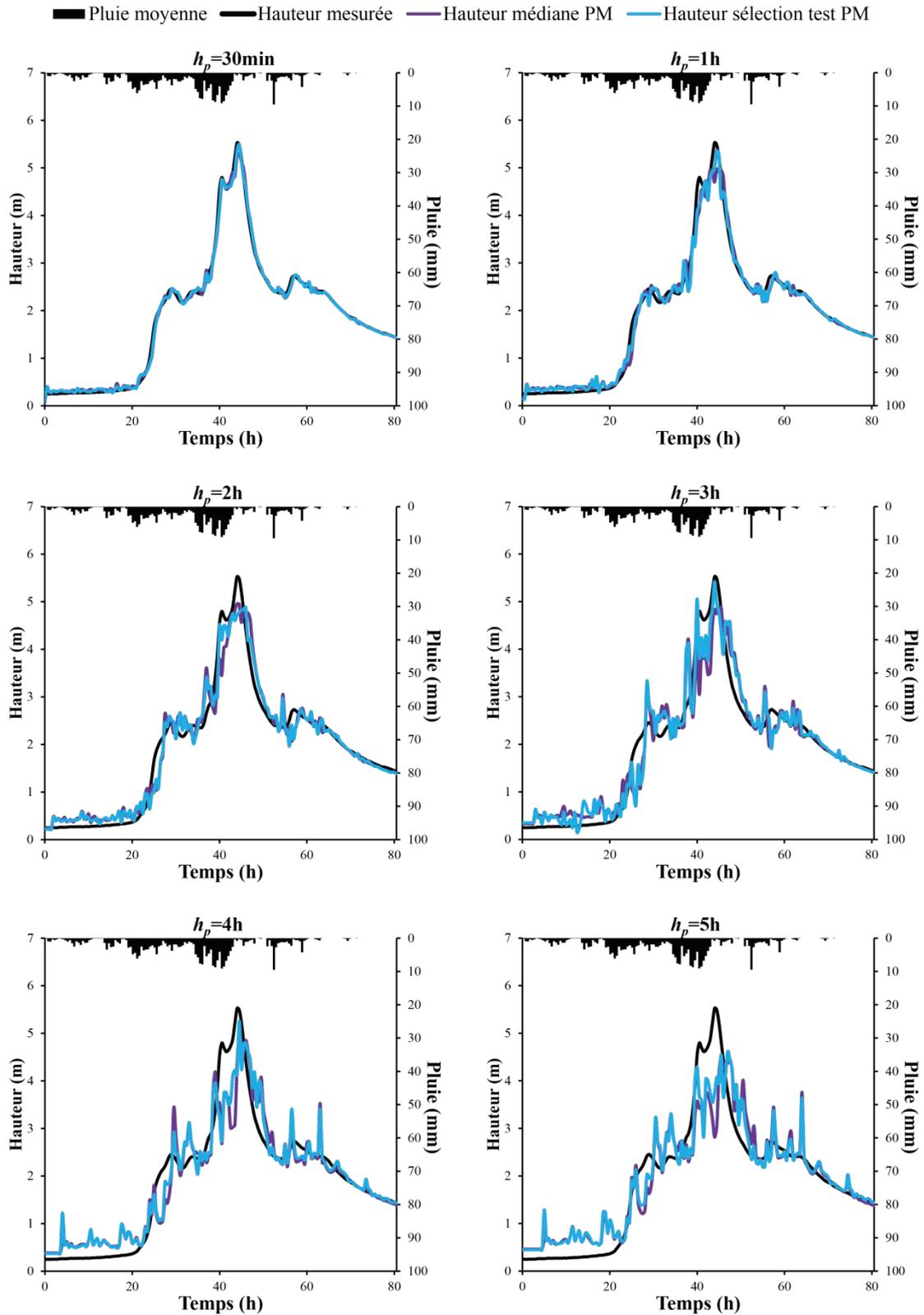
■ Pluie moyenne ■ Hauteur mesurée — Hauteur médiane PM — Hauteur sélection test PM



Pour l'événement 26 :



Pour l'événement 27 :



Annexe 6 : Tableaux des critères pour les différentes fenêtres d'adaptativité (chapitre 4)

Pour l'événement 19 :

C_P	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	0,34	0,02	0,46	0,66	0,61	-11,12
$F_a=30\text{min}$	0,31	0,04	0,25	0,64	0,59	-11,15
$F_a=1\text{h}$	0,49	0,12	0,11	0,44	-2,98	-11,61
$F_a=2\text{h}$	0,23	0,07	0,23	0,32	-3,37	-11,52
$F_a=3\text{h}$	0,14	0,21	0,33	0,24	-3,66	-11,33
$F_a=4\text{h}$	0,06	0,09	0,23	-0,11	-3,88	-11,47

Critère de Persistance

C_N	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	0,97	0,85	0,78	0,79	0,67	-12,49
$F_a=30\text{min}$	0,97	0,85	0,70	0,78	0,65	-12,52
$F_a=1\text{h}$	0,98	0,86	0,64	0,67	-2,29	-13,12
$F_a=2\text{h}$	0,97	0,86	0,69	0,59	-2,62	-13,04
$F_a=3\text{h}$	0,96	0,88	0,73	0,54	-2,89	-12,85
$F_a=4\text{h}$	0,96	0,86	0,69	0,25	-3,12	-13,01

Critère de Nash

PPD	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	90	76	58	65	67	356
$F_a=30\text{min}$	90	77	77	64	67	355
$F_a=1\text{h}$	93	82	99	91	171	368
$F_a=2\text{h}$	91	75	94	89	176	367
$F_a=3\text{h}$	91	73	70	81	174	357
$F_a=4\text{h}$	91	88	65	55	73	361

PPD

<i>SPPD</i>	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	90	60	57	63	57	42
$F_a=30\text{min}$	90	61	48	63	57	41
$F_a=1\text{h}$	93	69	54	77	62	61
$F_a=2\text{h}$	91	67	64	69	59	52
$F_a=3\text{h}$	91	73	60	63	59	52
$F_a=4\text{h}$	91	67	50	32	59	45

SPPD

Retard H_{pic} (h)	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	0	1	0,5	-0,5	0,5	5
$F_a=30\text{min}$	0	1,5	3	-0,5	0,5	5
$F_a=1\text{h}$	0	1	2	3,5	4,5	4,5
$F_a=2\text{h}$	0	2	3	4,5	7	5,5
$F_a=3\text{h}$	0	0,5	3,5	6	7,5	4,5
$F_a=4\text{h}$	0	1	1,5	8	7	5

Retard H_{pic}

Pour l'événement 26 :

C_p	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	-0,69	-0,65	0,20	0,17	0,43	0,26
$F_a=30\text{min}$	-2,13	-0,91	0,16	0,15	0,41	0,24
$F_a=1\text{h}$	-0,72	-1,61	-0,14	-0,09	0,06	-0,40
$F_a=2\text{h}$	-0,66	-2,26	-0,26	-0,20	0,00	-0,72
$F_a=3\text{h}$	-0,89	-2,13	-0,38	-0,20	-0,08	-0,88
$F_a=4\text{h}$	-1,25	-2,88	-0,48	-0,58	-0,18	-0,74

Critère de Persistance

C_N	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	0,91	0,70	0,64	0,46	0,53	0,24
$F_a=30\text{min}$	0,83	0,66	0,62	0,44	0,52	0,22
$F_a=1\text{h}$	0,91	0,53	0,48	0,28	0,24	-0,44
$F_a=2\text{h}$	0,91	0,41	0,43	0,21	0,20	-0,76
$F_a=3\text{h}$	0,90	0,44	0,37	0,21	0,13	-0,93
$F_a=4\text{h}$	0,88	0,30	0,33	-0,04	0,05	-0,79

Critère de Nash

<i>PPD</i>	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	102	92	72	79	68	74
$F_a=30\text{min}$	106	94	72	75	69	73
$F_a=1\text{h}$	104	91	88	94	95	103
$F_a=2\text{h}$	104	85	80	91	94	103
$F_a=3\text{h}$	104	85	44	92	84	100
$F_a=4\text{h}$	104	87	78	93	83	98

PPD

<i>SPPD</i>	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	94	68	62	59	61	41
$F_a=30\text{min}$	99	72	64	62	62	39
$F_a=1\text{h}$	97	63	67	80	79	54
$F_a=2\text{h}$	96	53	67	74	64	38
$F_a=3\text{h}$	96	60	67	77	53	37
$F_a=4\text{h}$	96	74	62	53	53	44

SPPD

Retard H_{pic} (h)	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	1	0,5	2,5	4	5	1
$F_a=30\text{min}$	1	0,5	1,5	4	5	1
$F_a=1\text{h}$	1	2	3	4,5	5,5	1
$F_a=2\text{h}$	1	3	2,5	5,5	6,5	1
$F_a=3\text{h}$	-0,5	3,5	2	7	7,5	-0,5
$F_a=4\text{h}$	-0,5	6	6	7,5	8,5	-0,5

Retard H_{pic}

Pour l'événement 27 :

C_p	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	0,09	0,21	0,30	0,40	0,41	0,52
$F_a=30\text{min}$	-5,02	-0,17	0,24	0,39	0,40	0,50
$F_a=1\text{h}$	0,14	0,15	0,12	-0,06	-0,04	0,22
$F_a=2\text{h}$	-0,05	-0,10	-0,02	-0,20	-0,21	0,08
$F_a=3\text{h}$	-0,07	-1,41	0,46	-0,36	-0,27	0,00
$F_a=4\text{h}$	-0,06	-0,15	-0,33	-0,31	-0,39	-0,02

Critère de Persistance

C_N	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	0,99	0,97	0,92	0,88	0,81	0,79
$F_a=30\text{min}$	0,94	0,96	0,92	0,88	0,81	0,78
$F_a=1\text{h}$	0,99	0,97	0,90	0,78	0,67	0,66
$F_a=2\text{h}$	0,99	0,96	0,89	0,76	0,62	0,60
$F_a=3\text{h}$	0,99	0,92	0,94	0,72	0,60	0,56
$F_a=4\text{h}$	0,99	0,96	0,85	0,73	0,57	0,55

Critère de Nash

PPD	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	98	93	92	96	105	77
$F_a=30\text{min}$	102	95	99	97	105	80
$F_a=1\text{h}$	100	98	99	111	113	106
$F_a=2\text{h}$	100	101	100	112	116	104
$F_a=3\text{h}$	99	101	100	108	95	104
$F_a=4\text{h}$	103	103	98	109	95	93

PPD

$SPPD$	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	98	83	92	95	77	54
$F_a=30\text{min}$	102	84	92	95	77	57
$F_a=1\text{h}$	100	86	99	109	67	62
$F_a=2\text{h}$	100	90	98	99	66	56
$F_a=3\text{h}$	99	100	91	93	78	47
$F_a=4\text{h}$	103	90	91	88	72	47

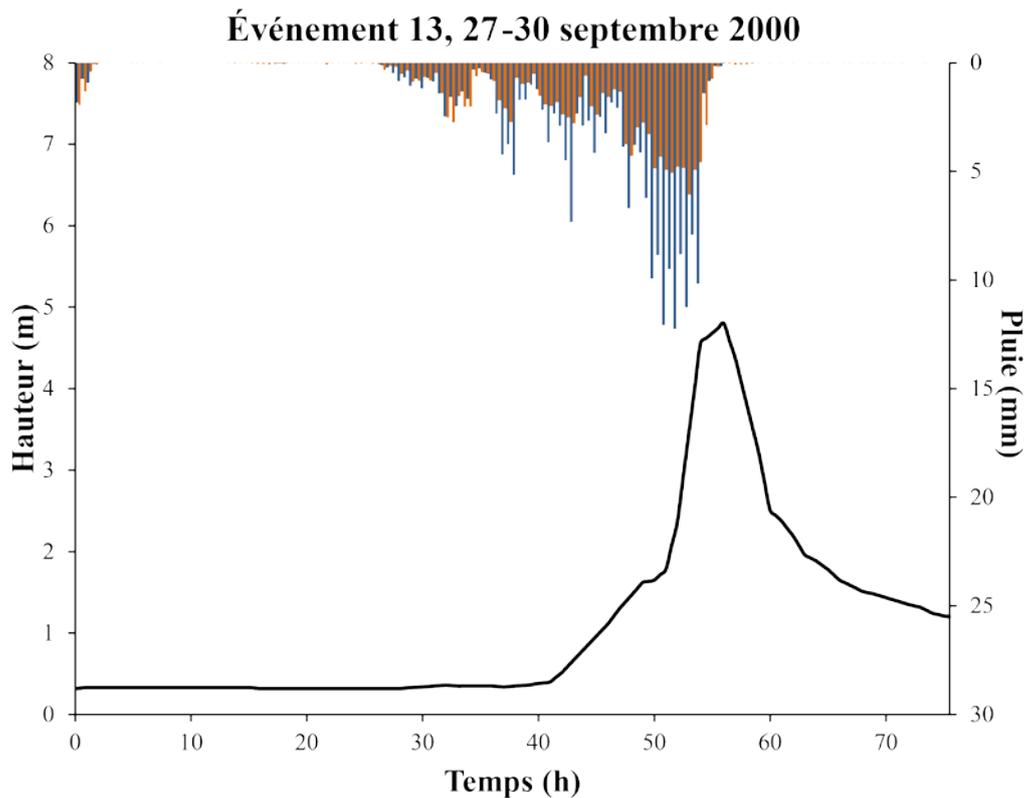
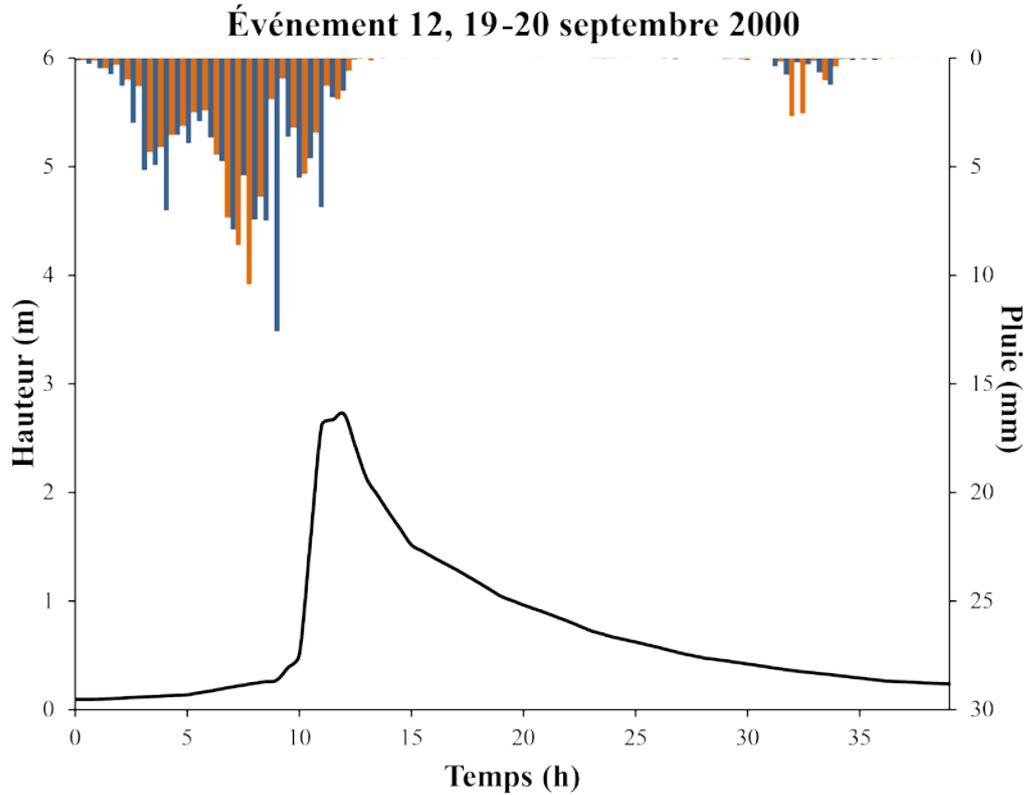
$SPPD$

Retard H_{pic} (h)	$h_p=30\text{min}$	$h_p=1\text{h}$	$h_p=2\text{h}$	$h_p=3\text{h}$	$h_p=4\text{h}$	$h_p=5\text{h}$
Sans adaptativité	0	0,5	0	0,5	2	1,5
$F_a=30\text{min}$	0	0,5	3	0,5	2	1,5
$F_a=1\text{h}$	0	-1,5	0	0,5	2	2
$F_a=2\text{h}$	0	-0,5	2,5	3	2	3
$F_a=3\text{h}$	0	0,5	2,5	3	4	4
$F_a=4\text{h}$	0	0,5	3	3	4	5,5

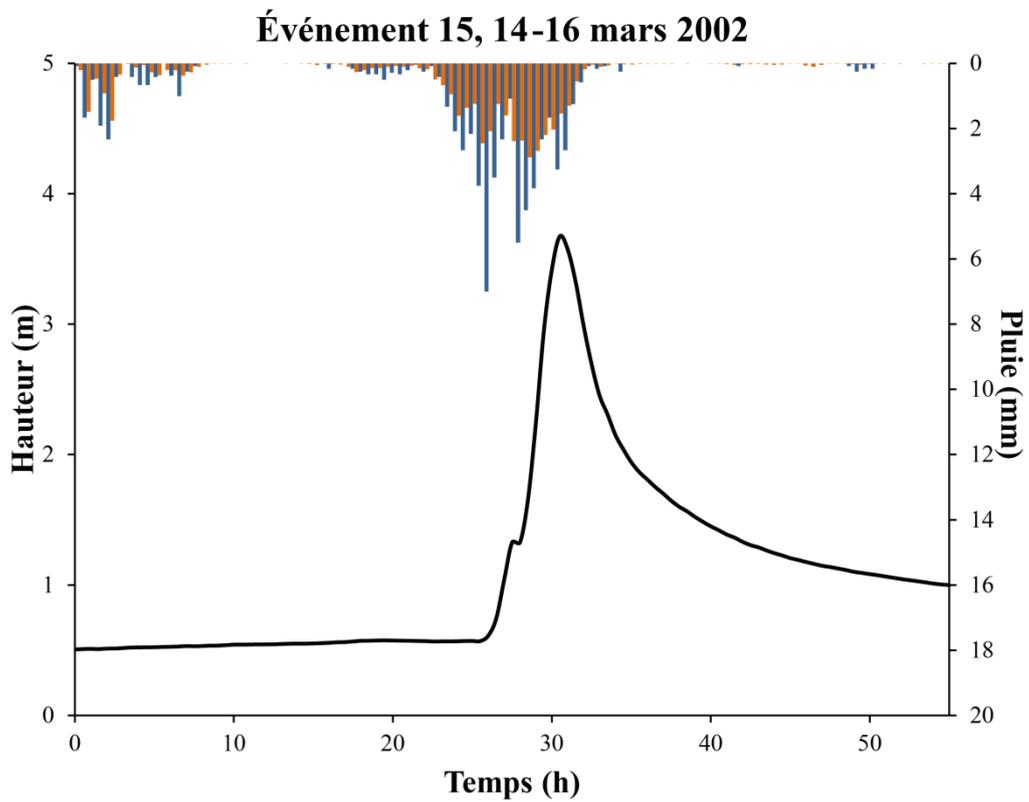
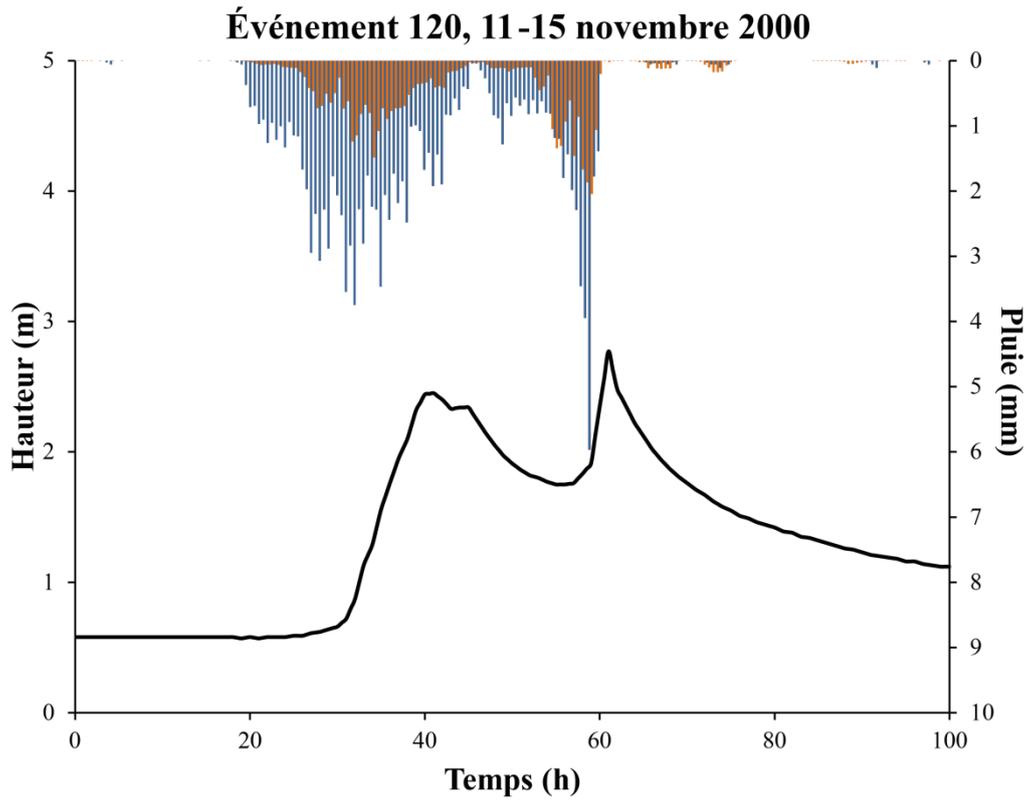
Retard H_{pic}

Annexe 7 : Linnigrammes et hyétogrammes (pluie issue des pluviomètres et pluie obtenue par le radar) des événements compris dans la base de données pour le Gardon d'Anduze (chapitres 5)

