

## Proposition de Sujet de thèse 2019

Laboratoire (et n° de l'unité) dans lequel se déroulera la thèse :  
CNRM - UMR 3589

Titre du sujet proposé : **Prévision Probabiliste Immédiate des Précipitations**

Nom et statut du responsable de thèse : Bouttier François, IGPEF et HDR, équipe CNRM/GMME/PRECIP  
coencadrant: Hugo Marchal, équipe CNRM/GMME/PRECIP

Coordonnées du responsable de thèse : francois.bouttier@meteo.fr (tél:05 6107 9825)

### Résumé du sujet de la thèse

La prévision à quelques heures d'avance des précipitations est une information essentielle pour la prévision hydrologique, notamment l'anticipation des crues rapides et la gestion des réseaux d'assainissement. Les précipitations à échelle kilométrique sont souvent peu prévisibles à cause de leur structure spatio-temporelle et de leurs causes physiques multiples. Leur prévision utilise le plus souvent la modélisation numérique de l'atmosphère (déterministe ou ensembliste), et la prévision immédiate (qui est une extrapolation empirique des précipitations observées). Ces 2 types de techniques comportent un fossé qualitatif aux échéances typiquement comprises entre 1 et 3 heures, en particulier les prévisions numériques convergent rarement vers les observations aux très courtes échéances. Il n'existe pas non plus de technique établie pour assortir la prévision immédiate d'information probabiliste.

La thèse proposée vise à combler ces lacunes en étudiant des techniques de prévision probabiliste sans couture entre les échéances de 0 à 6 heures d'échéances. L'objectif est d'élaborer des prévisions de qualité optimale, rafraîchissables avec les dernières observations, et assorties d'une information probabiliste. On s'intéressera particulièrement à la cohérence temporelle des prévisions (importante pour la prévision des cumuls de précipitations susceptibles de produire des crues) et à la prise en compte d'enjeux de sécurité (qui influent sur les taux de fausses alertes et de non-détection admissibles dans les prévisions).

On étudiera des post-traitements innovants des prévisions immédiates et numériques, notamment ensemblistes; des algorithmes d'apprentissage automatique permettront d'extrapoler optimalement les mesures disponibles en temps réel, et tenir à jour les probabilités des prévisions. Ces techniques seront testées sur une combinaison d'historiques longs et de cas de crues rapides remarquables, reconstituées par prévision rétrospective sur plusieurs années.

Nature du travail attendu et compétences souhaitées: travail de recherche appliquée en modélisation numérique, nécessitant des compétences en physique atmosphérique et statistiques, travail en équipe, expression scientifique écrite et orale, programmation python.

### Principales références bibliographiques

Cloke, H.L., Pappenberger, F., 2009: Ensemble flood forecasting: A review. J. Hydrol. (2009), doi:10.1016/j.jhydrol.2009.06.005

Nipen, Thomas N., Greg West, Roland B. Stull, 2011: Updating Short-Term Probabilistic Weather Forecasts of Continuous Variables Using Recent Observations. Wea. Forecasting, 26, 564-571. doi: http://dx.doi.org/10.1175/WAF-D-11-00022.1

Vincendon, B., Ducrocq, V., Nuissier, O. and Vié, B., 2011: Perturbation of convection-permitting NWP forecasts for flash-flood ensemble forecasting. Nat. Hazards and Earth Sys. Sci., 11, 1529-1544., doi:10.5194/nhess-11-1529-2011.

## Précisions sur le sujet 'Prévision Probabiliste Immédiate des Précipitations':

### 1. Contexte du sujet

En météorologie, **prévision immédiate** et **prévision d'ensemble** constituent deux communautés établies mais qui échangent peu, parce que les techniques sont historiquement différentes, et que les prévisions d'ensemble opérationnelles n'ont atteint que récemment des résolutions et des fréquences de rafraîchissement intéressantes pour les échéances inférieures à 6 heures. La fusion entre les deux disciplines est devenu un problème d'actualité, parce (1) les utilisateurs de prévisions immédiates réclament de plus en plus d'informations probabiliste pour lesquelles les techniques actuelles d'extrapolation (traitement d'image et approches objet) sont peu adaptées, et (2) les systèmes opérationnels d'assimilation de données raccordent mal les premières échéances des prévisions aux valeurs observées des nuages et précipitations, notamment parce que leurs algorithmes (analyse variationnelle et filtre de Kalman) reposent sur des hypothèses de linéarité incompatibles avec la complexité physique de ces phénomènes.

