

ANALYSE STATISTIQUE DE DONNÉES CAPTEURS DANS DES LOGEMENTS HABITÉS

Clément Leroux¹ Syrielle Montariol² Laure Nicollet³ Florian Rochet⁴

ENSAI (École Nationale de la Statistique et de l'Analyse de l'Information)

¹*clement.leroux@ensai.fr* ²*syrielle.montariol@ensai.fr*
³*laure.nicollet@ensai.fr* ⁴*florian.rochet@ensai.fr*

Résumé.

Le but de ce projet est d'analyser les séries temporelles provenant de capteurs de température, d'humidité, de CO₂ et d'ensoleillement. Deux éléments sont étudiés : l'évolution conjointe de deux séries, et les ruptures détectables dans les séries.

Pour pallier au problème des régressions fallacieuses dû à la non-stationnarité des séries, la méthode de Johansen est utilisée pour détecter des relations de cointégration. Par exemple, les séries de l'humidité et de la température sont souvent cointégrées dans une maison avec des pièces de petite taille et peu d'habitants, donc peu d'interactions. De même, les séries d'humidité dans deux pièces adjacentes d'une petite maison ont souvent des évolutions similaires. À l'inverse, aucune cointégration n'est détectée entre les séries dans une maison de grande taille.

Pour la détection de ruptures dans les séries, trois méthodes sont comparées : l'approche de Bai Perron, la méthode de Lavielle, et une fonction de détection spécifique implémentée pour compléter ces méthodes. Les deux dernières peuvent détecter immédiatement une ouverture de fenêtre, par la chute brusque de l'humidité, tandis que la première détecte une lente hausse de température due à l'ensoleillement ou les fluctuations du taux de CO₂ selon la présence des habitants dans la maison.

Mots-clés. Séries temporelles, cointégration, détection d'évènements, ruptures, données capteurs, qualité de l'air

Abstract.

The aim of this project is to analyse time series from diverse sensor readings such as temperature, relative humidity, CO₂, and sunshine. Two kind of methodologies are used : cointegration between series, and breakpoints detection.

The fact that the series are serially correlated and non-stationary brings up the problem of spurious regressions : the Johansen method is employed to address this issue and detect co-integration between series. The series of absolute humidity and temperature are most often co-integrated in a small house, with few inhabitants. Likewise, in a small house, absolute humidity has similar evolutions in adjacent rooms. On the contrary, in a large

house, co-integration can be found neither between series of measures (temperature or humidity) nor between close rooms.

Three methods are considered to detect events : The method of Lavielle, The Bai & Perron approach, and our own detection function (based on the selection of top 5% of slopes of the series). The Lavielle method and our function detect the opening of a window immediately, which implies a sharp drop in humidity, or the use of the shower, which implies a rise in humidity in adjoining rooms. The Bai & Perron approach detects low temperature increase due to sunshine and even the fluctuations of CO₂ whether or not the inhabitants are present.

Keywords. Time series, cointegration, event detection, breakpoints, sensor data, air quality

Présentation

Ces travaux ont été menés dans le cadre d'un projet académique réalisé en deuxième année (BAC+4) à l'ENSAI (Ecole Nationale de la Statistique et de l'Analyse de l'Information). Le projet a été encadré par Marie Ruellan en lien avec les travaux de recherche du laboratoire d'études SATIE et par Lionel Truquet, enseignant-chercheur de l'ENSAI. Ce projet est présenté en tant que lauréat du prix de la SFdS du meilleur projet statistique de deuxième année de l'ENSAI. Tous les traitements ont été effectués avec le logiciel R.

Des logements équipés de capteurs

De nos jours, respirer un air sain et consommer moins d'énergie sont des préoccupations majeures. La température, l'humidité ou encore le taux de CO₂ au sein de votre logement sont directement liés à ces problématiques. Ces grandeurs ont été relevées en janvier et février 2015 par des capteurs installés dans une maison et un appartement. Des événements spécifiques ont également été relevés comme l'ensoleillement ou l'ouverture des fenêtres. L'objectif premier de l'étude est de savoir si la température et l'humidité évoluent de façon similaire dans une même pièce ou entre deux pièces adjacentes. L'enjeu suivant est de déterminer quels événements impactent ces grandeurs ou le taux de CO₂.

Des séries intégrées d'ordre 1

Une analyse univariée des données permet d'avoir une première idée des ordres de grandeur et du comportement des mesures étudiées, et éventuellement de détecter certains problèmes dans nos données. La phase exploratoire a mis en avant plusieurs capteurs défectueux, ou en tout cas non exploitables. Nous avons ensuite analysé les propriétés des mesures considérées en tant que séries temporelles : saisonnalité et stationnarité.