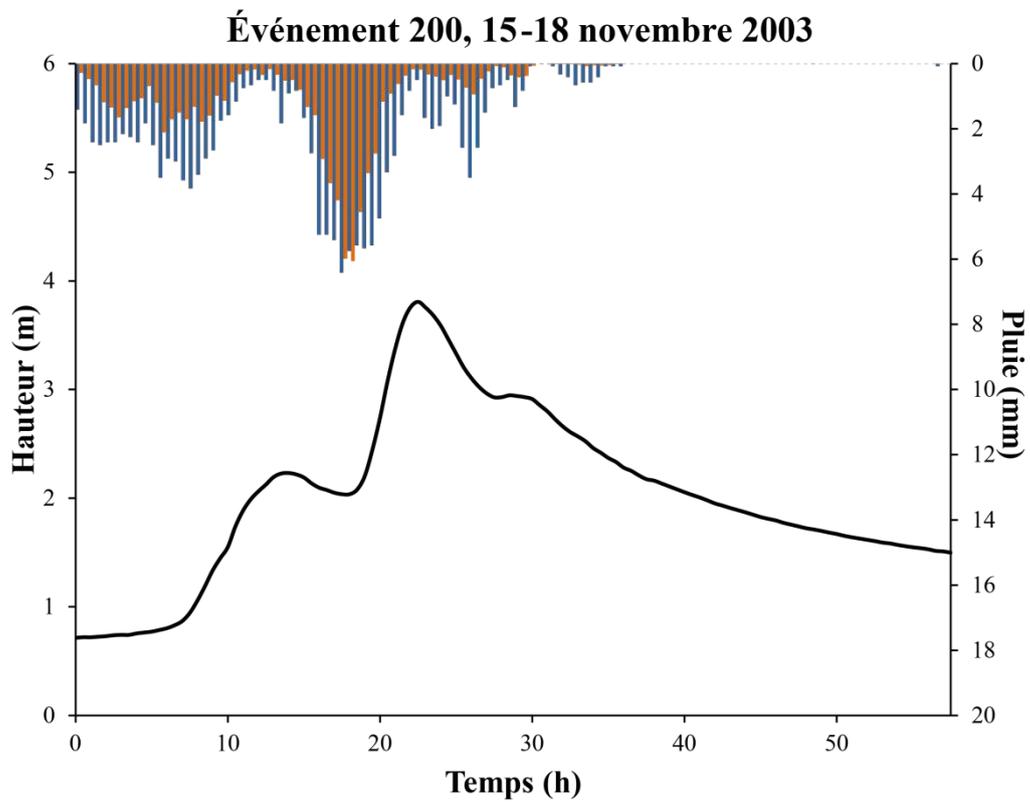
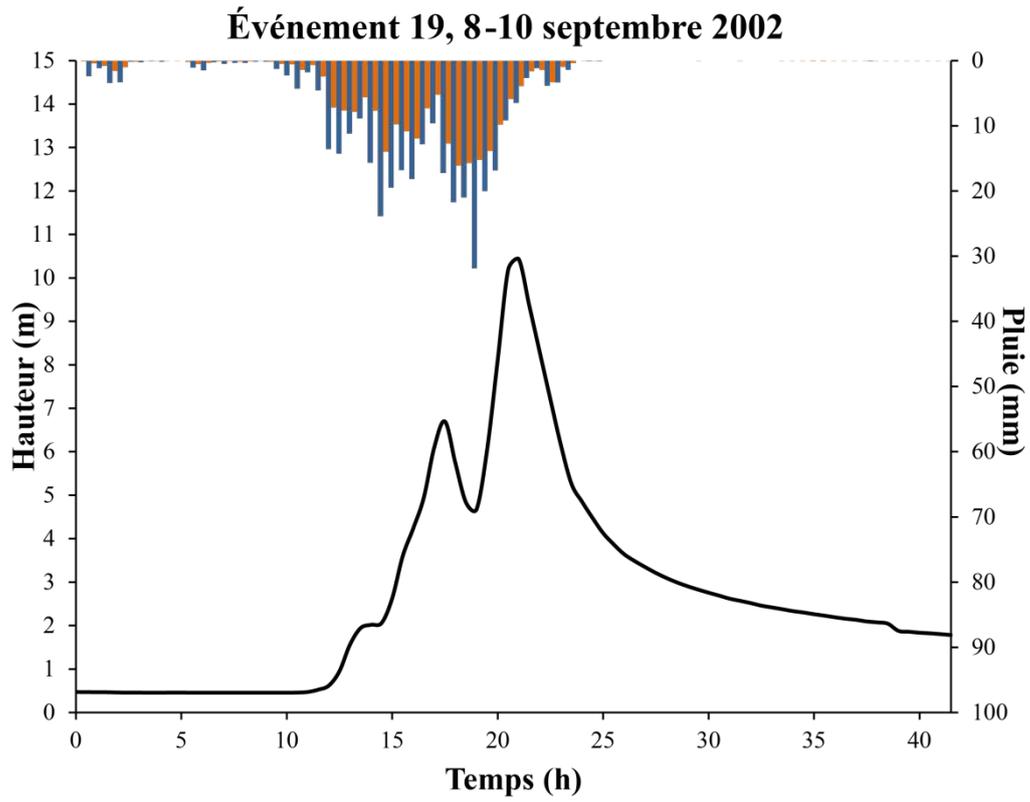
■ Pluie moyenne pluviomètres ■ Pluie moyenne radar — Hauteur à Anduze



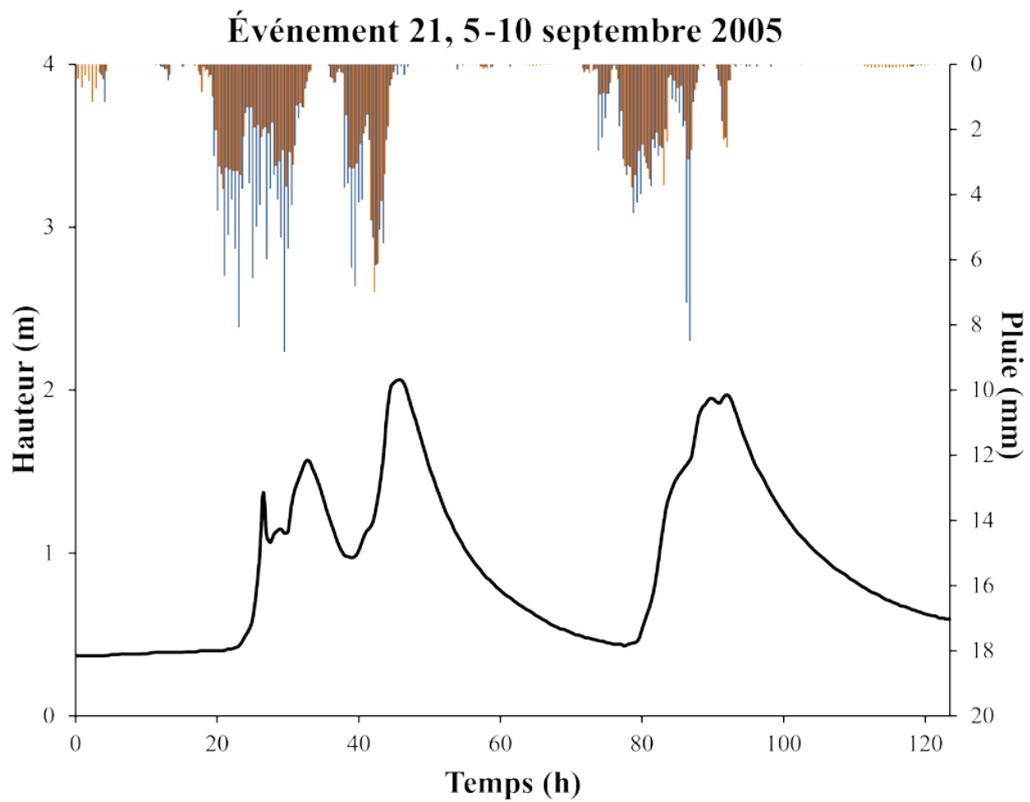
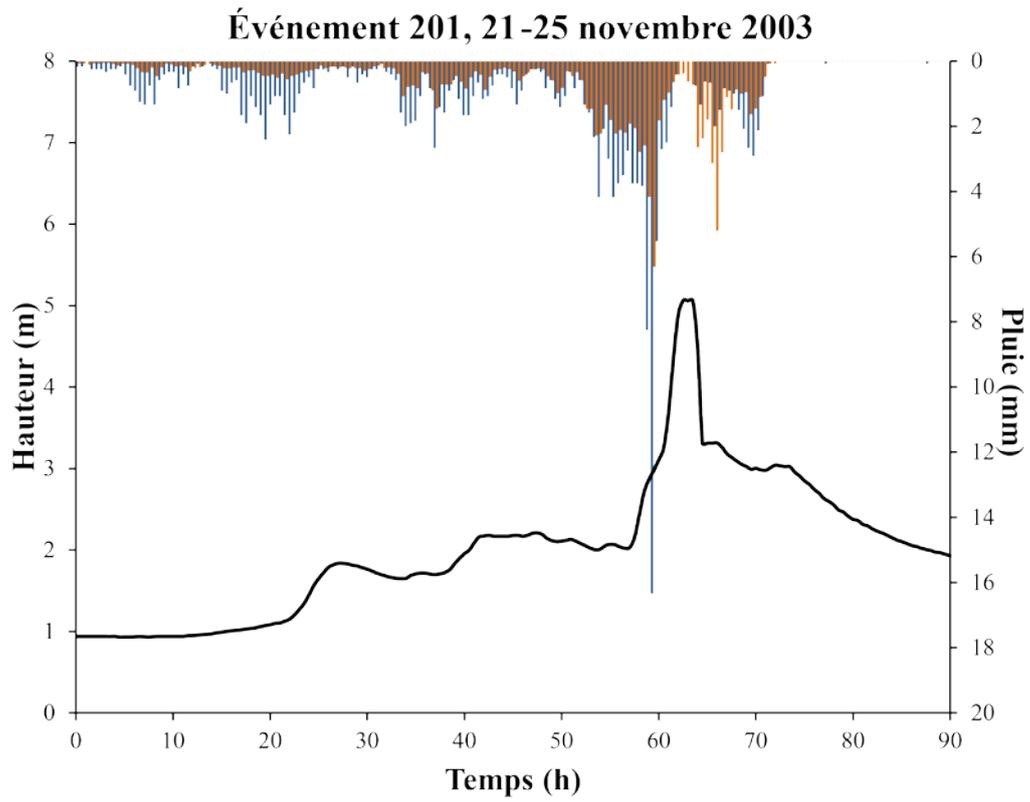
■ Pluie moyenne pluviomètres ■ Pluie moyenne radar — Hauteur à Anduze



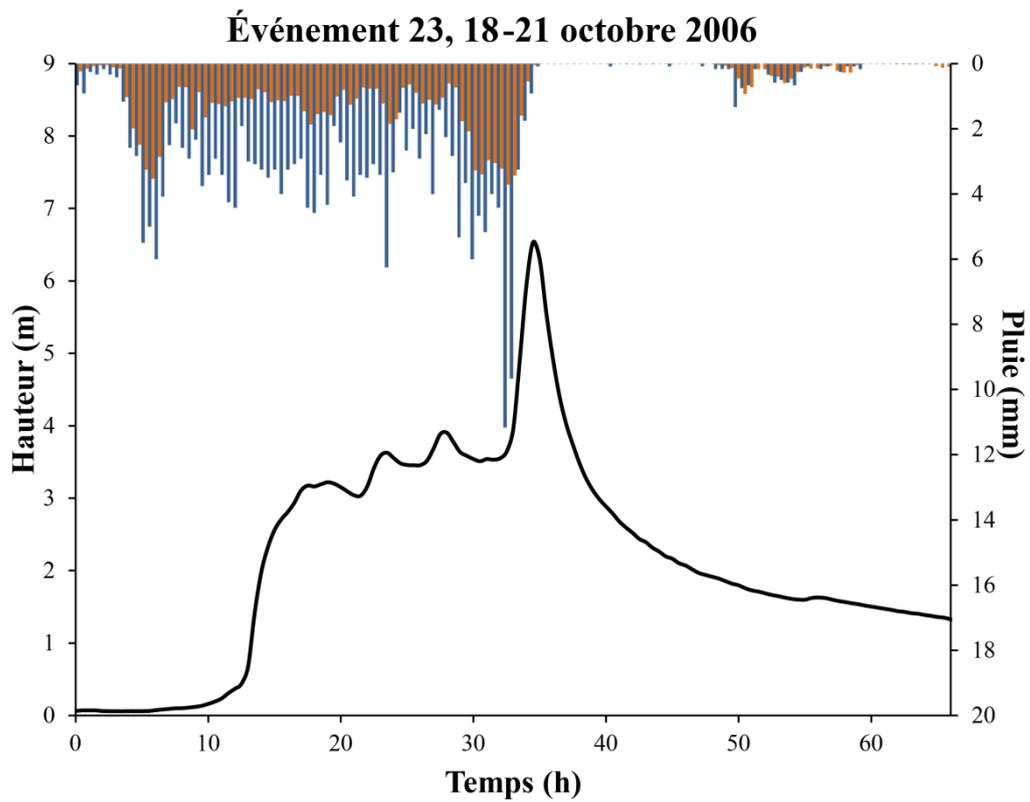
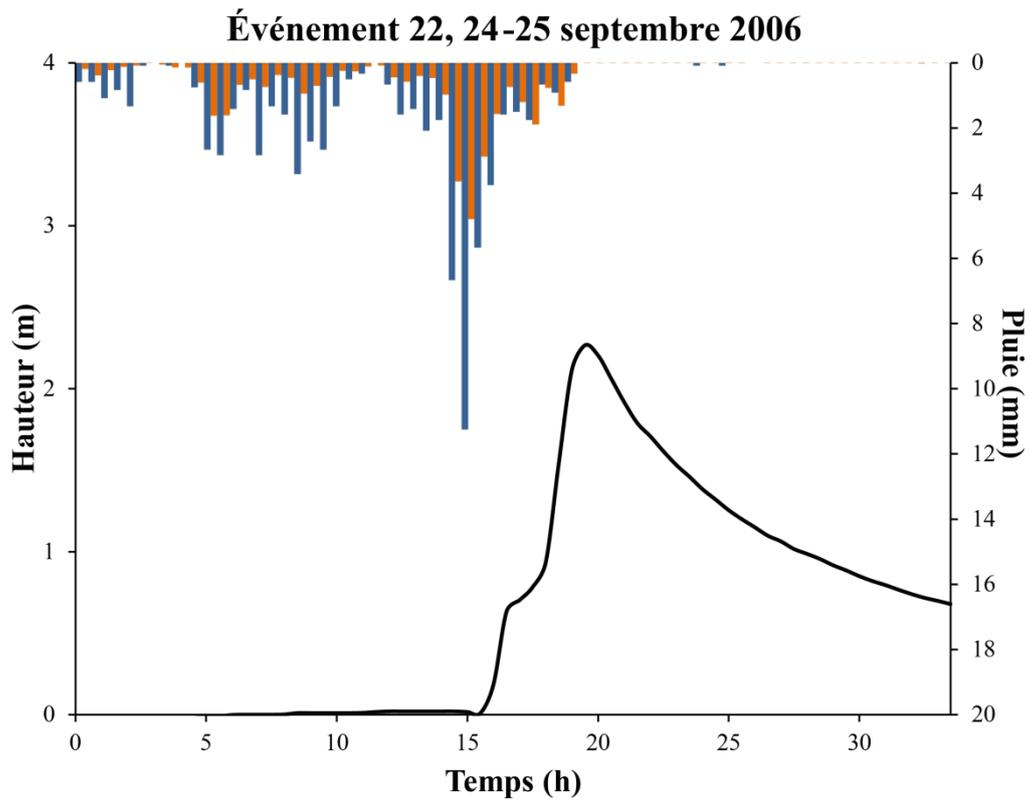
■ Pluie moyenne pluviomètres ■ Pluie moyenne radar — Hauteur à Anduze

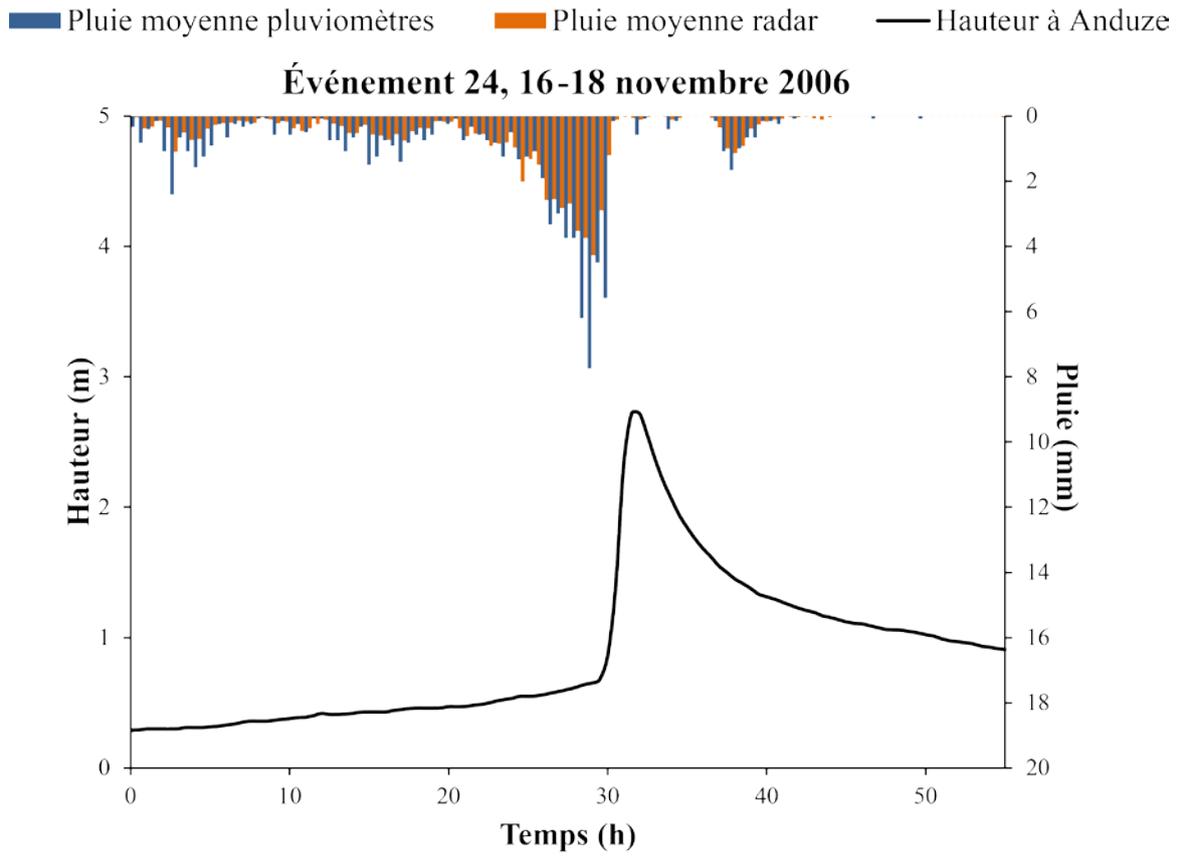


■ Pluie moyenne pluviomètres ■ Pluie moyenne radar — Hauteur à Anduze



■ Pluie moyenne pluviomètres ■ Pluie moyenne radar — Hauteur à Anduze

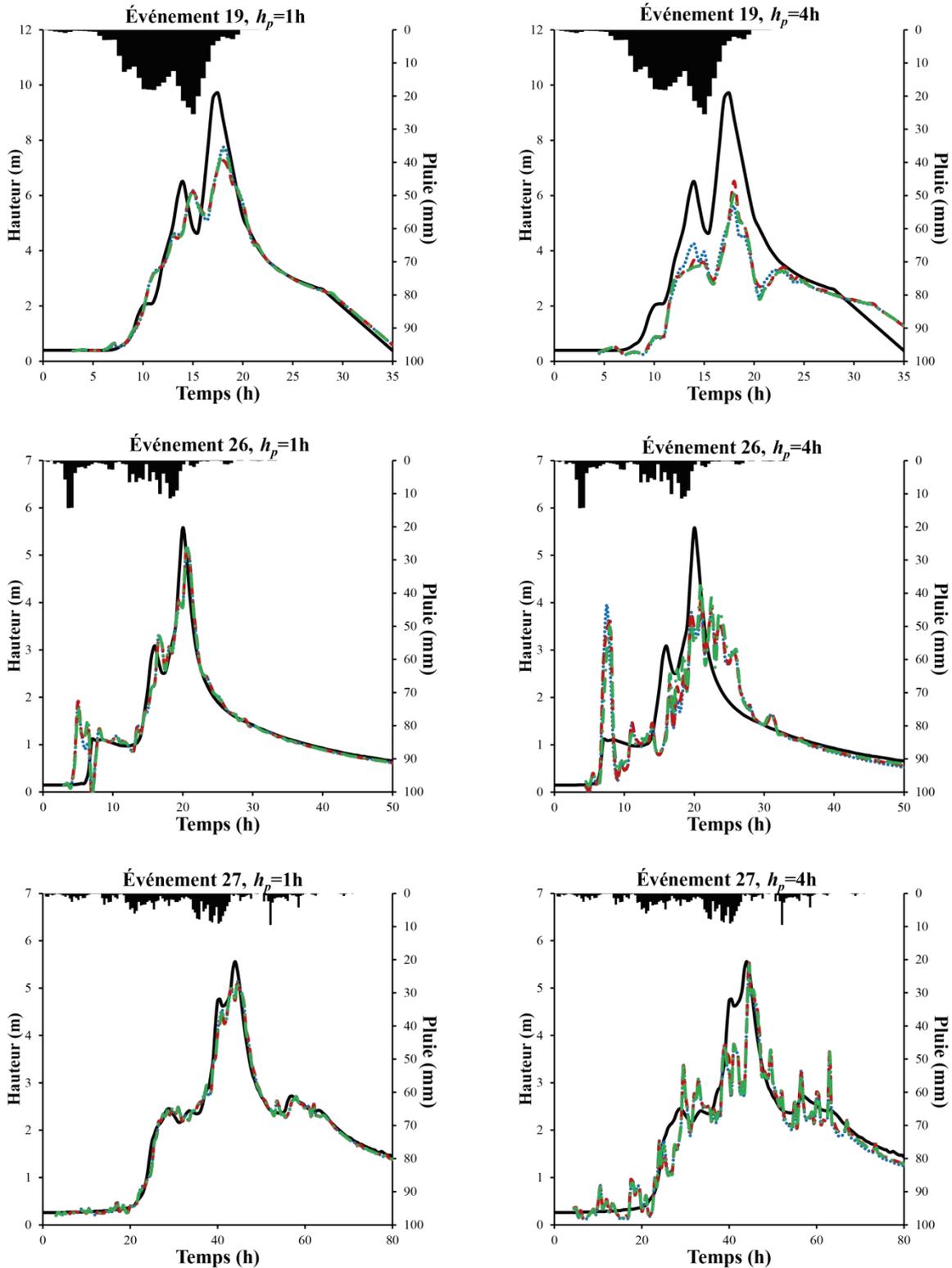




Annexe 8 : Comparaison des hauteurs prévues selon le choix des modèles sélectionnés

