

MODÈLE GÉNÉRATIF HIÉRARCHIQUE POUR LA RENTRÉE ATMOSPHERIQUE

Pierre Minvielle¹ & Audrey Giremus² & Vivien Loridan³

¹ *CESTA, DAM, CEA, France, pierre.minvielle@cea.fr*

² *IMS (Université de Bordeaux, CNRS, Bordeaux INP), France, audrey.giremus@u-bordeaux.fr*

³ *CESTA, DAM, CEA, France, vivien.loridan@cea.fr*

Résumé. Lorsqu'une navette spatiale ou une capsule rentre dans les couches denses de l'atmosphère, elle décélère fortement tandis qu'une onde de choc se forme en amont du véhicule, provoquant une montée en température et d'importants transferts de chaleur à la paroi. Prévoir l'évolution aérodynamique du véhicule à la rentrée est alors un enjeu essentiel, par exemple pour prédire la zone d'atterrissage. Pour ce faire, il est possible de s'appuyer sur de la simulation numérique multi-physique ainsi que sur quelques expériences de rentrée atmosphérique. Elles renseignent quant à l'erreur de simulation de l'évolution aérodynamique et à sa variabilité. Pour prévoir une future rentrée atmosphérique, il faut recourir à un modèle génératif qui, à partir de quelques courbes observées, en produit de nouvelles. La difficulté vient du nombre faible de données. L'approche proposée s'inspire du modèle KOH (Kennedy O'Hagan), largement répandu en calibration bayésienne de code et reposant sur des processus gaussiens. Cela conduit à un modèle bayésien hiérarchique qui tient compte de la non-stationnarité de l'erreur de modèle. La distribution prédictive peut être efficacement échantillonnée au moyen d'un échantillonneur Monte Carlo Hamiltonien (HMC) dynamique, dénommé No-U-Turn Sampler (NUTS), amplement exploité en statistiques computationnelles. Adapté à des problèmes de dimension conséquente, il va bénéficier des expressions analytiques de la densité cible et de son gradient pour explorer efficacement l'espace. Les échantillons générés, faiblement corrélés, peuvent être par la suite exploités pour prédire l'évolution aérodynamique et la zone d'atterrissage, via l'approximation de Monte Carlo d'intégrales multidimensionnelles en grande dimension, et en particulier, pour produire des intervalles de crédibilités.

Mots-clés. apprentissage statistique, calibration bayésienne, processus gaussien, Monte Carlo Hamiltonien

Abstract. When a space shuttle or capsule re-enters the dense layers of the atmosphere, it decelerates sharply while a shock wave is formed upstream the vehicle, causing a rise in temperature and significant heat transfer to the wall. Predicting the aerodynamic evolution of the vehicle on re-entry is therefore an essential challenge, for example in determining the landing zone. To do this, it is possible to rely on multi-physics numerical simulation and on atmospheric re-entry experiments. They provide information on the simulation error of aerodynamic evolution and its variability. To predict future atmospheric re-entry, we need to use a generative model which, based on the few curves observed, generates new ones. The difficulty lies in the small amount of data available. The proposed approach is inspired by

the KOH (Kennedy O’Hagan) model, widely used in Bayesian code calibration and based on Gaussian processes. This leads to a hierarchical Bayesian model that takes account of the non-stationarity of the model error. The predictive distribution can be efficiently sampled by a dynamic Monte Carlo Hamiltonian (HMC) sampler, the so-called No-U-Turn Sampler (NUTS), widely exploited in computational statistics. Adapted to high-dimensional problems, it will benefit from analytical expressions of the target density and its gradient to efficiently explore space. The weakly correlated generated samples can be used afterwards to predict aerodynamic evolution and the landing zone, via Monte Carlo approximation of high-dimensional multidimensional integrals, and in particular to produce credibility intervals.

Keywords. statistical learning, Bayesian calibration, Gaussian process, Hamiltonian Monte Carlo

1 Introduction

Un aéronef effectuant une rentrée atmosphérique évolue typiquement dans des régimes hypersoniques. Sous de telles conditions, une onde de choc apparaît autour du véhicule, au sein de laquelle l’air incident est fortement comprimé et ralenti. Il en résulte une augmentation drastique de la température, qui génère d’importants transferts de chaleur entre la surface de l’objet et l’écoulement. Les flux thermiques intenses auxquels sont soumis les matériaux de paroi peuvent potentiellement mettre en péril l’intégrité du véhicule. Afin de prévenir ce type de dommages, le système de protection thermique est spécifiquement conçu pour absorber une partie importante de l’énergie incidente via sa dégradation graduelle durant la phase de rentrée atmosphérique. Ce phénomène d’ablation modifie la géométrie du bouclier thermique au cours de la rentrée, ce qui a des conséquences sur les propriétés aérodynamiques du module et, par conséquent, sur sa trajectoire globale. La prédiction de la zone probable d’atterrissage repose ainsi sur deux aspects. D’une part, il s’agit de restituer le plus fidèlement possible les phénomènes multi-physiques impliqués lors de la rentrée atmosphérique (écoulements hypersoniques, turbulence, ablation, évolution thermique au sein des matériaux etc. . .) via le développement et l’utilisation de codes dédiés aux calculs aérothermiques. D’autre part, un couplage à la mécanique du vol est nécessaire afin de mettre à jour la trajectoire conformément aux changements aérodynamiques subis par le véhicule.

