

ELECTRICAL PHASE IDENTIFICATION USING A CLUSTERING ALGORITHM

[0001] This patent application claims benefit of U.S. non-provisional patent application serial no. 17/732,505, titled “ELECTRICAL PHASE IDENTIFICATION USING A CLUSTERING ALGORITHM”, filed on 04/28/2022, which claims benefit of priority to U.S. provisional patent application serial no. 63/246,269, titled “ELECTRICAL PHASE IDENTIFICATION USING A CLUSTERING ALGORITHM”, filed on 09/20/2021, both of which are incorporated herein by reference.

TECHNICAL FIELD

[0002] The present disclosure generally relates to the field of phase-balancing, and more specifically to methods, apparatus, and system for providing phase identification for electricity meters connected to an electrical grid.

BACKGROUND

[0003] Phase-balancing, such as balancing the electrical load on each of three electrical phases, in an electrical grid is an important consideration for utility providers. Phase-balancing may be complicated if electrical utility company records are incorrect, which frequently happens, for example, when linemen move a customer’s connection, i.e., the customer’s electricity meter, from one phase to another to better balance the load, but fail to record their actions and update the phase information of the customer’s meter. Poorly balanced load reduces operational efficiency, and increases the likelihood of equipment failure and delays power outage management.

[0004] An improved and/or correct understanding of the topology has many advantages for advanced metering infrastructure (AMI) having automated meter reading (AMR) and other features associated with smart electricity meters. Correct understanding of the phase of a meter can help to improve an understanding of the topology of the electrical grid.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] The detailed description is set forth with reference to the accompanying figures. In the figures, the left-most digit(s) of a reference number identifies the figure in which the reference number first appears. The use of the same reference numbers in different figures indicates similar or identical items or features.

[0006] FIG. 1 illustrates an example power distribution environment.

[0007] FIG. 2 illustrates a schematic diagram of an example transformer.

[0008] FIG. 3 illustrates an example set of voltage graphs for various phases.

[0009] FIG. 4A and 4B illustrate an example process for identifying a phase at an electricity meter level.

5 [0010] FIG. 5 illustrates an example detail process of one of the blocks of FIG. 4.

[0011] FIG. 6 illustrates an example detail process of another block of the blocks of FIG. 4.

[0012] FIG. 7A illustrates an example display of the clusters of electricity meters.

[0013] FIG. 7B illustrates example display of a phase map of the electricity meters of FIG. 7A.

10

[0014] FIG. 8 illustrates an example block diagram of a system for identifying electrical phase.

DETAILED DESCRIPTION

15 [0015] Systems and methods for identifying a phase connected to electricity meters are disclosed.

[0016] Electricity is generated in three phases, A, B, and C, and on each phase, voltage oscillates in a sine wave, for example, at 60 Hz. Each of three phases of electricity is transmitted on a separate power line and there may be a fourth line, N, a ground or neutral wire with no voltage on it. These lines interact with each other at transformers, or where power is consumed.

[0017] FIG. 1 illustrates an example power distribution environment 100. In this example, a power plant 102 generates electricity, which is carried by high voltage lines 104 to a power substation 106. The power substation 106 provides electricity via a feeder 108 to a transformer 110. The feeder 108 is a power line consisting of individual powered lines with phase A, B, and C servicing a plurality of premises connected via the transformer 110 and electricity meters 112A, 112B, and 112C providing electricity to associated premises 114A, 114B, and 114C.

