

Calibration probabiliste d'un modèle de propagation d'avalanches pour le calcul de périodes de retour, le zonage et l'optimisation d'ouvrages de protection

Nicolas Eckert, Eric Parent, Mohamed Naaim

► To cite this version:

Nicolas Eckert, Eric Parent, Mohamed Naaim. Calibration probabiliste d'un modèle de propagation d'avalanches pour le calcul de périodes de retour, le zonage et l'optimisation d'ouvrages de protection. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. 2009. <inria-00386583>

HAL Id: inria-00386583

<https://hal.inria.fr/inria-00386583>

Submitted on 22 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Calibration probabiliste d'un modèle de propagation d'avalanches pour le calcul de périodes de retour, le zonage et l'optimisation d'ouvrages de protection

Nicolas Eckert[1] , Eric Parent[2] & Mohamed Naaim[1]

[1] UR ETNA, Cemagref Grenoble, BP 76, 38 402 Saint Martin d'Hères, France

Tel : +33(0)476762222 +33(0)476513803, e-mail : nicolas.eckert@cemagref.fr

[2] Equipe MORSE, UMR 518 AgroParisTech/INRA, 19 avenue du Maine
F-75732 Paris cedex 15, France

Résumé

La prédétermination des avalanches est un problème d'extrêmes multivariés. En ingénierie, la notion d'aléa de référence et son corollaire, la période de retour, sont souvent utilisés de manière peu académique pour y répondre. L'objet de ce travail est de développer un cadre rigoureux grâce au couplage entre modélisation numériques et données de terrain. L'assemblage des différentes sources d'information et la quantification des incertitudes sont effectuées grâce au formalisme bayésien hiérarchique. Celui ci permet la calibration d'un modèle numérique de propagation complexe grâce aux méthodes MCMC. En prédiction, la période de retour est définie en tant que fonction univoque de la distance d'arrêt. Elle permet d'évaluer l'ensemble des scenarii de référence correspondant aux valeurs de projet choisies. Enfin, dès lors que les conséquences dommageables du phénomène en fonction de son intensité sont évaluées, un calcul décisionnel peut être mené pour le dimensionnement d'ouvrages de protection. Celui-ci est illustré par l'optimisation d'une digue paravalanche.

Abstract

Avalanche predetermination is a multivariate extreme statistical problem. For engineering, design values and return periods are often used in a questionable manner. The aim of this work is to propose a rigorous framework coupling numerical modelling and field data. The different information and uncertainty sources are brought together and quantified using hierarchical Bayesian modelling schemes. MCMC simulations are used to perform the on site calibration of a complex numerical avalanche model. In the predictive phase, return period is defined as a one-to-one mapping of the runout distance. This allows evaluating all reference scenarios corresponding to the chosen design values. Finally, as soon as hazard consequences are quantified as a function of its magnitude, a decisional approach can be used for the design of defence structures. This is illustrated by the optimisation of an avalanche dam.

Mots clés

Avalanches, Extrêmes, Période de Retour, Risque, Méthodes Statistiques-Dynamiques, Distance d'Arrêt, Optimisation.

Aléa de référence et période de retour

En ingénierie des risques naturels, le problème de la prédétermination est généralement traité par le biais de la définition d'un aléa de référence, c'est à dire une réalisation d'intensité élevée du phénomène étudié, par exemple un débit de crue de projet. L'aléa de référence est souvent exprimé sous la forme d'une période de retour. Celle-ci quantifie, sous hypothèse de répétabilité du phénomène et d'indépendance des occurrences, l'intervalle de temps moyen en années séparant deux événements d'intensité au moins égale à celle de l'aléa de référence. Elle permet de définir des seuils d'aléas pour le dimensionnement d'ouvrages de protection et le zonage, c'est-à-dire le découpage de l'espace en secteurs assortis de prescriptions en matière d'urbanisme.

