# Subgradient Methods for Averaging Household Load Profiles under Local Permutations

Marcus Voß, Brijnesh Jain, and Sahin Albayrak
Technische Universität Berlin (DAI-Labor)
marcus.voss@dai-labor.de

***Abstract*—The sample mean is one of the most fundamental concepts in statistics with far-reaching implications for data mining and pattern recognition. Household load profiles are compared to the aggregated levels more intermittent and a specific error measure based on local permutations has been proposed to cope with this when comparing profiles. We formally describe a distance based on this error, the local permutation invariant (LPI) distance, and introduce the sample mean problem in the LPI space. An existing exact solution has exponential complexity and is only tractable for very few profiles. We propose three subgradient-based approximation algorithms and compare them empirically on 100 households of the CER dataset. We find that stochastic subgradient descent can approximate the mean best, while the majorize-minimize mean is a good compromise for applications as no hyperparameter-tuning is needed. We show how the algorithms can be used in forecasting and clustering to achieve more appropriate results than by using the arithmetic mean.**

***Index Terms*—Approximation methods, Distance measurement, Load modeling**



## I. Introduction

The smart meter roll-out in many countries leads to energy providers, energy value-added service providers, and grid operators having access to increasing amounts of household-level load profiles. These load profiles will be analyzed within diverse data mining tasks such as classification, clustering, and forecasting (see [1], [2] for a review of use cases and methods). In general, the notion of a *distance* and the notion of a *sample mean* are within the most elemental concepts in statistics, underlying many data mining algorithms, e.g. k-means clustering, time series forecasting, or classification. A sample mean is defined as the minimizer of the sum of squared distances in terms of a specific metric space [3]. In general, the most commonly used distance is the Euclidean distance (also known as the L2-norm), due to its simplicity and well-understood analytical properties. The arithmetic mean is the sample mean in terms of the Euclidean distance.

Individual household load profiles are intermittent and have less structure to be exploited by data mining algorithms [4], [5]. If point-wise distances, such as the Euclidean distance, are used to compare household load profiles they inflict a double-penalty if a peak has about the correct height, but is shifted in time. Haben et al. [6] introduce the *adjusted p-norm error* as a measure to assess household load forecasts. It copes with intermittent load profiles by allowing locally restricted permutations of the profile. Charlton et al. [7] introduce the *permutation merge* algorithm to find a sample mean using the adjusted p-norm as a distance. The algorithm finds an exact solution in exponential runtime [7], thus only means of few profiles with small displacement are tractable. Even if tractable, they are too computationally expensive for application on many profiles or limited home energy management hardware. Therefore, we explore approximations of a sample mean for household load profiles.

Our contributions are as follows: First, we formally describe a distance based on the adjusted p-norm error, the local permutation invariant (LPI) distance and then formally describe and characterize the sample mean problem in the LPI space (Section II). We propose three subgradient-based algorithms for approximating the LPI sample mean: the subgradient mean (SGM) and its special case the majorize-minimize mean (MMM), as well as the stochastic subgradient descent mean (SSG), a stochastic version of SGM (Section III). We empirically compare the algorithms (Section IV) and evaluate them within two data mining applications (forecasting and clustering).

## II. Sample Means in LPI-Spaces

We formally define the LPI-space and its sample mean.

### A. Related Work

The sample mean of a finite sample of numbers minimizes the sum of squared errors (variance). Fréchet [3] generalized the concept of a sample mean to any metric space as the minimizer of the sum of squared distances. Since then, the sample mean of various distance spaces has been studied, e.g. shape spaces and Riemannian manifolds [8]–[10], tree-structures [11], graphs [12], [13], cluster ensembles [14], sequences [15], [16], and time series [17]–[19].

Computing a sample mean in arbitrary distance spaces is often intractable. Solving the sample mean problem of sequences under the edit distance and of time series under the dynamic time warping distance (DTW) is NP-hard [16], [20]. Hence, approaches resort to approximate solutions. For example, subgradient mean algorithms have been studied for graphs under optimal alignments [12], [21] and time series under DTW [19]. The majorize-minimize mean has been proposed for graphs [12], consensus clustering [22], and time series under DTW [18].

