

# FILTRE DE KALMAN ROBUSTE AVEC COVARIABLES STOCHASTIQUES

Jean-Luc MAHOROMEZA <sup>1,2</sup> & Olivier WINTENBERGER <sup>1</sup> & Adeline FERMANIAN <sup>2</sup>  
& Joseph DE VILMAREST <sup>3</sup>

<sup>1</sup> *Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, Paris, France*

<sup>2</sup> *LOPF, Califrais' Machine Learning Lab, Paris, France*

<sup>3</sup> *Viking Conseil, Paris, France*

**Résumé.** Le filtre de Kalman-Bucy permet l'estimation et la prédiction des modèles espace-état. Il est souvent confronté à des observations incertaines, par exemple quand elles proviennent de mesures bruitées d'un signal physique (capteurs, météo, etc.). Plutôt que de considérer les covariables comme des quantités déterministes, ce qui est fait classiquement, notre objectif est de prendre en compte leurs incertitudes et d'étudier la robustesse du filtre. Plus précisément, le filtre de Kalman est une approche bayésienne qui donne l'estimation de la loi *a posteriori* de l'état, conditionnellement aux observations passées. Nous raffinons cette approche en intégrant les incertitudes des covariables dans le modèle, c'est-à-dire en les considérant comme des variables aléatoires dont la loi doit être estimée.

**Mots-clés.** Robustesse, Séries temporelles, Filtre de Kalman, Statistiques bayésiennes.

**Abstract.** The Kalman-Bucy filter is used for estimating and predicting state-space models. It frequently faces uncertain observations, such as those derived from noisy measurements of physical signals (including sensors, weather data, etc.). Instead of considering covariates as deterministic quantities, as is conventionally done, our purpose is to consider their uncertainties into account and address a robust filter. More specifically, the Kalman filter is a Bayesian approach that estimates the posterior distribution of the state conditionally on past observations. We refine this approach by integrating the uncertainties of the covariates into the model, i.e, considering them as stochastic variables whose distributions need to be estimated.

**Keywords.** Robustness, Time series, Kalman Filter, Bayesian statistics.

## 1 Introduction

Les modèles espace-état se sont imposés comme un cadre incontournable pour analyser les modèles dynamiques. Ces modèles reposent sur l'idée qu'un vecteur observable  $y_t$ , souvent appelé "vecteur d'observation", est une fonction du vecteur non observé (latent)  $\theta_t$ , désigné comme le "vecteur d'état". La dynamique régissant le vecteur d'observation et le vecteur d'état est encapsulée au sein du modèle espace-état. Dans la seconde moitié du 20ème siècle,

le filtre de Kalman and Bucy (1961) a considérablement avancé la modélisation des modèles espace-état. Il a fourni un cadre d'estimation des modèles linéaires, fondé sur le filtre de Wiener (Wiener, 1949). Le filtre de Kalman-Bucy se distingue par son mécanisme récursif pour estimer et prédire l'état des modèles dynamiques linéaires dans un contexte de données bruitées. Cet avancement a été crucial pour la prédiction précise (l'espérance conditionnelle de l'état sachant le passé est calculée de façon exacte), l'estimation et le contrôle dans les modèles dynamiques linéaires, marquant une réalisation significative dans l'ingénierie, y compris le suivi de cibles, le contrôle d'avions (Ray and Stengel, 1991), l'ingénierie électrique (Vilmarest, 2022) et les sciences environnementales.

Au fil du temps, le filtre de Kalman a connu un succès généralisé, menant à diverses interprétations. Ollivier (2018) a établi un lien entre le filtre de Kalman étendu (Fahrmeir, 1992) et le gradient naturel en ligne de Murata and Amari (1999). De plus, Durbin and Koopman (2012) définit le problème de filtrage dans une configuration gaussienne linéaire comme un problème de recherche d'un estimateur linéaire non biaisé de variance minimale. Plusieurs variantes ont été développées pour adapter le filtre de Kalman à une gamme de problèmes. Ces adaptations ont exploré de nombreuses directions, incluant : (1) l'extension du filtrage aux scénarios non linéaires (Wan and Van Der Merwe, 2000; Fahrmeir, 1992); (2) le raffinement du filtre pour les cas où les hyper-paramètres du modèle sont inconnus (de Vilmarest and Wintenberger, 2021); (3) l'amélioration de la robustesse du filtre face aux perturbations du modèle, incluant la prise en compte à la fois du bruit statistiquement inconnu au sein du modèle espace-état et des incertitudes dans les variables du modèle dues à des erreurs déterministes ou stochastiques (Yedavalli, 2014; Zeng et al., 2012; Ra and Whang, 2008; Theodor and Shaked, 1996; Stengel and Ray, 1991; Ray and Stengel, 1991); et (4) la prise en compte des erreurs de modélisation où le modèle idéal est souvent mal défini, nécessitant parfois des hypothèses trop fortes qui peuvent s'écarter nettement des conditions réelles (Toda and Patel, 1980; Nishimura, 1970).