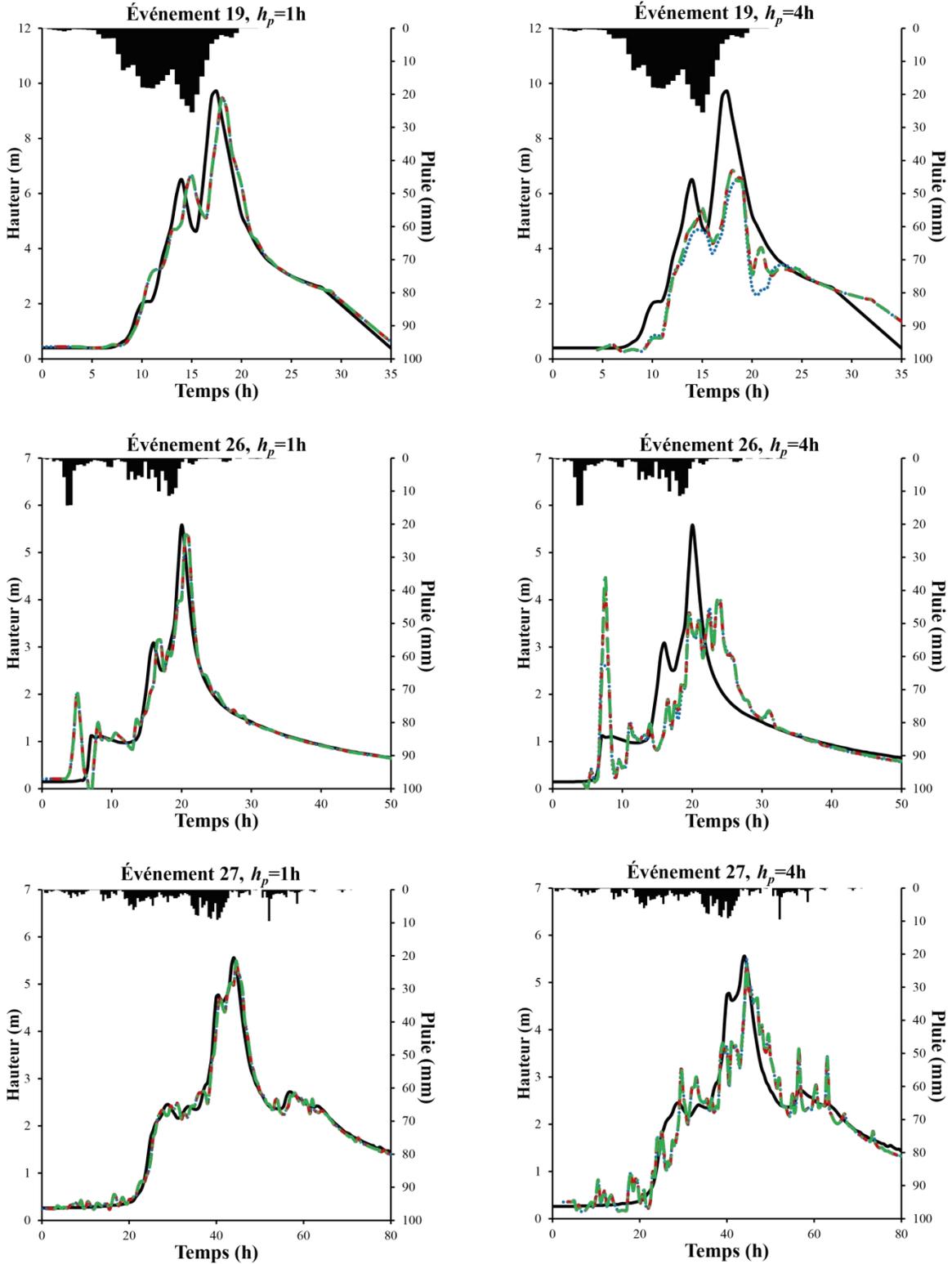
Pour les modèles à perceptron multicouche :

Pluie moyenne
 Hauteur mesurée
 Hauteur médiane 20 modèles vcp
 Hauteur médiane 20 modèles quelconques
 Hauteur médiane 200 modèles

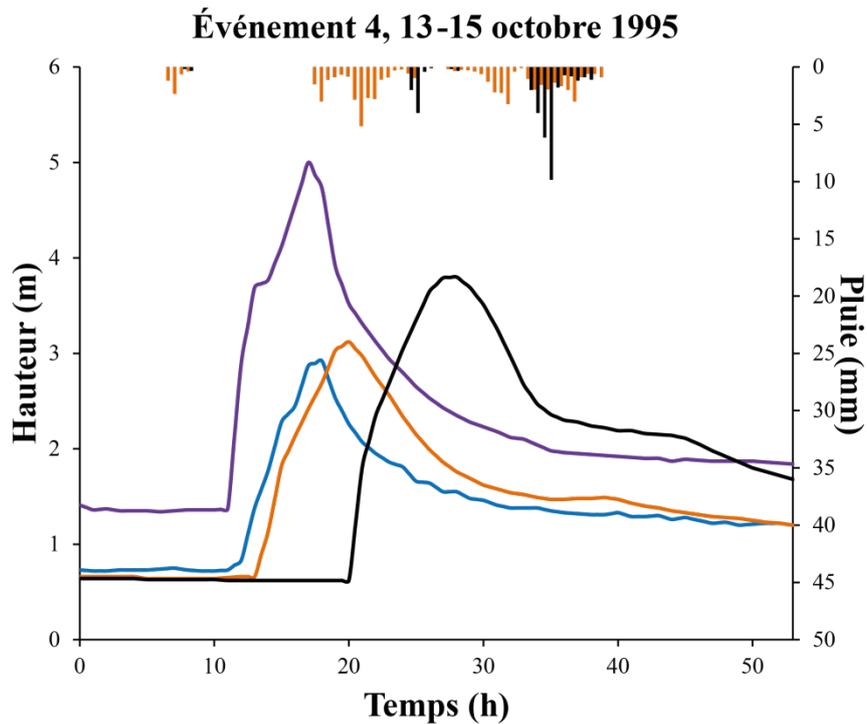
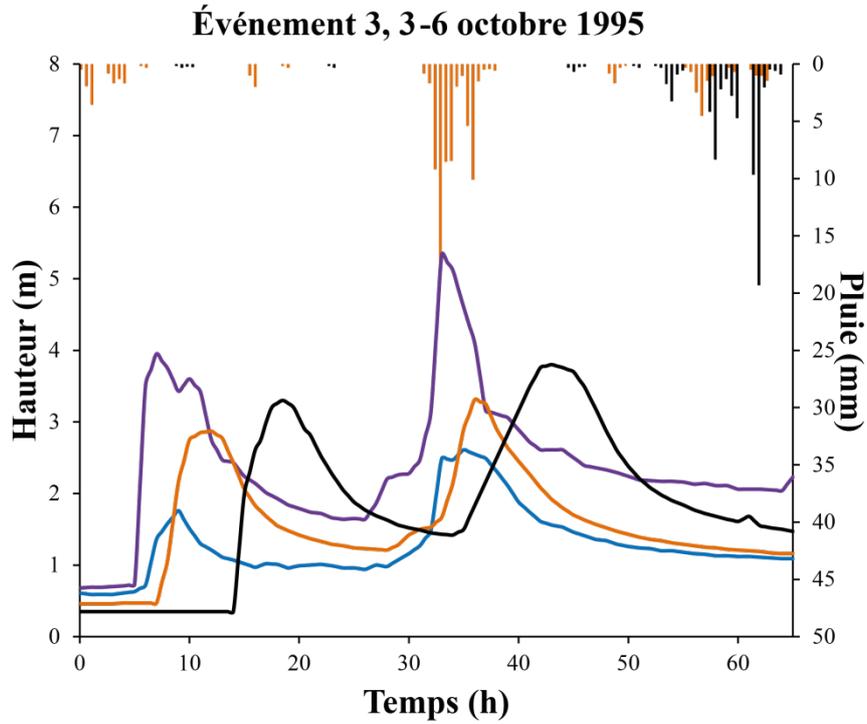
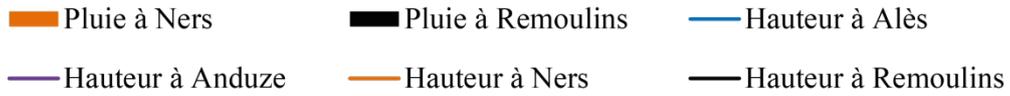


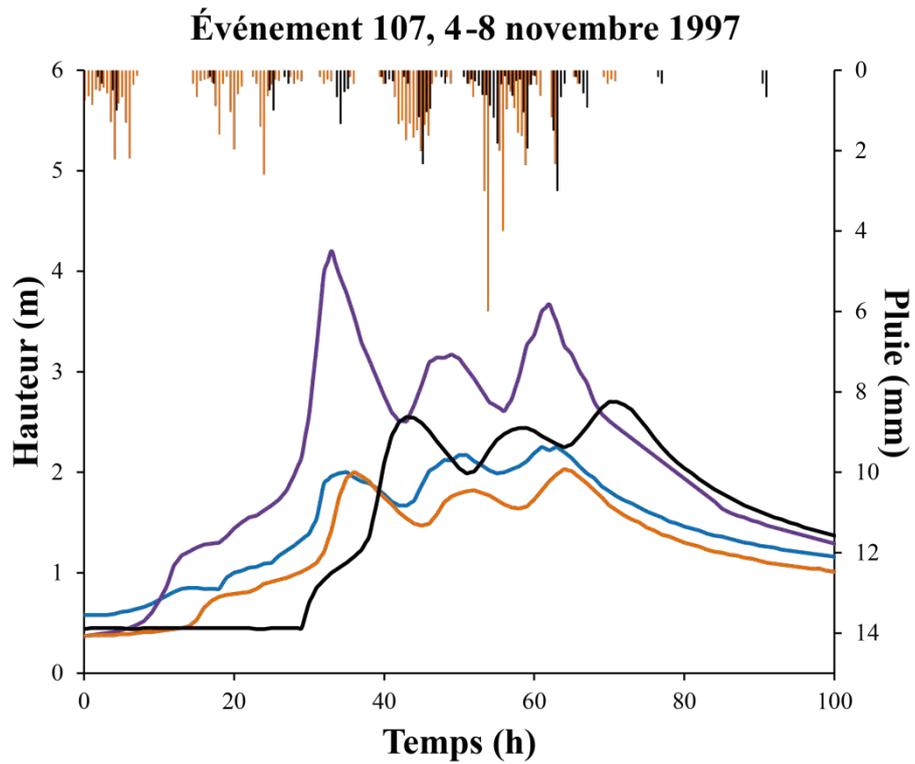
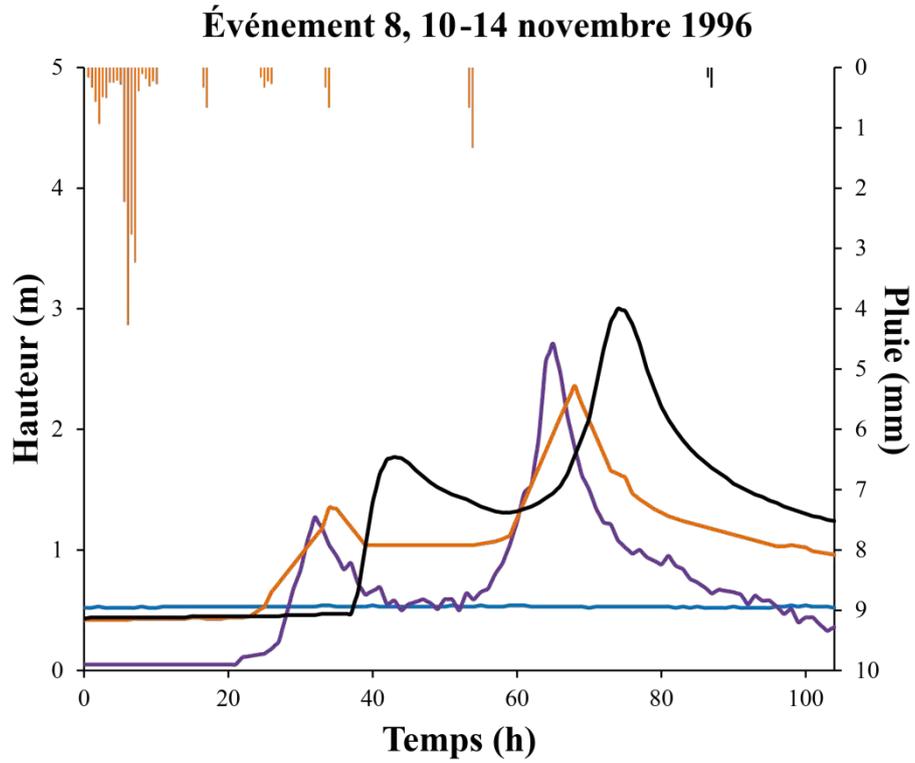
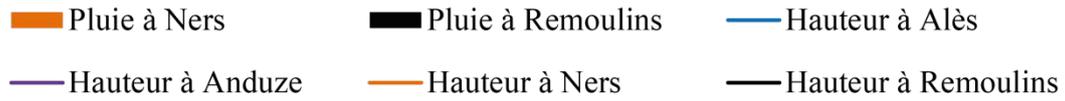
Pour les modèles à perceptron multicouche avec partie linéaire :

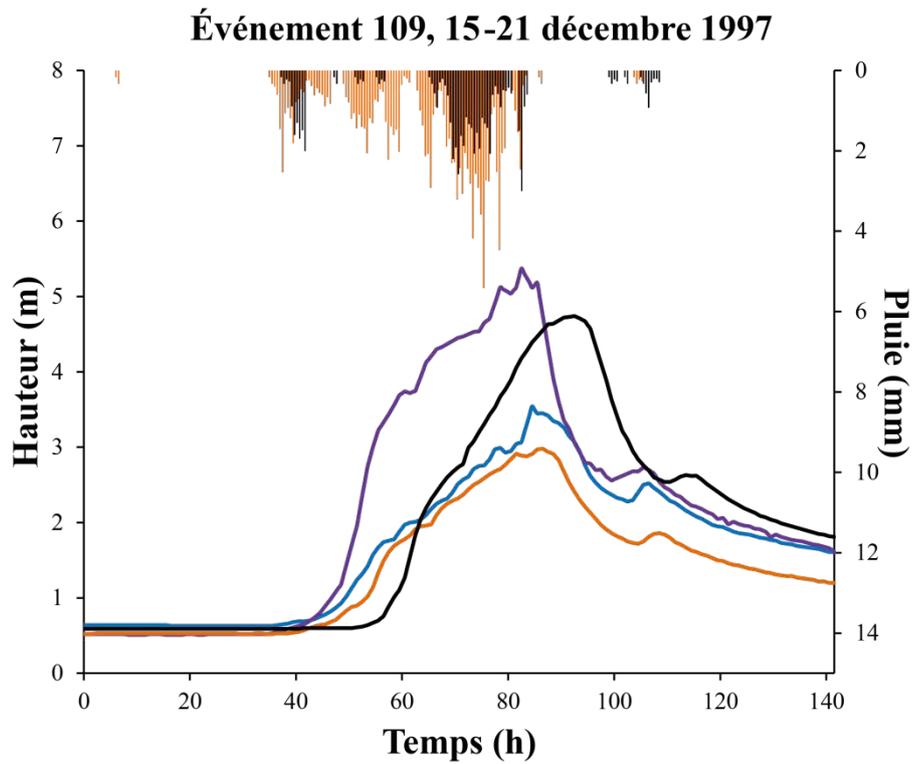
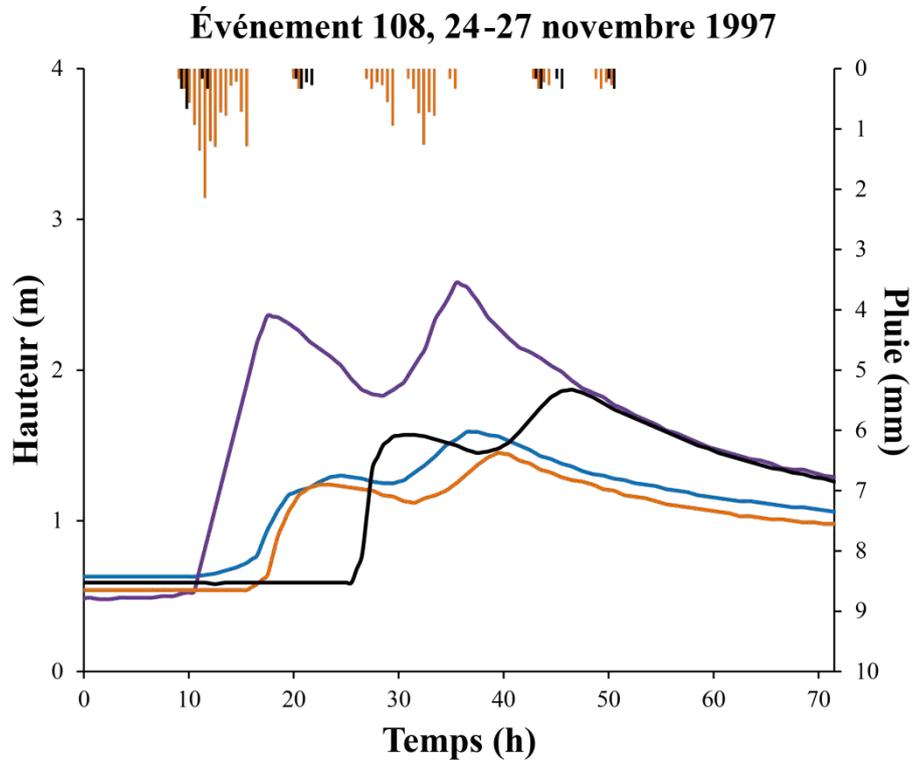
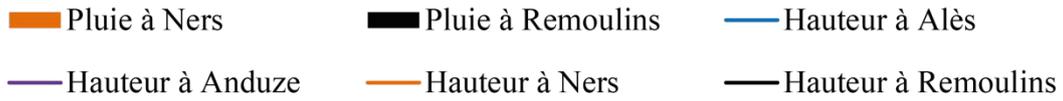
Pluie moyenne
 Hauteur mesurée
 Hauteur médiane 20 modèles vcp
 Hauteur médiane 20 modèles quelconques
 Hauteur médiane 200 modèles

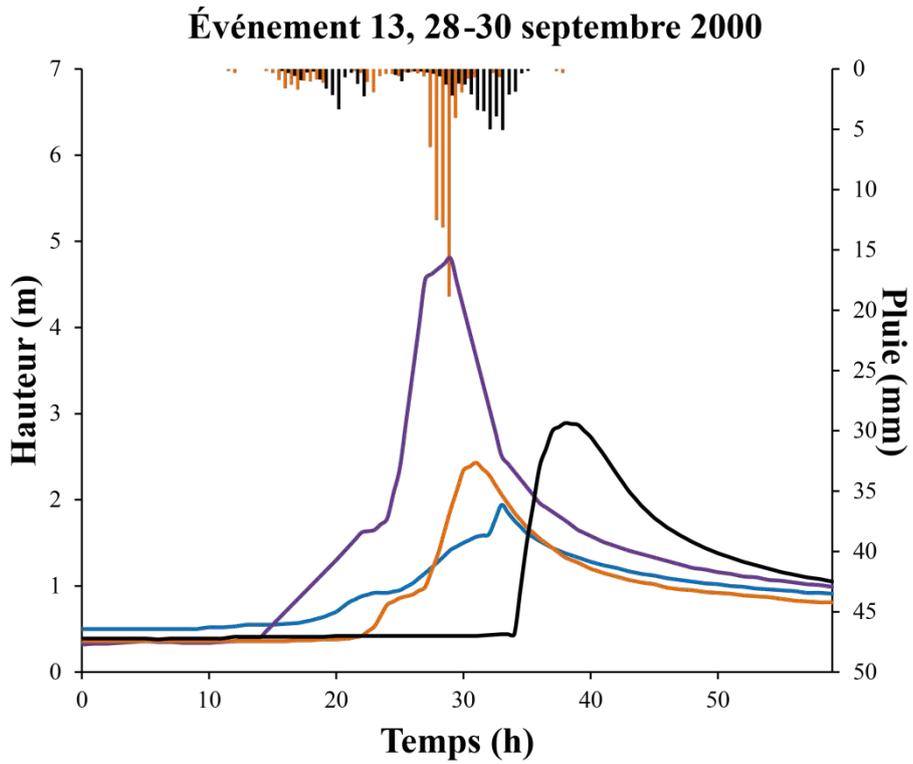
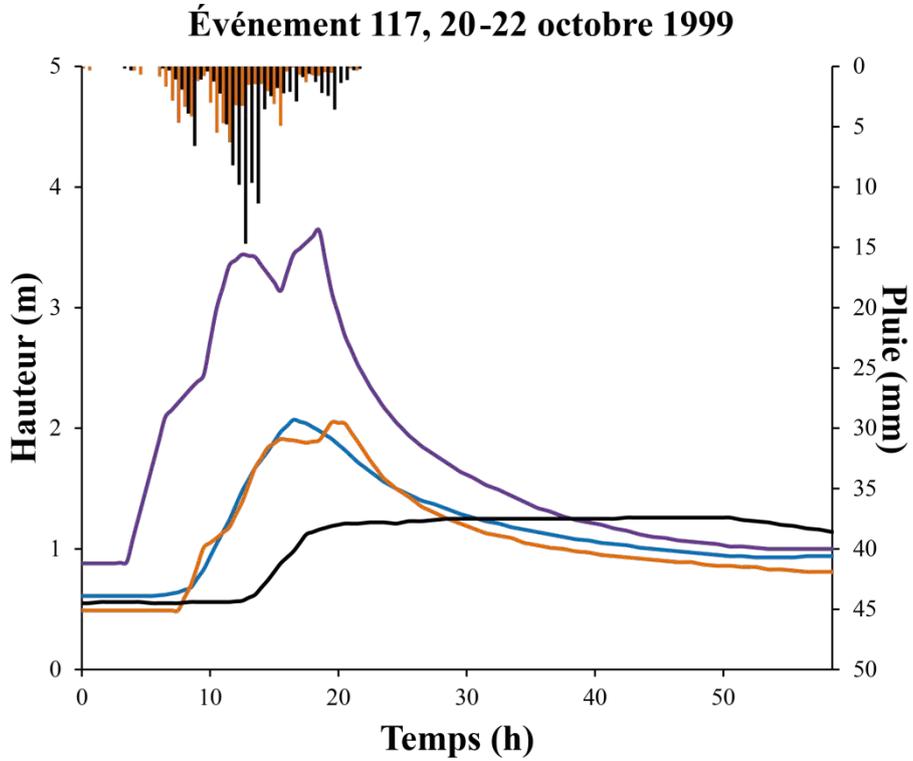
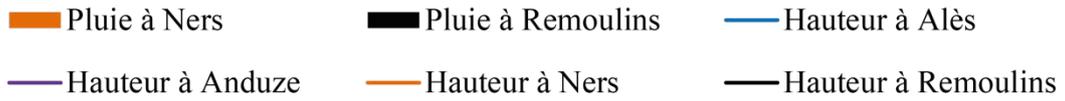


