Prévision du rayonnement solaire : classification en journées types et combinaisons de modèles statistiques

Emmanuelle Héritier

AGROCAMPUS OUEST

☐ **CFR Angers**

☑ **CFR Rennes**

Année universitaire : 2013-2014

Spécialité :

Agronome

Spécialisation (et option éventuelle) :

Statistiques Appliquées

Mémoire de Fin d'Études

☑ d'Ingénieur de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage

☐ de Master de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage

☐ d'un autre établissement (étudiant arrivé en M2)

Prévision du rayonnement solaire: Classification en journées types et combinaisons de modèles statistiques

Par : Emmanuelle HERITIER

CONFIDENTIEL pendant 5 ans

Soutenu à Rennes le* 9 Septembre 2014

Devant le jury composé de :

Président : Julie Josse

Maître de stage : Laurent Huet

Enseignant référent : David Causeur

Remerciements

Je tiens tout d'abord à remercier Nicolas Schmutz pour m'avoir accueilli dans son entreprise.

Merci également à mon maître de stage, Laurent Huet pour ses conseils.

Je tiens aussi à remercier Pierre-Julien Trombe pour son encadrement, ainsi qu'Olivier Liandrat pour son aide technique.

Sans oublier tout le reste de l'équipe, salariés et stagiaires : Nicolas Sébastien, Ines (miam les gateaux !) et Sam pour m'avoir, entre autres, prêté son bras gauche pour me défouler. La team atelier : Clément pour ses conseils en rando et en règle de 3 ainsi que PE, compagnon de route presque infaillible, pour sa bienveillance. La coloc trop stylée : Etienne, Mehdi, Simon et Hugo. Philippine pour notre amour de la RMSE, nos fous rires et j'en passe. Adri, Benji, TDog, Fred, Matthieu, Nico Nitche et Caro pour sa solidarité féminine et agronomienne.

Un grand merci au labo de statistiques de Rennes qui a su me transmettre le goût et même la passion des stat. Un merci plus particulier à David Causeur pour son écoute et pour avoir toujours su trouver les mots pour me motiver.

Merci également à tous mes amis de la Réunion qui ont bien rempli mes heures en dehors du travail et fait de ce stage une expérience humaine très enrichissante.

Mention particulière pour Batman…pour… mince je crois que la liste est longue ! Merci pour ton intégrité, de m'avoir aidé dans l'affirmation de mes valeurs et de moi-même. Yes we can !

Merci à mes amis de métropole pour leur soutien à distance. Enfin, un grand merci à ma famille (de sang et de cœur), pilier infaillible, pour m'avoir soutenu et laissé voler de mes propres ailes.

Tables des matières

Glossaire des abréviations

Sigles

ANN	Artificial Neural Network
ARER	Agence Régionale Energie Réunion
ARMA	AutoRegressive Moving Average
CGGD	Commissariat Général au Développement Durable
EPIA	European Photovoltaic Industry Association
GHI	Global Horizontal Irradiance
KC	Indice de Ciel Clair
NRMSE	Normalised Root Mean Square Error
OPECST	Office Parlementaire d'Evaluation des Choix Scientifiques et Technologiques
RMSE	Root Mean Square Error

Unités

$W.m^{-2}$: Watt par mètre carré.

MW : MégaWatt

Liste des illustrations

Listes des tableaux

Liste des annexes

Introduction

De nos jours, une réelle volonté de développer les énergies renouvelables existe. En effet, nous parlons de transition énergétique qui désigne le passage du système énergétique actuel utilisant des ressources non renouvelables vers un bouquet énergétique basé principalement sur des ressources renouvelables. Ainsi, l'insertion d'énergies renouvelables dans le réseau électrique est une des problématiques actuelles. L'énergie photovoltaïque connait d'ailleurs une réelle expansion comme le prouve la Figure 1.

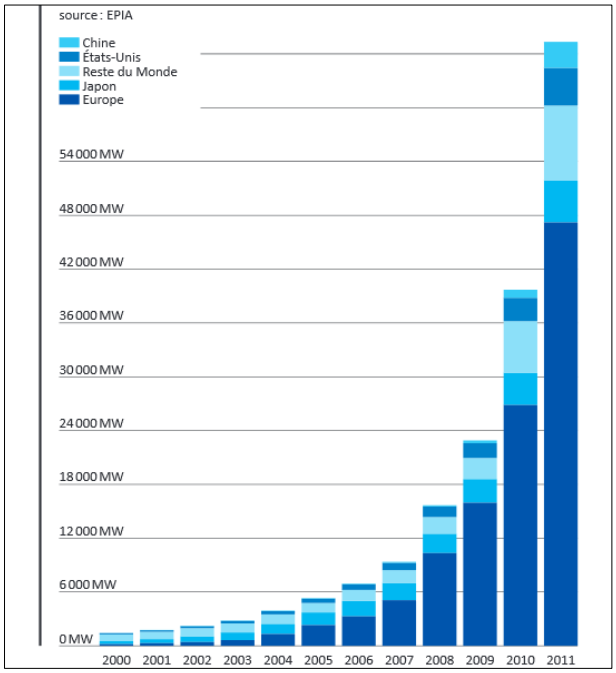


Figure 1: Développement de la capacité photovoltaïque mondiale (MW). Source : EPIA, 2012

Maintenir l'équilibre du réseau implique qu'à chaque instant la production électrique soit égale à la consommation électrique. Or l'énergie photovoltaïque, tout comme l'énergie éolienne, est une énergie dite intermittente. En effet elle dépend fortement de la météo. Afin de pouvoir utiliser les énergies renouvelables il faut donc trouver des solutions à l'intermittence. Le développement des technologies de stockage, le contrôle de la consommation, le développement du réseau électrique sont ainsi envisagés [OPECST, 2011]. De même la prévision de la production photovoltaïque aura un rôle primordial dans le futur.

Reuniwatt, l'entreprise dans laquelle j'ai effectué ce stage, s'intéresse particulièrement à cette prévision. Leur but étant de proposer un outil d'aide à la décision. Ainsi, ils orientent leur travail sur le fait d'atteindre l'état de l'art dans la prévision photovoltaïque, voire de le dépasser. C'est dans ce contexte que j'ai été amenée à travailler sur plusieurs modèles de prévision.

Ce rapport détaille donc un peu plus le contexte dans un premier temps, il présente ensuite les matériels et méthodes utilisés, puis les résultats pour finir par une discussion sur les améliorations possibles.

I. Contexte général de l'étude

A. La prévision photovoltaïque et état de l'art

Une étape primordiale pour prévoir la quantité d'électricité produite par un panneau solaire est la prévision de la quantité d'énergie reçue par le panneau. Partant du principe que l'inclinaison des panneaux est faible, nous parlons de prévision de GHI (Global Horizontal Irradiance) qui est l'irradiance totale reçue par une surface horizontale au sol.

La méthodologie employée pour prédire l'irradiance dépend de la durée de prévision souhaitée [DIAGNE et al. PELLAND et al.]. En effet, il est possible d'utiliser :

- Des modèles statistiques en considérant les mesures de GHI comme des séries temporelles. Il s'agit par exemple de modèles autorégressifs (AR), de modèles autorégressifs à moyennes mobiles (ARMA) ou encore de réseaux de neurones artificiels (ANNs). Ces approches sont généralement utilisées pour des horizons de prévision assez court, de quelques minutes à 6 heures.
- L'observation des nuages et de leur déplacement pour prévoir la quantité d'énergie reçue sur Terre. Celle-ci peut se faire par l'analyse d'images prises par des caméras au sol (prévision jusqu'à 30 minutes), ou bien par imagerie satellite (prévision jusqu'à 5 heures).
- Les modèles météorologiques proposent une prévision de 6 heures à une journée. Ils sont basés sur des équations dynamiques qui prévoient l'évolution de l'atmosphère à partir des conditions initiales.
- Des combinaisons de ces différentes méthodes afin de tirer profit de leur complémentarité. Ces méthodes ont fait leurs preuves dans de nombreux domaines depuis plus de 30 ans, elles améliorent la précision de la prévision. [De MENEZES et al., 2000]. Elles sont notamment utilisées en éolien, en économétrie ou en finance mais assez peu en prévision photovoltaïque. La combinaison entre des prévisions météorologiques et des données satellites proposée par Lorenz et al. (2012) est une des rares applications dans le domaine de l'énergie solaire.

B. Reuniwatt et objectifs du stage

1) Reuniwatt : objectifs et activités de l'entreprise

Reuniwatt est une jeune start-up d'une dizaine de salariés fondée par Nicolas Schmutz en 2009 spécialisée dans l'énergie photovoltaïque. Elle est implantée à l'île de La Réunion, la région la plus ensoleillée de France. La Réunion est la 10ème région de France en matière de puissance photovoltaïque raccordée au réseau [CGDD, 2014] et 35% de sa production électrique provient des énergies renouvelables [ARER, 2012]. L'activité de Reuniwatt s'articule autour de trois métiers :

- L'expertise en mix énergétique : accompagnement et assistance à la maîtrise d'ouvrage, brevet d'installations photovoltaïques particulièrement adaptées aux zones cycloniques.
- Les systèmes d'informations climatiques : acquisition de données climatiques du territoire à travers des réseaux de capteurs.

- La prévision de la production photovoltaïque avec l'outil d'aide à la décision Soleka. Ce dernier constitue la solution proposée par Reuniwatt au problème de l'intermittence de l'énergie solaire.

2) Objectifs du stage

Mon stage s'effectue dans le cadre du développement de Soleka. Soleka est destiné à faciliter l'insertion d'énergies intermittentes dans le mix énergétique. Cet outil permettra, à terme, de prévoir sur différents horizons temporels (30 minutes, 6 heures ou une journée) le rayonnement solaire qui va être reçu et donc, la quantité d'énergie photovoltaïque produite.

