

# Approche bayésienne hiérarchique et dynamique pour la prédiction et l'agrégation de scores

---

Adrien Todeschini,  
Vincent Couallier, Marie Chavent



29 Mai 2018

50èmes Journées de Statistique, EDF Lab Paris Saclay

1. Scorelab
2. Global Wine Score
3. Projet PEPS
4. Filtre de Kalman multi-capteur
5. Modèle bayésien hiérarchique

Scorelab

---

- Startup bordelaise créée en avril 2016.
- Spécialisée en Data Science et Machine Learning.
- Des produits innovants.
- Des prestations sur mesure.



**Guillaume Forcade**

CEO & Co-founder



**Jean-Baptiste Pautrizel**

PhD Physics & Co-founder



**Adrien Todeschini**

PhD Stat. ML & Data Scientist



**Wassek Al Chahid**

Full Stack Dev.



**Kevin Baudin**

MSc. Stat. & Data Scientist



**Yang Zheng**

Stagiaire Data Scientist



**Ngoc-Phi Tran**

Stagiaire Data Scientist



**Lauren Flynn**

Marketing and Business  
development

- [globalwinescore.com](https://globalwinescore.com)
- Moteur de recommandation de vin
- Credit scoring dans le courtage immobilier
- Modèles prédictifs pour la finance (commodités, crypto-monnaies)
- ML4U : Recommandation personnalisée pour les conférences de Machine Learning
- [noodle.scorelab.io](https://noodle.scorelab.io) : Doodle for notation



# Global Wine Score

---

Scorelab collecte et structure une base de données importante d'évaluations de critiques de vins (~200k vins, 1M notes).

Elle a développé le site [www.globalwinescore.com](http://www.globalwinescore.com) qui fournit un score unique par vin millésimé à partir de l'ensemble des notes collectées.

[Rankings](#) [Critics](#) [Blog](#)



One single score, aggregated from critics.

COLOR: Red  SCORE GWS 70 - 100  COUNTRY (OPTIONAL): Choose a country

### LATEST GWS RANKINGS

#### TOP RED BORDEAUX PRIMEURS 2016

1. Chateau Latour, Pauillac	99,00
2. Chateau Haut Brion, Pessac Leognan	98,92
3. Chateau Mouton Rothschild, Pauillac	98,69
4. Petrus, Pomerol	98,66
5. Chateau Ausone, Saint Emilion Grand Cru	98,49

#### TOP WHITE BORDEAUX PRIMEURS 2016

1. Chateau Haut Brion, Blanc, Pessac Leognan	95,40
2. Chateau Smith Haut Lafitte, Blanc, Pessac Leognan	94,22
3. Chateau Margaux, Pavillon Blanc Du Chateau Margaux, Blanc, Bordeaux	94,02
4. Chateau La Mission Haut Brion, Blanc, Pessac Leognan	93,82
5. Domaine De Chevalier, Blanc, Pessac Leognan	93,34

#### TOP SWEET PRIMEURS 2016

1. Chateau D Yquem, Blanc, Sauternes	97,29
2. Chateau Climens, Blanc, Barsac	95,41
3. Chateau Doly Daene, L Extravagant De Doly Daene, Blanc, Sauternes	95,37
4. Chateau De Fargues, Blanc, Sauternes	94,80
5. Chateau Suduiraut, Blanc, Sauternes	94,65



Search a wine


[Ranking](#)
[Critics](#)
[Blog](#)
[Home](#) / [Ranking](#)

Only Bordeaux en Primeurs are available while we update our database.

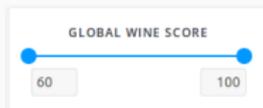
[Primeurs](#) × [Red](#) × [2016](#) ×

 Show **25** entries

 Search: 

**QUICK LINKS**

- Bordeaux Primeurs Red 2016
- Bordeaux Primeurs White 2016
- Bordeaux Primeurs Sweet 2016



