Strathprints Institutional Repository

Blair, Steven M. and Booth, Campbell D. and Williamson, Gillian and Poralis, Alexandros and Turnham, Victoria (2016) Automatically detecting and correcting errors in power quality monitoring data. IEEE Transactions on Power Delivery. ISSN 0885-8977 (In Press) , http://dx.doi.org/10.1109/TPWRD.2016.2602306

This version is available at http://strathprints.strath.ac.uk/57466/

Automatically Detecting and Correcting Errors in Power Quality Monitoring Data

Steven M. Blair, *Member, IEEE*, Campbell D. Booth, Gillian Williamson, Alexandros Poralis, and Victoria Turnham

Abstract—Dependable power quality (PQ) monitoring is crucial for evaluating the impact of smart grid developments. Monitoring schemes may need to cover a relatively large network area, yet must be conducted in a cost-effective manner. Real-time communications may not be available to observe the status of a monitoring scheme or to provide time synchronization, and therefore undetected errors may be present in the data collected. This paper describes a process for automatically detecting and correcting errors in PQ monitoring data, which has been applied in an actual smart grid project. It is demonstrated how to: unambiguously recover from various device installation errors; enforce time synchronization between multiple monitoring devices and other events by correlation of measured frequency trends; and efficiently visualize PQ data without causing visual distortion, even when some data values are missing. This process is designed to be applied retrospectively to maximize the useful data obtained from a network PQ monitoring scheme, before quantitative analysis is performed. This work therefore ensures that insights gained from the analysis of the data—and subsequent network operation or planning decisions—are also valid. A case study of a UK smart grid project, involving wide-scale distribution system PQ monitoring, demonstrates the effectiveness of these contributions. All source code used for the paper is available for reuse.

Index Terms—Distribution systems, harmonics, power quality, time synchronization.

I. Introduction

POWER quality (PQ) monitoring is important for measuring the effectiveness of smart grid implementations and trial projects. Wide-scale monitoring is increasingly feasible, and these schemes can provide utilities and customers with information regarding the operation of converter-connected devices, non-conventional loads, energy storage, and novel automation and control systems [1], [2].

The validity of PQ monitoring data must be verified before any conclusions can be confidently drawn from these data; this is particularly important for trial projects, where it is critical to maximize the understanding of novel technologies and methods. Furthermore, although accurate PQ monitoring devices are becoming more cost-effective [3], large-scale monitoring schemes may be expensive to implement, may operate for a limited time, and may not have real-time communications for managing the operation of the scheme. A relatively simple error, such as an incorrect current sensor polarity, can affect

S.M. Blair and C.D. Booth are with the Institute for Energy and Environment, Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1RD, UK (e-mail: steven.m.blair@strath.ac.uk).
G. Williamson and A. Poralis are with WSP | Parsons Brinckerhoff, Manchester Technology Centre, Manchester, M1 7ED, UK.
V. Turnham is with Electricity North West Limited, Technology House, Salford, M6 6AP, UK.

multiple derived measurements and can entail a complex and ad hoc post-processing effort to correct the initial error, to avoid distorting the results; as noted in [4], the complexity of PQ data management and analysis can be more costly than the equipment. If undetected, such an error could result in invalid conclusions and actions being taken. It is therefore important that these errors can be detected and, where possible, a clear and robust procedure must be defined to correct these errors in order to maximize the useful data extracted from the monitoring scheme. It is also important that PQ monitoring data can be readily visually inspected and therefore that the data are graphically represented faithfully, without distortion or aliasing.

There is a significant body of work which discusses methods for: automated event detection, classification, and feature extraction from PQ monitoring data [5]–[8]; the removal of outliers or "bad" data [9], [10]; and indexing and selecting appropriate PQ monitoring data for later analysis work [11]. However, this paper focuses on a different, but fundamental, issue: comprehensively validating that the raw PQ data are correct. Existing standards such as IEC 61000-4-30 describe appropriate PQ measurement methods, but this paper addresses salvaging the maximum useful data from a monitoring scheme which has already been completed.

There are three main contributions in this paper: 1) a novel process for counteracting accidental monitoring device installation errors is defined in Section II; 2) validation and correction of time synchronization for multiple PQ devices, as described in Section III-C; and 3) the analysis of methods for correctly and efficiently visualizing PQ data, given in Section IV. The successful application of these methods is demonstrated in Section V in the context of a recent large-scale smart grid trial project.

II. Correcting PQ Monitoring Device Installation Errors

A. Feasibility

It is possible for PQ monitoring device installation errors to significantly affect the collected data, thereby producing misleading results from subsequent analysis of the data if undetected. This problem is especially pertinent to large-scale monitoring schemes which—to minimize cost—do not have remote communications to validate the device functionality during the course of the monitoring scheme. Potential installation errors include [12]: wrong sensor polarity; wrong voltage or current sequence; correct voltage and current sequence, but inconsistent phase order (e.g., V_a, V_b, V_c and I_b, I_c, I_a);

TABLE I
FEASIBILITY OF AUTOMATED DETECTION AND CORRECTION OF
INSTALLATION ERRORS

Potential installation error	Can be detected?	Can be corrected?
Voltage phase A does not match system phase A	No	No
Incorrect voltage sequence	Yes	Yes[1]
Current phase A does not match voltage phase A	Yes[2]	Yes[2]
Incorrect current sequence	Yes[2]	Yes[2]
Incorrect current polarity	Yes[2]	Yes[2]

[1]If excessive negative sequence voltage is measured, the measurements (both voltage and current) from phases B and C can be swapped to correct the sequence. However, it is not possible to evaluate if the phases are connected to the correct system phases without additional information (such as correlation of a specific triggered event with another device).
[2]With certain assumptions and limitations, as described in Section II-B.

or combinations of these issues. These issues may not be identified until the monitoring scheme has been completed. However, it is possible to detect and correct some of these problems retrospectively, based on assumptions of the normal operation of three-phase systems. For example, it can be assumed that three-phase voltages and currents should be predominantly positive-sequence, and that the direction of real power flow should be consistent for all phases. Other rules could be applied based on knowledge of the system being monitored.