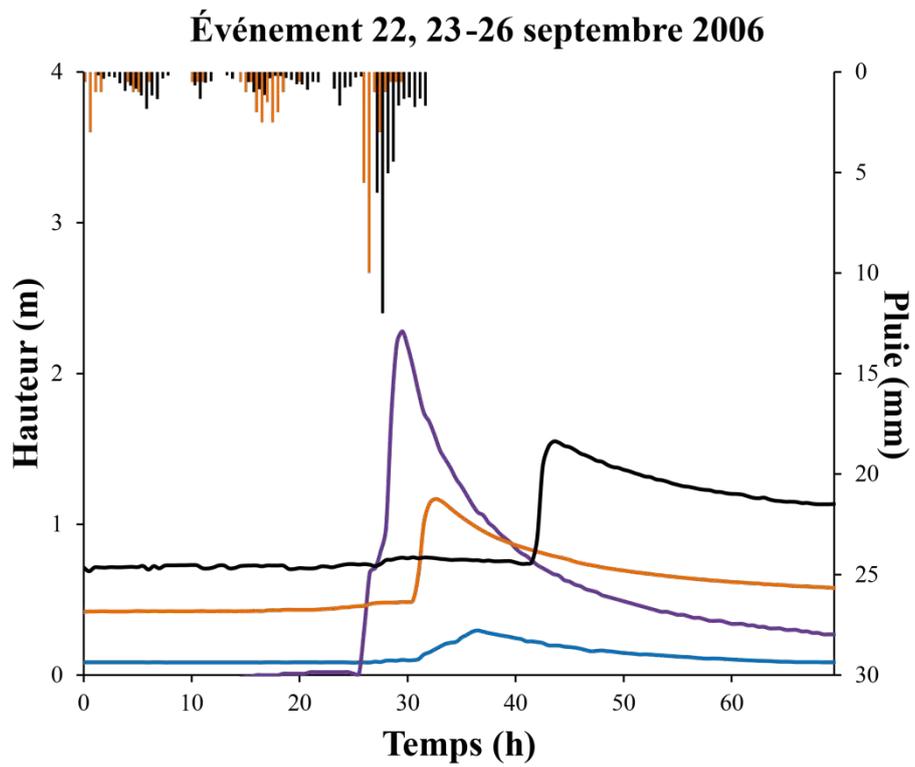
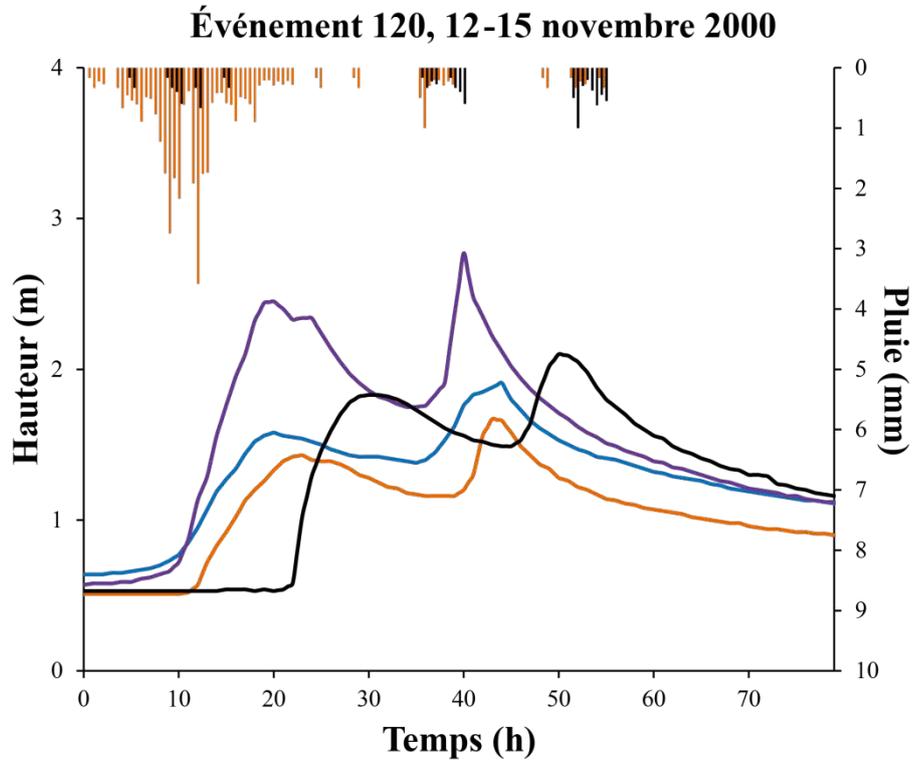
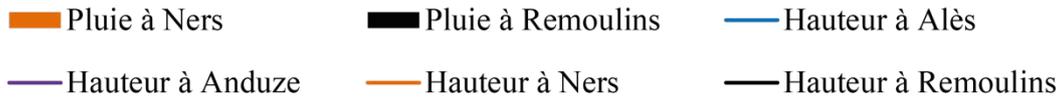
Annexe 9 : Limnigrammes et hyétogrammes des événements compris dans la base de données pour le Gardon à Remoulins (chapitres 6)

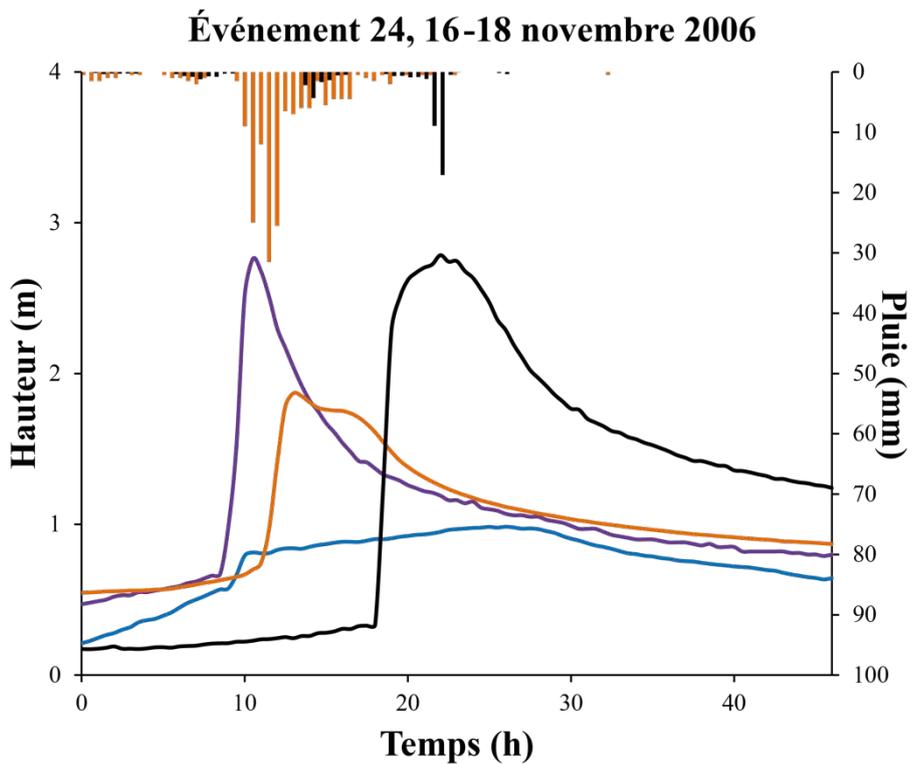
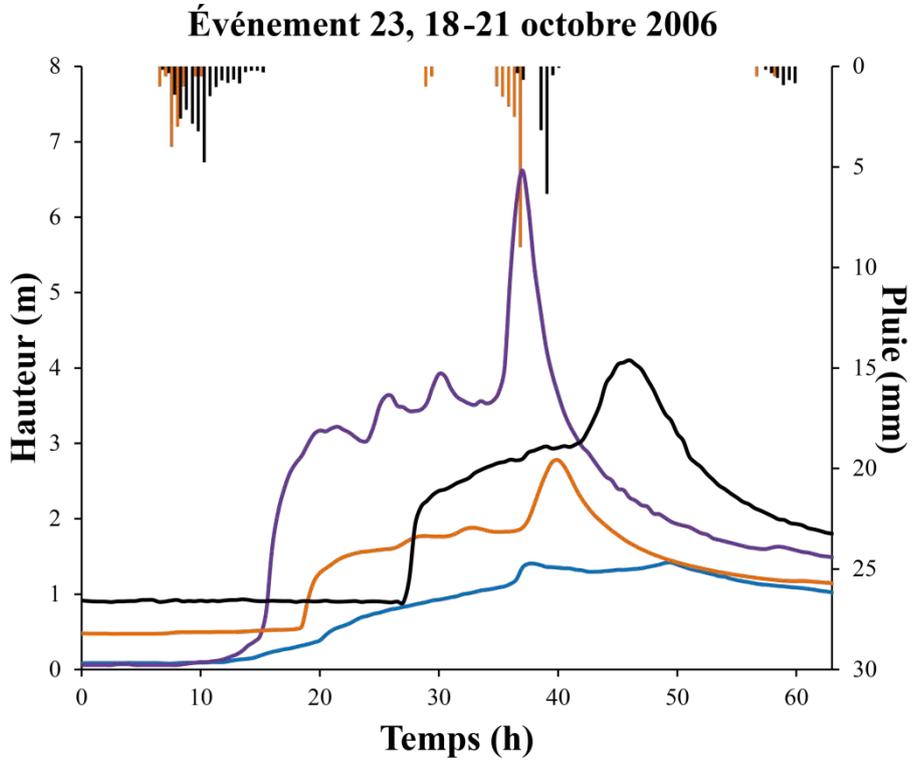
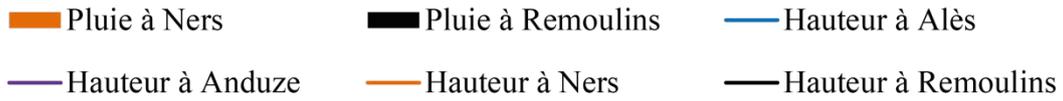


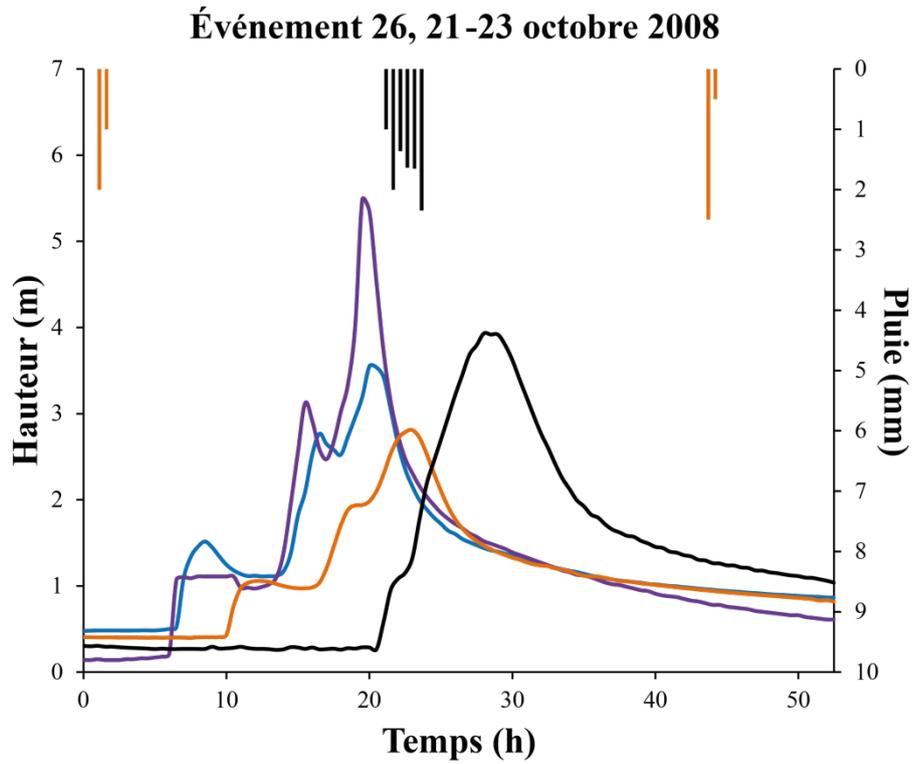
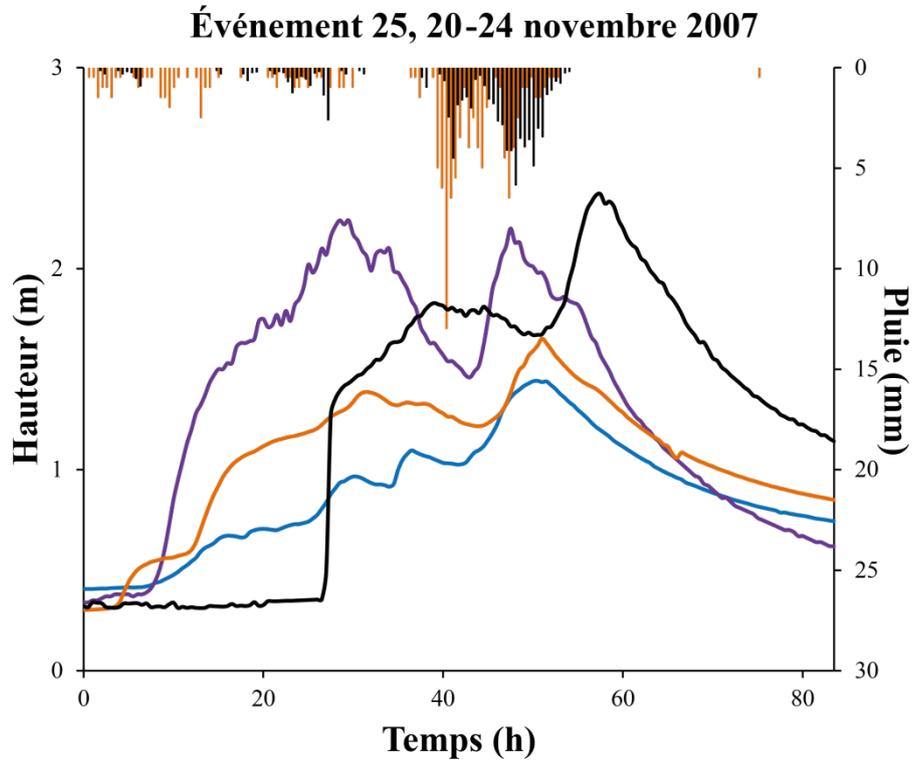
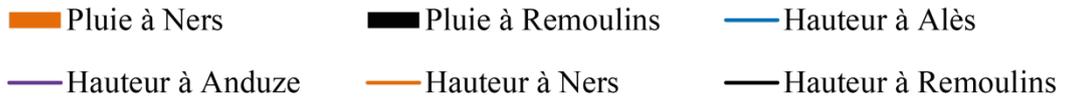


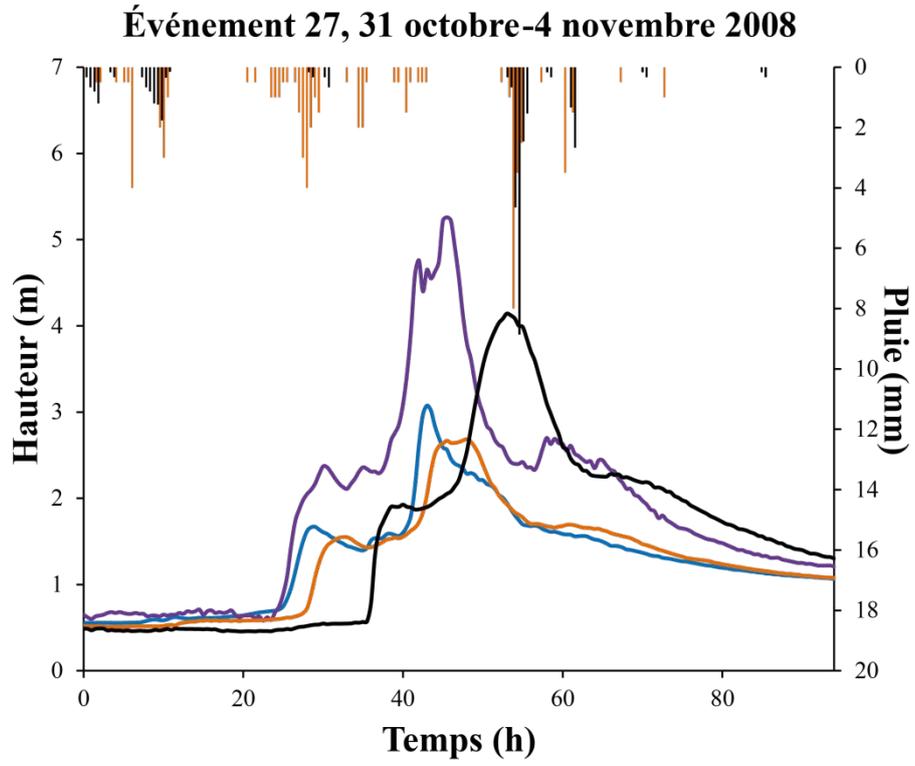
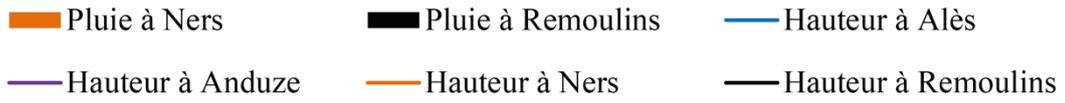






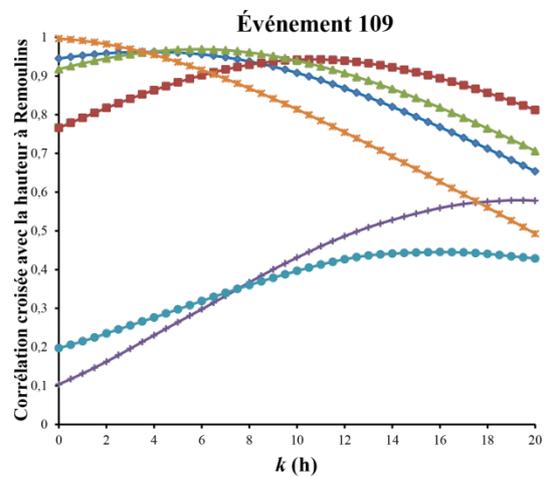
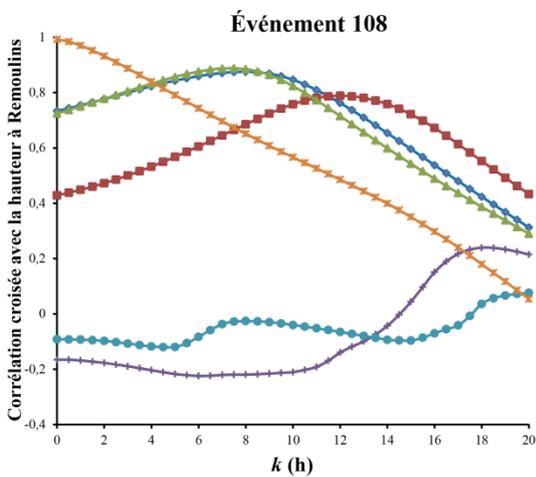
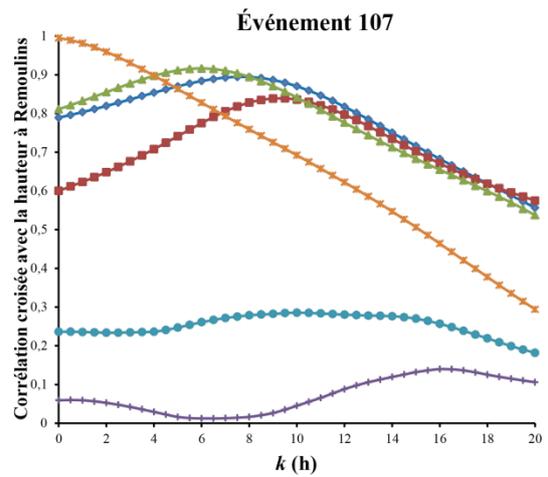
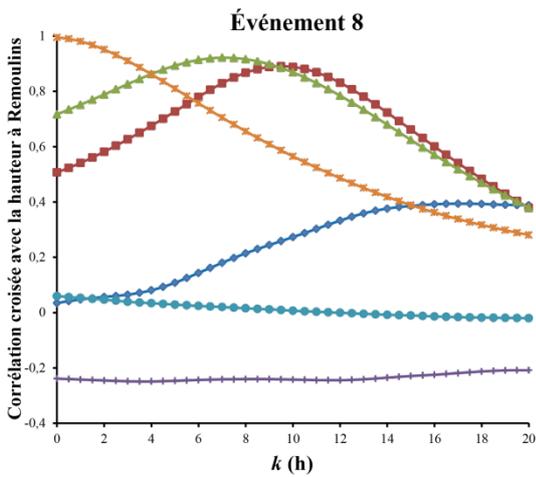
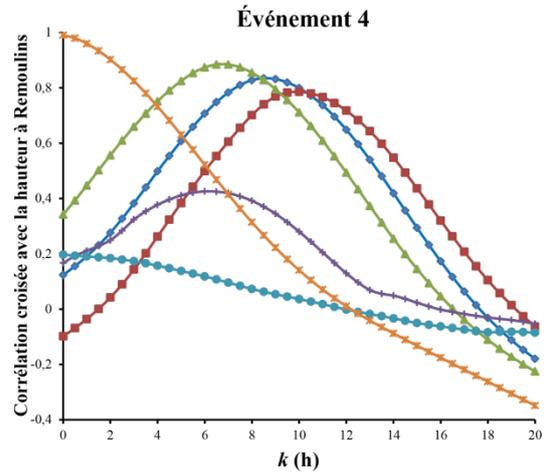
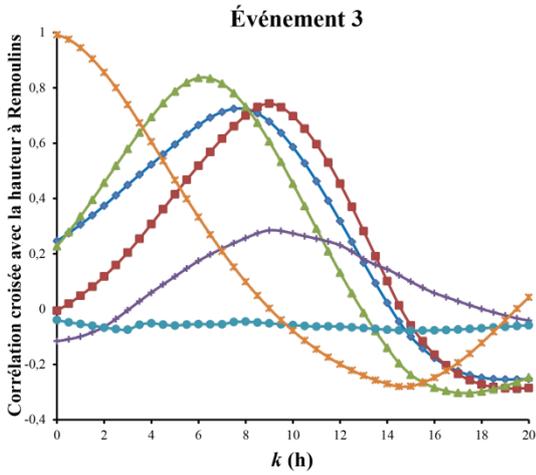


