Tâche 2 : Chaine hydrométéorologique

Anne Johannet, Pierre-Alain Ayral, Bernard Vayssade, Audrey Bornancin-Plantier, Yann Visserot

Laboratoire LGEI, ARMINES

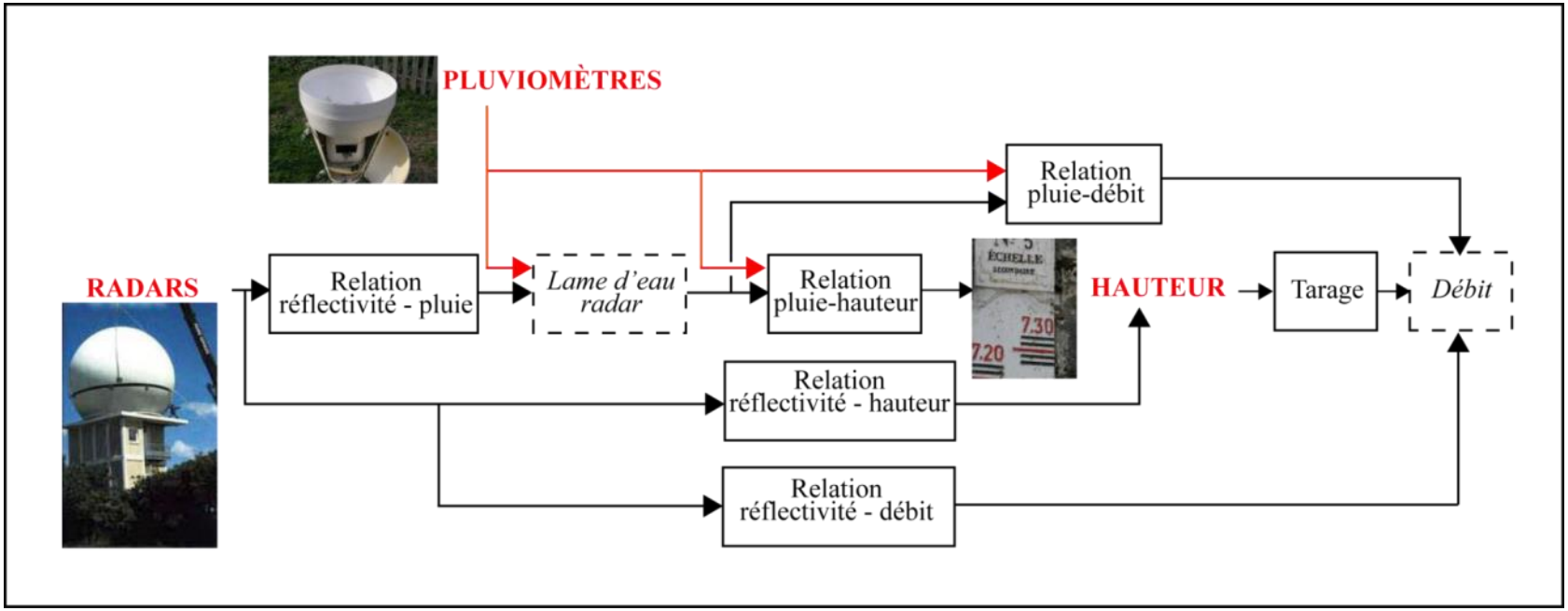
Pierre Roussel Ragot, Gérard Dreyfus

Laboratoire SIGMA, ESPCI ParisTech

Première partie, la relation pluie-hauteur à Anduze et à Remoulins

Thèse de Audrey Bornancin-Plantier

Modélisation systémique de la chaine hydrométéorologique

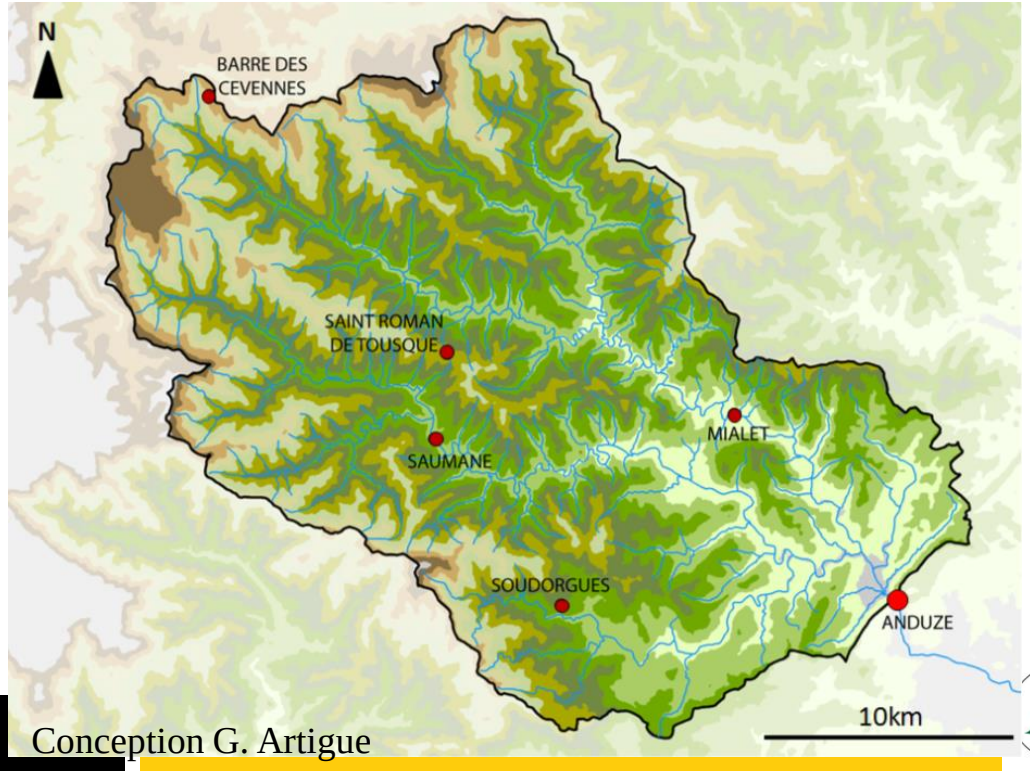


Présentation des données et difficultés de modélisation

- **Bassin versant du Gardon d'Anduze**
  - Superficie =545 km²
  - Altitude max = 1060m,,min = 130m

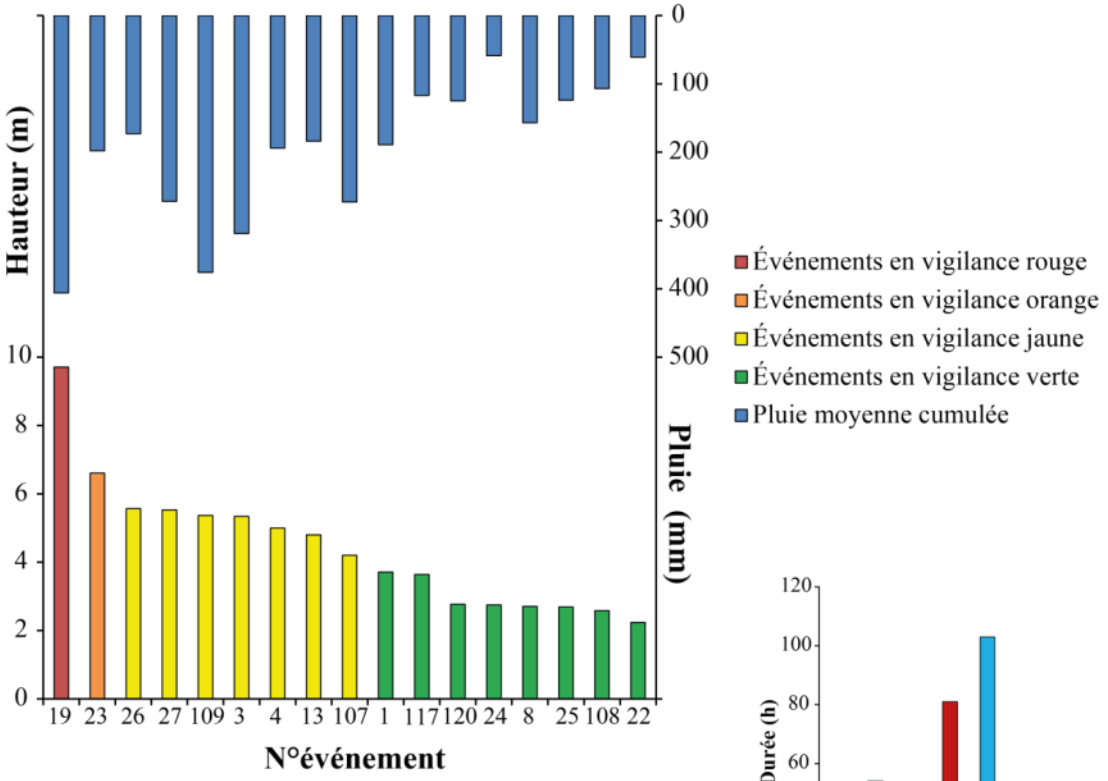
- **Erreurs et incertitudes sur les données**
  - 7 pluviomètres
  - 4 limnimètres

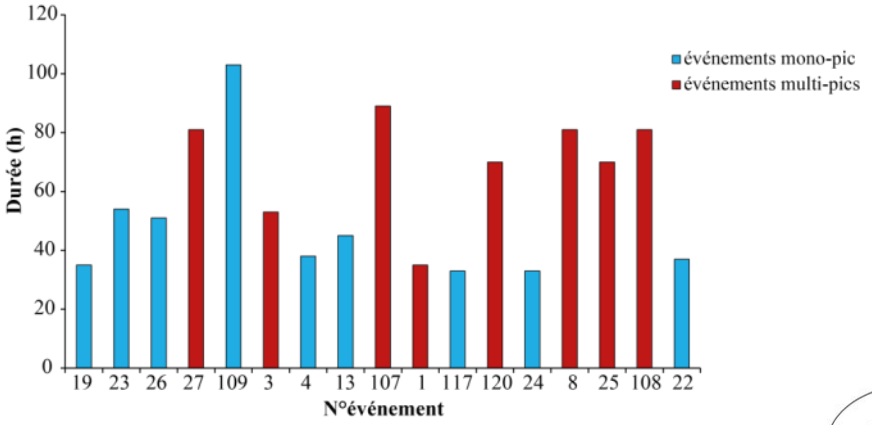
- **Base de données**
  - 17 événements



Conception G. Artigue

17 événements de crue





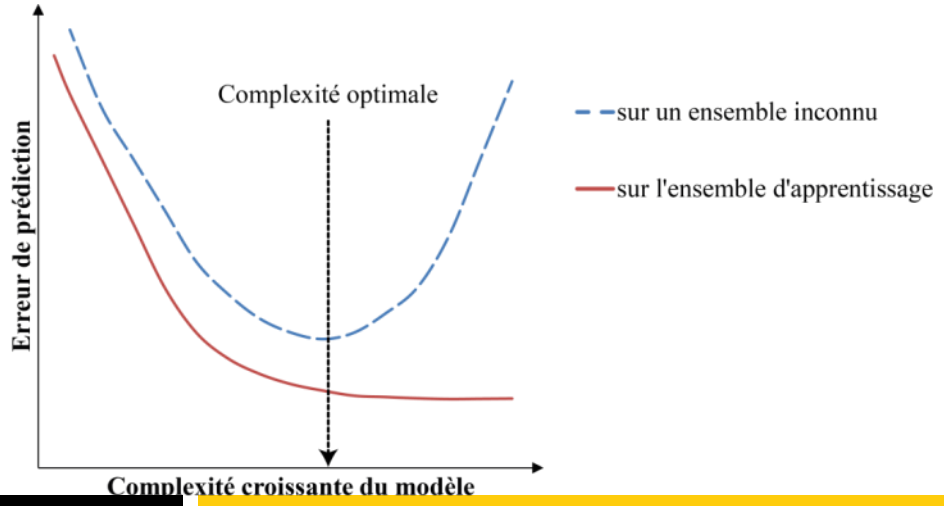
Importance de la généralisation

- **Objectif des modèles utilisant l'apprentissage statistique**

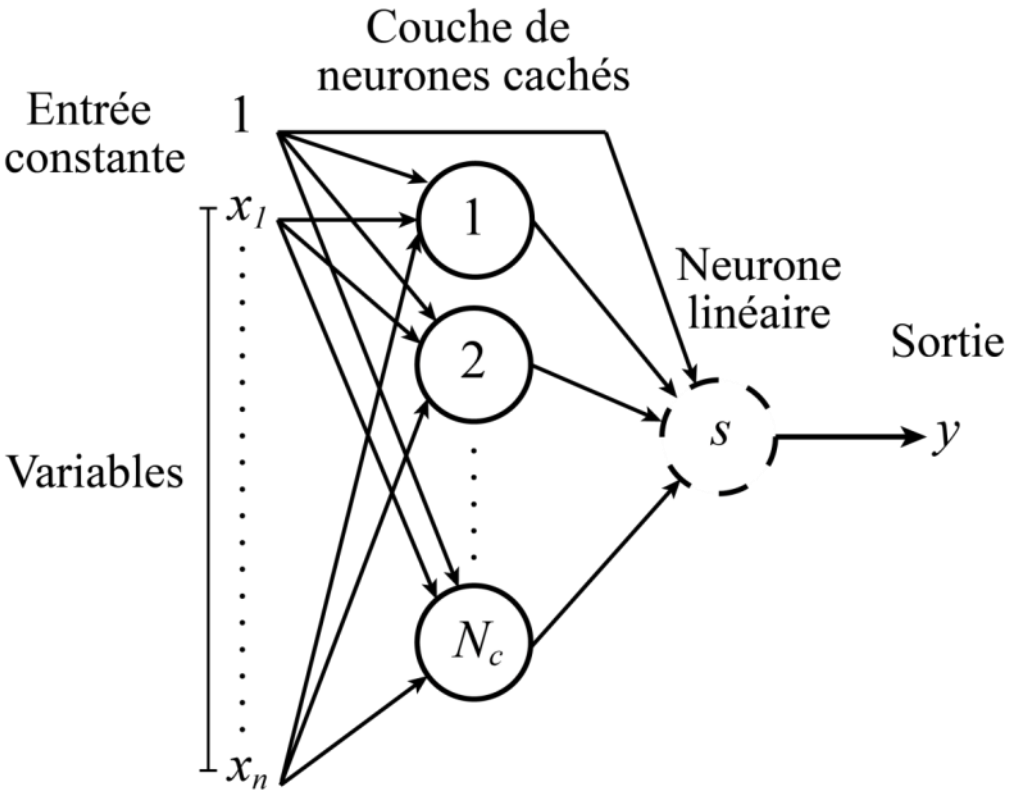
  → bonne *généralisation* à une situation inconnue