Table I summaries several types of installation errors and identifies those which can be automatically detected and corrected. Without additional information, it is not possible to validate that the voltage connections are absolutely correct; only the voltage sequence can be validated. However, it is assumed that voltage connection errors are less likely than for current sensor connections (due to current sensors possibly lacking color-coding and the ease with which they can be "clamped-on" with the wrong polarity), and that voltage polarity errors are not possible for systems with a star/wye configuration where there is a common neutral point. The process for resolving current sensor connection errors is described in Section II-B.

B. Detailed Current Phase Validation and Correction Process

It is assumed that each PQ monitoring device records the magnitude and angle of the fundamental component of the voltage and current in each phase. Typically, the phase angles will be given relative to the measured phase A voltage. Therefore, the current phasors, as depicted in Fig. 1, can be tested for their expected locations. There are six permutations of current phase connections (e.g., I_a, I_b, I_c; I_b, I_c, I_a; etc.) and a further 2^3 possibilities for the polarity of each phase, leading to a total of 48 configurations (two of which are illustrated in Fig. 1a). It is possible to unambiguously restore the correct phase configuration from any possible erroneous configuration. The detailed process is described in Fig. 2. It is important that the current magnitude measurements are within the sensor's rated range because otherwise the phase response of the sensor may be unreliable. The initial presence of a phase connection error or sensor polarity error can be detected by comparing

(a) Unity power factor

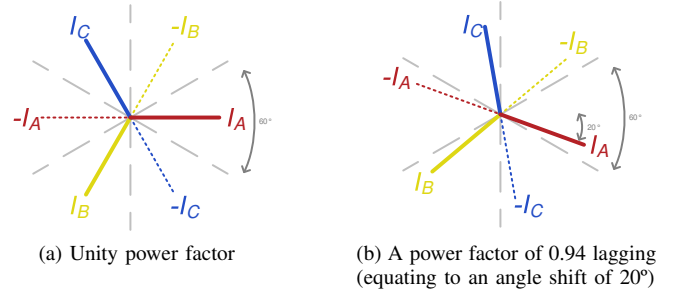(b) A power factor of 0.94 lagging (equating to an angle shift of 20º)

Fig. 1. Expected regions for current phasors with different power factors (for both correct and incorrect sensor polarity)

the negative sequence current magnitude to a threshold (such as >0.5 pu, relative to the rated current) or by the angle for any phase being outside the expected region given in Fig. 1a (i.e. within $\pm 30°$ for phase A, 210° to 270° for phase B, and 90° to 150° for phase C). The sensitivity of this check can be improved by ignoring the magnitudes of the current phasors in the negative sequence calculation (by assuming 1 pu magnitude), thereby excluding the impact of current unbalance which is common in three-phase LV networks [13]. If an error exists, the phase can be reassigned based on the expected locations given in Fig. 1a.

This approach applies to systems where, for all three phases:

- The real power flow is consistently in one direction i.e. there are no significant distributed generation connections on just one or two phases,
- The power factor is greater than approximately 0.866 (30º) leading or lagging, and
- The power factor is similar.

As demonstrated in Fig. 1b, it is not possible to distinguish between a correct phase configuration with relatively low power factor (<0.866) and certain erroneous configurations.

This process can be repeated for every time-step in the sampled data or, if it is known to represent the remaining data consistently (i.e., there were no changes to the sensor configuration over time), the first sample can be used to define the pattern for all of the data. In general, it is advisable to remove the initial samples and to delete "isolated" samples at the start of the capture. This is because the device may be switched on and off several times during the initial commissioning and testing (during which the connections may not be correct). If the power factor is known to be greater than the threshold of 0.866 (30º), which is typical for residential demand, this process can be fully automated; otherwise manual reassignment of the phase connections and polarities may be required.