### B. The LPI-Distance

The LPI-distance is a dissimilarity measure on load profiles that accounts for local permutations in temporal order and was introduced for day-ahead forecasting [6]. To describe the LPI-distance, we assume that $\mathcal{P}_n$ be the set of $n \times n$ permutation matrices. Consider the subset

$$\mathcal{L}_n^\omega = \{P = (p_{ij}) \in \mathcal{P}_n \,:\, p_{ij} = 1 \Rightarrow |i-j| \leq \omega\},$$

of $n \times n$ permutation matrices whose non-zero entries are confined to a diagonal band (adjustment limit) of width $\omega > 0$. Then the function $\delta \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$\delta(x,y) = \min\left\{\|Px - y\| \,:\, P \in \mathcal{L}_n^\omega\right\}$$

is a *local permutation invariant* (LPI) *distance* induced by the Euclidean norm $\|\cdot\|$.



We are particularly interested in permutation matrices that minimize the Euclidean norm. We say, $P \in \mathcal{L}_n^\omega$ is an *optimal alignment* of $x$ and $y$ if

$$\|Px - y\| \leq \|P'x - y\|$$

for all $P' \in \mathcal{L}_n^\omega$. In this case, $Px$ is optimally aligned to $y$. Note that optimal alignments are not unique, in general. Computing the LPI-distance and finding an optimal alignment can be cast to an assignment problem for which efficient solutions exists [6].

Figure 1 (top) depicts two load profiles $x$ and $y$ in 30 minute resolution over one day together with an optimally aligned profile $Px$ at bandwidth $\omega = 3$. The morning peaks of both load profiles occur at different times resulting in a relatively large Euclidean distance $\|x - y\| = 4.0$. In contrast, the LPI-distance $\delta(x,y) = \|Px - y\| = 2.2$. is substantially lower, because the optimal alignment $P$ adjusts the temporal variations of both peaks.

### C. Sample Means

We introduce sample means in LPI-spaces $(\mathbb{R}^n, \delta)$, where $\delta$ is an LPI-distance on $\mathbb{R}^n$. For this, we assume

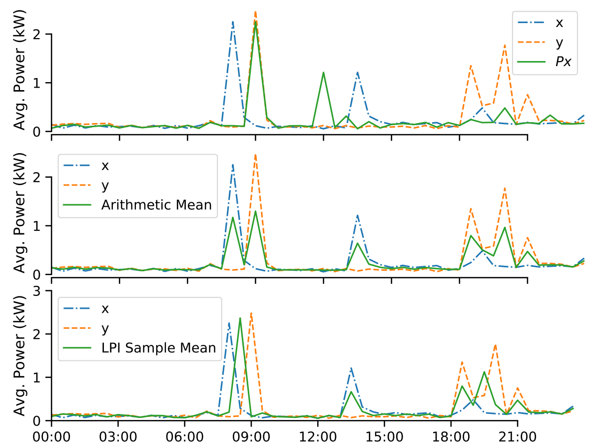


Figure 1. Load profiles $x$ and $y$ over one day and an optimally aligned profile $Px$ (top), arithmetic mean of $x$ and $y$ (middle), and sample mean of $x$ and $y$ in the LPI space (bottom).

that $\mathcal{X} = \{x_1, \ldots, x_k\}$ is a sample of $k$ load profiles $x_i \in \mathbb{R}^n$. The *Fréchet function* of $\mathcal{X}$ is defined by

$$F \colon \mathbb{R}^n \to \mathbb{R}, \quad x \mapsto \frac{1}{k}\sum_{i=1}^k \delta(x_i, x)^2\,.$$

A *sample mean* of $\mathcal{X}$ is any global minimum of $F(x)$. The *sample mean problem* of $\mathcal{X}$ is the problem of finding a global minimizer of the corresponding Fréchet function.

