

# Identification boîte noire et simulation de systèmes non-linéaires à temps continu par prédiction linéaire de processus aléatoires

Sylvain VINET, Emmanuel VAZQUEZ

SUPELEC, 91192 Gif-sur-Yvette, France

{sylvain.vinet, emmanuel.vazquez}@supelec.fr

**Résumé**— Cet article propose des méthodes de prédiction linéaire de processus aléatoires pour l'identification boîte noire et la simulation de systèmes dynamiques non-linéaires à temps continu. La méthode d'identification proposée utilise des observations bruitées du vecteur d'état à des instants quelconques. Elle comporte deux étapes distinctes. La première est l'estimation des dérivées temporelles de l'état et la deuxième est l'approximation du champ de vecteurs gouvernant la dynamique. Pour la simulation du système, nous proposons un nouveau schéma d'intégration numérique permettant de prendre en compte de manière consistante l'erreur d'approximation du champ de vecteur.

**Mots-clés**— Identification de systèmes non-linéaires à temps continu ; méthodes non-paramétriques ; processus gaussien.

## I. INTRODUCTION

Cet article propose une approche par processus aléatoires pour l'identification boîte noire et la simulation de systèmes dynamiques non-linéaires à temps continu définis par une équation différentielle ordinaire (EDO)

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), \quad (1)$$

avec  $\mathbf{x} \in \mathbb{R}^d$  (l'état du système),  $\mathbf{u} \in \mathbb{R}^q$  (la commande), et  $\mathbf{f} : \mathbb{R}^d \times \mathbb{R}^q \mapsto \mathbb{R}^d$  une application non-linéaire de classe  $C^1$  (le champ de vecteur). Supposons  $\mathbf{f}$  inconnue et  $\mathbf{x}$  observé aux instants  $t_1 < \dots < t_n$ . Nous nous fixons deux objectifs dans cet article. Le premier est d'estimer  $\mathbf{f}$  à partir des observations  $\mathbf{x}(t_1), \dots, \mathbf{x}(t_n)$ . Lorsque  $\mathbf{f}$  a été estimée, un second objectif est de simuler le système, c-à-d de prédire l'état pour tout  $t > t_n$ , connaissant  $\{\mathbf{u}(s) ; s \leq t\}$ .

La méthode des *délais* (*delay embedding* en anglais) est une approche classique pour identifier de manière non-paramétrique un système à temps continu. Concrètement, supposons les observations régulièrement espacées et notons  $\mathbf{x}^k$  le vecteur  $(\mathbf{x}_{k-\tau}, \dots, \mathbf{x}_k)$ , avec  $\mathbf{x}_i := \mathbf{x}(t_i)$ , et  $\mathbf{u}^k$  le vecteur  $(\mathbf{u}_{k-\tau}, \dots, \mathbf{u}_k)$ , avec  $\mathbf{u}_i := \mathbf{u}(t_i)$ . Sous certaines conditions, l'état du système aux instants  $t_i$  peut être approché en utilisant une équation de récurrence

$$\mathbf{x}_k = \mathbf{g}(\mathbf{x}^{k-1}, \mathbf{u}^{k-1}), \quad (2)$$

où  $\mathbf{g}$  est une fonction choisie dans un certain espace de fonctions, typiquement un espace de splines ou plus généralement un espace de Hilbert à noyau reproduisant (voir par exemple [4, 7]).

Dans de nombreux cas cependant, il serait préférable de conserver la représentation à temps continu du système,

c'est-à-dire d'estimer  $\mathbf{f}$  directement à partir des observations. En identification paramétrique il existe un certain nombre d'approches classiques : modèle dépendant de l'état [13], estimation fondée sur des techniques d'algèbre différentielle [3], la méthode de la fonctionnelle de moment de Poisson (voir par exemple [6]), etc. À notre connaissance, aucune approche *non-paramétrique* n'a encore été proposée en temps continu. Dans cet article, nous proposons une méthode d'estimation de  $\mathbf{f}$  en deux étapes. La première étape consiste à approximer la dérivée de l'état à partir des observations. La seconde étape consiste à construire une approximation de  $\mathbf{f}$  à partir des observations et des estimées de la dérivée. Comme indiqué ci-dessus, nous utilisons une modélisation par processus aléatoires (PA) : un premier PA, fonction du temps, est utilisé pour modéliser l'état du système et un second PA, fonction de l'état  $\mathbf{x}$  et de la commande  $\mathbf{u}$ , modélise le champ de vecteurs  $\mathbf{f}$ . L'utilisation d'un tel cadre probabiliste permet de quantifier facilement et de manière consistante les erreurs d'estimation produites à chaque étape. De plus, les modèles aléatoires considérés permettent de proposer une méthode de simulation originale d'EDO.

Cet article est organisé de la manière suivante. La section II rappelle des notions fondamentales de prédiction linéaire de processus aléatoires. Les sections III et IV présentent nos méthodes d'identification et de simulation. Pour finir, la section V fournit un exemple illustratif.

## II. PRÉDICTION LINÉAIRE DE PROCESSUS ALÉATOIRES

Dans cette section, nous rappelons brièvement les principes de *prédiction linéaire* (PL) et de *kriegerage intrinsèque* (KI) [5]. Le KI est apparu en géostatistique (voir par exemple [2]) et sera utilisé ici pour construire des approximations de fonctions non-linéaires. Considérons une fonction  $z : \mathbb{T} \rightarrow \mathbb{R}$ , où  $\mathbb{T}$  est un ensemble de paramètres ( $\mathbb{T} \subseteq \mathbb{R}^d$ , par exemple). Nous souhaitons construire une approximation de  $z$  à partir d'un nombre fini d'observations  $z(t_i)$ ,  $i \in \{1, \dots, n\}$ . Une idée classique est de modéliser  $z(t)$  par un PA et de calculer la *moyenne* des réalisations du PA qui interpolent les observations. En général, on ne sait pas calculer cette moyenne de manière efficace mais la théorie de la PL et du KI permet d'en obtenir une approximation linéaire par rapport aux observations (On parle aussi de *meilleur prédicteur linéaire non-biaisé*).

