

# APPORT DE LA CONNAISSANCE EXPERTE POUR LA CLASSIFICATION SPATIALE DE COMMUNES DES ALPES FRANÇAISES EN DEUX ZONES CLIMATIQUES.

Aurore Lavigne <sup>1,2</sup>, Nicolas Eckert <sup>2</sup>, Eric Parent <sup>1</sup> & Liliane Bel <sup>1</sup>

<sup>1</sup> *Equipe MORSE, UMR518 AgroParisTech/INRA, 16 rue C. Bernard, 75231 Paris CEDEX 05, France.*

<sup>2</sup> *IRSTEA, UR ETGR, 2 rue de la Papeterie-BP 76, F-38402 St-Martin-d'Hères, France.*

**Résumé.** Dans un contexte de changement climatique analyser l'évolution de l'activité avalancheuse dans le passé est une première étape primordiale pour prédire correctement l'évolution future à long terme. Deux modèles ont été proposés pour modéliser l'évolution des fréquences d'avalanches dans les Alpes en tenant compte des deux zones climatiques présentes (une au nord et l'autre au sud). Les deux modèles utilisent les mêmes données, mais leur approche tout comme leurs résultats diffèrent. Alors que le premier, considère les zones climatiques comme fixes, dans le second elles sont le résultat d'un modèle de classification spatiale des communes sur la base de leur série temporelle. Nous proposons ici de concilier ces deux approches, en ajoutant au modèle de classification spatiale de la seconde approche, la connaissance d'expert sur la localisation des zones, sous la forme d'un prior, sous le paradigme bayésien. L'élicitation des paramètres est réalisée en proposant des cartes à l'expert, qu'il juge en faisant appel à des éléments climatiques mais aussi à la réponse de l'activité avalancheuse à des conditions météorologiques données. Cette approche nous permet de retrouver des classes spatialement cohérentes avec des évolutions temporelles distinctes: une zone nord-ouest d'activité décroissante et une zone sud-est d'activité croissante. En ajoutant l'altitude dans le modèle, nous montrons que ces zones résultent de l'interaction du climat local avec l'altitude.

## **Elicitation d'expert, classification spatiale, avalanches**

**Abstract.** Estimating local snow avalanche evolution under climate change is crucial to provide accurate long term forecast. In the French Alps, it is assumed that two climatic zones (one at the north and one at the south) exist and have different temporal evolution of avalanche frequencies. Both used the same data, but their approaches, just as their results are different. One considers fixed zones, whereas for the second, the zones are the result of a township clustering on the base of their temporal evolutions. Here, we propose to conciliate the two approaches, by adding to the spatial-clustering model of the second approach extra-data knowledge on the form of an expert prior on the zones location, under the Bayesian paradigm. The expert is asked to make a judgement about maps, he considers for that the climatic zones but also the avalanche activity response of

townships to specific weather conditions. This intermediate approach proves to be benefit since we obtain spatially coherent cluster, with still two divergent trends: a north-west zone with decreasing avalanche activity and a south-east zone of increasing activity. By considering the altitude in the model, we show that the zones obtained result from the interaction of local climate with altitude.

## Expert elicitation, spatial clustering, avalanche

# 1 Contexte

Prévoir l'évolution des risques naturels dans un contexte changements climatiques est un sujet de préoccupation actuel, et l'estimation et l'analyse des tendances passées constitue une première étape vers une prévision à long terme réaliste de l'aléa climatique. Nous nous intéressons ici à l'activité avalancheuse et plus particulièrement aux fréquences annuelles des avalanches depuis le milieu du XX<sup>e</sup> siècle, sur l'ensemble des Alpes françaises.