We present an equivalent but analytically more convenient formulation of the Fréchet function. Using the definition of $\delta(x_i, x)$ gives

$$F(x) = \frac{1}{k}\sum_{i=1}^k \delta(x_i, x) = \frac{1}{k}\sum_{i=1}^k \min_{P_i \in \mathcal{L}_n^\omega} \|P_i x_i - x\|^2\,.$$

Interchanging summation and minimization yields

$$F(x) = \min_{P_1,\ldots,P_k \in \mathcal{L}_n^\omega} \frac{1}{k}\sum_{i=1}^k \|P_i x_i - x\|^2\,.$$

Let $\mathcal{C} = \mathcal{C}_{n,k}^\omega$ be the set of $k$-tuples $C = (P_1, \ldots, P_k)$ with elements $P_i \in \mathcal{L}_n^\omega$. Then the Fréchet function $F$ can be equivalently expressed as a pointwise minimum

$$F(x) = \min_{C \in \mathcal{C}} F_C(x), \qquad (1)$$

of finitely many component functions

$$F_C(x) = \frac{1}{k}\sum_{i=1}^k \|P_i x_i - x\|^2,$$

where $C = (P_1, \ldots, P_k)$. A $k$-tuple $C \in \mathcal{C}$ is *optimal* at $x \in \mathbb{R}^n$ if $F_C(x) = F(x)$. In this case, we say that

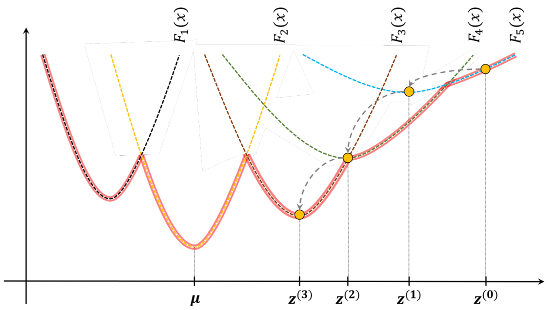


Figure 2. Simplified depiction of the Fréchet function $F$ (bold red curve) as a pointwise minimum of five differentiable convex components $F_1, \dots, F_5$ (colored dotted curves). The $x$-axis gives a one-dimensional representation of $\mathbb{R}^n$ and the $y$-axis shows the function values. The function $F$ attains its global minimum at the sample mean $\mu$. The solutions $z^{(t)}$ illustrate the MMM algorithm.

**Algorithm 1** Subgradient Mean Algorithm

```
1: procedure SGM(𝒳)
2:     initialize solution z ∈ ℝ^n
3:     z_* ← z
4:     repeat
5:         for i ← 1 to k do
6:             P_i ← optimal alignment of x_i and z
7:         end for
8:         z ← z + η ∑_{i=1}^k P_i x_i − z
9:         z_* ← argmin {F(z_*), F(z))}
10:        adjust step size η
11:    until termination
12:    return z
13: end procedure
```

Lines 3 and 9 keep track of the best solution $z_*$ found so far. Lines 5–7 construct an optimal $C = (P_1, \dots, P_k)$ such that $F_C(z) = F(z)$ and the $P_i$ are optimal alignments of $x_i$ and $z$.

$F_C$ is *active* at $x$. If $C = (P_1, \dots, P_k) \in \mathcal{C}$ is optimal at $x \in \mathbb{R}^n$, then the $P_i$ are optimal alignments of $x_i$ and $z$.

The components $F_C$ of $F$ are continuously differentiable and convex. Due to the min-operation, the Fréchet function is neither differentiable nor convex. Figure 2 schematically depicts the form of a Fréchet function.

We conclude this section with an intuitive justification of the sample mean. One core task in statistics is to summarize the data by a single representative that captures their characteristic features. A characteristic feature of the load profiles $x$ and $y$ from Figure 1 is a high morning peak. The arithmetic mean fails to capture this characteristic feature by introducing a second morning peak and flattening both peaks. In contrast, the sample mean correctly captures the characteristic feature of a single high morning peak.

## III. Sample Mean Algorithms

We propose three subgradient-based methods for approximately solving the sample mean problem.

### A. Subgradient Mean Algorithm

The subgradient mean (SGM) algorithm generalizes gradient descent methods to certain classes of non-differentiable functions such as the Fréchet function. The subgradient update rule is of the form

$$z' \leftarrow z - \eta \cdot \nabla F_C(z),$$

where $z$ is the current and $z'$ the updated solution, $\eta > 0$ is the step size, and $F_C$ is an active component of $F$ at $z$. The subgradient $\nabla F_C(z)$ is of the form

$$\nabla F_C(z) = -\frac{2}{k} \sum_{i=1}^{k} (P_i x_i - z),$$

where the $P_i$ are optimal alignments of $x_i$ and $z$. Thus, the subgradient update rule can be rewritten as

$$z' \leftarrow z + \eta' \cdot \sum_{i=1}^{k} (P_i x_i - z), \tag{2}$$

where the step size $\eta' = 2\eta/k$ absorbs the factor $2/k$. Algorithm 1 outlines the SGM method.