**CATEGORY** ▾

[Primeurs](#) ×

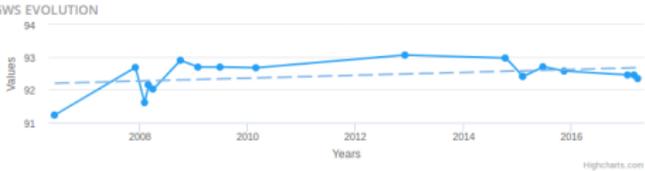
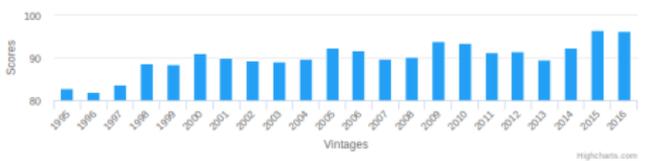
**COUNTRY** ▾

France

Italy

	#	Wine	Vintage	GWS	CI	NBJ
1	1.	Chateau Latour, Pauillac	2016	99,00	A+	17
2	2.	Chateau Haut Brion, Pessac Leognan	2016	98,92	A+	17
3	3.	Chateau Mouton Rothschild, Pauillac	2016	98,69	A+	18
4	4.	Petrus, Pomerol	2016	98,66	A+	18
5	5.	Chateau Ausone, Saint Emillion Grand Cru	2016	98,49	A+	18
6	6.	Chateau Lafite Rothschild, Pauillac	2016	98,42	A+	18
7	7.	Chateau Lafleur, Pomerol	2016	98,40	A+	16
8	8.	Chateau Leoville Las Cases, Grand Vin De Leoville, Saint Julien	2016	98,16	A+	19
9	9.	Chateau Margaux, Margaux	2016	97,91	A+	18
10	10.	Le Pin, Pomerol	2016	97,89	A+	14
11	11.	Chateau Cheval Blanc, Saint Emillion Grand Cru	2016	97,87	A+	18

Home / France / Saint Emilion Grand Cru / Chateau Canon, Saint Emilion Grand Cru, 2005

NAME Chateau Canon, Saint Emilion Grand Cru, 2005		
COUNTRY France	APPELLATION Saint Emilion Grand Cru	COLOR Red
GLOBAL WINE SCORE  <b>92,34</b> /100 		
Updated: 1 avril 2017		Share 
GWS CALCULATION BASED ON <b>10</b> Wine Critics 		RANKING Saint Emilion Grand Cru, 2005 Grand Cru 92,40 14. Chateau Pavle Macquin, Saint Emilion Grand Cru 92,38 15. Chateau Larcis Ducasse, Saint Emilion Grand Cru 92,37 16. Chateau Canon, Saint Emilion Grand <b>92,34</b>
CONFIDENCE INDEX <b>A</b> 		
GWS EVOLUTION 		
Scores 		

- l'hétérogénéité des échelles de notation et de leur granularité.
- hétérogénéité de la perception de la qualité de chaque évaluateur.
- hétérogénéité des ensembles de vins dégustés.
- l'évolution temporelle des notes.

- Collaboration avec le Larefi (Jean-Marie Cardebat)
- Collaboration avec l'IMB (Vincent Couallier, Marie Chavent)
- Projet PEPS en collaboration avec l'IMB, financé par l'AMIES.

# Projet PEPS

---

- Projet PEPS-2 de neuf mois de novembre 2016 à août 2017.
- Les missions ont tourné autour de trois axes :
  1. l'étude et la validation de l'algorithme d'agrégation de notes du Global Wine Score.
  2. l'exploration d'un nouveau modèle de scoring pour réduire l'incertitude sur les vins ayant reçu peu de notes.
  3. l'exploration d'algorithmes de machine learning pour la recommandation personnalisée de vins.
- La suite de l'exposé se concentre sur le point 2.

- Le Global Wine Score n'est calculé que pour les vins millésimés ayant reçu **au moins trois notes** de critiques.
- Scorelab souhaite enrichir le score d'un *a priori* lié
  - aux évaluations du vin sur des millésimes antérieurs
  - ainsi que l'évaluation de vins similaires (appellation, millésime).

- Un **modèle bayésien hiérarchique** [Gelman et al., 2014] permettant d'estimer le score moyen ainsi que la variance d'une appellation millésimée.  
La formulation hiérarchique permet d'emprunter de l'information aux appellations voisines et ainsi régulariser l'estimation des appellations ayant peu de notes.
- Les scores résiduels associés à chaque vin ou château sont modélisés par un **filtre de Kalman multi-capteur** [Willner et al., 1976] dont les bruits d'évolution et de mesure sont calibrés par maximum *a posteriori*.

