



HAL
open science

Emergency Department Admissions Overflow Modeling by a Clustering of Time Evolving Clinical Diagnoses

Gregory Soler, Guillaume Bouleux, Eric Marcon, Aymeric Cantais, Sylvie
Pillet, Olivier Mory

► **To cite this version:**

Gregory Soler, Guillaume Bouleux, Eric Marcon, Aymeric Cantais, Sylvie Pillet, et al.. Emergency Department Admissions Overflow Modeling by a Clustering of Time Evolving Clinical Diagnoses. 14th IEEE International Conference on Automation Science and Engineering (CASE 2018), Aug 2018, Munich, Germany. hal-01885975

HAL Id: hal-01885975

<https://hal.science/hal-01885975v1>

Submitted on 2 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Emergency Department Admissions Overflow Modeling by a Clustering of Time Evolving Clinical Diagnoses

Gregory Soler¹, Guillaume Bouleux¹, Eric Marcon¹, Aymeric Cantais², Sylvie Pillet² and Olivier Mory³

Abstract—Emergency Department (ED) of hospitals are greatly impacted by winter epidemics due to respiratory diseases and patient flow has long been essential to detect the underlying overcrowding. In this paper we propose to model the admission flow corresponding to clinical diagnoses encoded with ICD-10 which are more likely linked with respiratory diseases. To achieve this, clustering algorithms are applied on time evolving diagnosis in the adult ED of Saint-Etienne and benchmarked regarding a time series of laboratory-confirmed influenza data. For both K-Means and Hierarchical algorithms, the cluster containing the laboratory-confirmed series is composed of ICD-10 codes representing respiratory diseases and diseases linked with cardiac disorders, showing that these diseases present similar variations overtime. The information contained in such a cluster allow to plot the average number of arrivals and the average length of stay of the patients in ED only with these diagnoses.

Index Terms—Average length of stay, Diagnoses clustering, Hierarchical Clustering, K-Means, Influenza, Overcrowding, Patient flow, Respiratory diseases

I. INTRODUCTION

Emergency departments (ED) sometimes fail to meet patient expectations regarding the care that they provide. Such a failure is often concomitant with excess hospital admissions which lead to overcrowding [1]. There is no definition or quantification of overcrowding making consensus [2], the most used is the quantification of the ambulatory diversion when ED can not afford new patient any more [3]. Other criteria include occupancy rate, staff perception waiting time or the number of patients who leave without being seen [4] [5] [6] [7]. To better understand this phenomenon, it is essential that we study ED patient flow. The patient flow can be modeled with variables such as the number of arrivals, length of stay in ED or waiting time [8] [3]. Warning score with updates in real time or forecasting overcrowding have already been proposed, either based on mathematical tools or simulation [9] [10].

It has already been demonstrated that respiratory diseases are those that most greatly impact ED admissions, particularly during winter [11], [12]. Respiratory diseases that arise during winter epidemics are well diagnosed by physicians

in children. Most children admitted to the ED are affected by bronchiolitis caused by the Respiratory Syncytial Virus (RSV) [13]. This, however, is not the case for adults, in whom respiratory diseases are not easily diagnosed. That such misdiagnosis occurs for the latter is evinced by the small number of adult flu cases reported in emergency data base. Studies have quantified the degree of implication of Influenza or RSV in cardiorespiratory diseases inducing hospitalisation and deaths for adults, by using both ED data and laboratory-confirmed data, showing that both viruses impact adults [14] [15]. However, a thorough study of diagnoses coded with the 10th revision of International Classification of Disease (ICD-10) involved in respiratory diseases in ED in winter epidemic has not been done. A study of ED patient flow by diagnosis is necessary to figure out those that are linked with influenza-like illness seasonal epidemic, and thereby overcrowding. We decided to use clustering of diagnoses in ED overtime. It will allow to highlight diseases which behave in the same way over time. Diagnoses clustering has been done, sometimes with human learning in order to improve ICD-9 or with portioning techniques allowing better medical decision rule [16] [17]. In section II, we present our data base for the study and clustering algorithms applied to it and in section III and IV we discussed clustering results.

