

Smooth nonnegative tensor factorization for multi-sites electrical load monitoring

Amaury Durand^{*†}, François Roueff^{*}, Jean-Marc Jicquel[†], and Nicolas Paul[‡]

^{*} LTCI, Telecom Paris, Institut Polytechnique de Paris. 19 Place Marguerite Perey, 91120 Palaiseau, France.

[†] EDF R&D, TREE, E36, Lab Les Renardières. Ecuelles, 77818 Moret sur Loing, France.

[‡]EDF R&D, PRISME. 6 quai Watier, 78400 Chatou, France.

Abstract—The analysis of load curves collected from smart meters is a key step for many energy management tasks ranging from consumption forecasting to customers characterization and load monitoring. In this contribution, we propose a model based on a functional formulation of nonnegative tensor factorization and derive updates for the corresponding optimization problem. We show on the concrete example of multi-sites load curves disaggregation how this formulation is helpful for 1) exhibiting smooth intra-day consumption patterns and 2) taking into account external variables such as the outside temperature. The benefits are demonstrated on simulated and real data by exhibiting a meaningful clustering of the observed sites based on the obtained decomposition.

I. INTRODUCTION

Since the beginning of consumption data collection, the comparison and disaggregation of load curves have been two popular domains of study. While the former focuses on clustering rather than extracting meaningful features, the latter consists in decomposing a load curve as a sum of curves each representing a particular consumption pattern. In both cases, it is assumed that a panel of load curves $\{X_j(u) : u \in \mathcal{U}, 1 \leq j \leq J\}$ is observed, where \mathcal{U} is an interval of \mathbb{R} (e.g. $\mathcal{U} = [0, 24]$) and j is an index (e.g. the day, the customer, the sensor, etc). Numerous clustering methods have been applied to compare such curves (see [4, 11, 14] for an overview). On the other hand, load curves disaggregation is usually done in the context of Non Intrusive Load Monitoring (NILM) where the goal is to decompose the globally observed consumption into a sum of the consumption of several devices. Over the past decade, Blind source separation methods such as Nonnegative matrix factorization (NMF) have gained in popularity in the NILM community (see e.g. [8] and the references therein). The idea is to decompose the load curve of the r -th device as a *signature* curve $a_r(u) \geq 0$ which is modulated

across observations by an *activation* $b_{j,r} \geq 0$, that is $X_j(u) \approx \sum_{r=1}^R a_r(u) b_{j,r}$.

In a multi-site context, disaggregating the load curves can be a way to extract features to describe the load profiles and cluster the sites. Assume we observe a panel of daily load curves $\{X_{j,n}(u) : u \in [0, 24], 1 \leq j \leq J, 1 \leq n \leq N\}$ where u represents the intra-day time, j the observed day and n the site. Generalizing the NMF model of NILM gives the Nonnegative Tensor Factorization (NTF) model, $X_{j,n}(u) \approx \sum_{r=1}^R a_r(u) b_{j,r} c_{n,r}$, where $a_r(u) \geq 0$ is the *signature* which is modulated across days by the *day activation* $b_{j,r} \geq 0$ and across sites by the *site activation* $c_{n,r} \geq 0$. The signature and activations are referred to as the *factors*.

Tensor factorization models are very popular in chemometrics and psychometrics but their use for electrical load curves analysis is still recent [6, 7, 12]. Multi-sites load curves disaggregation has been proposed in [12] using PARAFAC, which is the same as NTF without positivity constraints (see [5, 9]). In particular, the authors use the site activations for clustering but do not give any interpretation of the factors nor the clusters. In this paper, we propose several modifications of the NTF model in order to orient the decomposition into more interpretable factors. We also use additional knowledge on the outside temperature and different consumption regimes. The model is presented in Section II and is validated on simulated and real data in Section III.

The notations used throughout this paper are the following. For all integer $N \geq 1$, $[N] = \{1, \dots, N\}$. For tensor operations, we use the notations of [5], that is $\|\cdot\|_F$ is the Froebenius norm, $[\cdot]_+$ the positive part, \otimes the Hadamard product, \otimes the Kronecker product, \odot the Khatri-Rao product and \circ the tensor product. Note that $\mathbf{A}^{\otimes 2} = \mathbf{A} \otimes \mathbf{A}$. The unfolding $\mathbf{X}_{(k)}$ of a tensor \mathbf{X} is defined in [9].