# Filtre de Kalman multi-capteur

---

- Soit  $z_{ijk}$  la note du critique  $j$  attribuée au vin  $i$  de millésime  $k$ .
- Soit  $n_{ik}$  le nombre de critiques ayant noté le vin  $i$  millésime  $k$ .
- On s'intéresse ensuite au comportement du résidu normalisé de la note  $z_{ijk}$  par rapport au niveau moyen de son appellation au millésime  $k$ .
- On note  $\tilde{t} = \sqrt{\tau_a} (t - \nu_a)$  le changement de variable où  $\nu_a$  et  $\tau_a$  sont le score moyen et la précision de l'appellation du vin  $i$  au millésime  $k$ .
- Soit  $\tilde{x}_{ik}$  le score normalisé à estimer du vin  $i$  millésime  $k$ .

On suppose le modèle suivant pour tout  $k \geq 1$  et  $j \in J_{ik}$  :

$$\tilde{x}_{i1} \sim \mathcal{N}(\tilde{\mu}_{1|0}, \tilde{\sigma}_{1|0}^2) \quad (\text{initialisation})$$

$$\tilde{x}_{ik} | \tilde{x}_{i,k-1} \sim \mathcal{N}(\tilde{x}_{i,k-1}, \tilde{\sigma}_{xi}^2), \quad k > 1 \quad (\text{évolution})$$

$$\tilde{z}_{ijk} | \tilde{x}_{ik} \sim \mathcal{N}(\tilde{x}_{ik}, \tilde{\sigma}_{zi}^2) \quad (\text{mesure})$$

où  $\tilde{\mu}_{1|0} = \mathbf{0}$  et  $\tilde{\sigma}_{1|0}^2 = \mathbf{1}$  sont des paramètres initiaux,  $\tilde{\sigma}_{xi}^2$  et  $\tilde{\sigma}_{zi}^2$  sont les variances de bruit d'évolution et de mesure du vin  $i$ .

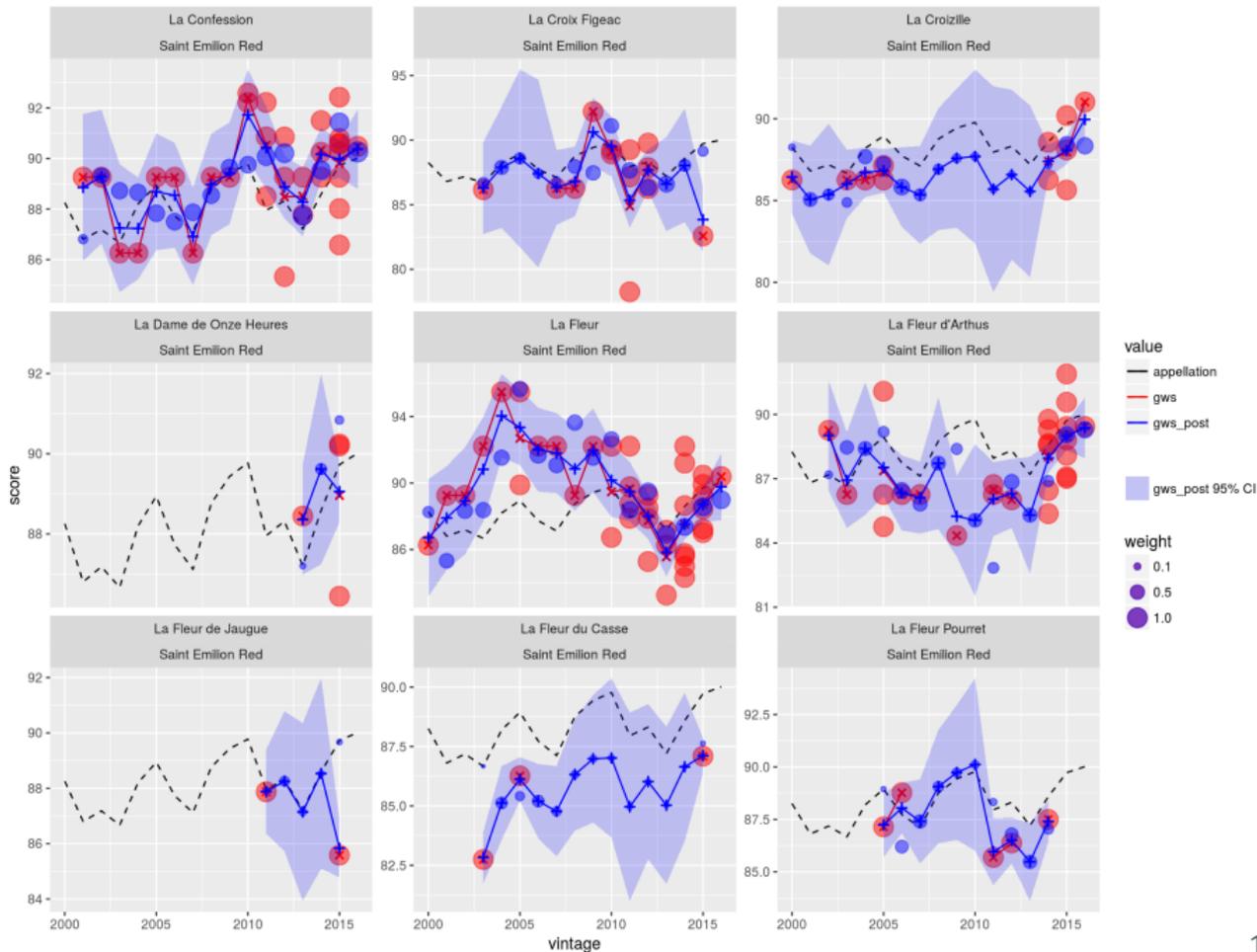
- Les équations de Kalman, fournissent de manière analytique une **prédiction** sur le score d'un vin