### B. The Majorize-Minimize Mean Algorithm



The majorize-minimize mean (MMM) algorithm is a subgradient method with constant step-size $\eta = 1/2$. Equation (2) together with $\eta' = 2\eta/k = 1/k$ gives the update rule

$$z' \leftarrow \frac{1}{k} \sum_{i=1}^{k} P_i x_i. \tag{3}$$

where $C = (P_1, \dots, P_k)$ is optimal at $z$. This update rule gives rise to an alternative interpretation and additional insight into the MMM algorithm (see Fig. 2). Observe that the MMM algorithm repeatedly alternates between majorization and minimization until convergence after finitely many iteration:

*Majorization*: Minimizing $F(x)$ is difficult. The majorization step constructs a surrogate function $F' : \mathbb{R}^n \to \mathbb{R}$ at the current solution $z$ with the following properties: (i) $F'(x) \geq F(x)$ for all $x \in \mathbb{R}^n$, (ii) $F'(z) = F(z)$, and (iii) $F'$ is simpler to minimize than $F$.

A function with properties (i) and (ii) majorizes $F$ at $z$. Lines 5 – 7 of Algorithm 1 construct an optimal $k$-tuple $C = (P_1, \dots, P_k)$ such that $F_C$ is active and majorizes $F$ at $z$.

**Algorithm 2** Stochastic Subgradient Mean Algorithm

```
1: procedure SSG(𝒳, η)
2:     initialize solution z ∈ ℝ^n
3:     z_* ← z
4:     repeat
5:         shuffle 𝒳
6:         for i ← 1 to k do
7:             P_i ← optimal alignment of x_i and z
8:             z ← z + η (P_i x_i − z)
9:         end for
10:        z_* ← argmin {F(z_*), F(z))}
11:        adjust step size η
12:    until termination
13:    return z
14: end procedure
```

Lines 3 and 10 keep track of the best solution $z_*$ found so far. Line 5 shuffles the sample $\mathcal{X}$ to randomize the order of its elements.

*Minimization*: The majorizing function $F_C(x)$ of the previous step has a unique minimum at

$$z' = \frac{1}{k}\sum_{i=1}^{k} P_i x_i.$$

Update rule (3) replaces the current solution $z$ with the recomputed solution $z'$.

The MMM algorithm is a descent method that converges after a finite number of updates. In every iteration, we have

$$F(z') \leq F_C(z') \leq F_C(z) = F(z)$$

where $z'$ is the new and $z$ is the previous solution. Once $F(z') = F(z)$ holds, we can terminate the algorithm, because further improvements are not possible. Thus, the MMM method strictly decreases $F$ in every iteration until it converges after finitely many iterations (because the set $\mathcal{C}$ is finite).

### C. The Stochastic Subgradient Mean Algorithm

The stochastic subgradient mean (SSG) algorithm simplifies the subgradient algorithm. Instead of computing the subgradient $\nabla F_C$ of an active component, each iteration approximates $\nabla F_C$ on the basis of a single randomly selected example $x \in \mathcal{X}$:

$$z \leftarrow z + \eta(Px - z),$$

where $P \in \mathcal{L}_n$ is an optimal alignment of $x$ and $z$. Algorithm 2 outlines the SSG method.

### D. Algorithmic Design Decisions

The performance of the algorithms introduced in the former section depends on the initial solution (SGM, MMM, SGG), the choice of step size rule $\eta$ (SGM, SGG), and the termination criterion (SGM, MMM, SGG). As initial solution, we suggest to compute the medoid of $\mathcal{X}$.A medoid is an example $x_* \in \mathcal{X}$ such that $F(x_*) \leq F(x)$ for all $x \in \mathcal{X}$. The difference between a medoid $x_*$ and a sample mean $\mu$ is that $x_*$ minimizes $F$ over the sample $\mathcal{X}$, whereas $\mu$ minimizes $F$ over the entire space $\mathbb{R}^n$. However, the medoid calculation needs $N^2$ LPI-distance calculations, hence for larger samples, a random instance or the medoid of a sub-sample may be more suitable.

