# Identifying Metering Hierarchies with Distance Correlation and Dominance Constraints

Tak-Shing T. Chan
*Department of Mathematics and Statistics*
*Lancaster University*
Lancaster, U.K.
t.t.chan@lancaster.ac.uk

Alex Gibberd
*Department of Mathematics and Statistics*
*Lancaster University*
Lancaster, U.K.
a.gibberd@lancaster.ac.uk

***Abstract*—In this paper, we consider observations from a series of smart meters that are either completely or partially aggregated, and our aim is to estimate the metering hierarchy. We propose to estimate this important metadata through a novel adaptation of the Chow–Liu tree learning procedure. Our approach takes into account prior knowledge from a set of dominance conditions that are easily elicited from the consumption data. In addition to more traditional correlation-based approaches we also introduce a distance-correlation-based method for detecting edges. Synthetic experiments show the benefits of distance correlation and the dominance conditions in recovering tree structure. The paper concludes with a real-world application of the method to infer energy metering hierarchies in a library building.**

***Index Terms*—smart meter, metadata, spanning trees and arborescences, distance correlation, aggregation**

## I. Introduction

Modern buildings often incorporate multiple energy sensors which are typically laid out according to a hierarchy [1]. For instance, we might have a main meter at the "edge" of a building, and then meters on individual floors, or associated with risers. The positioning and relationships between these sensors form an important class of metadata which enables us to add context to observations of energy consumption associated with the meters. Figure 1 provides an example of how such metadata might be canonically recorded as a wiring diagram. We also observe, associated with each meter, a time series of energy consumption from these meters. Considerable work is required to encode the information contained in the wiring diagrams into a taxonomy, however, and in many cases the wiring diagram may not even be accessible to the practitioners. Our aim in this paper is to estimate the hierarchy of the meters without access to ground-truth knowledge on the wiring diagram—that is, we use the data itself to directly infer the aggregation tree.

One of the barriers to effective inference of metadata is the heterogeneity of the environments that one can face when implementing monitoring solutions. For example, a university campus is radically different from a steel works. To aid with this task, recent work [2], [3] has considered the construction of a unifying schema for building metadata. While standards are important to develop, in practice, energy managers, or the systems they commissioned may not store data alongside meaningful identifiers—identifiers that are understandable by a human. For example, in our case study, meters are associated with an ID "`MC065-L04/M13R45099`", instead of "`Library, Main-Riser 3`".

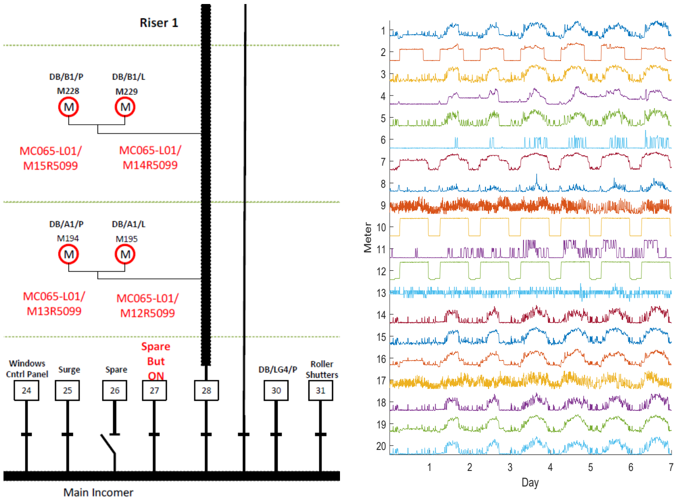

Fig. 1. Left: example of a wiring diagram for meter installation. Right: example of electrical consumption across meters within a hierarchy—can we infer the meter hierarchy from the consumption data directly? Under what situations can we expect to estimate this structure accurately?

Converting metadata into a structured form via a schema is known as *metadata normalization* [3]. Although this is an important task, it is not quite the objective of the present paper. Instead, we focus on a somewhat lower-level task, *i.e.*, can we recover the relationships between meters in absence of any labelled data on these relationships? Specifically, we present a range of methods that can automatically (without manual labeling) infer metadata related to the consumption hierarchy directly from the consumption patterns. Our approach is statistical in nature, predicating on two possible measures of association between meters. Given this statistical approach, it is prudent to verify the methods in a simulated setting in addition to a real-world application. For the former, we

The authors would like to acknowledge support from Jan Bastiaans and the Energy Lancaster team who curated and enabled access to the smart meter data used in this study. This work was funded by EPSRC grant EP/T025964/1.

consider a range of models aggregating dependent and non-Gaussian data over random trees, and for the latter we consider a case study relating to the library building on a university campus for which we have ground-truth data via the wiring diagram.

### A. Statistical Background

The aggregation of data is often performed deliberately, for instance, to compress or to enhance the interpretability of data. Knowledge of aggregation hierarchies is known to be useful in a prediction setting; for example, Hyndman et al. [4]–[6] use knowledge of the aggregation hierarchies to more accurately model and forecast future observations. When we observe data that are aggregated but do not have access to the aggregation hierarchy, the estimation of such a hierarchy forms a structure learning task over the space of feasible aggregation trees.