La prévision ensembliste à très courte échéance est encore un défi ouvert, que peu de centres météorologiques ont affronté. **Météo-France** possède un système de prévision numérique dédié à la prévision immédiate (Arome-PI), mais ses échéances très courtes ne permettent pas vraiment de l'utiliser de manière ensembliste, en outre il est à peine plus fidèle aux observations que le système traditionnel de prévision numérique dont il est tiré (Arome-France); la prévision d'ensemble PE-Arome n'est rafraîchie que toutes les 6 heures avec plusieurs heures de délais de production, elle est donc inutilisable telle quelle pour la prévision immédiate. Le **UK Met Office** a récemment configuré son système de prévision d'ensemble pour produire 3 membres toutes les heures, par volonté de fusionner prévisions immédiates et ensemblistes, mais les performances de ce système semblent limitées en termes de précipitations (Migliorini et al 2011). Le **DWD allemand** produit depuis 12 ans un ensemble à haute résolution toutes les 3 heures, qui a été précurseur pour la prévision d'ensemble à courte échéance, mais leur stratégie de prévision immédiate repose sur une approche par filtre de Kalman dont l'aboutissement semble lointain à cause de son coût. **Les USA** produisent un ensemble à haute résolution toutes les heures, mais avec un apport modeste pour les échéances de la prévision immédiate.

Dans le **monde hydrologique**, où la mise à jour des prévisions de crue est un enjeu très important, les modèles numériques peuvent en principe être rafraîchis très fréquemment, mais les performances sont limitées par la qualité des prévisions de pluie fournies par les centres météorologiques (Cloke et Pappenberger 2009). Certains centres de prévisions des crues ont donc tenté de contourner le problème avec des algorithmes de mise à jour à partir d'observations de pluie, de hauteur d'eau, voire des modèles de prévision statistiques tels que des réseaux de neurones, avec un succès limité (Tiwari et Chatterjee 2011).

### 2. Etat des recherches en cours sur le sujet

Au CNRM, la prévision probabiliste des crues rapides a été surtout explorée dans le contexte des **précipitations méditerranéennes intenses**, notamment dans le projet HyMeX. B. Vincendon et al (2011) ont publié une méthode statistique de perturbation des prévisions de pluie destinée à alimenter des prévisions hydrologiques probabilistes. A. Lovat termine actuellement un doctorat sur la

prévision à très courte échéance de crues méditerranéennes. Les services opérationnels de Météo-France développent des algorithmes sophistiqués de prévision immédiate des précipitations fusionnant observations et modèles numériques; ils sont demandeurs de progrès sur la production d'information probabiliste. Compte tenu des enjeux sociétaux très forts sur la prévision hydrologique et la prévention des crues rapides, plusieurs **partenariats de recherche** appliquée ont permis d'entretenir des collaborations avec des services hydrologiques, notamment le SCHAPI (service central de prévision des crues), des centres hydrologiques régionaux, des instituts tels qu'EDF, IRSTEA, IFSTTAR.

L'équipe participe actuellement au projet ANR PICS qui explore précisément le sujet de la thèse sous l'angle de la prévision des crues. En parallèle, elle développe des algorithmes de prévision automatique des précipitations qui intègrent les aspects probabilistes et prévision immédiate, grâce à une approche multimodèle qui fusionne modèles déterministes, prévisions d'ensemble, et prévision immédiate. **La thèse proposée** vise se concentre sur la fusion entre prévision d'ensemble et observations, sujet sur lequel les publications existantes sont limitées à des tentatives simples de pondération bayésienne (Ben Bouallègue 2013, Raynaud et al 2014) et de mise à jour de probabilités univariées (Nipen et al 2011). On utilisera des techniques plus générales, dont la faisabilité a été récemment établie lors du projet ANR PICS, avec des observations, des métriques et des scores mieux adaptés aux précipitations.

### **3. Ebauche de méthodologie et plan de travail**

Des méthodes de post-traitement des champs de précipitation prévus (par les systèmes existants de prévision immédiate, numérique et ensembliste) seront développées (sur des bases logicielles existantes) et testées objectivement sur des échantillons combinant de longs historiques (afin d'asseoir la significativité statistique des conclusions) et des études de cas extrêmes (notamment, des crues marquantes, comme l'Aude en octobre 2018). La profondeur des archives de cas extrêmes sera optimisée en recalculant au besoin des cas remontant jusqu'en 2015 environ, avec l'aide de l'encadrement.