For the large-scale monitoring scheme described in Section V, installation errors were automatically detected in 21 out of 77 PQ monitoring devices. In 18 cases, the errors can be resolved automatically and the data can be confidently used for further analysis. In the three remaining cases, the presence of low (or cyclically-low) demand affects the accuracy of the current sensors (which were rated for a minimum of 50 A) and therefore distorts the measurements, despite the phase connections likely being correct.
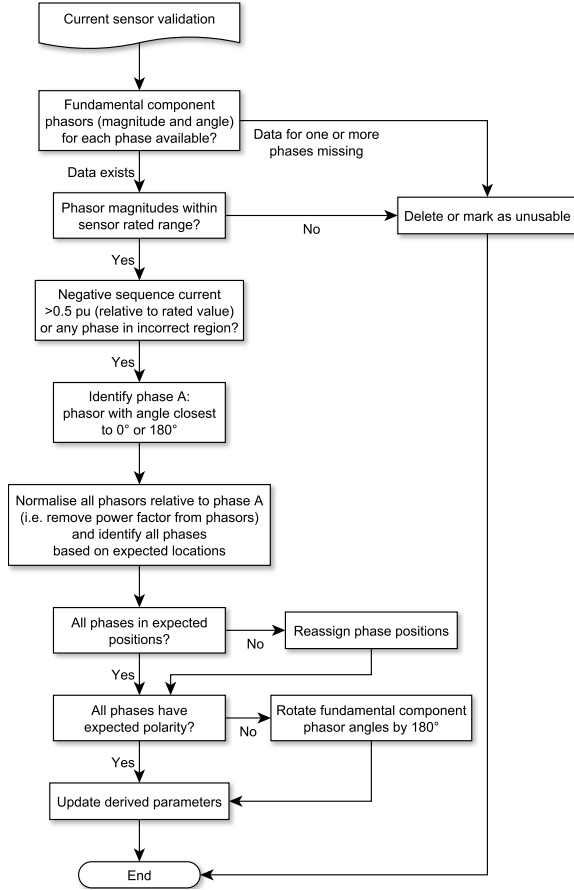
Fig. 2. Current sensor connection validation and correction

C. Updating Derived Parameters

Parameters which are derived from the raw voltage and current measurements will also be affected by device installation errors but, in some cases, can similarly be corrected automatically. Depending on the type of error, the following corrective action should be applied:

1) If a phase connection was determined to be incorrect (e.g., connected in negative sequence), then the individual phase data values for all captured data (e.g., THD, real power, etc.) should be reassigned. Furthermore, the positive and negative sequence values, if recorded, should be recalculated from the new individual phase values.

2) If a current sensor polarity is determined to be incorrect:

   a) Invert the measured current values for the affected phase and recalculate all derived parameters, such as real power and power factor.

   b) If the monitoring device records minimum and maximum values of the affected phase over the sampling period, these should be inverted and swapped. An illustrative example for a single phase is given in Fig. 3.

   c) The three-phase average values for each parameter, if recorded, should also be recalculated from the new individual phase values. Some derived parameters, such as the minimum and maximum negative sequence current, cannot be recovered for certain

types of errors because they are calculated over the entire measurement reporting period; these values should be discarded.

III. VALIDATING AND CORRECTING TIME SYNCHRONIZATION

A. Overview of Method

It is important to have confidence in the time-stamps of PQ data when comparing data from multiple measurement locations, or when relating measurements to other network events such as smart grid automation and control interventions (as required for the case study in Section V). To reduce cost, each monitoring device may not be fitted with accurate GPS-based or communications-based time synchronization methods, and instead will rely on a local clock within the device for time-stamping measurements. The clock may have been set incorrectly and may drift over time.

However, assuming that the monitoring devices record measured frequency, it is possible to retrospectively verify the time synchronization by correlation of the measured frequency trends over time [14], because measured frequency should be approximately the same at all locations throughout an interconnected AC power system at any instant in time. Furthermore, if a device's clock is found to be inaccurate, it is possible to calculate the time offset and compensate for this when reading data. A simplified illustration of this method is given in Fig. 4, where the clock time offset, Δt, exhibited by Monitor 3 can be corrected by comparison with the measured frequency trends from Monitor 1 and Monitor 2.

B. Frequency Trend Correlation

By assuming that, at a given instant in time, the majority of PQ monitoring devices retain accurate clocks, it is possible to correlate the frequency measurements from a single device with the mean value across all devices. If the frequency measurement from device i out of n at time t is $f_{i,t}$, the mean frequency measurement from all devices at time t is m_t as follows:

$$m_t = \frac{\sum_{i=1}^{n} f_{i,t}}{n} \qquad (1)$$

In other words, m_t is the mean of each "column" of frequency measurements at an instant in time (as shown graphically in Fig. 5). Therefore, for convenience, vectors of all device frequency measurements (F_i) and the corresponding mean values (M) within a window size, w, can be calculated as follows:

$$F_i = \begin{bmatrix} f_{i,1} & \cdots & f_{i,w} \end{bmatrix} \qquad (2)$$
$$M = \begin{bmatrix} m_1 & \cdots & m_w \end{bmatrix} \qquad (3)$$

A window size of 1 day gives $w = 288$ samples for 5-minute sampling. The correlation coefficient [14], ρ, between the 288 frequency measurements for a given monitoring device and the 288 mean values can be calculated as follows, where $\overline{X}$ is the mean of a given vector X, and σ_X is the standard deviation of X:
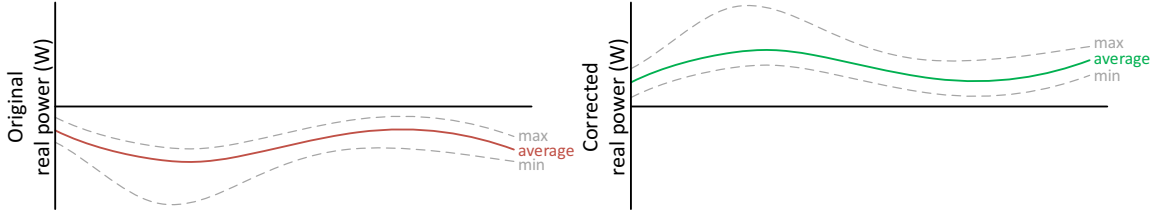
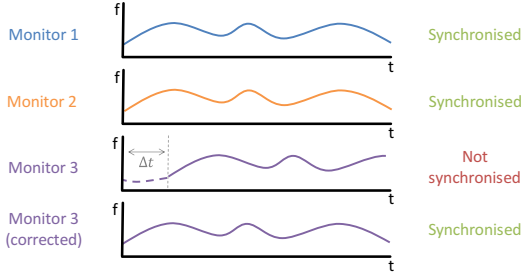Fig. 3. Example of real power average, minimum, and maximum correction