[0018] FIG. 2 illustrates a schematic diagram of an example transformer 200, such as a distribution transformer. Transformers are used to adjust voltage, and can be wired between a powered line and the neutral line, in which case, an output phase corresponds to the phase of the powered line. The transformers can also be wired between two powered lines, in which case, an output phase differs from all three powered lines and may be referred as a hybrid

30

phase. Therefore, there are six possible phases at the metering level: Phase A-N (or A), Phase B-N (or B), Phase C-N (or C), Phase A-B, Phase B-C, and Phase A-C. In this example, Phases A, B, and C are shown as on lines 202, 204, and 206, respectively. In this example, there are 2400 V between Phases A 202 and B 204 and between Phases B 204 and C 206, and a first connection 208 of a primary winding 210 of the transformer 200 is connected to Phase C 206 and a second connection 212 of the primary winding 210 is connected to Phase B 204. A secondary winding 214 has three outputs, a first output 216, a second output, or a center tap, 218, and a third output 220, which are connected to a line-a 222, a neutral line 224, and a line-b 226, respectively. The transformer 200 in this example is a step-down transformer that reduces the voltage of the powered lines, in this case, Phases B 204 and C 206, from 2400 V to 120 V between the line-a 222 and the neutral 224 and between the line-b 226 and the neutral 224, and to 240 V between the line-a 222 and the line-b 226.

[0019] FIG. 3 illustrates an example set of voltage graphs 300 for various phases. In this example, Phase A 302 is set as a reference, Phase B 304 is 120° ahead of Phase A 302, and Phase C 306 is 240° ahead of Phase A 302. Each phase has a voltage of 120 V_{RMS} and a frequency of 60 Hz. Phase A-B 308, Phase B-C 310, and Phase C-A 312 are also shown. As shown, depending on the phase, the voltage behaves differently in magnitude over time. Poor phase balancing, such as overloading one phase, overloading equipment connected to a phase, or connecting to an incorrect phase, may cause operational inefficiency and equipment overheating, for example, increase in early equipment failure, delays in power outage response/management, and safety hazard.

[0020] FIGs. 4A and 4B illustrate an example process 400 for identifying a phase at an electricity meter level.

[0021] At block 402, voltage time series data collected from every electricity meter on a feeder is entered. A feeder is a power line consisting of individual powered lines with phase A, B, and C servicing a plurality of premises connected via electricity meters. The distinct powered lines are presumed to experience different fluctuations in RMS voltage as a result of differing loads. Those fluctuations are expected to be seen by all electricity meters connected to that line, and voltage readings on the same phase of the feeder are expected to be highly correlated compared to voltage readings on other phases. Accordingly, voltage readings collected from each electricity meter on the feeder over a preselected collection time period, such as from January 1, 2020 to December 31, 2020, may be entered as the voltage time series data. The voltage readings may be taken at a preselected interval, such as every five minutes

with accuracy of ± 0.15 V. With smart electricity meters in advanced metering infrastructure (AMI) having automated meter reading (AMR), the voltage time series data may be automatically transmitted from each electricity meter to, and collected by, a central office of the utility service provider or a third party. Additionally, an existing meter-phase connectivity record, which is the current record of information regarding each meter's connection to phase connections, may also be entered. As discussed above, the existing meter-phase connectivity record may not be up to date due to, for example, when linemen move a customer's electricity meter from one phase to another to better balance the load, but fail to record their actions and update the phase information of the customer's meter.

10 **[0022]** At block 404, the voltage time series data of each electricity meter for a preselected analysis period of the preselected collection time period, such as each month over January 1, 2020 to December 31, 2020, is filtered to omit problematic data or electricity meter. For example, expected average voltages (RMS) may be 120 V, 208 V, 240 V, 277 V, and 480 V for the feeder, then values that are more than $\pm 5\%$ out of the expected average voltages may be omitted. Frozen periods, identified as extended periods of time with constant voltage on a given meter, may be omitted. Jump outliers, identified as large interval-to-interval voltage changes outside of a preselected threshold, may be omitted. Electricity meters with insufficient amount of data over the collection time period may be omitted. Electricity meters having location information inconsistent with actual geographical locations of the electricity meters may be omitted, or the location information may be corrected and the voltage time series data of those electricity meters with the corrected location information may be used.