II. PROPOSED MODEL

The proposed model is based on NTF with the additional assumption that the consumption $X_{j,n}(u)$ only depends on the day of the year j through two variables : the daily temperature $T_{j,n}$ and the consumption regime $\epsilon_{j,n}$ (e.g. business and non-business days). This leads to the following model

$$X_{j,n}(u) \approx \sum_{r=1}^R a_r(u) b_r(T_{j,n}) c_{n,r}^{(\epsilon_{j,n})}, \quad (1)$$

where $a_r(u) \geq 0, b_r(t) \geq 0, c_{n,r}^{(\epsilon)} \geq 0$. We also assume that the functions a_r and b_r are smooth and that a_r is periodic with a period of 24 hours (since it represents intra-day behavior). By analogy to the NMF and NTF cases we call the functions b_r the *thermal activations*.

A. The optimization problem

Let us consider two grids $(u_1, \dots, u_I) \subset [0, 24]$ and $(t_1, \dots, t_K) \subset \mathbb{R}$ which contain all the observed intra-day times and temperatures. Let also E be the number of consumption regimes so that $\epsilon_{j,n} \in [E]$ for all j, n . Then the factors $a_r, b_r, c_r^{(\epsilon)}$, are estimated by solving

$$\begin{aligned} & \min \{F + P\} \\ & \text{such that for all } r, i, t, n, \epsilon, \\ & \begin{cases} a_r(u_i) \geq 0, b_r(t_k) \geq 0, c_{n,r}^{(\epsilon)} \geq 0 \\ \int_0^{24} a_r = \int_{t_1}^{t_K} b_r = 1, \\ a_r(0^+) = a_r(24^-) \\ a_r'(0^+) = a_r'(24^-) \\ a_r''(0^+) = a_r''(24^-), \end{cases} \end{aligned} \quad (2)$$

and

$$\begin{aligned} F &= \sum_{i,j,n} \left(X_{j,n}(u_i) - \sum_{r=1}^R a_r(u_i) b_r(T_{j,n}) c_{n,r}^{(\epsilon_{j,n})} \right)^2, \\ P &= \alpha \sum_{r=1}^R \int_0^{24} (a_r'')^2 + \beta \sum_{r=1}^R \int_{t_1}^{t_K} (b_r'')^2. \end{aligned}$$

The scaling constraints prevent the factors from diverging since the error F is not affected by multiplying one of the factors by a constant as soon as the other factors are scaled accordingly. These scaling constraints also make sure that we can compare the site activations.

Finally, the penalizations on the L^2 -norm of the second derivatives imply that the solutions of Problem (2) are necessarily smooth spline functions. Namely, for all $r \in [R]$, the function a_r must be a 24-periodic cubic spline and the function b_r must

be a natural cubic spline. Since spline functions are characterized by their sample points, we can reformulate the problem as a weighted NTF.

B. Formulation as a weighted NTF problem

Classical results on cubic splines imply that there exist $\mathbf{v}_1 \in \mathbb{R}^I$ and $\mathbf{Q}_1 \in \mathbb{R}^{I \times I}$ positive definite such that for any 24-period cubic spline a on $[0, 24]$, we have $\int_0^{24} a = \mathbf{v}_1^\top \mathbf{a}$ and $\int_0^{24} (a'')^2 = \mathbf{a}^\top \mathbf{Q}_1 \mathbf{a}$, with $\mathbf{a} = [a(u_1), \dots, a(u_I)]^\top$. Similarly there exist $\mathbf{v}_2 \in \mathbb{R}^K$ and $\mathbf{Q}_2 \in \mathbb{R}^{K \times K}$ positive definite such that for any cubic spline b on $[t_1, t_K]$, we have $\int_{t_1}^{t_K} b = \mathbf{v}_2^\top \mathbf{b}$, and $\int_{t_1}^{t_K} (b'')^2 = \mathbf{b}^\top \mathbf{Q}_2 \mathbf{b}$, with $\mathbf{b} = [b(t_1), \dots, b(t_K)]^\top$. We now define $\mathbf{A} \in \mathbb{R}_+^{I \times R}$, $\mathbf{B} \in \mathbb{R}_+^{K \times R}$ and $\mathbf{C} \in \mathbb{R}_+^{EN \times R}$ by $\mathbf{A}_{i,r} = a_r(u_i)$, $\mathbf{B}_{k,r} = b_r(t_k)$, $\mathbf{C}_{(\epsilon-1)N+n,r} = c_{n,r}^{(\epsilon)}$, and $\mathbf{W} \in \mathbb{R}^{I \times K \times EN}$, $\mathbf{X} \in \mathbb{R}^{I \times K \times EN}$ by $\mathbf{W}_{i,k,(\epsilon-1)N+n} = \left(\sum_{j=1}^J \mathbb{1}_{T_{j,n}=t_k} \mathbb{1}_{\epsilon_{j,n}=\epsilon} \right)^{1/2}$ and