- **Dilemme biais-variance [Geman 1992]**

  → trouver une complexité optimale pour un compromis entre le biais et la variance



Le perceptron multicouche et ses propriétés fondamentales



- ✓ **Approximation universelle [Hornik 1989]**
- ✓ **Parcimonie [Barron 1993]**

Étapes de la création d'un prédicteur neuronal

- ✓ **Prétraiter les données**
- ✓ **Sélectionner les variables pertinentes**
- ✓ **Réaliser l'apprentissage sur des modèles d'architectures différentes**
- ✓ **Sélectionner le meilleur modèle**
- ✓ **Évaluer la performance de ce modèle sur ensemble de test**

Évaluation sur l'ensemble de test de la pertinence des modèles sélectionnés : critères de performances

- ✓ Root Mean Squared Error

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

- ✓ Peak Weighted Root Mean Squared Error [Amengual 2007]

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

y_k^p Sortie mesurée au temps k
y_k Sortie calculée au temps k
$\overline{y^p}$ Moyenne de la sortie mesurée

$y_{k+h_p}^p$ Sortie mesurée au temps $k+h_p$
y_{k+h_p} Sortie calculée au temps $k+h_p$

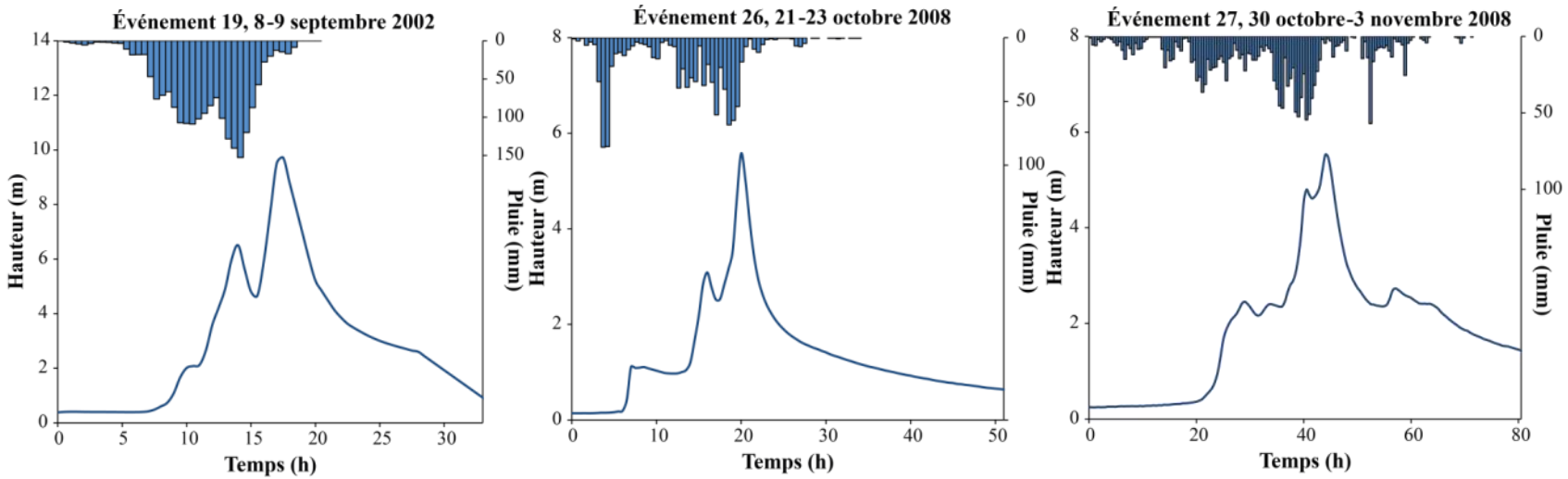
- ✓ Nash [Nash 1970]

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

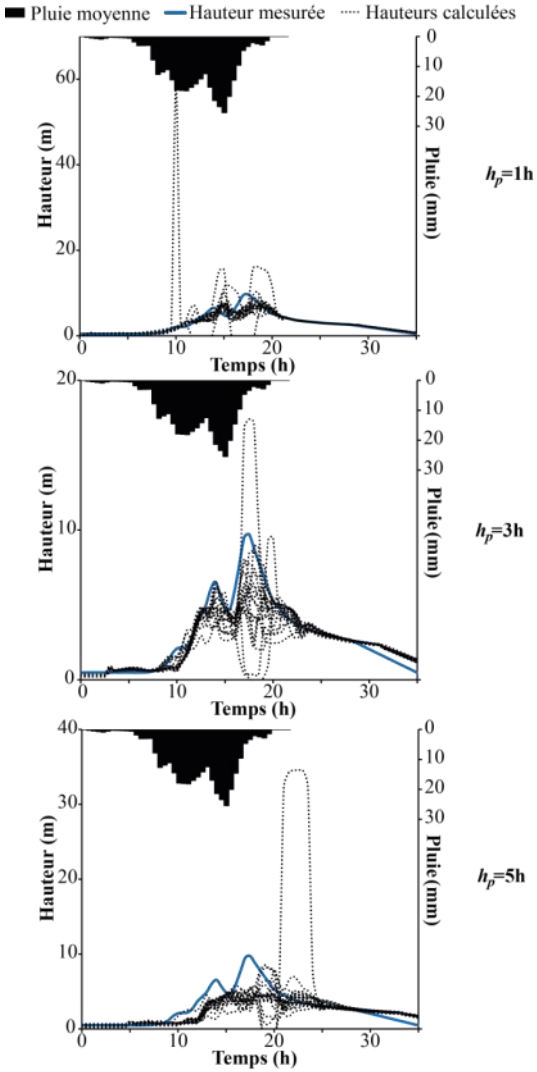
- ✓ Persistance [Kitadinis 1980]

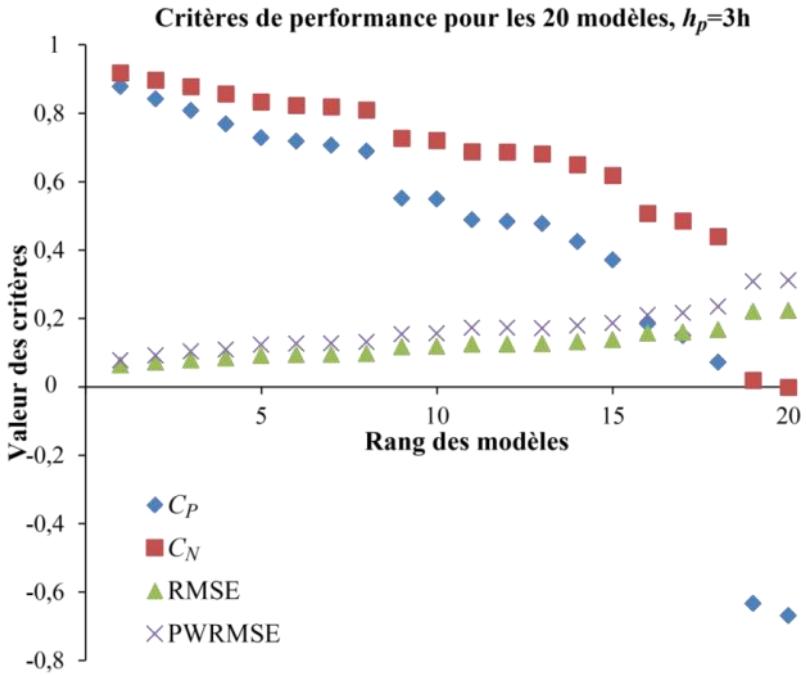
$$C_p = 1 - \frac{\sum_{k=1}^{n}\left(y_{k+h_p}^p - y_{k+h_p}\right)^2}{\sum_{k=1}^{n}\left(y_{k+h_p}^p - y_k^p\right)^2}$$

Événements en test



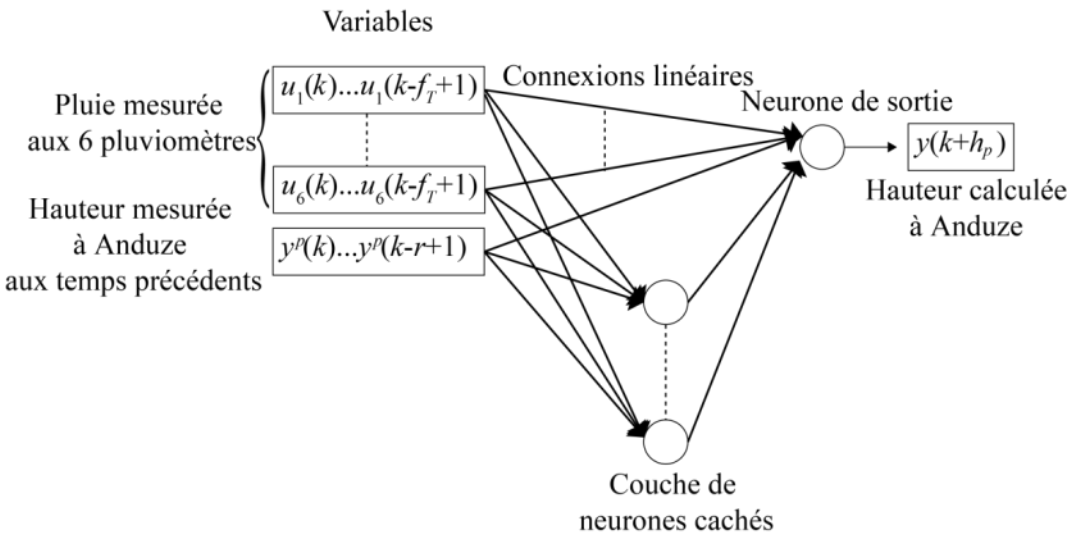
Hauteurs prévues pour l'événement 19 par les 20 modèles sélectionnés





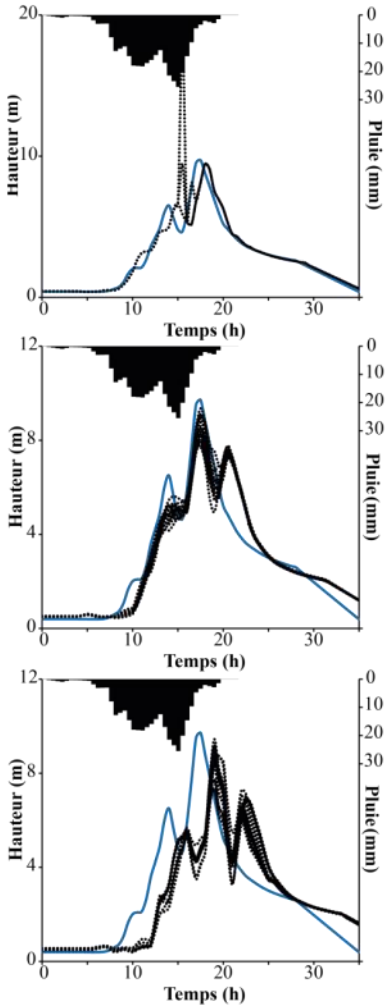
✓ Importante dispersion des prévisions

Ajout de termes directs



- ✓ **Ajout d'une partie linéaire au perceptron multicouche**
  - ⇒ **meilleure prise en compte de la composante linéaire du processus**
  - ⇒ **moins de dispersion dans les prévisions des modèles issus d'initialisations des paramètres différentes**

Hauteurs prévues pour l'événement 19 par les 20 modèles sélectionnés

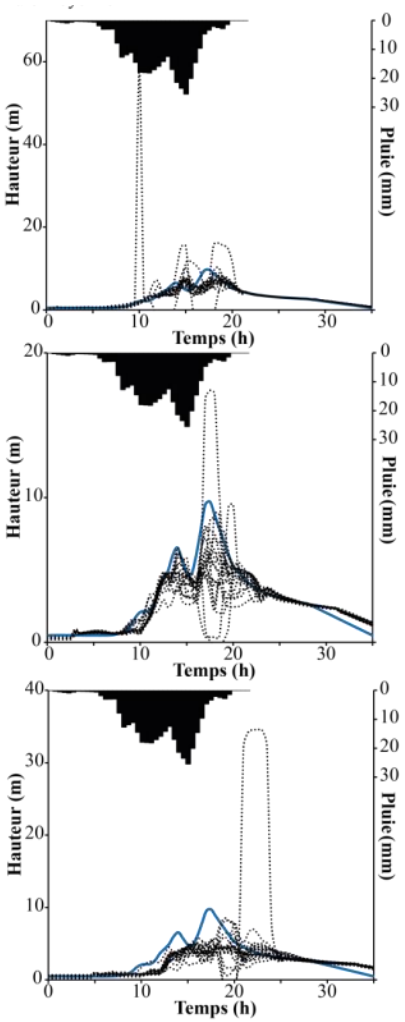


Pluie moyenne — Hauteur mesurée ····· Hauteurs calculées

h_p=1h

h_p=3h

h_p=5h

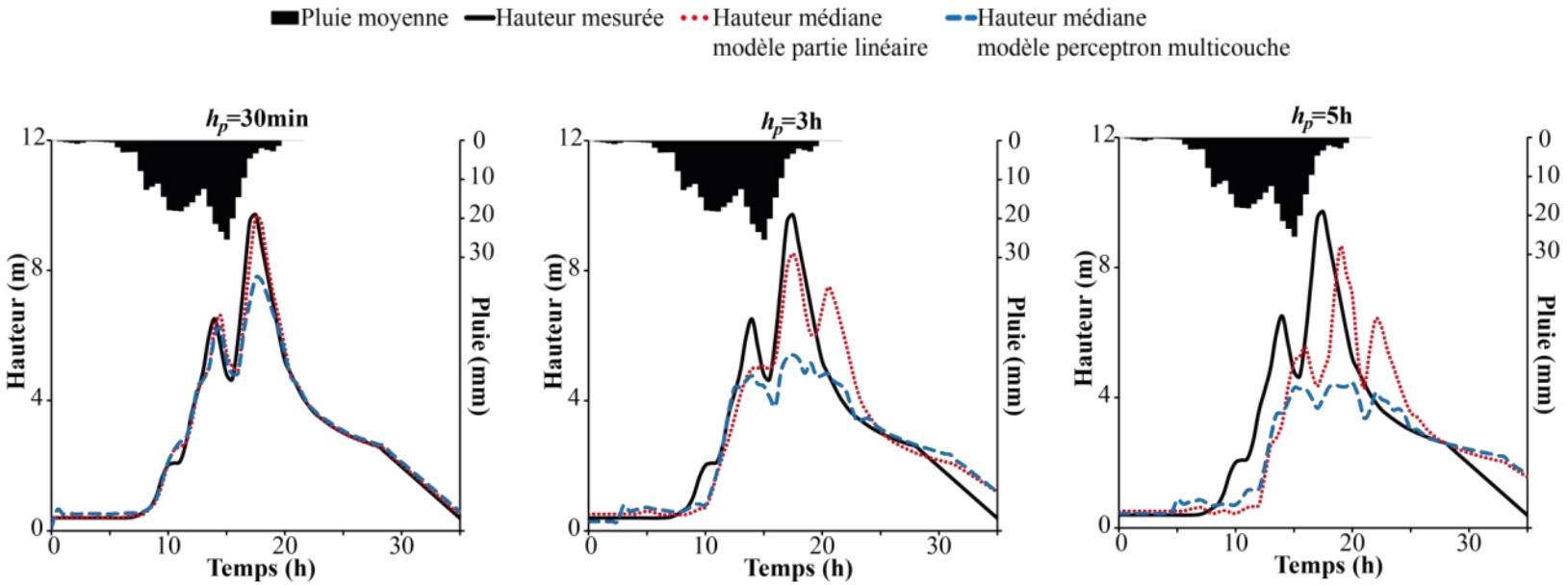


Avec partie linéaire

- **Dispersion des hauteurs prévues beaucoup plus faible pour les modèles avec partie linéaire**