Existing literature on hierarchical time series typically assumes that the hierarchy is known. For example, van Erven and Cugliari [7] represented electricity demand at increasing levels of aggregation such that hierarchical time series prediction can be used to 1) improve the forecasts and 2) to ensure that the forecasts are aggregate consistent (*i.e.*, they add up nicely), cf. [4], [5], [8]. Chow and Liu [9] demonstrated how to recover tree models using the maximum-weight spanning tree (MWST) with mutual information as edge weights. More recently, Quinn et al. [10], [11] proposed to use Edmonds' algorithm with directed information for the estimation of directed trees, while Tan et al. [12] looked at bounding the error for estimating tree-structured Gaussian graphical models with a Chow–Liu approach. For more general graphical models [13], recent innovations focus on efficient model selection via regularization [14], including harnessing prior knowledge about the absence of edge structure [15]. In power systems research, Liao, Weng *et al.* [16], [17] used the Chow–Liu algorithm to reconstruct power grid topologies.

### B. Meter Aggregation Trees

To formalize the problem, let $x_t^{(j)}$ denote the observations on *meter* $j$ at discretely-indexed time points $t = 1, \ldots, n$, for simplicity we assume that $n$ is the same for all $j = 1, \ldots, J$ meters. We further assume that the interval between measurements is the same across all meters. Let the meters be associated with vertices $V = \{1, \ldots, p\}$ in a graph $G(V, E)$ with edge set $E$. We wish to find a tree structure over which the measurements are aggregated, *i.e.*,

$$x_t^{(j)} = \sum_{i \in \mathrm{child}(j)} x_t^{(i)}, \tag{1}$$

where $\mathrm{child}(j) = \{i : (i, j) \in E\}$. An example of such a tree and its edge set is given in Figure 2. Technically, the trees we consider are arborescences (rooted directed trees), and these represent a subset of the directed acyclic graphs that could feasibly be constructed with vertex set $V$.

We assume that the disaggregated consumption is present on the leaves of the tree (red meters in Figure 2). We then consider two different observational settings:

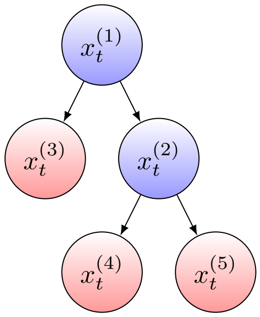

Fig. 2. An example of a meter hierarchy where $x_t^{(1)}$ is the root meter and $x_t^{(3)}$, $x_t^{(4)}$ and $x_t^{(5)}$ are the leaf meters. The signals are aggregated hierarchically such that $x_t^{(1)} = x_t^{(3)} + x_t^{(4)} + x_t^{(5)}$ and $x_t^{(2)} = x_t^{(4)} + x_t^{(5)}$. Note that some of the meters may be unobservable in real life.

1) Every meter on the network is observed, *e.g.*, associated with a time series. For instance, in Figure 2 we would observe the series for all five meters.
2) We only observe time series for a subset of meters. In this case, the series may not add up exactly in that the parents consumption may appear to be different from its child. The differing amount is due to unobserved child meters.

### C. Our Contributions

This paper contributes to the literature in two key ways. First, we demonstrate how a range of correlation-based approaches can be used to infer smart meter hierarchies, and the effectiveness of these procedures on real-world data. Second, we provide a technical contribution that extends existing tree-learning methods with a specific focus on inference for aggregation hierarchies.

Given our interest in learning trees, we naturally focus on Chow–Liu type algorithms [9]. However, instead of estimating the mutual information directly, we use the correlation to determine edge costs which can then be fed to either Kruskal's or Edmonds' algorithm to compute the final tree estimates. Importantly, we propose two extensions to this approach:

1) We consider the use of distance correlation [18], rather than correlation, as a measure of dependence between time series. To our best knowledge, this is the first application of distance correlation to MWST.
2) We introduce a straightforward dominance condition onto the hierarchy which requires child meters to be smaller in magnitude than their parents.

The motivation behind these extensions is to improve the robustness of the procedure when faced with either non-Gaussian, or non-linear relationships between the meters. The dominance condition is motivated by our application to energy time series, whereby the non-negative nature of the observations allows us to very quickly rule out infeasible edges.

Section 2 develops the methodology including the dominance and distance correlation adaptations. Section 3 presents a range of synthetic experiments to verify that our method can recover aggregation trees under a range of conditions, while Section 4 presents a case study on data obtained from a university library.

## II. Methodology

An important contribution to the tree learning literature is that of Chow–Liu [9] who demonstrate that the maximum-likelihood estimate of the tree structure for a discrete distribution is given by minimizing the K-L divergence of a graph restricted to trees (with distribution $Q$). Extending this idea to the Gaussian setting, Tan et al. [12] consider the K-L divergence between a tree-structured Gaussian graphical model and the normal distribution parameterized by the empirical covariance, *i.e.*, $\hat{P}(x) = \mathcal{N}(x; 0, \hat{\Sigma})$ such that

$$\hat{E} = \arg \min_{E_Q : Q \in \mathcal{P}_{\mathcal{N}}(\mathbb{R}^J, \mathbb{T}^J)} D_{\mathrm{KL}}(\hat{P} \parallel Q). \tag{2}$$

This is equivalent to maximizing the sum of mutual information between each meter and its parent and therefore we can obtain the optimal tree using a MWST algorithm. In general, it is not efficient to compute mutual information empirically when the number of samples is small (as these are typically based on discretization), so an alternative is to assume Gaussian distribution, which yields [12]:

$$I(x; y) = -\frac{1}{2} \log \left(1 - \rho(x, y)^2\right), \tag{3}$$

where $I$ denotes mutual information and $\rho(x, y)$ denotes the correlation between $x$ and $y$. Since $I$ is monotonic with respect to $\rho^2$, the simplest algorithm is to feed the squared correlation matrix into a MWST algorithm.