En complément, la prédiction peut s’appuyer sur des expériences représentatives de rentrée atmosphérique. Elles permettent de recalibrer les modèles et informent quant à la variabilité aérodynamique. Ci-après, on se place dans un formalisme proche de Kennedy et O’Hagan (2001), répandu en calibration bayésienne de codes. Ce travail présente la prédiction de l’évolution aérodynamique par un modèle génératif hiérarchique. La première partie expose le problème de calibration de code. Les deux parties suivantes décrivent le modèle bayésien hiérarchique et l’approche numérique d’échantillonnage. Enfin, des résultats illustratifs sont fournis.

1.1 Calibration de code

Lors d'une rentrée atmosphérique, une des principales caractéristiques aérodynamiques est le coefficient balistique $\beta = S \cdot C_A / m$, avec S la surface de référence, m la masse du véhicule et C_A le coefficient aérodynamique. Intervenant dans la force de traînée, β détermine la décélération de l'objet lors de la rentrée atmosphérique. Le coefficient β dépend du véhicule : il est fonction de l'entrée $\mathbf{x} \in \mathbb{N}$ dans l'espace des entrées (géométrie du véhicule, matériaux, etc.). Il est de plus fonction de l'altitude h . On peut donc le noter $\beta(\mathbf{x}, h)$. En faisant intervenir un code multi-physique déterministe \mathcal{T} , $\beta(\mathbf{x}, h)$ peut se décomposer de la manière suivante :

$$\beta(\mathbf{x}, h) = \mathcal{T}(\mathbf{x}, h) + \Delta\beta(\mathbf{x}, h) \quad (1)$$

où $\mathcal{T}(\mathbf{x}, h)$ est le coefficient balistique calculé pour l'entrée $\mathbf{x} \in \mathbb{N}$ et l'altitude h , et $\Delta\beta(\mathbf{x}, h)$ est l'erreur de modèle associée (inconnue). En transposant plus spécifiquement à chacune des expériences de rentrée atmosphérique, impliquant un véhicule différent décrit par l'entrée $\mathbf{x}_m \in \mathbb{N}$ ($m = 1 \cdots M$), il en découle :

$$\beta(\mathbf{x}_m, h) = \mathcal{T}(\mathbf{x}_m, h) + \Delta\beta(\mathbf{x}_m, h) \quad (2)$$

A chaque expérience, on observe *indirectement* $\beta(\mathbf{x}_m, h)$, c'est-à-dire après estimation, reconstruction ou restitution, au moyen de capteurs proprioceptifs et extéroceptifs. On suppose que ces observations se font à certaines altitudes discrètes : h_1, h_2, \dots, h_K . Soit \mathbf{y}_m le vecteur des coefficients balistiques observés, il peut se décomposer de la manière suivante :

$$\mathbf{y}_m = \mathcal{T}(\mathbf{x}_m) + \Delta\boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m \quad (3)$$

où $\Delta\boldsymbol{\beta}_m \triangleq [\Delta\beta(\mathbf{x}_m, h_1) \quad \Delta\beta(\mathbf{x}_m, h_2) \quad \cdots \quad \Delta\beta(\mathbf{x}_m, h_K)]^T$ est le vecteur des erreurs de modèle et $\boldsymbol{\varepsilon}_m \triangleq [\varepsilon_m(\mathbf{x}_m, h_1) \quad \varepsilon_m(\mathbf{x}_m, h_2) \quad \cdots \quad \varepsilon_m(\mathbf{x}_m, h_K)]^T$ est le vecteur des erreurs de restitution aux altitudes h_1, h_2, \dots, h_K .

Enfin, on considère une future expérience de rentrée correspondant à l'entrée $\mathbf{x}_\infty \in \mathbb{N}$:

$$\beta(\mathbf{x}_\infty, h) = \mathcal{T}(\mathbf{x}_\infty, h) + \Delta\beta(\mathbf{x}_\infty, h) \quad (4)$$

Comme l'expérience n'a pas encore eu lieu, on ne dispose pas encore d'observation associée. Le problème est le suivant : il s'agit de "calibrer" l'erreur de modèle $\Delta\beta(\mathbf{x}, h)$, c'est-à-dire déterminer quelles sont ses possibles valeurs, et prévoir l'erreur de modèle $\Delta\beta(\mathbf{x}_\infty, h)$, ou de manière équivalente $\beta(\mathbf{x}_\infty, h)$ au regard des observations $\mathbf{y}_{1:M} \triangleq \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M\}$. Ces dernières sont illustrées sur la figure 1 pour $M = 5$ où l'on représente en particulier les courbes d'écart relatifs entre la simulation et l'expérience.

1.2 Problème génératif à partir d'un faible nombre de données

Ce problème peut être reformulé sous la forme d'un problème génératif. Prévoir l'erreur de modèle $\Delta\beta(\mathbf{x}_\infty, h)$ revient à être capable de générer, à partir de ces quelques courbes,