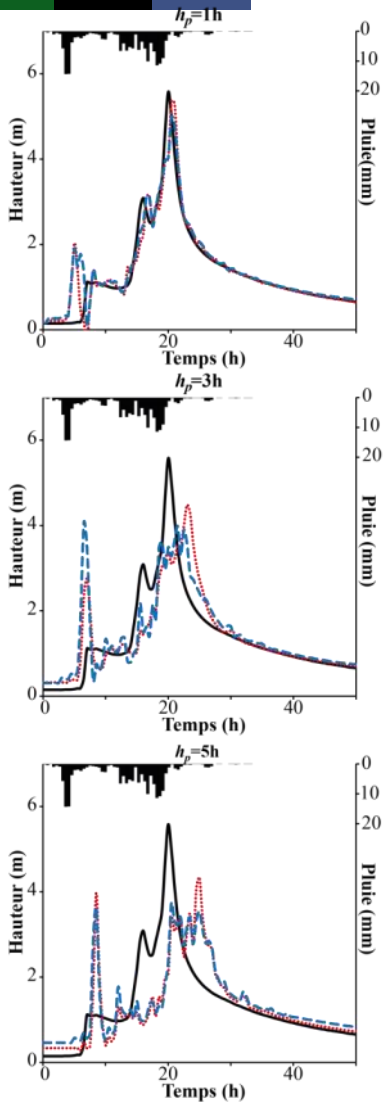
Sans partie linéaire

Hauteurs médianes prévues pour l'événement 19



Médianes		h_p=0,5h	h_p=3h	h_p=5h
***SPPD* (%)**	PM+PL	99	88	49
	PM	80	56	41
***R* (h)**	PM+PL	0	0	1,5
	PM	0	0	2,5

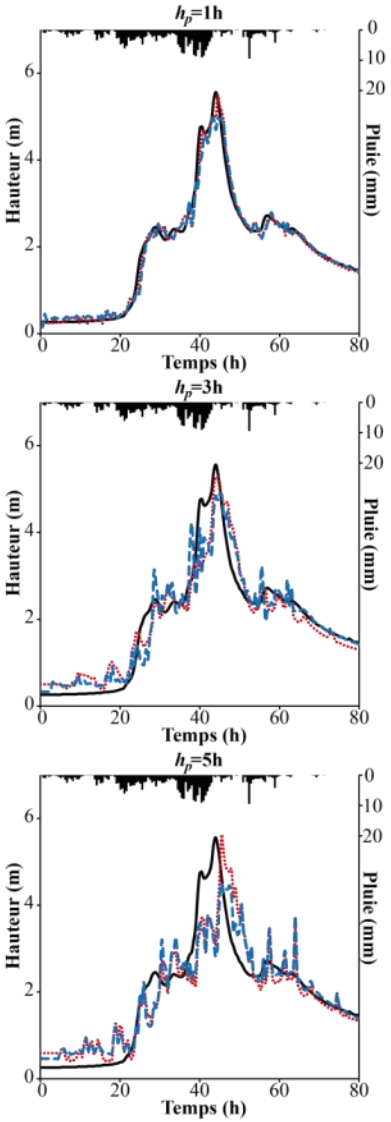
Hauteurs médianes prévues pour les événements 26 et 27



Médianes 26		h_p=1h	h_p=3h	h_p=5h
***SPPD* (%)**	PM+PL	72	61	37
	PM	72	63	42
***R* (h)**	PM+PL	0,5	3	5
	PM	0,5	1,5	0,5

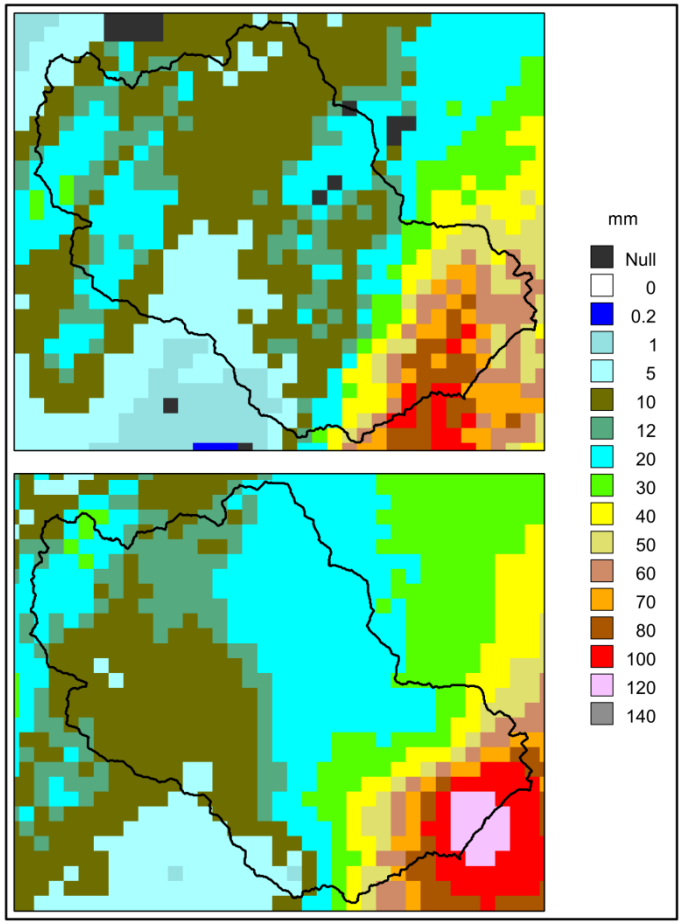
Événement 26 - Événement 27

Médianes 27		h_p=1h	h_p=3h	h_p=5h
***SPPD* (%)**	PM+PL	90	95	55
	PM	85	87	51
***R* (h)**	PM+PL	0,5	0,5	1,5
	PM	0,5	1	3,5



Synthèse

- **Sélection des modèles par validation croisée partielle**
- **Faible sensibilité de la capacité de généralisation à N_c, f_t et (μ_{ini};μ_{dec}) / influence importante de l'initialisation des paramètres**
- **Pour chaque horizon de prévision, sélection de 20 modèles issus d'initialisations différentes → dispersion des prévisions en test**
- **Ajout de termes directs au perceptron multicouche**
  - ✓ moins de dispersion des prévisions provenant des 20 modèles
  - ✓ amélioration notable de l'amplitude du pic
  - ✓ moins bonne synchronisation du pic pour certains h_p lointains
- **Anticipation jusqu'à h_p=2h voire 3h**



Précipitations sur le Gardon d'Anduze le 09/09/2002 à 02h00 (TU)
En haut : image RADAR sans traitement
En bas : image de la réanalyse

Seconde partie, la relation pluie-hauteur à Anduze et à Remoulins

Imagerie radar pour les Réseaux de Neurones

Yann VISSEROT [M2GCRN]
yann.visserot@gmail.com

Pierre-Alain AYRAL [EMA-LGEI]
Pierre-Alain.Ayral@mines-ales.fr

Audrey BORNANCIN-PLANTIER [EMA-LGEI]

Anne JOHANNET [EMA-LGEI] ***Anne.Johannet@mines-ales.fr***

SOMMAIRE

1. Rappel sur les données

2. Comparaison des différentes lames d'eau

3. Résultats avec RNF

4. Depuis décembre 2012…

5. Conclusion

1. Rappel sur les données

- Images radar de Nîmes et Bollène pour 28 épisodes de 1994 à 2009
- Réflectivité, au pas de temps 5 min et avec une résolution de 1 km²
- Une table de correspondance entre le niveau radar, la réflectivité et l'intensité de pluie

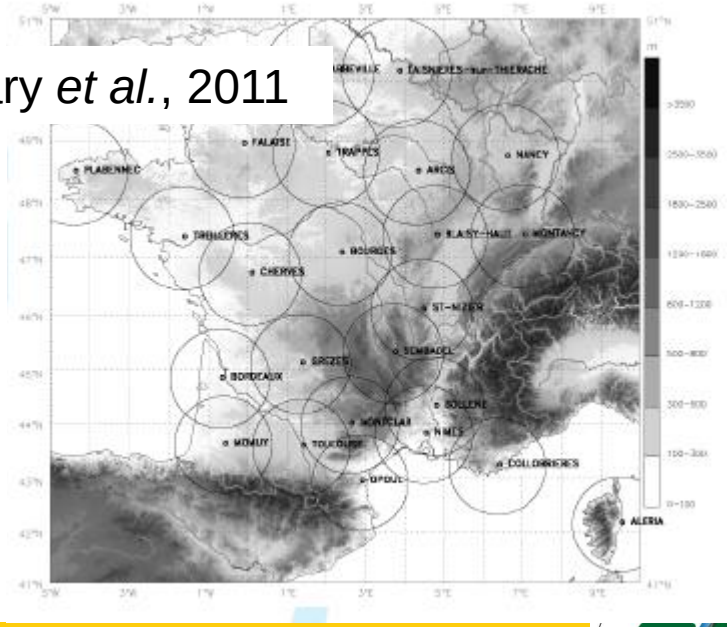
Tabary *et al.*, 2011

Correspondance entre Niveaux radar, Réflectivité et Intensité

(Echelle sur 54 niveaux)

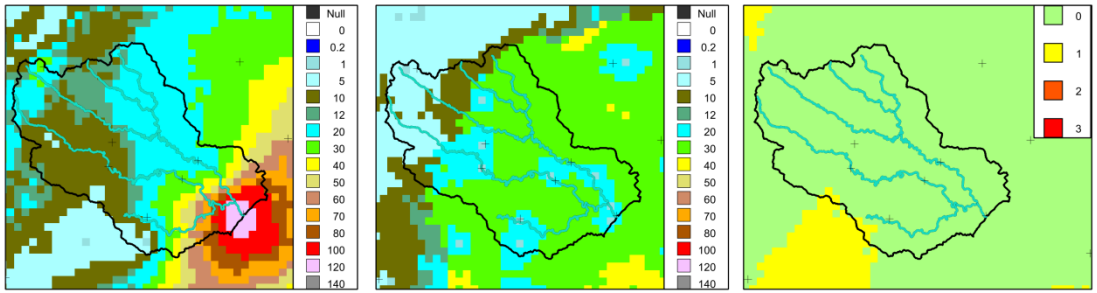
Niveau Radar (sans unité)	Réflectivité (dBz)	Intensité (mm/h)
0	inf. à 8 dBz	inf. à 0.2 mm/h
1	8 <= < 16	0.2 <= < 0.4
2	16 <= < 20	0.4 <= < 0.6
3	20 <= < 21	0.6 <= < 0.7
4	21 <= < 22	0.7 <= < 0.9
5	22 <= < 23	0.9 <= < 1.0
6	23 <= < 24	1.0 <= < 1.2
7	24 <= < 25	1.2 <= < 1.3
8	25 <= < 26	1.3 <= < 1.5

Niveau Radar (sans unité)	Réflectivité (dBz)	Intensité (mm/h)
30	47 <= < 48	31.6 <= < 36.5
31	48 <= < 49	36.5 <= < 42.1
32	49 <= < 50	42.1 <= < 48.6
33	50 <= < 51	48.6 <= < 56.2
34	51 <= < 52	56.2 <= < 64.8
35	52 <= < 53	64.8 <= < 74.9
36	53 <= < 54	74.9 <= < 86.5
37	54 <= < 55	86.5 <= < 99.9



1. Rappel sur les données

- Réanalyse des images RADAR de 2000 à 2006
- Carte à l'échelle de la France, au pas de temps horaire (TU)
- Ces images donnent trois informations :
  - Le cumul horaire (en 1/10ème de mm)
  - L'incertitude sur la valeur (en 1/10ème de mm)
  - Un « code qualité » (de 0 à 3)

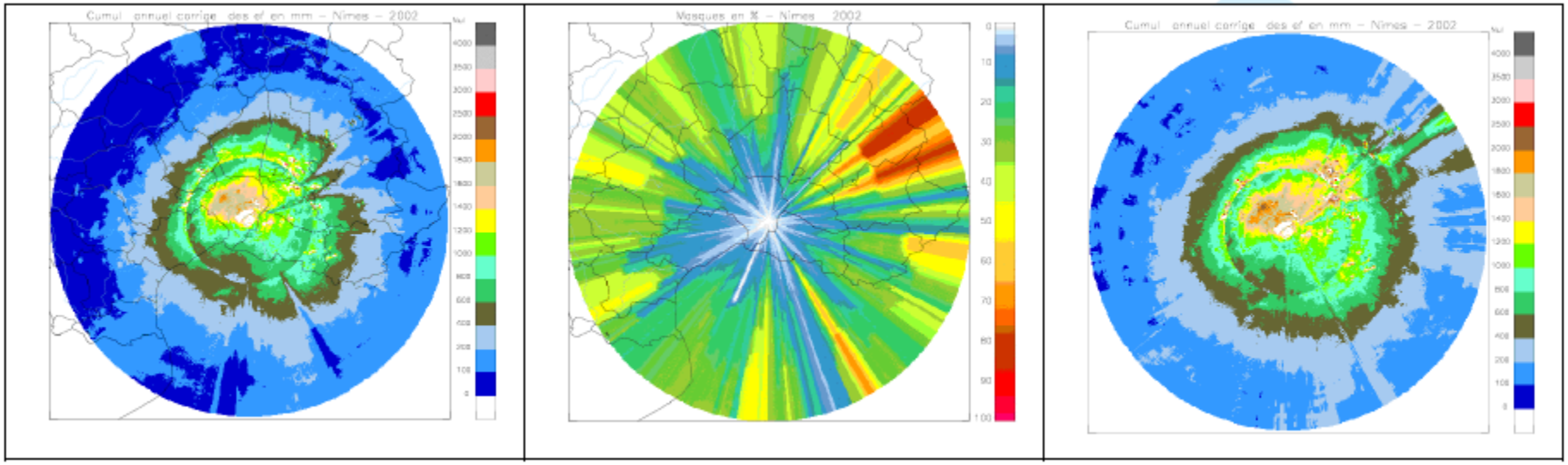


Géotiff 3 bandes - 09/09/2002 à 02h00 (TU)