### A. Distance Correlation (dCor)

In many situations, the Gaussian assumption previously discussed may not hold. However, we may still wish to use the Chow–Liu type argument to obtain a tree estimate. In this setting we propose to use the distance correlation, which mimics the construction of a correlation but uses an inner product form which looks at the pairwise distances between all points observed. Recall that the sample correlation for $x, y \in \mathbb{R}^{\ltimes}$ has the form

$$\rho(x, y) = \sigma(x, y)(\sigma(x, x)\sigma(y, y))^{-1/2}, \tag{4}$$

where $\sigma(x, y) = \langle x - \bar{x}, y - \bar{y} \rangle / n$. The correlation only measures linear independence, not statistical independence. A better measure of the latter can be obtained via the sample distance correlation [18], [19], defined as

$$\rho_d(x, y)^2 = \sigma_d(x, y)^2 / \sigma_d(x, x)\sigma_d(y, y), \tag{5}$$

where $\sigma_d(x, y)^2 = \langle CXC, CYC \rangle_F / n^2$, $C$ is the centering matrix, and $X_{ij} = |x_i - x_j|$ and $Y_{ij} = |y_i - y_j|$ for $i, j = 1, \ldots, n$ are the pairwise distance matrices for $x$ and $y$, respectively. In particular, $0 \leq \rho_d(x, y)^2 \leq 1$ with equality to 0 if and only if $x$ and $y$ are statistically independent [18]. Fast algorithms for distance correlation are available from [20], [21]. To create the optimal tree using distance correlation, we simply feed the distance correlation matrix into a MWST algorithm.

### B. Graph Inference with Dominance Constraints

Given the aggregation property of the tree hierarchy, and under the assumption that $x_t^{(i)} \geq 0$, no meter can be the parent of meters with larger readings. Therefore, under such conditions it is natural to define a dominance matrix as:

$$dom_{ij} = \mathbf{1}(x_t^{(i)} \geq x_t^{(j)}, \, \forall t). \tag{6}$$

Next, we remove all non-dominant edges before graph estimation. In doing so, we reduce the probability of including the wrong edges in the graph. If a directed graph is not required, we can use $dom|dom^T$ instead. Here | denotes the logical OR, *i.e.*, we symmetrize the dominance matrix.

TABLE I
Comparison of this work and previous work.

| Algorithm | Input Matrix | Output Tree |
|---|---|---|
| Chow and Liu [9] | $[I(x_i; x_j)]_{i,j=1}^p$ | Undirected |
| Mantegna [22] | $[\sqrt{2(1 - \rho(x_i, x_j))}]_{i,j=1}^p$ | Undirected |
| Quinn et al. [10] | $[I(x_j \to x_i)]_{i,j=1}^p$ | Directed |
| This work | $\rho^2$ or $\rho_d^2$ matrix with $dom$ | Either |

### C. Maximum Spanning Tree with Distance Correlation and Dominance

We propose to use the meter with the largest cumulative reading $\arg\max_i \sum_{t=1}^n x_t^{(i)}$ as the root meter. For an undirected tree, the root is not so important; however, for our aggregation tree, we wish to disaggregate from the root down to the leaves. The goal is thus to find a directed rooted tree (arborescence) based on our measures of dependence (either squared correlation, or distance correlation). Edmonds' algorithm is a standard approach for this task, with efficient $O(p \log p + |E|)$ computation [23]. It is worth noting that with our dominance conditions in place, the number of possible edges $|E|$ can be drastically reduced if $\sum_{ij} dom_{ij} \ll p^2$. If we do not have such conditions, then Edmonds' algorithm will achieve complexity $O(p^2)$. For comparison purposes, we also run our experiments with the classical Kruskal's algorithm which simply performs a greedy search based on the dependence weights, followed by a breadth-first search from the root to turn it into a directed tree (polytree). Pseudo-code for replicating the method is given in Algorithm 1.

To summarize, we consider eight methods in total, including all combinations of either: squared correlation (SC) or distance correlation (DC); undirected (polytree, P) or directed (arboresecence, A); and either dominance (D) or no dominance conditions imposed. In the experiments that follow, we identify the methods by combinations of the above acronyms, *e.g.*, SCAD corresponds to Squared Correlation + directed Arboresecence + Dominance conditions imposed, whereas SCP correspond to Squared Correlation + undirected Polytree (no dominance conditions imposed).

**Algorithm 1** Maximum Spanning Tree with Dominance

```
Input: $\mathbf{X} \in \mathbb{R}^{n \times k}$, $algo \in \{cor^2, dCor\}$, $directed \in \{T, F\}$
Output: $tree$
 1: Compute $dom$ from $\mathbf{X}$ using Eq. (6)
 2: $\mathbf{C} = -algo(\mathbf{X})$ // negation turns MinST into MaxST
 3: if $directed$ then
 4:     $\mathbf{C}(\neg dom) \leftarrow 0$
 5:     $tree \leftarrow \text{Edmonds}(\mathbf{C})$
 6: else
 7:     $dom \leftarrow dom|dom^T$ // see Subsection 2.2
 8:     $\mathbf{C}(\neg dom) \leftarrow 0$
 9:     $tree \leftarrow \text{BFS}(\text{Kruskal}(\mathbf{C}))$ // from the largest meter
10: end if
```