$$\mathbf{X}_{i,k,(\epsilon-1)N+n} = \frac{\sum_{j=1}^J \mathbb{1}_{T_{j,n}=t_k} \mathbb{1}_{\epsilon_{j,n}=\epsilon} X_{j,n}(u_i)}{\mathbf{W}_{i,k,(\epsilon-1)N+n}^2},$$

with the convention that $0/0 = 0$. Then Problem (2) is equivalent to

$$\begin{aligned} & \min_{\mathbf{A} \geq 0, \mathbf{B} \geq 0, \mathbf{C} \geq 0} f_{\mathbf{W}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \\ & \text{such that for all } r, \mathbf{v}_1^\top \mathbf{a}_r = \mathbf{v}_2^\top \mathbf{b}_r = 1, \end{aligned} \quad (3)$$

where $\mathbf{a}_r, \mathbf{b}_r$ and \mathbf{c}_r are the r -th columns of \mathbf{A}, \mathbf{B} and \mathbf{C} respectively, and

$$f_{\mathbf{W}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = L_{\mathbf{W}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) + P(\mathbf{A}, \mathbf{B}),$$

with

$$L_{\mathbf{W}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \left\| \mathbf{W} \circledast \left(\mathbf{X} - \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \right) \right\|_F^2,$$

and

$$P(\mathbf{A}, \mathbf{B}) = \alpha \text{Tr}(\mathbf{A}^\top \mathbf{Q}_1 \mathbf{A}) + \beta \text{Tr}(\mathbf{B}^\top \mathbf{Q}_2 \mathbf{B}).$$

C. Fast HALS algorithm

We solve Problem (3) using a Fast HALS algorithm. The idea of HALS is to minimize $f_{\mathbf{W}}$ alternatively in the columns of $\mathbf{A}, \mathbf{B}, \mathbf{C}$. The update in \mathbf{a}_r is obtained by solving

$$\min_{\substack{\mathbf{a}_r \geq 0 \\ \mathbf{v}_1^\top \mathbf{a}_r = 1}} \left\| \mathbf{W} \circledast (\mathbf{X}^{(r)} - \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r) \right\|_F^2 + \alpha \mathbf{a}_r^\top \mathbf{Q}_1 \mathbf{a}_r,$$

where $\mathbf{X}^{(r)} = \mathbf{X} - \sum_{s \neq r} \mathbf{a}_s \circ \mathbf{b}_s \circ \mathbf{c}_s$. This leads to the following update steps

$$\mathbf{a}_r \leftarrow \mathbf{M}_r^{-1} \left(\mathbf{W}_{(1)} \circledast \mathbf{X}_{(1)}^{(r)} \right) (\mathbf{c}_r \otimes \mathbf{b}_r)$$

$$\mathbf{a}_r \leftarrow \frac{[\mathbf{a}_r]_+}{\mathbf{v}_1^\top [\mathbf{a}_r]_+}.$$

The Fast HALS algorithm uses a clever way to write $\mathbf{X}^{(r)}$ and \mathbf{M}_r to avoid repeating computations unnecessarily. The updates in \mathbf{b}_r and \mathbf{c}_r are obtained similarly.