20 **[0023]** At block 406, voltage correlation of every meter-to-meter combination is calculated. In one example, the voltage correlation may be calculated using Pearson correlation coefficient (PCC) to determine the correlation between voltage at meter A and voltage at meter B, that is, how a change in voltage at meter A affects a change in voltage at meter B. Pearson correlation coefficient, ρ , has a value between -1 and 1, and is given by, for the correlation between X and Y:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}, \quad \text{where:}$$

cov is the covariance,

30 σ_X is the standard deviation of X, and
 σ_Y is the standard deviation of Y.

PCC may be calculated for every meter-to-meter pairing, and the results may be stored in a matrix.

[0024] At block 408, three initial kernels, K1 containing most of the electricity meters for Phases A, B, and C, is determined. For the process of block 408, an agglomerative cluster loop or method may be utilized to determine the three initial kernels. Examples of the agglomerative cluster method include analyses based on a single-linkage distance, Ward linkage distance, dendrogram step-through, and the like. Additionally, or alternatively, a Gaussian mixture model may be utilized to perform the clustering.

[0025] At block 410, a median first order difference voltage for each preselected interval is determined for each of the initial kernels. Correlation, PCC_1^{K1} , PCC_2^{K1} , and PCC_3^{K1} , with each of the three initial kernel for each meter are calculated at block 412. At block 414, a hybrid index for the three initial kernels may be calculated based on a median of the correlations, PCC_1^{K1} , PCC_2^{K1} , and PCC_3^{K1} . The hybrid index may be defined as the ratio of the second highest (median) correlation to the highest correlation,

$Hybrid\ Index^{K1} = \frac{median(PCC_1^{K1}, PCC_2^{K1}, PCC_3^{K1})}{max(PCC_1^{K1}, PCC_2^{K1}, PCC_3^{K1})}$. Alternatively, the hybrid index may also be defined, or calculated as:

$$alt_Hybrid\ Index^{K1} = \left\{ \frac{[-PCC_1^{K1} + PCC_2^{K1}]^2}{2} + \frac{[PCC_1^{K1} + PCC_2^{K1} - 2*PCC_3^{K1}]^2}{6} \right\}^{1/2}.$$

The hybrid index is used to separate out the line-to-line connections from the line-to-neutral connections as described later in more detail. Based on the $Hybrid\ Index^{K1}$, new kernels for each phase, K2, are determined at block 416. The correlation between each of the electricity meters and the new kernels, PCC_1^{K2} , PCC_2^{K2} , and PCC_3^{K2} , are calculated at block 418, and $Hybrid\ Index^{K2}$ is calculated at block 420.

[0026] At block 422, for each preselected analysis period, average correlation with each phase for the new kernels, K2, $mean(PCC_1^{K2})$, $mean(PCC_2^{K2})$, and $mean(PCC_3^{K2})$, and average hybrid index for K2, $mean(Hybrid\ Index^{K2})$, are calculated. The electricity meters are then clustered into three groups based on the average hybrid index for K2, $mean(Hybrid\ Index^{K2})$ at block 424. The three groups include a group with a high hybrid index, which is considered to be the line-to-line phase group, a group with low hybrid index, which is considered to be line-to-neutral phase group, and a group with in-between hybrid index values is used as a band separating the high and low hybrid index groups. At block 426, the electricity meters of the high hybrid index group, X, are grouped into three line-to-line phases, A-B, B-C, and C-A,

based on the average correlation, $mean(PCC_1^{K2})$, $mean(PCC_2^{K2})$, and $mean(PCC_3^{K2})$. For the clustering processes of blocks 424 and 426, the agglomerative cluster method as described above may be utilized.