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité et notons  $\xi : \Omega \times \mathbb{T} \rightarrow \mathbb{R}$  un PA du second ordre, c-à-d un ensemble de variables aléatoires de variances finies indexées par les

éléments de  $\mathbb{T}$ . Notons  $m(t) := \mathbb{E}[\xi(t)]$  la moyenne de  $\xi(t)$  et  $k(t, s) := \text{Cov}(\xi(t), \xi(s))$ ,  $t, s \in \mathbb{T}$ , sa fonction de covariance. Comme  $z$  est supposée être une réalisation inconnue du processus  $\xi$ , chaque observation  $z(t_i)$  correspond à une réalisation de la variable aléatoire  $\xi(t_i)$ . De nombreuses propriétés découlent des caractéristiques de la covariance, notamment dans le cas de processus aléatoires gaussiens. En particulier, lors de la mise en place du modèle aléatoire, il est essentiel de choisir  $k$  selon la régularité, la différentiabilité, les propriétés spectrales, etc. de  $z$ . Par manque de place, nous ne parlerons pas de ces aspects, qui sont largement étudiés dans la littérature statistique. En pratique, la covariance est choisie dans une classe de fonctions définies positives paramétrées. Les paramètres sont habituellement estimés en utilisant un critère d'adéquation aux observations, comme le *maximum de vraisemblance* [8].

### A. Éléments de prédiction linéaire

Les processus aléatoires à moyenne nulle sont tels que leur fonction de covariance est un produit scalaire de  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ . En effet si  $\forall t \in \mathbb{T}$ ,  $m(t) = 0$ , alors  $\forall t, s \in \mathbb{T}$ ,  $(\xi(t), \xi(s))_{L^2} = \mathbb{E}[\xi(t)\xi(s)] = k(t, s)$ . Notons  $\mathcal{H}$  l'espace de Hilbert engendré par le processus à moyenne nulle  $\xi$ , c-à-d le complété de l'espace vectoriel dont les éléments sont des combinaisons linéaires des variables aléatoires  $\xi(t)$ ,  $t \in \mathbb{T}$ , muni du produit scalaire  $(\xi(t), \xi(s))_{L^2} = k(t, s)$ .

Soit  $\xi_0$  une variable aléatoire dans  $\mathcal{H}$ . Supposons observé un ensemble fini de variables aléatoires  $\xi_i \in \mathcal{H}$ ,  $i = 1, \dots, n$ . Le meilleur prédicteur linéaire de  $\xi_0$  à partir de  $\xi_1, \dots, \xi_n$  correspond à la meilleure approximation pour la norme  $L^2$  de  $\xi_0$  par une variable aléatoire  $\hat{\xi}_0$  de  $\mathcal{H}_S = \text{vect}\{\xi_1, \dots, \xi_n\}$ .  $\mathbb{E}[(\xi_0 - \hat{\xi}_0)^2]$  (la moyenne du carré de l'erreur entre la variable et son prédicteur) est minimisée si et seulement si  $\hat{\xi}_0$  est la projection orthogonale de  $\xi_0$  sur  $\mathcal{H}_S$ . Par conséquent,  $\hat{\xi}_0$  est une combinaison linéaire des VA  $\xi_1, \dots, \xi_n$  et les coefficients de cette combinaison s'obtiennent comme solution d'un système d'équations linéaires. Nous donnerons plus de détails dans la section suivante.

### B. Krigeage intrinsèque

Le krigeage intrinsèque permet d'obtenir un prédicteur linéaire lorsque la moyenne de  $\xi(t)$  est une fonction linéairement paramétrée inconnue. Nous rappelons dans cette section les principaux résultats [5].

Tout PA  $\xi$  peut s'écrire comme somme d'un PA de moyenne nulle et d'une fonction déterministe (la moyenne de  $\xi$ ). Dans cet article, nous restreignons la présentation du KI au cas où la moyenne de  $\xi$  est polynomiale. Dans ce cas

$$\xi(t) = \sum_{i=1}^m \alpha_i p_i(t) + \eta(t), \quad (3)$$

où  $\eta(t)$  est un PA de moyenne nulle et où les  $p_i$  forment une base de polynômes à  $d$  variables. Notons  $\mathcal{P}$  l'espace de dimension  $m$  engendré par les fonctions  $\{p_i\}_{i=1, \dots, m}$  et  $\mathcal{H}_\eta$  l'espace de Hilbert engendré par  $\eta$ . Lorsque  $m(t)$  est connue, les résultats de la section précédente peuvent être réutilisés en considérant la prédiction de  $\xi(t) - m(t)$ . La difficulté de l'extension de la PL au cas où  $m(t)$  est inconnue

vient du fait que les espaces  $\mathcal{H}_\eta$  et  $\mathcal{P}$  sont de natures différentes, et donc, il n'y a pas de produit scalaire canonique dans l'espace engendré par  $\xi$  pour appliquer le théorème de projection orthogonal.

Pour remédier à cette difficulté et travailler aisément avec les PA ayant une moyenne inconnue dans  $\mathcal{P}$ , Matheron [5] introduit la notion de *processus aléatoire généralisé* (PAG) qui étend celle de PA. Soit  $\tilde{\Lambda}$  l'espace vectoriel des mesures à support fini, c-à-d l'espace des combinaisons linéaires  $\sum_{i=1}^n \lambda_i \delta_{t_i}$ , où  $\delta_t$  est la mesure de Dirac en  $t$ . Soit  $\tilde{\Lambda}_{\mathcal{P}^\perp}$  le sous-ensemble des éléments de  $\tilde{\Lambda}$  qui s'annulent sur  $\mathcal{P}$ . Donc,  $\lambda \in \tilde{\Lambda}_{\mathcal{P}^\perp}$  implique

$$\langle \lambda, z \rangle := \sum_{i=1}^n \lambda_i z(t_i) = 0, \quad \forall z \in \mathcal{P}.$$