1. Rappel sur les données

- Métadonnées des images radars de 2000 à 2006
- Carte annuelle des échos fixes et du taux de masquage
- Carte quotidienne du facteur correctif (radar/pluviomètre)

Tabary *et al.*, 2011



1. Rappel sur les données

Pour pouvoir utiliser la réflectivité (et les précipitations associées) dans le modèle de prévision, cette méthodologie a été retenue :

Transformation du format d'origine en .asc

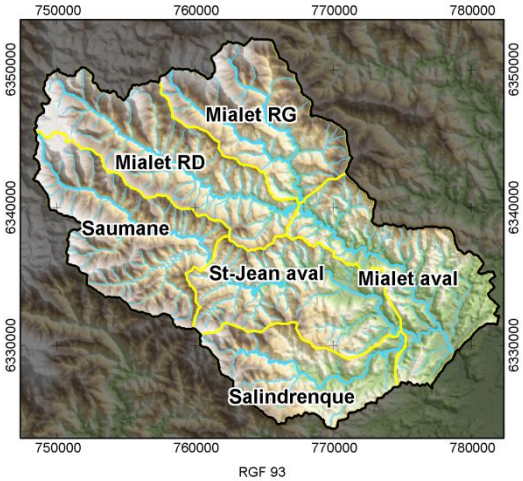
Découpage à l'échelle du rectangle englobant le BV

Suppression des échos fixes et traitement des taux de masquages

Passage au pas de temps 30min et spatialisation par 6 zones de la donnée

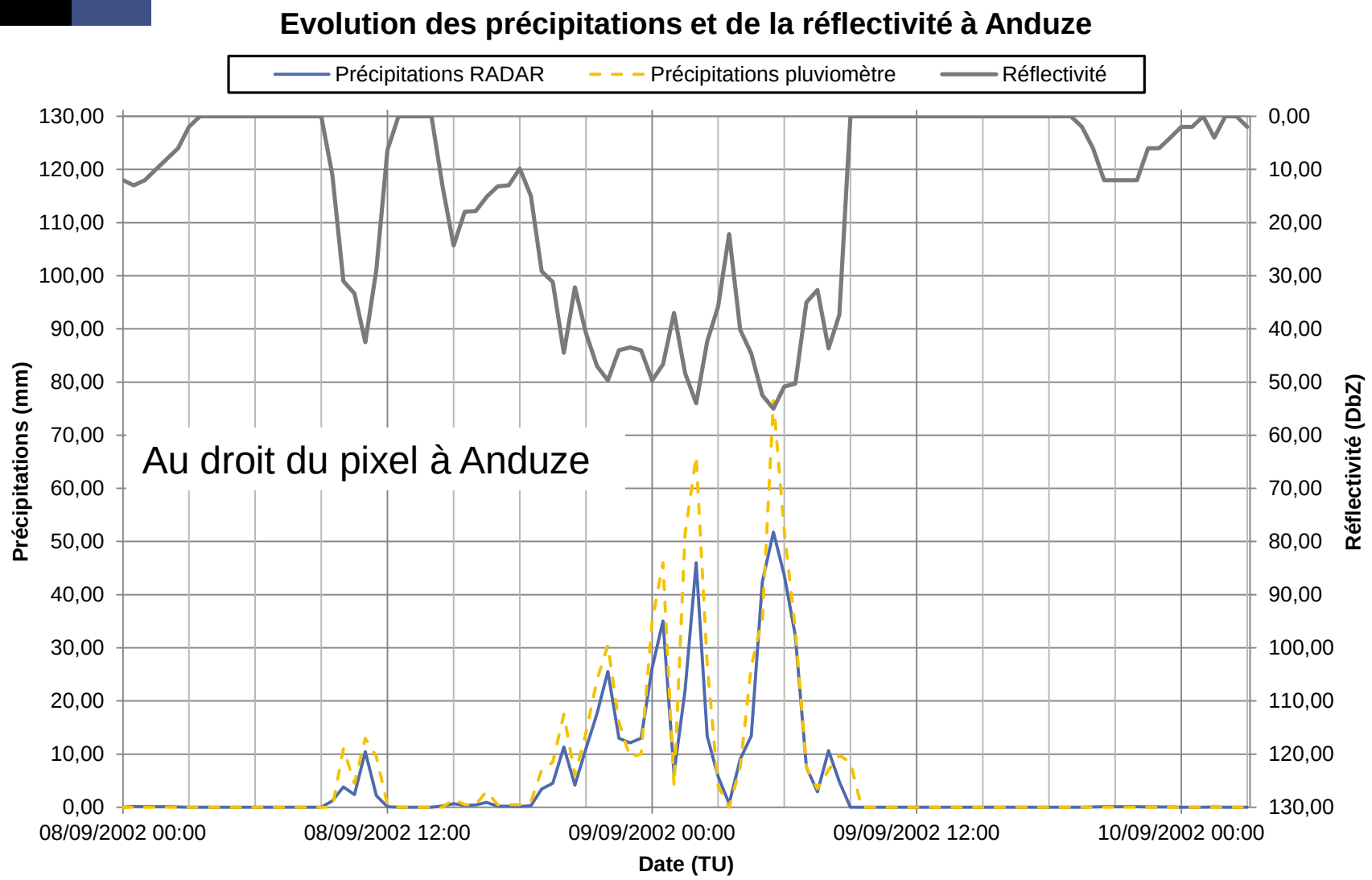
Sortie : un fichier .xls

RNF Pro

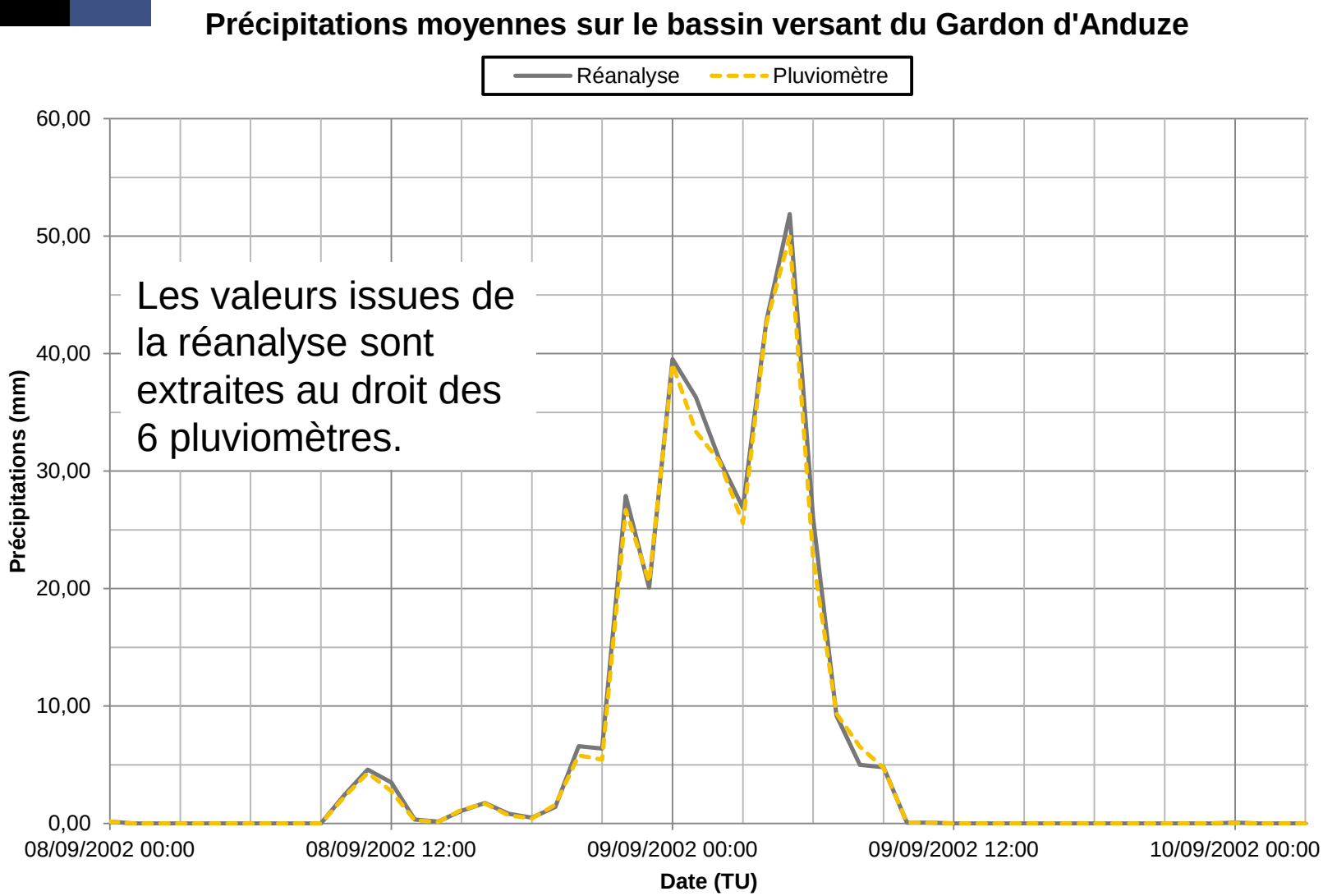


Bassin versant du Gardon d'Anduze découpé en 6 zones

2. Comparaison des différentes lames d'eau

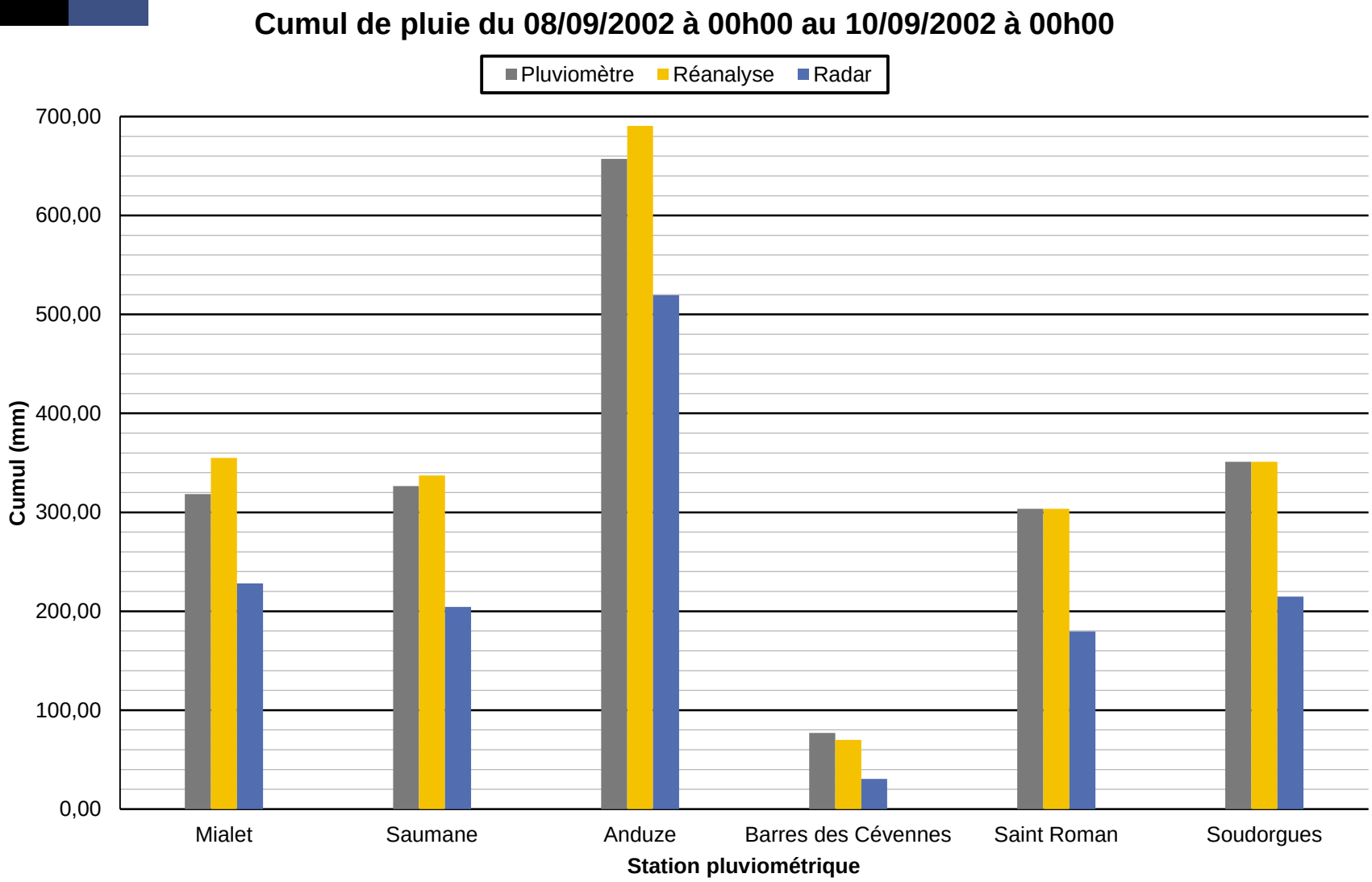


2. Comparaison des différentes lames d'eau



Les valeurs issues de la réanalyse sont extraites au droit des 6 pluviomètres.