II. MATERIAL AND METHODS

A. Data Base

We used a French National Database dedicated to emergencies, called Emergency Visit Summary (EVS, RPU in French). We collected the data from the EVS database of the Academic Hospital of Saint Etienne, France. Arrivals of adult patients at ED from June 2013 to August 2017 were then analyzed. The data base for clustering is composed of weekly time series of numbers of arrivals patients, for each diagnosis. Each series refers to one diagnosis, encoded with ICD-10. These codes were truncated after the second number. Only times series whose diagnosis appeared more than 200 times were conserved which corresponds to the average occurrence of a diagnosis during the observation period, whatever the nature of the ICD-10 code. A series of the weekly number of positive laboratory-confirmed test of influenza performed on children and adults in ED is also included in our data base. All the analysis are made with Python 3.6 with machine learning packages *scipy* and *sklearn*.

B. Times-series clustering

Clustering is a data mining exploratory method to form object groups and identify structures of unlabeled data set.

This work was supported by the PREPS-15-15-0668 grant of the french ministry of health and social affairs

¹GS, GB and EM are with the Univ Lyon, INSA de Lyon, DISP, EA 4570, F-69621, Villeurbanne, France {gregory.soler, guillaume.bouleux, eric.marcon}@insa-lyon.fr

²AC and SP are with Univ Lyon, UJM-Saint Etienne, GIMAP, EA 3064, F-42000, Saint Etienne, France {aymeric.cantais, sylvie.pillet}@chu-st-etienne.fr

³OM is with the University Hospital of Saint Etienne, F-42000, Saint Etienne, France olivier.mory@chu-st-etienne.fr

It is one of the most famous unsupervised learning methods. These methods form groups of objects or individuals with maximizing their within-group similarity and their between-group dissimilarity. Time-series clustering is an application of clustering to non-static data, i.e. data depending on time. Most of clustering algorithms used for times series are algorithms derived from those used in static data like partitioning methods or hierarchical methods. The main step of a clustering algorithm is to define the distance between objects or time-series to quantify their similarity during the algorithm. The most famous distance is the euclidean distance. The distance is chosen according to times-series characteristics, for example, if times series have equal or unequal length [18]. Very good reviews of state-of-the-art clustering methods have been published [19] [20] [21]. As a time series can be high dimensional, a dimension reduction is used to reduce the length of it. Principal Component Analysis (PCA), Kernel Principal Component Analysis when data have non linear patterns, or Discrete Fourier Transformation allow such reduction [22] [23].

We used both partitioning and hierarchical methods to form clusters. The partitioning methods computed was k-means and the hierarchical method was an agglomerative hierarchical method.

K-means: Given the set of n time-series $\{x_i | i = 1, \dots, n\}$, with $x_i = (x_{i1}, \dots, x_{ip})$ each of length p , and $\{v_k | k = 1, \dots, K\}$ center clusters with K , number of clusters, known, we assign each time series x_i to one, and only one cluster v_k . For short, after initialization of center clusters, we assign each time-series to the cluster whose its center cluster is the nearest with respect to the euclidean distance. Next, center clusters are updated, being average time-series of all time-series' cluster. Then, the processus is repeated until stabilization [19]. In our scenario, n refers to the number of ICD-10 codes in our data plus the laboratory-confirmed test series and p to the numbers of weeks for the study. This method can only be used with time series of equal length.

Each time series, which corresponds to either one ICD-10 code or the influenza test series, is standardized, by subtracting from it its mean and by dividing by the classical L^2 -norm. This standardisation is absolutely necessary because we need to make clusters of diagnosis having similarities in terms of variations and not in terms of occurrence [22]. The main disadvantage of K-means is that the number of clusters has to be specified. As we do not know the numbers of clusters K , this is chosen with criteria such as elbow variance or silhouette score.