Si $F(z)$ est la fonction de répartition annuelle du phénomène Z considéré, la période de retour T_z associée à l'aléa de référence z est l'inverse de la probabilité annuelle de dépassement de la valeur z (eq. 1). Le calcul de périodes de retour est un problème d'extrêmes dans la mesure où il s'agit d'extrapoler les événements les plus graves à partir de séries de données relativement courtes. Les modèles classiques permettant d'y répondre, de types POT ou GEV (Coles, 2001), sont bien connus en hydrologie.

$$T_z = \frac{1}{1 - F(z)} \text{ (eq. 1)}$$

L'aléa avalancheux et ses spécificités

Le problème d'extrêmes posé par la prédétermination des avalanches est particulier dans la mesure où la variable la plus critique, la distance d'arrêt, c'est-à-dire la distance atteinte par l'avalanche, n'est pas une variable intrinsèque. Le comportement de sa queue de distribution dépend de la topographie. Il n'est donc pas possible d'utiliser une approche purement statistique de type GEV ou POT pour prédire les distances d'arrêt correspondant à des périodes de retour très élevées.

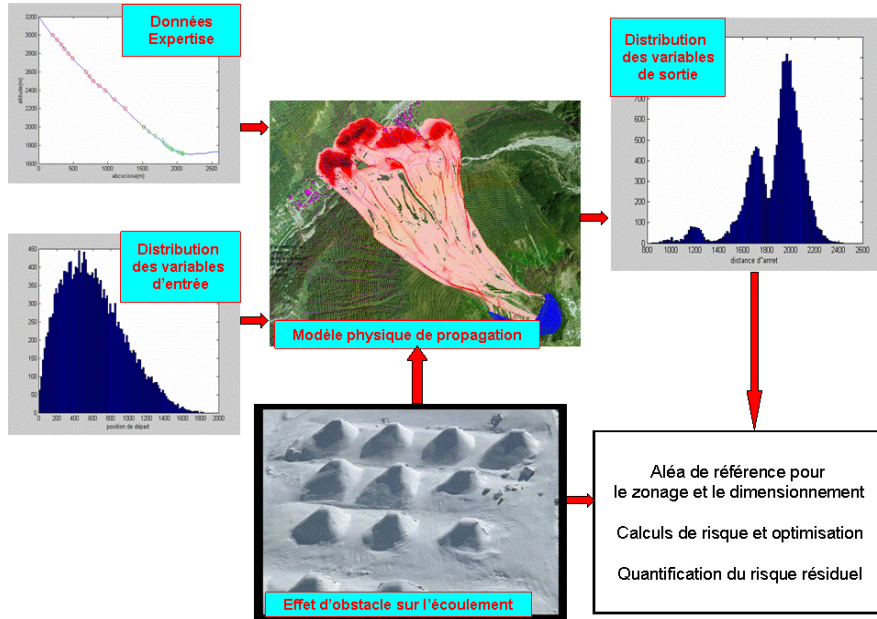
En outre, l'aléa avalancheux est réellement multivarié. L'extension de l'aléa, les champs de vitesse et de pression, le volume de dépôt, etc. doivent tous être quantifiés, et ce alors que les observations montrent que ces variables sont souvent mal corrélées. Une avalanche à faible distance d'arrêt peut, par exemple, infliger de très fortes pressions d'impact avec une neige dense et très cohésive.

Enfin, les avalanches sont particulièrement mal connues en tant que phénomène physique. La connaissance du lien entre contrainte et déformation reste par exemple encore très imparfaite du fait d'observations de terrain rares et entachées de fortes incertitudes. En revanche, la propagation de la neige sur une topographie donnée est relativement bien maîtrisée grâce à des codes numériques déterministes basés sur la résolution des équations élémentaires de la mécanique. Une approche purement physique n'est donc également pas envisageable, mais la modélisation numérique peut être associée à la modélisation probabiliste de façon à évaluer les périodes de retour recherchées en tenant compte de la topographie.