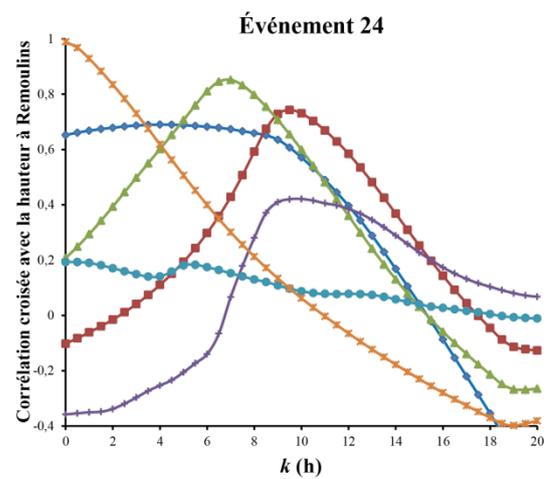
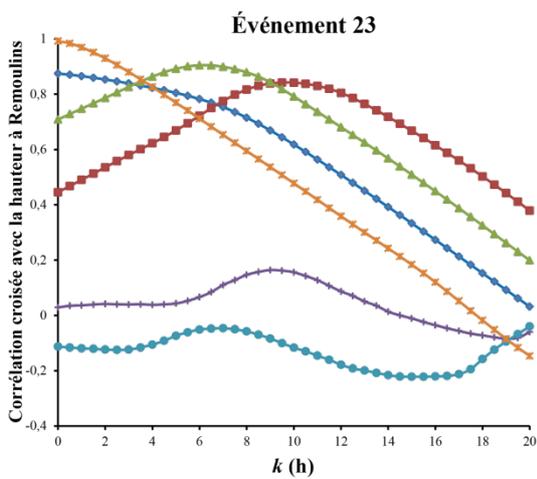
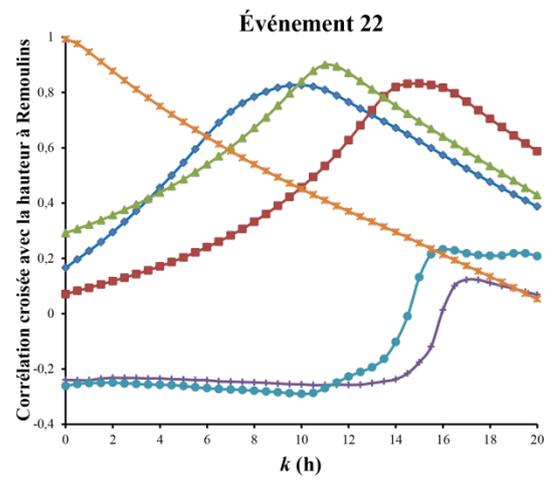
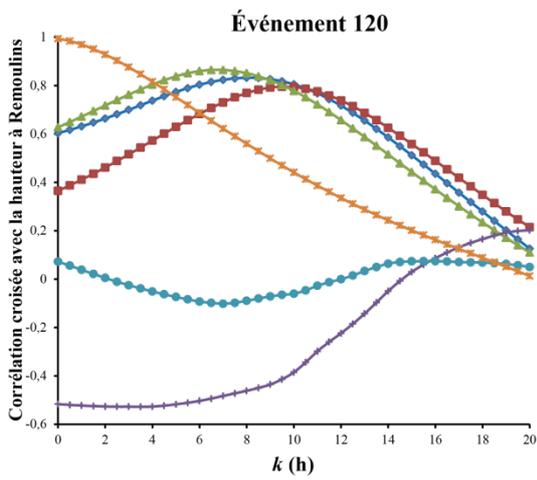
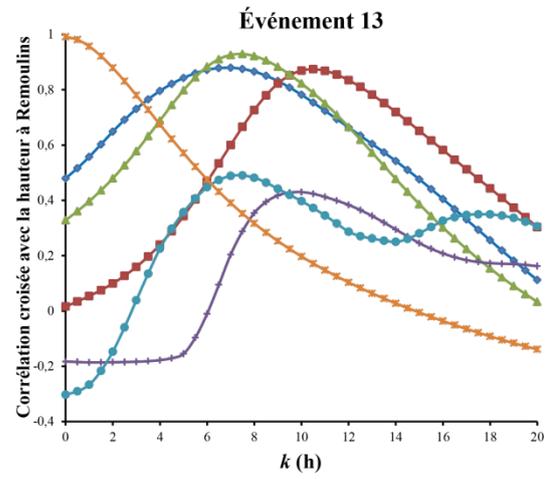
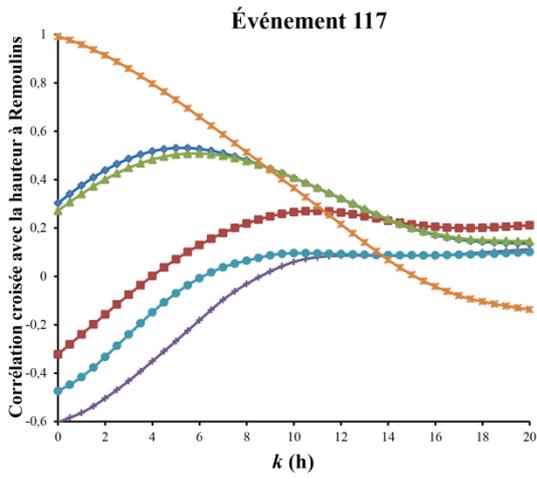
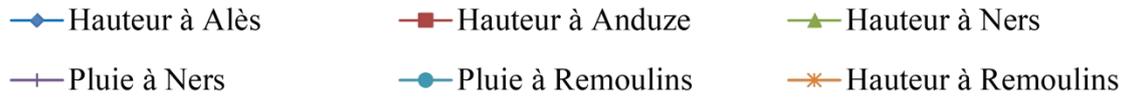


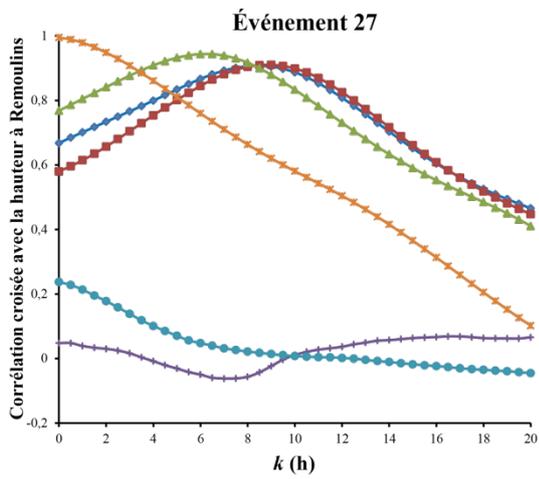
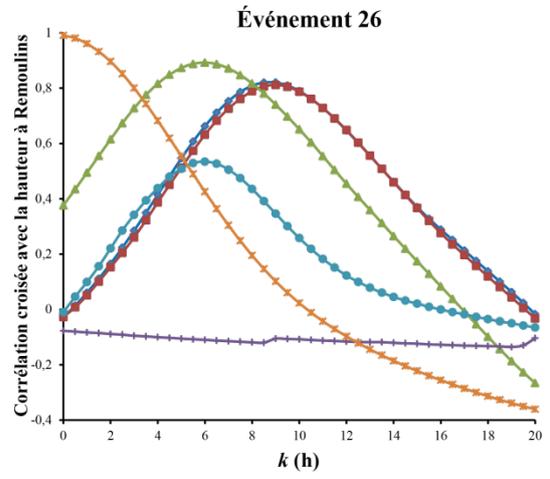
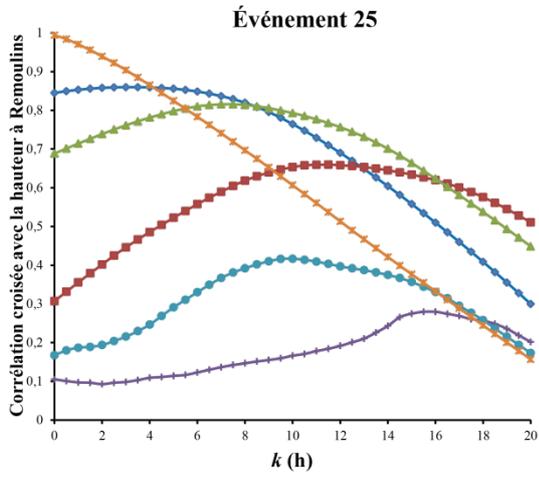
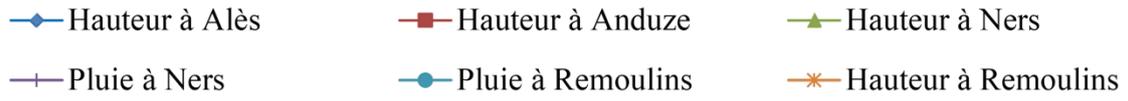


Annexe 10 : Pour chaque événement, corrélogrammes croisés calculés par rapport à la hauteur à Remoulins (chapitre 6)

- ◆— Hauteur à Alès
- Hauteur à Anduze
- ▲— Hauteur à Ners
- +— Pluie à Ners
- Pluie à Remoulins
- *— Hauteur à Remoulins







Annexe 11 : Résumé de la présentation orale effectuée pendant la conférence European Geosciences Union 2011 à Vienne

Flash Flood Forecasting using Neural Networks without Rainfall Forecasts: Model Selection and Generalization capability

Audrey Bornancin Plantier^a, Anne Johannet^a, Pierre Roussel Ragot^b, Gérard Dreyfus^b

^a EMA, Centre des Matériaux de Grande Diffusion, 6 av. de Clavières 30319 Alès cedex, France

^b ESPCI ParisTech, SIGnal processing and MACHine learning (SIGMA) laboratory, 10 rue Vauquelin, 75005 Paris, France

Abstract

The necessity of developing efficient forecasting tools for flash floods has been highlighted by the recent occurrences of catastrophic floods in the south of France such as in *Vaison-la-Romaine* (1992), *Nîmes* (1988), *Gardons* (2002), *Arles* (2003), *Var* (2010). These disasters result from intense rainfalls on small (a few hundreds of km²) and high-slope watersheds, resulting in flows of thousands of m³/s with concentration times of a few hours only [Delrieu 2004]. The death toll (over 100) in these circumstances in the southeast of France [Gaume 2004], and the cost of more than 1.2 billion euros in 2002 [Le Lay 2007] showed that a reliable tool to forecast such phenomena was mandatory.

Real time flood forecasting is a complex task with a growing economic and societal impact. Its complexity arises from the coupling of atmospheric, hydrological and geo-hydrological models; furthermore, the experimental data often lack reliability, which is an additional factor of complexity. Gathering more accurate data, increasing the accuracy of the current physics-based models, and implementing them on increasingly powerful computers, are very useful efforts, but they have limitations. In this context, the FLASH project [FLASH 2010], funded by the French Agency of Research (ANR)¹ proposes an alternative solution, which complements the above-mentioned approach by designing models in a machine-learning perspective.

The watershed of interest is the *Gardon d'Anduze* watershed (540 km²). Neural networks were developed to forecast the water level at *Anduze* for various forecasting horizons from 30 minutes to 5 hours. The database includes 17 flash floods, which occurred between 1994 and 2008. The experimental measurements were supplied by six rain gauges and three gauge stations [Toukourou 2010].

As time plays a functional role in the rainfall-runoff relation, discrete-time dynamic models must be designed. In this work a nonlinear function implemented by a multilayer perceptron with time delays was chosen. The least squares cost function is optimized with respect to the parameters by Levenberg Marquardt optimization, the gradient of the cost function being computed by backpropagation. The water level was predicted as a time series, with rainfalls as exogenous inputs: the model variables were past values of the water levels and of the rainfall. The water level was preferred to the water flow as predicted quantity, because it

¹ FLASH project (*Flood forecasting with machine Learning, data Assimilation and Semi-pHysical modeling*)

makes the prediction independent from the rating curve, which is not known accurately for high outflows.

Due to the high heterogeneity of rainfall, the rainfall forecasts are not yet available at the required small spatial and temporal scales. A specific model is thus adjusted for each forecasting horizon (half of an hour to 5 hours) without future rainfall information.

As the model complexity control is a particularly critical issue, due (i) to the lack of accurate estimations of rainfalls, and (ii) to the high noise level in water level measurements, the traditional early stopping regularization method was used. Model complexity selection was performed by a variant of cross-validation [Dreyfus 2005] using various validation scores. The sliding window width for rainfalls and for past levels, as well as the number of hidden neurons, and the hyperparameters of the optimization algorithm were estimated similarly.

The quality of the generalization is assessed by performance criteria calculated on three test events (independent from the training and validation sets): September 2002, October 2008 and November 2008.

Hydrographs at several forecasting horizons are displayed. Very satisfactory results are obtained up to a forecasting horizon of three hours (Nash criteria evolving between 0.95 for half of an hour, and 0.50 for three hours), thereby allowing early warnings to be issued to the public.

Keywords: Flash flood, forecasting, neural network, machine learning, model selection, cross validation

References

[Delrieu 2004]

G. Delrieu, V. Ducrocq, E. Gaume, J. Nicol, O. Payrastre, E. Yates, P. Kirstetter, H. Andrieu, P.A. Ayral, C. Bouvier, J. Creutin, M. Livet, S. Anquetin, M. Lang, L. Neppel, C. Obled, J. Parent-du- Chatelet, G.M. Saulnier, A. Walpersdorf, W. Wobrock "The catastrophic flash-flood of 8-9 september 2002 in the Gard region, France: A first case study for the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory". *Journal of Hydrometeorology*, 6, 34-52 2004

[Dreyfus 2005]

G. Dreyfus. "Neural networks: methodology and applications". Springer, 497 p. 2005

[FLASH 2010]

<http://edytem.univ-savoie.fr/spip/programmes-et-contrats-de/flash/>

[Gaume 2004]

E. Gaume, M. Livet, M. Desbordes, J.P. Villeneuve "Hydrological analysis of the river Aude, France, flash flood on 12 and 13 November 1999". *Journal of Hydrology*, Volume 286, Issues 1-4, 30, p ; 135-154, 2004

[Le Lay 2007]

M. Le Lay, G.M. Saulnier "Exploring the signature of climate and landscape spatial variabilities in flash flood events: Case of the 8-9 September 2002 Cévennes-Vivarais catastrophic event". *Geophysical Research Letters*, 2007, 34, L13401, 2007

[Toukourou 2010]

M. Toukourou, A. Johannet, G. Dreyfus, P. A. Ayrat "Rainfall-runoff Modeling of Flash Floods in the Absence of Rainfall Forecasts: the Case of "Cévenol Flash Floods". *Applied Intelligence*, ISSN 1573-7497, (in line 2010)

Annexe 12 : Article pour Neurocomputing (en cours de finalisation, sera rajouté dans la version du mémoire remise au jury le jour de la soutenance)

Sensitivity analysis of neural network's parameters and hyperparameters for Flash Flood Forecasting: Model Selection and Generalization capability.

Audrey Bornancin-Plantier^{1,2}, Anne Johannet¹, Pierre Roussel Ragot², Gérard Dreyfus².

(1) Ecole des Mines d'Alès, CMGD 6 av. de Clavières, 30319 Alès Cedex, France.

(2) ESPCI-Paristech, Laboratoire SIGMA, 10 rue Vauquelin, 75005 Paris, France.

Abstract: Flash flood forecasting has important issues in Europe, where such flooding causes regularly casualties and damages. These phenomena are difficult to forecast because of lack of knowledge about involved physical processes, inability of short-term rainfalls forecasts and uncertainties of the measurements, above all concerning the rainfalls. Because of misreading about the physical phenomenon, machine learning appeared as a suitable method in order to estimate and forecast water level in such watersheds. Faced with such issues, this paper presents the neural networks design of predictors devoted to the forecasting of water level of the *Gardon d'Anduze* at *Anduze*, archetypal basin of mountainous Mediterranean basins, with a prediction horizon from 30 min to 5 hours. Thanks to a method of selection derived from the cross-validation, the number of hidden neurons, temporal window of the rainfalls, training hyperparameters and initialization of the parameters are determined to optimize the models capacities. Surprisingly the predominant step in model design was not the estimation of complexity as suggested by the bias variance tradeoff but the random initial choice of parameters. This sensitivity is tested on three events, one of whose is the most intense of the database. It appeared in a first approach that, due to the great variability of cross-validation scores generated by the parameters initializations, criteria used for model selection were unable to select models efficient for all events. In order to diminish such variability, new architecture were proposed including linear and non-linear contribution. The linear part diminishing the role of parameters initialization. Based on this approach and using the same selection methodology than in the first part of the paper, designed predictors prove excellent results up to 3 hours and useful anticipation up to 4 hours, allowing thus to design a flood prediction prototype, currently in test in the French Flood Surveillance Service (SCHAPI). As the method is generic it could be applied on others basins; conditioned by the availability of sufficient dataset.

Introduction

Europe is a temperate region however subjected to great water-related disasters causing casualties and material damages. Face to this hazard each country provides its own early warning system [Alfieri 2012]. Concerning flash flooding hazard, the south of France (and more generally, Mediterranean region) is regularly subjected to catastrophic events as it occurred in *Vaison-la-Romaine* (1992), *Nîmes* (1988), *Gardons* (2002), *Arles* (2003), *Var* (2010). The causes of these phenomena are the intense rainfalls on small and high-slope

watershed which lead to flows of thousands of m^3/s associated with water concentration times of a few hours only. The consequences of these flash floods are over 100 fatalities during the last 20 years in the southeast of France [Gaume] and important material damages: for example more than 1.2 billion euros in the sole event of 2002 [Le Lay].

Faced with the necessity to warn and protect the population, the French service of flood warning (SCHAPI (*Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations*)) provides real time *vigicrues* map feeding (<http://www.vigicrues.ecologie.gouv.fr>) and participates to several projects. Amongst them, the FLASH project [FLASH 2010], funded by the French Agency of Research (ANR), proposes to complement the hydrological approach by designing models in a machine-learning perspective. The present work takes place in this project and aims to study the neural model design in order to provide real-time reliable forecasting for the *Gardon d'Anduze* watershed.

As the phenomena involved on such fast response to huge rainfall is not currently well apprehend, and because of the unavailability of rainfall forecasts to the useful scale of time (the hour) and space (the dozen of square kilometers), the real time performances of physically-based hydro-meteorological models are currently unreliable in real time.

For this reason, machine learning appeared as a suitable solution to investigate the feasibility of water level, or discharge, prediction in a framework of black box modeling. Focusing on neural networks because of their ability to implement dynamical hydrosystems, a lot of work has been devoted to forecast watershed behavior with, or without future rainfall forecasts [Abrahart, Toth, Ju, Kong A Siou 2011]. Generally it was shown that neural models can perform forecasting without rainfall forecasts and that the more the lead-time increases, the more they can outperform linear modeling [Bazartseren]. Considering overfitting which can potentially be exacerbated by noisy data, regularization methods must be used and contribute to provide good forecasts on intense events [Toukourou 2009, Kong A Siou 2012].

After the introduction, the presentation of the flash floods on the study place, the *Gardon d'Anduze* watershed will be provided in the second part. An analysis of both the database and the uncertainties associated will be presented, added with a description of current physic-based models involved in this basin. Then the models design for lead-time from half an hour to five hours will be presented in third part, associated with the methodology used to select the adequate architecture. In the fourth part the selection process is detailed drawing the unusual sensibility of the quality criteria to the model's parameters initialization. An extensive study thus follows in order to examine and range the importance of each architecture element and training hyperparameter. Best models are then assessed on three intense events drawing the difficulty to automatically find one unique model for each lead-time and reliable for all flash flood events. A second approach is then proposed in section 5 based on a combination of linear and non-linear models in order to diminish the sensitivity to parameters' initialization. This approach was proved efficient. The resultant models were shown as useful to predict the water level up to 2-3 hours in advances. For this reason, a prototype implementing this approach is currently in test in SCHAPI before to be used for *vigicrues* map feeding.

Gardon d'Anduze Flash floods

Gardon d'Anduze Watershed

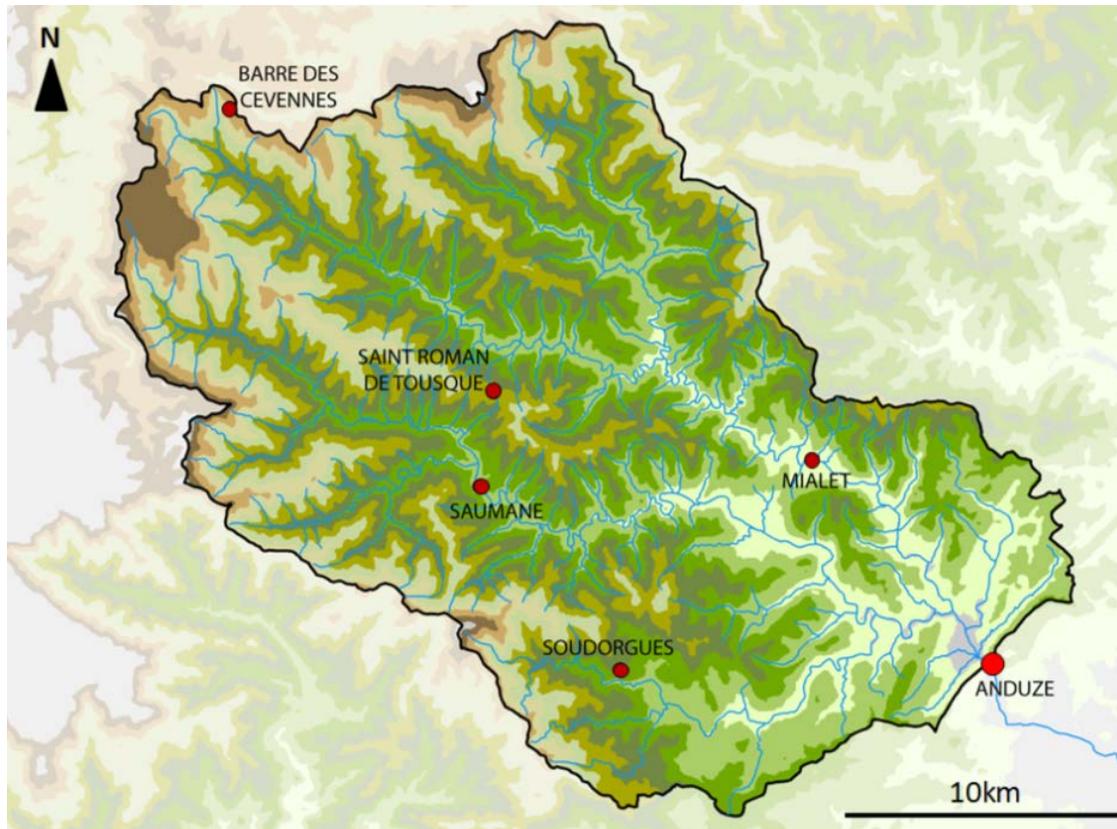


Fig. 1. Gardon d'Anduze watershed [Artigue 2010]

The *Gardon d'Anduze* watershed is a sub-watershed of the *Gardon*, a *Rhône* river tributary, located in the south-east of France. This basin has an area of 545km² and a mean slope of 10%: the highest altitude is 1060 m and the outlet altitude is 103 m. This important altitude gradient promotes a great velocity of the flows.

The *Gardon d'Anduze* watershed is regularly subjected to intense rainfall events, especially in autumn. For example during the storm of September 2002, the *Anduze* rain gauge recorded a cumulated rainfall of 650 mm in 20 hours (as comparison element the yearly mean cumulative rainfall registered in Paris is 660 mm).

Faced with this major risk, the French Ministry in charge of Sustainable Development (currently MEDDE) created in 2003 the national center for flood forecasting and warning SCHAPI (*Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations*) and the 22 regional forecasting services (*Services de Prévision des Crues*), which are in charge of the “vigicrues” surveillance service. The system is based on a 24 h watch of the main rivers and the production of warning messages towards the Civil Security and the citizens in the form of an Internet map (<http://www.vigicrues.ecologie.gouv.fr>). In order to provide accurate prediction of the *Gardon d'Anduze* water level to feed the *vigicrues* map in real time, SCHAPI was involve in the FLASH project [FLASH 2012], funded by the French

Agency of Research (ANR) which aims at improving real-time Flood prediction using machine learning. This work takes place in this project.

Due to these stakes, and thanks to the ancient and great work of monitoring performed by SCHAPI, and local SPC Grand Delta (Flood Forecasting Service for the downstream *Rhône* region), the *Gardon d'Anduze* basin is archetypal of *Cévenol* flash floods and is the field work of numerous studies [Europeans IMPRINTS, FLOODsite projects, ANR FLOODScale and Prediflood, ...].

Data base

The database used in this work includes 17 flash floods, which occurred between 1994 and 2008. The experimental measurements were supplied by six rain gauges and three gauges stations.

Currently, the rainfall collected by rain gauges and the water level are remotely transmitted every five minutes. Nevertheless in the database, the sampling periods depend on the events: 1 hour for the events from 1994 to 2002 and 5 minutes after 2002.

Table 1. Available database.