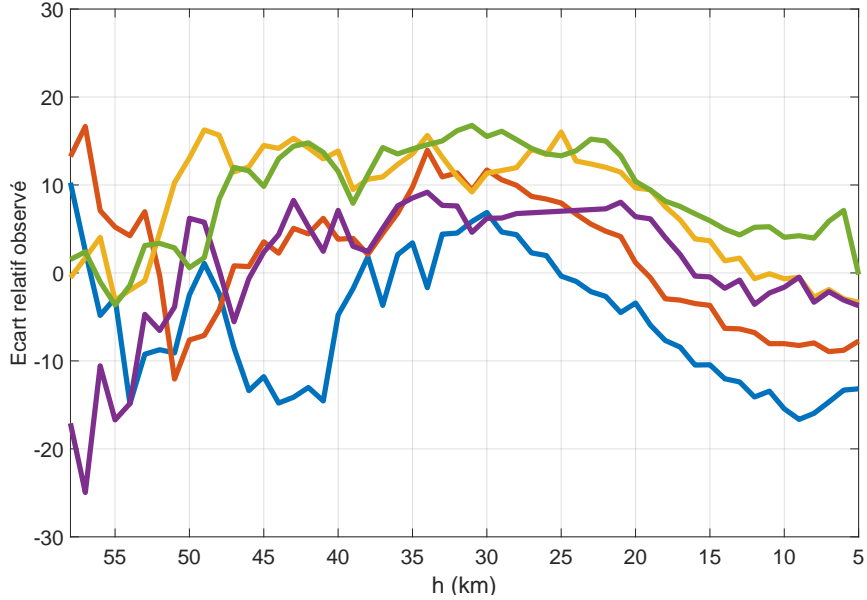


Figure 1: Ecart relatif observé du problème génératif ($M = 5$)

de nouvelles courbes que l'on souhaite représentatives. Autrement dit, d'un point de vue bayésien, on cherche à échantillonner aussi fidèlement que possible la densité *a posteriori* $p(\Delta\beta_\infty | \mathbf{y}_{1:M})$. Ce problème génératif est difficile, en raison du faible nombre de données, et il ne peut être résolu au moyen des modèles génératifs à base de réseaux neuronaux (GAN, VAE, etc.). Sans information supplémentaire, on ne saurait dire par exemple si les seules M courbes sont licites et issues d'une loi discrète, ou si elles sont des réalisations d'une loi continue. Cela nécessite donc de préciser la connaissance *a priori* et de faire le lien entre l'écart futur et les écarts passés $\Delta\beta_m$. Tel est bien le rôle du modèle génératif hiérarchique, décrit ci-après.

2 Modèle génératif hiérarchique

2.1 Processus gaussien

Pareillement au modèle KOH de Kennedy O' Hagan (2001), on opte pour une modélisation par un processus aléatoire gaussien :

$$\Delta\beta(\mathbf{x}, h) \sim \mathcal{PG}(\mu(\mathbf{x}, h), k(\mathbf{x}, h, \mathbf{x}', h')) \quad (5)$$

où $\mu(\cdot, \cdot)$ est la fonction moyenne et $k(\cdot, \cdot, \cdot, \cdot)$ est le noyau.

Stationnarité selon \mathbf{x} Le processus gaussien est supposé stationnaire au sens strict selon \mathbf{x} , c'est-à-dire que $\Delta\beta(\mathbf{x}, h)$ et $\Delta\beta(\mathbf{x}', h)$ ont même distribution de probabilité. A $\mathbf{x} = \mathbf{x}_m$ fixé,

l'erreur de modèle $\Delta\beta(\mathbf{x}_m, h)$ suit un processus gaussien indexé par $h \in \mathbb{R}$ ($m = 1, \dots, M$) :

$$\Delta\beta(\mathbf{x}_m, h) \sim \mathcal{PG}(\mu(h), k(h, h')) \quad (6)$$

où $\mu(\cdot)$ est la fonction moyenne et $k(\cdot, \cdot)$ est le noyau ou fonction de covariance. Ainsi, le processus est stationnaire en fonction de l'expérience, c'est-à-dire que toutes les densités de probabilités sont supposées invariantes selon m . Formulé autrement, les véhicules ont des caractéristiques suffisamment proches pour que les écarts aient des distributions similaires. De même, on a naturellement :

$$\Delta\beta(\mathbf{x}_\infty, h) \sim \mathcal{PG}(\mu(h), k(h, h')) \quad (7)$$

Non-stationnarité selon h Au vu de la figure 1, on choisit de modéliser l'écart par un processus gaussien non-stationnaire en h , au premier et second ordre. Concernant la fonction de covariance, on adopte plus spécifiquement la généralisation non-stationnaire du noyau SE (Squared Exponential) :

$$k(h, h') = \sigma(h)\sigma(h') \sqrt{\frac{2\ell(h)\ell(h')}{\ell(h)^2 + \ell(h')^2}} \exp\left(-\frac{(h-h')^2}{\ell(h)^2 + \ell(h')^2}\right) \quad (8)$$

où $\sigma(\cdot)$ et $\ell(\cdot)$ correspondent respectivement à l'écart-type et à la longueur de corrélation qui varient en fonction de l'altitude h . Un tel modèle fut tout d'abord introduit par Gibbs (1997), puis étendu ultérieurement par Paciorek et Schervish (2004).

On définit les vecteurs aléatoires $\Delta\boldsymbol{\beta}_m$ et $\Delta\boldsymbol{\beta}_\infty$, indépendants, tels que :

$$\Delta\boldsymbol{\beta}_m \triangleq [\Delta\beta(\mathbf{x}_m, h_1) \quad \Delta\beta(\mathbf{x}_m, h_2) \quad \dots \quad \Delta\beta(\mathbf{x}_m, h_K)]^T \quad (9)$$

$$\Delta\boldsymbol{\beta}_\infty \triangleq [\Delta\beta(\mathbf{x}_\infty, h_1) \quad \Delta\beta(\mathbf{x}_\infty, h_2) \quad \dots \quad \Delta\beta(\mathbf{x}_\infty, h_K)]^T \quad (10)$$

avec K le nombre d'altitudes discrétisées. On constate alors que $\Delta\boldsymbol{\beta}_m \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{P})$ et $\Delta\boldsymbol{\beta}_\infty \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{P})$ où $\boldsymbol{\mu} \triangleq [\mu(h_1) \quad \mu(h_2) \quad \dots \quad \mu(h_K)]^T$ est le vecteur moyenne et $\mathbf{P} \triangleq [k(h_i, h_j)]_{i=1 \dots K, j=1 \dots K}$ est la matrice de covariance dont les composantes sont :

$$\mathbf{P}_{ij} = \sigma_i \sigma_j \sqrt{\frac{2\ell_i \ell_j}{\ell_i^2 + \ell_j^2}} \exp\left(-\frac{(h_i - h_j)^2}{\ell_i^2 + \ell_j^2}\right) \quad (11)$$

en notant $\boldsymbol{\sigma} \triangleq [\sigma(h_1) \quad \sigma(h_2) \quad \dots \quad \sigma(h_K)]^T$ et $\boldsymbol{\ell} \triangleq [\ell(h_1) \quad \ell(h_2) \quad \dots \quad \ell(h_K)]^T$.