Les autocorrélogrammes révèlent en général une saisonnalité journalière des séries, moins marquée sur les relevés d'humidité. Elle est néanmoins extrêmement difficile à éliminer complètement. On se restreint donc à étudier les séries sur des plages d'une journée, quitte à les différencier pour tenter de les stationnariser.

Il est aussi utile d'étudier la stationnarité des séries de la température, de l'humidité, et du CO₂. Dans la plupart des cas, ces trois grandeurs évoluent de façon non stationnaire sur une journée. Cependant, les différences premières de ces séries, qui reflètent directement l'évolution de ces grandeurs, sont stationnaires à l'échelle d'une journée. Leur étude permettra de rechercher une éventuelle liaison entre les évolutions des valeurs relevées.

Des similitudes dans les évolutions des différentes séries de données

Des liens entre différentes séries (température, humidité, CO₂) peuvent être mis en évidence. A cet effet, nous étudions les corrélations entre ces séries ainsi que la présence de phénomènes dits de cointégration.

L'étude des corrélations croisées entre les séries différenciées permet de déterminer quelles grandeurs évoluent dans le même sens. Elle montre ici que la température et l'humidité d'une même pièce évoluent dans le même sens et en même temps ; en revanche, aucun lien n'est détecté avec les variations du CO₂. Dans des pièces de grandes tailles, un retard de quelques minutes de la température sur l'humidité est constaté.

Une recherche de cointégration peut conduire à un résultat plus fort, en déterminant si les mesures évoluent conjointement ou non, avec un éventuel retard. La méthode de Johanssen, présentée par COLLETAZ 2015, a mis en évidence des liens entre les grandeurs d'intérêt : Au sein d'une même pièce, il arrive que la température et l'humidité absolue soient cointégrées. D'autre part, dans des pièces proches, l'humidité absolue évolue de manière uniforme. Enfin, en raison de la bonne isolation des logements, les données relevées à l'extérieur sont rarement cointégrées avec les données intérieures.

Différentes méthodes permettent de détecter des changements brusques ou structurels

L'absence de cointégration n'indique toutefois pas nécessairement qu'elles n'évoluent jamais de manière commune, même ponctuellement. L'étude de l'impact d'évènements spécifiques permet de mettre en évidence des pics ou des tendances communes à plusieurs séries.

La méthode présentée par LAVIELLE 2004, appliquée aux différences premières des séries permet le plus souvent de détecter une chute brusque de l'humidité et la température, causée par l'ouverture d'une fenêtre.

Ces séries sont globalement stables et ponctuées par de brusques fluctuations, ce qui est un évolution courante pour l'humidité et la température au sein d'un logement. Dans ce cas, une méthode plus directe, détectant simplement les 5% des pentes les plus fortes de la série, suffit à relever chaque chute de température ou d'humidité due à une ouverture de fenêtre.

Ces deux méthodes permettent aussi de détecter dans une pièce attenante à la salle de bain une hausse de la température, probablement due à la mise en route du chauffe-eau, et la hausse de l'humidité causée par l'utilisation de la douche.

L'approche de BAI et PERRON 2003, qui partitionne la série en utilisant le modèle de régression linéaire, permet de déceler les évolutions lentes même en présence de bruit. Par exemple, la hausse lente de température dans les pièces d'un logement suite à une exposition aux rayons du soleil est mise en évidence avec précision, distinguant la phase de croissance de température par rapport au reste de la série. Cette approche conduit aussi à un partitionnement de l'évolution du CO₂ correspondant à la présence ou non des individus dans la maison, ainsi que leur degré d'activité telles les phases de sommeil.

Pour conclure, l'étude des données relevées par les capteurs de logements permet de mieux comprendre comment les grandeurs liées à la qualité de l'air évoluent à l'intérieur d'une habitation. Ainsi, les grandeurs varient souvent de la même façon dans l'ensemble du logement. La recherche d'évènements influents sur ces grandeurs nécessite d'adopter des méthodes de détection différentes en fonction de la durée de ces évènements.

Références

- COLLETAZ, Gilbert (2015). *Une présentation simplifiée de l'approche de Johansen - La cointégration dans un cadre multivarié*.
- LAVIELLE, Marc (2004). *Using penalized contrasts for the change-point problem*. RR-5339, INRIA. p.18.
- BAI, Jushan et Pierre PERRON (2003). "Computation and Analysis of Multiple Structural Change Models". In : *Journal of Applied Econometrics*, p. 1-22.