Notre étude aborde le problème (3), se concentrant sur la construction d'un estimateur robuste aux perturbations qui se manifestent comme du bruit dans les covariables. De même, notre étude s'inscrit dans la direction du problème (4) ; l'hypothèse courante que les covariables sont déterministes néglige leur nature intrinsèquement bruitée ou estimée. L'étude de la robustesse d'un estimateur se divise en deux grandes catégories : la première suppose une perturbation déterministe, sans aucun modèle probabiliste attribué à cette perturbation (Yedavalli, 2014). La seconde catégorie couvre les systèmes soumis à des perturbations stochastiques (Ra and Whang, 2008; Theodor and Shaked, 1996; Stengel and Ray, 1991).

Dans notre étude, nous considérons un modèle bien spécifié mais dans lequel les covariables sont bruitées. Ce cas des covariables soumises aux perturbations stochastiques se rencontre dans un large spectre d'applications. Par exemple, dans les sciences environnementales, elle est essentielle pour modéliser la dispersion des polluants et les prévisions du changement climatique, où les covariables comme les concentrations de polluants ou les mesures de température viennent avec leur propre ensemble d'incertitudes. Le secteur financier s'appuie sur des méthodologies similaires pour modéliser les indicateurs économiques et les prix des actifs, en intégrant la volatilité et les évaluations des risques pour prendre des décisions d'investissement éclairées. Dans le domaine de la santé, en particulier dans le

domaine de l'épidémiologie, la modélisation de la propagation des maladies avec des données incertaines est critique pour les stratégies de planification et de réponse. De plus, l'ingénierie, en particulier les systèmes de contrôle et la robotique, applique ces principes pour compenser les imprécisions des capteurs et pour affiner les réponses du système. La météorologie se démarque également, où la précision des prédictions météorologiques est considérablement améliorée en modélisant la perturbation dans les données atmosphériques.

## 2 Énoncé du problème

Nous explorons le modèle espace-état linéaire gaussien défini comme suit :

$$\begin{aligned}
 \text{Équation d'état :} & & \theta_t &= A\theta_{t-1} + \eta_t, & \eta_t &\sim \mathcal{N}(0, Q_t), \\
 \text{Équation des covariables :} & & x_t &= \tilde{x}_t + \Delta x_t, & \Delta x_t &\sim \mathcal{N}(0, R_t), \\
 \text{Équation d'espace :} & & y_t &= x_t^T \theta_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, \sigma_t^2).
 \end{aligned} \tag{1}$$

Ici,  $y_t$  et  $\varepsilon_t$  sont des nombres réels, tandis que  $\theta_t$ ,  $x_t$ ,  $\tilde{x}_t$ ,  $\Delta x_t$  et  $\eta_t$  sont des vecteurs dans  $\mathbb{R}^d$ . La transition de  $\theta_{t-1}$  à  $\theta_t$  est donnée par une matrice de transition  $A \in \mathbb{R}^{d \times d}$  supposée connue, plus un bruit gaussien  $\eta_t$  centré de matrice de covariance  $Q_t \in \mathbb{R}^{d \times d}$ . L'équation d'espace régit la relation entre l'observation  $y_t$ , les variables explicatives  $x_t$  et l'état  $\theta_t$ . Les hyperparamètres du modèle,  $Q_t$ ,  $R_t$  et  $\sigma_t^2$ , sont supposés être connus. Nous supposons que  $\Delta x_t$  suit une loi gaussienne centrée de matrice de covariance  $R_t$ .