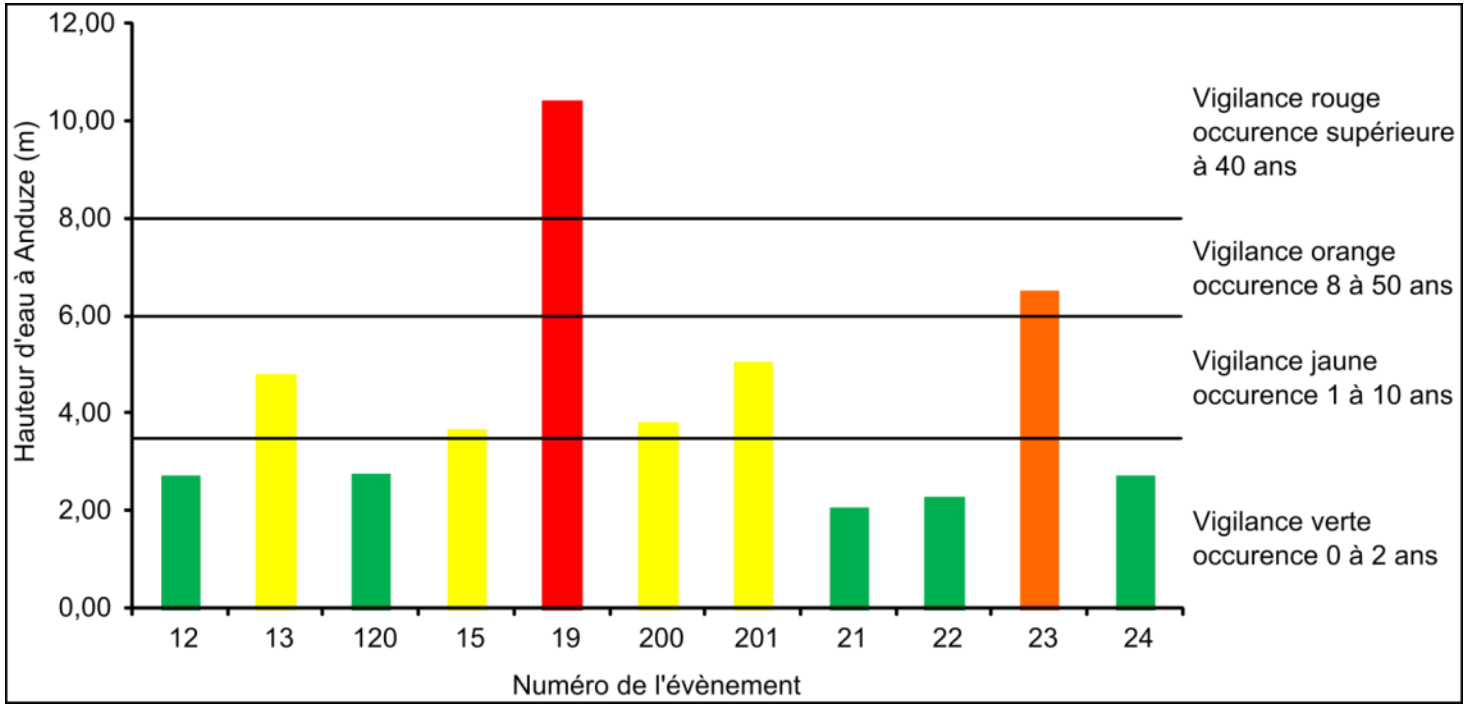
2. Comparaison des différentes lames d'eau



2. Comparaison des différentes lames d'eau – **Sur la base de données**

Caractérisation des évènements :

- Base d'apprentissage restreinte (11 évènements)
- Peu d'épisodes intenses



2. Comparaison des différentes lames d'eau – **Sur la base de données**

Comparaison entre :

- Les mesures ponctuelles et le pixel radar correspondant
- Les mesures ponctuelles et les valeurs des zones radar

Objectif : évaluer la différence entre les différentes variables d'entrées des modèles

Outils de comparaison :

- L'écart algébrique relatif : $\varepsilon = \frac{L1-L2}{L2} \times 100$
- Le coefficient de corrélation linéaire : $r^2 = \frac{\sum_{i=1}^{N}(x_i-\bar{x}).(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}.\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}}$

2. Comparaison des différentes lames d'eau – **Sur la base de données**

Comparaison mesures ponctuelles / pixels radar :

- Ecart relatif par évènement :

	12	13	120	15	19	200	201	21	22	23	24
ε(%)	-28	-39	**-66**	-42	-35	-47	-34	-33	-54	-54	**-22**

Sous-estimation moyenne des mesures radar de -41 %.

- Coefficient de corrélation linéaire par évènement :

	12	13	120	15	19	200	201	21	22	23	24
r^2	**0,61**	0,80	0,71	0,71	**0,95**	0,84	0,76	0,86	0,91	0,88	0,75

Coefficient moyen : 0,80.

Sous-estimation des mesures mais signal similaire.

2. Comparaison des différentes lames d'eau – **Sur la base de données**

Comparaison mesures ponctuelles / zones radar :

- Ecart relatif par évènement

	12	13	120	15	19	200	201	21	22	23	24
ε(%)	-26	-40	**-63**	**-21**	-26	-49	-32	-25	-54	-51	-27

Diminution moyenne de -37,5 %

- Coefficient de corrélation linéaire par évènement :

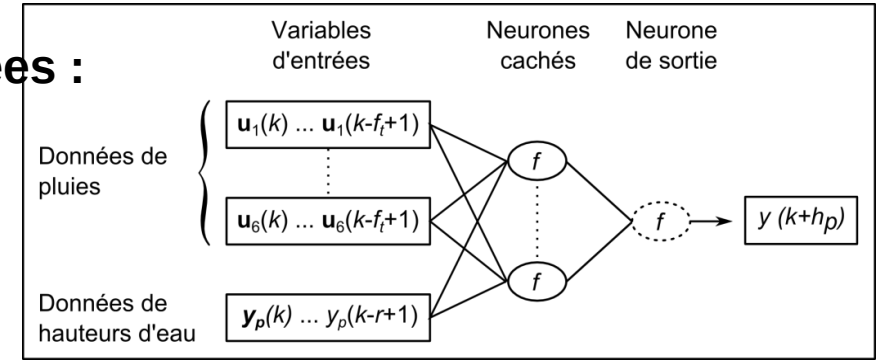
	12	13	120	15	19	200	201	21	22	23	24
r^2	0,55	**0,85**	0,68	0,61	0,80	**0,85**	0,70	0,79	0,81	0,74	**0,44**

Coefficient moyen : 0,71

Signal diminué mais similaire.

3. Résultats avec RNF – 6 Pluviographes Vs 6 Zones radar

Les sous-ensembles de la base de données :



- Evènement de 2002 en test
- Evènement de septembre 2000 en arrêt
- Sous-ensemble d'apprentissage : le reste des évènements

Sélection de l'architecture :

- Sélection de la fenêtre temporelle des pluies
- Sélection de l'historique des hauteurs observées (ordre)
- Sélection du nombre de neurones cachés (complexité du modèle)

3. Résultats avec RNF – 6 Pluviographes Vs 6 Zones radar

Architecture du modèle « 6 pluviographes »

h_p	0,5 h	1 h	2 h	3 h	4 h	5 h
Fenêtre temporelle f_t (heure)	0,5	5	4,5	4,5	5	4,5
Retard r (heure)	1,5	4,5	4,5	1	0,5	1
Complexité N_c	2	2	2	2	1	1

Architecture du modèle « 6 zones radar »

h_p	0,5 h	1 h	2 h	3 h	4 h	5 h
Fenêtre temporelle f_t (heure)	0,5	0,5	0,5	0,5	0,5	4,5
Retard r (heure)	1	1	1	2	2,5	1
Complexité N_c	1	3	4	3	4	1

3. Résultats avec RNF – **6 Pluviographes Vs 6 Zones radar**

Critères d'évaluation des prévisions :

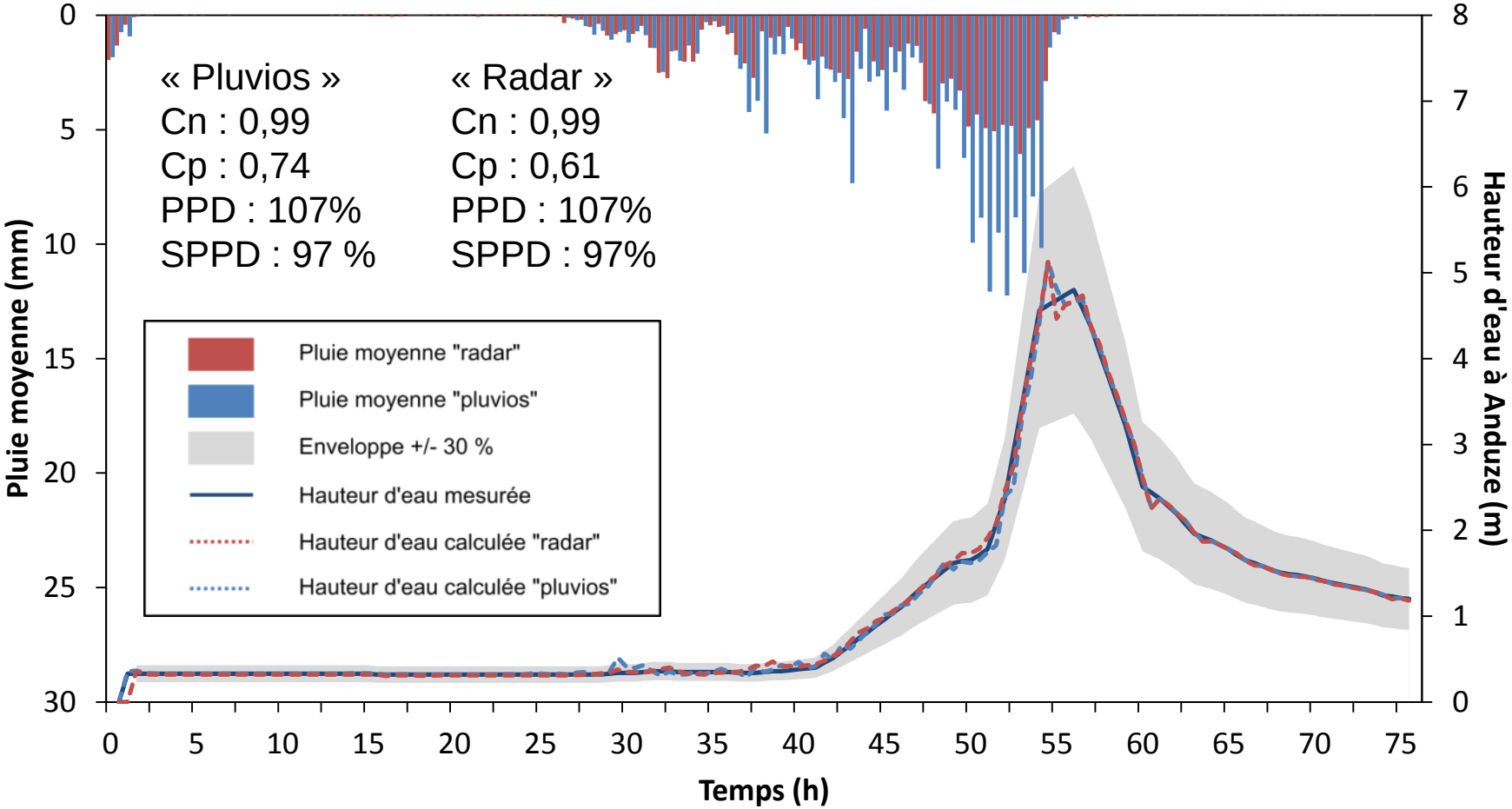
- Timing du pic de crue
- Pente de la montée de l'hydrogramme de crue
- Intensité du pic de crue (précision de 30 %)

Critères de qualité utilisés :

- Critère de Nash : adéquation des valeurs prévues avec les valeurs mesurées
- Critère de persistance : performance du modèle par rapport à une prévision naïve
- *Percentage Peak Discharge* (PPD)
- *Synchronous Percentage Peak Discharge* (SPPD)

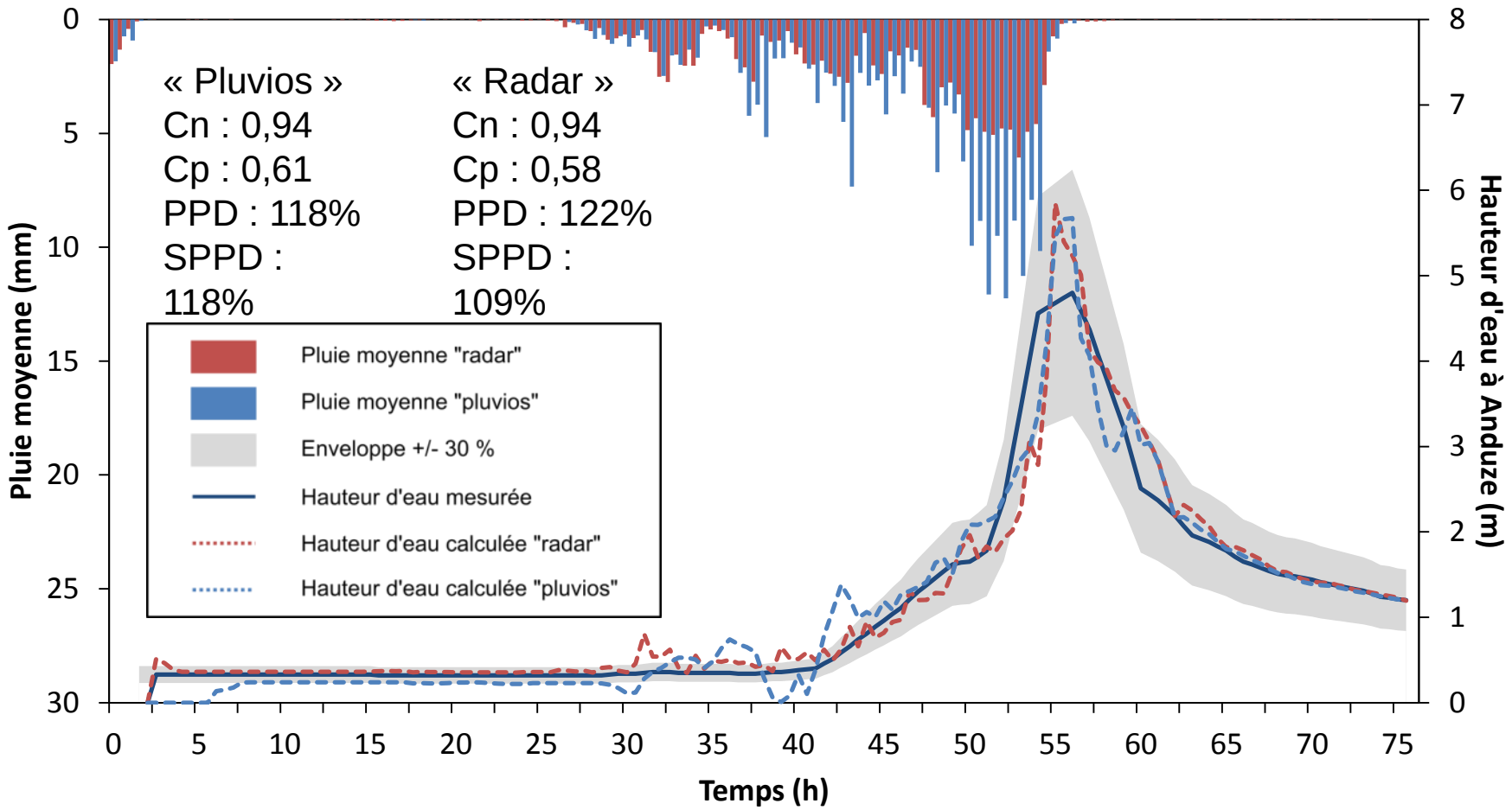
3. Résultats avec RNF – Evénement de septembre 2000 (1 000 m³/s)