For each number of clusters possible (from 1 to n), after computing the K-means algorithm, the elbow variance and the silhouette score are calculated. The elbow method consists in plotting the ratio of variance explained (variance between center clusters divided by total variance) in fonction of the number of clusters chosen. The location of a bend (or the beginning of an asymptote) only give an indication of the appropriate number of clusters K [24]. The silhouette score is calculated using the mean intra-cluster distance (a_i) and the mean nearest-cluster distance (b_i) for the time series

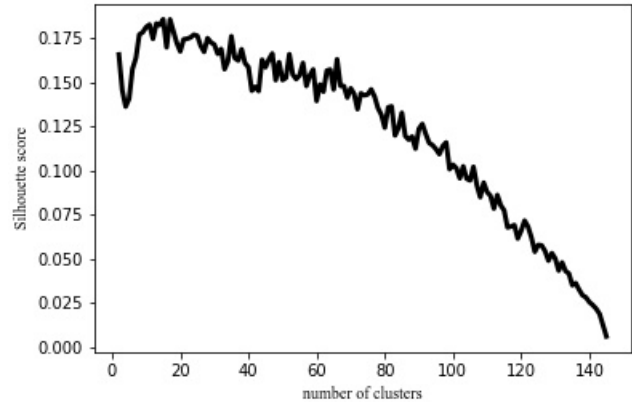


Fig. 1. Silhouette score by number of clusters

x_i . The silhouette score is $\frac{1}{n} \sum_{i=1}^n \frac{b_i - a_i}{\max(a_i, b_i)}$. The maximum score corresponds to the number of clusters K . The silhouette score is also a indicator of the performance of the algorithm, it is bounded between -1 for incorrect clustering and 1 for good clustering [25].

Before clustering, a reduction of dimension is also performed with PCA to delete data noise. The number of principal components retained q is chosen with AIC criteria. As the number of diagnoses retained (n) is far lower than time-series length (p), standard AIC criteria is modified and calculated for each number of principal components retained possible, with λ_l eigenvalues of the l component principal, by decreasing order [26] :

$$AIC^*(j) = (n - 1 - j) \log\left(\frac{1}{n - 1 - j} \sum_{l=j+1}^{n-1} \lambda_l\right) - \log\left(\prod_{l=j+1}^{n-1} \lambda_l\right) - \frac{(n - j - 2)(n - j + 1)}{p} \quad (1)$$

The number of components q retained corresponds to $\arg \min_{1 \leq j \leq n-2} AIC^*(j)$. Thus, the K-means procedure is applied on time-series projected on components retained.

Agglomerative Hierarchical Clustering: The Agglomerative Hierarchical Clustering starts by placing each time series in its own cluster. Then, a distance between each time series is calculated, e.g. the euclidean distance. Then, the two most similar clusters are merged into one cluster and the distance between the new cluster and others are updated by a formula chosen depending on the distance. We used the iterative Ward formula to calculate the distance between clusters C and C' where C is the new cluster formed by the merging of clusters U et V (2). The processus is iterated until the merging of all time series into one cluster. All the merge are summarized in a tree plot called dendrogram where the nodes represent a merge [27].

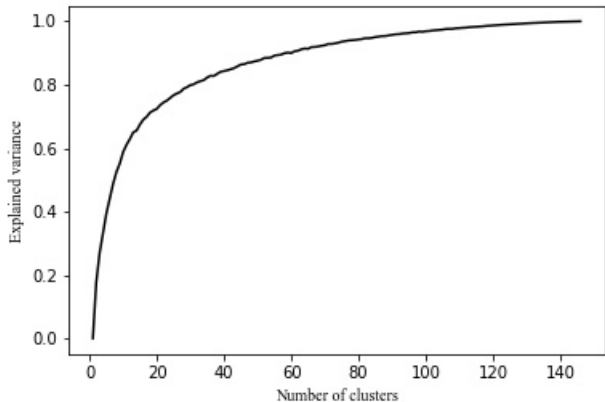


Fig. 2. Explained Variance by number of clusters

$$\begin{aligned}
 d_W(C, C')^2 &= \frac{|C'| + |U|}{|C'| + |U| + |V|} d_W(C', U)^2 \\
 &+ \frac{|C'| + |V|}{|C'| + |U| + |V|} d_W(C', V)^2 \\
 &- \frac{|C'|}{|C'| + |U| + |V|} d_W(U, V)^2
 \end{aligned} \quad (2)$$

Clusters can be identified with the use of a threshold. All time-series (or merge) with a distance lower than seven tenths belong to the same cluster. This is represented with colors in a dendrogram where each color refers to a cluster, except blue.