Fig. 4. Overview of frequency trend synchronization

$$\rho\left(F_i, M\right) = \frac{\sum_{t=1}^{w}\left(f_{i,t} - \overline{F_i}\right)\left(m_t - \overline{M}\right)}{(w-1)\,\sigma_{F_i}\,\sigma_M} \quad (4)$$

A correlation coefficient value of $\rho\left(F_i, M\right) = 1.0$ signifies that the two frequency trends are fully positively linearly correlated. A threshold of $\rho\left(F_i, M\right) > 0.9$ has been used to determine the condition for monitoring device i being sufficiently correlated with the mean, such that the clock of device i can be trusted. The output of this process is a day-by-day assessment of the "trustworthiness" of the clock for each device. A minimum number of monitoring locations must have data available for each window before the time synchronization detection process is attempted; a threshold of 20 locations has been used for the case study in Section V.

Fig. 5 illustrates typical results for a window size of 1 day (with most columns not shown for brevity) of real field data from the case study in Section V. The frequency measurements are 5-minute average aggregate values, and each shaded column represents a 5-minute period. The entire distribution of frequency measurements is color-coded as a heat map between white (relatively low frequency) and dark-orange (relatively high frequency). The mean of each "column" of frequency measurements is also shown. The final row contains "third-party" measurements of the UK system frequency, obtained from [15]; the time-stamps for both sources of data have been aligned (although the capture rates do not precisely match which affects the correlation results). Note that the mean and third-party measurement rows each have independent heat maps from the device frequency measurement window.

In the majority of cases, the frequency measurements are well-correlated—both with each other (i.e., with the mean), and with the third-party frequency measurement. This is indicated by each column being approximately the same shade, and this pattern is consistent across the full time range of data. Therefore, this confirms that these monitoring device internal

clocks have been set correctly and are reliable within approximately 5 minutes (i.e. the accuracy of the time synchronization depends on the measurement period), which is sufficiently accurate for the power quality analysis tasks described in Section V. In some cases, such as monitoring locations 1, 2, and 16 in Fig. 5, the frequency measurement is not consistently aligned. The following section describes how the data can be salvaged by automatically correcting the time-stamps. A further improvement to the yield of usable measurements can be obtained by assuming that relatively short gaps in time synchronization validity (which may occur due to the device being briefly unpowered), between periods which have been validated, can also be "trusted".

C. Automatically Correcting Time Offsets

As noted in Section III-A, the measured frequency trends for some devices may not correlate with the other measurements, implying that these devices' internal clocks have not been set consistently. It is possible to correct the clocks for these devices by detecting the clock offset. This can be achieved by sliding the window of frequency measurements, as described in Section III-B, backwards or forwards in time until the correlation threshold is met; i.e., the window used for calculating F_i is adjusted, while M is held constant. To reduce execution time, the method used in the case study searches for the first occurrence which satisfies the correlation threshold. This avoids the need for a time-consuming exhaustive search of all possible offsets, particularly given that the majority of time offsets are within the relatively small offset of 1 hour.

A benefit of the proposed approach is that correlation of frequency measurements provides an absolute time reference (within the accuracy of the measurement period). As described in Section III-B, the process is performed continually to calculate the clock offset over time (e.g. an interval of one day). Therefore, the proposed method includes protection against clock drift due to a new offset being calculated at regular intervals for each PQ monitoring device.

IV. VISUALIZATION OF LARGE-SCALE PQ TIME-SERIES DATA

A. Requirements

It is important to be able to graphically view recorded PQ data in a simple and consistent way; along with the automated methods described in Sections II and III, this is important for validating and understanding PQ data. Analysis of the PQ data may also require visualization of short periods (e.g., one day) or long-term trends (e.g., multiple years).

Fig. 5. Heat map of typical frequency trend correlation results from the case study field data

For visualizing trends over a relatively long period of time, it is critical to properly re-sample the time-series data. For example, a sampling period of 5 minutes should result in over 100,000 captured values per year, per measurand, which must be down-sampled for plotting annual trends. This is essential for practical reasons (even high-resolution computer displays have a relatively small number of horizontal pixels and cannot meaningfully represent very granular data) and for efficient computational performance (to avoid time-consuming operations on unnecessarily-large data sets).

Therefore, the re-sampling algorithm must be capable of substantially reducing the number of data points, without excessive computational time or complexity. The algorithm must avoid aliasing and must deal with missing data elegantly. Furthermore, some PQ monitoring devices record mean, maximum, and minimum values over the sampling period, which are important for understanding actual system characteristics such as accurately determining maximum demand; each data type must be re-sampled appropriately to preserve this information. Care must be taken to ensure that the re-sampling algorithm correctly "centers" each data point on the appropriate time-stamp. An appropriate and efficient data storage mechanism is also required for achieving responsive, real-time data queries for visualization applications involving relatively large quantities of PQ data [16].