Septembre 2000 – Hp = 30 min



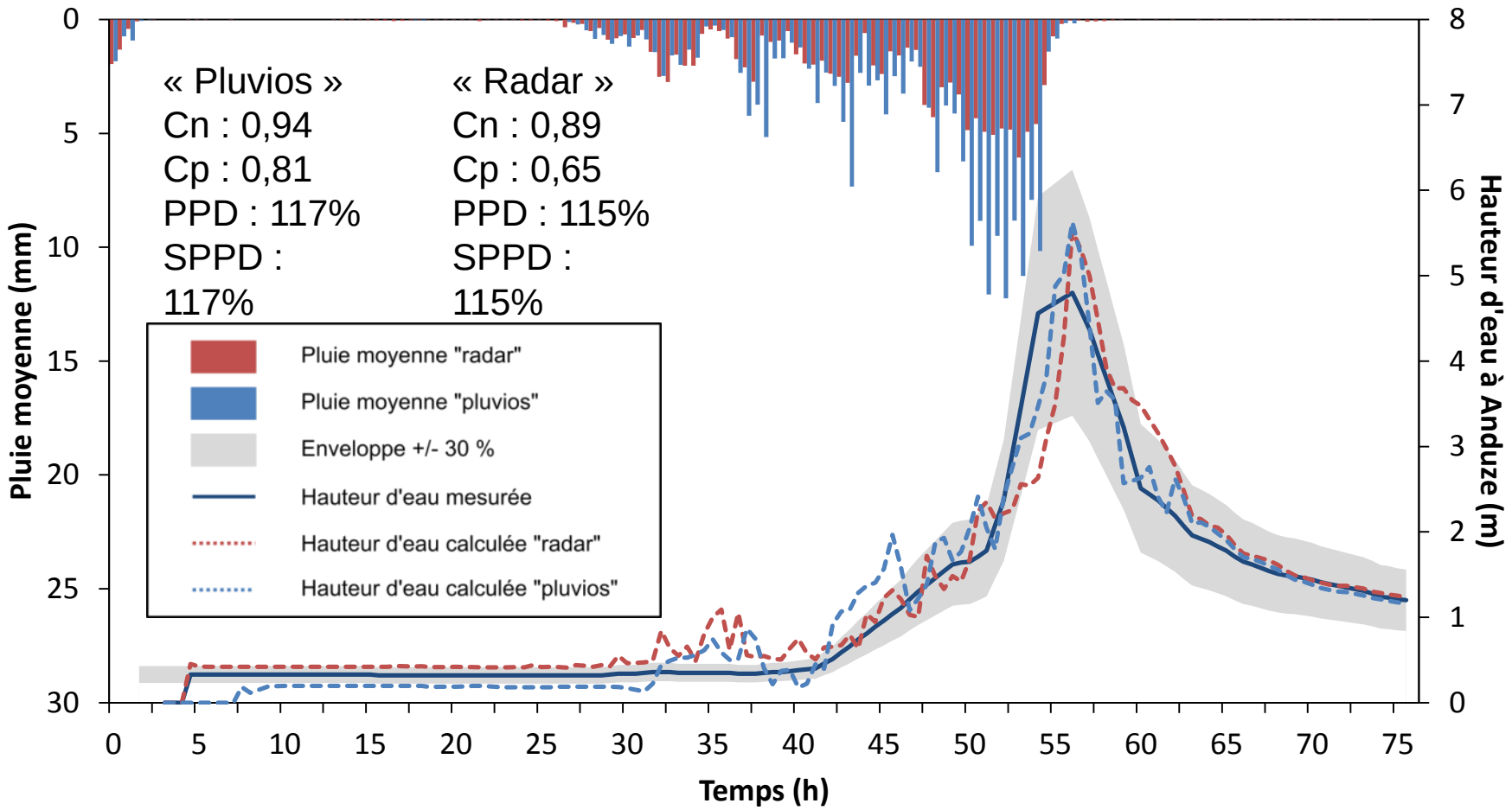
3. Résultats avec RNF – Evénement de septembre 2000 (1 000 m³/s)

Septembre 2000 – H_p = 2 h



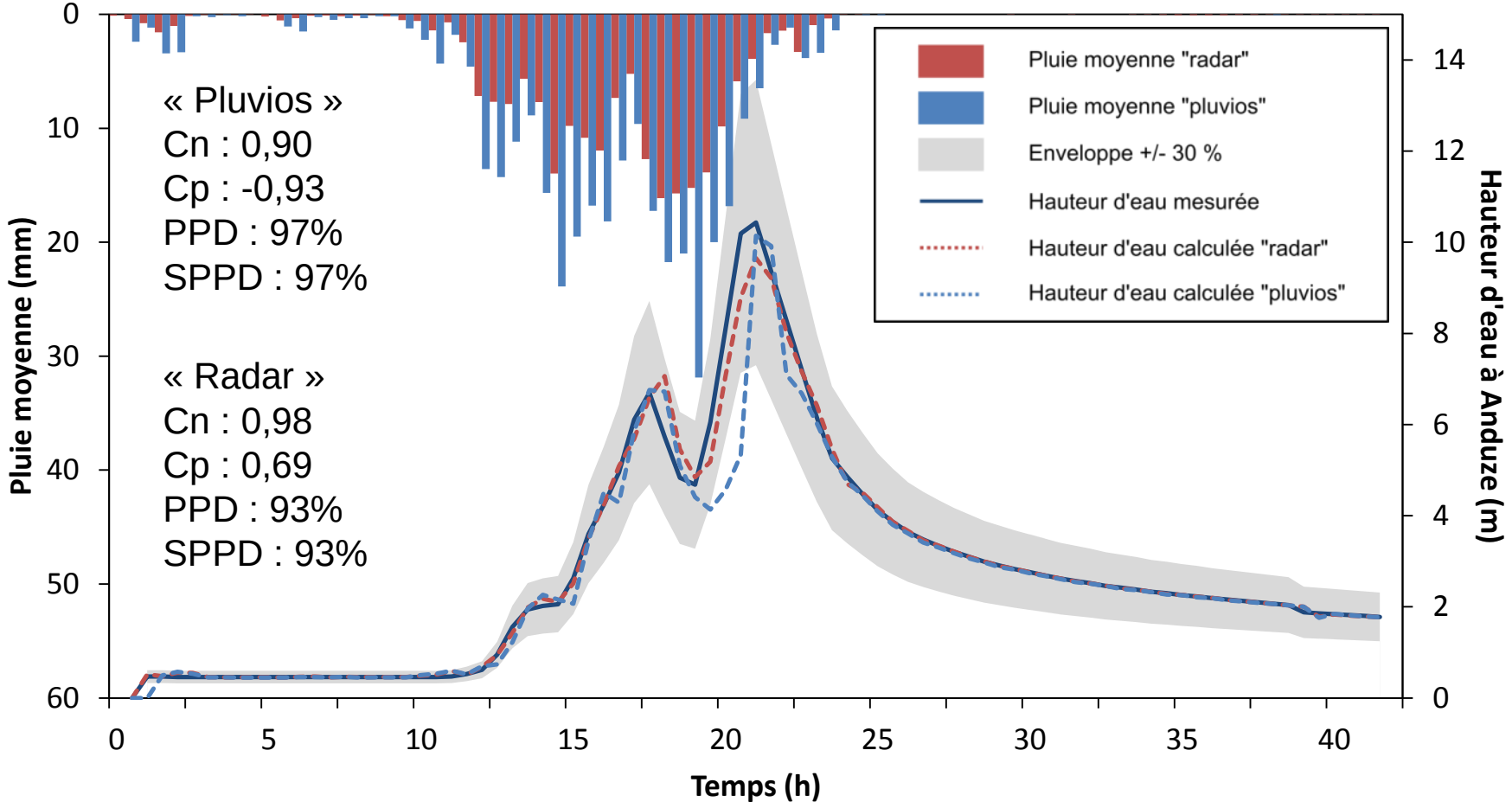
3. Résultats avec RNF – Evénement de septembre 2000 (1 000 m³/s)

Septembre 2000 – Hp = 3 h



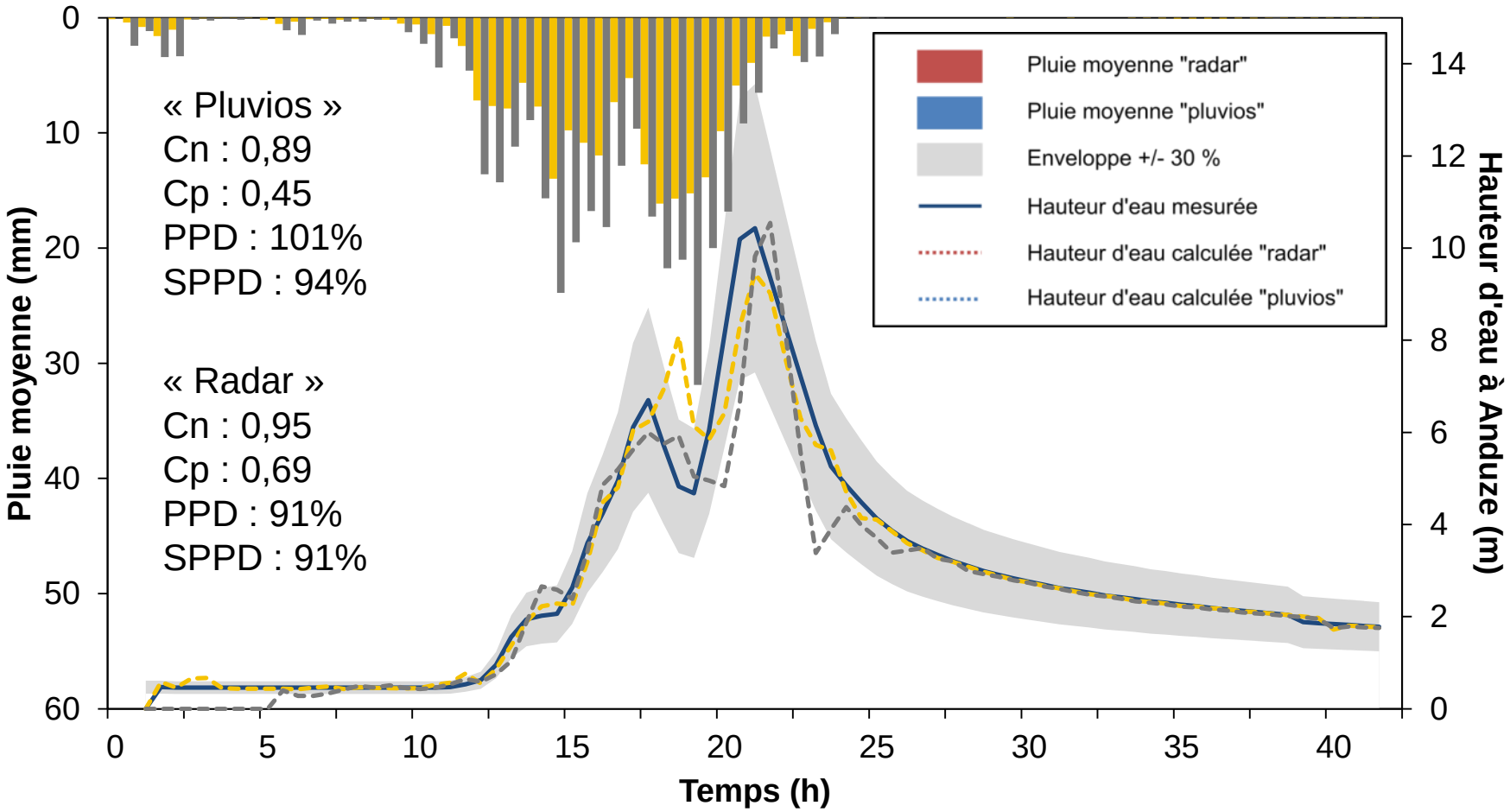
3. Résultats avec RNF – Evénement de septembre 2002 (3 500 m³/s)