$$\mu_{ik} = \mathbb{E} [x_{ik} | z_{ik-1}, \dots, z_{i1}]$$

avant que celui-ci n'ait reçu la moindre notation.

- Le **score bayésien final (*a posteriori*)** est une moyenne pondérée des notes de critiques observées et de la prédiction dont le poids  $\omega_{ik}$  est relatif à son degré de confiance

$$s_{ik} = \mathbb{E} [x_{ik} | z_{ik}, \dots, z_{i1}] = \frac{1}{\omega_{ik} + n_{ik}} \left( \omega_{ik} \mu_{ik} + \sum_{j \in J_{ik}} z_{ijk} \right).$$



- Par ailleurs, grâce à l'ajout d'un *hyperprior*

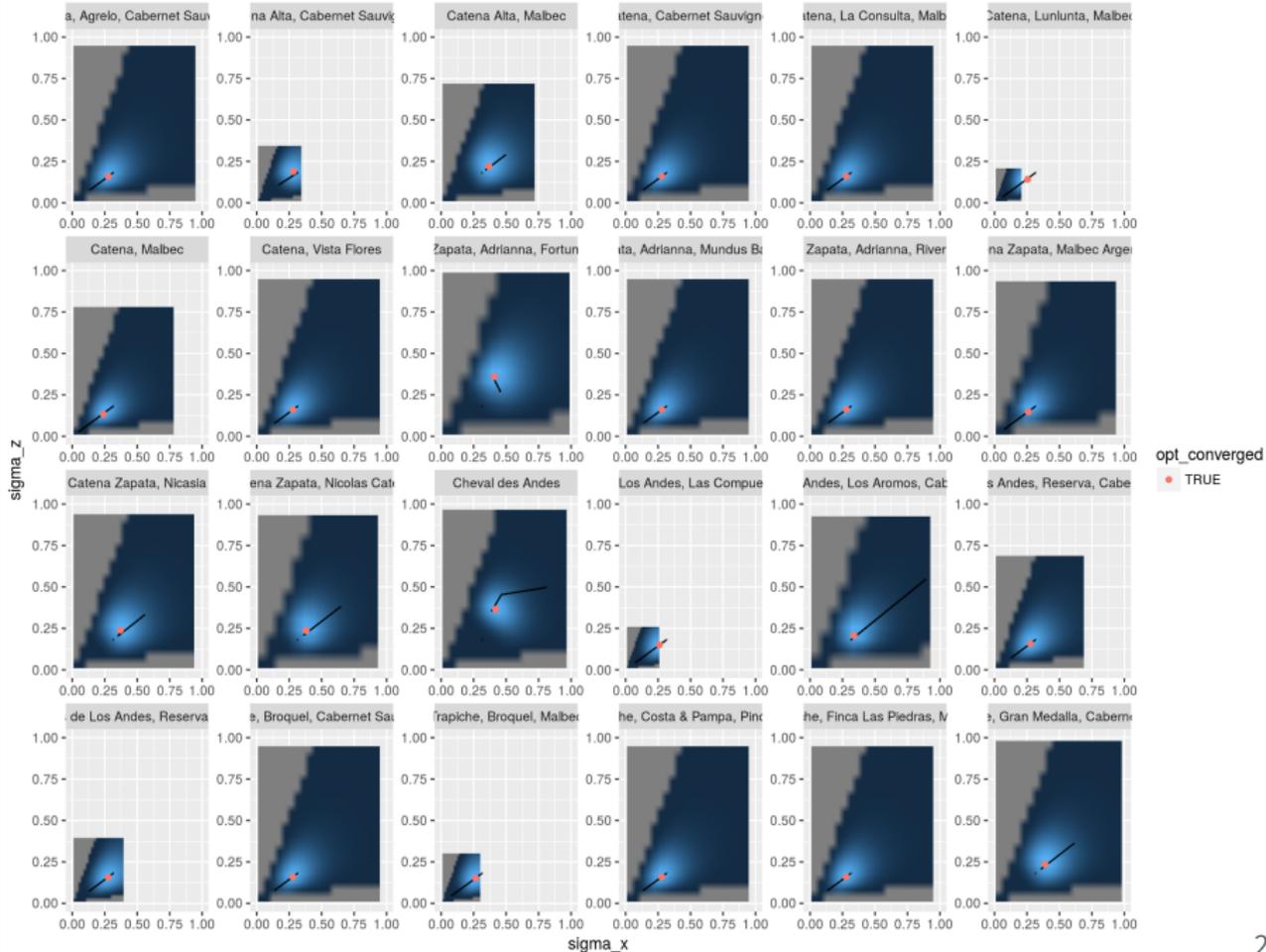
$$p(\tilde{\sigma}_{xi}^2, \tilde{\sigma}_{zi}^2)$$

et le calcul de la **vraisemblance marginale** (analytique par les équations de Kalman)

$$p(z_{i1}, \dots, z_{ik_{\max}} | \tilde{\sigma}_{xi}^2, \tilde{\sigma}_{zi}^2)$$

nous pouvons estimer les deux paramètres de bruit d'évolution de mesure par **maximum a posteriori**.

- Nous avons en outre adopté une approche **bayésienne empirique** pour définir l'*hyperprior*.



# Modèle bayésien hiérarchique

---

- Pour estimer les scores moyens d'appellations, on suppose la vraisemblance suivante

$$z_{ijk} | \nu_a, \tau_a \sim \mathcal{N}(\nu_a, \tau_a^{-1}), j = 1, \dots, n_{ik}$$

où  $\nu_a$  et  $\tau_a$  sont le score moyen et la précision de l'appellation du vin  $i$  au millésime  $k$ .

- Les appellations  $a$  sont considérées comme les nœuds d'une hiérarchie basée sur les terroirs (e.g. Margaux < Médoc < Bordeaux < France < Monde).