III. EXPERIMENTAL RESULTS

We validate our model on two datasets. The first was extracted from energy demand data simulated for 1 year and with hourly rate by the Office of Energy Efficiency & Renewable Energy (EERE)¹. We took a total of 775 sites in California, Arizona, Nevada, Utah, Oregon, Idaho and Washington, gathering 5 different building types. The second dataset consists of energy demand collected by EDF from 108 supermarkets across France over a period of 1 year with a sample rate of 10 minutes. The average external temperature of each day is observed for each site², EERE's data have one consumption regime and EDF's data have two consumption regimes (closing and opening days). Because of space constraints, we will not discuss the detection of regimes. In order to compare sites of different sizes we scale the observed load curves by the average daily consumption.

For both NTF and our model, we use the Fast HALS updates and the algorithm is stopped when the relative improvement of the loss reaches 10^{-5} . Each factors was initialized by taking the positive part of the singular vectors of the corresponding unfolding of \mathbf{X} (e.g. $\mathbf{a}_r = [\mathbf{u}_r]_+$ where \mathbf{u}_r is the r -th singular vector of $\mathbf{X}_{(1)}$). Since there is no foolproof method to select the number of components R (see [1, 3, 13] for ad-hoc methods) and since cross-validation is not straightforward with tensor data (see [2, 10]), we take $R = 6$ and $\alpha = \beta = 3000$ for EERE's data and $\alpha = \beta = 50$ for EDF's data. This choice gives a good balance between goodness of fit and interpretation of the factors. For clustering, we run K -means on the sites activations for all regimes. This means that site n is represented by the feature vector $(c_{n,1}^{(1)}, \dots, c_{n,R}^{(1)}, \dots, c_{n,1}^{(E)}, \dots, c_{n,R}^{(E)})$ where we recall that E is the number of regimes.

A. Results on EERE's data

The factors obtained by NTF and our model are represented in Figure 1 where the colors correspond to the building types. The dependence on the temperature is justified by the fact that the

day activations of NTF mainly indicate season changes. Moreover, the advantage of smoothing the signatures is that only the most important peaks are kept, which is valuable for interpretation. The radar plots of the site activations presented in Figure 2 show a better separation of the building types with our model (especially between the restaurants). To quantify this observation, we ran K-means with 5 clusters and compared the clusters with the true labels using the adjusted random index. Our model gives a perfect fit with an adjusted random index of 1 compared to 0.75 for NTF which mostly fails to separate the two types of restaurants. In our model, both hotels have a high site activation in Component 3 whose signature is typical of a hotel (high for breakfast and dinner and medium for lunch). Small Hotels tend to heat at night (Component 4) while Large Hotels tend to heat during the day (Component 1). Apartments have a high site activation in Component 4 whose signature and thermal activation are typical of heating in residential buildings. Components 5 and 6 also characterize the Apartments and can be interpreted as holidays. Indeed, the thermal activation of Component 5 is high when it is very cold and very hot (winter and summer holidays) where people are more at home in the middle of the day (where the signature is high) while the thermal activation of Component 6 peaks for medium temperatures and its signature presents a typical working day profile. Finally, the signature of Component 2, which peaks before lunch and dinner (and a bit before breakfast), is characteristic of restaurants. The two types of restaurants differ by the importance of lunch (more important for Quick Service Restaurants) and the amount of heating used (Component 1).

B. Results on EDF's data

The results obtained for EDF's data are presented in Figure 3 where the colors correspond to the clusters obtained by K-means. The number of clusters was selected by the silhouette index (between 2 and 9 clusters). The two consumption regimes can be seen in the day activations of NTF which justifies taking them into account in our model. The site activations are not very diverse in NTF and the two clusters obtained mainly differ from each other by Components 1 and 5 which respectively represent the standard profiles of opening and closing days. On the contrary, our model seems to extract more variable site activations (see also Figure 4) thus exhibiting more clusters. The interpretation of the components is also easier because of smoothness and of the dependence on tem-

¹Available at <https://openecr.org/doi-10.21203/rs.3.rs-1000000/v1>

²EERE's data can be obtained from the TMY3 weather stations using the `eewweather` Python package.