Désignons par  $\mathcal{F}_m = \sigma(x_1, y_1, \dots, x_m, y_m)$  la filtration canonique représentant l'information disponible jusqu'au temps  $m$ . On se concentre sur les scénarios où le modèle devient incomplet ; c'est-à-dire que les valeurs de  $x_t$  manquent à partir d'un certain temps  $m$ , et seul  $\tilde{x}_t$  est observé. Nous souhaitons donner l'expression des moments d'ordre un et deux de la loi *a posteriori* de l'état. Formellement, nous exprimons ces quantités de la manière suivante :

$$\begin{aligned}
 \tilde{\theta}_{t|m} &= \mathbb{E}[\theta_t | \mathcal{F}_{m-1}, y_m], \\
 \tilde{P}_{t|m} &= \mathbb{E}[(\theta_t - \tilde{\theta}_{t|m})(\theta_t - \tilde{\theta}_{t|m})^T | \mathcal{F}_{m-1}, y_m].
 \end{aligned} \tag{2}$$

## 3 Présentation de l'approche

Commençons par rappeler l'expression du filtre de Kalman standard où les covariables  $x_t$  sont observées à chaque instant. Dans le cas du modèle espace-état linéaire gaussien, l'espérance de  $\theta_t$  conditionnellement à la filtration  $\mathcal{F}_t$  est une gaussienne de moyenne  $\hat{\theta}_{t|t}$  et de matrice de covariance  $P_{t|t}$  (Kalman and Bucy, 1961). Les quantités  $\hat{\theta}_{t|t}$  et  $P_{t|t}$  sont données dans le théorème 3.1.

**Théorème 3.1 (Filtre de Kalman)** *Soit  $t > 0$ , sous les hypothèses du modèle espace-état (1), la loi de  $\theta_t$ , conditionnellement à la filtration  $\mathcal{F}_t$ ,  $\theta_t$  suit une gaussienne de moyenne  $\hat{\theta}_{t|t}$*

et de matrice de covariance  $P_{t|t}$  dont les expressions sont :

$$\begin{aligned} P_{t|t} &= P_{t|t-1} - K_t x_t^T P_{t|t-1}, & P_{t+1|t} &= \Phi P_{t|t} \Phi^T + Q_t, \\ \hat{\theta}_{t|t} &= \hat{\theta}_{t|t-1} - K_t (y_t - x_t^T \hat{\theta}_{t|t-1}), & \hat{\theta}_{t+1|t} &= \Phi \hat{\theta}_{t|t}, \end{aligned}$$

où

$$K_t = \frac{P_{t|t-1} x_t}{x_t^T P_{t|t-1} x_t + \sigma_t^2}.$$

Ici,  $K_t$  est une matrice  $\mathbb{R}^{d \times 1}$ , appelée **matrice de gain**. Elle combine de manière optimale les nouvelles observations avec les estimations antérieures pour mettre à jour les prédictions de l'état en fonction de l'erreur en  $y_t$ . Utiliser  $\tilde{x}_t$  dans le théorème 3.1 induit des écarts qui deviennent plus grandes en fonction de la variance de la perturbation  $\Delta x_t$  (Figure 1).

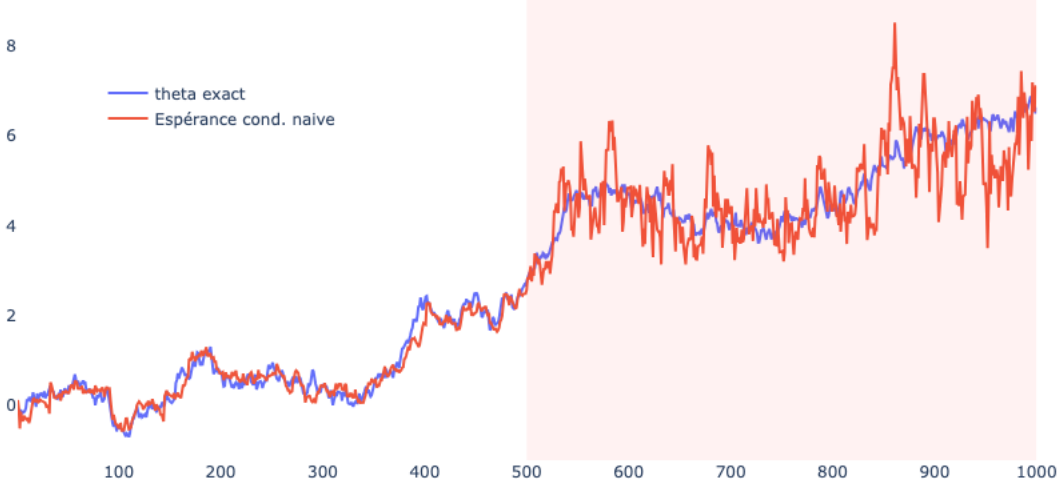


Figure 1: Simulation avec  $\tilde{x}_t = \sin(10\pi t)$ ,  $\Delta x_t \sim \mathcal{N}(0, 0.1)$ ,  $\sigma_t^2 = 0.1$ ,  $Q_t = 0.01$ ,  $\theta_{0|0} = 1$ ,  $P_{0|0} = 1$  et  $A = I$ . La valeur exacte de  $\theta_t$  est représentée en bleu, la courbe rouge représente la moyenne  $\hat{\theta}_{t|t}$ .  $x_t$  est observé jusqu'à  $t = 500$ . On observe qu'à partir  $t = 500$ , le filtre devie.