B. Re-sampling of PQ Data

Table II compares various approaches for re-sampling PQ data, and graphical results are given in Fig. 6 for a target number of samples of 500. The preferred approaches (algorithms 2 and 3), which are accurate and practical, use the "pandas" library [17] for the Python programming language which is designed to support the robust manipulation of time-series data. Similar results can also be obtained using the Time Series functionality in MATLAB. Any method which involves decimation of the data (i.e., blindly keeping every n^{th} sample, such as algorithm 1) is liable to introduce aliasing; trends which are not present in the original data may be extracted. However, this can be mitigated by first applying a moving average filter or a low-pass filter (LPF). LPF-based methods require the filter parameters (i.e., the order and cut-off frequency) to be specified, and a zero-phase digital filter must be used to avoid erroneous phase distortion (i.e., a time shift in the resulting re-sampled PQ data).

Algorithms 2 and 3 are recommended for use in PQ visualization tools, and an example implementation is available at [18]. This visualization tool, which applies algorithm 2, is illustrated in Fig. 7 and can be accessed online at [19]. The tool is able to dynamically and efficiently re-sample PQ data based on the desired time range; for example, seasonal variation in demand is evident in Fig. 7. By substantially reducing the number of sampling points, depending on the required view of the data, both the network data transfer latency and the software plotting times are significantly reduced.

C. Missing and Spurious Data Points

Following verification of monitoring device installation and time synchronization, there are other potential errors in the time-series data recorded by each monitoring device which can be addressed:

- Missing values, due to loss of communications, device failure, or other factors. Depending on the extent of the missing data, it is possible to interpolate the missing values; however, care must be taken to observe daily, weekly, and annual trends when interpolating. Single missing values can be safely interpolated from adjacent (healthy) values, and a linear interpolation should be sufficiently accurate. Only mean measurements can be

TABLE II
COMPARISON OF PQ DATA RE-SAMPLING ALGORITHMS

Algorithm	Number of data points can be selected	Applicable to different plot types (mean, min, and max)	Relative computational speed	Suitable for representing both short and long periods	Implementation complexity	Deals with missing data elegantly	Visually acceptable results
1) Decimation	Yes	Yes, but results are distorted	Very fast	Yes	Very simple	Yes	No
2) "pandas" re-sampling	Yes, by selecting an appropriate re-sampling rate	Yes	Medium	Yes	Simple; must ensure data are correctly centered on time-stamps	Yes	Yes
3) "pandas" moving average, decimated	Yes, by specifying the window size	Yes (different functions are used for mean, min, and max)	Fast	Yes	Simple; must ensure data are correctly centered on time-stamps	Yes	Yes
4) LPF, decimated	Yes	No	Fast	Yes, but appropriate filter parameters must be specified	Medium; filter parameters must be specified and zero-phase filtering must be used	Yes	Yes (for mean plot types only)
5) Cubic spline interpolation	Yes	No	Slow	Not suitable for long periods	Complex	No	No
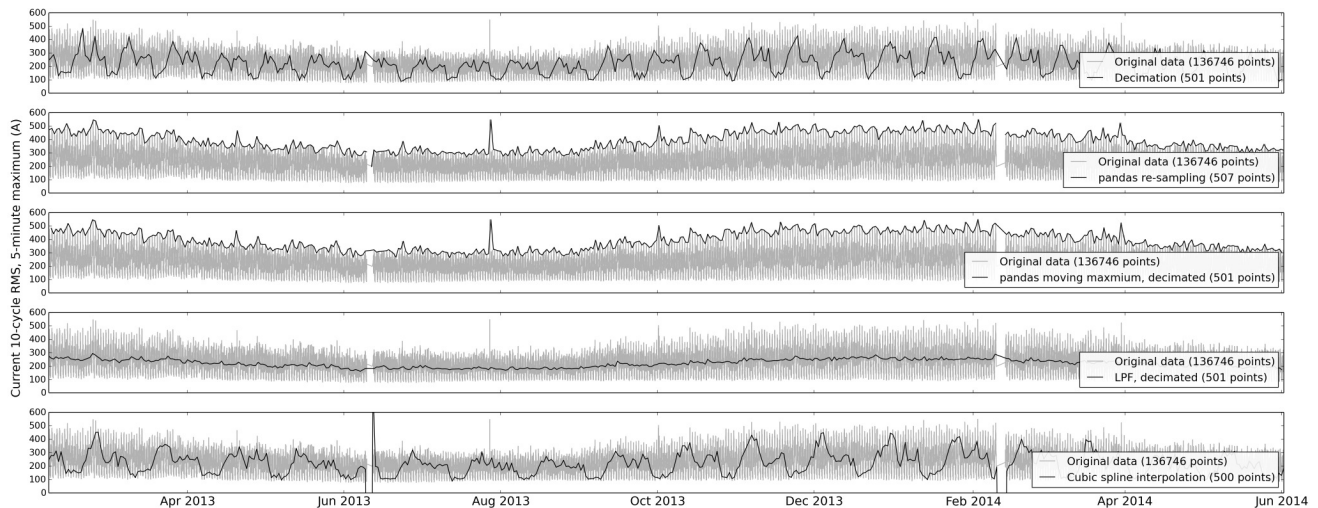
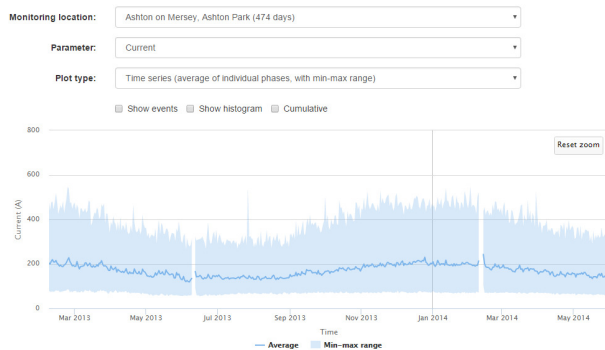Fig. 6. Re-sampling algorithm results