C. Highlighting clusters representative of respiratory diseases

After clustering the diagnoses, the objective is to find a cluster that depicts the flow of respiratory diseases, leading cause of overcrowding during winter epidemics. We define the cluster containing the series of the weekly number of positive influenza tests as the representative cluster of respiratory diseases. With these results, for each clustering algorithm, we make two series evident. We plot the weekly average time series of number of arrivals of patients with the ICD-10 codes contained in our representative cluster. We plot also the weekly average length of stay of patients in ED with these ICD-10 codes. These series are representative time series of influenza-like illness patient flow, responsible for overcrowding during winter epidemics.

III. RESULTS

Regarding K-Means, after computing PCA, the AIC criteria gives a number of principal components to retain equal to six. The Silhouette Score of K-Means procedure is calculated and the maximum gives a number of clusters equals to 15 (Fig 1), which refers to a score near to 0.2. The ratio of explained variance corresponding to this partition is equal to 0.70 (Fig 2). The cluster whose the influenza test series belongs to contains a majority of ICD J code, which refers

to respiratory disorders. The ICD-10 codes of this cluster are listed in Table I. In addition to ICD-10 code for "Bacterial Pneumonia" (J15), "Pneumonia" (J18), "Bronchitis" (J20) and "Other chronic obstructive pulmonary disease" (J44), the K-Means cluster contains also "Heart Failure" (I50).

TABLE I

TIME SERIES OF CLINICAL DIAGNOSES CONTAINED IN THE CLUSTER WITH INFLUENZA TEST SERIES BY CLUSTERING ALGORITHMS

| | |
|-------------------------|---|
| K-Means clustering | I50 J15 J18 J20 J44 |
| Hierarchical clustering | I10 I25 I48 I50 J15 J18 J20 J44 J96 R04 |

About Hierarchical clustering, the resulting dendrogram is plotted in Fig. 3. The threshold gives a number of clusters equal to six. The cluster with the influenza test series is mostly composed of respiratory and cardiac diseases. This is the purple cluster in Fig. 3. Summarized in Table I, it contains, all diagnosis in K-Means cluster but also "Essential hypertension" (I10), "Chronic ischemic heart disease" (I25), "Fibrillation" (I48), "Respiratory failure" (J96) and "Haemorrhage from respiratory passages" (R04). The hierarchical algorithm gives also information in retained cluster : we remark that the influenza series is really similar to the "J20" series and the cluster is formed by two subgroups, one composes of respiratory diseases and the other composes of cardiac diseases and haemorrhage. Other clusters do not show a clearly linked diagnostic pattern, except the cyan cluster composed of many traumatologic diagnoses, showing that the respiratory diseases, and those which are linked with, have clearly a particular pattern.

The average time series of arrivals, for each algorithm, is plotted in Fig. 4. This weekly time series clearly present an annual seasonality with a huge increase during winter. This time series is our definition of patient flow that overcrowds ED. They clearly show annual peaks of ED visits; during winter 2013-2014, in week 01 (i.e. first week of 2014), during winter 2014-2015, in week 07 (i.e. mid-february), during winter 2015-2016, in week 50 and 07 (two average peaks), and during winter 2016-2017 in week 01 (huge peak). Regarding the average length of stay, the annual seasonality, i.e. the increasing during winter, is also visible but less pronounced (Fig. 5).

IV. DISCUSSION

Both clustering algorithms give a cluster containing a majority of diagnoses of respiratory diseases (code J), in addition to diseases linked with them as haemorrhage from respiratory passages and cardiac diseases. Surveys have already been done about the concomitance between cardiac and respiratory diseases [28] [29], showing that cardiac diseases can be a consequence of respiratory diseases. Moreover, the diagnoses of retained clusters of both algorithms are similar, while totally independent: all diagnoses in cluster retained with K-Means are in the cluster of hierarchical algorithm. It implies that there is no ambiguity for the existence of a significant seasonal patient flow linked to respiratory