Si  $\xi(t)$ ,  $t \in \mathbb{T}$ , est un PA du second ordre, l'application linéaire

$$\begin{aligned} \xi : \tilde{\Lambda} &\rightarrow \mathcal{H} = \text{vect}\{\xi(t) ; t \in \mathbb{T}\} \\ \lambda = \sum_{i=1}^n \lambda_i \delta_{t_i} &\mapsto \xi(\lambda) = \sum_{i=1}^n \lambda_i \xi(t_i), \end{aligned}$$

étend  $\xi(t)$  sur  $\tilde{\Lambda}$ .  $(\lambda, \mu)_{\tilde{\Lambda}} := (\xi(\lambda), \xi(\mu))_{\mathcal{H}}$  définit un produit scalaire sur  $\tilde{\Lambda}$ . Soit  $\Lambda$  le complété de  $\tilde{\Lambda}$  muni de ce produit scalaire. Nous pouvons étendre  $\xi(\lambda)$  sur  $\Lambda$  par continuité et nous obtenons alors un PAG. Notons que pour tout  $\lambda \in \Lambda$ ,  $\mathbb{E}[\xi(\lambda)] = \langle \lambda, m(\cdot) \rangle$ , et pour tout  $\lambda = \sum_i \lambda_i \delta_{t_i} \in \tilde{\Lambda}$  et  $\mu = \sum_j \mu_j \delta_{s_j} \in \tilde{\Lambda}$ ,

$$k(\lambda, \mu) := \text{Cov}[\xi(\lambda), \xi(\mu)] = \sum_{i,j} \lambda_i \mu_j k(t_i, s_j). \quad (4)$$

$k(\lambda, \mu)$  peut être étendu sur  $\Lambda$  par continuité de l'opérateur de covariance. Notons aussi  $\Lambda_{\mathcal{P}^\perp}$  le complété de  $\tilde{\Lambda}_{\mathcal{P}^\perp}$  muni de la norme issue du produit scalaire  $(\cdot, \cdot)_{\Lambda}$ . Puisque  $\lambda \in \Lambda_{\mathcal{P}^\perp}$  filtre toute fonction de  $\mathcal{P}$ , si la moyenne de  $\xi(t)$  est dans  $\mathcal{P}$ , alors  $\forall \lambda \in \Lambda_{\mathcal{P}^\perp}$ ,  $\xi(\lambda)$  est une variable aléatoire de moyenne nulle. Remarquons également que  $\forall \lambda, \mu \in \Lambda_{\mathcal{P}^\perp}$ ,  $k(\lambda, \mu) = (\lambda, \mu)_{\Lambda}$ .

Nous rappelons maintenant les principaux résultats du KI. Soit  $\xi$  un PA de moyenne  $m(t) \in \mathcal{P}$  et de covariance  $k(t, s)$ . Soient  $\xi_0, \xi_1, \dots, \xi_n$  des variables aléatoires dans  $\mathcal{H} = \text{vect}\{\xi(t) ; t \in \mathbb{T}\}$ . Pour tout  $i \in \{0, 1, \dots, n\}$ , il existe un élément  $\lambda_i \in \Lambda$  tel que  $\xi_i = \xi(\lambda_i)$ . Considérons un modèle d'observations avec bruit additif de sorte que les observations correspondant à des réalisations des VA  $\xi_i^{\text{obs}} = \xi_i + \varepsilon_i$ ,  $i = 1, \dots, n$ , où les  $\varepsilon_i$  sont des VA, de moyennes nulles, indépendantes de  $\xi(t)$ , et de matrice de covariance  $\mathbf{K}_\varepsilon$ .

Le prédicteur du krigeage intrinsèque de  $\xi_0$  à partir des  $\xi_i^{\text{obs}}$  est la projection linéaire

$$\hat{\xi}_0 = \sum_{i=1}^n a_{0,i} \xi_i^{\text{obs}} = \xi(\hat{\lambda}_0) + \sum_{i=1}^n a_{0,i} \varepsilon_i \quad \left( \hat{\lambda}_0 = \sum_{i=1}^n a_{0,i} \lambda_i \right) \quad (5)$$

de  $\xi_0$  sur  $\mathcal{H}_S = \text{vect}\{\xi_i^{\text{obs}}, i = 1, \dots, n\}$ , telle que la variance de l'erreur de prédiction  $\xi_0 - \hat{\xi}_0$  est minimisée sous la contrainte

$$\lambda_0 - \hat{\lambda}_0 = \lambda_0 - \sum_{i=1}^n a_{0,i} \lambda_i \in \Lambda_{\mathcal{P}^\perp}. \quad (6)$$

Les coefficients  $a_{0,i}$ ,  $i = 1, \dots, n$ , sont solutions d'un système d'équations linéaires [5], qui peut s'écrire sous forme matricielle

$$\begin{pmatrix} \mathbf{K} + \mathbf{K}_\varepsilon & \mathbf{P}^\top \\ \mathbf{P} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{a}_0 \\ \boldsymbol{\mu}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{k}_0 \\ \mathbf{p}_0 \end{pmatrix}, \quad (7)$$

où  $\mathbf{K}$  est la matrice carré d'ordre  $n$  dont les coefficients sont  $k(\lambda_i, \lambda_j)$ ,  $i, j = 1, \dots, n$ ,  $\mathbf{P}$  est une matrice de dimension  $m \times n$  dont les coefficients sont  $\langle \lambda_j, p_i \rangle$  pour  $j = 1, \dots, n$  et  $i = 1, \dots, m$ ,  $\boldsymbol{\mu}_0$  est un vecteur de coefficients de Lagrange,  $\mathbf{k}_0$  est un vecteur de taille  $n$  avec comme coefficients  $k(\lambda_i, \lambda_0)$  et  $\mathbf{p}_0$  est un vecteur de taille  $m$  de coefficients  $\langle \lambda_0, p_i \rangle$ ,  $i = 1, \dots, m$ .