2.2 Vraisemblance

Les erreurs de mesure, précédemment introduites dans l'eq. (3), sont modélisées par des vecteurs gaussiens : $\boldsymbol{\varepsilon}_m \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{R})$ où \mathbf{R} est la matrice de covariance de mesure. On peut alors écrire :

$$\mathbf{y}_m | \Delta\boldsymbol{\beta}_m \sim \mathcal{N}(\mathcal{T}(\mathbf{x}_m) + \Delta\boldsymbol{\beta}_m, \mathbf{R}) \quad (12)$$

2.3 Modèle bayésien hiérarchique

Pour garantir la positivité des hyperparamètres du modèle de processus gaussien, on introduit les vecteurs aléatoires de travail suivants :

$$\tilde{\boldsymbol{\sigma}} \triangleq [\text{Log}\sigma(h_1) \quad \text{Log}\sigma(h_2) \quad \cdots \quad \text{Log}\sigma(h_K)]^T \quad (13)$$

$$\tilde{\boldsymbol{\ell}} \triangleq [\text{Log}\ell(h_1) \quad \text{Log}\ell(h_2) \quad \cdots \quad \text{Log}\ell(h_K)]^T \quad (14)$$

Après regroupement, on aboutit à l'hyper-paramètre : $\boldsymbol{\theta} \triangleq [\boldsymbol{\mu}^T \quad \tilde{\boldsymbol{\sigma}}^T \quad \tilde{\boldsymbol{\ell}}^T]^T$ de dimension $3 \cdot K$. Il faut souligner que les vecteurs aléatoires $\Delta\boldsymbol{\beta}_m$ ($m = 1, \dots, M$) et $\Delta\boldsymbol{\beta}_\infty$ sont liés entre eux par l'intermédiaire des paramètres $\boldsymbol{\theta}$ du processus gaussien. Cela permet d'apprendre la loi de $\Delta\boldsymbol{\beta}_\infty$ à partir des $\Delta\boldsymbol{\beta}_m$ et donc des données \mathbf{y}_m ($m = 1, \dots, M$). Avec l'approche hiérarchique bayésienne, il est possible de modéliser le manque d'information sur les paramètres d'une distribution *a priori* en recourant au paradigme de Bayes, c'est-à-dire en spécifiant une autre distribution sur ces paramètres (loi *hyper a priori*) : $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$, où $p(\boldsymbol{\theta})$ est la densité de probabilité de $\boldsymbol{\theta}$. On choisit les lois *a priori* indépendantes suivantes : $\boldsymbol{\mu} \sim \mathcal{N}(0, \mathbf{P}_{\beta_\mu, \gamma_\mu}^{SE})$, $\tilde{\boldsymbol{\sigma}} \sim \mathcal{N}(\alpha_\sigma \cdot \mathbf{1}_K, \mathbf{P}_{\beta_\sigma, \gamma_\sigma}^{SE})$ et $\tilde{\boldsymbol{\ell}} \sim \mathcal{N}(\alpha_\ell \cdot \mathbf{1}_K, \mathbf{P}_{\beta_\ell, \gamma_\ell}^{SE})$, avec $\mathbf{P}_{\beta, \gamma}^{SE} \triangleq [k_{\beta, \gamma}^{SE}(h_i, h_j)]_{i=1 \dots K, j=1 \dots K}$. Le noyau SE (Squared Exponential) $k_{\beta, \gamma}^{SE}$, aussi dénommé gaussien, s'exprime classiquement par :

$$k_{\beta, \gamma}^{SE}(h, h') = \beta^2 \exp\left(-\frac{(h - h')^2}{2\gamma^2}\right) \quad (15)$$

Les paramètres $\beta_\mu, \gamma_\mu, \alpha_\sigma, \beta_\sigma, \gamma_\sigma, \alpha_\ell, \beta_\ell$ et γ_ℓ sont fixés de telle manière à induire des lois faiblement informatives. Sur la figure 2, on présente quelques réalisations *a priori* de $\Delta\boldsymbol{\beta}_m$ ($m = 1, \dots, M$) ou de manière équivalente de $\Delta\boldsymbol{\beta}_\infty$, à partir du modèle hiérarchique, pour des valeurs fixées plus loin dans la partie 4. Cela illustre la connaissance, notamment la non-stationnarité, qui est injectée dans la loi *a priori* du processus aléatoire.