Fig. 7. Example of PQ monitoring data visualization tool

interpolated; missing minimum and maximum values cannot be meaningfully estimated. An iterative process may be needed: first determine all suspicious data, then attempt interpolation from healthy data.

- Other obvious erroneous values, such as invalid values

during device initialization and single or multiple data points "frozen" at a constant value. A method such as Chebyshev's inequality can be used to detect such erroneous values, and its use for filtering demand data is described in [20]. This is especially useful if the original data set is relatively unreliable [21]. However, care must be taken not to confuse genuine anomalies, such as fault currents or voltage sags, as errors.

V. CASE STUDY SMART GRID PROJECT

A. Project Overview

The Capacity to Customers (C$_2$C) project has been completed by Electricity North West Limited (ENWL), a distribution network operator (DNO) in the UK, with several industrial and academic partners. The project combined large-scale demand-side response contracts, new post-fault automation schemes, and interconnected network operation to maximize the potential for new load and generation connections. A trial
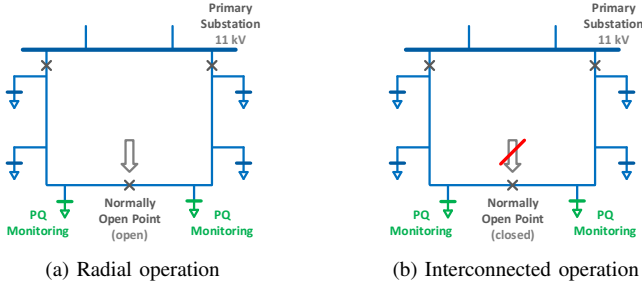
(a) Radial operation   (b) Interconnected operation

Fig. 8. C$_2$C trial circuit operation modes

has been conducted over an extensive area of ENWL's network [22]. The operation of 6.6/11 kV circuits in interconnected mode (Fig. 8b) rather than conventional radial mode (Fig. 8a) has the potential to affect PQ, particularly voltage total harmonic distortion (THD), due to the change in effective impedance of the circuits and the aggregation of harmonic distortion from the two radial circuits. Therefore, it was important to thoroughly establish the impact of the interconnected circuit operation on PQ, such that the potential wide-scale deployment of the C$_2$C approach can be accepted.

B. PQ Comparison Methodology

A total of 77 "PQube" [23] PQ monitoring devices have been deployed on a representative subset of the C$_2$C project trial circuits for approximately 18 months. The devices have been installed on the LV side of 6.6/11 kV to 400 V transformers in secondary (or distribution) substations. Using the PQ monitoring data captured during the trial, several measured system parameters have been compared to ascertain any differences that are apparent as a result of operating in either radial configuration or interconnected configuration. Therefore, the effects of interconnected operation on PQ—if any—can be quantified. It is assumed that week-by-week demand is typically similar, and therefore that the relevant power quality metrics can be compared fairly by selecting one week of data before and after the change of state of a Normally Open Point (NOP). By analyzing a log of NOP open/close commands from the DNO's control room, "valid" events have been extracted where the state of the NOP is consistent for one week before and one week after the NOP state change.

C. Data Validation Requirements

It is essential to validate the monitoring data so that any conclusions being drawn from the measurements are sound and fair. In particular, it is critical that the internal clock of each PQ monitoring device is relatively accurate—within a few minutes of a known, absolute time reference such as Coordinated Universal Time (UTC)—and reliable. This is because comparisons must be made between radial and interconnected network configurations; this involves aligning monitoring data to independently time-stamped switching events from the DNO's control room logs. The following steps have been performed to validate the monitoring data:

1) Extract NOP switching events (i.e., transitions between radial and interconnected circuit operation modes) from

TABLE III
VALID EVENTS AND CORRESPONDING MONITORING DATA

Stage	Description	Number of suitable events	Number of monitoring locations included
1	NOP state change events with valid date range	123	n/a
2a	Monitoring data available within date range	114	57
2b	Continuous data available from at least one monitoring location on circuit	83	51
3a	Valid time synchronization	78	49
3b	Valid time synchronization, with additional "trusting"	81	49
4	Demand variation check	52	34

the DNO's control room logs for each monitored circuit and determine the events which occur in valid date ranges.
2) Determine the data availability profile for each monitoring location, i.e., the dates where the PQ monitor was operational and correctly recorded data.
3) Validate clock synchronization for all PQ monitoring devices using the method in Section III. Where possible, realign clocks, and the corresponding measurement data, which exhibit a time offset. In general, the PQ devices which experienced clock errors were typically off by one hour, most likely due to the clock being initially set to local daylight saving time, rather than being set according to UTC.
4) Elimination of monitoring locations for NOP state change events where the mean difference in demand between the two weeks to be compared is greater than 5%. This ensures that no unusual events occurred during one of the two weeks, which may skew the results. This threshold has been chosen by examining the distribution of the differences in mean demand between the two weeks of data; in the majority of cases, the difference in demand is within 5%.