La variance de l'erreur de prédiction, qui correspond à l'incertitude de la prédiction, est donnée par

$$\begin{aligned} \widehat{\sigma}_0^2 &:= \text{Var}[\xi_0 - \widehat{\xi}_0] \\ &= k(\lambda_0, \lambda_0) - 2\mathbf{a}_0^\top \mathbf{k}_0 + \mathbf{a}_0^\top (\mathbf{K} + \mathbf{K}_\varepsilon) \mathbf{a}_0 \\ &= k(\lambda_0, \lambda_0) - \mathbf{a}_0^\top \mathbf{k}_0 - \mathbf{p}_0^\top \boldsymbol{\mu}_0. \end{aligned} \quad (8)$$

Cette variance, également appelée *variance de krigeage*, permet de donner des intervalles de confiance sur la prédiction. Dans les paragraphes suivants, nous utiliserons les notations suivantes

$$\begin{cases} \mathcal{K}(\xi_0 | \xi_1^{\text{obs}}, \dots, \xi_n^{\text{obs}}) := \mathbf{a}_0^\top, \\ \mathcal{V}(\xi_0 | \xi_1^{\text{obs}}, \dots, \xi_n^{\text{obs}}) := \text{Var}[\xi_0 - \widehat{\xi}_0]. \end{cases} \quad (9)$$

### C. Prédiction des dérivées

Dans cette section, nous rappelons comment utiliser le krigeage intrinsèque pour estimer les dérivées d'un PA à partir d'observations ponctuelles [10]. Pour simplifier l'exposé, supposons  $\mathbb{T} = \mathbb{R}$ . Rappelons qu'un PA  $\xi(t)$  du second ordre de moyenne nulle et de fonction de covariance  $k(t, s)$  est différentiable en moyenne quadratique en  $t$  si

$$\xi_h(t) = \frac{1}{h}(\xi(t+h) - \xi(t)) \quad (10)$$

converge en moyenne quadratique lorsque  $h \rightarrow 0$ . La limite existe si et seulement si  $\partial^2 k(u, s) / \partial u \partial s$  existe au point  $(t, t)$ . Si  $\xi(t)$  est différentiable en moyenne quadratique pour tout  $t$ , le processus limite est appelé le *processus dérivé* et nous le notons  $\dot{\xi}$ . Les dérivées d'ordre supérieur sont obtenues par itération et sont notées  $\xi^{(r)}$ . On vérifie facilement que

$$\text{Cov}[\xi^{(q)}(t), \xi^{(r)}(s)] = \frac{\partial^{q+r}}{\partial t^q \partial s^r} k(t, s). \quad (11)$$

Nous étendons maintenant la notion de différentiabilité aux PAG. Soit  $\tau_h : \Lambda \rightarrow \Lambda$  l'opérateur de translation tel que pour  $\lambda = \sum_i \lambda_i \delta_{t_i} \in \tilde{\Lambda}$ ,  $\tau_h \lambda = \sum_i \lambda_i \delta_{t_i+h}$ . Posons

$$\lambda_h := \frac{1}{h}(\tau_h \lambda - \lambda).$$

Un PAG  $\{\xi(\lambda) ; \lambda \in \Lambda\}$  est dit *différentiable en moyenne quadratique* en  $\lambda \in \Lambda$  si  $\xi(\lambda_h)$  converge en moyenne quadratique lorsque  $h \rightarrow 0$ . Quand la limite existe, on la note  $\dot{\xi}(\lambda)$ . Si  $\partial^2 k(u, s) / \partial u \partial s$  existe en  $(t, t)$  pour tout  $t \in \mathbb{R}$ , alors  $\xi(\lambda)$  est différentiable en moyenne quadratique pour tout  $\lambda$  [voir 10]. Remarquons que si  $\dot{\xi}(\lambda)$  existe, il existe également  $\dot{\lambda} \in \Lambda$  tel que  $\xi(\lambda) = \xi(\dot{\lambda})$ .

Les dérivées d'ordre  $r$  sont notées  $\xi^{(r)}(\lambda)$  et les éléments de  $\Lambda$  tels que  $\xi^{(r)}(\lambda) = \xi(\lambda^{(r)})$  sont notés  $\lambda^{(r)}$ . Étant donné  $\lambda = \sum_i \lambda_i \delta_{t_i}^{(q_i)}$  et  $\mu = \sum_j \mu_j \delta_{s_j}^{(r_j)}$  dans  $\Lambda$ , il est facile de vérifier que

$$\text{Cov}[\xi(\lambda), \xi(\mu)] = \sum_{i,j} (-1)^{r_j} \lambda_i \mu_j \frac{\partial^{q_i+r_j}}{\partial t_i^{q_i} \partial s_j^{r_j}} k(t_i, s_j).$$

Nous invitons le lecteur à se reporter à [10] pour plus de détails.

### III. MÉTHODE D'IDENTIFICATION EN DEUX ÉTAPES

Dans cette section, nous présentons une méthode d'identification boîte noire de systèmes non-linéaires à temps continu à partir d'un nombre fini d'observations (pouvant être bruitées) du vecteur d'état<sup>1</sup>.

Pour éviter de compliquer les notations, la présentation sera faite dans le cas d'un état scalaire ( $d = 1$ ). La généralisation à plusieurs dimensions est immédiate (comme nous le verrons dans l'exemple de la section V). Comme indiqué dans l'introduction, la méthode comporte deux étapes. L'objectif de la première étape est d'estimer la dérivée de l'état  $x$ . La seconde étape consiste à calculer une approximation  $\hat{f}$  de  $f$ .

#### A. Estimation des dérivées de l'état

Soit  $\{x_i^{\text{obs}}, i = 1 \dots n\}$  un ensemble d'observations bruitées de l'état aux instants  $t_i$ , qui peuvent être irrégulièrement espacés. Notre objectif est d'estimer  $\dot{x}(t)$ ,  $t \in \mathbb{R}$ , à partir des observations. Considérons un PA  $X \in \mathbb{R}$  différentiable en moyenne quadratique, indicé par  $\mathbb{R}$ , de moyenne  $m$  constante mais inconnue, et de fonction de covariance  $k_X$  (au moins) deux fois différentiable. Supposons de plus que  $x$  est une réalisation de  $X$ , et que le bruit est modélisé par des variables aléatoires  $V_i$ ,  $i = 1, \dots, n$ , i.i.d, de moyenne nulle et variance connue  $\sigma_V^2$ . Donc, pour tout  $i$ ,  $x_i^{\text{obs}}$  est une réalisation de la variable aléatoire

