

Classification et prévision des habitudes de consommation dans un réseau d'eau intelligent

Doctorant : M. Leyli-Abadi

Directeur de thèse : A. Samé

Co-directrice de thèse : L. Oukhellou

Université Paris-Est Créteil, IFSTTAR, COSYS, GRETTIA, France

3èmes Journées de l'Automatique du GDR MACS (JAMACS 2018)

Le 16 novembre, 2018

Plan

1 Introduction

- Ressources en eau
- État de l'art
- Données

2 Méthodologie

3 Modélisation probabiliste

- Mélange de modèles de Markov à variables exogènes

4 Résultats

5 Discussion

6 Conclusion

Ressources en eau

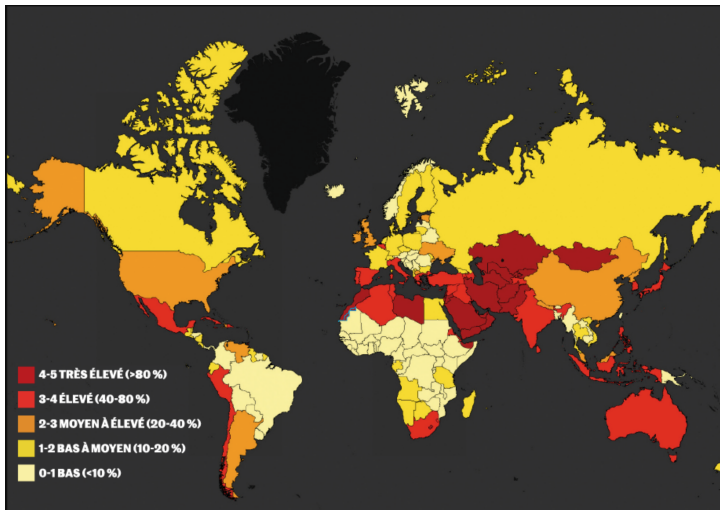
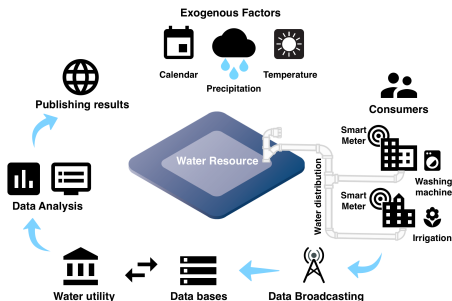


Figure: Carte du stress hydrique

Réseau d'eau intelligent



Réseaux d'eau intelligents

- Relevé à distance de données sur la consommation (smart meters et réseau sans fil)
- Exploitation des données pour une gestion plus efficace des réseaux

État de l'art sur la prévision de la demande en eau

- Regression

- Approche de régression pour modéliser la demande en eau en fonction des facteurs exogènes (tarification, facteurs climatiques, etc) (Kenney et al., 2008)
- Modèle de régression linéaire en conjonction avec l'algorithme de K-means pour la prédiction de consommation hebdomadaire dans (Aksela et al., 2010)

- Réseaux de neurones

- Prédiction de la demande en eau utilisant les méthodes statistiques et ANN dans (Altunkaynak et al., 2017) et (Walker et al., 2015)

- Modèle de Markov

- Méthode basée sur les chaines de Markov pour la prévision à court terme de la demande (Gagliardi et al., 2017)
- Méthode basée sur le SAX et les chaines de Markov pour la classification de la dynamique des habitudes de consommation en électricité (Wang et al., 2016)

Description des données

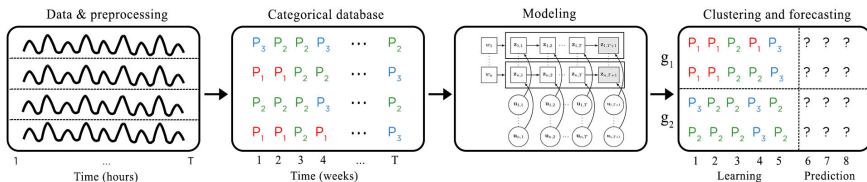
- 2 000 compteurs intelligents
- Type de résident:
 - individuels
 - collectifs
- Fréquence: horaire
- La durée: 19 mois
 - De Janvier 2015
 - À Juillet 2016
- Variables exogènes utilisées:
 - Variables climatiques
 - Calendrier
- Données exogènes non-étudiées:
 - Données socioéconomique
 - Données démographiques