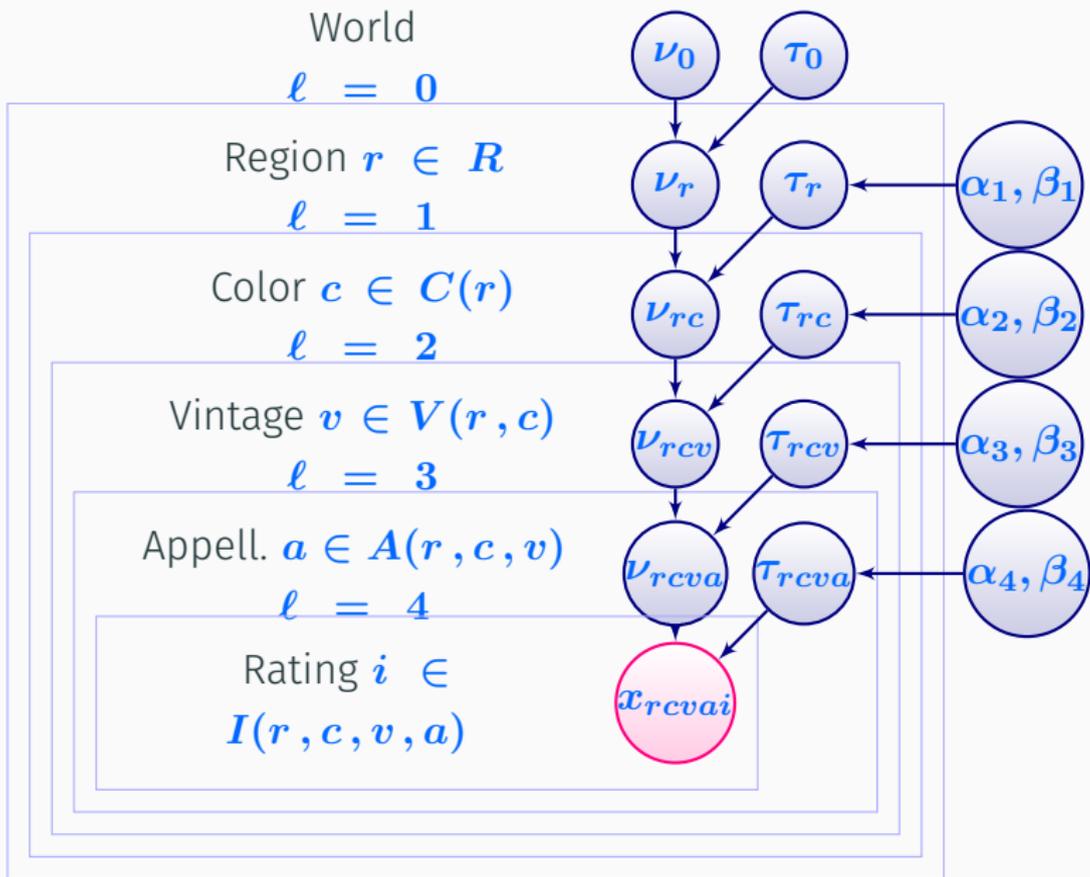
- Pour tout nœud  $a$  appartenant à l'arbre hiérarchique