## III. Experiments

To investigate the performance of our methods, we construct a simple testbed whereby we simulate a random tree with exactly $p$ meters according to a Galton–Watson process, with branchings governed by the discrete uniform distribution $\mathcal{U}\{2, k\}$. Here $k$ is the maximum number of children for each meter, and for reconstructability, we exclude the possibility of having meters with only a single child. For each leaf of the tree, we simulate a random signal and then aggregate this over the tree according to $X = X_{\text{leaf}}A$ where $A$ is a binary matrix with entries:

$$A_{ij} = \mathbf{1}(\text{leaf}(i) \in \text{path}(root, j)). \tag{7}$$

As an example of use, please refer to the aggregation tree in Figure 2. It can be readily verified that:

$$\underbrace{\begin{bmatrix} x_{11} & \dots & x_{15} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{n5} \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} x_{13} & x_{14} & x_{15} \\ \vdots & \vdots & \vdots \\ x_{n3} & x_{n4} & x_{n5} \end{bmatrix}}_{X_{\text{leaf}}} \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}}_{A}. \tag{8}$$

We investigate two scenarios, the first represents very idealistic settings where data is simulated independently and identically across time, with the same distribution for each leaf meter. In these cases, the focus is on the performance of the method as a function of the number of data point $n$ and the structure of the graphs, which we vary through the maximum number of children $k$. For each setting of the experimental parameters, we simulate 2,000 random trees, and compare the estimation performance using Matthews' correlation coefficient [24] between the estimated trees and the true trees,

$$\text{MCC} = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}. \tag{9}$$

where $tp$, $tn$, $fp$, and $fn$ are the number of true positives, true negatives, false positives, and false negatives, respectively. The MCC has a similar interpretation as a correlation coefficient, 1 is the best fit, while a value of around 0 indicates the estimated tree is not informative.

### A. Synthetic Data Generation

Since real meter data may possess occasional outliers that result in heavy-tailed distributions, we should be careful to consider a range of processes when evaluating our proposed methodology. We therefore consider a family of stable distributions that can mimic these scenarios—with the Gaussian setting forming an edge case.

*1) Symmetric $\alpha$-Stable Distribution:* This distribution we simulate from is formally defined by the characteristic function $E[e^{iuX}] = \exp(-\gamma^\alpha |u|^\alpha)$ where $0 < \alpha \leq 2$ and $\gamma > 0$ is a scale parameter [25]. We write $X \sim S\alpha S(\gamma)$ if $X$ follows the symmetric $\alpha$-stable distribution. $S2S(\gamma)$ is the Gaussian distribution $\mathcal{N}(0, 2\gamma^2)$, whereas for $\alpha < 2$ the tail distribution approximately follows the power law $P(X > x) \propto x^{-\alpha}$. In general, the stable density is not expressible in closed form, but there are a few exceptions including $\alpha = 1$ (Cauchy), $\alpha = 1.5$ (Holtsmark), and $\alpha = 2$ (Gaussian).

*2) Fractionally Integrated Process with $S\alpha S$ Innovations:* The fractionally integrated FI($d$) process is similar to its non-fractional counterpart except that the differencing parameter $d \in \mathbb{R}$ is fractional:

$$(1 - B)^d x_t = \varepsilon_t. \tag{10}$$

Here $x_t$ is a time series we wish to simulate, $\varepsilon_t$ are the i.i.d. innovations, and $B$ is the backshift operator. In this paper, the innovations are assumed to be $S\alpha S$, which in turn requires two convergence conditions: $1 < \alpha \leq 2$ and $\alpha(d - 1) < -1$ [26]. The advantage of fractional $d$ is that when $d > 0$ it can model long-range dependence where the autocorrelation decays slower than exponentially, and the degree of long-range dependence is captured by $d$. To simulate $x_t$, multiply (10) by $(1 - B)^{-d}$ on both sides then use the binomial expansion to get $x_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ with coefficients $c_0 = 1$ and $c_j = \Gamma(j + d)/(\Gamma(d)\Gamma(j + 1))$ for $j \geq 1$ [26]. To make this infinite sum computable, we truncate the sum as follows:

$$x_t = \sum_{j=0}^{n-1} c_j \varepsilon_{t-j}. \tag{11}$$

### B. Experiment 1: Complete Observations

Consider $X_i \sim FI(d)$ with $S\alpha S(1)$ innovations for each leaf $i$ of tree $T$. We generate $n$ samples each then subtract the global minimum to ensure that all data are non-negative. This is then aggregated via (7) to produce the observed data. We use the following parameters: $\alpha \in \{1.5, 2\}$, with and without long-range dependence ($d = 1 - 1/\alpha$ and $d = 0$), $n \in \{10, 20, 50, 100, 200, 500\}$, $p = 31$, $k = 3$, with and without dominance, with 2,000 experiments per scenario, via Algorithm 1. When not using dominance, we remove lines 1, 4, 7–8 from Algorithm 1. The results are shown in Figure 3. Results demonstrates that methods with dominance are clearly best, and squared correlation comes out on top for $\alpha = 2$ (Gaussian) whereas distance correlation comes out best for $\alpha = 1.5$ (Holtsmark). Furthermore, we can see that $\alpha = 1.5$