**La première partie** du doctorat sera dans la continuité du travail réalisé dans l'équipe pour le projet ANR PICS: constituer un ensemble exploitable à très courte échéance par combinaison de prévisions numériques multiples (Arome-PI, Arome-France, PE-Arome) perturbées avec une méthode développée dans l'équipe (Vincendon et al 2011), et optimiser sa qualité probabiliste par un rééchantillonnage des membres comparable aux méthodes bayésiennes de type filtre particulaire (Raynaud et al 2014). L'idée physique de base est que les prévisions les plus proches des observations ont des chances d'être les meilleures quelques heures plus tard. Dans un contexte de prévision immédiate, on peut exploiter en temps réel les performances les plus récentes du système de prévision, et donc ajuster ses paramètres par un processus d'auto-apprentissage. Ces idées seront critiquées en développant des métriques adaptées à la prévision des précipitations intenses, et en examinant l'homogénéité spatio-temporelle des statistiques. En particulier, on regardera s'il est intéressant de distinguer différents types de temps (blocage orographique de flux océanique, convergence cévenole, ligne de grain méditerranéenne, etc). Cela rejoint la notion d'analogie, c'est à dire l'hypothèse que les erreurs tendent à se reproduire dans des configurations météorologiques similaires (Diomedé et al 2014). Les scores utilisés tiendront compte de la sévérité

des événements (c'est à dire de leur caractère climatologiquement plus ou moins exceptionnel).

**Dans une seconde partie**, on s'intéressera à la compatibilité entre information observée et des produits de prévision numérique: quelles sont les échelles et intensité pertinentes (et peut-on les identifier automatiquement) ? Les corrélations spatiales d'erreurs sont-elles isotropes, et peut-on tirer parti de leur géométrie préférentielle lors des épisodes de pluies intenses méditerranéennes (cette question entraîne celle de la justesse des corrélations des pluies prévues) ? Ce sera l'occasion d'étudier en détail les structures des erreurs de prévision des précipitations. Des techniques de réseaux de neurones pourront être expérimentées pour voir si on peut automatiser l'identification de structures précipitantes intéressantes pour leur potentiel prédictif (cellules en V, organisation en ligne...).

**En troisième partie**, en fonction des résultats précédents, on pourra s'intéresser à la notion d'impact des précipitations intenses: l'idée est qu'en cas de risque avéré de crue rapide, l'information importante se trouve dans les extrêmes des prévisions d'ensemble; il est donc utile d'orienter les perturbations des prévisions d'ensemble de manière à échantillonner au mieux ces extrêmes. Cela nécessitera une modélisation rigoureuse de la notion d'impact, dont une difficulté est qu'elle peut être associée à différentes durées d'accumulation des précipitations suivant les cas et les bassins versants. Ce travail pourra déboucher sur le test d'algorithmes innovants de prévisions d'ensemble et d'extraction de scénarios de prévision.

La valorisation des résultats météorologiques obtenus pourra être enrichie au-delà des observations de pluie, par des validations de prévisions de crues dans des modèles hydrologiques opérés par des partenaires.

#### **4. Insertion de la thèse dans un cadre collaboratif**

Comme expliqué ci-dessus (section 3), le sujet sera fortement intégré avec les autres activités de l'équipe en prévision d'ensemble, il bénéficiera des contacts déjà établis avec les services opérationnels de Météo-France, les instituts SCHAPI, EDF, IRSTEA et IFSTTAR pour les validations hydrologiques. La thèse donnera lieu à des contacts avec les autres équipes travaillant sur la prévision d'ensemble météorologique (centres ALADIN et HIRLAM, réseau européen SRNWP, équipe CNRM/GMAP/RECYF) ainsi qu'à des présentations en conférences météorologiques internationales (par ex. HyMeX, EMS, EWGLAM), et participation à des conférences en en prévision immédiate et hydrologie pour stimuler les perspectives professionnelles en fin de thèse.

Les spécialistes toulousains en observation radar, filtrage particulière et apprentissage automatiques seront sollicités pour expertiser la stratégie algorithmique.

Une collaboration interne entre l'équipe et les services de production de Météo-France est en place et permettra d'instruire la future valorisation opérationnelle des résultats obtenus (fiche interface ' approche probabiliste autour de PIAF', production d'alertes précipitations et de scénarios hydrométéorologiques).

#### **5. Compétences et outils de l'équipe encadrante sur le sujet**

F. Bouttier (HDR) a encadré et publié de nombreux travaux de recherche sur la prévision d'ensemble AROME, il encadre actuellement un CDD sur la prévision hydrométéorologique ensembliste en collaboration avec des agences de prévision

des crues. Il maîtrise un environnement de validation des expériences de prévision d'ensemble AROME (scores probabilistes, comparaison aux observations in situ, radar, foudre, satellite, hydrométéorologie). Il entretient des collaborations internationales sur la prévision d'ensemble et ses applications (réseau de coopération ALADIN-HIRLAM, projet spécial du CEPMMT).