Les méthodes statistiques-dynamiques

A partir de 2001, des méthodes de simulation de type Monte Carlo ont été proposées pour la prédétermination des avalanches (Barbolini et Keylock, 2002; Meunier et Ancey, 2004; Bozhinskiy, 2004). Leur principe est d'alimenter un modèle de propagation déterministe avec des données d'entrée tirées dans des distributions aléatoires supposées parfaitement identifiées et d'obtenir ainsi la distribution des distances d'arrêt et des autres variables de sortie. Outre l'obtention aisée d'avalanches de période de retour élevée, les applications des méthodes statistiques dynamiques sont très nombreuses : calcul de risque, dimensionnement d'ouvrages, quantification du risque résiduel (Figure 1). Néanmoins, les résultats obtenus sont étroitement dépendants des hypothèses probabilistes effectuées, notamment le choix des distributions des données d'entrée et de leur éventuelles corrélations.

Figure 1 : Approche statistique-dynamique pour la prédétermination et le calcul de risque



Modélisation numérique bidimensionnelle de l'avalanche de Montroc, Haute-Savoie, France.
Tas freineurs de Tacconnaz, Haute-Savoie, France, photo F. Valla/ Cemagref.

Distance d'arrêt, période de retour et scénarii de référence

De façon formelle, dans une approche statistique-dynamique, chaque avalanche i est décrite par un couple de vecteurs (x_i, y_i) . Les x_i sont les variables « d'entrées » caractérisant le déclenchement et la propagation : surface de départ, altitude de départ, volume de neige mobilisé, qualité de la neige..., etc. Les y_i correspondent aux variables dites « de sortie » présentant un intérêt pour le zonage et le dimensionnement : distance d'arrêt, champs de vitesse et de pression..., etc. Le calcul de probabilité annuelle nécessite en outre les nombres d'avalanches annuels a_i sur une période aussi longue que possible. Par simulation, la loi de probabilité jointe $p(x, y, a | \hat{\theta})$ est obtenue en

déconditionnant selon la physique du phénomène (eq. 2). Le conditionnement par rapport à $\hat{\theta}$ indique que la distribution obtenue dépend d'un vecteur de paramètres (coefficients de frottement du modèle d'avalanche et paramètres indexant les distributions des différentes variables d'entrée) pour lequel un estimateur ponctuel $\hat{\theta} = \left(\hat{\theta}_M, \hat{\theta}_F \right)$ est supposé connu. L'estimation ponctuelle de la période de retour $T_{x_{stop}}$ associée à la distance d'arrêt x_{stop} (eq. 3) s'obtient aisément grâce aux méthodes statistiques-dynamiques par combinaison du nombre moyen estimé d'avalanches par an pour le couloir considéré $E \left[A | \hat{\theta}_F \right]$, et celui de la fonction de répartition de la distance d'arrêt

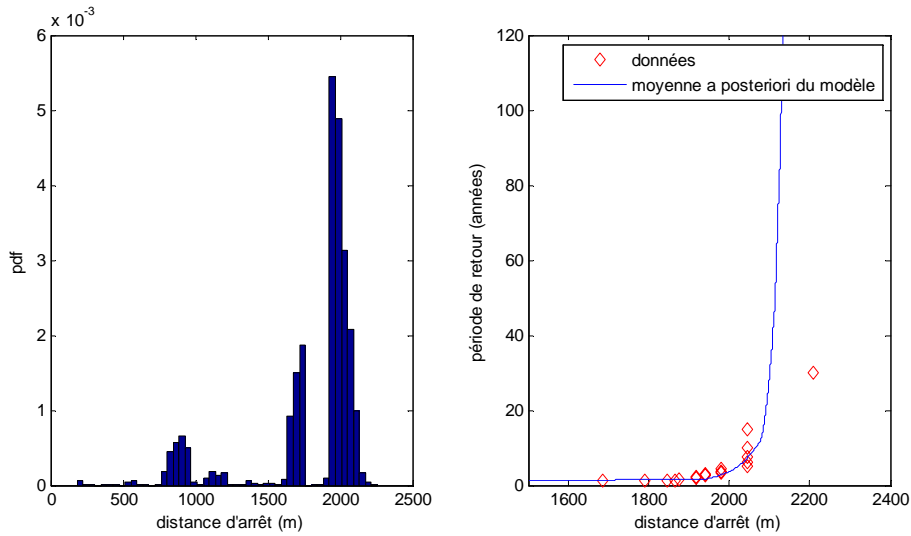
$\hat{F}(x_{stop}) = P \left(X_{stop} \leq x_{stop} | \hat{\theta}_M \right)$. La Figure 2 présente un exemple de calcul sur un cas concret.