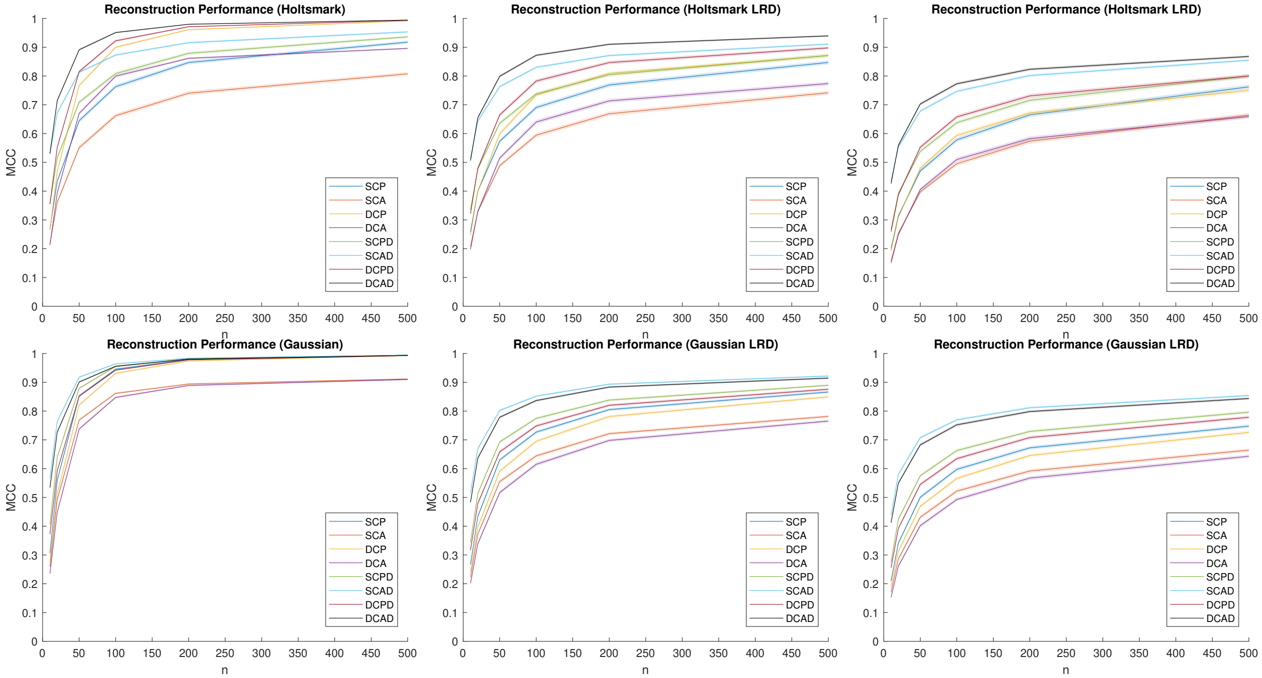

Fig. 3. Symmetric $\alpha$-stable distribution simulation results. Plots of MCC vs $n$ for all methods with shaded confidence intervals (1st–99th percentile). Left: no long-range dependence. Middle: with long-range dependence. Right: with long-range dependence and 16 missing meters. Top: $\alpha$ = 1.5. Bottom: $\alpha$ = 2.

is harder than $\alpha = 2$ and data with long-range dependence is harder than those without.

To assess the robustness to different tree shapes, we also look at different $k$-restricted trees and tabulate their relative rankings and MCCs for each algorithm as $k$ varies from 2 to 5. The resulting rankings of the methods are relatively stable, as can be seen from the means and standard deviations in Tables II and III.

TABLE II
MCCs FOR $\alpha$ = 1.5, $n$ = 500.

| | Means (Standard Deviations) | | | |
|---|---|---|---|---|
| Method | $k = 2$ | $k = 3$ | $k = 4$ | $k = 5$ |
| SCP | .863 (.115) | .847 (.122) | .832 (.126) | .803 (.134) |
| SCA | .729 (.129) | .742 (.127) | .744 (.127) | .720 (.132) |
| DCP | .887 (.098) | .871 (.106) | .847 (.113) | .806 (.120) |
| DCA | .764 (.117) | .774 (.116) | .763 (.118) | .728 (.123) |
| SCPD | .877 (.107) | .871 (.110) | .866 (.112) | .852 (.117) |
| SCAD | .912 (.073) | .911 (.074) | .909 (.076) | .896 (.084) |
| DCPD | .904 (.090) | .898 (.097) | .889 (.100) | .868 (.107) |
| DCAD | **.936 (.057)** | **.939 (.056)** | **.937 (.057)** | **.925 (.065)** |

### C. Experiment 2: Missing Data

This experiment is motivated by real-life situations where a number of meters are unobservable. Tree generation is identical to Experiment 1 except that we randomly remove 16 meters after aggregating from a random tree of size 47, such that only 31 meters are available for reconstruction. We conduct experiments for $\alpha \in \{1.5, 2\}$, $n = 500$, $p = 47$, $k = 3$, with 2,000 experiments per each $\alpha$. The results are given in Figure 3 show the importance of the dominance condition in maintaining a reasonable level of performance, where we see SCAD and DCAD consistently outperforming the other methods. Generally in the setting with missing meters we should not expect to consistently recover the true graph; however, the proposed dominance condition still enables us to recover the majority of the graph structure in an acceptable manner (peaking around MCC ≈ 0.85).

TABLE III
MCCs FOR $\alpha$ = 2, $n$ = 500.

| | Means (Standard Deviations) | | | |
|---|---|---|---|---|
| Method | $k = 2$ | $k = 3$ | $k = 4$ | $k = 5$ |
| SCP | .872 (.077) | .866 (.080) | .839 (.080) | .801 (.086) |
| SCA | .764 (.093) | .781 (.086) | .765 (.084) | .733 (.089) |
| DCP | .860 (.081) | .849 (.084) | .818 (.085) | .777 (.090) |
| DCA | .752 (.097) | .765 (.090) | .745 (.089) | .708 (.094) |
| SCPD | .887 (.072) | .890 (.073) | .874 (.072) | .853 (.076) |
| SCAD | **.918 (.057)** | **.922 (.056)** | **.919 (.055)** | **.904 (.057)** |
| DCPD | .876 (.077) | .876 (.079) | .858 (.078) | .838 (.080) |
| DCAD | .911 (.060) | .915 (.058) | .910 (.058) | .896 (.060) |

## IV. Estimation of Metering Hierarchy

In this section, we present a real-world application of our aggregation tree estimators. The task is based on the observations of 20 meters measuring electricity consumption at different points throughout a university library.