Géolocalisation des compteurs

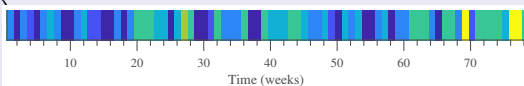
Démarche adoptée

- Prétraitement
- Modélisation
- Classification et prévision

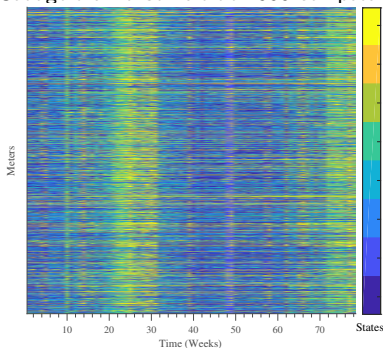


Construction d'une base de données

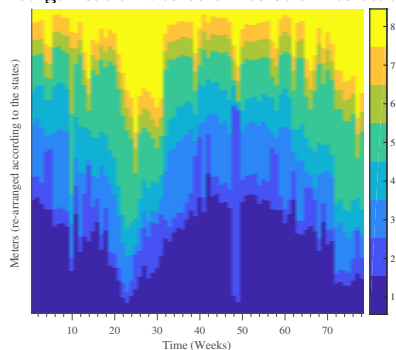
Représentation de chaque compteur par la série catégorielle de ses labels d'habitudes hebdomadaire de consommation (8 modalités appelées également "états")



Codage d'un ensemble de 2000 compteurs



Réorganisation des colonnes selon les états



Modélisation

- **Objectifs** : analyse prédictive des habitudes de consommation hebdomadaire sur un réseau d'eau potable
 - Identifier la dynamique temporelle des habitudes de consommation
 - Identifier les classes de consommateurs (clusters)
 - Réaliser des prévisions de futures habitudes (états)

Mélange de modèles de Markov (MixJNHMM)

• Notation

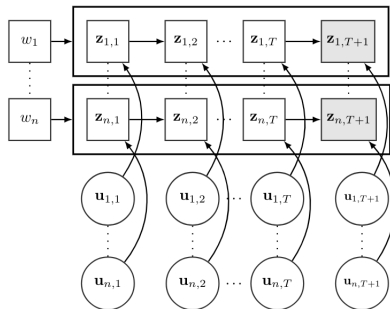
- \mathbf{w}_i : label associé à une série temporelle \mathbf{z}_i

• Modèle

$$f(\mathbf{z}_i | \mathbf{u}_i; \psi) = \sum_g p_g f_g(\mathbf{z}_i | \mathbf{u}_i; \theta_g)$$

où

$$f_g(\mathbf{z}_i | \mathbf{u}_i; \theta_g) = \prod_{\ell=1}^K \left[\frac{e^{\alpha_{g\ell}^\top \mathbf{u}_{i1}}}{\sum_{\ell=1}^K e^{\alpha_{g\ell}^\top \mathbf{u}_{i1}}} \right]^{z_{i1\ell}} \prod_{t=2}^T \prod_{\ell=1}^K \prod_{k=1}^K \left[\frac{e^{\beta_{g\ell k}^\top \mathbf{u}_{it}}}{\sum_{k=1}^K e^{\beta_{g\ell k}^\top \mathbf{u}_{it}}} \right]^{z_{itk} z_{it-1\ell}}$$



Mélange des modèles de Markov (suite)