To improve the performance of SGM and SSG, we can apply more sophisticated update rules, for example Momentum, Adagrad, Adadelta, and Adam. For an overview, we refer to [23].

## IV. Empirical Evaluation

The goal of the experiments is to assess the performance and usability of the proposed subgradient mean algorithms. For this, we used the dataset of the CER Smart Metering Project [24].[1]

### A. Experimental Setup

We initialized MMM, SGM, and SSG with the medoid profile (see Section III-D). We used Adam [25] in SSG with a linear decay of the step size from an initial to a final value. For each run we chose the step size determined through grid search of $\eta \in [0.01, 0.025, 0.05, 0.075, 0.1]$.



### B. Sample Mean Benchmark

In this experiment we compared the performance of the proposed subgradient methods SGM, MMM, and SSG against the arithmetic mean, the medoid, and the permutation merge (PM) [7]. We set the bandwidth of the LPI-distance to $\omega = 3$. We constructed samples of varying size $N$ consisting of profiles from the 2010 CER-data with a 30-minute resolution of 100 households. Every sample includes $N$ profiles of the same weekday from the same household over different periods of time. We considered a month ($N = 4$), three months ($N = 12$), half a year ($N = 24$) and almost the full year ($N = 48$). Then we applied all mean algorithms to all samples.

Figure 3 summarizes the Fréchet variations of all mean algorithms vs. the sample size $N$. As an exact algorithm, PM exhibited the best performance but was computationally tractable only for samples of size $N = 4$. This observation justifies the construction of approximate algorithms. For $N = 4$, the subgradient methods deviated $5.1\%$ (SSG), $6.5\%$ (SGM), and $7.8\%$ (MMM) from the exact solution. This result suggests that there is

[1] Script reproduce data: https://github.com/marcus-voss/cer-to-hdf5

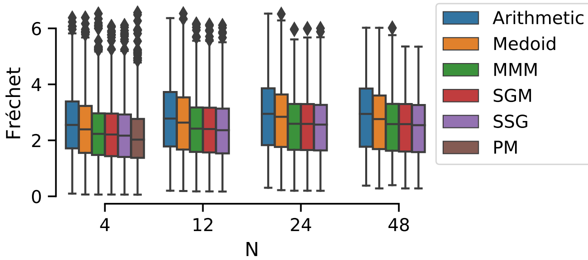


Figure 3. Comparison of different sample mean algorithms.

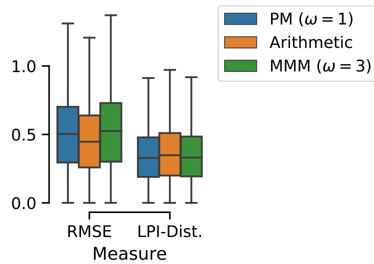


Figure 4. Comparison of sample means in permutation merge forecaster.

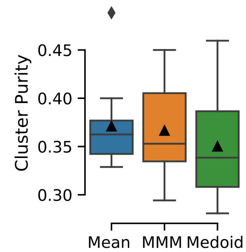


Figure 5. Comparision of sample means in k-means.

room for further improvements. For all sample sizes $N$, the three subgradient methods returned substantially lower Fréchet variations on average than the arithmetic mean and medoid. Furthermore, among the subgradient methods, SSG performed best and MMM worst, but only by a small margin. Figure 6 exemplarily contrasts traces of the sample means obtained by PM, MMM, SSG, and the arithmetic mean with the original profiles (left). The plot shows that the arithmetic mean smoothes peaks out, whereas LPI-based means better preserve sharp peaks.

### *C. Day-Ahead Forecasting*

In this experiment, we assessed the performance of the sample mean in day-ahead forecasting.

We used the MMM as a parameter-free approximation within the permutation merge forecaster [7] and compared it against the PM method with bandwidth $\omega = 1$ (as in [7]) and the simple seasonal moving average with the arithmetic mean. A weekly seasonality with an average of $N = 6$ profiles works well, with diminishing returns when including more [7]. We note that for $N = 6$ and the bandwidth $\omega = 3$, PM is already intractable. As discussed in [6], minimizing the LPI-distance for $\omega = 3$ is a reasonable goal for many applications. Hence, we configured MMM with $\omega = 3$.