J'ai travaillé sur le volet de la prévision à J+1, c'est-à-dire la prévision pour le lendemain. N'ayant que très peu eu affaire à l'étude des séries temporelles, j'ai d'abord appréhendé les problématiques de ce type de données avec le modèle ARMA, modèle de référence pour l'analyse des séries temporelles. Puis, l'objectif du stage étant de travailler sur la combinaison de modèles statistiques pour la prévision photovoltaïque, j'ai travaillé sur un modèle de prévision à J+1, la classification en journées types, et sur sa combinaison avec d'autres modèles de prévision basiques à J+1, modèle de persistance et modèle de climatologie.

II. Matériels

A. Les données

Les données dont nous disposons proviennent de différents sites en Corse. Nous en répertorions cinq : Corte, Letia, Lucciana, Moltifao et Vescovato. Ce rapport présente essentiellement les résultats obtenus sur le jeu de données Corte.

Les données sont constituées d'une mesure, celle du GHI, exprimée en $W.m^{-2}$. Cette variable prend des valeurs différentes le long d'une journée mais aussi selon la saison. En été, le GHI en Corse peut dépasser les 1000 $W.m^{-2}$, alors qu'il dépasse rarement le 500$W.m^{-2}$ en hiver.

Elles sont accompagnées de la date et l'heure de la mesure. Le jeu de données est complété par deux autres variables : le GHI issu du modèle de ciel clair et l'indice de ciel clair (KC).

- Le modèle de ciel clair est un modèle qui donne une estimation du GHI dans l'hypothèse d'un ciel sans nuages. Nous utilisons les données issues du modèle McClear. Elles sont gratuites et accessibles en ligne [LEFEVRE et al., 2013]**.** Cette variable sert au calcul du KC présenté ci-dessous.

- L'indice de ciel clair se définit de la façon suivante :

  KC= GHI mesuré / GHI modèle ciel clair

  Le calcul du KC sert à normaliser la série pour enlever la composante journalière et saisonnière. Typiquement, cela sert à s'affranchir de la différence d'amplitude du GHI entre été et hiver. Le KC est généralement compris entre 0 et 1, cependant il arrive qu'il dépasse 1. Ce cas se présente quand l'irradiance reçue est supérieure à celle qu'il y aurait eu avec un ciel sans nuages, en effet ceci est dû à la réflexion des nuages.

La Figure 2 illustre ces différentes variables pour une journée d'hiver et une journée d'été.

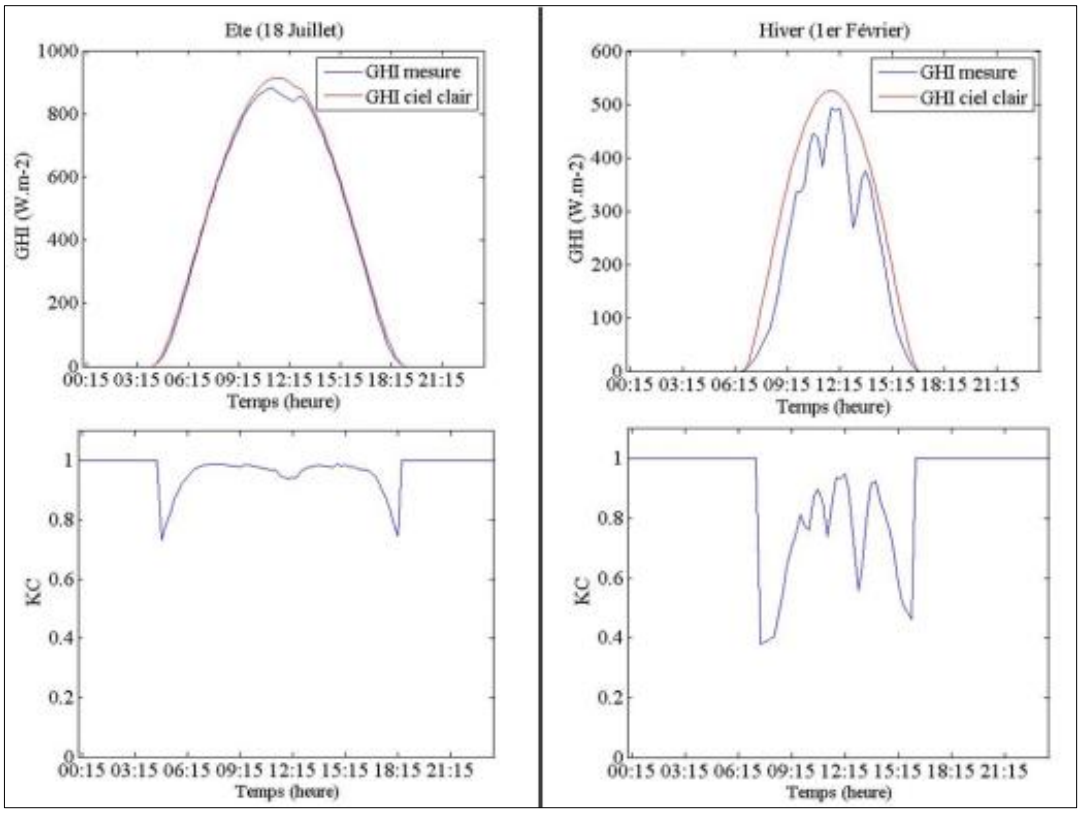


Figure 2: Illustration des séries GHI mesuré, GHI par ciel clair et de KC sur une journée d'hiver (1er Février) et d'été (18 Juillet)

Nous disposons de ces données entre le 1er janvier 2004 et le 31 décembre 2012, avec une résolution temporelle de 15 minutes (chaque donnée correspond à une moyenne sur 15 minutes). Le jeu de données est découpé de la façon suivante :

- Données d'apprentissage : Du 1er janvier 2004 au 31 décembre 2010
- Données d'initialisation : Du 1er janvier 2012 au 31 janvier 2012
- Données de validation : Du 1er février 2012 au 31 décembre 2013
- Données de réserve : Du 1er janvier 2011 au 31 décembre 2011

B. Les logiciels utilisés

Les employés de Reuniwatt travaillent beaucoup avec Matlab R2013, de ce fait il était plus facile que j'utilise aussi ce logiciel. Ceci permet notamment de s'affranchir des problèmes de compatibilité de format, et de pouvoir s'échanger les scripts facilement. Au sein de l'entreprise, le principe des toolbox certifiées est aussi apprécié.

Cependant, R commence aussi à être utilisé, en particulier par ceux qui travaillent sur les modèles statistiques pour la prévision. J'ai pu donc continuer à m'exercer sur ce logiciel, surtout en ce qui concerne ma découverte du modèle ARMA.

Dans chacun des cas, un soin a été apporté pour rendre mes scripts et mes fonctions utilisables par les autres employés et les clients.

III. Méthodes

A. Un modèle de prévision utilisant des données récentes : le modèle ARMA

Les modèles ARMA sont souvent utilisés pour traiter les séries temporelles.

Ces modèles reposent sur le fait que la valeur suivante dépend uniquement des valeurs précédentes et d'un facteur correctif qui prend en compte les erreurs commises précédemment. Un modèle ARMA d'ordres (p,q) vérifie :

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \phi i \, X_{t-i} + \sum_{i=1}^{q} \theta i \, \varepsilon_{t-i}$$

Où Φi et θi sont les paramètres à estimer.

Pour ajuster un modèle ARMA, il est préférable que la série soit stationnaire. Dans le cas contraire, des transformations peuvent être apportées, comme le passage au logarithme ou encore le recours à la différenciation de la série.

En prévision photovoltaïque, nous ne cherchons pas seulement à prévoir la valeur suivante du GHI mais aussi celles à des horizons temporels plus élevés pour faire de la prévision à plus long terme. Mais du fait de la construction du modèle, la prévision accumule de plus en plus d'erreurs avec le temps. En effet, au-delà de quelques pas de temps, la prévision dépend des prévisions faites précédemment. De ce fait, l'horizon de prévision du modèle ARMA de base est assez limité, de l'ordre de quelques pas de temps. Ainsi lorsque les données sont échantillonnées à un pas de temps horaire, la prévision peut être faite pour la journée. Lorsque les données sont échantillonnées toutes les 15 minutes, le modèle n'est pas adéquat pour des horizons de prévision dépassant l'ordre de 2 à 3 heures.

Un autre problème se pose avec les données photovoltaïques, celui du cas des données de nuit. Pendant la nuit, le GHI est de 0, on ne peut pas appliquer simplement le modèle ARMA pour prédire les données du début du jour. En effet, les fonctions implémentées dans les logiciels Matlab ou R ne prennent pas en charge les données de nuit. Un des moyens de s'en affranchir est de les enlever des données au moment de l'utilisation de la fonction.

Un modèle ARMA a été implémenté, mais les résultats de l'application ne sont pas intégrés dans ce rapport.

B. Prévision par un modèle de classification en journées types

1) Principe général de la méthode

Ce type de prévision est basée sur une classification en journées types. Les étapes sont les suivantes :

- Classification en k journées types des jours du jeu de données d'apprentissage à partir d'un critère.

- Pour chaque classe, caractérisation de chaque journée type par une série journalière de KC, établie à partir de la moyenne des séries de KC de l'ensemble des jours de cette classe du jeu de données d'apprentissage.
- Etablissement d'une matrice de transition d'un type de journée à l'autre.
- Prévision de la série journalière de KC du jour J+1 en fonction de la classe de la journée J, de la matrice de transition et de la matrice des séries de KC caractéristiques des classes. La Figure 3 présente cette étape.