Table III summarizes the number of valid events at each stage of the validation process. Although a relatively small percentage (42%) of the initial events are ultimately usable in the numerical analysis, the results in Section V-D can be considered to be very robust due to the comprehensive validation process.

D. Data Analysis and Results

The data from all valid events have been used to quantify of the extent of the change between radial and interconnected operation, if any, for various PQ metrics. These metrics are: THD, Total Demand Distortion (TDD), short-term flicker (Pst), and long-term flicker (Plt). For example, the difference in the weekly mean THD measurements, δTHD, has been calculated as follows:

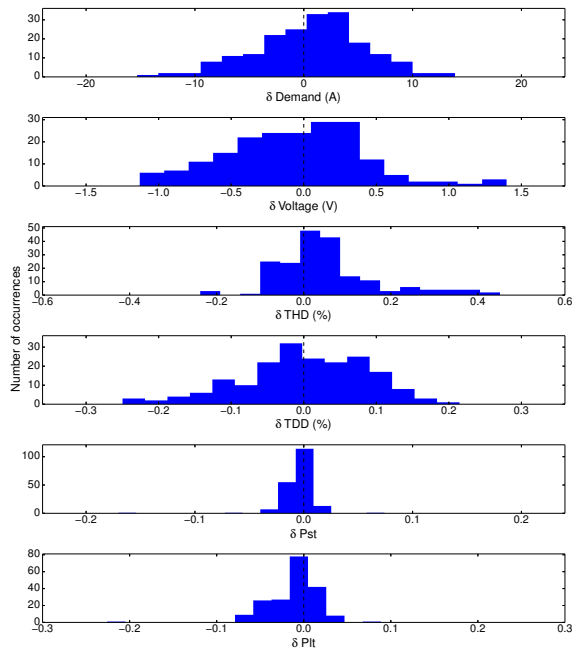$$\delta THD = mean(THD_{interconnected}) - mean(THD_{radial}) \quad (5)$$

Fig. 9. Comparison of change in PQ measurements (interconnected operation "minus" radial operation)

Therefore, a positive δTHD value represents generally higher THD for interconnected operation compared with radial operation. The results for all circuits, for all PQ metrics, are aggregated as a histogram in Fig. 9. For simplicity, the results for each phase—which are calculated individually—are combined in the distributions given in Fig. 9. It can be observed that interconnected operation has no significant impact on the PQ of the trial circuits; these results are described in more detail in [21].

The error detection and correction process defined in this paper has maximized the useful data captured during the case study project. In particular, the successful validation of each PQ monitoring device's clock was crucial for associating monitoring data with time-stamped network automation events, with a high level of confidence.

VI. CONCLUSIONS

When the results of PQ monitoring may affect decision-making, the actions of automation and control systems, or the learning from trial projects, it is important to assume the worst-case and to treat measurements as potentially erroneous—until comprehensively verified. This paper has described a process for automatically detecting and correcting several types of errors which may be present in PQ monitoring data. All source code used in this paper is available at [18], [24], [25]. Although some of the errors described in this paper can be avoided by adopting a rigorous installation procedure for PQ monitoring devices, it is possible that some errors may unintentionally be present—and this may not be known until the completion of the monitoring scheme. It is essential that PQ visualization tools correctly re-sample data and cater for missing values. The effectiveness of the proposed methods has been demonstrated through its application in a smart grid project in the UK.

REFERENCES

[1] C. Foote, G. Burt, I. Wasiak, R. Mienski, R. Pawelek, P. Gburczyk, and M. Thoma, "A Power-Quality Management Algorithm for Low-Voltage Grids With Distributed Resources," *IEEE Trans. Power Deliv.*, vol. 23, no. 2, pp. 1055–1062, 2008.

[2] CIGRE/CIRED Joint Working Group C4.112, *Guidelines for Power Quality Monitoring: Measurement Locations, Processing and Presentation of Data*, 2014.

[3] A. J. Roscoe, T. Sklaschus, G. Oldroyd, S. M. Blair, and G. M. Burt, "Measurement of 40 power system harmonics in real-time on an economical ARM Cortex-M3 platform," *Electron. Lett.*, vol. 49, no. 23, pp. 1475–1476, nov 2013.

[4] J. Kilter, S. Elphick, J. Meyer, and J. V. Milanovic, "Guidelines for Power quality monitoring - Results from CIGRE/CIRED JWG C4.112," in *2014 16th Int. Conf. Harmon. Qual. Power*. IEEE, may 2014, pp. 703–707.

[5] W. Anis Ibrahim and M. Morcos, "Artificial intelligence and advanced mathematical tools for power quality applications: a survey," *IEEE Trans. Power Deliv.*, vol. 17, no. 2, pp. 668–673, apr 2002.

[6] R. C. Dugan, M. F. McGranaghan, S. Santoso, and H. W. Beaty, *Electrical Power System Quality*. McGraw-Hill Education, 2012.

[7] M. Tuljapurkar and A. Dharme, "Wavelet Based Signal Processing Technique for Classification of Power Quality Disturbances," in *2014 Fifth Int. Conf. Signal Image Process*. IEEE, jan 2014, pp. 337–342.