$$X_i^{\text{obs}} := X(t_i) + V_i. \quad (12)$$

En utilisant les résultats de la section II-C, la dérivée de  $x$  en  $t_i$ ,  $i = 1 \dots n$ , peut être estimée en calculant le prédicteur  $\hat{X}(t_i)$  par KI de  $\dot{X}(t_i)$  à partir de  $X_j^{\text{obs}}$ ,  $j = 1, \dots, n$ . Concrètement, pour tout  $i = 1, \dots, n$ ,  $\dot{x}(t_i)$  peut être estimé par

$$\hat{\dot{x}}(t_i) = \mathcal{K}_X(\dot{X}(t_i) | \mathbf{X}^{\text{obs}}) \mathbf{x}^{\text{obs}}, \quad (13)$$

avec

$$\begin{aligned} \mathbf{X}^{\text{obs}} &= (X_1^{\text{obs}}, \dots, X_n^{\text{obs}})^\top \text{ et} \\ \mathbf{x}^{\text{obs}} &= (x_1^{\text{obs}}, \dots, x_n^{\text{obs}})^\top. \end{aligned}$$

La notation  $\mathcal{K}_X$  indique que la covariance  $k_X$  est utilisée pour calculer les coefficients du krigeage. La variance de l'erreur d'estimation est donnée par  $\mathcal{V}_X(\dot{X}(t_i) | \mathbf{X}^{\text{obs}})$ .

<sup>1</sup>Cette méthode pourrait également être appliquée si le vecteur d'état n'était que partiellement observé à condition que la partie manquante se déduise par opérations linéaires (par exemple intégration ou dérivation) sur les composantes observées. En effet le krigeage permet de prédire toute opération linéaire sur un PA. Par manque de place nous n'aborderons pas ce point.

## B. Approximation du système

L'étape suivante de notre méthode d'identification est l'approximation de la fonction  $f : \mathbb{R} \times \mathbb{R}^q \rightarrow \mathbb{R}$  en utilisant les dérivées estimées par (13). (Si  $d > 1$ , chaque composante  $f_i$ ,  $i = 1, \dots, d$ , de  $f$  doit être estimée séparément.) Supposons que  $f$  est une réalisation d'un PA  $F$ , indicé par  $x$  et  $\mathbf{u}$ , de moyenne  $m(x, \mathbf{u})$  et de covariance  $k_F\{(x, \mathbf{u}), (x', \mathbf{u}')\}$ , où  $(x, \mathbf{u}), (x', \mathbf{u}') \in \mathbb{R} \times \mathbb{R}^q$ . Comme mentionné à la section II-B,  $m(x, \mathbf{u})$  peut être une constante inconnue ou un polynôme en  $x$  et  $\mathbf{u}$ , ce qui permet d'inclure des connaissances a priori sur  $f$  (par exemple, on peut imposer à  $F$  une tendance linéaire suivant  $x$  ou  $\mathbf{u}$  [11]). Pour l'approximation de  $f$ , deux cas sont à considérer suivant que le bruit d'observation est faible ou fort.

Lorsque le bruit d'observation est faible, nous supposons que l'erreur d'observation (la différence entre  $x_i^{\text{obs}}$  et  $x(t_i)$ ) peut être négligée. Alors, pour tout  $i$ , la dérivée estimée (13) peut être vue comme une réalisation de  $F$  au point  $(x_i^{\text{obs}}, \mathbf{u}(t_i)) \approx (x(t_i), \mathbf{u}(t_i))$ . L'erreur d'estimation de la dérivée de l'état est prise en compte en introduisant une variable gaussienne  $W_i$  indépendante de moyenne nulle et de variance égale à  $\mathcal{V}_{\mathcal{X}}(\mathcal{X}(\lfloor \cdot \rfloor) | \mathbf{X}^{\text{obs}})$ . Ainsi, pour tout  $i$ , définissons la variable aléatoire  $F_i^{\text{obs}} := F(x_i^{\text{obs}}, \mathbf{u}(t_i)) + W_i$ . Pour tout  $x$  et  $\mathbf{u}$ , le prédicteur  $\hat{F}(x, \mathbf{u})$  par KI de  $F(x, \mathbf{u})$  à partir des variables  $F_i^{\text{obs}}$ ,  $i = 1, \dots, n$ , peut être utilisé pour obtenir une approximation de  $f(x, \mathbf{u})$ , s'écrivant

$$\hat{f}(x, \mathbf{u}) = \mathcal{K}_F(F(x, \mathbf{u}) | \mathbf{F}^{\text{obs}}) \hat{\mathbf{x}} \quad (14)$$

avec

$$\begin{aligned} \mathbf{F}^{\text{obs}} &= (F_1^{\text{obs}}, \dots, F_n^{\text{obs}})^{\top} \quad \text{et} \\ \hat{\mathbf{x}} &= (\hat{x}(t_1), \dots, \hat{x}(t_n))^{\top}. \end{aligned}$$

La notation  $\mathcal{K}_F$  indique que la covariance  $k_F$  est utilisée pour calculer les coefficients du krigeage.

Dans le cas d'un bruit fort, l'erreur sur la position de l'observation de  $F$  doit être prise en compte. Le modèle d'observation peut s'écrire  $F_i^{\text{obs}} := F(x(t_i) + V_i, \mathbf{u}(t_i)) + W_i$ , avec  $V_i$  le bruit d'observation de l'état. Bien sûr, la covariance entre  $F_i^{\text{obs}}$  et  $F(x', \mathbf{u}')$  n'est pas  $k_F\{(x(t_i), \mathbf{u}(t_i)), (x', \mathbf{u}')\}$ . Cependant, la covariance peut être calculée et un prédicteur similaire à (14) peut être obtenu à nouveau. Dans la littérature du krigeage, la modification de la fonction de covariance due à des incertitudes sur la position d'observation est un problème classique (voir [2], p. 74–80). Bien qu'important en pratique, le cas du bruit fort ne sera pas développé dans cet article par manque de place.