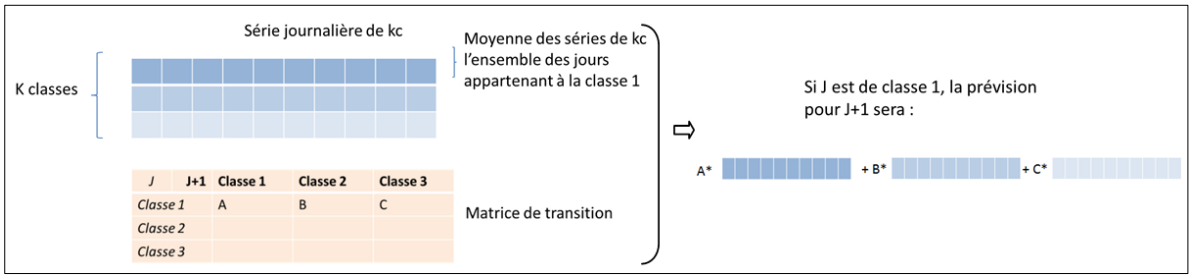


Figure 3: Description de la méthode de prévision par classification

Ainsi, les critères sur lesquels nous pouvons intervenir sont :

- Choix des critères servant à l'élaboration des classes
- Choix du nombre de classes
- Choix de l'algorithme de classification
- La distance utilisée pour l'affectation dans les classes
- Choix de la méthode de calcul de la matrice de transition

2) Point de départ

Soubdhan et al.(2009) proposent une méthode de classification utilisant une méthode de mélange de distribution de Dirichlet. Pour ce faire, ils résument les variations journalières d'irradiance solaire par des histogrammes. Cette idée a été testée par Reuniwatt.

Une des grandes problématiques de la classification de ce type de données est le problème de de la différence de valeur de GHI mesuré/ciel clair le long de l'année et de la différence de durée du jour. Ces différences sont visibles dans la Figure 4.

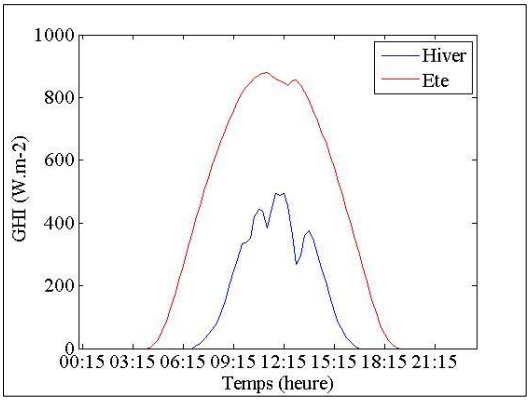


Figure 4: Série de GHI mesuré pour une journée d'hiver (1er Février) et une journée d'été (18 Juillet)

La méthode de classification proposée par Reuniwatt prend en compte ces deux difficultés. Tout d'abord, nous travaillons avec le KC. Le fait de travailler avec cette variable normalise les données pour ne plus avoir de différence entre l'été et l'hiver en termes d'amplitude de GHI. En effet, un GHI élevé en été est reste plus fort qu'un GHI élevé en hiver.
Ensuite, le critère pour classer les journées est l'histogramme de répartition du KC le long de la journée. Travailler avec des histogrammes permet de s'affranchir du problème de la différence de la durée du jour. La Figure 5 présente la construction des histogrammes de KC.

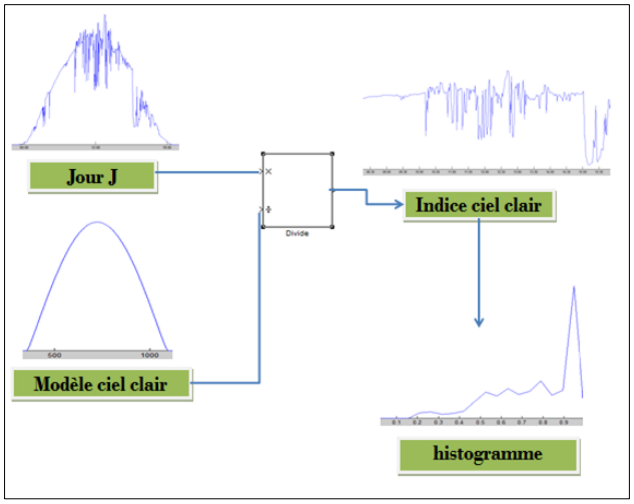


Figure 5: Construction de la variable sur laquelle se base la classification. Source : LIANDRAT, 2012.

Une matrice de transition est calculée pour connaitre la probabilité de passer d'une classe à une autre. Elle est calculée de façon empirique en comptant les transitions entre les différentes classes d'un jour à l'autre du jeu de données d'apprentissage.

La distance au centroide est une distance euclidienne, calculée de la façon suivante :

$$D = \sqrt{\sum_{i} (Cci - Ci)^2}$$

Dans cette méthode de classification, la distance calculée est celle entre deux histogrammes de KC. Cc faisant référence à l'histogramme de KC du centroide de la classe, C faisant référence à l'histogramme de KC de la journée à classer.

L'algorithme utilisé est celui du k-means (il est présenté en Annexe I). Afin de compenser la sensibilité aux choix des classes initiales du k-means, nous avons répété le processus cinq fois. La solution gardée est celle qui minimise la somme des variances intra-classe.

Le choix du nombre de classe ne suit pas de règles bien précises. Celui-ci se fait en prenant appui sur les valeurs de variabilité intra-classe et inter-classe tout en gardant du recul. Cette démarche sera illustrée dans la partie IV (résultats) du rapport de stage.

En ce qui concerne la prévision, il est difficile de s'affranchir de la différence de durée du jour dans ce cas. Les séries de KC caractéristiques utilisées pour la prévision sont la moyenne des séries de KC de l'ensemble des jours appartenant à la classe k. Mais ceci entraine un biais assez important, puisqu'au moment de la prévision nous allons prédire des journées trop longues en hiver et des journées d'été avec des valeurs de début et de fin de jour bien inférieures à la normale.

3) Recherche d'amélioration

Le fait de raisonner en termes d'histogrammes présente une limite assez importante : cela ne conserve pas la structure temporelle des données. En effet, une journée où il fera beau le matin et nuageux l'après-midi et une journée où il fera nuageux le matin et beau l'après-midi peuvent présenter le même histogramme, et donc se retrouver dans la même classe. Or la caractérisation des classes se faisant en moyennant les séries de KC, une perte d'information importante peut être observée.

L'idée d'amélioration est donc de conserver la dimension temporelle dans la classification. Il reste important de travailler sur le KC pour les mêmes raisons que celles présentées précédemment. Nous ne pouvons pas comparer les séries de KC telles quelles en raison du problème de la durée du jour qui est illustré en Figure 6. Nous travaillons avec une distance euclidienne point par point, si nous ne transformons pas les séries de KC nous allons comparer des points qui ne sont pas comparables : des données de nuits avec des données de milieu de matinée par exemple.

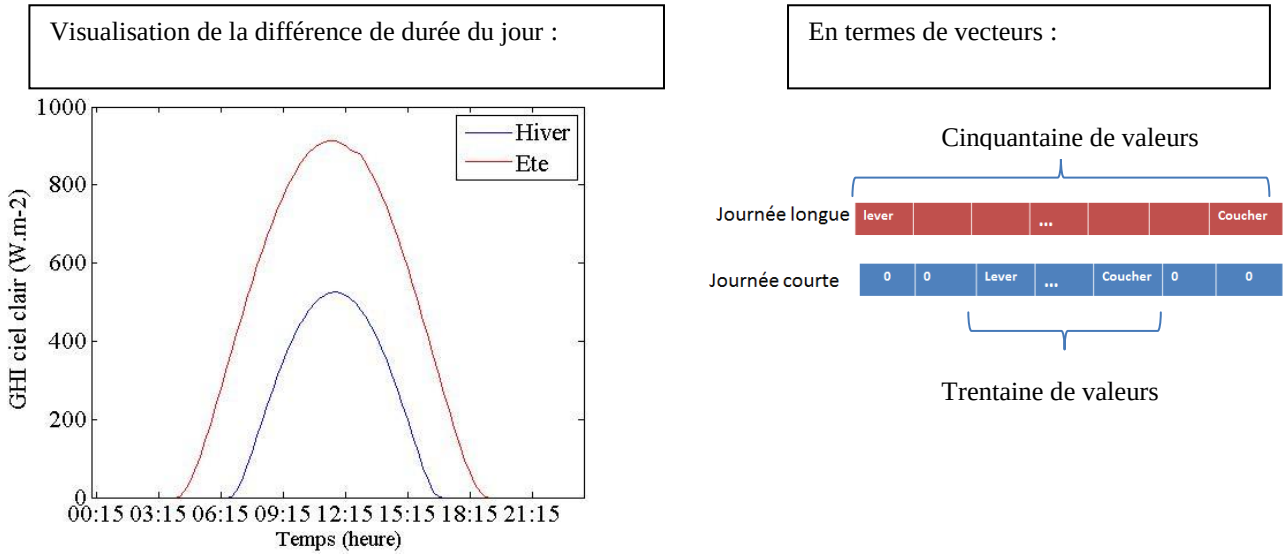


Figure 6: Schématisation du problème de la différence de durée du jour

L'idée est de rééchantillonner toutes les journées selon une journée de taille unique. La première composante du vecteur contient la valeur de KC au lever du jour, la dernière celle au coucher du soleil. Entre les deux, chaque composante i de l'ensemble des vecteurs concerne le « même moment de la journée ».

Nous choisissons de calibrer toutes les journées selon un vecteur de taille 21.

Pour l'ensemble des journées, le nouveau vecteur temps est calculé de la façon suivante :

$$\Delta t = \text{Heure de coucher du Soleil} - \text{Heure du lever du Soleil} / 20$$

Le Tableau 1 explicite l'élaboration de ce nouveau vecteur temps.

Tableau 1: Etalonnage des données selon un vecteur de taille unique

Vecteur temps initiale	Lever du Soleil=Ls	Ls+15min	Ls+30min	…	Coucher du Soleil
Nouveau vecteur temps	Lever du Soleil=Ls	Ls+ Δt	Ls+ 2Δt	…	Coucher du Soleil

Le nouveau vecteur de KC, correspondant à la série de KC du nouveau vecteur temps est obtenu par interpolation linéaire.