Maximisation de la log-vraisemblance complétée

$$\begin{aligned}
 \mathcal{L}_c(\boldsymbol{\theta}, \mathbf{w}) &= \log P(\mathbf{z}, \mathbf{w} \mid \mathbf{u}; \boldsymbol{\theta}) \\
 &= \sum_{g=1}^G \left[\sum_{i=1}^n w_{ig} \log p_g + \sum_{i=1}^n \sum_{\ell=1}^K w_{ig} z_{i1\ell} \log \pi_{g\ell}(\mathbf{u}_{i1}; \boldsymbol{\alpha}_g) \right. \\
 &\quad \left. + \sum_{i=1}^n \sum_{t=2}^T \sum_{k=1}^K \sum_{\ell=1}^K w_{ig} z_{itk} z_{i(t-1)\ell} \log \pi_{g\ell k}(\mathbf{u}_{it}; \boldsymbol{\beta}_{g\ell}) \right]
 \end{aligned}$$

Proposition d'un nouvel algorithme CEM : maximisation alternée par rapport à \mathbf{w} et $\boldsymbol{\theta}$

Algorithm CEM proposé

- Étape-E : estimation des probabilités a posteriori

$$s_g^{(c)}(\mathbf{z}_i) = \frac{p_g^{(c)} f_g(\mathbf{z}_i, \boldsymbol{\theta}_g^{(c)})}{\sum_{g'=1}^G p_{g'}^{(c)} f_{g'}(\mathbf{z}_i, \boldsymbol{\theta}_{g'}^{(c)})}$$

- Étape-C : attribution des labels \mathbf{w}_i aux observations \mathbf{z}_i , suivant

$$\mathbf{w}_i^{(c+1)} = \arg \max_{1 \leq g \leq G} s_g^{(c)}(\mathbf{z}_i)$$

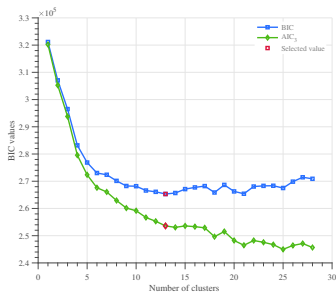
- Étape-M : mise à jour des paramètres utilisant les estimations du maximum de vraisemblance $(\boldsymbol{\alpha}_g^{(c+1)}, \boldsymbol{\beta}_{g\ell}^{(c+1)})$

$$p_g^{(c+1)} = \frac{\sum_{i=1}^n \mathbf{w}_{ig}^{(c)}}{n}, \quad \forall g = \{1, \dots, G\}$$
$$\mathcal{L}_1(\boldsymbol{\alpha}_g) + \sum_{\ell=1}^K \mathcal{L}_{2,\ell}(\boldsymbol{\beta}_{g\ell})$$

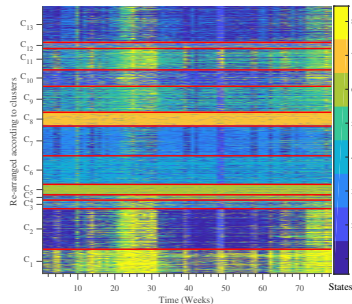
Sélection de modèles et résultat de clustering

Critère ICL

$$ICL(G, K, m) = \mathcal{CL}(\hat{\theta}) - \frac{\vartheta(G, K, m)}{2} \log(n \times T)$$