Figure 4 summarizes the results of the three forecasters in terms of the RMSE and mean LPI-distance with $\omega = 3$. It shows that—as expected— the arithmetic mean better minimizes the RMSE but PM and MMM are better in terms of the LPI-distance. While on average PM and MMM perform similarly, we see in Figure 6 that the MMM better models the load pattern, as PM with $\omega = 1$ results in an oscillating pattern (middle).

### *D. K-Means Clustering*

In our last two experiments, we compared different k-means clustering schemes: (i) regular k-means based on the Euclidean distance (mean), (ii) k-medoid in LPI-spaces, and (iii) k-means in LPI-spaces again using parameter-free MMM for mean computation.

In the first experiment, the samples are all daily load profiles per household. We set the bandwidth of the LPI-distance to $\omega = 3$ and the number of clusters to $k = 2$ assuming that two inherent clusters exist (weekday and weekend). We assessed the quality of a clustering by measuring its *cluster purity*. The cluster purity is computed as follows: For every cluster $c$, count the number $n_c$ of profiles from the most common class. Then the cluster purity is the ratio of the sum of the numbers $n_c$ over all clusters $c$ and the total number of profiles. While on average we find only a slight improvement, 72.4% for using MMM-based k-means compared to 72.1% for regular k-means and 72% for k-medoids, we see in Figure 6 the resulting centroids for an example household, showing that the mean flattens out the peaks (right).

In a second experiment, we built 10 groups of 10 households each and applied clustering with $k = 10$ to cluster all profiles of the same weekday, assuming that the clustering can separate the households. Again we computed the cluster purity. Figure 5 shows the distribution of cluster purities of the three k-means schemes, showing that results differ across the groups, while the median (line) for MMM-based k-means is lower than the arithmetic mean, the mean is similar (triangle) and the upper quartile is higher.

## V. Conclusion

The arithmetic mean fails to cope with temporal variation inherent in household load profiles. As an alternative, the concept of sample mean in LPI-spaces has been proposed. Solving the sample mean problem is computationally excessive and existing exact solutions can only be applied to the smallest sample sizes and bandwidths. To alleviate this problem, we proposed three related subgradient methods for averaging household load profile in LPI-spaces. The subgradient methods are fast heuristics that return approximate solutions within an acceptable period of time. Empirical results show that SSG and SGM exhibited better performance than the

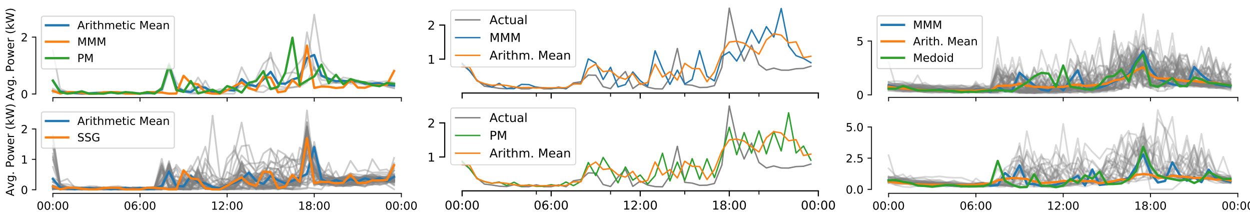


Figure 6. Example traces of means and the averaged profiles (left, top N=4 and bottom N=48) and of forecasts with arithmetic mean and actuals (middle), as well as cendroids after k-means clustering (right, top weekday and bottom weekend).

MMM algorithm. However, as opposed to SSG and SGM, MMM requires no parameter tuning.

Detailed run-time comparisons where out of scope for this study, but would be an interesting future work so that their scaling properties can be analyzed on larger datasets. Then also algorithmic adjustments are necessary, e.g. avoiding unnecessary calculation of the Fréchet function in each iteration or the reuse of optimal alignments. As shown in the forecasting application the adjusted p-norm does not capture all properties of similarity of load profiles well (the oscillation for $\omega = 1$), hence extensions and alternatives based the earth mover distance have been proposed [26]. The proposed algorithms of this work could also be applied for these distances if a subgradient can be obtained. Further research is needed to determine what distance, or what configuration of a distance, is most suitable for a specific application.