- [0027]** At block 428, the electricity meters of the low hybrid index group, Y, are grouped into three line-to-neutral phases, A, B, and C, based on the phase having the highest average correlation, $mean(PCC_1^{K2})$, $mean(PCC_2^{K2})$, and $mean(PCC_3^{K2})$ with the meter. The three line-to-line groups of electricity meters and the three line-to-neutral groups of electricity meters are combined as new kernels, K3, having six phases, A, B, C, A-B, B-C, and C-A, at block 430.
- [0028]** At block 432, the filtered data from block 404 is used to calculate correlation of each electricity meter with each of the six kernels of K3, PCC_1^{K3} , PCC_2^{K3} , PCC_3^{K3} , PCC_4^{K3} , PCC_5^{K3} , and PCC_6^{K3} , and hybrid index based on the correlation with line-to-neutral kernels, PCC_1^{K3} , PCC_2^{K3} , and PCC_3^{K3} , is calculated at block 434. At block 436, average correlation with each of six phases are calculated as $mean(PCC_i^{K3})$, for $i = 1, 2, 3, 4, 5, 6$, where i represents each of the six phases, A, B, C, A-B, B-C, and C-A. At block 438, an average hybrid index, $mean(Hybrid Index^{K3})$ is calculated. The electricity meters are grouped into two groups, a line-to-line group and a line-to-neutral group at block 440. The agglomerative cluster method described above may be utilized to group electricity meters with a high average hybrid index into the line-to-line group and electricity meters with a low average hybrid index into the line-to-neutral group. A predicted phase is assigned to each meter based on the highest correlation at block 442. For the line-to-line group, the predicted phase is the one with a highest correlation in $mean(PCC_i^{K3})$, for $i = 4, 5, 6$, and for the line-to-neutral group, the predicted phase is the one with a highest correlation in $mean(PCC_i^{K3})$, for $i = 1, 2, 3$. The predicted phase may then be output for comparison with the existing meter-phase connectivity record.

- [0029]** FIG. 5 illustrates an example detail process of block 408 of FIG. 4.

- [0030]** At block 502, three largest clusters of electricity meters are determined. For all possible number of clusters from three to the number of meters in the sample, the largest three clusters, from large to small, L1, L2, and L3 are determined. At block 504, a ratio of the third largest cluster size to the largest cluster size, $R_{1to3} = \frac{size_{L3}}{size_{L1}}$, is calculated. At block 506, the lowest possible number of clusters, $\min N_{Clusters}$, such that R_{1to3} is greater than a preselected criteria, is determined, which ensures that the three initial kernels obtained are not too imbalanced. For example, for the preselected criteria of 0.5, the largest cluster is no larger than

twice the size of the smallest cluster. At block 508, the agglomerative cluster method may be utilized to group the electricity meters into the min $N_{Clusters}$ calculated in block 506. At block 510, the three largest clusters are selected as the three initial kernels, and the process proceeds to block 410.

5 [0031] FIG. 6 illustrates an example detail process of block 416 of FIG. 4.

[0032] At block 602, a predetermined range, for example from 0.75 to 0.85, of *Hybrid Index^{K1}* is evaluated in a predetermined increment, for example, 0.01, and a cutoff value of *Hybrid Index^{K1}* is determined at block 604. The cutoff value of *Hybrid Index^{K1}* may be defined as a value of *Hybrid Index^{K1}* below which there exist a first sufficient number of electricity
10 meters for each phase and, above which there exist a second sufficient number of electricity meters, where the first and second sufficient numbers may be preselected. At block 606, electricity meters with *Hybrid Index^{K1}* value lower than the cutoff value are selected as the elements for three new kernels, K2, and median of each phase is calculated and defined as three new kernels, K2, at block 608. The process then proceeds to block 418.

15 [0033] FIG. 7A illustrates an example display 700 of the clusters of electricity meters. Clusters of electricity meters may be displayed when each meter is plotted in 3D coordinates based on its correlation to phases A, B and C. The display 700 is a 2D view of the 3D plot viewed from the point (1,1,1) facing the origin (0,0,0) as shown by a graphical representation 702. FIG. 7B illustrates an example display 704 of phases of the electricity meters of FIG. 7A
20 plotted over the locations of the electricity meters on a map.