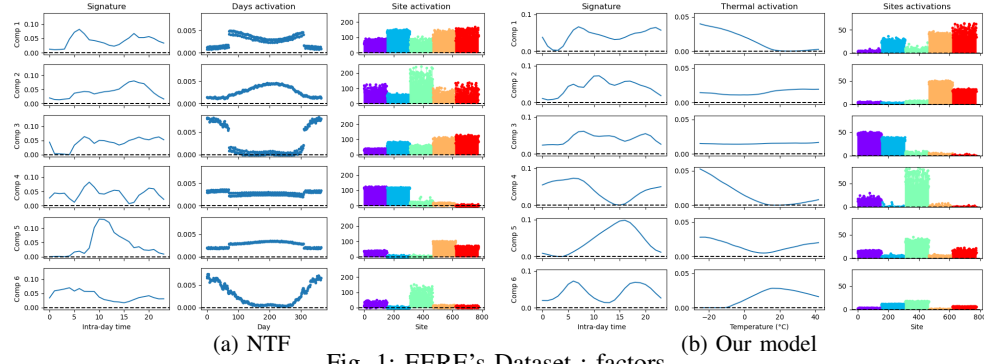


Fig. 1: EERE's Dataset : factors

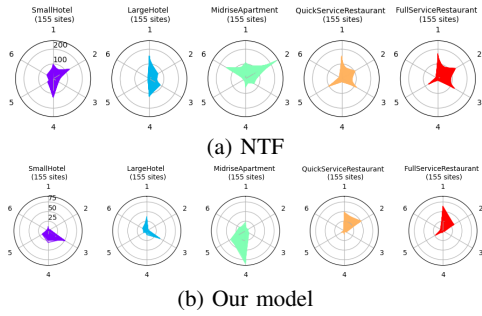


Fig. 2: EERE's Dataset : site activations

perature. For example, the first and last components are independent of the temperature and represent some constant behavior. A high site activation in the first component implies small difference between opening and closing hours (and therefore is high for regime 1 which represents closing days) and a high site activation in Component 6 implies big difference between opening and closing hours. The other components represent various heating (Components 2 and 3) and cooling (Components 4 and 5) profiles. The signature of Component 2 is characteristic of heater since it peaks in the morning and then slowly decays because of inertia. The signature and thermal activation of Component 5 indicate that it represents air conditioning which is activated for high temperatures and in the middle of the day. Finally, the thermal activation of Component 4 indicates that it represents food refrigeration which is activated for lower temperatures than air conditioning and reaches a saturation level for high temperatures. The clusters can be interpreted as follows. Cluster 1 : Small variation between opening and closing hours (Component 1) and small impact of air conditioning (Components 5). Cluster 2 : Use a lot of air conditioning (Components 5) and medium heating profile (Component 3). Cluster 3 : High heating (Component 2) and food refrigeration (Component 4). Cluster 4 : Not much impacted by

the temperature but the difference between opening and closing hours is high (Components 6).

Giving a physical meaning to the factors and clusters is highly valuable for monitoring or maintenance purpose. For example, sites in Cluster 2 should reduce their use of air conditioning while for sites in Cluster 3, reducing heating is a priority.

IV. CONCLUSION

We proposed a model based on a functional formulation of the nonnegative tensor factorization model and an associated optimization algorithm for the disaggregation of multi-sites load curves. By taking into account additional information such as the outside temperature and smoothness of the factors, we showed that this model exhibits more meaningful features and clusters than previously used NTF models.

REFERENCES

- [1] R. Bro and H. A. L. Kiers. A new efficient method for determining the number of components in parafac models. *Journal of Chemometrics*, 17(5):274–286, 2003. doi: <https://doi.org/10.1002/cem.801>.
- [2] R. Bro, K. Kjeldahl, A. Smilde, and H. Kiers. Cross-validation of component models: A critical look at current methods. *Analytical and Bioanalytical Chemistry*, 390(5):1241–1251, Mar. 2008. ISSN 1618-2642. doi: 10.1007/s00216-007-1790-1.
- [3] E. Ceulemans and H. A. L. Kiers. Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology*, 59(1):133–150, 2006. doi: <https://doi.org/10.1348/000711005X64817>.
- [4] G. Chicco. Overview and performance assessment of the clustering methods for electrical