$$p(x, y, a | \hat{\theta}) = p(y | a, x, \hat{\theta}) \times p(x | a, \hat{\theta}) \times p(a | \hat{\theta}) \quad (\text{eq. 2})$$

$$T_{x_{stop}} = \frac{1}{E \left[A | \hat{\theta}_F \right] \times \left(1 - \hat{F}(x_{stop}) \right)} \quad (\text{eq. 3})$$

N'évaluer la période de retour que pour la distance d'arrêt ne dispense pas de quantifier sérieusement les autres variables nécessaires au zonage et au dimensionnement, en particulier la vitesse v , la pression pr et le volume de dépôt v_{stop} . Pour ce faire, il est possible, avec les méthodes statistiques-dynamiques, d'obtenir la distribution $p(x_{stop}, v, pr, v_{stop} | x_{stop} > x_{stop_T}, \hat{\theta})$ pour n'importe quelle distance d'arrêt x_{stop_T} en ne considérant que les événements qui dépassent la distance d'arrêt considérée. On quantifie ainsi tous les scénarii de référence correspondent à une même période de retour. Bien entendu, il y en, au moins en théorie, une infinité. En contrepartie, étant donné qu'ils sont tous évalués, il est facile de balayer leur gamme de variabilité, ce qui rend possible, pour une expertise, de retenir les plus pertinents, par exemple ceux correspondant aux bornes de l'intervalle de confiance à 95% de la loi jointe $p(x_{stop}, v, pr, v_{stop} | x_{stop} > x_{stop_T}, \hat{\theta})$.

Figure 2 : Distribution des distances d'arrêt et période de retour associée



Couloir numéro13 de la commune de Bessans, Savoie, France.
Distribution des distances d'arrêt simulées et comparaison aux observations.

Calibration et prédiction bayésienne

La solution la plus naturelle au problème du choix crucial de la distribution des variables d'entrées $p(x | a, \hat{\theta})$ est d'estimer θ à partir de données terrain. L'inférence bayésienne est bien adaptée à ce problème grâce aux méthodes dites MCMC (Gilks et al., 2001 ; Brooks, 1998) d'estimation par simulation (Eckert et al., 2007 ; Eckert et al., 2008). Dans le théorème de Bayes, (eq. 4), $p(\theta)$ est la distribution de probabilité *a priori* du vecteur de paramètres θ . L'appellation « *a priori* » exprime le fait qu'elle a été établie préalablement à l'observation des données notées *data* et qu'elle permet ainsi éventuellement l'utilisation de connaissance « experte » de type naturaliste. $l(x, y, a | \theta)$ est la fonction de vraisemblance quantifiant la crédibilité des différentes observations effectuées au travers du modèle numérique de propagation utilisé. $p(\theta | data)$ est la distribution de probabilité *a posteriori* de θ quantifiant l'incertitude résiduelle sur les paramètres liée à la quantité limitée de données disponibles.

Dans le cadre Bayésien, l'incertitude entourant la distance de projet x_{stop_T} correspondant à la période de retour T est quantifiée par sa loi prédictive $p(x_{stop_T} | data)$. Celle-ci s'obtient en moyennant x_{stop_T} sur la loi *a posteriori* des paramètres (eq. 5). Elle prend en compte à la fois

l'incertitude touchant l'intensité et la fréquence des avalanches sur le site étudié. Sa dispersion dépend de la variabilité de la loi *a posteriori* des paramètres (θ_M, θ_F) et donc de la quantité de données disponibles pour l'inférence. Asymptotiquement, elle converge vers la vraie valeur ponctuelle de x_{stop_T} .