En France, l'enquête permanente sur les avalanches (EPA) relève les évènements avalancheux depuis le début du XX<sup>e</sup> siècle en Savoie, puis s'étend à l'ensemble des Alpes à partir des années 50. Par sa taille, cette base est unique, mais des erreurs d'observations persistent. L'analyse conjointe de l'ensemble des données permet alors de limiter l'influence du bruit de mesure, et ainsi d'obtenir des tendances robustes. Cependant les Alpes constituent en France une grande région hétérogène, s'étendant du nord au sud sur environ 260 km et présentant des altitudes supérieures à 4800m. De plus les climatologues reconnaissent deux zones climatiques d'influences différentes dans les Alpes, une zone nord, influencée par des flux de nord-ouest et une zone sud influencée par un mélange de flux d'ouest et de sud (Durand et al, 2009 [1]).

Devant cette constatation, deux modèles pour le nombre d'avalanches par commune et par an  $Y_{ct}$ , où  $c \in \{1, \dots, N\}$  indice la commune et  $t \in \{1, \dots, T\}$  indice l'année, ont été proposés. Dans ces modèles,  $Y_{ct}$  est supposé suivre une distribution binomiale négative d'espérance  $\lambda_{ct}$ , somme d'un terme fixe  $E_c$ , lié au nombre de couloirs dans la commune  $c$ , et de termes spatial et temporel indépendants. De plus, ils supposent qu'il existe deux zones climatiques caractérisées par deux évolutions temporelles distinctes et indépendantes,  $\eta_t^{TN}$  pour la zone nord et  $\eta_t^{TS}$  pour la zone sud, avec  $B_c$  la variable binaire d'affectation de la commune  $c$  dans la zone nord,

$$\log(\lambda_{ct}) = \log(E_c) + \mu + \eta_c^S + B_c \eta_t^{TN} + (1 - B_c) \eta_t^{TS}. \quad (1)$$

Alors que Eckert et al, (2012) [2] considèrent les zones fixes et délimitées par la frontière climatique donnée par Durand et al (2009) [1], Lavigne et al, (2012) [3] laissent les communes se répartir en deux groupes selon leur évolution temporelle et leur localisation. La variable  $B_c$  est alors modélisée par une régression probit dépendant des coordonnées

$(x_c, y_c)$  du centroïde de la commune  $c$ ,

$$B_c \sim \text{Bernoulli}(p_c) \quad (2)$$

$$\text{probit}(p_c) = \beta_1 + \beta_2 x_c + \beta_3 y_c + \beta_4 x_c^2 + \beta_5 y_c^2.$$

Bien que ces deux études aient le même objectif, les résultats obtenus divergent. Avec le modèle de classification, des groupes de communes disloqués sont obtenus et les zones sous-jacentes, représentées par le champ spatial  $p_c$ , sont séparées par un axe SSO-NNE. Les évolutions temporelles de chacun des groupes sont opposées, dans la zone nord-ouest les fréquences diminuent depuis les années 1980, alors que dans la zone sud-est les fréquences augmentent fortement sur la même période. Au contraire, les tendances obtenues pour chacune des zones par Eckert et al, (2012) [2] sont corrélées.

Nous proposons d’associer ces deux approches en ajoutant au modèle développé par Lavigne et al, (2012)[3], la connaissance d’expert sur la localisation des zones climatiques.

## 2 Une méthode d’éllicitation de la connaissance d’expert sur la localisation des zones

Nous nous intéressons ici à la localisation des deux zones climatiques, dans un premier temps nous montrons l’importance du paramètre  $\beta$  dans la classification des communes, puis nous proposons une méthode d’éllicitation pour ce paramètre.

La régression spatiale (Eq. 2) représente la limite entre les zones nord et sud. Pour chaque vecteur de paramètre  $\beta$  une carte de la probabilité d’appartenance à la zone nord  $p_c$  peut-être dessinée (Fig. 1), elle se caractérise par

- le tracé de la limite entre nord et sud, c’est à dire l’ensemble des points pour lesquels  $p_c = 0.5$ ,
- la largeur de la zone de transition, c’est à dire la région pour laquelle l’attribution d’un point à une des deux classes est incertaine,  $p_c$  appartient à l’intervalle  $[0.1, 0.9]$ .