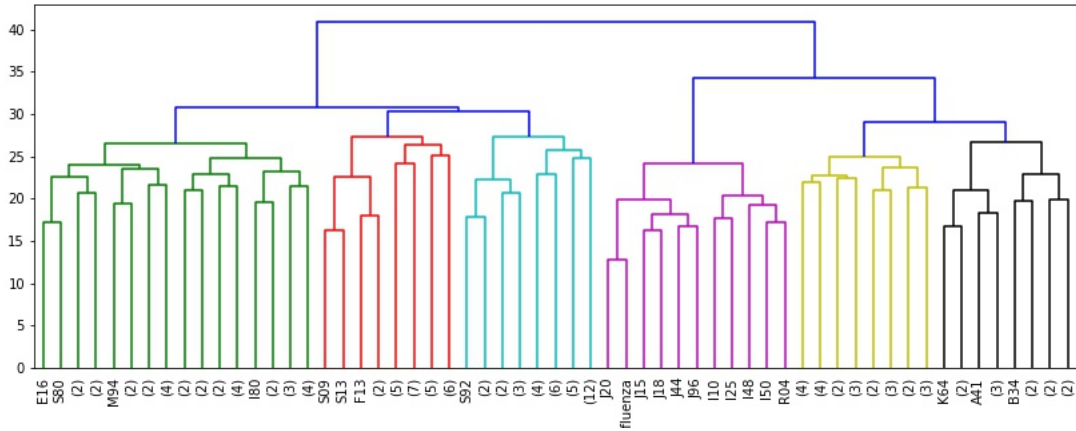


Fig. 3. Dendrogram of Agglomerative Hierarchical Clustering

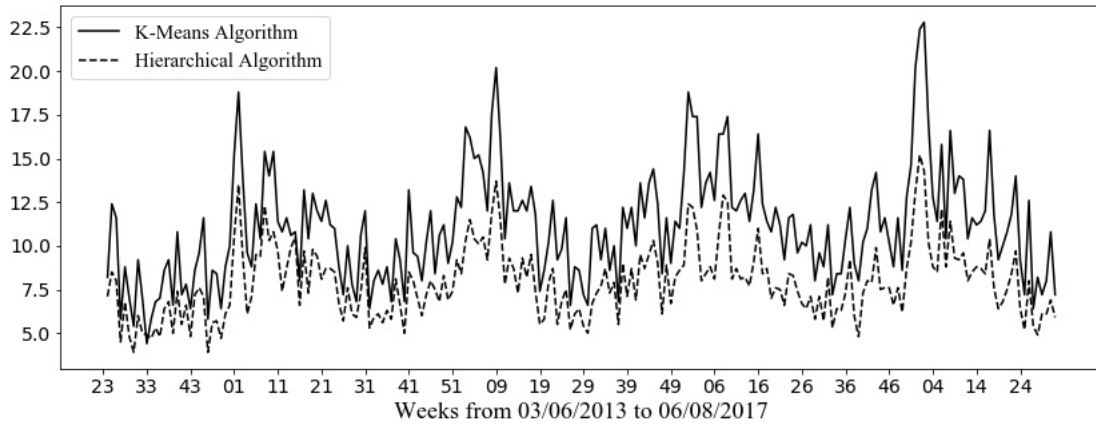


Fig. 4. Average time series of number of arrival of patients with diagnoses in retained cluster