Septembre 2002 – Hp = 30 min



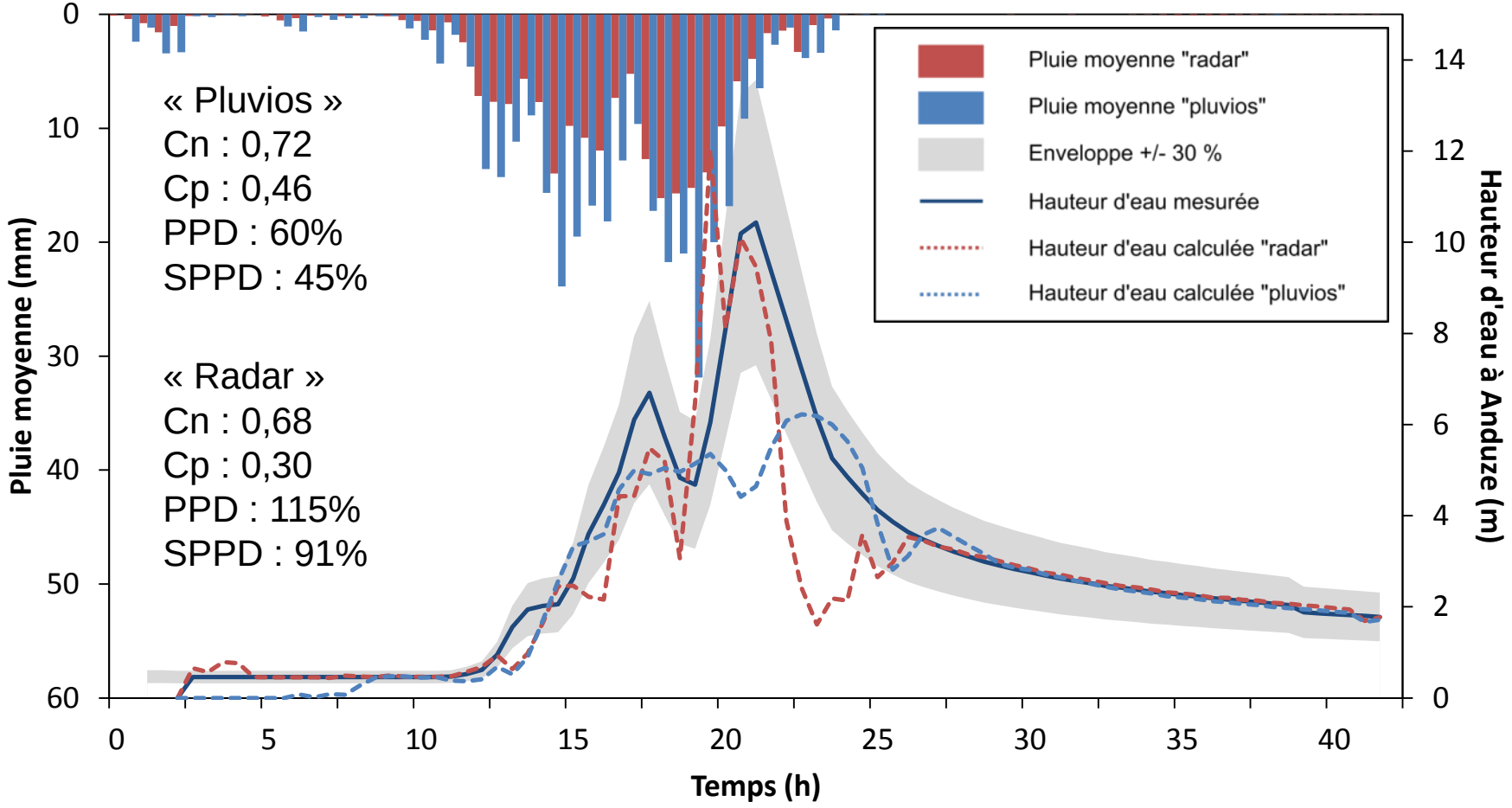
3. Résultats avec RNF – Evénement de septembre 2002 (3 500 m³/s)

Septembre 2002 – Hp = 1 h



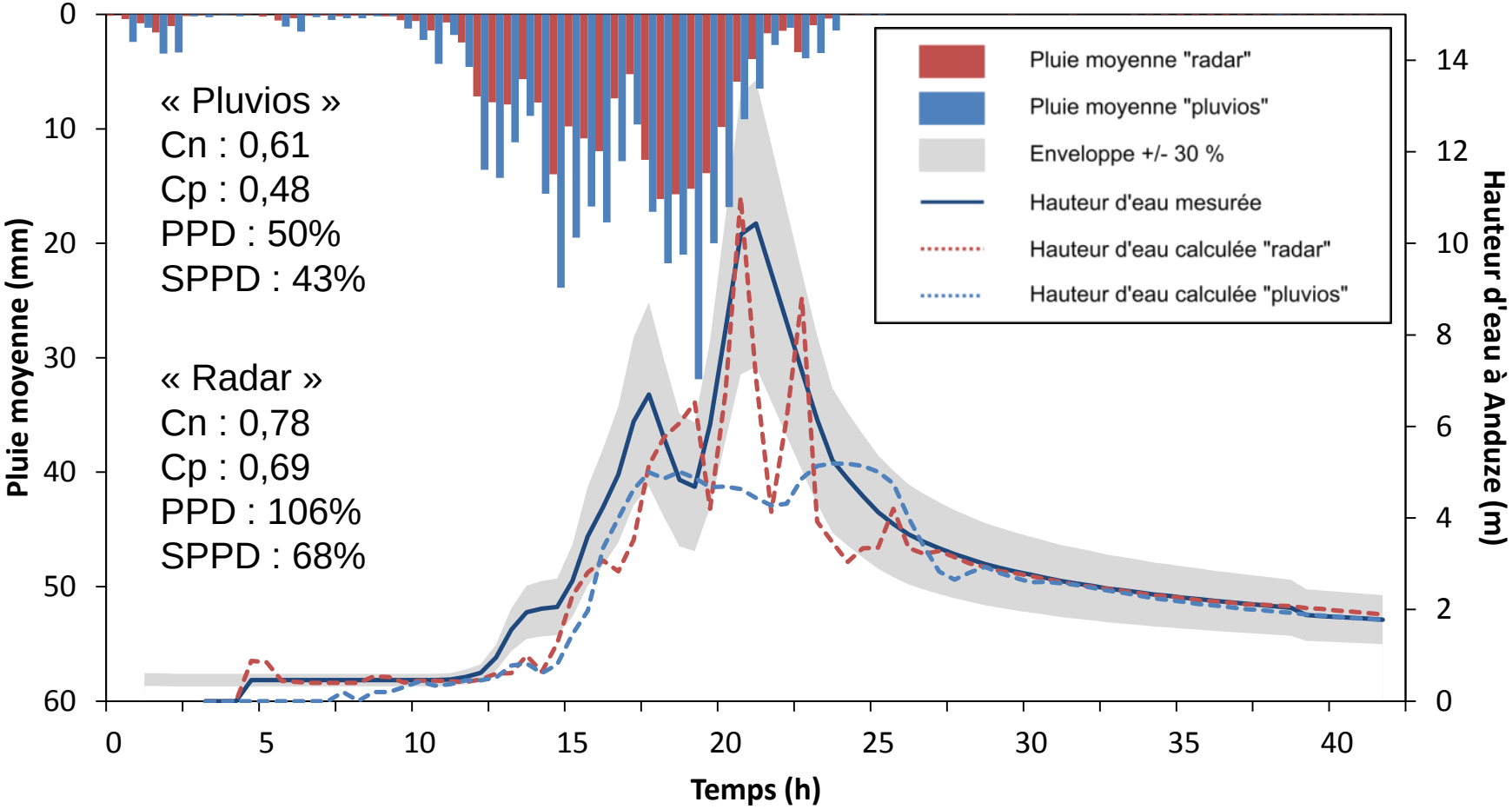
3. Résultats avec RNF – Evénement de septembre 2002 (3 500 m³/s)

Septembre 2002 – Hp = 2 h



3. Résultats avec RNF – Evénement de septembre 2002 (3 500 m³/s)

Septembre 2002 – Hp = 3 h

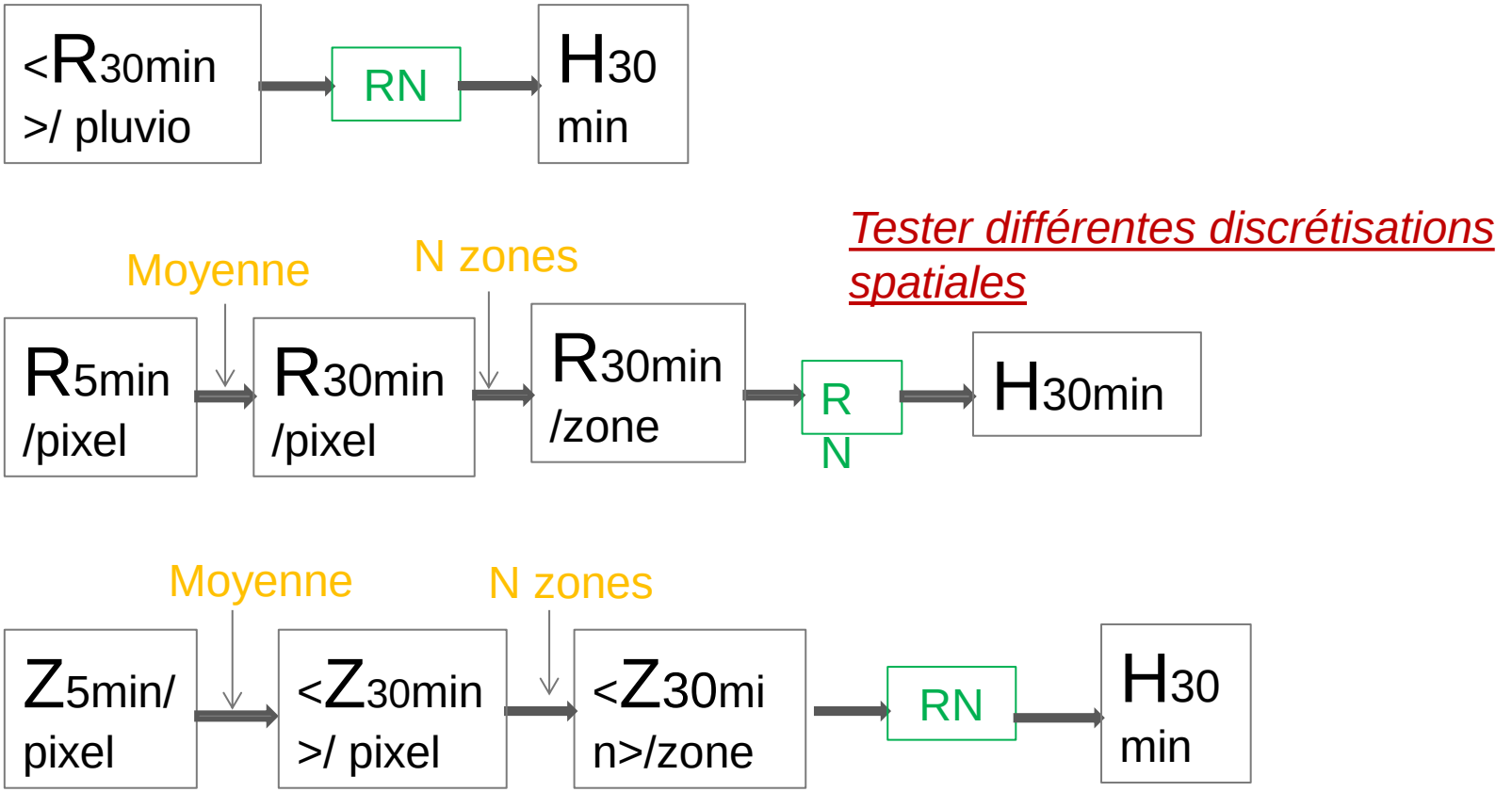


3. Résultats avec RNF – **Synthèse**

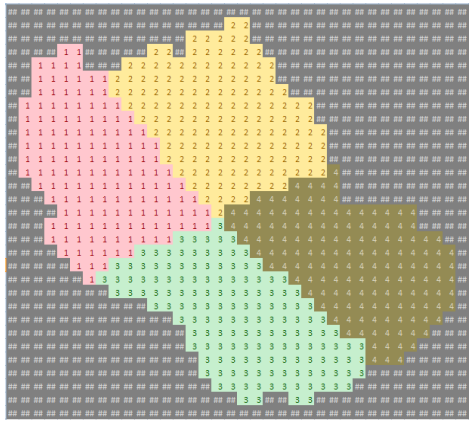
- Base de données tronquée pour permettre la comparaison avec les travaux d'Audrey
- Mesures radar sous-estiment les précipitations
- Modèle « pluvios » meilleur que le modèle « radar » pour un évènement non intense
- Modèle « radar » meilleur que le modèle « pluvios » sur l'évènement le plus intense
- Traitement des taux de masquage non effectif (effet mineur toutefois sur l'image de Nîmes)
- **Représentativité spatiale des « zones radar »**
- **Et la réflectivité….**

4. Depuis décembre 2012

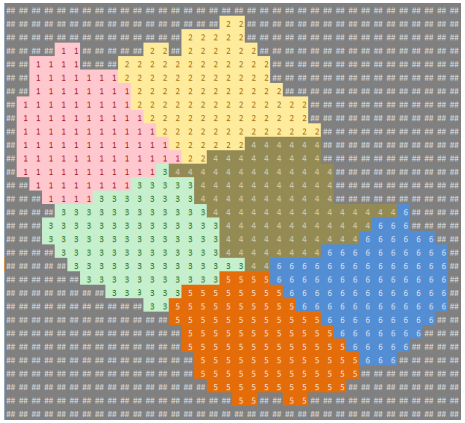
Une mission R&D – 2 élèves de M1 de l'EMA pendant 5 semaines