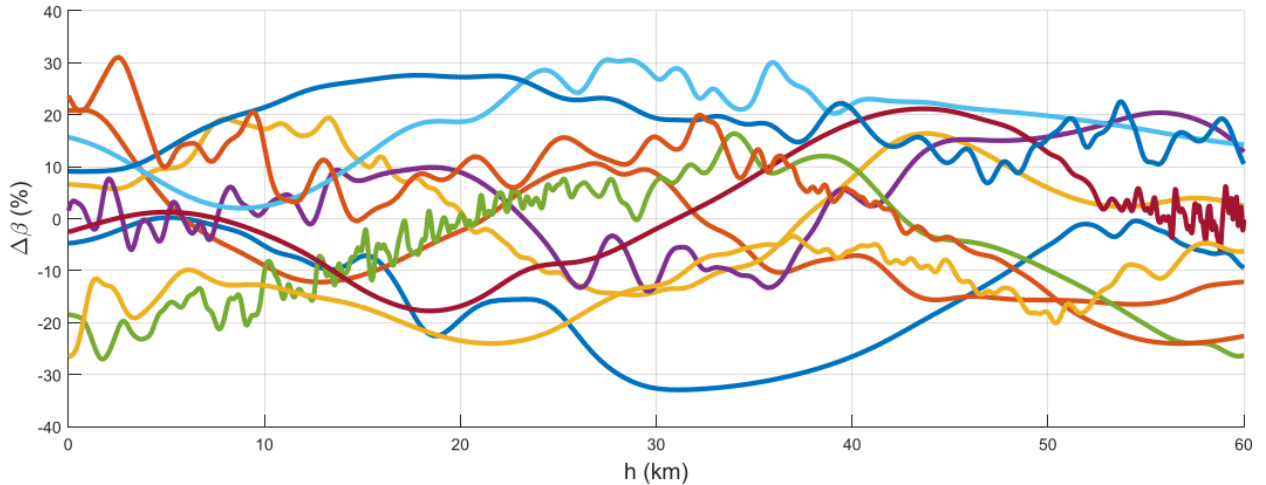


Figure 2: Réalisations du modèle hiérarchique

En complément, on choisit de simplifier drastiquement le modèle hiérarchique en faisant l'hypothèse d'indépendance entre erreurs de modèle $\Delta\beta_m$ et $\Delta\beta_{m'}$ ($m \neq m'$), conditionnellement à θ . De plus, de manière naturelle, on considère qu'il y a indépendance entre l'erreur de modèle $\Delta\beta_m$ et l'erreur de mesure ε_m .

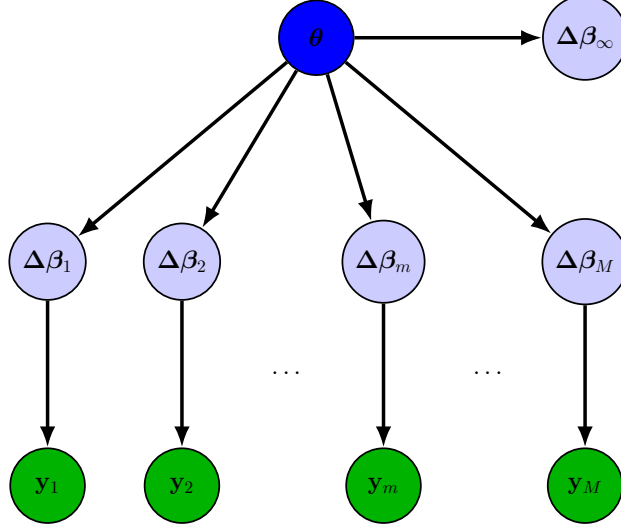


Figure 3: Modèle hiérarchique (propriété de Markov globale)

Modèle graphique probabiliste Le modèle graphique est constitué par le graphe orienté acyclique de la figure 3 qui représente les dépendances conditionnelles. On en déduit la décomposition suivante de la loi jointe :

$$p(\mathbf{y}_{1:M}, \Delta\beta_{1:M}, \Delta\beta_\infty, \theta) = p(\mathbf{y}_{1:M} | \Delta\beta_{1:M}) p(\Delta\beta_{1:M} | \theta) p(\Delta\beta_\infty | \theta) p(\theta) \quad (16)$$

$$= \prod_{m=1}^M p(\mathbf{y}_m | \Delta\beta_m) p(\Delta\beta_m | \theta) \cdot p(\Delta\beta_\infty | \theta) p(\theta) \quad (17)$$

Plus spécifiquement, la loi conditionnelle $p(\mathbf{y}_m | \theta)$ s'écrit :

$$p(\mathbf{y}_m | \theta) = \int p(\mathbf{y}_m | \Delta\beta_m) p(\Delta\beta_m | \theta) d\Delta\beta_m \quad (18)$$

Par linéarité du modèle de mesure (12), on montre aisément que la loi est gaussienne :

$$p(\mathbf{y}_m | \theta) = \mathcal{N}(\mathbf{y}_m, \mathcal{T}(\mathbf{x}_m) + \boldsymbol{\mu}, \mathbf{P} + \mathbf{R}) \quad (19)$$

3 Inférence bayésienne

S'appuyant sur le modèle hiérarchique proposé, on précise ci-après le problème inférentiel que l'on cherche à résoudre ainsi que l'approche numérique que l'on met en œuvre.