Au préalable de ce traitement, les données sont lissées avec une moyenne mobile d'ordre 3 sans pondération. L'ordre 3 est choisi visuellement. Nous cherchons à lisser les données pour ne pas accorder trop d'importance aux fortes fluctuations brusques, correspondant par exemple au passage rapide d'un nuage, mais nous souhaitons quand même garder de la variabilité.

Pour affecter les individus aux classes, nous utilisons toujours la distance euclidienne :

$$D = \sqrt{\sum_i (Cci - Ci)^2}$$

Cette fois ci, Cc fait référence au vecteur de KC calibré du centroide, C fait référence au vecteur de KC calibré de la journée à classer.

La taille du vecteur étalonné est choisie de façon à ne pas être trop court, synonyme de perte d'information, mais pas trop long pour ne pas accorder trop d'importance aux petits décalages dans le temps. En effet, nous utilisons la distance euclidienne pour affecter les individus dans les classes, celle-ci se calcule donc point par point. Lorsque nous travaillons sur les histogrammes, nous travaillons avec un vecteur de taille 25, d'où la volonté de rester dans cet ordre de grandeur.

Cette nouvelle méthode, développée pendant ce stage, permet également d'améliorer l'étape de prévision qui souffrait de la différence de durée du jour dans la méthode précédente. En effet, nous pouvons aussi travailler avec les séries de KC calibrées. Pour chaque jour, nous prévoyons sa journée calibrée. Les heures du début et de fin du jour peuvent être connues à l'avance. Pour prédire la série de KC de la journée il suffit de décomposer le vecteur de KC calibré prédit afin de lui redonner la taille initiale.

C. La combinaison de modèles

La combinaison de modèles est une piste très intéressante à explorer pour l'amélioration de la précision de prévision. Il existe plusieurs techniques permettant de faire de la combinaison [De MENEZES et al., 2000]. L'étude suivante se présente comme une étude préliminaire du comportement de la combinaison de modèles.

L'étude de la combinaison de modèles s'est faite en binôme.

1) Les modèles utilisés pour étudier la combinaison de modèles

Pour étudier la combinaison de modèles nous avons travaillé avec trois modèles différents.

- La persistance : ce modèle consiste à partir du principe qu'il fera le même temps que la veille lorsque l'on travaille sur l'horizon de prévision J+1. Il s'agit d'une prévision naïve. Elle constitue souvent le modèle de référence à battre.

- Un modèle de climatologie que nous avons appelé « climatologie 7 jours ». Ici, la prévision du KC est égale à la moyenne des sept derniers jours. Dans notre cas, nous avons des données tous les quarts d'heure, donc la prévision pour 8h15 sera la moyenne des KC obtenus les sept derniers jours à 8h15.

- La classification développée dans la partie B.2. Nous avons utilisé la prévision obtenue avec une classification en quatre classes.

Pour chacun de ces modèles, une prévision est faite avec le jeu de données Corte. Chaque modèle offre donc une prévision correspondant à la prévision des « données de validation ». Ces prévisions sont ensuite combinées deux à deux.

2) Les différents types de combinaisons testées

Pour appréhender la combinaison de modèles statistiques pour la prévision photovoltaïque, nous avons testé trois types de combinaisons :

- La moyenne : la prévision est la moyenne des deux prévisions proposées.
- La régression linéaire : chaque prévision constitue la série de valeurs prises par une variable. Dans notre cas, l'équation du modèle est la suivante :

$$Y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

Y_i étant la prévision au temps i. X_{i1} la prévision au temps i pour le modèle 1. β_o, β_1 et β_2 les paramètres à estimer et ε_i les erreurs qui sont supposées suivre une loi normale.

- Une méthode d'optimisation de l'erreur de prévision sous contrainte afin que la somme des poids affectés à chaque modèle de la combinaison soit égale à un. L'algorithme utilisé pour faire cette optimisation a été développé par mon binôme et est présenté en Annexe II.

Ces techniques de combinaisons ont été testées sur deux modèles mais ils sont généralisables à plus de modèles.

Pour comparer les modèles de combinaison, nous utiliserons la RMSE (Root Mean Square Error ou Erreur Quadratique Moyenne en français), ainsi qu'une RMSE normalisée. Cette NRMSE se calcule de la façon suivante :

$$NRMSE = RMSE / \text{moyenne des observations}$$

Ceci n'est pas la définition théorique de la NRMSE qui elle, divise la RMSE par l'écart entre la valeur maximale et minimale des observations. Néanmoins il s'agit de celle utilisée par Reuniwatt. Cette méthode de calcul étant utilisée par leurs clients pour comparer les résultats à ceux des concurrents, par soucis de cohérence nous l'avons également utilisée. Aussi, cette méthode se justifie par le fait qu'en prévision photovoltaïque, la valeur minimale est 0. De ce fait cela reviendrait à normaliser par la valeur maximale, or la moyenne reflète mieux la distribution que la seule valeur du maximum.

IV. Résultats

A. Comparaison des classifications

Les résultats présentés sont ceux issus du jeu de données Corte. La comparaison des performances des modèles se fait sur la prévision du GHI. Les modèles utilisés sont basés sur la prévision du KC, mais les données sont transformées en données de GHI avec le modèle de ciel clair que nous prenons pour acquis.

Résultats avec la méthode 1 (Méthode de Reuniwatt) :

Dans un premier temps, nous traçons le graphique, en Figure 7, de l'inertie intra et inertie inter en fonction du nombre de classes.

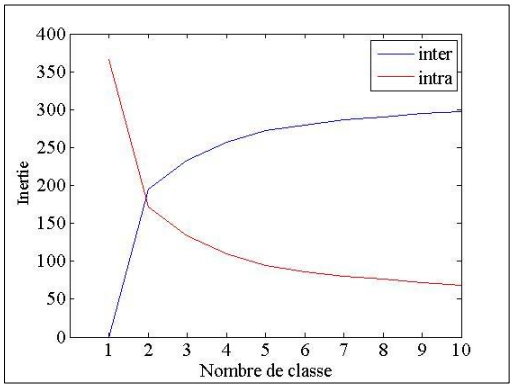


Figure 7: Inertie inter et intra pour la méthode des histogrammes en fonction du nombre de classes

Au vue de notre problème, le choix de deux classes ne semble pas pertinent, car il apparaît comme réducteur de limiter la complexité de la météo à deux journées types. L'utilisation de ce graphe pour choisir le nombre de classe étant controversé, nous choisissons d'effectuer la classification pour deux, trois et quatre classes. Nous décidons ensuite du nombre de classes à choisir en fonction de la RMSE. Les résultats sont présentés en Figure 8.

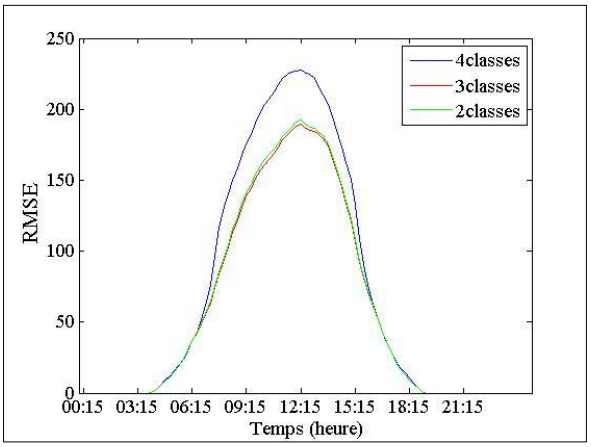


Figure 8: Comparaison de la RMSE en fonction du nombre de classes (classification par histogramme)

Les résultats de RMSE pour quatre classes étant nettement moins bon que pour trois classes, nous choisissons de travailler avec trois classes pour ce type de classification.

Afin de mieux comprendre à quoi correspondent les classes, nous regardons les histogrammes des centroides des classes et l'allure de la courbe de GHI d'une journée de chaque classe. Ils sont présentés dans la Figure 9.

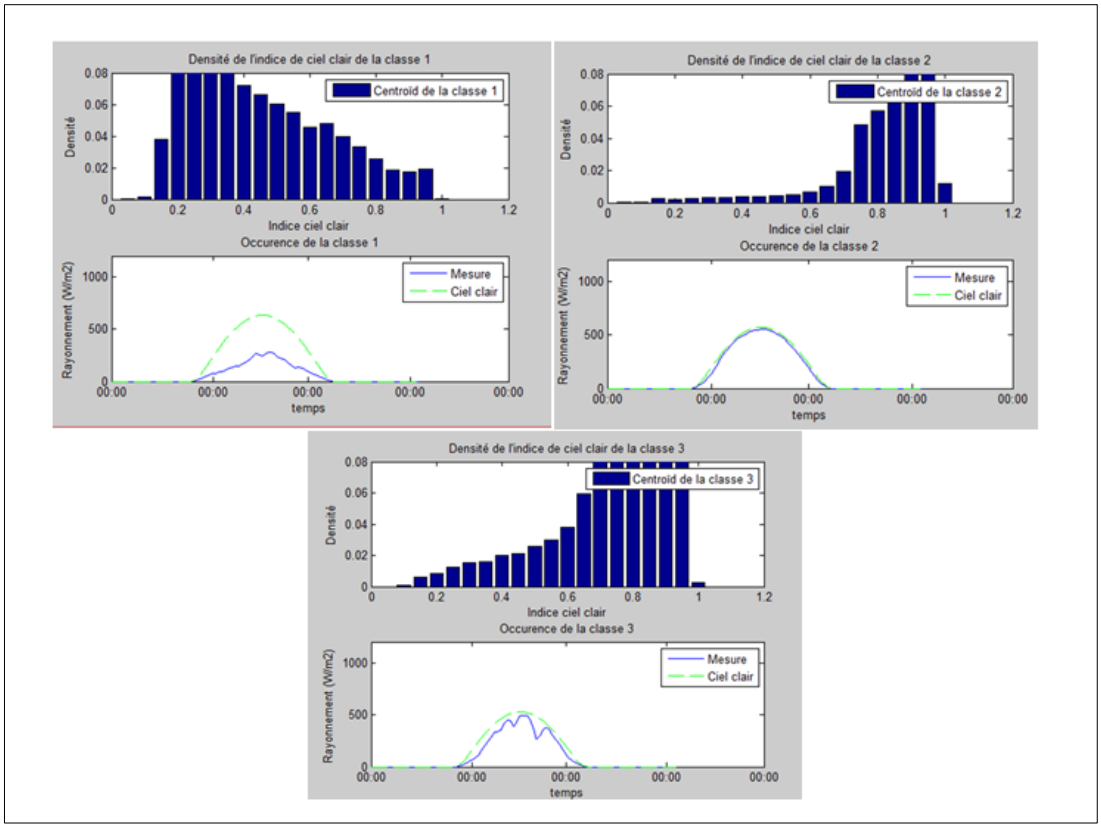


Figure 9: Description des classes dans le cas de classification par histogramme de KC. Source : Liandrat, 2012.