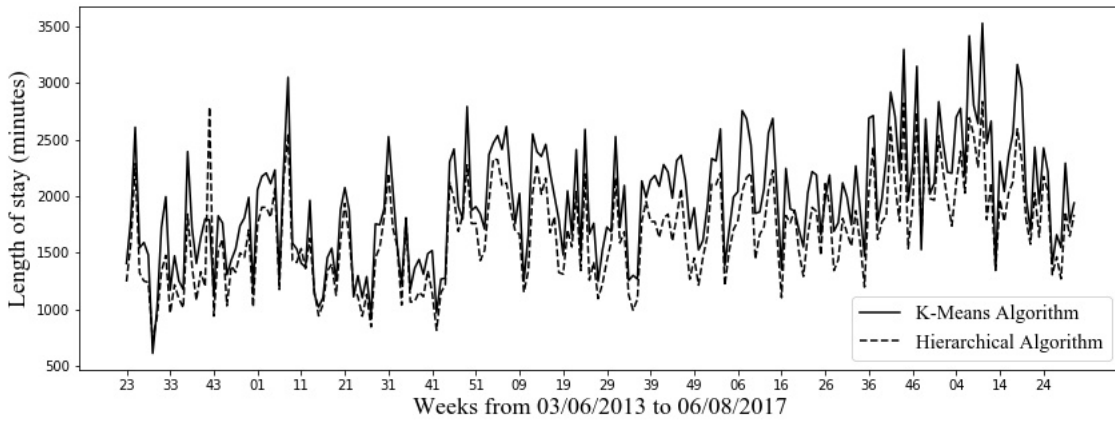


Fig. 5. Time series of average length of stay in ED of patients with diagnoses in retained cluster

diseases, responsible of overcrowding during winter. These results are significant and powerful because of the used of a learning technique without any consideration for clusters or diagnosis patterns.

The ICD-10 codes "J10", "J11", "J12" coding for Influenza illnesses are not enough diagnosed to be retained in the study, showing that there is an under-diagnostic of that pathology, whereas it is known to be a leading cause of hospitalisation and death in adults and especially the elderly [14]. It could be explained by the fact that for some patients, either there is a misdiagnosis (the patient has a flu but not diagnosed, he is diagnosed with other respiratory disease for example) or only a consequence of the flu (e.g. haemorrhage, heart failure or even bacterial pneumonia).

According to Fig. 4, the average number of arrivals in resultant time series implies that the epidemics in 2014-2015 and 2015-2016 do not have huge peak but epidemics seem to be longer. These observations coincide with description of the length influenza season by "Sante Publique France" where they affirm that epidemic winter season 2014-2015 and 2015-2016 were far longer than 2013-2014 [30]. The average length of stay of respiratory and associated diseases, which depends naturally on the severity of patient illness, seems to be impacted by winter epidemic and the increasing of arrivals.

V. CONCLUSION

We performed a hierarchical and partitioning clustering algorithms on number of arrivals of adults in ED which both give a cluster composed of respiratory diseases, and confirm that it exists diseases linked with as cardiac disorders and pulmonary haemorrhage. There is a real seasonal pattern of patient flow with these diagnoses, with a peak during epidemic winter. The resulting time series are like a model or an example that describes patient flow which often leads to overcrowding. Prediction or warning score on such time series can be used to detect overcrowding and thus to reduce length of stay.

ACKNOWLEDGMENT

The authors are grateful to Pr. A. Viallon, University Hospital of Saint Etienne and head of the adult emergency department for having authorized the processing of the data.