[8] S. Li and X. Wang, "Cooperative Change Detection for Online Power Quality Monitoring," dec 2014. [Online]. Available: http://arxiv.org/abs/1412.2773

[9] G. Heydt and E. Gunther, "Post-measurement processing of electric power quality data," *IEEE Trans. Power Deliv.*, vol. 11, no. 4, pp. 1853–1859, 1996.

[10] V. Terzija and V. Stanojevic, "Power quality indicators estimation using robust Newton-type algorithm," *IEE Proc. - Gener. Transm. Distrib.*, vol. 151, no. 4, p. 477, 2004.

[11] Z. Yang, J. Cao, Y. Xu, H. Zhang, P. Yu, and S. Yao, "Data Cleaning for Power Quality Monitoring," in *2013 Fourth Int. Conf. Netw. Distrib. Comput*. IEEE, dec 2013, pp. 111–115.

[12] K. D. Jones, A. Pal, and J. S. Thorp, "Methodology for Performing Synchrophasor Data Conditioning and Validation," *IEEE Trans. Power Syst.*, vol. 30, no. 3, pp. 1–10, 2014.

[13] A. Navarro, T. Gozel, L. Ochoa, R. Shaw, and D. Randles, "Data analysis of LV networks: Determination of key parameters from one year of monitoring over hundreds of UK LV feeders," in *Int. Conf. Electr. Distrib.*, 2015.

[14] M. Huijbregtse and Z. Geradts, "Using the ENF Criterion for Determining the Time of Recording of Short Digital Audio Recordings," in *Proc. 3rd Int. Work. Comput. Forensics*, ser. Lecture Notes in Computer Science, vol. 5718, Berlin, Heidelberg, aug 2009, pp. 116–124.

[15] GridWatch, "UK National Grid status download data," 2014. [Online]. Available: http://www.gridwatch.templar.co.uk/download.php

[16] M. P. Andersen, S. Kumar, C. Brooks, A. von Meier, and D. E. Culler, "DISTIL: Design and implementation of a scalable synchrophasor data processing system," in *2015 IEEE Int. Conf. Smart Grid Commun*. IEEE, nov 2015, pp. 271–277.

[17] "pandas - Python Data Analysis Library," 2015. [Online]. Available: http://pandas.pydata.org/

[18] S. M. Blair, "C2C Power Quality Monitoring Visualisation," 2016. [Online]. Available: https://github.com/stevenblair/pq_monitoring_visualisation

[19] ——, "C2C power quality monitoring," 2015. [Online]. Available: http://c2c.eee.strath.ac.uk

[20] X. Chen, C. Kang, X. Tong, Q. Xia, and J. Yang, "Improving the Accuracy of Bus Load Forecasting by a Two-Stage Bad Data Identification Method," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1–8, 2014.

[21] S. M. Blair and C. Booth, "Analysis of the Technical Performance of C2C Operation for HV Networks," University of Strathclyde, Tech. Rep., dec 2014.

[22] S. M. Blair, C. D. Booth, P. Turner, and V. Turnham, "Increasing Distribution Network Capacity using Automation to Reduce Carbon Impact," in *Dev. Power Syst. Prot.* IET, 2014.

[23] Power Standards Lab, "PQube - Power Quality and Energy Monitor." [Online]. Available: http://www.powerstandards.com/PQube.php

[24] S. M. Blair, "PQube Data Extraction," 2016. [Online]. Available: https://github.com/stevenblair/pqube_data_extraction

[25] ——, "C2C Power Quality Analysis," 2016. [Online]. Available: https://github.com/stevenblair/c2c_pq_analysis

Steven M. Blair (M'09) received the M.Eng. degree (with distinction) in computer and electronic systems and the Ph.D. degree in electrical engineering from the University of Strathclyde, Glasgow, UK in 2008 and 2013, respectively.

Dr Blair is presently a Research Associate at the Institute for Energy and Environment, Department of Electronic and Electrical Engineering, University of Strathclyde. He is a member of IEC Technical Committee 57 and CIGRE Task Force B5.02. His research interests include power system protection, communications, power quality, real-time simulation, and fault current limitation.

Campbell D. Booth received the B.Eng. and Ph.D. degrees in electrical and electronic engineering from the University of Strathclyde, Glasgow, UK in 1991 and 1996, respectively. He is currently a Reader with the Institute for Energy and Environment, Department of Electronic and Electrical Engineering, University of Strathclyde.

His research interests include power system protection; plant condition monitoring and intelligent asset management; applications of intelligent system techniques to power system monitoring, protection, and control; knowledge management; and decision support systems.

Gillian Williamson received the B.Eng. degree in electrical and electronic engineering from Salford University in 1991 and Ph.D. degree from the University of Manchester, UK in 1995.

She is currently a Principal Engineer within the Power Networks Department of WSP|Parsons Brinckerhoff, specialising in network innovation.

Alexandros Poralis received an engineering diploma (equivalent to an M.Eng. degree) in electrical and computer engineering from the National Technical University of Athens in 2010. In 2013, he received the M.Sc. degree in renewable energy systems technology from Loughborough University.

He is presently an assistant engineer in the Power Networks department at WSP | Parsons Brinckerhoff and has experience in innovation projects and power system studies. He is a member of the IET and the Technical Chamber of Greece.

Victoria Turnham biography not available at time of publication.