[0034] FIG. 8 illustrates an example block diagram of a system 800 for identifying electrical phase.

[0035] The system 800 may comprise one or more processors (processors) 802 communicatively coupled to memory 804. The processors 802 may include one or more
25 central processing units (CPUs), graphics processing units (GPUs), both CPUs and GPUs, or other processing units or components known in the art. The processors 802 may execute computer-executable instructions stored in the memory 804 to perform functions or operations, with one or more of components communicatively coupled to the one or more processors 802 and the memory 804, as described above with reference to FIGs. 4-7. For example, the memory
30 804 may store a phase analysis application 806 that is executed for analyzing the phases as described above with reference to FIGs. 4-7. Depending on the exact configuration of the system 800, the memory 804 may be volatile, such as RAM, non-volatile, such as ROM, flash memory, miniature hard drive, memory card, and the like, or some combination thereof. The

memory 804 may store computer-executable instructions that are executable by the processors 802.

[0036] The components of the system 800 coupled to the processors 802 and the memory 804 may comprise a user interface (UI) 806, including a display 808, and a communication module 810. The communication module 810 may communicate with a plurality of electricity meters 812 to receive the voltage time series data collected as discussed above with reference to FIGs. 4-6, as indicated by an arrow 814. Additionally, or alternatively, the electricity meters 812 may communicate with a central office 816 of the utility provider, or a third party, as shown by an arrow 818, and the central office 816 may collect the voltage time series data. The central office 816 may communicate the collected voltage time series data to the communication module 810 as shown by an arrow 820. While the communications 814, 818, and 820 between the communication module 810 and the electricity meters 812, the electricity meters 812 and the central office 816, and the central office 816 and the communication module 810, respectively, are shown as wireless communications, the communications 814, 818, and 820 may be established in various ways, such as via a cellular network, Wi-Fi network, cable network, landline telephone network, and the like.

[0037] While not shown, each of the electricity meters 812 may comprise one or more processors, memory coupled to the processors, a metrology module coupled to the processors, and a communication module coupled to the processors. The processors may include one or more central processing units (CPUs), graphics processing units (GPUs), both CPUs and GPUs, or other processing units or components known in the art. The processors may execute computer-executable instructions stored in the memory to perform functions or operations with one or more of components communicatively coupled to the one or more processors and the memory, such as measuring the voltage and storing voltage time series data in the memory or transmitting to the central office 816 or to the communication module 810 of the system 800.

[0038] Depending on the exact configuration of the electricity meter 812, the memory may be volatile, such as RAM, non-volatile, such as ROM, flash memory, miniature hard drive, memory card, and the like, or some combination thereof. The memory may store computer-executable instructions that are executable by the processors. The electricity meter 812 may receive instructions from the central office 816 regarding the preselected collection time period and the preselected interval, for example, changing the collection time period to two years and the interval to two minutes.

[0039] Some or all operations of the methods described above can be performed by execution of computer-readable instructions stored on a computer-readable storage medium, as defined below. The terms “computer-readable medium,” “computer-readable instructions,” and “computer executable instruction” as used in the description and claims, include routines, applications, application modules, program modules, programs, components, data structures, algorithms, and the like. Computer-readable and -executable instructions can be implemented on various system configurations, including single-processor or multiprocessor systems, minicomputers, mainframe computers, personal computers, hand-held computing devices, microprocessor-based, programmable consumer electronics, combinations thereof, and the like.

[0040] The computer-readable storage media may include volatile memory (such as random-access memory (RAM)) and/or non-volatile memory (such as read-only memory (ROM), flash memory, etc.). The computer-readable storage media may also include additional removable storage and/or non-removable storage including, but not limited to, flash memory, magnetic storage, optical storage, and/or tape storage that may provide non-volatile storage of computer-readable instructions, data structures, program modules, and the like.