Après la deuxième étape de la méthode d'identification, une approximation  $\hat{f}$  de  $f$  est obtenue. Nous pouvons ensuite utiliser un solveur classique d'EDO pour simuler le système  $\dot{\mathbf{x}} = \hat{f}(\mathbf{x}, \mathbf{u})$ . Cependant, nous pensons que l'erreur d'approximation de  $f$  doit être prise en compte durant la simulation. Dans la section suivante nous proposons une méthode d'intégration qui utilise le point de vue des PA présenté précédemment afin de tenir compte de cette erreur d'approximation.

## IV. SIMULATION

Dans cette section nous proposons une méthode d'intégration numérique de l'équation différentielle (1) fondée sur une modélisation aléatoire du vecteur d'état. Comme dans

la section III, nous supposons que  $d = 1$  pour simplifier la présentation. Le schéma d'intégration proposé est une méthode à pas multiples (voir par exemple [1]) dont un prédicteur de l'état peut s'écrire

$$x_{n+1} = \sum_{i=0}^p \alpha_i x_{n-i} + \sum_{j=0}^q \beta_j f(x_{n-j}, \mathbf{u}((n-j)h)), \quad n \geq 0, \quad (15)$$

où  $h$  est le pas d'intégration et  $x_n$  est la prédiction de  $x(nh)$  étant données les conditions initiales  $x_0 := x(0)$ ,  $x_{-1} := x(-h)$ ,  $x_{-2} := x(-2h)$ , ... Dans les méthodes classiques, les coefficients  $\alpha_i$  et  $\beta_j$  sont obtenus par minimisation de l'erreur de prédiction en supposant une approximation polynomiale de  $x$ . Dans la méthode proposée, les coefficients sont obtenus comme résultat de la meilleure prédiction linéaire du PA  $X$ .

Dans un premier temps, supposons  $f$  connue. Pour prédire  $x_{n+1}$  à partir de  $x_n, \dots, x_{n-q}$ ,  $q \geq 0$ , considérons le prédicteur  $\hat{X}((n+1)h)$  par KI de  $X((n+1)h)$  à partir de  $X(nh)$  et  $\dot{X}((n-q)h), \dots, \dot{X}(nh)$ . Alors, une prédiction de l'état au pas  $n+1$  s'écrit

$$x_{n+1} = \mathcal{K}_X(X((n+1)h) | X(nh), \dot{X}((n-q)h), \dots, \dot{X}(nh)) \mathbf{x}_n^{\sharp}, \quad (16)$$

avec  $\mathbf{x}_n^{\sharp} := (\mathbf{x}_n, \dot{\mathbf{x}}_{n-q}, \dots, \dot{\mathbf{x}}_n)^{\top}$  et, pour tout  $i$ ,

$$\dot{x}_i := f(x_i, \mathbf{u}(ih)). \quad (17)$$

**Remarque 1** Le prédicteur (16) est un cas particulier de (15) avec  $p = 0$  (voir Figure 1). Quand  $p = 0$ , la stabilité du schéma d'intégration (15) est assurée si  $\alpha_0 \leq 1$  [1]. Dans (16), on a  $\alpha_0 = 1$  en raison de (6). Ceci explique notre choix de considérer une prédiction linéaire sans les termes  $X((n-1)h), X((n-2)h)$ , etc.

**Remarque 2** Supposons la covariance stationnaire, ce qui implique qu'il existe une fonction  $k_{\text{iso}}$  telle que  $k(t, t') = k_{\text{iso}}(h)$ , avec  $h = |t - t'|$ . Alors la méthode d'intégration proposée est *consistante*. De plus si  $k_{\text{iso}}(h)$  est  $s$  fois différentiable à l'origine, alors  $E[(\hat{X}((n+1)h) - X((n+1)h))^2] = O(h^s)$ . Nous ne fournissons pas la preuve de ce résultat par manque de place.

Lorsque  $f$  est estimée, il est possible de prendre en compte l'erreur d'approximation de  $f$  dans le schéma d'intégration proposé précédemment. Dans ce cas, nous considérons à la place de (16) le prédicteur linéaire  $\hat{X}((n+1)h)$  par KI de  $X((n+1)h)$  à partir de  $X(nh)$  et  $\dot{X}((n-i)h) + \tilde{W}_i$ ,  $i = 0, \dots, q$ , où les  $\tilde{W}_i$  sont des variables aléatoires indépendantes de moyennes nulles et de variances égales à

$$\mathcal{V}_{\mathcal{F}}(\mathcal{F}(\mathcal{S}_{\setminus -}), \mathbf{u}(\setminus -)) | \mathcal{F}_1^{\text{obs}}, | = \infty, \dots, \setminus).$$

Ici,  $\tilde{W}_i$  représente l'incertitude de  $\dot{x}((n-i)h)$  due à l'erreur d'approximation de  $f$ . Bien sûr,  $f$  sera remplacé par  $\hat{f}$  pour le calcul de  $\dot{x}_i$  dans (17).

## V. EXPÉRIENCE

Dans cette section, nous présentons quelques résultats expérimentaux pour illustrer les procédures d'identification et de simulation proposées. Nous considérons une équation

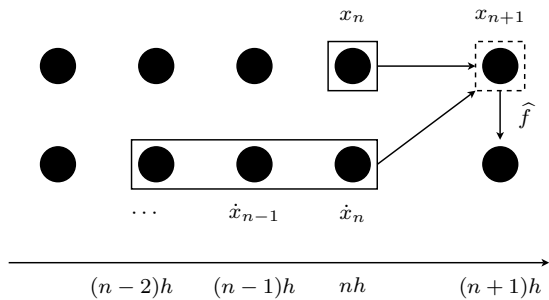


Fig. 1. Illustration du schéma d'intégration proposé

différentielle non-linéaire de type Lotka-Volterra commandée

$$\begin{cases} \dot{x}_1(t) = f_1(x_1, x_2, u_1) = \alpha x_1 - \beta x_1 x_2 + u_1, \\ \dot{x}_2(t) = f_2(x_1, x_2, u_2) = -\gamma x_2 - \delta x_1 x_2 + u_2, \end{cases} \quad (18)$$

avec  $\alpha = \gamma = 1$ ,  $\beta = 0.2$ ,  $\delta = 0.5$  et

$$\begin{aligned} u_1(t) &= 5 \cos^2(t/\exp(1)) \cos^2(t), \\ u_2(t) &= 5 \cos^2(t/\exp(1)) \sin^2(\frac{\pi}{2}t). \end{aligned}$$