$$p(\theta | data) \propto p(\theta) \times l(x, y, a | \theta) \quad (\text{eq. 4})$$

$$p(x_{stop_T} | data) = \int F_{x_{stop} | \theta_M}^{-1} \left(1 - \frac{1}{E[A | \theta_F] \times T} \right) \times p(\theta_M, \theta_F | data) \times d\theta_M \times d\theta_F \quad (\text{eq. 5})$$

Calcul de risque pour l'optimisation d'un ouvrage paravalanche

La mise en œuvre d'une analyse décisionnelle nécessite la spécification d'un coût $C(d, y, a)$ fonction de l'aléa en zone d'arrêt et d'une possible variable de décision d , par exemple la hauteur d'un ouvrage de protection h_d . Il dépend de la vulnérabilité des enjeux, de leur valeur ainsi que de leur exposition et peut éventuellement être pondéré par le comportement du décideur face au risque. Le coût s'exprime comme la somme d'un terme de construction $C_o(h_d)$ dépendant de la hauteur de l'ouvrage et d'un terme quantifiant les dommages totaux infligés aux enjeux par les avalanches successives se produisant à partir de la date de construction. Ce dernier dépend des effets de l'ouvrage sur l'écoulement (Faug et al., 2008) Sous des hypothèses peu restrictives, le coût total se ramène à l'équation 6 où A_{ct} représente l'actualisation totale, C_o le coût linéaire de construction, C_1 la valeur totale des enjeux considérés et $E_d[a | \theta_F, \theta_M]$ la fréquence moyenne de dépassement de la position de l'ouvrage. Enfin, pour une quantification facile de la pertinence de toute décision, un état de référence quantifiant le coût de ne rien faire est introduit (eq. 7), ce qui définit une fonction d'utilité linéaire $u(h_d, y, a)$.

$$C(h_d, y, a) = C_o h_d + A_{ct} \times E_d[a | \theta_F, \theta_M] \times C_1(h_d, y) \quad (\text{eq. 6})$$

$$u(h_d, y, a) = C_o h_d + A_{ct} \times E_d[a | \theta_F, \theta_M] \times (C_1(h_d, y) - C_1(0, y)) \quad (\text{eq. 7})$$

Le risque est l'espérance de l'utilité par rapport à la variabilité de l'aléa. C'est une fonction de la variable de décision et des paramètres du modèle quantifiant la perte moyenne attendue au point de l'espace considéré si la décision de construire un ouvrage de hauteur h_d est prise plutôt que de ne rien faire (eq. 8). Au sens de la théorie de la décision classique, la décision optimale h_C^* maximise

ce bénéfice (eq. 9). Pour ce faire, il est nécessaire d'utiliser les estimateurs ponctuels $(\hat{\theta}_M, \hat{\theta}_F)$,

puis de recommander par « plug-in » la hauteur minimisant le risque conditionnellement à ces estimateurs. Tout comme le calcul des aléas de référence, les calculs d'optimisation peuvent être menés numériquement par simulations.

$$R_C(h_d, \theta_M, \theta_F) = C_o h_d + A_{ct} \times E_d[a | \theta_F] \times \int (C_1(h_d, y) - C_1(0, y)) \times p(y | \theta_M) \times dy \quad (\text{eq. 8})$$

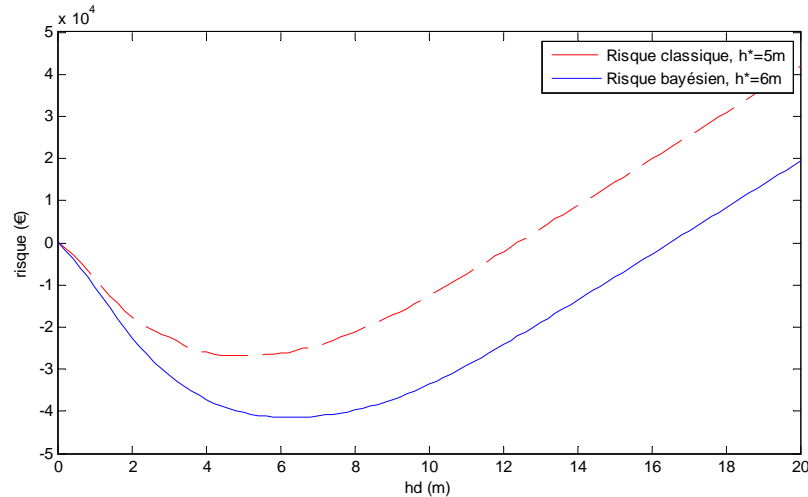
$$h_C^* = \text{Arg min}_{h_d} \left(R_C \left(h_d, \hat{\theta}_M, \hat{\theta}_F \right) \right) \quad (\text{eq. 9})$$