Dans le cas où  $x_t$  n'est pas observé, en conditionnant l'espérance conditionnelle  $\mathbb{E}[\theta_t | \mathcal{F}_{t-1}, y_t]$  par  $x_t$  (tower property), cela nous permet de reformuler l'espérance sous  $\theta_t$  définie dans l'équation (2) en une espérance de  $\hat{\theta}_{t|t}$  (Théorème 3.1) sous la loi *a posteriori* de  $x_t$ . Par la suite, on notera par  $f_{x_t | \mathcal{F}_{t-1}, y_t}$  la densité de probabilité de la loi *a posteriori* de  $x_t$ . Notre

démarche consiste d'abord à donner l'expression exacte de la densité de probabilité  $f_{x_t|\mathcal{F}_{t-1},y_t}$  à une constante de normalisation près. Cela nous permet par la suite de déterminer l'espérance de  $\hat{\theta}_{t|t}$  sous la loi *a posteriori* de  $x_t$ , et par conséquent  $\mathbb{E}[\theta_t|\mathcal{F}_{t-1},y_t]$ . Cette méthode offre un double avantage, car elle permet de filtrer à la fois le bruit sur  $x_t$  et sur  $\theta_t$  grâce à la nouvelle observation  $y_t$ . Cet avantage est particulièrement pertinent lorsque la variance de  $\Delta x_t$  est élevée, comme le montre la Figure 2.

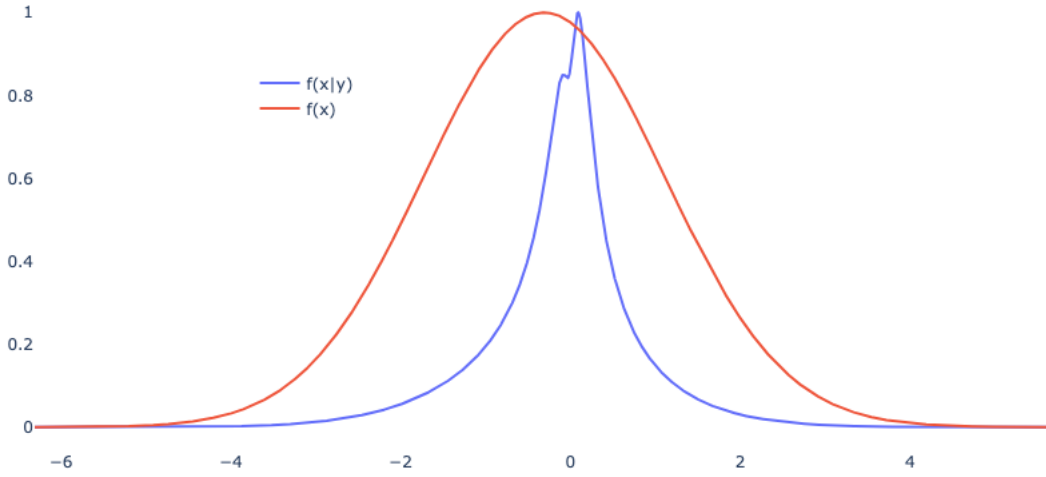


Figure 2: Simulation avec  $\tilde{x}_t = \sin(10\pi t)$ ,  $\Delta x_t \sim \mathcal{N}(0, 2)$ ,  $\sigma_t^2 = 0.1$ ,  $Q_t = 5$ ,  $\theta_{0|0} = 1$ ,  $P_{0|0} = 1$  et  $A = I$ . À  $t$  fixé et à une constante de normalisation près, le graphe rouge représente la densité  $f_{x_t} \sim \mathcal{N}(\tilde{x}_t, R_t)$  et le graphe bleu représente la densité de probabilité  $f_{x_t|\mathcal{F}_{t-1},y_t}$ .