[0041] A non-transitory computer-readable storage medium is an example of computer-readable media. Computer-readable media includes at least two types of computer-readable media, namely computer-readable storage media and communications media. Computer-readable storage media includes volatile and non-volatile, removable and non-removable media implemented in any process or technology for storage of information such as computer-readable instructions, data structures, program modules, or other data. Computer-readable storage media includes, but is not limited to, phase change memory (PRAM), static random-access memory (SRAM), dynamic random-access memory (DRAM), other types of random-access memory (RAM), read-only memory (ROM), electrically erasable programmable read-only memory (EEPROM), flash memory or other memory technology, compact disk read-only memory (CD-ROM), digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other non-transmission medium that can be used to store information for access by a computing device. In contrast, communication media may embody computer-readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave, or other transmission mechanism. As defined herein, computer-readable storage media do not include signals such as communication media.

5 [0042] The computer-readable instructions stored on one or more non-transitory computer-readable storage media, when executed by one or more processors, may perform operations described above with reference to FIGs. 4-7. Generally, computer-readable instructions include routines, programs, objects, components, data structures, and the like that perform particular functions or implement particular abstract data types. The order in which the operations are described is not intended to be construed as a limitation, and any number of the described operations can be combined in any order and/or in parallel to implement the processes.

10 **CONCLUSION**

[0043] Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or acts described. Rather, the specific features and acts are disclosed as exemplary forms of implementing the claims.

15

CLAIMS

WHAT IS CLAIMED IS:

1. A method comprising:
 - calculating voltage correlations of meter-to-meter combinations of a plurality of
5 electricity meters based on voltage time series data collected over a preselected collection time
period, each electricity meter of the plurality of electricity meters connected to one of six
phases comprising three line-to-neutral phases and three line-to-line phases;
clustering the plurality of electricity meters into three initial kernels representing the
three line-to-neutral phases based on voltage correlations;
10 for each of the three initial kernels, calculating correlation values with each electricity
meter of the plurality of electricity meters;
determining three new kernels based on the correlation values;
clustering the plurality of electricity meters into three groups based on a hybrid index
for the three new kernels calculated based on average correlation values associated with each
15 electricity meter;
forming six new kernels of electricity meters representing the six phases based on the
average correlation values associated with each electricity meter; and
assigning a predicted phase to an electricity meter of the plurality of electricity meters
based on corresponding correlation values of the electricity meter with each of the six new
20 kernels based on the voltage time series data.
2. The method of claim 1, wherein the voltage time series data is, at least one of:
 - collected by the plurality of the electricity meters,
 - stored by the plurality of the electricity meters,
 - 25 transmitted by the plurality of electricity meters to, and stored by, a central office of a
utility service provider, or
 - obtained from the central office.
3. The method of claim 1, wherein the voltage time series data includes a
30 preselected analysis period of voltage data associated with an electricity meter of the plurality
of electricity meters taken over a preselected collection time period at a preselected interval.

4. The method of claim 1, further comprising, prior to calculating the voltage correlations of meter-to-meter combinations:

omitting problematic data of the voltage time series data, wherein the problematic data include:

- 5 voltage values deviating more than a preselected value from an expected average value,
- constant voltage over longer than a predetermined period,
- missing data over more than a preselected number of preselected intervals, and
- 10 data from an electricity meter having location information inconsistent with an actual geographical location of the electricity meter.

5. The method of claim 1, wherein clustering the plurality of electricity meters into the three initial kernels comprises:

- 15 determining three largest clusters of electricity meters for all clusters having at least three electricity meters and up to all electricity meters of the plurality of electricity meters;
- determining a ratio of a third largest cluster size of the three largest clusters to a largest cluster size of the three largest clusters;
- determining a lowest number of clusters, such that the ratio is greater than a preselected
- 20 criteria;
- grouping the plurality of electricity meters into the lowest number of clusters; and
- selecting three largest clusters as the three initial kernels.