### A. Meter Data

The building we consider serves as the main library for a university campus (Figure 4). It has been extended in various stages over the years, with the original building dating back to the construction of the campus in the 1960's. The metering of the building has also developed over the years. Historically

Fig. 4. An image of the library studied in this paper. The majority of the building is brick construction (circa 1964).

only the consumption at the edge of the building had been monitored. The historic data (more than 10 years old) was observed via dial meters; however, more recently the university has transitioned to digital metering with much smaller meter constants. The meters themselves are queried by data loggers (associated with ID's as mentioned earlier). These data loggers monitor the consumption within each 10 minute interval, reporting the number of times $C$ kWh of energy has been consumed in this interval, where $C$ is the meter constant. With the digital meters $C$ can typically be very small; however, there will still be some quantization error at this stage of data collection.

Due to the infrastructure in place to query and store data from the sensors there are common outages which results in sporadic periods of missing data for many years of measurement. For this reason, we focus our analysis in 2021, which was the cleanest subset of data we could easily extract for this case study. As the university was periodically closed due to COVID-19 response the usage pattern, and therefore consumption across the library may therefore not represent a standard operational year. While this would be a challenge if we were interested in drawing insights on the patterns of consumption, it should not hinder our analysis here which relates to the hierarchy of meters themselves—we should not expect the metering hierarchy to change over the course of our analysis, or indeed from year to year (unless new meters are installed). Our observations are gathered every 10 minutes for the same 20 meters across a period of 105 days. In total, we have 151,200 observations.

## B. Tree Estimation Procedure

To perform the analysis, we consider 105 days worth of data and build a separate tree for each day. We repeat this procedure for all variations of the method, and report the recovery results relative to the ground-truth tree, as encoded via the wiring diagram. To remove the strong periodic component, we perform the analysis with and without differencing.

It is worth noting that we perform the analysis with and without a set of five virtual meters, which allows us to consider a more complex hierarchy. The use of such virtual meters is justified based on the wiring diagram (see Figure 1), where we would like to measure consumption when meters are combined, but given the meter placement these profiles are not physically measured. In the analysis below, we contrast the performance of the methods with and without these virtual meters.

## C. Daily Analysis

In this section we present a summary of the results taken across each day of measurements. The results of the study are presented in Figure 5. The results align with the synthetic experiments, except that it is harder to recover the tree with real data, for example the best performing method DCAD typically achieves an MCC of around 0.6–0.7 at a daily level. As expected, when we remove the virtual meters (which can be seen as meters 15–20 in Figure 1) the recovery of the tree becomes more difficult. This is likely due to further consumption occurring between the parent and child meters, so the aggregation is not tight. In general we see that the directed tree search wins out over the undirected tree, and the dominance constraints prove useful. Significantly, this allows the dominance aware methods to identify the correct root meter with high probability. Results on the non-differenced data (*i.e.*, consumption profiles seen in Figure 1) are typically slightly worse than on the differenced data which we present here. The general pattern of performance is similar however, with DCAD and SCAD winning the comparison.

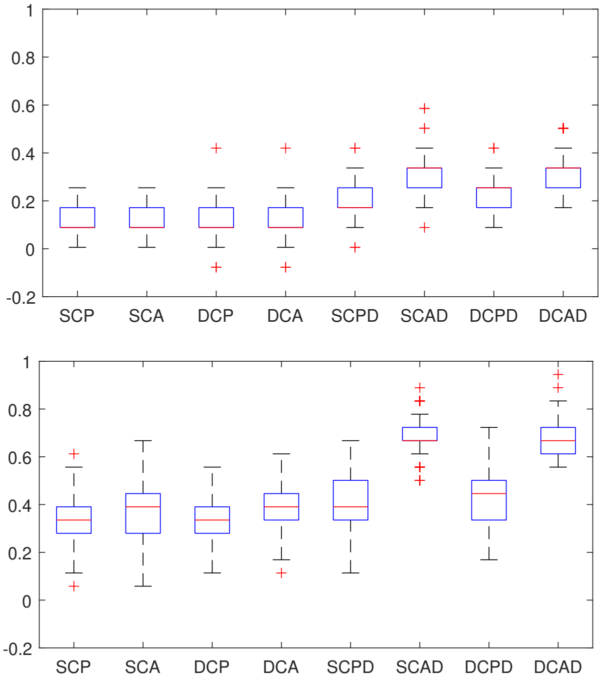

Fig. 5. Boxplots of performance as measured by the MCC for the differenced data. Top: performance without virtual meters. Bottom: performance with virtual meters.