Le paramètre  $\beta$  définit donc à la fois le tracé de la limite et l’importance du caractère spatial dans la classification.

Nous cherchons mettre une loi de probabilité représentant la connaissance de l’expert sur le paramètre  $\beta$ . Cependant le paramètre n’a pas de sens physique pour l’expert, et sa dimension, 5, nécessite l’éllicitation d’une loi conjointe, bien trop compliquée dès lors que la dimension dépasse 2.

Nous simplifions donc le problème en supposant que  $\beta$  suit une distribution de Zellner:

$$\beta \sim N(\mathbf{m}, d(\mathbf{X}'\mathbf{X})^{-1}), \quad (3)$$

avec  $\mathbf{X}$  la matrice de design de la régression probit (Eq. 2). Cette distribution a l'avantage de ne dépendre que de deux paramètres, et est souvent utilisée pour l'analyse de plans d'expériences, car elle permet de ne pas introduire de connaissance sur la structure de covariance.

L'expert avec lequel nous travaillons a participé à l'élaboration du modèle, il est donc d'accord avec l'hypothèse des deux zones. Pour définir ces zones, il utilise des éléments climatiques, mais aussi ses connaissances spécifiques sur les avalanches, notamment la réponse en terme d'activité avalancheuse à des conditions météorologiques données, il fournit en quelque sorte une climatologie des avalanches.

L'élicitation de la moyenne est réalisée à partir d'un système d'information géographique, l'expert dessine la carte de probabilité d'appartenance à la zone nord, à l'aide de polygones de valeurs 0.999, 0.9, 0.7, 0.5 et symétriquement, 0.3, 0.1 et 0.001. Ces valeurs représentent le degré de certitude d'appartenance à la zone nord: sans aucun doute (0.999), presque sûrement (0.9), certainement (0.7) et sans avis (0.5). En croisant cette couche de polygones avec la couche des communes, on obtient une valeur par commune  $p_c^{obj}$  et l'on détermine enfin  $\mathbf{m}$  en minimisant en  $m$   $\sum_c (\Phi(X_c m) - p_c^{obj})^2$ . La carte retenue est présentée à l'expert pour validation.

L'élicitation du paramètre de variance  $d$  est une étape plus délicate. Pour un jeu de cartes simulées d'après (3), plus  $d$  est grand, plus la variabilité entre les cartes est grande. L'expert est simplement capable de dire si une carte est acceptable ou non. Pour une valeur de  $d$  fixée, on présente à l'expert un jeu de 50 cartes, celles-ci sont classées en 3 catégories: acceptables, non acceptables et douteuses si l'expert hésite entre les deux catégories précédentes. On considère finalement que l'on accepte environ 5% de cartes non acceptables dans l'échantillon, et on prend alors  $d = 5$ .

Les valeurs obtenues pour  $c$  sont présentées dans le tableau 4 et la carte de l'espérance *a priori* sur la figure 1(a). Ce modèle défini en section 1 et utilisé avec ce prior est nommé *M1*.

### 3 Des zones d'activité avalancheuses marquées, et des évolutions temporelles différentes

On note une modification de la frontière sur la carte de probabilité d'appartenance à la zone nord (Fig. 1(b)), *a posteriori* la frontière est décalée au nord et a une direction SO-NE. Les évolutions temporelles correspondantes sont proches de celles observées par Lavigne et al, (2012)[3], au nord une diminution des fréquences à partir du milieu des années 1980, et au sud une augmentation régulière depuis les années 1950 (Fig. 2).