Dissocier l'inférence du problème décisionnel peut toutefois avoir des conséquences indésirables. En effet, le phénomène est ainsi traité pour la décision comme s'il était parfaitement connu, ce qui est très irréaliste dans la pratique dès lors que l'inférence a été effectuée avec une quantité faible de données. Le risque bayésien $R_B(h_d)$ prend en compte l'incertitude liée à la méconnaissance locale pour la décision en moyennant le risque classique sur toute la loi *a posteriori* des paramètres (eq. 10). L'erreur d'estimation est ainsi incorporée de façon claire dans la décision. Le risque bayésien est une fonction de la variable décisionnelle uniquement, ce qui rend la détermination de la

décision bayésienne optimale facile. La Figure 3 présente les fonctions de risque obtenues sur un cas concret, et notamment le caractère plus pessimiste du risque bayésien conduisant à un dimensionnement plus prudent (Eckert et al., 2009).

$$R_B(h_d) = \int R_C(h_d, \theta_M, \theta_F) \times p(\theta_M, \theta_F | data) \times d\theta_M \times d\theta_F \quad (\text{eq. 10})$$

Figure 3 : Optimisation d'une digue paravalanche verticale



Couloir numéro 13 de la commune de Bessans, Savoie, France. La digue est située à une abscisse bisannuelle. L'enjeu, un bâtiment isolé, est situé à une abscisse centennale. En ordonnée, le risque correspond à l'opposé du bénéfice moyen attendu de la digue considérée.

Bibliographie

- [1] Barbolini, M., Keylock, C.J. (2002). A new method for avalanche hazard mapping using a combination of statistical and deterministic models. *Natural Hazard and Earth System Sciences*. 2. pp 239-245.
- [2] Bozhinskiy, A.N. (2004). The Monte Carlo simulation of avalanche-type processes. *Annals of Glaciology* 38. pp 351-356.
- [3] Brooks, S. P. (1998). Markov Chain Monte Carlo Method and its application. *The Statistician*. Vol 47, 1. pp 69-100.
- [4] Coles, S. (2001). An introduction to statistical modelling of extreme values. Springer ed. 208p.
- [5] Eckert, N., Parent, E., Richard, D. (2007). Revisiting statistical-topographical methods for avalanche predetermination: Bayesian modelling for runout distance predictive distribution. *Cold Regions Science and Technology*. 49. pp 88-107.
- [6] Eckert, N., Parent, E., Naaim, M., Richard, D. (2008) Bayesian stochastic modelling for avalanche predetermination: from a general system framework to return period computations. *Stochastic Environmental Research and Risk Assessment*. 22. pp 185-206.
- [7] Eckert, N., Parent, E., Faug, T., Naaim, M. (2009). Bayesian optimal design of an avalanche dam using a multivariate numerical avalanche model. *Stochastic Environmental Research and Risk Assessment*. Accepted.
- [8] Faug, T., Gauer P., Lied K., Naaim, M. (2008). Overrun length of avalanches overtopping catching dams: Cross-comparison of small-scale laboratory experiments and observations from full-scale avalanches. *Journal of Geophysical Research* 113. F03009, doi:10.1029/2007JF000854.
- [9] Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (2001). *Markov Chain Monte Carlo in Practice*. Chapman & Hall ed. 486 p.
- [10] Meunier, M., Ancey, C. (2004). Towards a conceptual approach to predetermining high-return-period avalanche run-out distances. *Journal of Glaciology*. 50-169. pp 268-278.