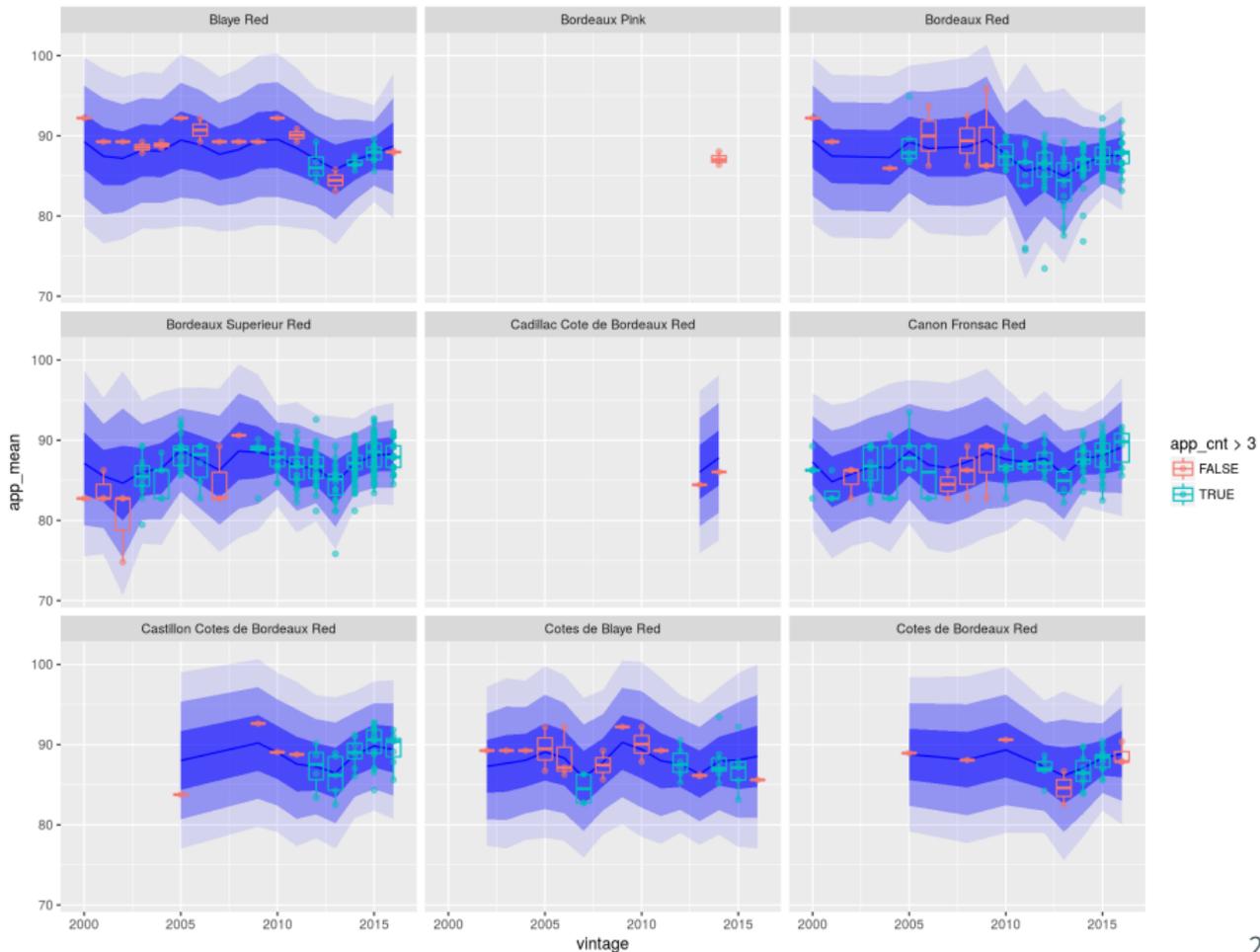
$$\nu_a | \nu_{p(a)}, \tau_{p(a)} \sim \mathcal{N} \left( \nu_{p(a)}, \tau_{p(a)}^{-1} \right)$$

$$\tau_a | \alpha_{\ell(a)}, \beta_{\ell(a)} \sim \Gamma \left( \alpha_{\ell(a)}, \beta_{\ell(a)} \right)$$

où  $p(a)$  est le nœud parent et  $\ell(a)$  est le niveau du nœud  $a$ .

- On suppose de plus un *a priori* vague sur les hyperparamètres du modèle graphique.
- Inférence via méthodes **MCMC** [Gilks, 2005] à l'aide du logiciel **JAGS** [Plummer et al., 2003]. Rapide car les *a priori* sont conjuguées.





## Conclusion

- Le modèle bayésien permet de capturer l'incertitude des quantités à estimer.
- Il est compatible avec un nombre variable d'observations à chaque temps.
- Le score *a posteriori* s'interprète facilement comme une moyenne pondérée des observations et de la prédiction.

## Perspectives

- Développer un modèle unique de type filtre de Kalman hiérarchique.
- Comparer les performances prédictives avec d'autres approches purement machine learning.



Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014).

***Bayesian data analysis, volume 2.***

CRC press Boca Raton, FL.



Gilks, W. R. (2005).

***Markov chain Monte Carlo.***

Wiley Online Library.



Plummer, M. et al. (2003).

**JAGS : A program for analysis of bayesian graphical models using Gibbs sampling.**

*In Proceedings of the 3rd international workshop on distributed statistical computing, volume 124, page 125. Vienna, Austria.*



Willner, D., Chang, C., and Dunn, K. (1976).

**Kalman filter algorithms for a multi-sensor system.**

*In Decision and Control including the 15th Symposium on Adaptive Processes, 1976 IEEE Conference on, volume 15, pages 570–574. IEEE.*

Merci de votre attention!



<https://scorelab.io>

 [scorelab\\_io](#)

[adrien@scorelab.io](mailto:adrien@scorelab.io)

 [adrtd](#)