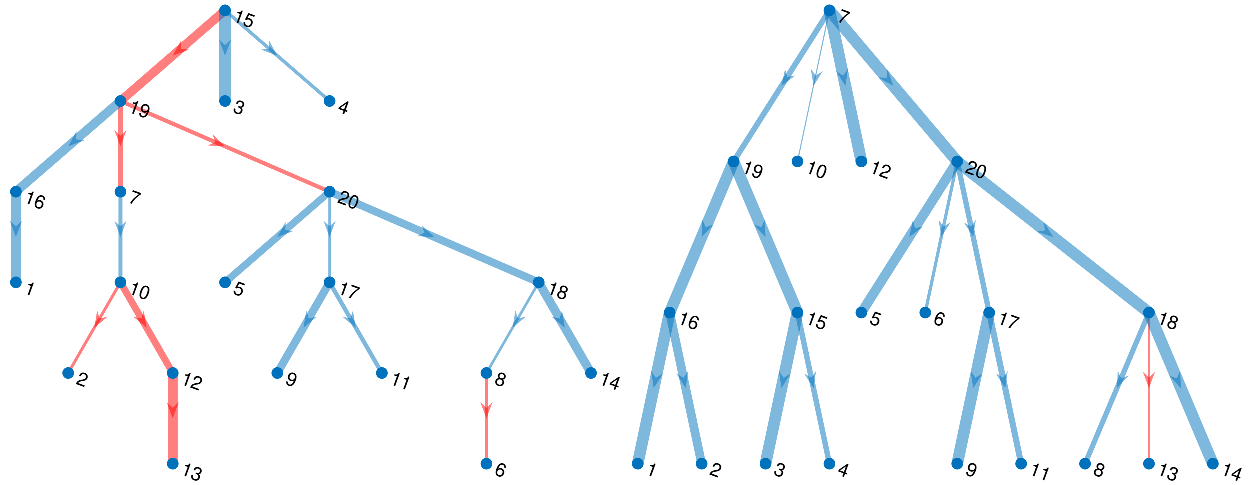

Fig. 6. Reconstructed trees using distance correlation. Weight of line is proportional to the number of times an edge was recovered over the 105 sub-samples (red indicates incorrect edges). Left: no dominance. Right: with dominance. *Note: the true tree is recovered in the right plot except for meter 13 which should be directly connected to meter 7.*

### *D. Composite Analysis*

Given that we do not expect the hierarchy to evolve over time, it is useful to further summarize the daily results. For visualization purposes here we plot a tree estimated on the whole data set, *i.e.*, instead of building one tree for each day, we estimate a single tree on the whole dataset. To assign some measure of confidence in the estimated edges we then weight the edges by the empirical probability of that edge being present across each day (cf. [27], [28]). In this application, we can easily elicit the dominance conditions, and to illustrate the effectiveness of these constraints we plot the trees with and without these conditions in Figure 6. Results show that dominance is very helpful for the reconstruction. The trees presented here are based on distance correlation; however, similar results are achieved using squared correlation. The results here include the virtual meters which help trace out the hierarchy (*i.e.*, the intermediate nodes) in the tree.

We remark that while the tree in this example is relatively small, it is still a challenging problem as there are effectively many missing meters in this application, *i.e.*, consumption can occur between meters in the network.

## V. Conclusion

In this paper we have investigated several ways to estimate aggregation trees from smart meter data. We have proposed a novel combination of methods, and importantly introduced constraints based on dominance conditions to the tree-learning algorithm. We have also extended Chow and Liu [9] style approaches by considering the use of distance correlation rather than correlation. In general, we find that the DCAD, or distance correlation + directed arboresecence + dominance conditions imposed, performs the best; however, our studies show the biggest difference in performance is due to the dominance condition. This makes sense in both theory and practice, as it allows one to easily identify feasible parent-child relationships. The real-world results verify this benefit and show that on real smart-meter data it is feasible to recover good estimates of metering hierarchies purely from the consumption data itself.

It is important to reflect on some of the weaknesses of these methods. In general, the problem of reconstructing the hierarchies is very difficult, and this is especially the case when the aggregation relationship is not tight, *i.e.*, the child meters do not add to something close to the parent meters. In practice, this may often happen, for instance, we may have a building that has meters on several, but not all floors. The aggregate consumption across all floors is thus not simply the sum of the individually metered floors. In the case study, we have an analogous situation. To solve this problem, we can introduce virtual meters in strategic places; however, in our example we used prior knowledge to place these virtual meters (based on the wiring diagram). In reality, this may not be possible. As such, our methods should be considered highly experimental—we believe they provide significant utility, but only when one suspects the aggregation hierarchy is tight. In a more advanced system, one could plausibly use further metadata about the meters to direct the tree-estimation procedure; however, given the heterogeneous nature of metering installations, we preferred to investigate the performance of simple tree learning procedures at this stage. Another topic for future research is to use the expected dominance, or $dom_{ij} = \mathbf{1}(\mathbb{E}[sign(x^{(i)} - x^{(j)})] \geq 0)$, which allows more leeway for missing data. One further application of the tree-learning algorithms proposed could be to improve energy consumption forecasting. Once we have a good grasp on the aggregation structure, we can use this to help reconcile time-series forecasts of consumption and the meter level, including at a highly granular scale, for instance, individual room/floor consumption.

# References

[1] A. K. Prakash, V. C. Prakash, B. Doshi, U. Arote, P. K. Sahu, and K. Ramamritham, “Locating and sizing smart meter deployment in buildings,” in *Proc. ACM 6th Int. Conf. Future Energy Syst.*, 2015, pp. 223–224.

[2] B. Balaji, A. Bhattacharya, G. Fierro, J. Gao, J. Gluck, D. Hong, A. Johansen, J. Koh, J. Ploennigs, Y. Agarwal, M. Bergés, D. Culler, R. K. Gupta, M. B. Kjærgaard, M. Srivastava, and K. Whitehouse, “Brick: Metadata schema for portable smart building applications,” *Appl. Energy*, vol. 226, pp. 1273–1292, Sep. 2018.