3.1 Distribution prédictive

Rappelons que le problème génératif requiert d'échantillonner $p(\Delta\boldsymbol{\beta}_\infty|\mathbf{y}_{1:M})$, la distribution prédictive. Après marginalisation, elle s'exprime par :

$$p(\Delta\boldsymbol{\beta}_\infty|\mathbf{y}_{1:M}) = \int p(\Delta\boldsymbol{\beta}_\infty|\boldsymbol{\theta}, \mathbf{y}_{1:M})p(\boldsymbol{\theta}|\mathbf{y}_{1:M})d\boldsymbol{\theta} \quad (20)$$

On peut alors faire l'approximation suivante : $p(\Delta\boldsymbol{\beta}_\infty|\mathbf{y}_{1:M}) \approx \frac{1}{N} \sum_{n=1}^N p(\Delta\boldsymbol{\beta}_\infty|\boldsymbol{\theta}^{(n)}, \mathbf{y}_{1:M})$, si l'on est capable de produire des échantillons $\boldsymbol{\theta}^{(n)}$ de la distribution *a posteriori* $p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$. Or, $p(\Delta\boldsymbol{\beta}_\infty|\boldsymbol{\theta}^{(n)}, \mathbf{y}_{1:M}) = \mathcal{N}(\Delta\boldsymbol{\beta}_\infty, \boldsymbol{\mu}^{(n)}, \mathbf{P}^{(n)})$ où l'on extrait directement l'échantillon "moyenne" $\boldsymbol{\mu}^{(n)}$ à partir de $\boldsymbol{\theta}^{(n)} = [\boldsymbol{\mu}^{(n)T} \quad (\tilde{\boldsymbol{\sigma}}^{(n)})^T \quad (\tilde{\boldsymbol{\ell}}^{(n)})^T]^T$ tandis que $\mathbf{P}^{(n)}$ peut être déterminé en fonction de $\tilde{\boldsymbol{\sigma}}^{(n)}$ et $\tilde{\boldsymbol{\ell}}^{(n)}$ au moyen de (11). Il en résulte que $p(\Delta\boldsymbol{\beta}_\infty|\mathbf{y}_{1:M})$ peut être approché par un mélange de gaussiennes. Cependant, il faut en premier lieu générer des échantillons $\boldsymbol{\theta}^{(n)} \sim p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$, pour $n = 1 \cdots N$. L'apprentissage de la loi $p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$ correspond à ce qu'on appelle la calibration. Comme on ne sait pas le faire directement, on va se tourner vers une méthode à chaîne de Markov Monte Carlo (MCMC), adaptée et efficace.

3.2 Expression du log-posterior et du gradient

Similairement à Heinonen (2016), on peut déduire, après développements, les expressions analytiques suivantes du log posterior $\log p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$ et du gradient du log posterior $\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$, en notant $\tilde{\mathbf{y}}_m \triangleq \mathbf{y}_m - (\mathcal{T}(\mathbf{x}_m) + \boldsymbol{\mu})$, l'observation centrée, et $\tilde{\tilde{\mathbf{y}}}_m \triangleq (\mathbf{P} + \mathbf{R})^{-1}\tilde{\mathbf{y}}_m$, l'observation centrée réduite.

Log posterior $\log p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$

$$\log p(\boldsymbol{\theta}|\mathbf{y}_{1:M}) = -\frac{1}{2} \sum_{m=1}^M \tilde{\mathbf{y}}_m^T (\mathbf{P} + \mathbf{R})^{-1} \tilde{\mathbf{y}}_m - \frac{1}{2} M \log \det(\mathbf{P} + \mathbf{R}) \quad (21)$$

$$+ \log p(\boldsymbol{\mu}|\beta_\mu, \gamma_\mu) + \log p(\tilde{\boldsymbol{\sigma}}|\alpha_\sigma, \beta_\sigma, \gamma_\sigma) + \log p(\tilde{\boldsymbol{\ell}}|\alpha_\ell, \beta_\ell, \gamma_\ell) + C \quad (22)$$

Gradient du log posterior $\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$

$$\frac{\partial \log p(\boldsymbol{\theta}|\mathbf{y}_{1:M})}{\partial \boldsymbol{\mu}} = \sum_{m=1}^M \tilde{\tilde{\mathbf{y}}}_m - (\mathbf{P}_{\beta_\mu, \gamma_\mu}^{SE})^{-1} \boldsymbol{\mu} \quad (23)$$

$$\begin{aligned} \frac{\partial \log p(\boldsymbol{\theta}|\mathbf{y}_{1:M})}{\partial \tilde{\boldsymbol{\sigma}}} &= 2 \text{diag} \left(\left[\sum_{m=1}^M \tilde{\tilde{\mathbf{y}}}_m \tilde{\tilde{\mathbf{y}}}_m^T - M(\mathbf{P} + \mathbf{R})^{-1} \right] \mathbf{P} \right) \\ &\quad - (\mathbf{P}_{\beta_\sigma, \gamma_\sigma}^{SE})^{-1} (\tilde{\boldsymbol{\sigma}} - \alpha_\sigma \cdot \mathbf{1}_K) \end{aligned} \quad (24)$$

$$\begin{aligned} \frac{\partial \log p(\boldsymbol{\theta}|\mathbf{y}_{1:M})}{\partial \tilde{\boldsymbol{\ell}}_i} &= \frac{1}{2} \text{tr} \left(\left[\sum_{m=1}^M \tilde{\tilde{\mathbf{y}}}_m \tilde{\tilde{\mathbf{y}}}_m^T - M(\mathbf{P} + \mathbf{R})^{-1} \right] \frac{\partial \mathbf{P}}{\partial \tilde{\boldsymbol{\ell}}_i} \right) \\ &\quad - \left[(\mathbf{P}_{\beta_\ell, \gamma_\ell}^{SE})^{-1} (\tilde{\boldsymbol{\ell}} - \alpha_\ell \cdot \mathbf{1}_K) \right]_i \quad \text{pour } i = 1 \cdots K \end{aligned} \quad (25)$$