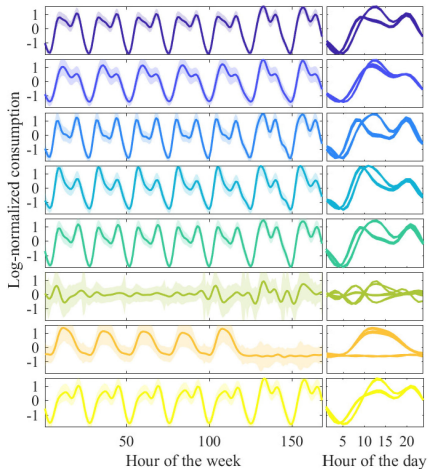


Choix du nombre de classes

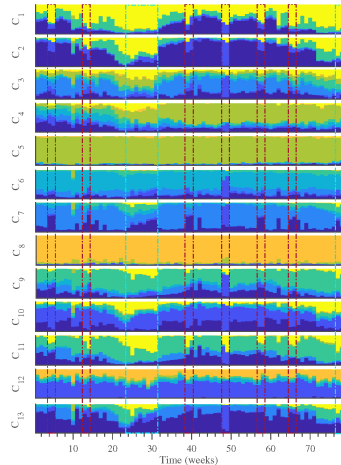


Résultat de classification obtenu utilisant le modèle de mélange. Classes (C_1, \dots, C_{13}) séparées par les lignes horizontales.

Interprétation des clusters



Habitudes de consommation principales
(états hebdomadaires)

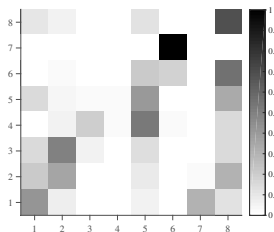


Re-organisation des compteurs selon
les états pour chaque cluster

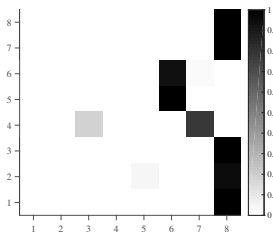
Résultats (matrices de transition globale)

Matrice de transition

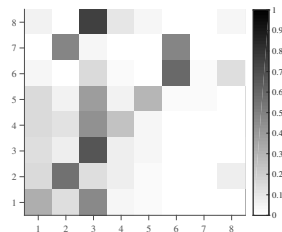
- Contient les probabilités de transition entre les états
- Représentation des probabilités en couleur (couleur sombre : probabilité proche à un)



Cluster 1



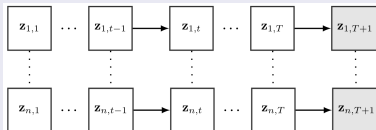
Cluster 5



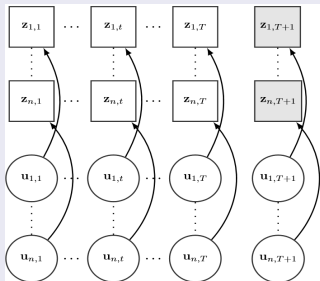
Cluster 7

Méthodes évaluées

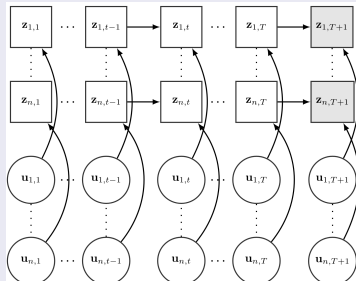
Modèle de Markov homogène



Modèle à états indépendants



Modèle de Markov non-homogène



Résultats (prédiction)

Prédiction

$$\hat{\mathbf{z}}_{iT+1} = \arg \max_k P(\mathbf{z}_{iT+1} = k \mid \mathbf{z}_{iT}, w_i = g, \mathbf{u}_{iT})$$

SIM: modèle à états indépendants, MM: modèle de Markov homogène, JNMM: modèle proposé

T_t : Température au temps t , P_t : Précipitation au temps t , C_t : Consommation au temps t