[3] J. Koh, D. Hong, R. Gupta, K. Whitehouse, H. Wang, and Y. Agarwal, “Plaster: An integration, benchmark, and development framework for metadata normalization methods,” in *Proc. 5th ACM Int. Conf. Syst. Built Environ. (BuildSys)*, 2018, pp. 1–10.

[4] R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang, “Optimal combination forecasts for hierarchical time series,” *Comput. Statist. Data Anal.*, vol. 55, no. 9, pp. 2579–2589, Sep. 2011.

[5] S. L. Wickramasuriya, G. Athanasopoulos, and R. J. Hyndman, “Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization,” *J. Amer. Statist. Assoc.*, vol. 114, no. 526, pp. 804–819, Apr. 2019.

[6] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, “Hierarchical probabilistic forecasting of electricity demand with smart meter data,” *J. Amer. Statist. Assoc.*, vol. 116, no. 533, pp. 27–43, Jan. 2021.

[7] T. van Erven and J. Cugliari, “Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts,” in *Modeling and Stochastic Learning for Forecasting in High Dimensions*, A. Antoniadis, J.-M. Poggi, and X. Brossat, Eds., 2015, pp. 297–317.

[8] Y. Pang, B. Yao, X. Zhou, Y. Zhang, Y. Xu, and Z. Tan, “Hierarchical electricity time series forecasting for integrating consumption patterns analysis and aggregation consistency,” in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 3506–3512.

[9] C. Chow and C. Liu, “Approximating discrete probability distributions with dependence trees,” *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 462–467, May 1968.

[10] C. J. Quinn, T. P. Coleman, and N. Kiyavash, “Approximating discrete probability distributions with causal dependence trees,” in *Proc. 2010 Int. Symp. Inf. Theory Appl.*, 2010, pp. 100–105.

[11] C. J. Quinn, N. Kiyavash, and T. P. Coleman, “Efficient methods to compute optimal tree approximations of directed information graphs,” *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3173–3182, Jun. 2013.

[12] V. Y. F. Tan, A. Anandkumar, and A. S. Willsky, “Learning Gaussian tree models: Analysis of error exponents and extremal structures,” *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2701–2714, May 2010.

[13] S. L. Lauritzen, *Graphical Models*. Oxford, U.K.: Oxford Univ. Press, 1996.

[14] O. Banerjee, L. El Ghaoui, and A. d’Aspremont, “Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data,” *J. Mach. Learn. Res.*, vol. 9, no. 15, pp. 485–516, 2008.

[15] M. Grechkin, M. Fazel, D. Witten, and S.-I. Lee, “Pathway graphical lasso,” in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2617–2623.

[16] Y. Liao, Y. Weng, M. Wu, and R. Rajagopal, “Distribution grid topology reconstruction: An information theoretic approach,” in *Proc. 2015 North Amer. Power Symp. (NAPS)*, 2015, pp. 1–6.

[17] Y. Weng, Y. Liao, and R. Rajagopal, “Distributed energy resources topology identification via graphical modeling,” *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2682–2694, Jul. 2017.

[18] G. J. Székely, M. L. Rizzo, and N. K. Bakirov, “Measuring and testing dependence by correlation of distances,” *Ann. Statist.*, vol. 35, no. 6, pp. 2769–2794, Dec. 2007.

[19] M. Koc, B. Akinci, and M. Bergés, “Comparison of linear correlation and a statistical dependency measure for inferring spatial relation of temperature sensors in buildings,” in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Buildings (BuildSys)*, 2014, pp. 152–155.

[20] X. Huo and G. J. Székely, “Fast computing for distance covariance,” *Technometrics*, vol. 58, no. 4, pp. 435–447, Oct. 2016.

[21] A. Chaudhuri and W. Hu, “A fast algorithm for computing distance correlation,” *Comput. Statist. Data Anal.*, vol. 135, pp. 15–24, Jul. 2019.

[22] R. N. Mantegna, “Hierarchical structure in financial markets,” *Eur. Phys. J. B Condens. Matter Complex Syst.*, vol. 11, no. 1, pp. 193–197, Sep. 1999.

[23] H. N. Gabow, Z. Galil, T. Spencer, and R. E. Tarjan, “Efficient algorithms for finding minimum spanning trees in undirected and directed graphs,” *Combinatorica*, vol. 6, no. 2, pp. 109–122, Jun. 1986.

[24] D. Chicco, N. Tötsch, and G. Jurman, “The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation,” *BioData Mining*, vol. 14, no. 1, p. 13, Feb. 2021.

[25] J. P. Nolan, “Basic properties of univariate stable distributions,” in *Univariate Stable Distributions: Models for Heavy Tailed Data*. Cham, Switzerland: Springer, 2020, pp. 1–23.

[26] P. S. Kokoszka and M. S. Taqqu, “Fractional ARIMA with stable innovations,” *Stochastic Process. Appl.*, vol. 60, no. 1, pp. 19–47, Nov. 1995.

[27] M. Tumminello, C. Coronnello, F. Lillo, S. Miccichè, and R. N. Mantegna, “Spanning trees and bootstrap reliability estimation in correlation-based networks,” *Int. J. Bifurcation Chaos*, vol. 17, no. 7, pp. 2319–2329, Jul. 2007.

[28] F. Musciotto, L. Marotta, S. Miccichè, and R. N. Mantegna, “Bootstrap validation of links of a minimum spanning tree,” *Physica A: Statist. Mech. Appl.*, vol. 512, pp. 1032–1043, Dec. 2018.