3.3 Echantillonneur HMC dynamique

Echantillonner $p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$ est ardu en raison de la dimension $3 \cdot K$ de l'espace qui est de l'ordre de quelques centaines. Les algorithmes MCMC basiques, tels que Metropolis-Hastings, sont inadaptés, car beaucoup trop lents. Exploitant les expressions analytiques du log-posterior $\log p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$ et de son gradient $\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$, on se tourne vers des méthodes Monte Carlo Hamiltonien (HMC), décrites par Neal (2011). Ces méthodes vont autoriser une exploration plus rapide de l'espace. Initialement connu sous le terme "Hybrid Monte Carlo" et appliqué en dynamique moléculaire, HMC est populaire dans la communauté de l'apprentissage automatique. Il est fondé sur la mécanique hamiltonienne, une branche de la mécanique classique, et la géométrie différentielle afin de simuler la dynamique d'un système physique. Cela permet à l'algorithme d'explorer plus efficacement l'espace des paramètres et de fournir dans un schéma de type Metropolis-Hastings une proposition distante et pertinente, c'est-à-dire avec un faible taux de rejet. Cela conduit à une convergence plus rapide et à un meilleur mélange de la chaîne de Markov, notamment lorsque la dimension est élevée. Toutefois, certains paramètres de contrôle de l'algorithme HMC s'avèrent délicats à régler.

Pour surmonter cette difficulté, Hoffman (2014) a proposé un échantillonneur HMC dit "dynamique", dénommé No U-Turn Sampler (NUTS). Il détermine dynamiquement les longueurs successives L des pas, en utilisant une heuristique pour éviter les retours en arrière (No U-Turn), effectuant ainsi une exploration plus efficace de l'espace de recherche. Quant au pas d'intégration ε , il est réglé lors de la phase de "chauffe" par une méthode adaptative, dite de "dual averaging". Nous exploitons dans le papier une version originelle développée par Matthew D. Hoffman en Matlab^{©1}. D'autres versions plus avancées sont au cœur de langages de programmation probabiliste, tels que *Stan*, *PyMC3* et *Turing*,

Synthèse En entrée, on dispose des courbes $\mathbf{y}_{1:M}$ de la figure 1, auxquelles il faut adjoindre les incertitudes de mesure associées, via la donnée de la matrice de covariance \mathbf{R} . Une fois spécifiée la distribution *a priori* $p(\boldsymbol{\theta})$, l'échantillonneur HMC dynamique produit en régime stationnaire des échantillons $\boldsymbol{\theta}^{(n)} \sim p(\boldsymbol{\theta}|\mathbf{y}_{1:M})$, pour $n = 1 \dots N$. Un des intérêts de l'échantillonneur HMC est que ces échantillons sont faiblement corrélés. La distribution prédictive $p(\Delta\boldsymbol{\beta}_{\infty}|\mathbf{y}_{1:M})$ est alors échantillonné à partir du modèle de mélange gaussien.

4 Résultats

Le tableau 4 fait la synthèse des hypothèses et paramètres de fonctionnement. En cohérence avec les données de la figure 1, le modèle génératif hiérarchique produit des échantillons *a posteriori*, tels que représentés sur la figure 5. On retrouve la non-stationnarité en fonction de l'altitude, avec des fluctuations rapides au dessus de 30 km et plus lentes en dessous. Si les courbes produites s'écartent parfois des données, cela témoigne de la masse de probabilité présente dans la queue de distribution. Finalement, l'approche est apte à générer une grande variété de courbes, qui sont autant de réalisations possibles au vu des données ob-

¹distribué par la société The MathWorks (voir le site www.mathworks.com).

servées et considérant l'*a priori* $p(\boldsymbol{\theta})$. Les échantillons générés, faiblement corrélés, peuvent être ensuite exploités pour prédire l'évolution aérodynamique et la zone d'atterrissage, via l'approximation de Monte Carlo d'intégrales multidimensionnelles en grande dimension.

Données	Courbes de la figure 1, $K = 54$ ($m = 1 \dots 5$) $\mathbf{R} = 0.3^2 \cdot \mathbf{I}$, soit des erreurs non corrélées de $\pm 1\%(3\sigma)$
<i>A priori</i>	Densité $p(\boldsymbol{\mu})$: $\beta_\mu = 5, \gamma_\mu = 5$ Densité $p(\tilde{\boldsymbol{\sigma}})$: $\alpha_\sigma = 0.5, \beta_\sigma = 0.25, \gamma_\sigma = 0.5$ Densité $p(\tilde{\boldsymbol{\ell}})$: $\alpha_\ell = 0.5, \beta_\ell = 0.5, \gamma_\ell = 0.5$
NUTS	Nombre d'itérations : $N = 10000$ Initialisation : $\boldsymbol{\theta}_0 = [0 \cdot \mathbf{1}_K^T \quad \alpha_\sigma \cdot \mathbf{1}_K^T \quad \alpha_\ell \cdot \mathbf{1}_K^T]^T$ (<i>a priori</i> moyen) Phase de chauffe : $N_0 = 1000/100$ (dual averaging) Taux d'acceptation cible : $\tau = 60\%$ Profondeur maximale de l'arbre binaire : $N_{depth} = 10$ <i>Pas d'intégration "leapfrog" après adaptation</i> : $\varepsilon = 6 \cdot 10^{-4}$ <i>Nombre d'évaluations de $\nabla \log p(\boldsymbol{\theta} \mathbf{y}_{1:M})$</i> : 4184664 <i>Temps de calcul mono-chaîne</i> ~ 12 h

Figure 4: Données et fonctionnement de l'échantillonneur HMC

5 Conclusion et perspectives

En rentrée atmosphérique, la prédiction de l'évolution aérodynamique du véhicule peut être menée au moyen d'un modèle génératif hiérarchique opérant sur un faible nombre de données. La distribution prédictive peut être échantillonnée via un algorithme de Monte Carlo Hamiltonien dynamique, fondé sur la géométrie différentielle.