## 4 Résultats théoriques

Commençons par considérer que l'instant  $t$  représente la première fois que nous n'observons pas  $x_t$  (c'est à dire  $t = m$ ). Pour simplifier les équations, nous considérons que la matrice de transition  $A$  est une matrice identité. Puisqu'à  $t - 1$ , nous avons observé  $x_{t-1}$ , l'espérance conditionnelle  $\hat{\theta}_{t|t-1} = \mathbb{E}[\theta_t|\mathcal{F}_{t-1}]$  et sa matrice de covariance  $P_{t|t-1} = \mathbb{E}[(\theta_t - \hat{\theta}_{t|t-1})(\theta_t - \hat{\theta}_{t|t-1})^T|\mathcal{F}_{t-1}]$  sont données de façon exacte par le filtre de Kalman. La densité de probabilité

$f_{\theta_t|\mathcal{F}_{t-1},y_t}$  de la loi *a posteriori* de  $\theta_t$  se dérive comme suit :

$$\begin{aligned} f_{\theta_t|\mathcal{F}_{t-1},y_t}(z) &= \int_{\mathbb{R}^d} f_{\theta_t|\mathcal{F}_{t-1},y_t,x_t=x}(z) \cdot f_{x_t|\mathcal{F}_{t-1},y_t}(x) dx \\ &= \int_{\mathbb{R}^d} f_{\mathcal{N}(\hat{\theta}_{t|t},P_{t|t})}(z) \cdot f_{x_t|\mathcal{F}_{t-1},y_t}(x) dx. \end{aligned} \quad (3)$$

La densité de probabilité résultante est une loi de mélange non finie. Les lois de mélange ont l'avantage d'avoir une forme assez simple et explicite des leur moments.

**Hypothèse 4.1** Soit  $t > 0$ , sous les hypothèses du modèle espace-état (1), supposons que l'intégrale

$$\int_{\mathbb{R}^d} x^2 f_{x_t|\mathcal{F}_{t-1},y_t}(x) dx$$

est finie.

**Proposition 4.1** Soit  $t > 0$ , sous l'hypothèse (4.1) et du modèle espace-état (1), les moments d'ordre un ( $\tilde{\theta}_{t|t}$ ) et deux ( $\tilde{P}_{t|m}$ ) de la loi *a posteriori* de  $\theta_t$  sont données par les expressions :

$$\begin{aligned} \tilde{\theta}_{t|t} &= \mathbb{E}[\theta_t|\mathcal{F}_{t-1},y_t] = \int_{\mathbb{R}^d} \hat{\theta}_{t|t} f_{x_t|\mathcal{F}_{t-1},y_t}(x) dx = \mathbb{E}_{x_t|\mathcal{F}_{t-1},y_t}[\hat{\theta}_{t|t}], \\ \tilde{P}_{t|t} + \left(\tilde{\theta}_{t|m}\right) \left(\tilde{\theta}_{t|m}\right)^T &= \int_{\mathbb{R}^d} \left(P_{t|t} + \hat{\theta}_{t|t} \left(\hat{\theta}_{t|t}\right)^T\right) f_{x_t|\mathcal{F}_{t-1},y_t}(x) dx = \mathbb{E}_{x_t|\mathcal{F}_{t-1},y_t} \left[ P_{t|t} + \hat{\theta}_{t|t} \left(\hat{\theta}_{t|t}\right)^T \right]. \end{aligned}$$

La proposition 4.1 est intéressante, car elle reformule le calcul des moments *a posteriori* de  $\theta_t$ , spécifiquement l'espérance et la variance, en termes d'espérances de fonctions dépendant de  $x_t$  sous la loi  $f_{x_t|\mathcal{F}_{t-1},y_t}$ . L'évaluation de ces espérances nécessite la connaissance de l'expression de  $f_{x_t|\mathcal{F}_{t-1},y_t}$ , à une constante de normalisation près, que nous dérivons en appliquant la règle de Bayes.

**Proposition 4.2** Soit  $t > 0$ , sous les hypothèses du modèle espace-état (1), à une constante de normalisation près, la densité de probabilité  $f_{x_t|\mathcal{F}_{t-1},y_t}$  a pour expression :

$$f_{x_t|\mathcal{F}_{t-1},y_t} \propto \frac{(2\pi)^{-(d+1)/2} |R_t|^{-1/2}}{(x_t^T P_{t|t-1} x_t + \sigma_t^2)^{1/2}} \exp(C) \exp\left(-\frac{1}{2} [(x_t - \Sigma_t K_t)^T \Sigma_t^{-1} (x_t - \Sigma_t K_t)]\right),$$