6. The method of claim 1, wherein determining the three new kernels based on the correlation values comprises:

- 25 evaluating hybrid index values of a predetermined range for the three initial kernels, a hybrid index value calculated as a ratio of a median correlation value of the three initial kernels to a maximum correlation value of the three initial kernels;
- determining a cutoff value of the hybrid index below which there are more than a first
- 30 preselected number of electricity meters for each phase and above which there are more than a second preselected number of electricity meters; and
- selecting electricity meters with hybrid index values below the cutoff value as elements for the three new kernels.

7. The method of claim 1, wherein clustering the plurality of electricity meters into three groups comprises:

- 5 clustering a group of electricity meters with a high hybrid index as a line-to-line phase group,
- clustering electricity meters with a low hybrid index as a line-to-neutral phase group,
- and
- clustering electricity meters with in-between hybrid index values as a band separating the high and low hybrid index groups.

10

8. The method of claim 7, wherein forming the six new kernels of electricity meters representing the six phases comprises;

- grouping electricity meters having the high hybrid index into three line-to-line phase groups based on average correlations of a phase with each electricity meter;
- 15 grouping electricity meters having the low hybrid index into three line-to-neutral phase groups based on highest average correlations of a phase with each electricity meter; and
- forming the six new kernels of electricity meters by combining the three line-to-line phase groups and the three line-to-neutral phase groups.

20

9. A system comprising:

one or more processors; and

memory communicatively coupled to the one or more processors, the memory storing thereon computer executable instructions that, when executed by the one or more processors, cause the one or more processors to perform operations comprising:

25

calculating voltage correlations of meter-to-meter combinations of a plurality of electricity meters based on voltage time series data collected over a preselected collection time period, each electricity meter of the plurality of electricity meters connected to one of six phases comprising three line-to-neutral phases and three line-to-line phases;

30

clustering the plurality of electricity meters into three initial kernels representing the three line-to-neutral phases based on voltage correlations;

for each of the three initial kernels, calculating correlation values with each electricity meter of the plurality of electricity meters;

determining three new kernels based on the correlation values;
clustering the plurality of electricity meters into three groups based on a
hybrid index for the three new kernels calculated based on average correlation
values associated with each electricity meter;
5 forming six new kernels of electricity meters representing the six phases
based on the average correlation values associated with each electricity meter;
and
assigning a predicted phase to an electricity meter of the plurality of
electricity meters based on corresponding correlation values of the electricity
10 meter with each of the six new kernels based on the voltage time series data.

10. The system of claim 9, wherein the voltage time series data is, at least one of:
collected by the plurality of the electricity meters,
stored by the plurality of the electricity meters,
15 transmitted by the plurality of electricity meters to, and stored by, a central office of a
utility service provider, or
obtained from the central office.

11. The system of claim 9, wherein the voltage time series data includes a
20 preselected analysis period of voltage data associated with an electricity meter of the plurality
of electricity meters taken over a preselected collection time period at a preselected interval.

12. The system of claim 9, wherein the operations further comprise, prior to
calculating the voltage correlations of meter-to-meter combinations:
25 omitting problematic data of the voltage time series data, wherein the problematic data
include:
voltage values deviating more than a preselected value from an expected average value,
constant voltage over longer than a predetermined period,
missing data over more than a preselected number of preselected intervals, and
30 data from an electricity meter having location information inconsistent with an actual
geographical location of the electricity meter.

13. The system of claim 9, wherein clustering the plurality of electricity meters into the three initial kernels comprises:

determining three largest clusters of electricity meters for all clusters having at least three electricity meters and up to all electricity meters of the plurality of electricity meters;

5 determining a ratio of a third largest cluster size of the three largest clusters to a largest cluster size of the three largest clusters;

determining a lowest number of clusters, such that the ratio is greater than a preselected criteria;

grouping the plurality of electricity meters into the lowest number of clusters; and

10 selecting three largest clusters as the three initial kernels.