H. Marchal est un ingénieur travaillant sur la prévision probabiliste opérationnelle des précipitations, en collaboration avec F. Bouttier, il codéveloppe les outils et données utilisés dans la thèse proposée.

Le sujet s'inscrit dans le programme de recherches amont du groupe CNRM/GMME pour l'amélioration de la prévision d'ensemble PE AROME et l'étude de la prévisibilité des processus précipitants. Ce programme est complémentaire des activités des équipes DirOp/PI et Dirp/COMPAS/DOP de Météo-France qui travaillent sur des aspects plus directement opérationnels de la prévision immédiate des précipitations et de leurs corrections statistiques. Un travail d'intercomparaison et de coordination des résultats des différentes équipes sur ce thème sera effectué par les encadrants en parallèle de la thèse.

## 6. Références complémentaires

Ben Bouallègue Z, Theis SE, Gebhardt C. 2013. Enhancing COSMO-DE ensemble forecasts by inexpensive techniques. *Meteorol. Z.* 22: 49–59.

Ben Bouallègue, Z., T. Heppelmann, S. Theis and P. Pinson, 2016: Generation of scenarios from calibrated ensemble forecasts with a dual-ensemble copula-coupling approach. *Mon. Wea. Rev.*, 144: 4737-4750. doi: 10.1175/MWR-D-15-0403.1

Diomede Tommaso, Chiara Marsigli, Andrea Montani, Fabrizio Nerozzi, and Tiziana Paccagnella, 2014: Calibration of Limited-Area Ensemble Precipitation Forecasts for Hydrological Predictions. *Mon. Wea. Rev.*, 142, 2176-2197. doi: <http://dx.doi.org/10.1175/MWR-D-13-00071.1>

Ebert, Elizabeth E., 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Mon. Wea. Rev.*, 129, 2461-2480. doi: [http://dx.doi.org/10.1175/1520-0493\(2001\)129<2461:AOAPMS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2)

Hamill, T.M., E. Engle, D. Myrick, M. Peroutka, C. Finan, and M. Scheuerer, 2017: The U.S. National Blend of Models for Statistical Postprocessing of Probability of Precipitation and Deterministic Precipitation Amount. *Mon. Wea. Rev.*, 145, 3441-3463. doi:10.1175/MWR-D-16-0331.1

Lu, Chungu, Huiling Yuan, Barry E. Schwartz, Stanley G. Benjamin, 2007: Short-Range Numerical Weather Prediction Using Time-Lagged Ensembles. *Wea. Forecasting*, 22, 580-595. doi: <http://dx.doi.org/10.1175/WAF999.1>

Migliorini, S., Dixon, M., Bannister, R. and Ballard, S. (2011), Ensemble prediction for nowcasting with a convection-permitting model - I: description of the system and the impact of radar-derived surface precipitation rates. *Tellus A*, 63: 468-496. doi: 10.1111/j.1600-0870.2010.00503.x

Mittermaier, M. P., 2007: Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles. *Q. J. R. Meteorol. Soc.*, 133, 1487-1500.

Osinski, R. and F. Bouttier, 2018: Short-range probabilistic forecasting of convective risks for aviation based on a lagged-average-forecast ensemble approach. *Met. Apps*, 25, 105-118. doi:10.1002/met.1674

Peel, Syd, Laurence J. Wilson, 2008: Modeling the Distribution of Precipitation Forecasts from the Canadian Ensemble Prediction System Using Kernel Density Estimation. *Wea. Forecasting*, 23, 575-595. doi: <http://dx.doi.org/10.1175/2007WAF2007023.1>

Schwartz, C.S., J.S. Kain, S.J. Weiss, M. Xue, D.R. Bright, F. Kong, K.W. Thomas, J.J. Levit, M.C. Coniglio, and M.S. Wandishin, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, 25, 263-280.

Tiwari M. K. and C. Chatterjee, 2010 : Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs) J. Hydrol. 382:20-33  
doi:10.1016/j.jhydrol.2009.12.013

Raynaud, L., Pannekoucke, O., Arbogast, P. and Bouttier, F. (2014), Application of a Bayesian weighting for short-range lagged ensemble forecasting at the convective scale. Q.J.R. Meteorol. Soc., 141: 459-468. doi: 10.1002/qj.2366