Number	Date	Duration (hours)	Mean Cumulated rainfall (mm)	Maximum water level (m)
1	September, 21-24 1994	35	189	3,71
3	October, 4-5, 1995	54	319	5,34
4	October, 13-14,1995	92	194	5
8	November, 10-12, 1996	82	157	2,71
107	November, 5-7, 1997	74	273	4,20
108	November, 26-27,1997	66	107	2,58
109	December, 18-19, 1997	104	376	5,37
117	October, 20-21,1999	34	117	3,64
13	September, 28-29, 2000	46	184	4,80
120	November, 12-14, 2000	71	125	2,77
19	September, 8-9, 2002	29	406	9,71
22	September, 24-25, 2006	23	61	2,24

23	October, 19-20, 2006	55	198	6,61
24	November, 17-18, 2006	34	59	2,75
25	November, 20-23, 2007	70	124	2,69
26	October, 21-23, 2008	43	173	5,57
27	31 st October – 3 rd November 2008	81	272	5,53

Analysis of database

Events of the database correspond to various flash floods situations. Their lengths vary from 23 hours to 101 hours, 7 events are shorter than 48 hours and 5 events longer than 72 hours. Peak of water level reached values between 2.77m and 9.71m. One can note that the most intense event reaches 9.71 m (event 19) while the second one reaches only 6.61m (event 23), *i.e.* the second most intense events amplitude reaches only 68% of the amplitude of the most intense one.

Rainfalls are temporally and spatially very different from one event to another and, consequently, the response of the river can be complex: several events have one peak while others have several peaks.

Also the initial humidity of the watershed is a predominant factor to the hydrological response; unhelpfully this initial state is very difficult to obtain physically. It is easy to understand that all this variability, making each event unique and added to the small number of events in the database (including 17 events and roughly 2000 vectors), exhausted the difficulty to simulate and forecast such flash floods.

In order to assess the proposed methodology on several configuration of flash flood, extensive simulations were carried out on three different events: two recent and very different events: the 26 which occurs on very dry soils and the 27 which received two times more water than the 26 delivering however a similar peak height. Moreover, as the goal of this work is to provide an operational tool able to predict intense flood to warn the population, it was mandatory to evaluate the quality of predictions on the most intense (and thus the most dangerous) event of the database: event 19. If the model can well generalize to the event 19, it will probably be efficient for rare or extreme future events.

Errors and uncertainty in measurement

Compared with errors or uncertainties occurring in artificial devices, these quantities are extremely increased when addressing natural measurements of rainfalls and discharge, or water level. Concerning the rainfalls occurring during flash flood in the studied area, they are very spatially heterogeneous. If one considers for example the cumulative rainfall taken into account by a rain gauge, it can be very different from a rain gauge to another: as illustration, during the event referred in the database as event 26, the rain gauge in *Mialet* recorded a cumulative rainfall of 479 mm while the rain gauge in *Barre-des-Cévennes*, recorded only 9 mm. *Mialet* and *Barre-des-Cévennes* are less than 30 km away.

Following [Obled] it can be assumed that 6 rain gauges correctly distributed in the *Gardon d'Anduze* watershed can represent the rainfall variability. Nevertheless, it is possible that the most part of the rainfall during one event can also take place between two rain gauges and was thus not measured at all. Moreover, the accuracy of the rain gauges themselves is a source of uncertainty: their accuracy is quantified between 10% and 20% [Marchandise].

In order to estimate the rainfalls with a better spatial resolution, meteorological radars were recently investigated. The spatial resolution of the radar is usually 1 km². However other errors get involved by radars [Jordan]: physical biases as ground clutter or blockage of the radar beam, also inaccurate relation between the reflectivity measured and the rainfall associated, and random sampling errors. These errors can be reduced, and consequently the radar rainfalls aren't necessarily representative of the real rainfall. Therefore they are usually calibrated using the rain gauge information. Moreover, as radar technology for meteorological application is a recent technology, signal processing was not constant during the time of interest and caused non stationarity on radar estimation of rainfall. Thus because of the youthfulness of this technology and the associated drawbacks, rainfall estimation coming from rain gauges was chosen in this study.

Regarding the rivers, water level was preferred to discharge as predicted quantity to free oneself of the rating curve which isn't actually known with sufficient reliability for high flows (during flooding it is dangerous to approach the river). But the water level measurement is also submitted to uncertainty sources, the bed of the river can change of shape during floods, and works (earthworks or constructions work) can modify the calibration of the gauge station. Uncertainty on flash flood discharge measurements is generally estimated around 20 to 30% [Marchandise]. For example in the case of the event 19 for which the measurements were estimated erroneous a posteriori, a hydrological study estimated the discharge in a range comprised between 2700 m³/s and 4200 m³/s (20% error).

Moreover, concerning events before 2002, the sampling period was one hour, which was not sufficient (the flood peak is underestimated for these events). Currently, data acquisition is performed at the rate of 5mn, and the database sampling period was chosen to 30 minutes considering operational constraints of SCHAPI and SPC.

Hydrological modeling

Several hydrological models were designed for modeling the discharge-rainfall relation on the *Gardon d'Anduze* watershed, three of them are TOPMODEL, SCS and MARINE. TOPMODEL (TOPography based hydrological MODEL), whose first version was developed by Beven and Kirkby [beven], is a physically based, distributed watershed model. This rainfall-runoff model was realized thanks to two principles: the runoff happens on contributive saturated areas, which ones are influenced by the topography. In order to insert the geographical variability of watershed and rainfalls (which are consequently usually called spatialized rainfalls), the watershed was meshed [Saulnier]. It was thus shown that the predominant information to consider was the spatialized rainfalls.

SCS (Soil Conservation Services) [Bouvier] is a conceptual model inspired by the tank concept. Storage of water is thus represented by a tank. A spatialized version where the radar rainfalls were considered with meshes of 1 km² was designed. Only two parameters were

calculated: the soil transmissivity and the tank capacity, both parameters were the same for all meshes. This model gave interesting results in watershed simulation (at our knowledge, no publication was available with prediction results on the *Gardon d'Anduze* with this models) as regard to its simplicity.

MARINE represents the physical phenomena at different scales: digital elevation model (DEM or topographic representation of the watershed), land use (urban, fields, forests...), soil texture (which gives information about its infiltrability), description of the river beds (roughness and geometry). The predominant process taken into account is the surface runoff. Using MARINE on the Gardon d'Anduze, [Braud] showed equally that the predominant information to take into consideration was the spatially variation of rainfalls.

It can be synthetize that these models can provide reliable hydrographs in simulation mode (without forecasts). They are essential to better comprehend the physical process of flash flooding. They undermine the importance of spatial distribution of the rainfalls on the underlying process and the necessity to estimate the soil moisture during the event.

Design of neural model for flash flood prediction

The multilayer perceptron model was preferred in this study because of its two well-known interesting properties: the universal approximation [Hornik] and the parsimony [Barron]. In this study, the standard model was first implemented. It contains one hidden layer of neurons with hyperbolic tangent as activation function and one linear output neuron. The least squares cost function was optimized with respect to the parameters by Levenberg Marquardt optimization.

In real-time flood forecasting, real time data are known up to time k (k is the discrete time), and the water level is intended to be predicted at several steps in the future. The model must thus compute the water level at time $k+h_p$ (h_p is the lead-time or horizon of prediction).

The available variables are the rainfalls and the water level measured in *Anduze*, at previous times. Rainfalls were applied through a sliding temporal window, w_t . In order to take into account the spatial variability of rainfalls, rain is fed using the six rain gauges available on the basin. As future rainfalls measurements are unavailable, the output of the neural model is the prediction of the water level in *Anduze* without using prediction of rainfalls.

As the rainfall-discharge relation is dynamic, two kinds of models can be designed [Nerrand]. Denoting by $y^p(k)$ the observed value of the process (here, the water level) at discrete time k , $y(k, \mathbf{w})$ the forecast value of the process (water level) at discrete time k calculated with the set of parameters \mathbf{w} , by $\mathbf{u}(k)$ the vector of exogenous variables (here the rainfalls) at time k , h_p the horizon of prediction, and by g_m the nonlinear function implemented by the neural network, the following input-output predictors can be designed:

(i) recurrent model

$$y(k+h_p) = g_m(y(k, \mathbf{w}), \dots, y(k-r+1, \mathbf{w}), \mathbf{u}(k), \dots, \mathbf{u}(k-w_t+1)),$$

(ii) feedforward model

$$y(k+h_p) = g_m(y^p(k), \dots, y^p(k-r+1), \mathbf{u}(k), \dots, \mathbf{u}(k-w_t+1), \mathbf{w}),$$

r and w_t are integers adjusted during the selection of the model.

It has been shown that the first kind of model is optimal if the process is subjected to output noise (for example measurement noise); while the second type of model is optimal if the process is subjected to state noise (here, noise on rainfalls or lack of information about soil moisture). As highlighted by sections 2.3 and 2.5, the uncertainty concerning rainfalls and soil moisture is predominant for the physical process; the second type of predictor will thus be preferred. Moreover others studies on a similar watershed show that the recurrent model is less efficient than the feed-forward one [Artigue]. Such conclusion is otherwise well known in hydrological modeling by neural networks.

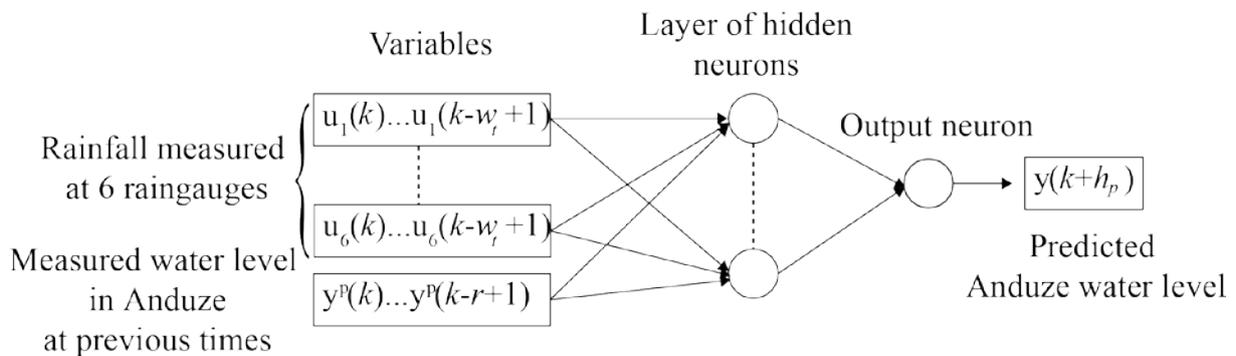


Fig. 2. Feedforward predictor designed for rainfall-predicted runoff relation.

Since the future rainfalls are unknown, one can consider that model can possibly perform rainfalls anticipation whose behavior depends on the lead-time. Thus a specific model was adjusted for each horizon of prediction: 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours and 5 hours. Generic model shown in Fig.2 was then used in this study.

Complexity selection

An important issue of the multilayer perceptron is the bias-variance dilemma [Geman]. During the training stage, overfitting can hampers the generalization of the model, and must be avoided. In previous works on the same basin, operating however with a smaller database (without recent events 26 and 27), it was shown [Toukourou 2009, 2011] that early stopping [Sjöberg] was more efficient than weight decay as regularization method for flash flood prediction. Nevertheless using early stopping made mandatory to set apart two different datasets: training set and stopping set (usually termed as validation set), which unhelpfully diminishes the size of training set.

The selection of the model is realized thanks to a variant of the cross-validation method [Stone], the partial cross-validation: each event of the training set corresponds to a subset K of K -fold cross-validation. Nevertheless in order to promote a model with good performances upon intense events, only specific events (the four most intense events of the training set)

were used to proceed to the selection by cross-validation. The validation score was calculated based on mean squared error and the model with the smallest validation score was chosen. Partial cross-validation score (S_{pcv}) allowed thus to select two components of the architecture: the number of hidden neurons (h) and the length of temporal sliding rainfall windows (w_t). Both contribute to increase the number of parameters of the models and consequently its complexity.

The choice of the stopping set is commonly cited as a difficulty; especially when the database is too limited for random sampling as it is in this study. It was thus proposed to use as stopping set the event which is the best predicted. Indeed if this event is well predicted, this means that it is the best representative subset of the training set. According to this thinking, the event 13 was chosen as stop set.

In order to assess the generalization performances on intense events, training set was composed by all the events of the database except four intense events. As previously indicated the event 13 was used as stopping set, moreover, events 19, 26, 27 were affected to the test set in order to evaluate the ability of models to generalize to events fast, intense, and for the event 19, most intense event of the database.

Application to the Gardon d'Anduze flash flood prediction

As indicated in previous section, the complexity selection intends to determine the exact optimal architecture regarding the partial cross-validation score (S_{pcv}). This means determining successively (all other possible variability being fixed constant) the number of hidden neurons and the number of variables. Such methodology was previously applied in the work of [Toukourou 2011], nevertheless the variability over parameters initialization was not sufficiently studied in this previous work; we will show in the present study that it has an unusual major importance. In order to be exhaustive, hyperparameters of Levenberg Marquardt algorithm which has a role during the convergence process were also adjusted using the same methodology.

Complexity selection

Number of hidden neurons

The number of hidden neurons determines the complexity of the model. If the number of hidden neurons is too low, the model isn't able to account for the complexity of the phenomenon. But otherwise, a number of hidden neuron too high could diminish the performance of the model.

The number of hidden neurons, h , was varied from one to seven. The temporal windows, w_t , of the rainfalls recorded at the 6 rain gauges and the hyperparameters of the training were set constant. 200 initializations of the network parameters were carried out for each value of h . The number of 200 was obtained by experience, it guaranty that a good minimum of the cost function can be reached. The same work was repeated for the various horizons of prediction.

The initialization of the network parameters was randomly generated with a normal distribution with an average equal to zero and a standard deviation equal to $\sqrt{\frac{1}{v+1}}$, v being the number of exogenous variables in order to prevent sigmoid saturation.

For each of the 200 initializations, the partial cross-validation score (S_{pcv}) was computed. Several descriptors were necessary in order to compare validation scores of the various numbers of hidden neurons: for each h , (from one to seven), the average, the median, the standard deviation, the minimum and the maximum of the 200 validation scores (each one for a particular initialization) were calculated. Based on these descriptors the selection of the number of hidden neurons must be done. One can see as illustration the evolution of these scores for the specific model designed for the prediction horizon of one hour (Fig. 3)

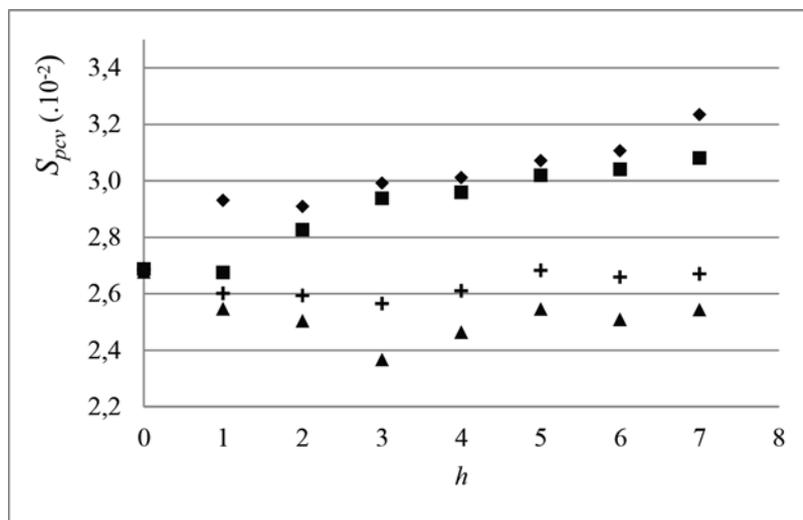


Fig. 3. Evolution of statistical descriptors versus the number of hidden neurons for $h_p=1h$. Average (rhomb), median (square), minimum of the partial cross-validation score on the 200 initializations (triangle), and average on the 20 best S_{pcv} (cross), versus the number of hidden neurons for a horizon of prediction of one hour. Zero hidden neuron corresponds to the linear model: a network with only one linear neuron. The maximum of the partial cross-validation score over the 200 initializations can't be visualized on the same figure than the others criteria because its value is ten times superior to the others. In this particular case the maximum reaches $S_{pcv}=0.11$ for $h=7$.

One can note on the Fig.3 that the best number of hidden neurons is three. Another point to notice is that few abnormal elements, having very high cross-validation score (the maximum can reach ten times the value of the average), affect greatly the computation of the average. Therefore the method is not robust and the calculation of the median could thus be a better estimator of the expectation.

The curve of the median has for minimum one hidden neuron; it promotes models with low complexity.

Considering the curve of the minimum of validation score, it tends to decrease when the number of hidden neurons increases which means that a moderate complexity allows finding more accurate models. Nevertheless, remembering that the database is sparse and that each event is very different one from the others, it is possible to consider that the minimum wouldn't be a good indicator too: it depends entirely of one particularly accurate initialization and isn't a reliable estimator of the quality of the model faced, in real time, to an event different from all the training events. A compromise seems thus to be necessary. Thereby, the average of the 20 best partial cross-validation scores on the 200 initializations (denoted as Average10% on the Figs.) was proposed to achieve the selection task. The number of hidden neurons corresponding to the minimum of the curve associated to this descriptor is thus chosen.

All figures for all horizons of prediction are not produced in the paper because of the huge number of figures that this will generate. For each experience only one figure is thus selected. Considering the selection of number of hidden neurons, one can generally stay that when the horizon of prediction increases, the gap grows between the S_{pcv} for the linear model and the non-linear one: using of neural networks was thus justified.

To quantify the dispersion of the results and thus the risk to select a model unable to deliver good predictions on a new unknown event (in real time), two dispersion criteria were introduced and calculated regarding the partial cross validation score:

$$S_1 = (\text{Maximum-Minimum})/\text{Average}$$

$$S_2 = \text{Standard Deviation}/\text{Average}.$$

These values are reported in Table 2 for all forecasting horizons.

Table 2.

Statistical description, for all number of hidden neurons (one to seven), of Average10% curve: Average, Minimum, Maximum, Standard Deviation, S_1 and S_2 for each horizon of prediction.

Average10%	$h_p=0,5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
Average (10^{-2})	1.56	2.63	3.90	4.86	5.67	6.69
Min (10^{-2})	1.53	2.57	3.83	4.81	5.63	6.59
Max (10^{-2})	1.59	2.68	3.96	4.92	5.73	6.80
Stand. Dev. (10^{-4})	2.13	4.44	4.39	3.60	4.32	9.06
S_1 (10^{-2})	4.1	4.5	3.4	2.3	1.8	3.1
S_2 (10^{-2})	1.4	1.7	1.1	7.4	7.6	1.4

It is noticeable on the Table 2 that the values of the Average10% are not very dispersed, the choice of the number of hidden neurons doesn't seem to be predominant for the performances of the model. This observation can be explained by the utilization of early stopping which decreases the impact of overfitting.

Moreover one can note that the average of the score increases when the forecasting horizon increases with is coherent with the fact that the function to perform is more difficult to find without future rainfalls and future observed water levels (for example $y^p(k+4)$ is not available for prediction of $y^p(k+5)$). Minimum and maximum evolve in the same range of values while standard deviation increases less than the average score. S_1 and S_2 seem roughly invariant to prediction horizon.

Concerning the variation of the partial cross validation score for each forecasting horizon, and for the best associated number of hidden neurons, the Table 3 presents the values of the statistical descriptors presented herebefore.

Table 3.

For each horizon of prediction and with the chosen number of hidden neurons (h): Average, Minimum, Maximum, Standard Deviation, S_1 and S_2 of the S_{pcv} , over 200 initializations of parameters.

S_{pcv}	$h_p=0.5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
h	2	3	2	3	4	3
Average (10^{-2})	1.86	2.99	4.36	5.37	6.14	7.15
Min (10^{-2})	1.49	2.37	3.63	4.55	5.40	6.33
Max (10^{-2})	14.5	4.58	5.55	7.52	8.65	8.38
Stand. Dev. (10^{-4})	101	34	36	40	38	36
S_1 (10^{-2})	700	74	44	55	53	29
S_2 (10^{-2})	54	11	8.3	7.4	6.3	5.1

One can first notice on Table 3 that the number of hidden neurons is not really correlated to the horizon of prediction. This is understandable considering that when the horizon of prediction increases, it is obvious that the difficulty to forecast the water level increases because of the lack of information about future rainfalls and not because of a more complex underlying function.

Also, the S_1 and S_2 criteria are ten to a hundred times higher than in the precedent Table: the various initializations of the network parameters create an important dispersion of the values of S_{pcv} . Between the better score and the worse one there is a great gap, as drawn by scores S_1 S_2 specially for short term prediction. For example considering the 0.5 hours prediction, the average is very close to the minimum and the standard deviation is great which implies that there were few very bad initializations. Unusually, parameters initialization is a factor more critical than the number hidden neurons.

Variable selection: width of the rainfall temporal windows

The width of the temporal windows for the rain gauges was varied from 30 minutes to 5h30. In order to diminish the computational complexity of the various adjustments to perform, the

width was chosen equal for all the rain gauges. The number of hidden neurons and the hyperparameters of the training step were set constant. As for the number of hidden neurons, for each horizon of prediction, 200 initializations of the network parameters were carried out for each value of w_t (window width); for each initialization the partial cross-validation score was computed. The same statistical descriptors than previously were calculated and the Average10% was used to choose the best w_t for each horizon of prediction. Numerical values are reported in Table 4.

Table 4. Description, for all width of the temporal window (one to six), of Average10% curve: Average, Minimum, Maximum, Standard Deviation, S_1 and S_2 for each horizon of prediction.

Average10%	$h_p=0.5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
Average (10^{-2})	1.54	2.62	3.91	4.86	5.78	6.75
Min (10^{-2})	1.52	2.56	3.85	4.78	5.66	6.61
Max (10^{-2})	1.57	2.66	3.95	4.92	5.91	6.84
Stand. Dev. (10^{-4})	1.38	3.44	2.99	4.04	9.04	7.44
S_1 (10^{-2})	3.3	3.6	2.6	2.8	4.4	3.4
S_2 (10^{-2})	8.9	1.3	7.7	8.3	1.6	1.1

The values of S_1 and S_2 are similar to the ones calculated for the Average10% for the number of hidden neurons. The influences of the variation of the width w_t and of h are in the same range of value on the cross-validation score.

Table 5.

For each horizon of prediction and with the chosen width of temporal windows (w_t): Average, Minimum, Maximum, Standard Deviation, S_1 and S_2 of the S_{pcv} , over 200 initializations of parameters.

S_{pcv}	$h_p=0.5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
w_t (h)	3	2.5	2.5	1.5	0.5	0.5
Average (10^{-2})	1.77	2.99	4.39	5.57	6.13	7.23
Min (10^{-2})	1.49	2.41	3.68	4.60	5.53	6.45
Max (10^{-2})	7.18	7.26	14.7	9.76	7.52	16.9
Stand. Dev. (10^{-4})	0.47	0.45	0.89	0.55	0.35	0.01
S_1 (10^{-2})	320	160	250	93	32	14
S_2 (10^{-2})	27	15	20	9.9	5.8	15

When considering the evolutions of statistical descriptors versus w_t , it can be noticed in the Table 5 that, logically, the width of the temporal windows decreases when the horizon of prediction increases. On the contrary of the number of hidden neurons, the evolution of the width of the temporal window is interpretable: larger is the horizon of prediction, smaller is the width of rainfall history. This “rule” is understandable considering the time delay that the water has to join the outlet. If the prediction horizon is temporally too far, the rainfall efficient for this horizon of prediction is not yet fallen, or just fallen, the width is thus minimal.