## Acknowledgment

This work was partially funded by the German government under funding ref. number 03SIN539 (WindNODE) and DFG Sachbeihilfe JA 2109/4-2.



## References

[1] B. Yildiz, J. Bilbao, J. Dore, and A. Sproul, "Recent advances in the analysis of residential electricity consumption and applications of smart meter data," *Applied Energy*, 2017.

[2] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, 2018.

[3] M. Fréchet, "Les éléments aléatoires de nature quelconque dans un espace distancié," *Ann. Inst. H. Poincaré*, vol. 10, 1948.

[4] C. Wu, W. Tang, K. Poolla, and R. Rajagopal, "Predictability, constancy and contingency in electric load profiles," in *IEEE SmartGridComm*. IEEE, 2016.

[5] R. Sevlian and R. Rajagopal, "A scaling law for short term load forecasting on varying levels of aggregation," *International Journal of Electrical Power & Energy Systems*, vol. 98, 2018.

[6] S. Haben, J. Ward, D. V. Greetham, C. Singleton, and P. Grindrod, "A new error measure for forecasts of household-level, high resolution electrical energy consumption," *International Journal of Forecasting*, vol. 30, no. 2, pp. 246–256, 2014.

[7] N. Charlton, D. V. Greetham, and C. Singleton, "Graph-based algorithms for comparison and prediction of household-level energy use profiles," in *IEEE International Workshop on Intelligent Energy Systems (IWIES)*. IEEE, 2013.

[8] R. Bhattacharya, V. Patrangenaru *et al.*, "Large sample theory of intrinsic and extrinsic sample means on manifolds," *The Annals of Statistics*, vol. 31, no. 1, pp. 1–29, 2003.

[9] I. L. Dryden and K. V. Mardia, *Statistical shape analysis: with applications in R*. John Wiley & Sons, 2016.

[10] X. Pennec, "Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, p. 127, 2006.

[11] A. Feragen, P. Lo, M. de Bruijne, M. Nielsen, and F. Lauze, "Toward a theory of statistical tree-shape analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, 2013.

[12] B. J. Jain, "Statistical graph space analysis," *Pattern Recognition*, vol. 60, pp. 802–812, 2016.

[13] X. Jiang, A. Münger, and H. Bunke, "On median graphs: properties, algorithms, and applications," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 10, 2001.

[14] B. J. Jain, "The mean partition theorem in consensus clustering," *Pattern Recognition*, vol. 79, pp. 427–439, 2018.

[15] D. Gusfield, *Algorithms on strings, trees and sequences: computer science and computational biology*. Cambridge univ. press, 1997.

[16] L. Wang and T. Jiang, "On the complexity of multiple sequence alignment," *Journal of computational biology*, vol. 1, 1994.

[17] M. Cuturi and M. Blondel, "Soft-dtw: a differentiable loss function for time-series," *arXiv preprint arXiv:1703.01541*, 2017.

[18] V. Hautamaki, P. Nykanen, and P. Franti, "Time-series clustering by approximate prototypes," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[19] D. Schultz and B. Jain, "Nonsmooth analysis and subgradient methods for averaging in dynamic time warping spaces," *Pattern Recognition*, vol. 74, pp. 340–358, 2018.

[20] L. Bulteau, V. Froese, and R. Niedermeier, "Hardness of consensus problems for circular strings and time series averaging," *arXiv preprint arXiv:1804.02854*, 2018.

[21] B. J. Jain and K. Obermayer, "Algorithms for the sample mean of graphs," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2009, pp. 351–359.

[22] E. Dimitriadou, A. Weingessel, and K. Hornik, "A combination scheme for fuzzy clustering," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 07, 2002.

[23] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[24] Commission for Energy Regulation (CER), "CER smart metering project - electricity customer behaviour trial 2009-2010 [dataset]. 1st edition," 2012, www.ucd.ie/issda/CER-electricity.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2014.

[26] K. Abdulla, K. Steer, A. Wirth, and S. Halgamuge, "Improving the on-line control of energy storage via forecast error metric customization," *Journal of Energy Storage*, vol. 8, 2016.