L'ensemble des observations est constitué de  $n = 40$  échantillons du vecteur d'état aux instants  $0, h, \dots, (n-1)h$ , avec  $h = 0.5$ , simulés en utilisant un solveur d'EDO standard et altéré par un bruit blanc gaussien de loi  $\mathcal{N}(t, \infty t^{-\Delta})$ . Les composantes de l'état sont modélisées par deux PA indépendants  $X_1$  et  $X_2$  de moyenne inconnue et constante et de covariance choisie dans la classe des fonctions de Wendland (voir par exemple [12]). La classe des fonctions de Wendland fournit des covariances stationnaires à support compact. Ici, la fonction de classe  $C^6$

$$k_X(t, s) = \sigma^2 (1 - h/\rho)_+^8 (32(h/\rho)^3 + 25(h/\rho)^2 + 8h/\rho + 1),$$

$h = |t - s|$ , a été utilisée comme fonction de covariance à la fois pour  $X_1$  et pour  $X_2$ . Les paramètres  $\sigma^2$  et  $\rho$  ont été estimés par maximum de vraisemblance [8]. La figure 2 montre les composantes du vecteur d'état et leurs dérivées, estimées suivant la première étape de la procédure d'identification par KI (Section III-A).

L'étape suivante de la procédure d'identification consiste à obtenir une approximation de  $\mathbf{f}$ . Ici, les deux composantes de  $\mathbf{f}$  sont modélisées par deux PA indépendants  $F_1$  et  $F_2$ , avec une moyenne inconnue de la forme  $\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{1,1} x_1 x_2 + \alpha_{2,0} x_1^2 + \alpha_{0,2} x_2^2$  et une fonction de covariance généralisée s'écrivant  $k_F(\mathbf{x}, \mathbf{x}') = a_1 h - a_3 h^3 + a_5 h^5$ , avec  $h = \|\mathbf{x} - \mathbf{x}'\|$ . Les paramètres  $a_i \geq 0$  sont estimés par maximum de vraisemblance. Les figures 3 et 4 illustrent l'approximation de la première composante de  $\mathbf{f}$  obtenue suivant la deuxième étape de notre méthode d'identification. Notons que les choix que nous avons effectués pour les covariances découlent de considérations classiques (voir [9]).

Après que  $\mathbf{f}$  a été estimée, nous prédisons la trajectoire du système à partir de  $t = 20$ . Dans la figure 2, nous montrons le résultat de la simulation obtenue par la méthode présentée à la section IV. L'erreur de prédiction reste petite et ne grandit pas avec le temps.

## VI. CONCLUSION

L'approche originale que nous proposons pour l'identification boîte noire et la simulation de systèmes non-linéaires

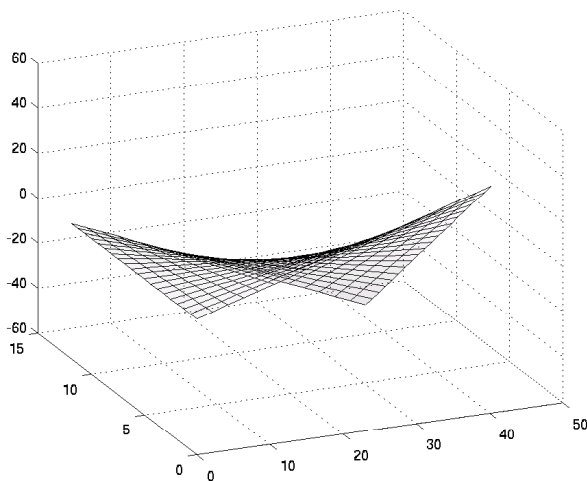


Fig. 3. Représentation de  $f_1(x_1, x_2, 0) = x_1 - 0.2x_1x_2$ .

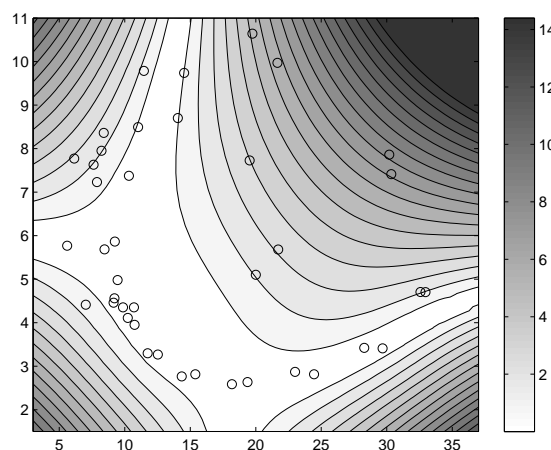


Fig. 4. Lignes de niveau dans le plan  $(x_1, x_2)$  de la différence en valeur absolue entre  $f_1(x_1, x_2, 0.47)$  et son approximation  $\hat{f}_1(x_1, x_2, 0.47)$  obtenu par notre méthode. Les cercles indiquent la position des observations.

à temps continu se fonde sur les processus aléatoires et le krigeage intrinsèque. Ce cadre probabiliste permet la quantification des erreurs d'approximation et l'élaboration naturelle d'un schéma d'intégration consistant. De plus des résultats expérimentaux prometteurs ont été obtenus.