où

$$\begin{aligned} C &= -\frac{1}{2} \left( \frac{y_t^2}{x_t^T P_{t|t-1} x_t + \sigma_t^2} + \tilde{x}_t^T R_t^{-1} \tilde{x}_t - K_t^T \Sigma_t^T K_t \right), \quad \Sigma_t = \left( \frac{\hat{\theta}_{t|t-1} \hat{\theta}_{t|t-1}^T}{x_t^T P_{t|t-1} x_t + \sigma_t^2} + R_t^{-1} \right)^{-1}, \\ K_t &= \frac{\hat{\theta}_{t|t-1} y_t}{x_t^T P_{t|t-1} x_t + \sigma_t^2} + R_t^{-1} \tilde{x}_t. \end{aligned}$$

La connaissance de l'expression de la densité  $f_{x_t|\mathcal{F}_{t-1},y_t}$  à une constante de normalisation près nous permet de construire un estimateur consistant de  $\mathbb{E}_{x_t|\mathcal{F}_{t-1},y_t}[\hat{\theta}_{t|t}]$ , qui est lui-même un estimateur de  $\mathbb{E}[\theta_t|\mathcal{F}_{t-1},y_t]$ . À chaque  $t > 0$  fixé, les méthodes de Monte-Carlo nous permettent de construire un estimateur consistant de  $\mathbb{E}_{x_t|\mathcal{F}_{t-1},y_t}[\hat{\theta}_{t|t}]$  dont les bornes de l'erreur sont en  $\mathcal{O}(n^{-1/2})$  où  $n$  est la taille de l'échantillon  $\{x_i\}_{0 \leq i \leq n}$  distribué selon la loi  $f_{x_t|\mathcal{F}_{t-1},y_t}$ . La borne de l'erreur en  $\mathcal{O}(n^{-1/2})$  peut-être améliorée en utilisant les méthode Quasi-Monte Carlo ou les méthodes de Monte-Carlo par chaînes de Markov. Les expressions des moments d'ordre un et deux dans le théorème 4.1 nous permettent d'optimiser le temps de calcul dû à l'échantillonnage en utilisant un même échantillon de  $f_{x_t|\mathcal{F}_{t-1},y_t}$ .

## References

- Joseph de Vilmarest and Olivier Wintenberger. Viking: Variational bayesian variance tracking. *arXiv preprint arXiv:2104.10777*, 2021.
- James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*, volume 38. OUP Oxford, 2012.
- Ludwig Fahrmeir. Posterior mode estimation by extended kalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association*, 87(418):501–509, 1992.
- Rudolph E Kalman and Richard S Bucy. New results in linear filtering and prediction theory. 1961.
- Noboru Murata and Shun-ichi Amari. Statistical analysis of learning dynamics. *Signal Processing*, 74(1):3–28, 1999.
- T Nishimura. Modeling errors in kalman filters. Technical report, 1970.
- Yann Ollivier. Online natural gradient as a kalman filter. 2018.
- Won-Sang Ra and Ick-Ho Whang. Stochastic robust kalman filtering for linear time-varying systems with a multiplicative measurement noise. *IFAC Proceedings Volumes*, 41(2):12546–12551, 2008.
- Laura Ryan Ray and Robert F Stengel. Application of stochastic robustness to aircraft control systems. *Journal of Guidance, Control, and Dynamics*, 14(6):1251–1259, 1991.
- Robert F Stengel and Laura R Ray. Technical notes and correspondence: Stochastic robustness of linear time-invariant control systems. *NASA. Langley Research Center, Joint University Program for Air Transportation Research, 1990-1991*, 1991.
- Yahali Theodor and Uri Shaked. Robust discrete-time minimum-variance filtering. *IEEE Transactions on Signal Processing*, 44(2):181–189, 1996.
- M. Toda and R. Patel. Bounds on estimation errors of discrete-time filters under modeling uncertainty. *IEEE Transactions on Automatic Control*, 25(6):1115–1121, 1980. doi: 10.1109/TAC.1980.1102502.
- Joseph de Vilmarest. *Modèles espace-état pour la prévision de séries temporelles. Application aux marchés électriques*. PhD thesis, Sorbonne université, 2022.
- Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, pages 153–158. Ieee, 2000.
- Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. The MIT press, 1949.

Rama K Yedavalli. Robust control of uncertain dynamic systems. *AMC*, 10:12, 2014.

Caibin Zeng, YangQuan Chen, and Qigui Yang. Robust controllability of interval fractional order linear time invariant stochastic systems. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 4047–4050, 2012. doi: 10.1109/CDC.2012.6425949.