REFERENCES

- [1] D. R. Eitel, S. E. Rudkin, M. A. Malvey, J. P. Killeen, and J. M. Pines, "Improving Service Quality by Understanding Emergency Department Flow: A White Paper and Position Statement Prepared For the American Academy of Emergency Medicine," *The Journal of Emergency Medicine*, vol. 38, no. 1, pp. 70–79, Jan. 2010. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0736467908003910>
- [2] S. S. Jones, T. L. Allen, T. J. Flottemesch, and S. J. Welch, "An Independent Evaluation of Four Quantitative Emergency Department Crowding Scales," *Academic Emergency Medicine*, vol. 13, no. 11, pp. 1204–1211, Nov. 2006. [Online]. Available: <http://doi.wiley.com/10.1197/j.aem.2006.05.021>
- [3] A. Kolker, "Process Modeling of Emergency Department Patient Flow: Effect of Patient Length of Stay on ED Diversion," *Journal of Medical Systems*, vol. 32, no. 5, pp. 389–401, Oct. 2008. [Online]. Available: <http://link.springer.com/10.1007/s10916-008-9144-x>
- [4] J. L. Wiler, R. T. Griffey, and T. Olsen, "Review of Modeling Approaches for Emergency Department Patient Flow and Crowding Research: MODELING APPROACHES FOR ED PATIENT FLOW AND CROWDING RESEARCH," *Academic Emergency Medicine*, vol. 18, no. 12, pp. 1371–1379, Dec. 2011. [Online]. Available: <http://doi.wiley.com/10.1111/j.1553-2712.2011.01135.x>
- [5] S. J. Weiss, R. Derlet, J. Arndahl, A. A. Ernst, J. Richards, M. Fernandez-Frankelton, R. Schwab, T. O. Stair, P. Vicellio, and D. Levy, "Estimating the degree of emergency department overcrowding in academic medical centers: results of the National ED Overcrowding Study (NEDOCS)," *Academic Emergency Medicine*, vol. 11, no. 1, pp. 38–50, 2004.
- [6] S. J. Weiss, A. A. Ernst, R. Derlet, R. King, A. Bair, and T. G. Nick, "Relationship between the National ED Overcrowding Scale and the number of patients who leave without being seen in an academic ED," *The American Journal of Emergency Medicine*, vol. 23, no. 3, pp. 288–294, May 2005. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0735675705000847>
- [7] O. El-Rifai, T. Garaix, V. Augusto, and X. Xie, "A stochastic optimization model for shift scheduling in emergency departments," *Health Care Management Science*, vol. 18, no. 3, pp. 289–302, Sept. 2015.
- [8] M. Laskowski, R. D. McLeod, M. R. Friesen, B. W. Podaima, and A. S. Alfa, "Models of Emergency Departments for Reducing Patient Waiting Times," *PLoS ONE*, vol. 4, no. 7, p. e6127, July 2009. [Online]. Available: <http://dx.plos.org/10.1371/journal.pone.0006127>
- [9] N. R. Hoot, S. K. Epstein, T. L. Allen, S. S. Jones, K. M. Baumbin, N. Chawla, A. T. Lee, J. M. Pines, A. K. Klair, B. D. Gordon, T. J. Flottemesch, L. J. LeBlanc, I. Jones, S. R. Levin, C. Zhou, C. S. Gadd, and D. Aronsky, "Forecasting Emergency Department Crowding: An External, Multicenter Evaluation," *Annals of Emergency Medicine*, vol. 54, no. 4, pp. 514–522.e19, Oct. 2009. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0196064409006155>
- [10] G. Bouleux, E. Marcon, and O. Mory, "Early index for detection of pediatric emergency department crowding," *IEEE journal of biomedical and health informatics*, vol. 19, no. 6, pp. 1929–1936, 2015.
- [11] G. Hall, T. Krahn, A. Majury, A. Van Dijk, G. Evans, K. Moore, and A. Maier, "Emergency department surveillance as a proxy for the prediction of circulating respiratory viral disease in Eastern Ontario," *Canadian Journal of Infectious Diseases and Medical Microbiology*, vol. 24, no. 3, pp. 150–154, 2013.
- [12] F. H. Y. Yap, P.-L. Ho, K.-F. Lam, P. K. S. Chan, Y.-H. Cheng, and J. S. M. Peiris, "Excess hospital admissions for pneumonia, chronic obstructive pulmonary disease, and heart failure during influenza seasons in Hong Kong," *Journal of Medical Virology*, vol. 73, 2004. [Online]. Available: <http://www.readcube.com/articles/10.1002/jmv.20135>
- [13] D. L. Schanzer, J. M. Langley, and T. W. S. Tam, "Hospitalization Attributable to Influenza and Other Viral Respiratory Illnesses in Canadian Children," *The Pediatric Infectious Disease Journal*, vol. 25, no. 9, pp. 795–800, Sept. 2006.
- [14] D. L. Schanzer, J. M. Langley, and T. W. Tam, "Role of influenza and other respiratory viruses in admissions of adults to Canadian hospitals," *Influenza and Other Respiratory Viruses*, vol. 2, no. 1, pp. 1–8, Jan. 2008. [Online]. Available: <http://doi.wiley.com/10.1111/j.1750-2659.2008.00035.x>
- [15] S. K. Schindeler, D. J. Muscatello, M. J. Ferson, K. D. Rogers, P. Grant, and T. Churches, "Evaluation of alternative respiratory syndromes for specific syndromic surveillance of influenza and respiratory syncytial virus: a time series analysis," *BMC Infectious Diseases*, vol. 9, no. 1, Dec. 2009. [Online]. Available: <http://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-9-190>
- [16] I. G. Stiell, G. H. Greenberg, R. D. McKnight, R. C. Nair, I. McDowell, and J. R. Worthington, "A study to develop clinical decision rules for the use of radiography in acute ankle injuries," *Annals of emergency medicine*, vol. 21, no. 4, pp. 384–390, 1992.
- [17] D. A. Travers, "Diagnosis Clusters for Emergency Medicine," *Academic Emergency Medicine*, vol. 10, no. 12, pp. 1337–1344, Dec. 2003. [Online]. Available: [http://doi.wiley.com/10.1197/S1069-6563\(03\)00536-0](http://doi.wiley.com/10.1197/S1069-6563(03)00536-0)
- [18] X. Zhang, J. Liu, Y. Du, and T. Lv, "A novel clustering method on time series data," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11 891–11 900, Sept. 2011. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0957417411004908>
- [19] T. Warren Liao, "Clustering of time series dataa