## RÉFÉRENCES

- [1] J.C. BUTCHER : *Numerical Methods for Ordinary Differential Equations*. Wiley, 2003.
- [2] J.-P. CHILÈS et P. DELFINER : *Geostatistics : Modeling Spatial Uncertainty*. Wiley, New York, 1999.
- [3] M. FLIESS, C. JOIN et H. SIRA-RAMIREZ : Complex continuous nonlinear systems : Their black box identification and their control. *In Proc. SYSID'2006*, 2006.
- [4] A. GIRARD, C. E. RASMUSSEN, J. Quiñero CANDELA et R. MURRAY-SMITH : Multiple-step ahead prediction for non linear dynamic systems – a Gaussian Process treatment with propagation of uncertainty. *In*

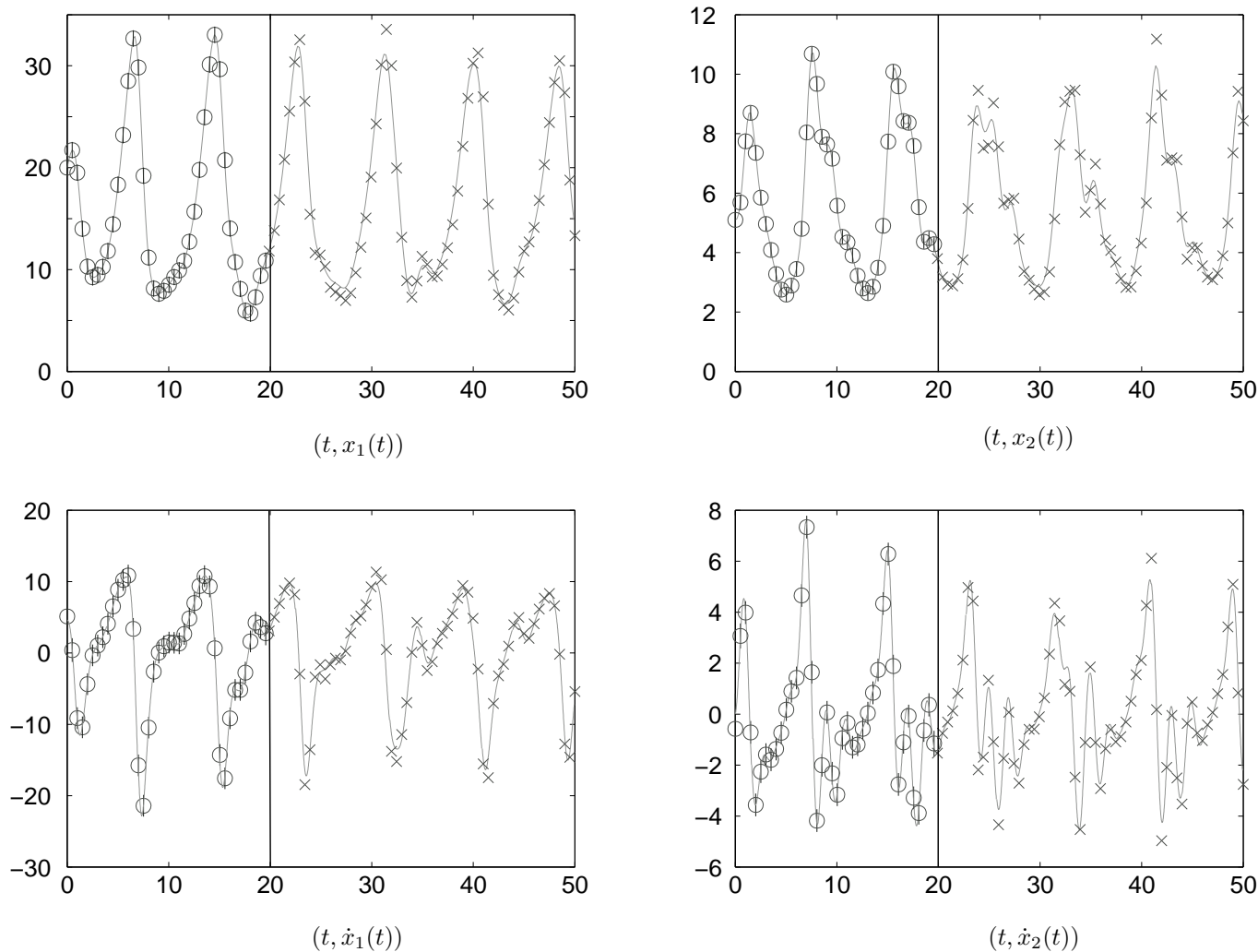


Fig. 2. Simulation du système (18). Trait plein :  $x_1(t)$ ,  $x_2(t)$ ,  $\dot{x}_1(t)$  et  $\dot{x}_2(t)$  calculé par un solveur ODE. Cercles : observations et dérivées estimées lors de la procédure d'identification. Croix : composantes du vecteur d'état et leurs dérivées obtenues par la méthode de simulation proposée. La trait vertical en  $t = 20$  représente le début de la simulation.

- S. BECKER, S. THRUN et K. OBERMAYER, éditeurs : *Advances in Neural Information Processing Systems*, volume 15, pages 529–536. MIT press, 2003.
- [5] G. MATHERON : The intrinsic random functions, and their applications. *Adv. Appl. Prob.*, 5:439–468, 1973.
- [6] G.P. RAO et H. UNBEHAUEN : Identification of continuous-time systems. *IEE Proceedings - Control Theory and Applications*, 153(2):185–220, 2006.
- [7] J. SJÖBERG, Q. ZHANG, L. LJUNG, A. BENVENISTE, B. DELYON, P.Y. GLORENNEC, H. HJALMARSSON et A. JUDITSKY : Nonlinear black-box modeling in system identification : a unified overview. *Automatica*, 31(12):1691–1724, 1995.
- [8] M. L. STEIN : *Interpolation of Spatial Data : Some Theory for Kriging*. Springer, New York, 1999.
- [9] E. VAZQUEZ : *Modélisation comportementale de systèmes non-linéaires multivariés par méthodes à noyaux et applications*. Thèse de doctorat, Univ Paris XI, Orsay, France, 2005.
- [10] E. VAZQUEZ et E. WALTER : Estimating derivatives and integrals with kriging. *In IEEE 44th Conference on Decision and Control*, 2005.
- [11] E. VAZQUEZ et E. WALTER : Intrinsic kriging and prior information. *Appl. Stoch. Models Bus. Ind.*, 2005.
- [12] H. WENDLAND : *Scattered Data Approximation*. Cambridge University Press, New York, 2005.
- [13] P. YOUNG, H. GARNIER et A. JARVIS : The identification of continuous-time linear and nonlinear models : a tutorial with environmental applications. *In In Proc. SYSID'2003*, 2003.