Models	Inputs (\mathbf{u}_{it})	Metrics			
		Accuracy	Recall	Precision	F-Measure
MM		0.60	0.61	0.61	0.61
MixHMM		0.60	0.62	0.61	0.62
JNMM	(T_{t-1}, P_{t-1})	0.64	0.65	0.65	0.65
	(W_{t-1})	0.69	0.70	0.70	0.70
	$(W_{t-1}, T_{t-1}, P_{t-1})$	0.69	0.70	0.70	0.70
	(W_{t-1}, C)	0.71	0.71	0.71	0.71
	$(W_{t-1}, C, T_{t-1}, P_{t-1})$	0.72	0.72	0.72	0.72
K-means+JNMM	(T_{t-1}, P_{t-1})	0.67	0.66	0.67	0.67
	(W_{t-1})	0.73	0.74	0.73	0.74
	$(W_{t-1}, T_{t-1}, P_{t-1})$	0.74	0.73	0.74	0.73
	(W_{t-1}, C)	0.73	0.73	0.72	0.73
	$(W_{t-1}, C, T_{t-1}, P_{t-1})$	0.75	0.74	0.75	0.75
MixJNHMM	(T_{t-1}, P_{t-1})	0.70	0.71	0.69	0.70
	(W_{t-1})	0.78	0.76	0.77	0.77
	$(W_{t-1}, T_{t-1}, P_{t-1})$	0.79	0.80	0.78	0.79
	(W_{t-1}, C)	0.78	0.77	0.75	0.76
	$(W_{t-1}, C, T_{t-1}, P_{t-1})$	0.80	0.80	0.79	0.79

Discussion

Modèle de Markov homogène (MM)

- Avantages
 - Complexité algorithmique faible
 - La matrice de transition résume la dynamique de toutes les données
- Désavantages
 - Pas de prise en compte des variables exogènes
 - Difficulté à capter les passages semaine-weekend

Modèle à états indépendants avec l'entrée

- Avantages
 - Complexité algorithmique faible
 - Erreur d'apprentissage assez faible
- Désavantages
 - Pas de prise en compte de la régularité temporelle
 - Erreur de prédiction assez élevée par rapport à l'apprentissage

Discussion (suite)

Modèle de Markov non-homogène

- Avantages
 - Prise en compte de la régularité temporelle et des variables exogènes
 - Prévisions plus précises
- Désavantages
 - Complexité algorithmique $\mathcal{O}(N \times (N - 1) \times d^2)$
 - Pas assez précis pour les séries temporelles présentant des dynamiques hétérogènes

Mélange des Modèle de Markov

- Avantages
 - Prise en compte de dynamiques locales
- Désavantages
 - Complexité algorithmique $\mathcal{O}(G \times N \times (N - 1) \times d^2)$

Conclusion et perspectives

Conclusion

- Proposition de méthodes probabiliste basées sur les modèles de Markov pour la prévision des habitudes de consommation
- Meilleure précision atteinte prenant en compte les dépendances avec les variables exogènes et les états précédants
- Amélioration atteinte utilisant le mélange des modèles de Markov sur les données réelles de consommation d'eau

Perspectives

- Introduction des variables exogènes spatiales au modèle
- Adaptation du modèle proposé pour la détection de changement du comportement de consommation

References I



Douglas S Kenney, Christopher Goemans, Roberta Klein, Jessica Lowrey, and Kevin Reidy.

Residential water demand management: lessons from aurora, colorado1, 2008.



K Aksela and M Aksela.

Demand estimation with automated meter reading in a distribution network.

Journal of Water Resources Planning and Management, 137(5):456–467, 2010.



Abdusselam Altunkaynak and Tewodros Assefa Nigussie.

Monthly water consumption prediction using season algorithm and wavelet transform–based models.

Journal of Water Resources Planning and Management, page 04017011, 2017.



David Walker, Enrico Creaco, Lydia Vamvakeridou-Lyroudia, Raziye Farmani, Zoran Kapelan, and Dragan Savić.

Forecasting domestic water consumption from smart meter readings using statistical methods and artificial neural networks.

Procedia Engineering, 119:1419–1428, 2015.

References II



Francesca Gagliardi, Stefano Alvisi, Zoran Kapelan, and Marco Franchini.

A probabilistic short-term water demand forecasting model based on the markov chain.

Water, 9(7):507, 2017.



Yi Wang, Qixin Chen, Chongqing Kang, and Qing Xia.

Clustering of electricity consumption behavior dynamics toward big data applications.

IEEE transactions on smart grid, 7(5):2437–2447, 2016.

Merci pour votre attention !

Questions ?