En ajoutant l'altitude moyenne  $h_c$  de la commune  $c$  à la régression Eq. 2 (modèle *M1A*), les espérances *a priori* et *a posteriori* des composantes de  $\beta$  coïncident (Tab. 4) et le coefficient lié à l'altitude, négatif, ajoute à la classe nord les communes de basses altitudes. Cela nous laisse supposer que la distribution des classes sur la base de l'activité

Paramètres	Prior	Posterior <i>M1</i>	Posterior <i>M1A</i>
$\beta_1$	1.373 (0.1918)	-1.312 ( $9.240 \cdot 10^{-2}$ )	1.191 ( $7.404 \cdot 10^{-2}$ )
$\beta_2$	$-2.325 \cdot 10^{-2}$ ( $5.843 \cdot 10^{-3}$ )	$-3.693 \cdot 10^{-2}$ ( $2.300 \cdot 10^{-3}$ )	$-2.535 \cdot 10^{-2}$ ( $1.799 \cdot 10^{-3}$ )
$\beta_3$	$5.152 \cdot 10^{-2}$ ( $2.105 \cdot 10^{-3}$ )	$6.238 \cdot 10^{-2}$ ( $1.184 \cdot 10^{-3}$ )	$4.684 \cdot 10^{-2}$ ( $8.008 \cdot 10^{-4}$ )
$\beta_4$	$-1.172 \cdot 10^{-4}$ ( $9.322 \cdot 10^{-5}$ )	$-1.171 \cdot 10^{-4}$ ( $5.116 \cdot 10^{-5}$ )	$-2.305 \cdot 10^{-4}$ ( $3.833 \cdot 10^{-5}$ )
$\beta_5$	$2.212 \cdot 10^{-4}$ ( $3.381 \cdot 10^{-5}$ )	$3.128 \cdot 10^{-4}$ ( $1.833 \cdot 10^{-5}$ )	$2.268 \cdot 10^{-4}$ ( $1.260 \cdot 10^{-5}$ )
$\beta_6$			$-1.503 \cdot 10^{-3}$ ( $2.517 \cdot 10^{-4}$ )

Table 1: Espérance *a priori* et *a posteriori* du paramètre  $\beta$ , pour les modèles *M1* et *M1A*. L'écart-type est donné entre parenthèse.

avalancheuse est le résultat de l'interaction entre la répartition des altitude avec les variations spatio-temporelle du climat. Ainsi les communes de basses altitudes principalement au nord-ouest verraient leur hauteur de neige décroître (Durand et al, 2009b [1]), en même temps que leur activité avalancheuse. Au sud, les fortes précipitations de neige ainsi que les vagues hivernales de chaleur causeraient plus d'avalanches. Bien sur, cette vision est très simplifiée, et nous devons aussi considérer un possible biais d'échantillonnage, ainsi que la présence de flux météorologiques spécifiques tels que les retours d'est.

## 4 Conclusion

Nous montrons dans cette étude que l'ajout d'information sous la forme de connaissance experte, permet d'obtenir des résultats plus cohérents, en limitant le poids des communes dont les données paraissent douteuses. Nous avons alors pu montrer l'importance de l'altitude dans la classification.

Les évolutions contraires mises en évidence soulignent l'importance de telles études sur l'évolution des risques naturels sous le changement climatique.

## Bibliographie

- [1] Y. Durand, G. Giraud, M. Laternser, P. Etchevers, L. Mérindol, and B. Lesaffre (2009), Reanalysis of 47 years of climate in the French Alps (1958-2005): Climatology and trends for snow cover. *Journal of Applied Meteorology and Climatology*, 48(12):2487-2512.
- [2] N. Eckert, C. J. Keylock, H. Castebrunet, A. Lavigne, and M. Naaim (2013), Temporal trends in avalanche activity in the French Alps and subregions: from occurrences and runout altitudes to unsteady return periods. *Journal of Glaciology*.
- [3] A. Lavigne, L. Bel, E. Parent, and N. Eckert (2012), A model for spatio-temporal clustering using multinomial probit regression: application to avalanche counts. *Environmetrics* 23(6):522-534.