Il apparait que la classe 1 correspond à une journée de mauvais temps, la classe 2 à une journée de beau temps et la classe 3 à une journée moyenne.

Résultats avec la méthode 2 (Nouvelle méthode développée lors de ce stage) :

Le graphe de l'inertie intra et de l'inertie inter nous donne sensiblement la même chose que précédemment. Il est tout de même présenté en Annexe III. L'idée initiale était d'améliorer la précision et de pouvoir analyser la classe comprenant les journées « moyennes ». Nous effectuons donc la classification pour trois et quatre classes et nous les comparons au moyen de la RMSE. Les résultats de la RMSE globale pour la prévision du GHI sont présentés dans le Tableau 2. La Figure 10 présente les résultats de la RMSE le long de la journée.

Tableau 2 : Résultat de RMSE globale pour la classification en 3 ou 4 classes

RMSE globale classification en 3 classes	RMSE globale classification en 4 classes
93,6681	93,4026

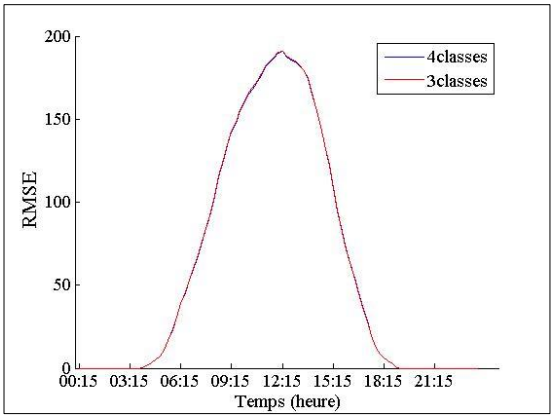


Figure 10: Comparaison de la RMSE en fonction du nombre de classe (classification par série de KC)

Nous remarquons ici que les résultats de la RMSE sont très proches pour les deux classifications contrairement à la méthode précédente. Il est intéressant de se concentrer davantage sur la caractérisation des classes. Ici nous pouvons le faire en traçant la série de KC calibrée du centroide de chaque classe. Nous le faisons pour les deux méthodes de classification, les graphiques sont présentés dans la Figure 11.

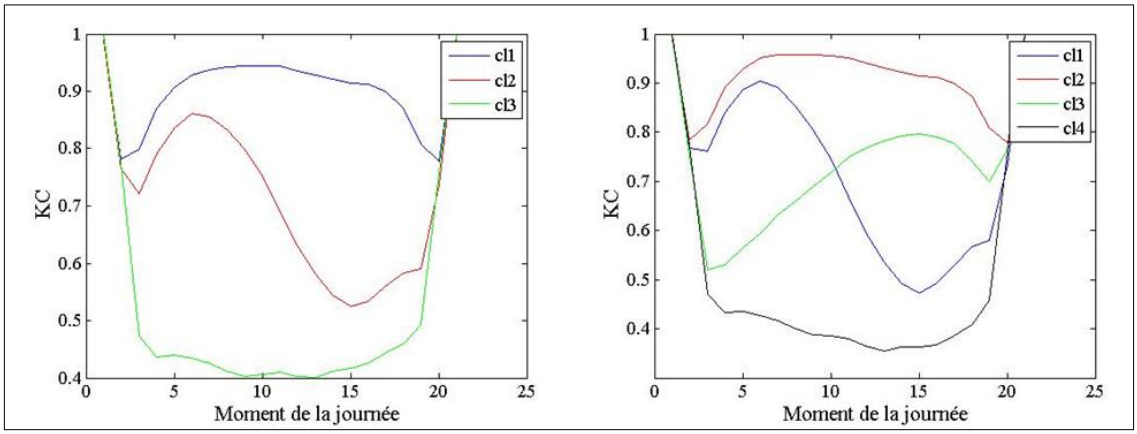


Figure 11: Description des classes dans le cas de la classification par série de KC

Avec la classification en quatre classes nous voyons apparaitre une distinction entre les jours où il fait beau uniquement le matin (cl2 à gauche, cl1 à droite) et ceux où le beau temps est progressif (cl3 à droite).

Pour comparer de plus près les deux méthodes, nous pouvons comparer les classements des journées dans les deux cas. Les résultats sont présentés dans le Tableau 3.

Tableau 3: Tableau des correspondances entre les classifications en 3 ou 4 classes

	Classification en 3 classes	Belle journée	Mauvaise journée	Belle le matin
Classification en 4 classes				
Belle journée		1444	0	34
Mauvaise		0	368	0
Belle le matin		0	18	357

Beau progressif	73	34	187

Ceci nous permet de donner davantage de crédit à la classification en quatre classes. En effet, celle-ci donne des résultats équivalents à la classification en trois classes lorsqu'il s'agit de prédire une belle journée, une mauvaise journée et une journée où il ne fait beau que le matin. La quatrième classe est quant à elle plus hétérogène.

Or les résultats de la RMSE sont presque équivalents pour les deux méthodes. Il est donc peut être préférable d'utiliser une classification en quatre classes car ceci peut nous donner une information supplémentaire. En effet, il est légitime de penser que dans le cas d'une prédiction de journées « belles », « mauvaises », et « belles le matin », la prévision du type de journée est plus précise qu'avec une classification en trois classes. Le classement des journées dans ces catégories est moins du au hasard qu'avec la classification en trois classes.

Pour vérifier cela, nous calculons l'écart type à la valeur de référence de chaque classe, qui est ici la valeur du centroide. Pour chaque classe, nous calculons l'écart type à la valeur de référence pour chaque point du vecteur « série de KC calibrée ». Puis nous faisons la moyenne. Les données sont rassemblées dans le Tableau 4.

Tableau 4 : Tableau des écart types à la valeur de référence

	3 classes	**4 classes**
Classe « beau »	0,0917	0,0786
Classe « mauvais »	0,1563	0,1367
Classe « beau matin »	0,1648	0,1562
Classe « beau progressif »		0,1487

Ceci confirme notre idée qu'en classant nos journées en quatre journées types nous gagnons en certitude pour prévoir les profils des journées « belles », « mauvaise » et « beau le matin et moins beau l'après-midi ».

Comparaison entre la méthode 1 et la méthode 2 :

Les résultats de la RMSE globale pour la prévision du GHI sont présentés dans le Tableau 5.

Tableau 5 : Récapitulatif des RMSE globale pour la méthode 1 et la méthode 2

	3 classes	4 classes
Méthode 1 : histogramme de KC	92.23	
Méthode 2 : série de KC calibrée	93,421	93,1563

Nous remarquons que les résultats de la RMSE sont assez proches bien que la classification par histogramme reste meilleure dans notre cas.

La Figure 12 présente les résultats de la RMSE le long de la journée pour les meilleurs résultats de chaque méthode : classification en trois classes dans le cas où celle-ci se fait sur

les histogrammes de KC et classification en quatre classes dans le cas où celle-ci se fait sur la série de KC calibrée.

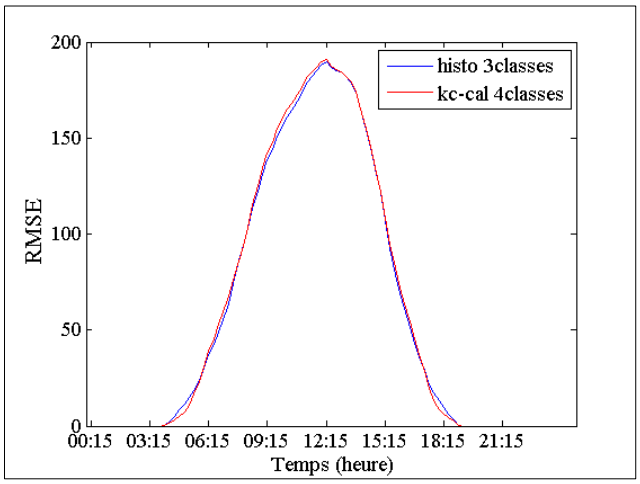


Figure 12 : Comparaison de la RMSE le long de la journée pour les 2 méthodes de classification

Nous pouvons remarquer que la classification par la série de KC calibrée présente de meilleurs résultats en début et fin de journée. Ceci s'explique facilement par le fait qu'avec cette méthode, nous maîtrisons mieux la différence de durée du jour au moment de la prévision.

Dans notre cas, la classification avec la série de KC calibrée ne semble pas plus précise que la classification avec les histogrammes de KC. Nous avons vu que lorsque nous opérions à ces deux classifications en trois classes, nous obtenions les mêmes types de journées : « belle journée », « mauvaise journée » et « journée moyenne ». Par ailleurs la classification en trois ou quatre classes pour la méthode utilisant la série de KC calibrée donne les mêmes résultats en termes de RMSE. De ce fait, il semble cohérent que les résultats pour la classification en trois classes par histogrammes de KC et la classification en quatre classes (ou trois) par série de KC calibrée soient sensiblement les mêmes.