14. The system of claim 9, wherein determining the three new kernels based on the correlation values comprises:

15 evaluating hybrid index values of a predetermined range for the three initial kernels, a hybrid index value calculated as a ratio of a median correlation value of the three initial kernels to a maximum correlation value of the three initial kernels;

determining a cutoff value of the hybrid index below which there are more than a first preselected number of electricity meters for each phase and above which there are more than a second preselected number of electricity meters; and

20 selecting electricity meters with hybrid index values below the cutoff value as elements for the three new kernels.

15. The system of claim 9, wherein clustering the plurality of electricity meters into three groups comprises:

25 clustering a group of electricity meters with a high hybrid index as a line-to-line phase group,

clustering electricity meters with a low hybrid index as a line-to-neutral phase group, and

30 clustering electricity meters with in-between hybrid index values as a band separating the high and low hybrid index groups.

16. The system of claim 15, wherein forming the six new kernels of electricity meters representing the six phases comprises;

grouping electricity meters having the high hybrid index into three line-to-line phase groups based on average correlations of a phase with each electricity meter;

grouping electricity meters having the low hybrid index into three line-to-neutral phase groups based on highest average correlations of a phase with each electricity meter; and

5 forming the six new kernels of electricity meters by combining the three line-to-line phase groups and the three line-to-neutral phase groups.

17. A non-transitory computer-readable storage medium storing thereon computer executable instructions that, when executed by one or more processors of an electricity meter,
10 cause the one or more processors to perform operations comprising:

calculating voltage correlations of meter-to-meter combinations of a plurality of electricity meters based on voltage time series data collected over a preselected collection time period, each electricity meter of the plurality of electricity meters connected to one of six phases comprising three line-to-neutral phases and three line-to-line phases;

15 clustering the plurality of electricity meters into three initial kernels representing the three line-to-neutral phases based on voltage correlations;

for each of the three initial kernels, calculating correlation values with each electricity meter of the plurality of electricity meters;

determining three new kernels based on the correlation values;

20 clustering the plurality of electricity meters into three groups based on a hybrid index for the three new kernels calculated based on average correlation values associated with each electricity meter;

forming six new kernels of electricity meters representing the six phases based on the average correlation values associated with each electricity meter; and

25 assigning a predicted phase to an electricity meter of the plurality of electricity meters based on corresponding correlation values of the electricity meter with each of the six new kernels based on the voltage time series data.

18. The non-transitory computer-readable storage medium of claim 17, wherein the
30 operations further comprise, prior to calculating the voltage correlations of meter-to-meter combinations:

omitting problematic data of the voltage time series data, wherein the problematic data include:

voltage values deviating more than a preselected value from an expected average value,

constant voltage over longer than a predetermined period,

missing data over more than a preselected number of preselected intervals, and

data from an electricity meter having location information inconsistent with an actual geographical location of the electricity meter.

19. The non-transitory computer-readable storage medium of claim 17, wherein clustering the plurality of electricity meters into three groups comprises:

clustering a group of electricity meters with a high hybrid index as a line-to-line phase group,

clustering electricity meters with a low hybrid index as a line-to-neutral phase group, and

clustering electricity meters with in-between hybrid index values as a band separating the high and low hybrid index groups.

20. The non-transitory computer-readable storage medium of claim 19, wherein forming the six new kernels of electricity meters representing the six phases comprises;

grouping electricity meters having the high hybrid index into three line-to-line phase groups based on average correlations of a phase with each electricity meter;

grouping electricity meters having the low hybrid index into three line-to-neutral phase groups based on highest average correlations of a phase with each electricity meter; and

forming the six new kernels of electricity meters by combining the three line-to-line phase groups and the three line-to-neutral phase groups.