- survey,” *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, Nov. 2005. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0031320305001305>
- [20] S. Aghabozorgi, A. Seyed Shirshorshidi, and T. Ying Wah, “Time-series clustering A decade review,” *Information Systems*, vol. 53, pp. 16–38, Oct. 2015. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0306437915000733>
- [21] S. Rani and G. Sikka, “Recent techniques of clustering of time series data: a survey,” *International Journal of Computer Applications*, vol. 52, no. 15, 2012.
- [22] M. Gavrilov, D. Anguelov, P. Indyk, and R. Motwani, “Mining the stock market (extended abstract): which measure is best?” in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000, pp. 487–496.
- [23] K. Kalpakis, D. Gada, and V. Puttagunta, “Distance measures for effective clustering of ARIMA time-series,” in *Proceedings 2001 IEEE International Conference on Data Mining*, 2001, pp. 273–280.
- [24] D. J. Ketchen and C. L. Shook, “The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique,” *Strategic Management Journal*, vol. 17, no. 6, pp. 441–458, June 1996. [Online]. Available: [http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0266\(199606\)17:6<441::AID-SMJ819>3.0.CO;2-G/abstract](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G/abstract)
- [25] L. Kaufman and P. J. Rousseeuw, “Partitioning Around Medoids (Program PAM),” in *Finding Groups in Data*. John Wiley & Sons, Inc., 1990, pp. 68–125, doi: 10.1002/9780470316801.ch2. [Online]. Available: <http://onlinelibrary.wiley.com/doi/10.1002/9780470316801.ch2/summary>
- [26] Z. D. Bai, F. Yasunori, and P. C. Kwok, “High-dimensional consistency of AIC and BIC for estimating the number of significant components in principal component analysis,” *The Annals of Statistics*, to appear, 2017.
- [27] C. Goutte, P. Toft, E. Rostrup, F. Nielsen, and L. K. Hansen, “On clustering fMRI time series,” *NeuroImage*, vol. 9, no. 3, pp. 298–310, Mar. 1999.
- [28] C. S. Kwok, S. Aslam, E. Kontopantelis, P. K. Myint, M. J. S. Zaman, I. Buchan, Y. K. Loke, and M. A. Mamas, “Influenza, influenza-like symptoms and their association with cardiovascular risks: a systematic review and meta-analysis of observational studies,” *International Journal of Clinical Practice*, vol. 69, no. 9, pp. 928–937, Sept. 2015. [Online]. Available: <http://doi.wiley.com/10.1111/ijcp.12646>
- [29] G. Bouleux, M. Biard, G. Soler, J.-S. Casalegno, A. Cantais, E. Marcon, and O. Mory, “Relation between paediatric bronchiolitis and elderly cardio-respiratory distress by a time series analysis of emergency admissions: a way to crowding anticipation,” manuscript submitted for publication.
- [30] E. de surveillance de la grippe, “Surveillance de la grippe en France, saison 2016-2017,” *Bulletin Epidemiologique Hebdomadaire*, vol. 22, pp. 466–75, 2017.