La classification par série de KC calibrée est à privilégier, car comme nous l'avons vu elle nous renseigne mieux sur le déroulement de la journée pour les journées « moyennes ».

Aussi ces deux méthodes de classification ont été comparées sur d'autres sites. Les résultats ont confirmé l'intérêt de la classification par série de KC calibrée. En effet les résultats de la RMSE étaient satisfaisants par rapport aux ceux obtenus avec la classification par histogrammes de KC. Les résultats de la RMSE sur le site de Lucciana sont présents en Annexe IV.

B. Comparaison avec les autres méthodes de prévisions

Nous avons amélioré la méthode de classification statique actuelle, il s'agit maintenant de la comparer à d'autres méthodes. Les autres modèles qui nous servent de comparaison sont : la persistance, la « climatologie 7 jours » ainsi qu'une classification dynamique.

Cette dernière repose sur les histogrammes de KC et est réalisée avec trois classes. La prévision se fait à partir de la moyenne des 20 derniers jours appartenant à la classe, et non pas à l'ensemble des jours de la classe, d'où la notion de « dynamique ».

Les résultats de la RMSE globale pour la prévision du GHI sont présentés dans le Tableau 6. La Figure 13 présente les résultats de la RMSE le long de la journée.

Tableau 6: Tableau récapitulatif des RMSE et NRMSE des différentes méthodes de prévision

	RMSE	NRMSE
Persistance	118,5711	0,631
Climatologie 7 jours	89,6387	0,477
Classification statique proposée	93,1563	0,496
Classification dynamique	93,046	0,4954

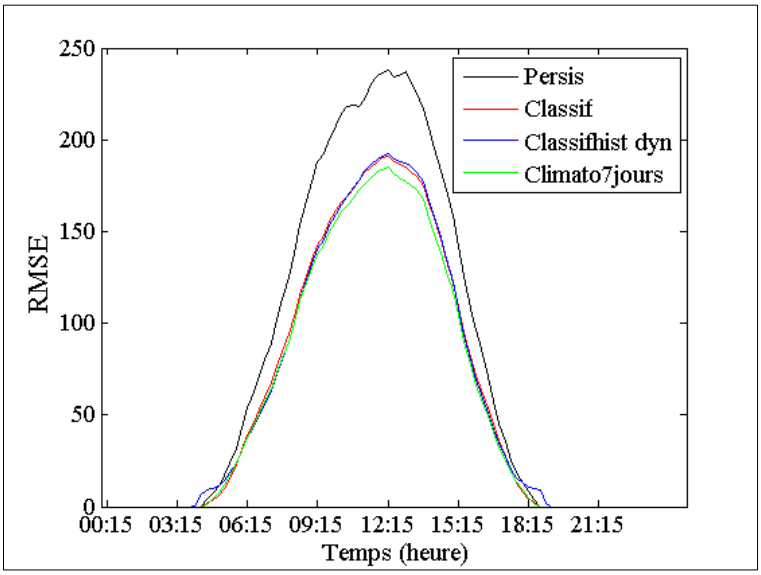


Figure 13: Comparaison des RMSE le long de la journée

Ainsi nous pouvons remarquer que la classification statique présentée précédemment est nettement meilleure que le modèle de persistance. Elle se rapproche très fortement de la méthode la plus développée actuellement à Reuniwatt, à savoir la classification dynamique à partir des histogrammes de KC. La bonne performance de la classification dynamique à partir des histogrammes de KC par rapport à la classification statique qui utilise le même critère s'explique par le fait que dans le cas de la dynamique nous nous affranchissons un peu du problème de la différence de durée du jour. En effet, la prévision se base sur une classification en trois classes et prend en compte les vingt derniers jours appartenant aux classes, ce qui correspond en moyenne aux deux derniers mois seulement.

Cependant, nous remarquons que le modèle de climatologie qui consiste à faire simplement la moyenne des sept derniers jours reste le meilleur. Ce modèle s'avère en effet très difficile à battre. Ceci pourrait être expliqué par le fait qu'il existe des dynamiques atmosphériques différentes en été et en hiver, or la classification ne différencie pas les journées d'été et

d'hiver. Il pourrait constituer le modèle de référence pour Reuniwatt et remplacer la persistance.

C. Apport de la combinaison de modèles

Tout d'abord, nous pouvons nous intéresser aux performances globales des combinaisons proposées. Le Tableau 7 présente les résultats de RMSE des combinaisons des différents modèles.

Tableau 7 : Résultats de RMSE des combinaisons des différents modèles

	Moyenne	Régression linéaire	Optimisation
Persistance/classification	98,182	92,363	92,698
Persistance/Climatologie	95,88	89,304	89,273
Classification/climatologie	87,96	87,607	87,72

Le couple « classification-climatologie » semble être le plus performant. Les résultats de la régression linéaire pour ce couple montrent d'ailleurs que ce couple est le meilleur modèle que nous ayons obtenu.

Nous remarquons aussi que les résultats de la régression linéaire et de l'optimisation sont assez proches. Aussi, faire la moyenne entre la prévision par classification et la prévision par climatologie semble apporter de très bons résultats.

Pour mieux comprendre ces résultats, nous pouvons regarder les coefficients de la régression ainsi que les poids optimaux calculés. Ils sont présentés dans le Tableau 8.

Tableau 8: Récapitulatif des coefficients et des poids obtenus pour les différentes combinaisons

Regression linéaire	Classif / persistance	Classif / climato	Persistance / climato
Classif	β1= 0,95	β1= 0,4	
Persistance	β2= 0,095		β1= 0,09
Climato		β2= 0,6	β2= 0,9
βo	-2,0858	-2,01	1,07
Optimisation	Classif / persistance	Classif / climato	Persistance / climato
Classif	0,889	0,37	
Persistance	0,111		0,094
Climato		0,63	0,906

Ainsi, nous pouvons remarquer que les coefficients de la régression et les poids issus de l'optimisation sont souvent très proches. Il n'est donc pas étonnant que ces deux types de combinaisons donnent des résultats proches en termes de prévision. Le tableau ci-dessus explique également la bonne performance de la moyenne concernant le couple

« classification-climatologie ». En effet l'optimisation et la régression apposent des poids/coefficients proches de 50/50 (0,6 et 0,4).

Si l'on compare les résultats de la RMSE globale pour la combinaison avec ceux obtenus individuellement dans le Tableau 6, nous remarquons que la combinaison non naïve améliore la prévision. En effet, la combinaison par régression ou par optimisation améliore dans tous les cas la prévision par rapport au modèle individuel. Ceci est un résultat général, nous allons nous intéresser à la RMSE le long de la journée.

La climatologie étant notre meilleur modèle individuel, nous allons comparer les combinaisons avec ce modèle à partir de la RMSE le long de la journée. Afin de ne pas alourdir le graphique qui se trouve en Figure 14, nous comparons la climatologie uniquement aux résultats de régression.

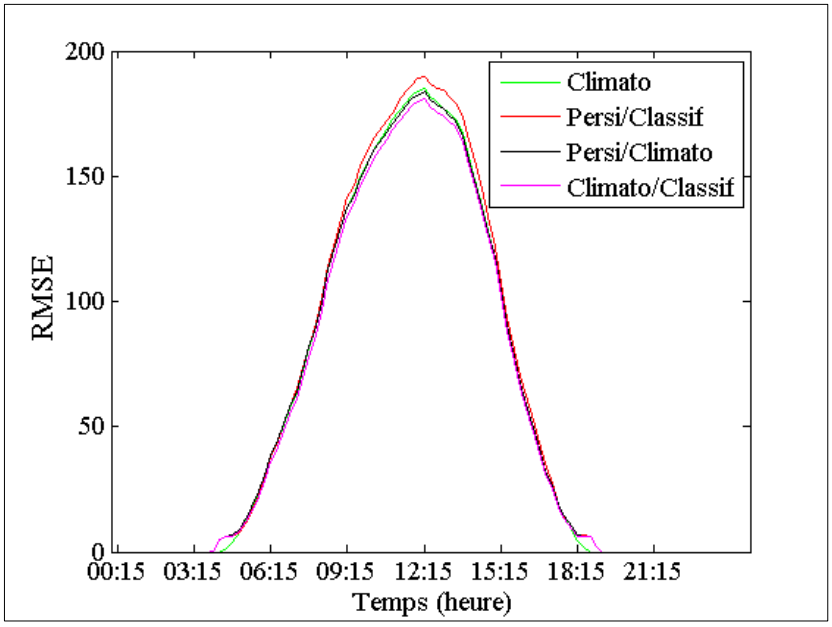


Figure 14: Comparaison de la RMSE le long de la journée entre la climatologie et la régression linéaire

Nous remarquons que les deux régressions à partir de climatologie sont très proches du modèle de climatologie seul hormis d'une part en tout début et toute fin de journée où elles sont moins performantes, et d'autre part aux alentours du midi solaire où elles sont légèrement meilleures. Les prévisions en tout début et en toute fin de journée n'étant généralement pas intéressantes, il ressort que la combinaison est meilleure tout au long de la journée. Ainsi il semblerait que même en combinant un modèle A avec un modèle B moins performant, nous obtenons une meilleure prévision que pour le modèle A seul.

V. Discussion et perspectives

A. Autres pistes pour l'amélioration de la classification

La distance :

La difficulté de la prévision photovoltaïque se situe dans la composante temporelle du problème. Il est difficile de déterminer un critère pour classer les journées qui conserve la donnée temporelle sans être trop dépendant des décalages ou des brusques variations.

Dans notre cas, nous avons choisi de raisonner sur une distance euclidienne, mais ceci peut être perfectible. Il serait notamment judicieux de raisonner avec une distance qui prenne en compte les décalages dans le temps. Ainsi la piste de la déformation temporelle dynamique peut être intéressante. Il s'agit d'un algorithme permettant de mesurer la similarité entre deux suites qui peuvent varier au cours du temps [FU et al., 2006].