As it was regarding the choice of hidden neurons number, the values of the S_{pcv} for the choice of temporal windows depend principally on the initialization of the network parameters.

Role of hyperparameters of Levenberg-Marquard algorithm

The same demarche than previously was followed in this section in order to define the couple of initial training step (μ_{ini}), and training step decrease factor (μ_{dec}), used in Levenberg Marquardt algorithm for each horizon of prediction. The initial training step varies from 10^{-4} to 10, every power of ten (6 possibilities), and the training step factor varies from 0.1 to 0.9 every 0.1 (9 possibilities). The couple of initial training step, and training step decrease factor, was then chosen among 54 possibilities.

The number of hidden neurons and the width of the temporal window step are set constant to the previously selected values. As before, for each horizon of prediction, 200 initializations of the network parameters are carried out for each hyperparameters couple. For each initialization a partial cross-validation score is computed. The same statistical descriptors are calculated and the Average10% is used to choose the best couple for each horizon of prediction. Numerical results are synthesized in Table 6 and 7.

Table 6.

Description, for all couple of training hyperparameters, of Average10% curve: Average, Minimum, Maximum, Standard Deviation, S_1 and S_2 for each horizon of prediction.

Average10%	$h_p=0.5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
Average (10^{-2})	1.52	2.56	3.90	4.90	5.66	6.75
Min (10^{-2})	1.50	2.51	3.80	4.79	5.58	6.59
Max (10^{-2})	1.57	2.62	4.08	4.97	5.73	6.82
Stand. Dev. (10^{-4})	1.45	2.84	7.64	4.07	3.37	5.23
S_1 (10^{-2})	4.5	4.3	7.0	3.7	2.6	3.4
S_2 (10^{-2})	9.6	11	20	8.3	6.0	7.8

Table 7.

For each horizon of prediction and with the chosen couple of training hyperparameters (μ_{ini} ; μ_{dec}): Average, Minimum, Maximum, Standard Deviation, S_1 and S_2 of the S_{pcv} , over 200 initialization of parameters.

S_{pcv}	$h_p=0.5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
$(\mu_{ini}; \mu_{dec})$	$(10^{-3}; 0.2)$	$(10^{-1}; 0.4)$	$(10^{-3}; 0.3)$	$(10^{-3}; 0.9)$	$(10^{-3}; 0.9)$	$(10^{-4}; 0.9)$
Average (10^{-2})	1.67	2.95	4.32	5.31	6.12	7.15
Min (10^{-2})	1.46	2.32	3.69	4.64	5.34	6.33
Max (10^{-2})	5.13	6.21	6.68	7.71	7.79	8.38
Stand. Dev. (10^{-4})	29	45	41	40	37	37
S_1 (10^{-2})	220	130	69	58	40	21
S_2 (10^{-2})	18	15	9.4	7.8	6	5

One can note that, compared to the other factors, the couple of hyperparameters has a lower effect on the dispersion of the cross-validation score. This phenomenon is conform to the usual knowledge. Nevertheless, as for the number of hidden neurons and the temporal window, the Average10% shows that the S_{pcv} are not very dispersed, so changing the values of the initial training step and the training step decrease factor hasn't great impact on the model performance. But once again the choice of the initialization has an importance on the S_{pcv} .

Given these results, because of the existence of local minima for the cost function, the initialization of the parameters takes an important part in the quality of the model. The important point to emphasize is that, unusually, the selection of initial parameters set is the most important factor to consider: all other factors are really less important.

Selection of initialization

After the selection of the architecture and of the training hyperparameters, the best initialization must be chosen for each horizon of prediction. To this end we proceed in the same way than for the complexity selection. 200 initialization were proceeded with the selected configuration found in section 4.1 to 4.2 and synthetized in Table 8. The partial cross-validation score was calculated once again. Statistical descriptors are presented in Table 9.

Table 8.

Synthesis or selected elements and hyperparameters

h_p (hour)	0.5	1	2	3	4	5
$(\mu_{ini}; \mu_{dec})$	$(10^{-3}; 0.2)$	$(10^{-1}; 0.4)$	$(10^{-3}; 0.3)$	$(10^{-3}; 0.9)$	$(10^{-3}; 0.9)$	$(10^{-4}; 0.9)$
w_t (hour)	3	2.5	2.5	1.5	0.5	0.5
h	2	3	2	3	4	3

It can be noticed that the variability of the partial cross-correlation score with the selected architecture elements and hyperparameters was already presented in Table 7 for the 200 initializations. Indeed this table presents the dispersion of score at the end of the selection process when only parameters initialization varies. It is still apparent that a great heterogeneity remains as the S_2 criterion which calculates Standard Deviation divided by the Average is the same order than the Average itself ($h_p = 2-5$ hours) or 10 times greater ($h_p = 0.5-1$ hour).

In order to dispose of a set of models able to provide forecast, the best 20 initializations were chosen as it was done during the selection process. Table 7 presents the statistical descriptors of this set of “best” models for each prediction horizon. It can be noticed that descriptors take better scores than in Table 7 which is absolutely normal.

Table 9. Statistical descriptors for the best 20 initializations, for each prediction horizon.

S_{pcv}	$h_p=0.5h$	$h_p=1h$	$h_p=2h$	$h_p=3h$	$h_p=4h$	$h_p=5h$
Average (10^{-2})	1.60	2.69	4.06	5.04	5.92	6.92
Min (10^{-2})	1.49	2.35	3.81	4.68	5.34	6.43
Max (10^{-2})	1.83	3.48	4.50	5.54	6.42	7.63
Stand. Dev. (10^{-4})	11	27	17	23	3	32
S_1 (10^{-2})	2	41	17	17	18	17
S_2 (10^{-2})	6.92	10	4.40	4.58	5.09	4.57

Test on independent events

Following the complexity selection of the model and the adjustment of the best couple of hyperparameters for each horizon of prediction, it is necessary to evaluate the performance of the models. To this end several quality criteria must be defined.

Quality criteria

Four performance criteria were calculated: the classical root mean square error (RMSE), the R^2 criterion (C_N), the persistency criterion (C_p) and the peak weighted root mean squared error (PWRMSE). Mathematical description of all criteria is following.

$$C_N = 1 - \frac{\sum_{k=1}^n (y_k^p - y_k)^2}{\sum_{k=1}^n (y_k^p - \bar{y}^p)^2}$$

y_k^p Output measured at time k

y_k Output computed at time k

\bar{y}^p Average of the measured output

The R^2 criterion, or Nash coefficient [Nash], commonly used by the hydrologists, is the coefficient of determination. The most the C_N is nearer than 1, the best are the results. Moreover, it is also well known that a curve translated in time of few time steps can have a

good R^2 whereas it doesn't provide prediction accurate in time. To avoid this problem, the persistence criterion and the peak weighted root mean squared error are also calculated.

The persistence criterion [Kitadinis] provides information on the prediction capability of the model compared to the naive forecast. The naive forecast postulates that the value of the process at time $k+h_p$ is the same than the value at time k . If C_p is positive, then the model prediction is better than the naive prediction. If C_p is negative that means that the model does not provide any information. As for the R^2 , the most the C_p is nearer than 1, the best are the results.

C_p is defined as:

$$C_p = 1 - \frac{\sum_{k=1}^n (y_{k+h_p}^p - y_{k+h_p})^2}{\sum_{k=1}^n (y_{k+h_p}^p - y_k^p)^2}$$

$y_{k+h_p}^p$ Output measured at time $k+h_p$

y_{k+h_p} Output forecast at time $k+h_p$

y_k^p Output measured at time k

Nevertheless, when calculating the R^2 , or C_p , only an average performance on the event is obtained while it is easy to understand that the peak of flood has a crucial importance.

For this reason the peak weighted root mean squared error [Amengual] was proposed to enhance errors on the peak values. To this end, it weights the squared error with a ratio which increases with the value of the studied quantity. This ratio is thus higher for the peak values giving more importance to the error at the peak. The definition is the following:

$$PWRMSE = \left[\frac{1}{2n} \sum_{k=1}^n \left[(y_k^p - y_k)^2 \left(1 + \frac{y_k^p}{\bar{y}^p} \right) \right] \right]^{\frac{1}{2}}$$

y_k^p Output measured at time k

y_k Output estimated at time k

\bar{y}^p Average of the measured output

Results in test

As indicated in the introduction it is mandatory to assess the quality of the model on test sets independent from both processes of training and stopping and also independent from model selection. Performances criteria are thus calculated on events set apart from database at the beginning of the study (see section 3.1.1.). After model selection, training is performed using all the training set without partial cross-validation. The best 20 initializations were set apart to study the variability of responses in test situation depending on the various models resulting from various initializations.

Three intense events were set apart for the test: 19, 26 and 27. They were chosen because they are all intense and very different events (see section 2.3 and 2.4). The test stage proceeds as if the model was operating in real time in a causal way: the future of the event is unknown.

The first question is thus, as we have shown that the selection of the model depends greatly on the initialization, what was the variability observed in test phase in relation with the arbitrary choice realized in the model selection stage?

Faced with this question, we report in Fig. 4. 20 hydrographs obtained with the 20 better models after the selection process (corresponding to the 20 better parameters initialization thanks to the S_{pcv} score). Unhappily, it is shown that the set of parameters, chosen by cross-validation, can bring models very different, one of them totally unrealistic, with water level of 18m, the double of the measured water level!

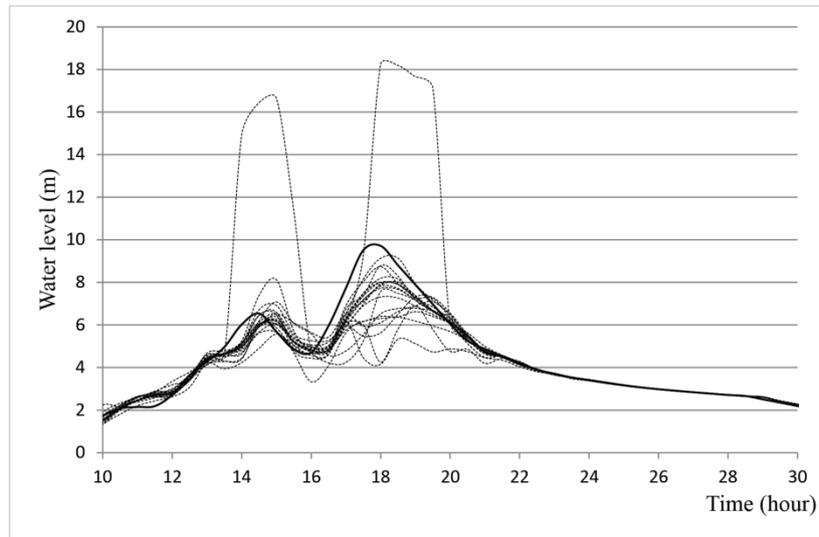


Fig. 4. Observed limnograph (solid line) and predictions over event 19 for an horizon of prediction 0.5h, with the 20 various initializations (dotted lines) of the network parameters chosen thanks to the partial cross-validation.

On another hand, three initializations allow getting almost 90% of the peak amplitude and are well synchronized. The other initializations lead to models forecasting quite well the peak amplitude.

In order to better quantify the quality of these tests hydrographs, quality criteria are reported in Fig. 5. for the event 19.

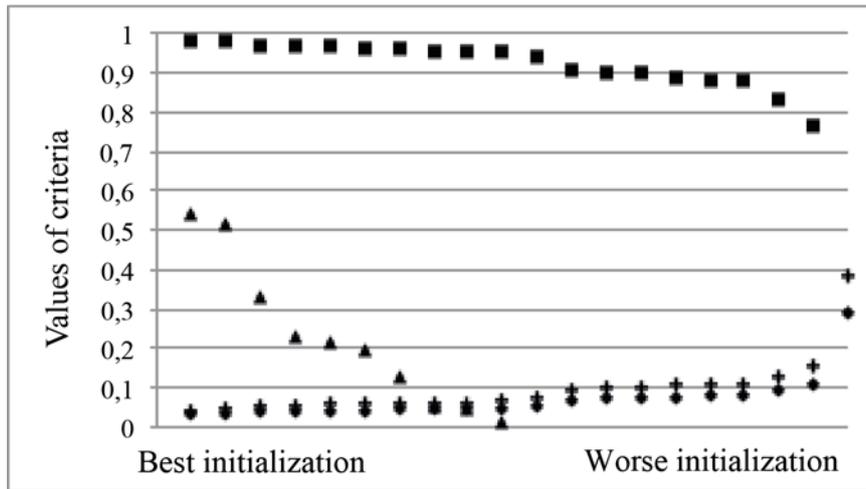


Fig. 5. Quality criteria for the event 19 in test (for $h_p=0.5h$). RMSE (rhomb), R^2 (square), Persistency (triangle) and PWRMSE (rhomb). 20 initializations of the network's parameters are ranked from the best to the worse. Persistency criterion becomes negative for half of initialization (worse value -37), they are not represented in the Fig. The same apply for the R^2 , which is -0.7 for the worse initialization.

The values of the performance criteria confirm the comments on the Fig 4. Three initializations give a good R^2 , none a good persistence. The quality of the persistence decreases rapidly and is even negative for half part of initializations. Theses scores can be examined in the light of the Table 10 where the partial cross-validation scores of these models are given. It is also possible to notice that the PWRMSE criterion do not provide any interest because it varies exactly as the quadratic error, this criteria will not be used in the following.

This huge diversity of solutions is clearly unacceptable. We propose then to use in real time forecasting the median of these 20 “best” initializations. For each time step, and in all this section, the forecast value of the water level is thus the median of the 20 values given by the selected initializations.

Table 10: Values of the R^2 and the persistence criteria for the three test events at different horizons of prediction.

Event	19			26			27		
	$R^2_{naïve}$	R^2	C_p	$R^2_{naïve}$	R^2	C_p	$R^2_{naïve}$	R^2	C_p
0.5h	0.95	0.96	0.00	0.94	0.96	0.31	0.99	~ 1	0.53
1h	0.84	0.86	0.08	0.82	0.86	0.23	0.97	0.98	0.48
2h	0.57	0.78	0.46	0.52	0.62	0.17	0.89	0.95	0.52
3h	0.34	0.78	0.64	0.29	0.56	0.33	0.78	0.90	0.49
4h	0.07	0.73	0.69	0.10	0.55	0.44	0.66	0.85	0.52
5h	-0.33	0.56	0.62	-0.18	0.35	0.36	0.51	0.80	0.54

Table 10 presents values of the used criteria: R^2 and persistency for each event in test and for the 6th horizons of prediction. It also contains the value of the R^2 criterion for the naive prediction (remember the naive forecast value is equal at $k+h_p$ to the observed value at time k). One can note that generally the R^2 criterion of the naive forecast is very high for short horizons of prediction (namely 0.5h and 1h), after that it decreases rapidly. The last observation indicates that first it is very difficult to perform better prediction than the naive one at short horizons, and second that consequently the stake is to obtain better forecasts than the naive ones at farthest horizons than 1h. The quality of the solution can be evaluated thanks to the persistency criterion.

Taking these considerations into account it appeared immediately that the forecasted value obtained with the median at short horizons is not very good: it is slightly better than the naive one. The persistence is always positive which is satisfying: models are more efficient than the naive prediction. Equally the quality of the forecasts decreases with the h_p . Nevertheless the persistency criterion increases with h_p meaning that the model is more efficient when the horizon increases: larger is the horizon, worst is the naive prediction, it is thus “easier” to provide better model.

Regarding the difference between the three tested events, one can note that the event 26 is the more difficult to deal with (possibly because of the heterogeneity of rainfalls and very dry soils as suggested in the following). Event 27 is easier to deal with because it is long and slower, it lasts about twice more than both the others, the rise of the peak is also slower, and thus the model can better adjust with the previous water level. Curiously, the event 19 appeared as easier to forecast than the 26. This is due to the simplicity of the event: a huge localized rainfall implies a huge discharge response.

Hydrographs for the three events of interest are provided in Fig.6.

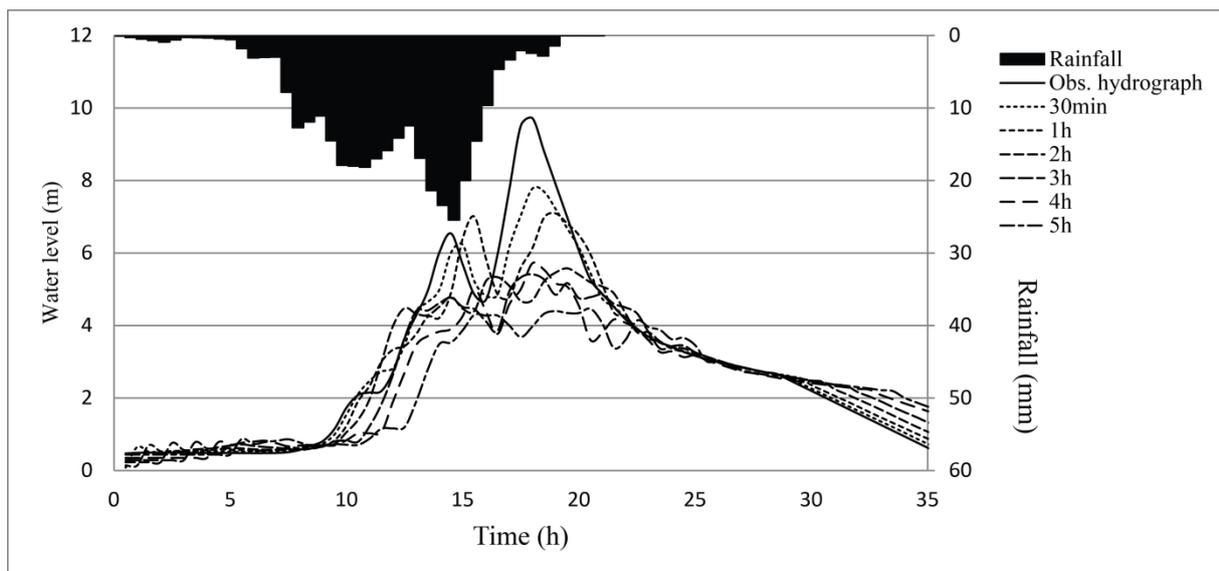


Fig. 6-a.

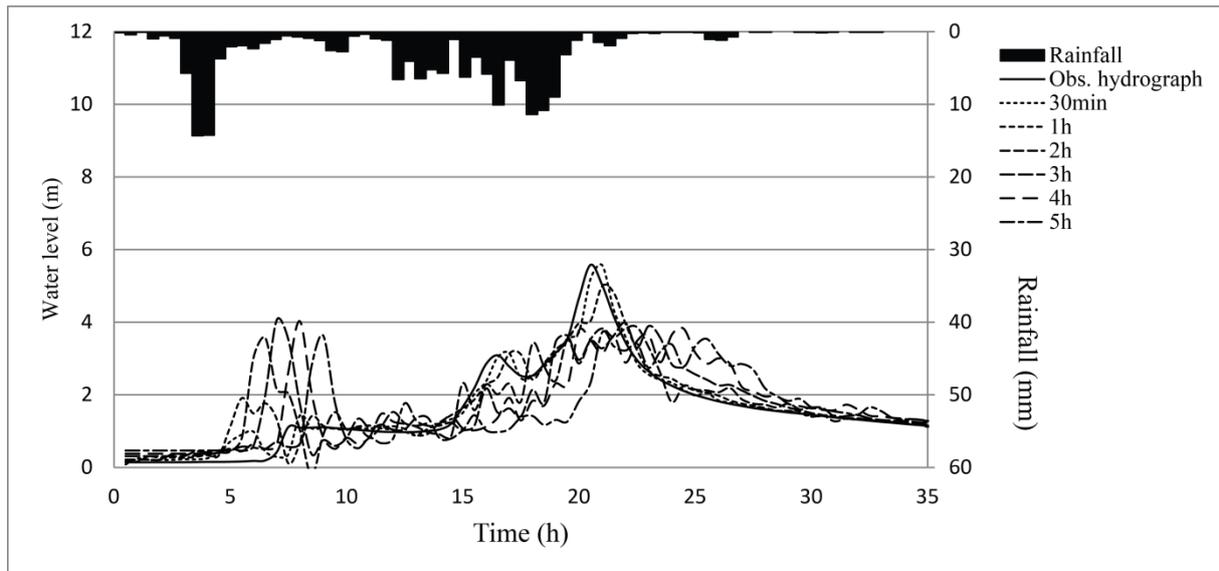


Fig. 6-b.

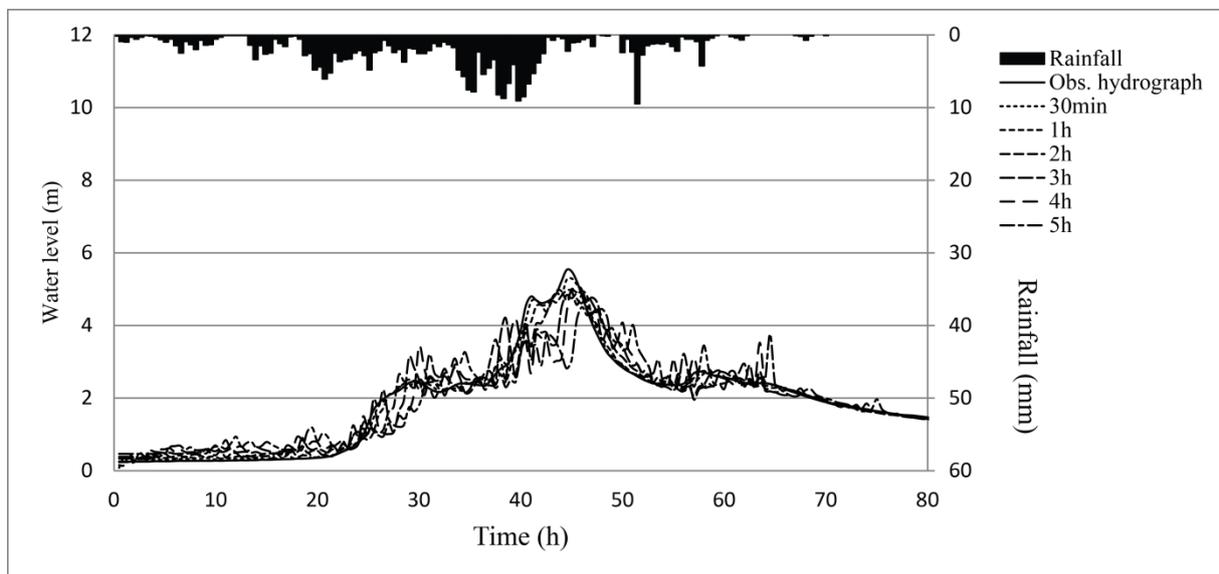


Fig6-c.

Fig.6. Hydrographs at *Anduze* for the event 19 (6-a), 26 (6-b), 27(6-c), for the different horizons of prediction from 30 minutes to 5 hours. Same scale is chosen for the three hydrographs in order to better apprehend the difference of intensity of rainfalls and water level. Drawn rainfalls are the mean rainfall on the basin (average on the 6 rain gauges), in order to simplify the graph.

Regarding the event 19, it first appears that this event is extraordinary intense and fast: the slope of the second peak is 5 meters in 1.5 hours! One can easily understand that without anticipation on rainfalls it is difficult to predict the peak longer than 2 hours lead-time because the time separating the peak of rainfall and the peak of water level is only roughly 2 hours. Predicting further than 2 hours needs thus to predict water level while the rainfall is not fallen.