En perspectives, on peut évoquer l'adaptation à des données manquantes ou l'exploitation intensive de chaînes de Markov en parallèle pour accélérer le calcul. Un axe particulièrement pertinent serait d'introduire un modèle dynamique qui pourrait tenir compte de la non-stationnarité éventuelle selon le véhicule. Cela permettrait de tenir compte d'un nombre supérieur d'expériences de rentrée atmosphérique et ainsi d'améliorer encore la prédiction.

Bibliographie

- Duane, S. et al (1987), Hybrid Monte Carlo, *Physics letters B*, 195, pp. 216-222.
- Gibbs, M. N. (1998), Bayesian Gaussian processes for regression and classification, Doctoral dissertation, University of Cambridge.
- Heinonen M. et al (2016), Non-stationary Gaussian process regression with Hamiltonian Monte Carlo, *Artificial Intelligence and Statistics*, pp. 732-740.
- Kennedy M. C. and O'Hagan A. (2001), Bayesian calibration of computer models, *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 63, pp. 425-464.

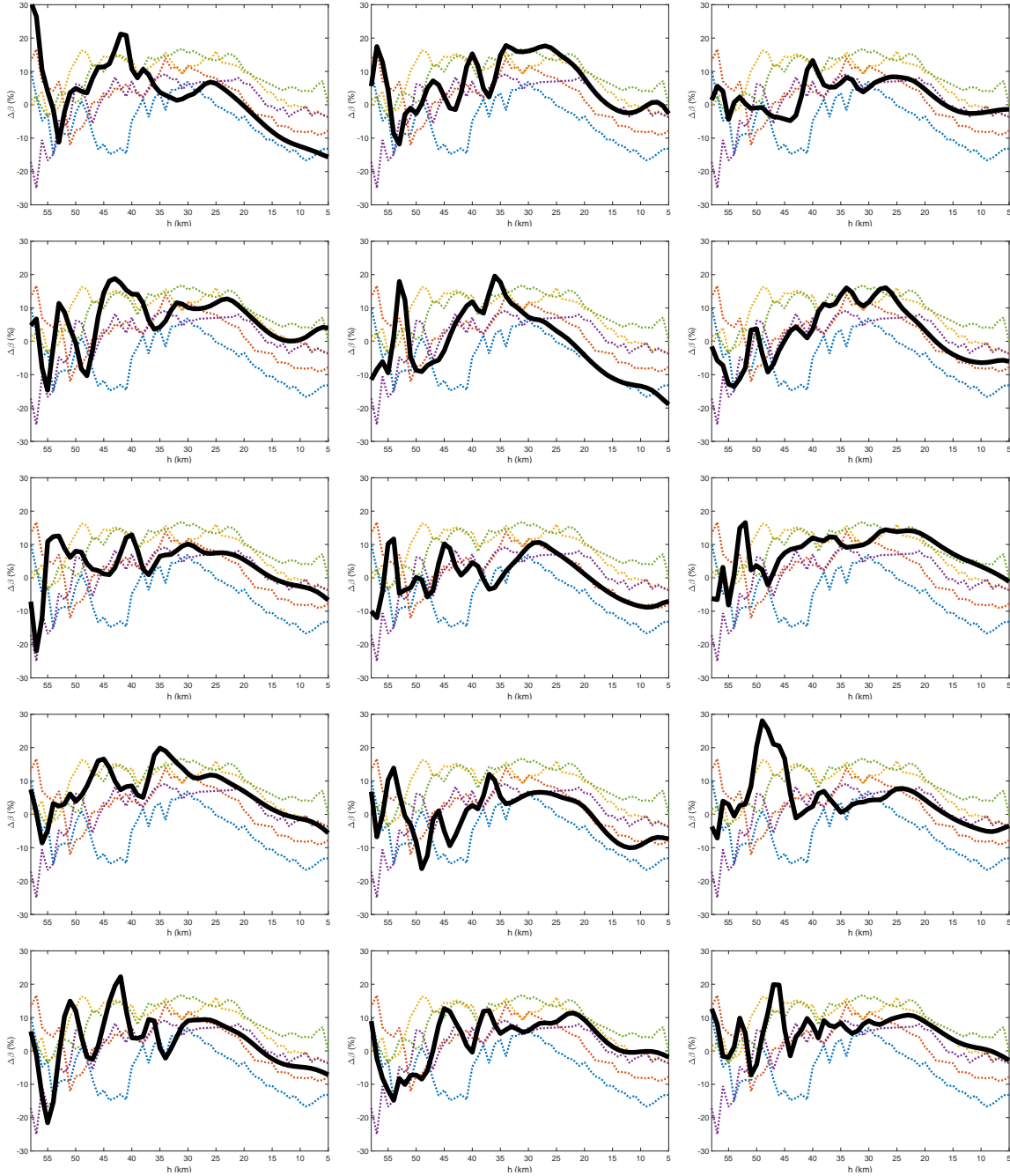


Figure 5: Echantillons *a posteriori* (trait épais noir) et origines (trait fin pointillé)

Hoffman M. D. and Gelman A. (2014), The No-U-Turn sampler : adaptively setting path lengths in Hamiltonian Monte Carlo, *J. Mach. Learn. Res.*, 15, pp. 1593-1623.

Neal R. M. (2011), MCMC using Hamiltonian dynamics, *Handbook of Markov Chain Monte Carlo*, 2., pp. 2.

Paciorek, C. et Schervish, M. (2003), Nonstationary covariance functions for Gaussian process regression, *Advances in Neural Information Processing Systems*, 16. ISO 690