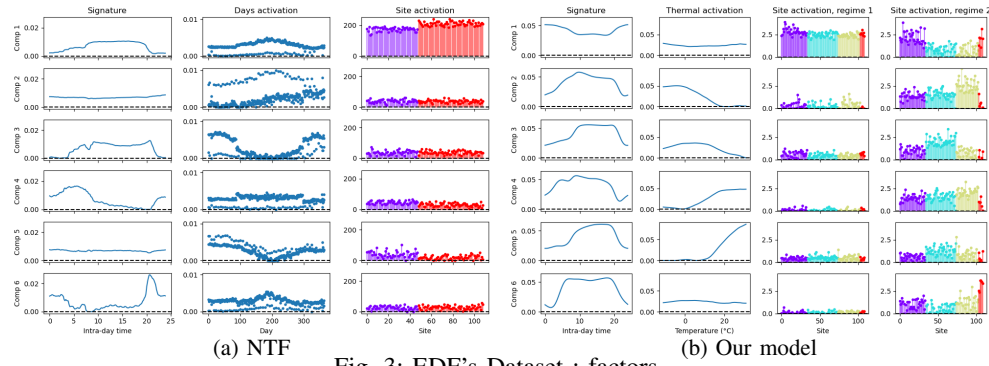


Fig. 3: EDF's Dataset : factors

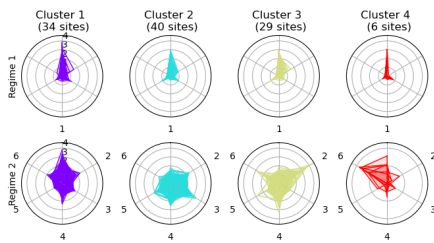


Fig. 4: EDF's Dataset : site activations of our model

- load pattern grouping. *Energy*, 42(1):68 – 80, 2012. ISSN 0360-5442. doi: <https://doi.org/10.1016/j.energy.2011.12.031>. 8th World Energy System Conference, WESC 2010.
- [5] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. Wiley Publishing, 2009. ISBN 0470746661.
- [6] M. Figueiredo, B. Ribeiro, and A. Almeida. Exploring the performance of non-negative multi-way factorization for household electrical seasonal consumption disaggregation. *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 899–906, 2014.
- [7] M. Figueiredo, B. Ribeiro, and A. Almeida. Analysis of trends in seasonal electrical energy consumption via non-negative tensor factorization. *Neurocomput.*, 170(C):318–327, Dec. 2015. ISSN 0925-2312. doi: [10.1016/j.neucom.2015.03.088](https://doi.org/10.1016/j.neucom.2015.03.088).
- [8] S. Henriët, U. Şimşekli, S. Dos Santos, B. Fuentes, and G. Richard. Independent-variation matrix factorization with application to energy disaggregation. *IEEE Signal Processing Letters*, 26(11):1643–1647, 2019. doi: [10.1109/LSP.2019.2941428](https://doi.org/10.1109/LSP.2019.2941428).
- [9] T. G. Kolda and B. W. Bader. Tensor de-

compositions and applications. *SIAM Review*, 51(3):455–500, September 2009. doi: [10.1137/07070111X](https://doi.org/10.1137/07070111X).

- [10] A. B. Owen and P. O. Perry. Bi-cross-validation of the SVD and the nonnegative matrix factorization. *The Annals of Applied Statistics*, 3(2):564 – 594, 2009. doi: [10.1214/08-AOAS227](https://doi.org/10.1214/08-AOAS227).
- [11] A. Rajabi, L. Li, J. Zhang, J. Zhu, S. Ghavidel, and M. J. Ghadi. A review on clustering of residential electricity customers and its applications. In *2017 20th International Conference on Electrical Machines and Systems (ICEMS)*, pages 1–6, 2017. doi: [10.1109/ICEMS.2017.8056062](https://doi.org/10.1109/ICEMS.2017.8056062).
- [12] B. Sandoval, E. Barocio, P. Korba, and F. R. S. Sevilla. Three-way unsupervised data mining for power system applications based on tensor decomposition. *Electric Power Systems Research*, 187:106431, 2020. ISSN 0378-7796. doi: <https://doi.org/10.1016/j.epsr.2020.106431>.
- [13] M. E. Timmerman and H. A. L. Kiers. Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology*, 53 (1):1–16, 2000. doi: <https://doi.org/10.1348/000711000159132>.
- [14] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, and X. Zhao. A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, 82:1027–1047, 2018. ISSN 1364-0321. doi: <https://doi.org/10.1016/j.rser.2017.09.108>.