Moreover the shown hydrograph would possibly be quite inaccurate because, as mentioned in section 2.4, instrumentation did not work reliably during the event. For a horizon of prediction of 30 minutes, the model reaches 80% of the peak amplitude and is synchronized with the observed peak; for 1 hour the model reaches 70% of the peak amplitude, but only 55% for 2, 3 hours and beyond. Only the forecast at 30 min horizon would be useful because others are too weak or too delayed. These delays are explainable: the network needs the rising of previous observed water level to enhance its hydrograph; it can't anticipate with only the rainfalls (which lacks of recent values).

Regarding the event 26 which is classified as intense even if it is two times weaker than the 19, forecasts are very bad. Actually this event is very difficult to forecast because rainfalls arrived on dry soils in a very inhomogeneous distribution: two records very intense on *Soudorgue* and *Mialet* rain gauges (80 mm in one hour, quasi nothing elsewhere) followed by a small increase of water at *Anduze* 4 hours later. Then the second peak arrived consequently to 150 mm fallen in 3 hours only in *Mialet* and *Anduze* rain gauges. It is thus very probable that a non-negligible amount of water fall outside of the rain gauges network, explaining by this way that models always underestimate the water level.

Forecasts for 30 minutes are satisfactory and the model reaches 70% of the peak amplitude for 2 and 3 hours. There are many oscillations because of the rainfall impulses (previously cited). A wrong peak appears as answer to the important rainfall impulse at the beginning of the event. This wrong peak is "false positive" response and could provide wrong warning. Nevertheless the experience of forecasters can filter this kind of bad response; they know that soils are dry at the beginning of this event, and that no specific variable provides this information to the network.

Regarding the event 27, the model works quite better, it reaches more than 90% of the peak amplitude for the 30 min and 1 hour lead-times and it is synchronized with the observed water level. For 2 and 3 hours, the peak amplitude is synchronously reached at more than 85%. Inopportune oscillations are due to the great time-variability of rainfalls.

In conclusion to this analysis of the quality of predictions in real-time test situation, it can be synthesized that the great variability of possible solutions, due to an excessive sensibility to the parameters initialization, breaks up the reliability of the method. We think that the intense level of noise and uncertainties in the variables, as rainfalls than water levels, hamper the model for converging toward the physical solution during training. The absence of future rainfalls forecasts would not be the preponderant defect because the heterogeneity in solutions can be observed even at the smallest prediction horizon (Fig 4). In order to overcome this drawback, several ways can be investigated: (i) find another method better appropriate for initialize parameters, (ii) make the model adaptive for select in real time the better initialization, (iii) take profit of the ability of neural networks to integrate in a unique formalism linear and nonlinear solutions. Indeed, concerning the third way, as linear minimization converges always toward a unique solution, one can reasonably think that linear and nonlinear joint models would take profit on advantages of both types of modeling: linear for the uniqueness of the solution, and nonlinear for the quality of nonlinear solutions.

Joint linear and nonlinear models

Model design

This part presents the design and results obtained in extension of the previous study, with the same goals and database. Subsets for training stop and test are identical; methodology is the same.

The template of the architecture is shown in Fig 7. It contains a nonlinear multilayer perceptron added with linear connections linking directly variables to the linear output neuron. During training both linear and nonlinear functions are trained simultaneously.

Regarding the variability previously observed in models selected by cross validation, the idea is using the linear part of the behavior to diminish this variability, and to complement the linear function adding a nonlinear contribution. Traditionally these kinds of approach prove difficulties to train efficiently the nonlinear part of the function. It will thus be interesting focusing on: (i) the variability of solutions selected by cross validation, (ii) the quality of obtained predictions, (ii) the relative contributions of linear and nonlinear behaviors.

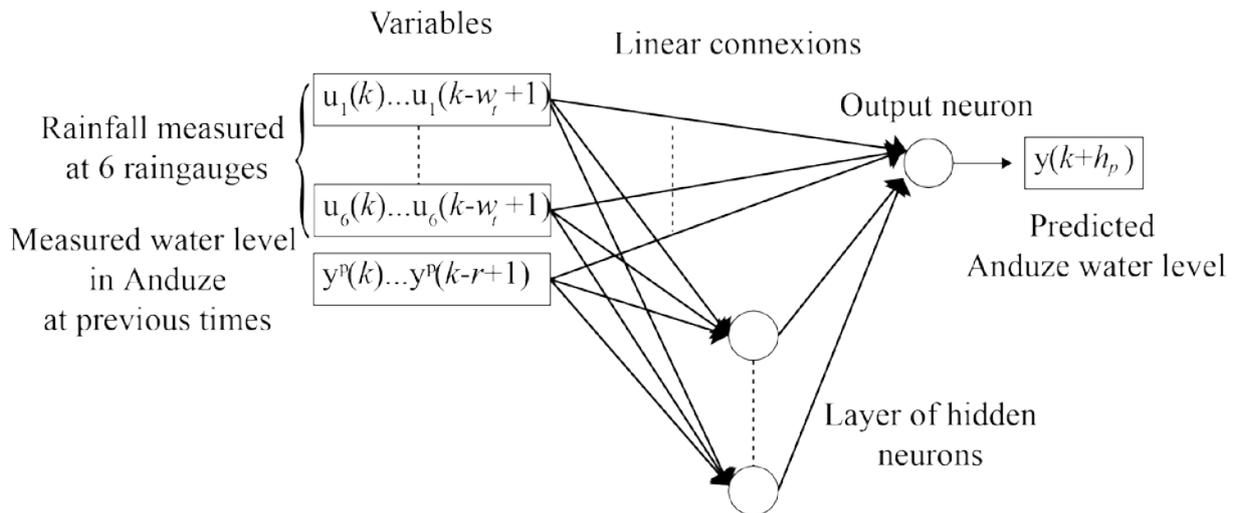


Fig.7. Feedforward linear and nonlinear joint model for water level prediction

As it was noticeable in sections 4.1-4.3 that the preponderant factor to take into account during the selection process was the initialization of parameters, only this factor was selected in this second stage. For each prediction horizon, architecture is thus the one selected previously (synthetized in Table 8). Only selection of initialization will be investigated in the following.

Initialization selection

Starting with the architecture selected in part 4, we first look at the variability of the partial cross-validation score using the same statistical descriptors than previously. It appeared in Table 11 that the evolution of the S_{pcv} score is more understandable using the linear contribution: its values and Standard Deviation increase with the horizon of prediction, as well as roughly the criteria S_1 and S_2 . Particularly for the horizon 0.5h and 1h, the dispersion of S_{pcv} is quasi null (Standard Deviation of 10^{-5}).

Table 11. Variability of partial cross-validation scores for each prediction horizon, for the 10 best initializations and the whole 200 initializations.

S_{pcv}		Average (10^{-2})	Min (10^{-2})	Max (10^{-2})	Stand. Dev. (10^{-4})	S_1 (10^{-2})	S_2 (10^{-3})
$h_p = 0.5$ h	20 best init.	1,01	1,01	1,01	$3 \cdot 10^{-3}$	0	0.03
	200 init.	1,01	1,01	1,01	$2 \cdot 10^{-3}$	0	0.02
$h_p = 1$ h	20 best init.	3,63	3,63	3,63	$58 \cdot 10^{-3}$	0	0.002
	200 init.	3,63	3,63	3,63	$63 \cdot 10^{-3}$	0	0.002
$h_p = 2$ h	20 best init.	10,25	9,84	10,41	15	4	15
	200 init.	11,15	9,84	13,02	58	28	53
$h_p = 3$ h	20 best init.	16	16	16	20	2	7.47
	200 init.	17	16	19	71	20	41.62
$h_p = 4$ h	20 best init.	21	21	21	23	4	10
	200 init.	23	20	35	190	63	83
$h_p = 5$ h	20 best init.	27	24	29	125	17	46
	200 init.	31	24	37	196	40	62

In order to compare the current architecture with the previous one, the median of the 20 best initializations is calculated, and the values of quality criteria are presented in Table 12 (to be compared with Table 10). It appeared that the quality of the joint model is greatly improved specifically for the persistency criterion with is thus always positive.

Event	19		26		27	
	R^2	C_p	R^2	C_p	R^2	C_p
0.5h	0.98	0.54	0.96	0.32	~ 1	0.82
1h	0.91	0.49	0.86	0.23	0.99	0.69
2h	0.87	0.73	0.69	0.32	0.94	0.48
3h	0.86	0.81	0.58	0.36	0.90	0.54
4h	0.77	0.78	0.55	0.42	0.87	0.62
5h	0.43	0.60	0.28	0.26	0.75	0.50

Table 12. Values of the R^2 and the persistence criteria for the three test events at different horizons of prediction.

Better appreciation of the quality must be proposed using the hydrographs. It appeared in Fig. 8-a, 8-b, 8-c that hydrographs are also really improved and provide reliable forecasts up to 3 hours for the events 19 and 27. Only at one hour for the event 26. For this last event, we think

that it would be necessary to dispose of information about the soil moisture, which is unfortunately unavailable in the database.

Regarding the event 19 which is really well predicted three hours in advance, this model will be very useful to warn the population and prevent casualties. For this reason the designed model is currently in test in the SCHAPI Service.

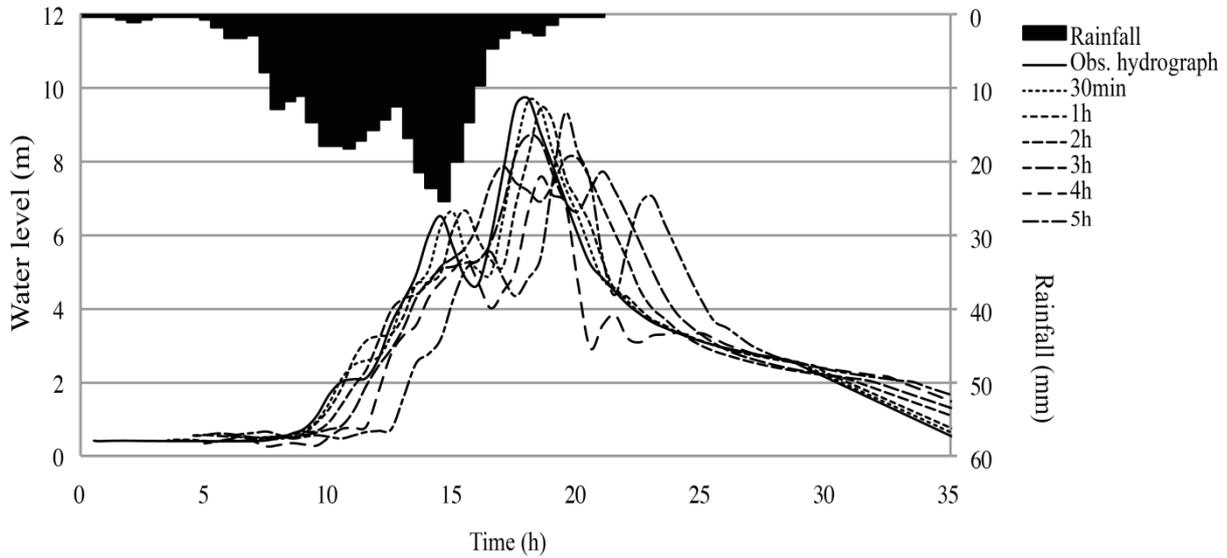


Fig. 8-a.

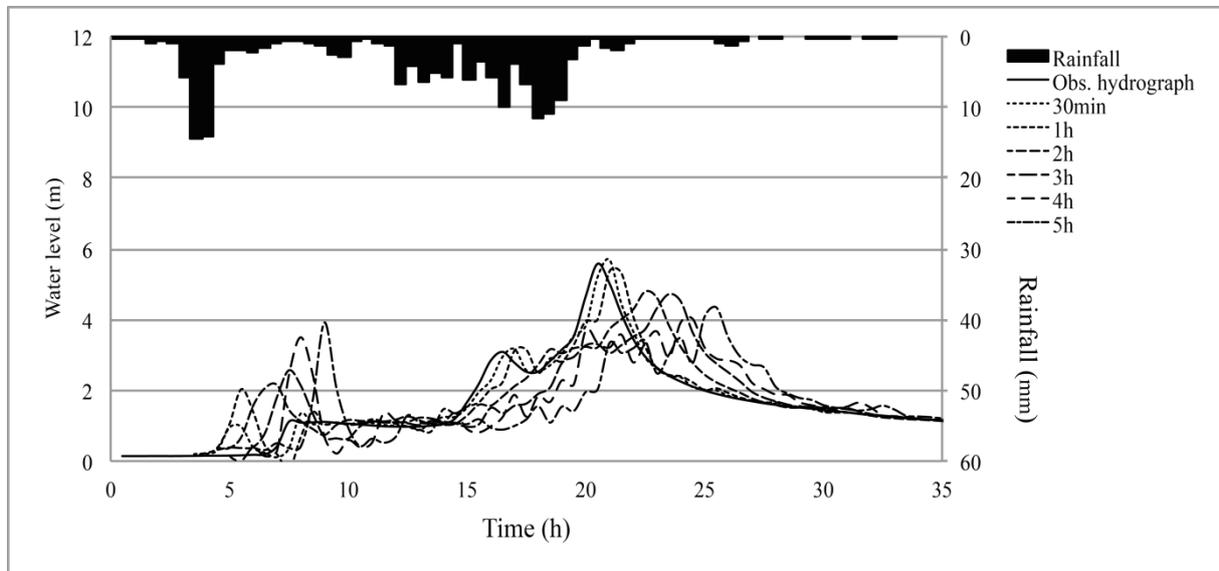


Fig. 8-b.

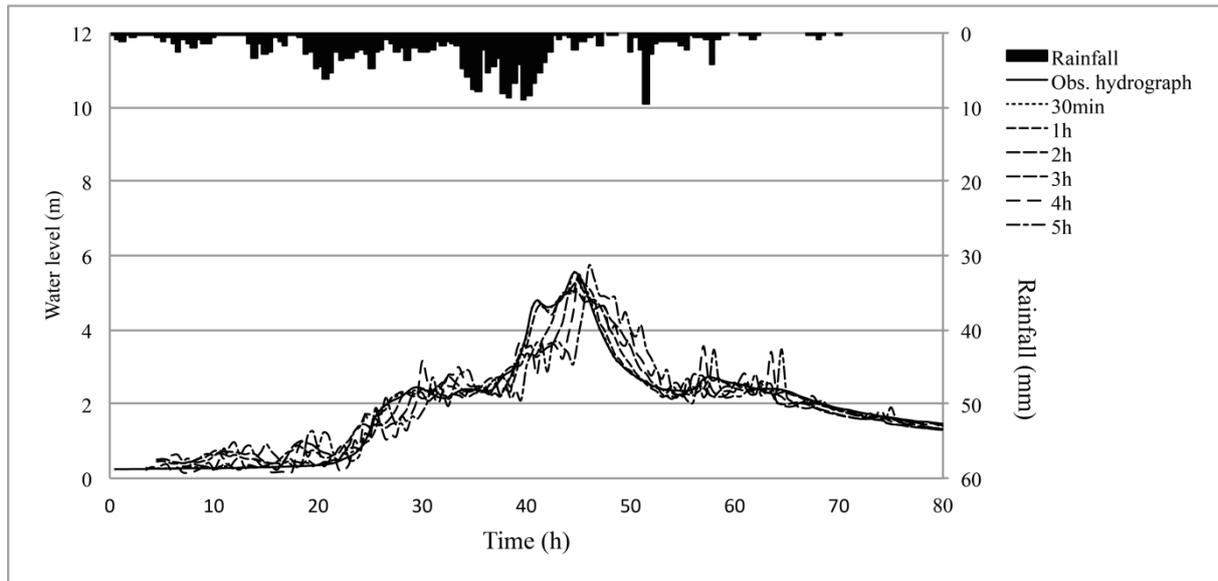


Fig. 8-c.

Fig. 8. Hydrographs at *Anduze* for the event 19 (8-a), 26 (8-b), 27(8-c), for the different horizons of prediction from 30 minutes to 5 hours. Same scale is chosen for the three hydrographs in order to better apprehend the difference of intensity of rainfalls and water level. Drawn rainfalls are the mean rainfall on the basin (average on the 6 rain gauges), in order to simplify the graph.

Conclusion

Flash flood forecasting without rainfall forecast is a difficult task because of the complex hydrometeorological phenomena, the errors and uncertainties of the measurements and also the small number of events of the data base. In this paper, the modeling is based on neural networks instead of hydrological model commonly used.

The partial cross-validation is used as selection method to determine the model design, the number of variables and to adjust some training parameters. One can note that surprisingly, the complexity of the model isn't essential in this study. On the contrary, the initialization of the networks parameters emerges as a determinant factor concerning the quality of the model performances. In addition, because of the heterogeneity of the solutions proposed by the model, in relation with the importance of initialization and the diversity of hydrometeorological events, it was shown difficult to choose automatically the best solution. The method thus provides insufficient predictions. In order to compensate the role of the initialization, another model including directly linear connections was then introduced in order to reduce the importance of the parameters initialization. It was shown that this architecture, novel in hydrology, provides significantly improved results delivering reliable forecast 3 hours ahead. The models presented in this paper are currently implemented for the SCHAPI French Flood Surveillance Service to provide the *vigicrue* website.

References

[Abrahart] R. J. Abrahart, L. M. See, Neural network modelling of non-linear hydrological relationships, *Hydrology and Earth System Sciences*, 11(2007) 1563–1579.

- [Alfieri 2012] L. Alfieri, P. Salamon, F. Pappenberger, F. Wetterhall, J. Thielen, “Operational early warning systems for water-related hazards in Europe”, *Environmental Science & Policy*, p. 35-49, 2012
- [Amengual] A. Amengual, R. Romero, M. Gomez, A. Martin, S. Alonso, A Hydrometeorological Modeling Study of a Flash-Flood Event over Catalonia, Spain, *Journal of hydrometeorology*, 8 (2007) 282-303.
- [Artigue] G. Artigue (NHES à compléter quand accepté)
- [Barron] A.R. Barron, Universal Approximation Bounds for Superpositions of a Sigmoidal Function, *IEEE Transactions on Information Theory*, IT-39 (1993) 930-945.
- [Bazartseren] B. Bazartseren, G. Hildebrandt, K.-P. Holz, Short-term water level prediction using neural networks and neuro-fuzzy approach, *Neurocomputing*, 55 (2003) 439 – 450.
- [Beven] K.J. Beven , M.J. Kirkby, A physical based, variable contributing areas model of basin hydrology, *Hydrological Sciences Bulletin*, 24 (1979) 43-69.
- [Braud] I. Braud, H. Roux, S. Anquetin, M.-M. Maubourguet, C. Manus, P. Viallet, D. Dartus, The use of distributed hydrological models for the Gard 2002 flash flood event: Analysis of associated hydrological processes, *Journal of Hydrology*, 394 1-2 (2010) 162-181.
- [Bouvier] C. Bouvier, P.A. Ayrat, P. Brunet, A. Marchandise, C. Martin, Recent advance in rainfall-runoff modelling : extrapolation to extreme floods in southern France, *International Workshop on Hydrological Externes, Rende 3-4* (2006).
- FLASH website: <http://edytem.univ-savoie.fr/spip/programmes-et-contrats-de/flash/>
- Vigicrue website: <http://www.vigicrues.gouv.fr/>
- [Gaume] E. Gaume, M. Livet, M. Desbordes, J.P. Villeneuve, Hydrological analysis of the river Aude, France, flash flood on 12 and 13 November 1999, *Journal of Hydrology*, 286 (2004) 135-154.
- [Geman] S. Geman, E. Bienenstock, R. Doursat, Neural Networks and the Bias/Variance Dilemma, *Neural Computation*, 4 (1992) 1-58.
- [Hornik] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*, 2 (1989) 359-366.
- [Jordan] P.W. Jordan, A.W. Seed, P.E. Weinmann, A stochastic model of radar measurement errors in rainfall accumulations at catchment scale. *Journal of Hydrometeorology*, 4 (2003) 841–855.
- [Ju] Q. Ju, Z. Yu, Z. Hao, G. Ou, J. Zhao, D. Liu, Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model, *Neurocomputing*, 72 (2009) 2873–2883.
- [Kitadinis] P. K. Kitadinis, R. Bras, Real-Time forecasting with a conceptual hydrologic model, applications and results, *Water Resources Research*, 16 (1980) 1034-1044.
- [Nerrand] O. Nerrand, D. Urbani, P. Roussel-Ragot, L. Personnaz, G. Dreyfus, Training recurrent neural networks: why and how? An illustration in process modeling. *IEEE Trans Neural Netw* 5 (1994) 178–184.

- [Kong A Siou 2011] Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France), L. Kong A Siou, A. Johannet, S. Pistre, V. Borrel, *Journal of Hydrology*, 403 3-4 (2011) 367-380.
- [Kong A Siou 2012] L. Kong A Siou, A. Johannet, S. Pistre, V. Borrel Estupina, Optimization of the generalization capability for rainfall-runoff modeling by neural networks: the case of the Lez aquifer (southern France), *Environmental Earth Sciences*, 65 8 (2012) 2365-2375.
- [Le lay] M. Le Lay, G.M. Saulnier, Exploring the signature of climate and landscape spatial variabilities in flash flood events: Case of the 8-9 September 2002 Cévennes-Vivarais catastrophic event, *Geophysical Research Letters*, 34 (2007) L13401.
- [Marchandise] A. Marchandise, Modélisation hydrologique distribuée sur le Gardon d'Anduze ; étude comparative de différents modèles pluie-débit, extrapolation de la normale à l'extrême et test d'hypothèses sur les processus hydrologiques, University of Montpellier II (France), Phd (2007).
- [Nash] J. E. Nash, J. V. Sutcliffe, River Flow Forecasting through Conceptual Models. Part I – A Discussion of Principles, *Journal of Hydrology*, 10 (1970) 282-290.
- [Obled] C. Obled, I. Zin, B. Hingray, Choix des pas de temps et d'espace pour des modélisations parcimonieuses en hydrologie des crues, *la Houille Blanche*. 5 (2009) 81-87.
- [Saulnier] G.M. Saulnier, M. Le Lay, Sensitivity of flash flood simulations on the volume, the intensity and the localization in the Cévennes-Vivarais region (France), *Water Resources Research*, (2009) 45.
- [Sjöberg] J. Sjöberg, L. Ljung, Overtraining regularization, and searching for minimum in neural networks, Preprint IFAC Symposium on Adaptive Systems in Control and Signal Processing (1992).
- [Stone] M. Stone, Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Society*, B36 (1974) 111-147.
- [Toth] E.Toth, A. Brath, A. Montanari, Comparison of short-term rainfall prediction models for real-time flood forecasting, *Journal of Hydrology*, 239 (2000) 132–147.
- [Toukourou 2009] M. Toukourou, A. Johannet, G. Dreyfus, Flash Flood Forecasting by Statistical Learning in the Absence of Rainfall Forecast: A Case Study, Engineering applications of neural networks, Book Series: Communications in Computer and Information Science 43 (2009) 98-107.
- [Toukourou 2011] M. Toukourou, A. Johannet, G. Dreyfus, P. A. Ayrat, Rainfall-runoff Modeling of Flash Floods in the Absence of Rainfall Forecasts: the Case of "Cévenol Flash Floods", *Applied Intelligence*, 35 2 (2011)178-189.