La Figure 15 illustre le principe de la déformation temporelle dynamique.

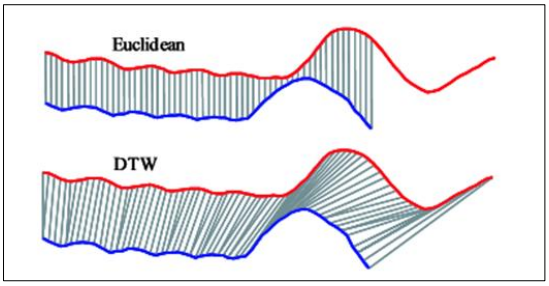


Figure 15: Illustration de la déformation temporelle dynamique

Cette méthode suppose de s'interroger sur la sensibilité aux décalages. Aussi n'étant pas une mesure de distance à proprement parler, des transformations sont à apporter pour l'intégrer dans un contexte de classification.

<u>La matrice de transition :</u>

Dans notre cas, nous n'utilisons pour la matrice de transition qu'un calcul empirique de probabilité de passage d'une classe à une autre. Aussi nous partons de l'affectation de la journée J à une unique classe. Or nous pourrions ajouter de l'incertitude à cette affectation de façon que ce que notre matrice de transition soit plus proche de la réalité. Le principe de la méthode est présenté en Figure 16.

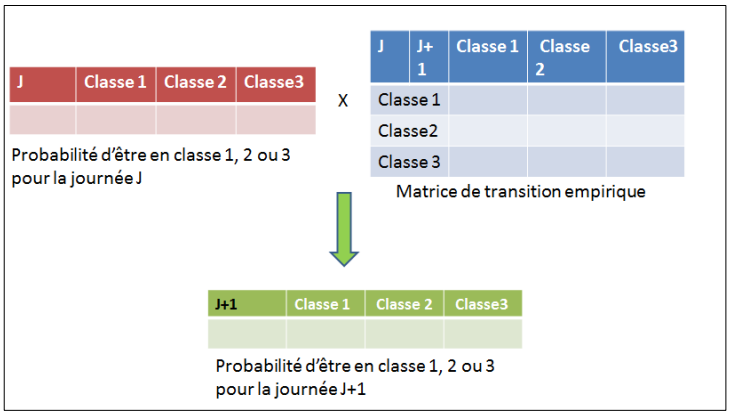


Figure 16: Description de la méthode d'amélioration de calcul de la matrice de transition

Pour ce faire, nous devons calculer les probabilités d'appartenance aux différentes classes pour la journée J. Nous disposons des distances aux centroides des classes. Il est alors possible de normaliser cette distance de façon à lui faire prendre la forme d'une probabilité.

Mise à jour dynamique de la classification :

La méthode proposée repose sur le découpage du jeu de données en plusieurs parties. Les données d'apprentissage sont bien délimitées dans le temps et peuvent être anciennes. Dans notre cas elles s'arrêtent en 2010. Nous avons aussi vu que la méthode de classification « dynamique » était meilleure que la statique. Nous pouvons donc penser que transformer la méthode proposée en une méthode dynamique pourrait s'avérer très avantageux. Ainsi il pourrait être intéressant d'opérer une mise à jour automatique de la classification au fur à et mesure de l'acquisition des données, afin de l'affiner.

Complexification de la caractérisation des journées par apports de données exogènes :

Enfin, nous pouvons réfléchir sur la bonne performance du modèle de climatologie simple pour continuer d'améliorer la classification. Cette méthode est une piste très intéressante et très flexible, elle n'est donc pas à abandonner. Cependant le résultat de la climatologie révèle une faille qu'il nous faut trouver. Il pourrait être intéressant d'intégrer des données météorologiques pour la caractérisation des journées types. Les classes obtenues seraient ainsi surement plus complexes, plus nombreuses mais aussi plus précises. Elles permettraient de révéler davantage la diversité des phénomènes météorologiques.

B. Réflexion sur la combinaison de modèles :

Affectation dynamique de coefficients/poids :

Dans notre étude, nous n'avons travaillé que sur une affectation de coefficients/poids unique pour la journée. Or d'une manière générale, la précision de la prévision des modèles n'est pas la même le long de la journée. Ces fluctuations dans la précision le long de la journée peuvent aussi changer d'un modèle à un autre. Ainsi il apparait intéressant d'étudier l'évolution des poids entre deux modèles le long de la journée pour mieux appréhender la combinaison de modèle.

La combinaison de modèle permet aussi d'envisager une amélioration de la précision de la prévision en utilisant des modèles performants pour des horizons de prévisions différents. Dans le cas où nous obtenons les données au fur et à mesure, nous pouvons envisager de combiner un modèle de prévision à J+1 avec un modèle ARMA pour des données échantillonnées toutes les 15 minutes. Ceci pourrait se faire à l'aide d'une affectation de poids dynamique le long de la journée. La prévision ARMA affinerait au fur et à mesure la prévision faite à J+1 par l'autre modèle. Pour cette affectation de poids dynamique, il pourrait être intéressant d'étudier la piste du filtre de Kalman pour notamment prendre en compte le bruit contenu dans les mesures.

Prise en compte de la corrélation entre les modèles :

Pour étudier la combinaison de modèle, nous avons utilisé trois modèles différents. Il serait intéressant d'étudier les corrélations entre ces modèles, leurs influences sur la combinaison et sur la précision de la prévision. Cette étude pourrait conduire à l'élaboration d'une technique de combinaison prenant en compte la corrélation entre les modèles, ou plus simplement à aider l'utilisateur dans son choix de modèles à combiner.

VI. Conclusion

Bilan technique :

Dans le cadre du projet Soleka, mes recherches avaient pour objectif de tester différentes méthodes de prévision solaire pour l'horizon J+1.

Pour cela, nous avons d'abord pu appréhender plusieurs méthodes de classification en journées types. D'une part via une caractérisation des classes avec les histogrammes de KC et d'autre part via une caractérisation des classes avec une série de KC calibrée. Une méthode de climatologie simpliste faisant une moyenne des sept derniers jours ainsi qu'un modèle de persistance ont aussi pu être étudiés. La comparaison de ces modèles a permis d'identifier plusieurs points importants. Nous pouvons insister sur la difficulté à battre le modèle simpliste de climatologie. Le critère de la RMSE le place en tête des modèles de prévision évoquées ci-dessus et suggère ainsi de le considérer comme modèle de référence. Par ailleurs, le développement d'une nouvelle méthode de classification en quatre classes n'a pas été vain. En effet celle-ci a non seulement permis de gagner en précision dans l'établissement des profils des journées types, mais aussi d'envisager la conservation de la structure temporelle des données. Ainsi il devient possible de prédire le déroulement de la journée.

Nous avons ensuite fait l'ébauche d'une analyse sur la combinaison de modèles. Cette étude préliminaire a essentiellement prouvé l'intérêt de la combinaison pour améliorer la précision de la prévision en termes de RMSE. En effet, qu'il s'agisse de combinaison par régression linéaire ou par optimisation, les modèles sont plus performants après hybridation que pris individuellement. Cette analyse a également permis d'identifier des pistes de recherches supplémentaires. Pour aller plus loin, l'une d'entre elles serait d'envisager une combinaison dynamique des modèles par utilisation d'un filtre de Kalman tenant compte des composantes aléatoires des signaux.

Ces découvertes ont fait émerger d'autres idées pour la recherche de méthodes performantes pour la prévision solaire. Ces idées pourront être exploitées par la suite par Reuniwatt.

Bilan personnel :

Ce stage dans une jeune entreprise telle que Reuniwatt m'a permis de mieux appréhender le travail de Recherche & Développement d'une start-up où les travaux menés s'inscrivent directement au cœur de son activité. Six mois chez Reuniwatt ont donc été enrichissants d'un point de vue personnel du fait de la proximité avec l'ensemble des employés qui a d'ailleurs facilité mon intégration. J'ai pu appliquer mes compétences en langage R et compléter celle en langage Matlab. Mes travaux m'ont aussi donné l'occasion d'approfondir mes connaissances en modélisation statistique notamment avec les modèles ARMA. Enfin, au

cours de ce stage il m'a plusieurs fois été demandé d'établir des notes techniques à l'attention des salariés ou futurs stagiaires susceptibles de reprendre mon travail ainsi que de restituer mes travaux réalisés par des présentations orales. Ces tâches m'ont donc permis de gagner en rigueur et clarté dans mes explications.

Annexes

Annexe I : L'algorithme du k-means

L'algorithme des k-moyennes (ou K-means en anglais) est un algorithme de partitionnement de données en un certain nombre de classes k. Le nombre de classe k doit être spécifié préalablement.

Initialisation :

- K individus sont tirés aléatoirement dans l'ensemble des individus I. Ils constituent les k centres de classes initiaux et sont provisoires.

- Chaque individu i de I est affecté à la classe la plus proche. Une première partition est ainsi formée.

Déroulement de l'algorithme (étape qui se répète):

- Au sein de chaque classe, le centre de gravité est recalculé et constitue le nouveau centre de classe.
- L'ensemble des individus sont réaffectés au centre de classe le plus proche. Une nouvelle partition est ainsi formée.

