CLAIMS

WHAT IS CLAIMED IS:

1. A method 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;

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;

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 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 electricity meters,

stored by the plurality of electricity meters,

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 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:

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.

5. The method of claim 1, 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;

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

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:

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

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:

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.

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;

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.

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:

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;

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;

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 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 electricity meters,

stored by the plurality of electricity meters,

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 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:

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.

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;

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

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:

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

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:

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.

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

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, 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;

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;

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 kernels based on the voltage time series data.

18. The non-transitory computer-readable storage medium of claim 17, wherein the 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.