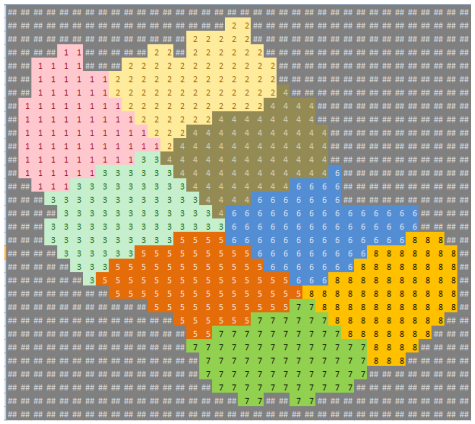
4. Depuis décembre 2012



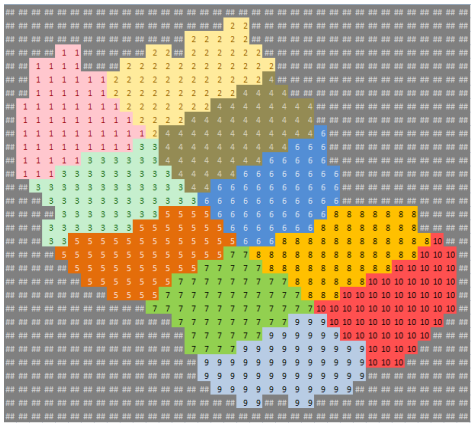
4 Zones



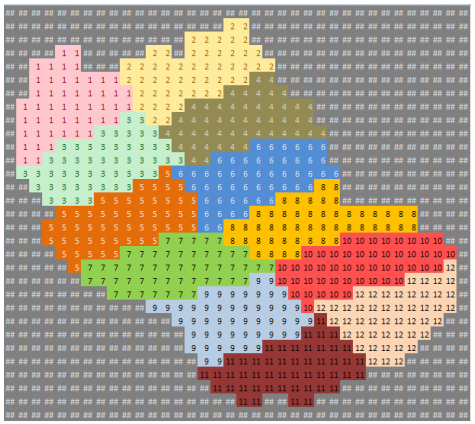
6 Zones



8 Zones



10 Zones

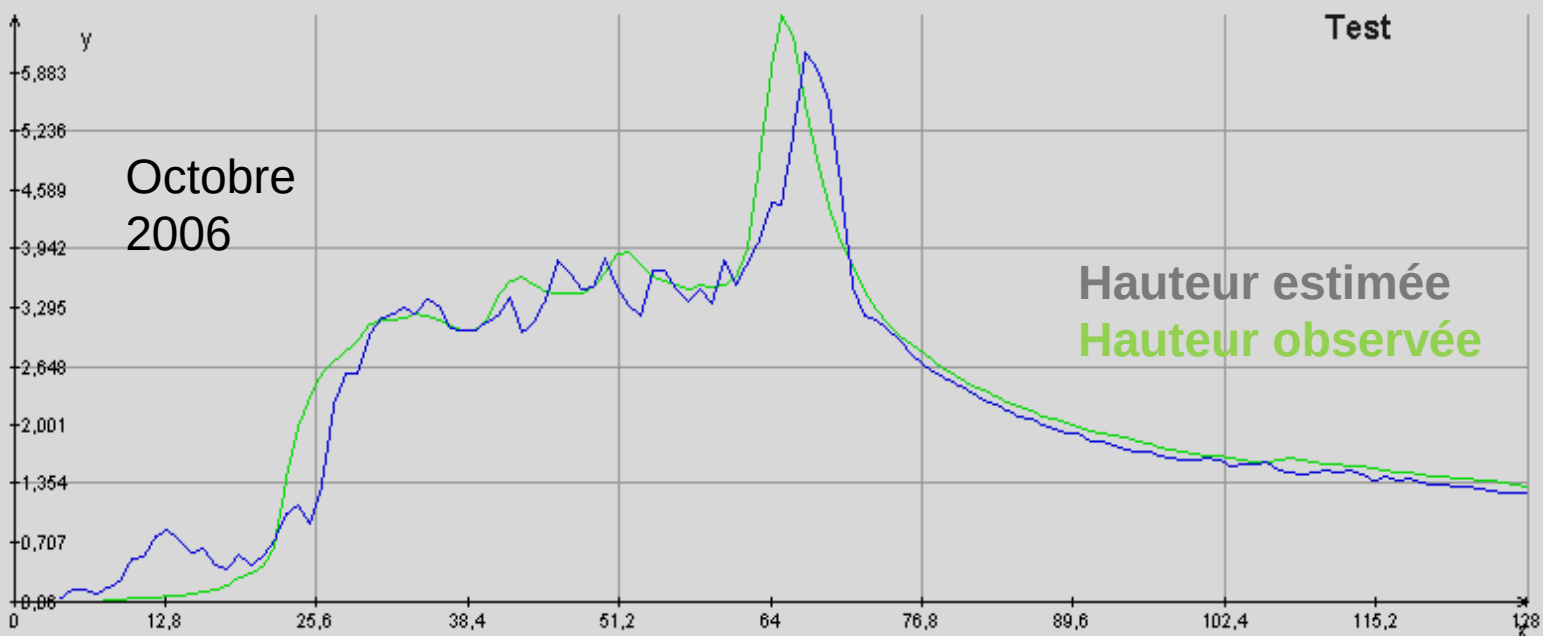


12 Zones

4. Depuis décembre 2012

Prévision à 2h00

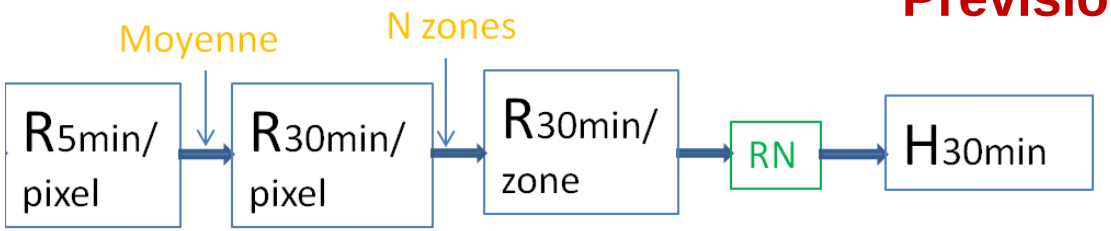
Nash	Persistance	Retard	Hauteur de pic
0,91	0,64	1h	-0,5m

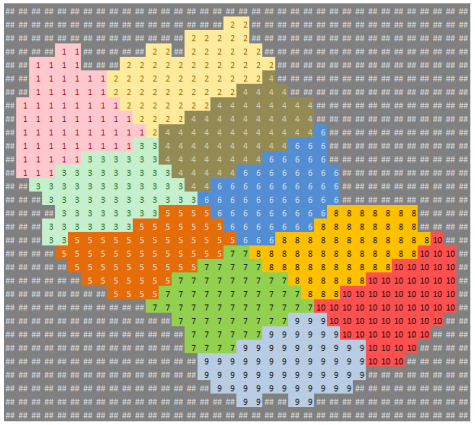


Nombre de pas de temps (30min)

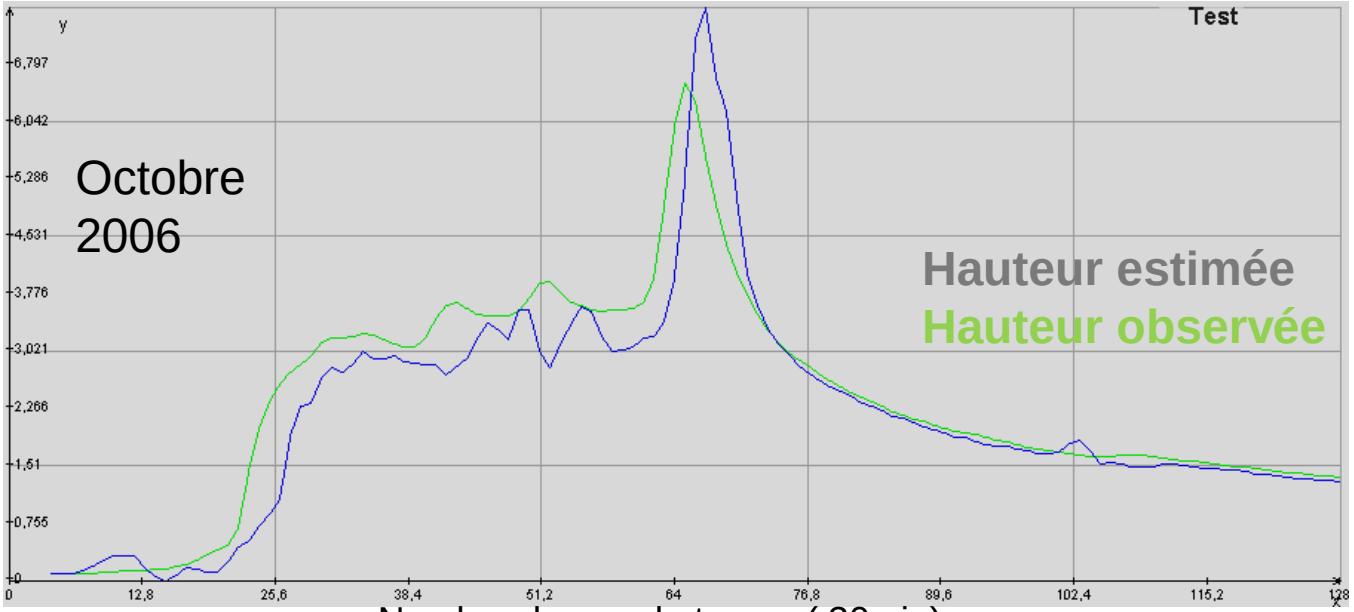
4. Depuis décembre 2012

Prévision à 2h00



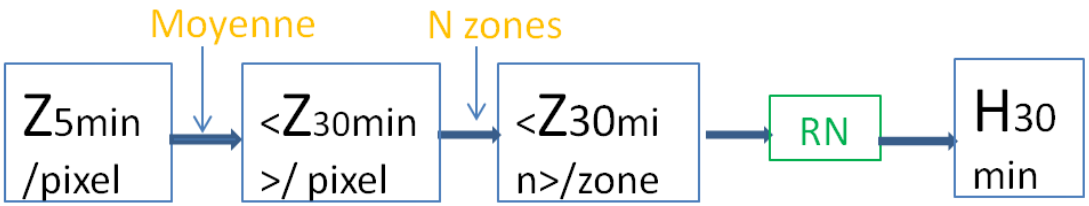


	Nash	Persistance	Retard	Hauteur de pic
10 zones	0,87	0,48	1h	+0,7m



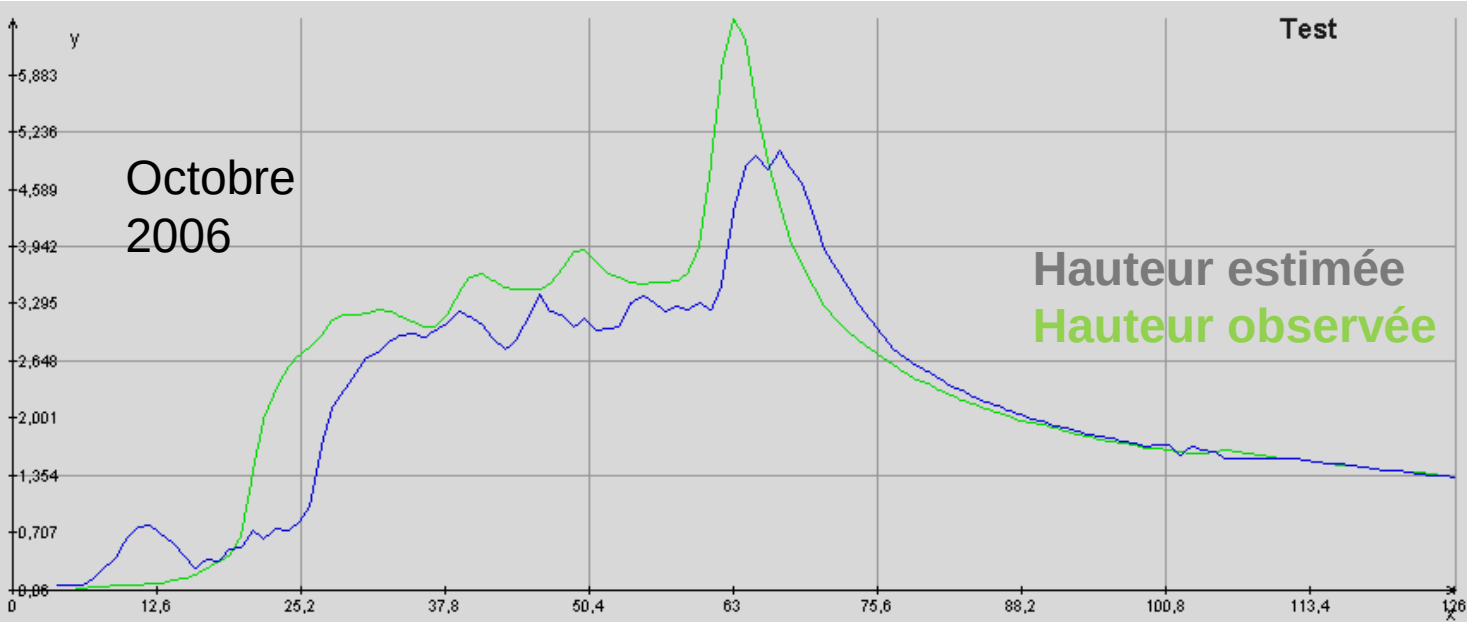
Nombre de pas de temps (30min)

4. Depuis décembre 2012



Prévision à 2h00

	Nash	Persistance	Retard	Hauteur de pic
4 zones	0,89	0,55	45 min	-1,6m

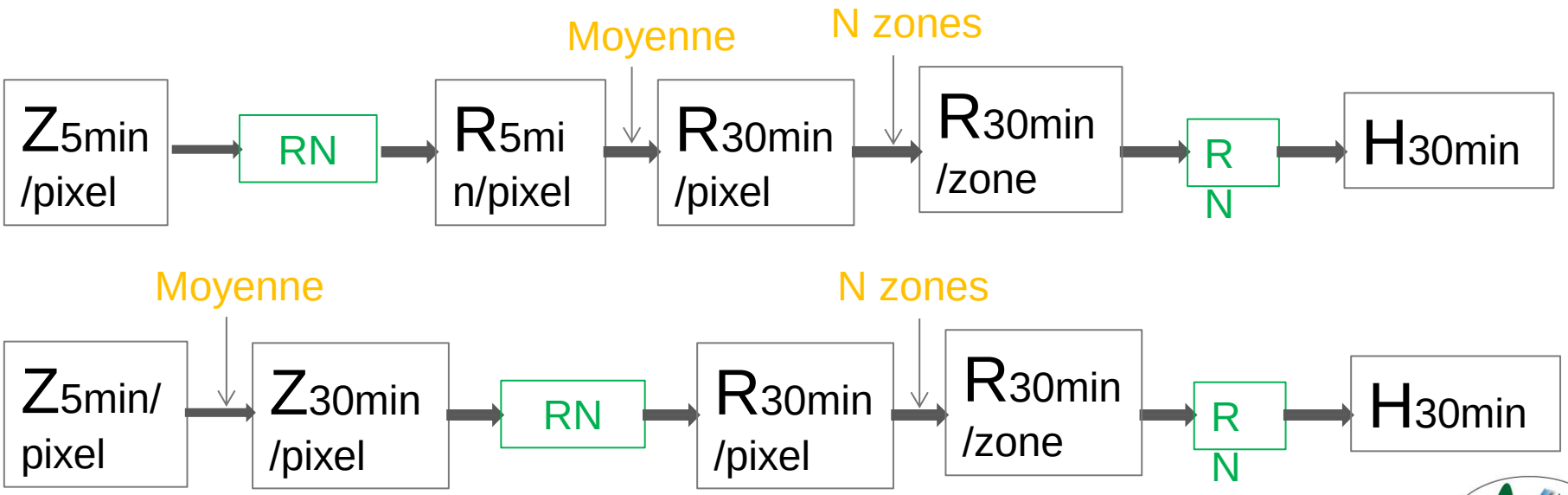


Nombre de pas de temps (30min)

4. Depuis décembre 2012

- Des résultats proches de ceux d'Audrey Bornancin-Plantier

- L'intégration de Z à 30 minutes pas optimale mais pas catastrophique
- Un optimum proche des 10 zones (*a priori* un découpage en 12 zones donne de moins bon résultats)



5. Conclusion

- Utilisation possible de l'imagerie radar (il n'y a pas souvent 6 pluviographes pour un bassin versant de 545 km²…)

- Discrétisation spatiale en 10 zones avec un découpage « hydrologique »

- Des résultats préliminaires sur l'intégration directe de la réflectivité

→ Poursuivre les tests sur l'intégration directe de la réflectivité

- Modifier le pas de temps du modèle : de 30 à 5 minutes

Et/ou

→ Déterminer une loi ZR par

le réseau de neurones (à 5 ou 30 minutes)

- Réaliser un pré-modèle