Arrêt de l'algorithme :

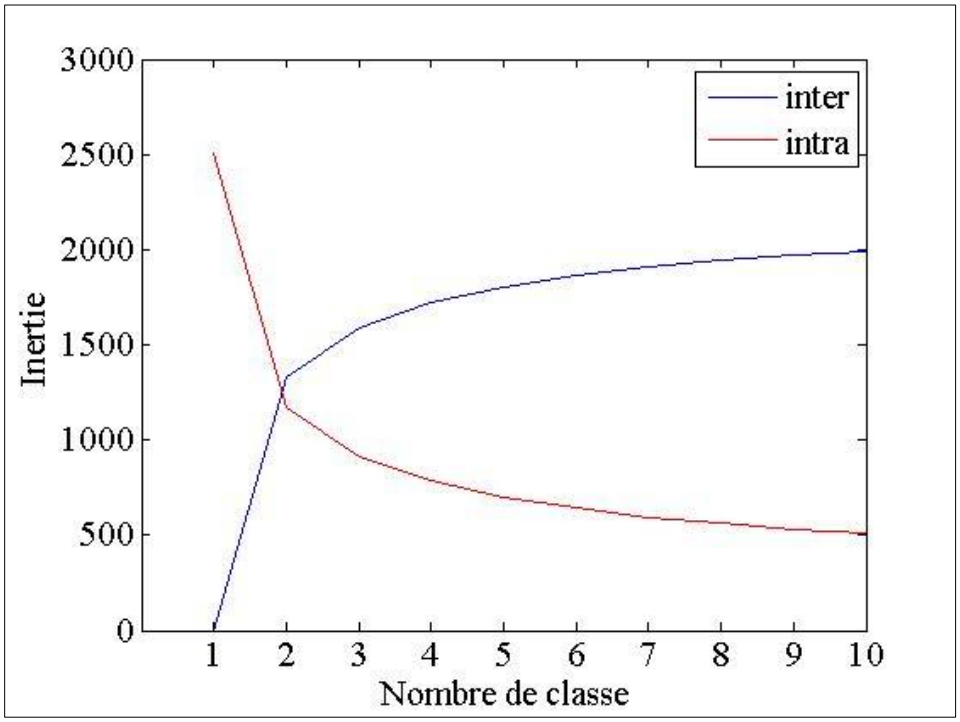
La réaffectation des individus s'arrête lorsque :

- deux itérations successives conduisent à une même partition.
- On fixe un critère d'arrêt tel que le nombre maximal d'itérations.

Remarque :

Le résultat de la classification risque de dépendre de l'étape d'initialisation.

Annexe II : Graphique de l'inertie intra et inter classe en fonction du nombre de classes pour la méthode 2



ANNEXE III : Explication de l'algorithme de minimisation sous contrainte

La fonction Matlab « *quadprog* » résout le problème suivant :

$$\min_x \frac{1}{2}x'Qx + f'x \ \ tel\ que \begin{cases} Ax \leq b \\ A_{eq}x = b_{eq} \end{cases}$$

Syntaxe Matlab :

$$[x, fval] = quadprog(Q, f, A, b, A_{eq}, b_{eq})$$

x : retourne le vecteur qui minimise le problème $\frac{1}{2}x'Qx + f'x$ sous contraintes

$fval$: retourne la valeur de la fonction à optimiser au point x de minimisation

Le programme d'optimisation nous intéresse pour minimiser la RMSE dans le cas d'hybridation de différents modèles de forecast. La RMSE est définie par :

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - F_i)^2}$$

O : série observée (dans notre cas, la puissance du GHI)

F : série de forecast (dans notre cas, la combinaison entre plusieurs modèles, en donnant un poids α_j à chacun des modèles sous contraintes $\forall\, j, \alpha_j \geq 0$ et $\sum \alpha_j = 1$)

Dans notre problème, le vecteur de poids α que l'on souhaite trouver doit être positif. Minimiser la RMSE revient donc à minimiser la MSE ($RMSE = \sqrt{MSE}$). Par souci de calcul matriciel, l'optimisation se fait donc sur la MSE.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(O_i - \sum_j \alpha_j F_{i,j}\right)^2$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} O_i^2 + \sum_{i=1}^{n}\left(\sum_j \alpha_j F_{i,j}\right)^2 + 2\sum_{i=1}^{n}\sum_j \alpha_j O_i F_{i,j}\right.$$

La matrice Q correspond aux termes quadratiques et le vecteur f correspond aux termes simples.

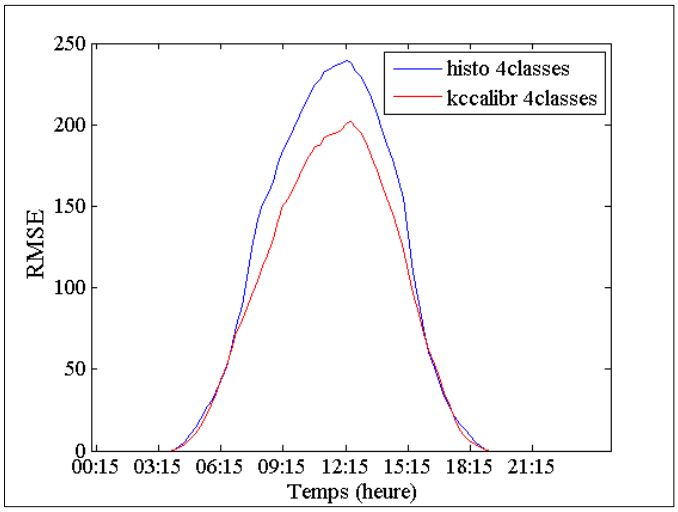
On pose :
F la matrice contenant les différents modèles de prévision
O la série des données observées que l'on veut approcher au mieux
Q et f sont définis par :

$$Q = 2 * \frac{1}{n} F'F$$

$$f = -2 * \frac{1}{n} F'O$$

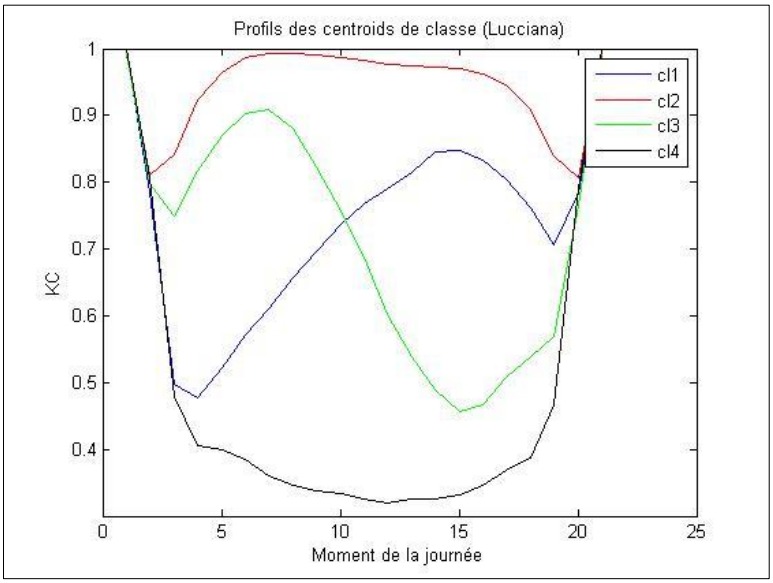
ANNEXE IV : Résultats de la RMSE pour les 2 modes de classification sur le site de Lucciana

Dans le cas de ce jeu de données, la meilleure performance pour le modèle de classification par histogrammes est celle obtenue avec 4 classes (elle est très proche de celle obtenue avec 3 classes).
Il en est de même avec le modèle de classification par la série de KC calibrée.
De ce fait, nous allons comparer les résultats de la RMSE obtenues avec une classification en quatre classes.



Dans ce cas, la classiification par série de KC est plus performante que celle par histogramme de KC.

La caractérisation des classes est sensiblement la même que pour le jeu de données Corte.



Bibliographie

ARER (2013). Bilan énergétique 2012 de la Réunion : Les chiffres clés. http://www.arer.org/IMG/pdf/BER_Technique_2012-2.pdf

Audition du Sénat par l'OPECST (17 novembre 2011). Comptes rendus de l'office parlementaire d'évaluation des choix scientifiques et technologiques, *Sécurité nucléaire et avenir de la filière nucléaire*.
http://www.senat.fr/compte-rendu-commissions/20111114/office.html#toc1

Reynaud D. (Mai 2014). Tableau de bord éolien-photovoltaïque au premier trimestre 2014. La Défense (FR) : Commissariat général au développement durable, 2014. Rapport d'études n°522.
http://www.statistiques.developpement-durable.gouv.fr/fileadmin/documents/Produits_editoriaux/Publications/Chiffres_et_statistiques/2014/chiffres-stats522-eolien2014t1-mai2014.pdf

De Menezes L. et al (2000). Review of guidelines for the use of combined forecast. *European Journal of Operational* Research, 120, pp. 190-204
http://forecastingprinciples.com/files/CombiningReview.pdf

Diagne M. et al (2013). Review of Solar irradiance forecatsion methodes and a proposetion for small-scale insular grids. *Renewable and Sustainable Energy Review*, n°27, pp. 65-76.

EPIA, Le développement du photovoltaïque dans le monde. Paris (FR) : Syndicat des énergies renouvelables SOLER, juin 2012.
http://www.enr.fr/docs/2010155958_SPV01Developpementmondemai2010.pdf

Fu Wai-Chee A. et al (2006). Scaling and time warping in time series querying. *The VLDB Journal*.
http://143.89.40.4/~raywong/paper/vldbj07-warping.pdf

Kalman R.E (1960) "A New Approach to Linear Filtering and Prediction Problems," Transactions of the ASME - Journal of Basic Engineering Vol. 82, pp. 35-45
https://www.cs.unc.edu/~welch/kalman/media/pdf/Kalman1960.pdf

Liandrat O. Approche hybride pour la prévision de la ressource solaire à La Réunion. Rapport de projet de fin d'études, INP Ensimag, 2012, 39 p.

Lefevre M. et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions., Atmos. Meas. Tech., 2013, pp. 2403-2418.
http://www.soda-pro.com/web-services/radiation/mcclear

Lorenz E. et al. (2012). Short term forecasting of solar irradiance by combining satellite data and numerical weather predictions. 27th European Photovoltaic Solar Energy conference and exhibition.

Pelland S. et al. Photovoltaic and Solar Forecasting: State of the art. IEA PVPS, 2013.

Soubdhan T. et al. (2009) Classification of daily solar radiation distributions using a mixture of Dirichlet distribution. Solar Energy 83.