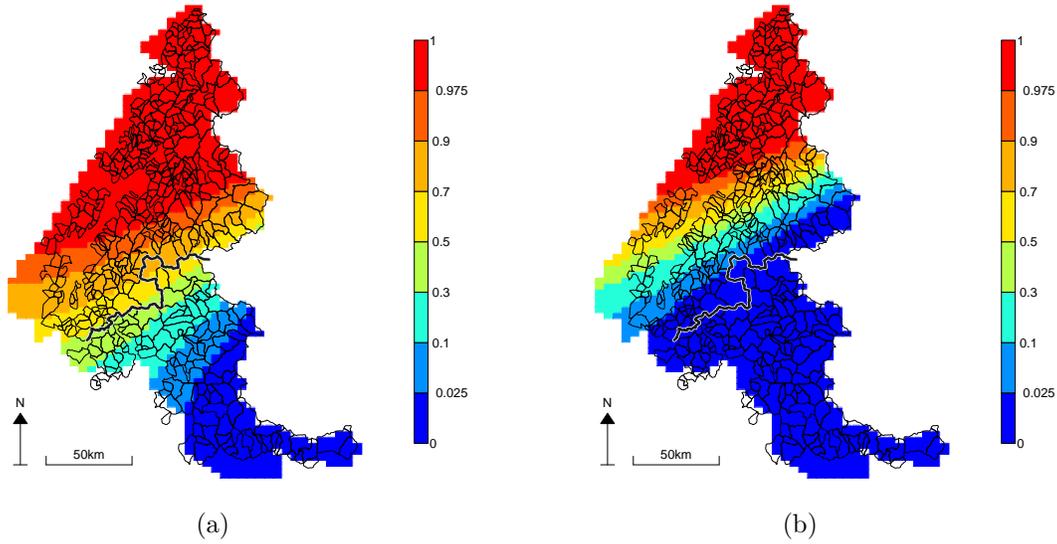


Figure 1: Carte de probabilité d'appartenance à la zone nord: (a) espérance *a priori*; (b) espérance *a posteriori* avec le modèle  $M1$ . La ligne épaisse représente la limite climatique nord/sud utilisée par Eckert et al, (2012) [2].

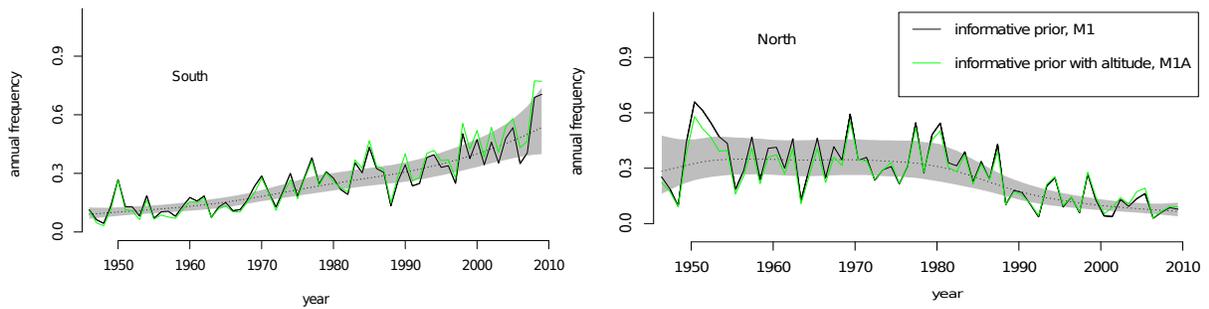


Figure 2: Fréquences annuelles d'avalanches par couloir et par année. En gris, l'intervalle de probabilité à 95% du terme lisse, en pointillé: l'espérance du terme lisse. En trait plein: l'estimation de l'évolution temporelle, en noir avec le modèle  $M1$ , en vert avec le modèle  $M1A$ .