# Online Power Quality Anomaly Detection and Diagnosis using Bayesian Changepoint and Waveform Similarity Methods

Amy Anderson, Dr Xu Jiang, Dr Bruce Stephen & Prof Stephen McArthur

University of Strathclyde Glasgow

Institute for Energy and Environment
Department of Electronic and Electrical Engineering
University of Strathclyde
Glasgow G1 1RD
United Kingdom

# Distribution Network Observability

- Distribution networks are extensive
  - 1000’s of km of service cables + 1000’s of connected loads + ??000 generators
- Want better quality of supply? Where to start? no monitoring.
- Now have micro phasor measurement units (µPMU)
  - MHz sampling rates
  - One monitoring point – look *down* the network for faults
  - Very high data rates – can easily overwhelm a human operator
- How to make operational sense of that data?
  - No time/expertise to label faults
  - Don’t want to employ someone to constantly watch the system…

# High-resolution fault analysis and diagnosis: solution outline

1. A novel anomaly segmentation method is proposed and developed, which can be used to pick up abnormal events from a high-frequency PQ data stream. This method is robust to the noise of non-linear loads. Compared to recent benchmarks, the detection accuracy is high, and the false alarm rate is low.

2. A new approach is used to diagnose fault cause using waveform data recorded at distribution level substations. This method is built on a novel waveform similarity metric and associated context, which does not require feature selection. It is shown to outperform conventional classifiers with only a minimal set of exemplar faults.

3. An innovative fault labelling method is proposed for distribution level fault records. This provides a possible way to form labelled historical fault records, which can then be used to train intelligent classifiers and also increase fault diagnostic accuracy.

   - The use of automatically generated labels only reduces classification accuracy by around 5% in comparison to manually labelled data on an operationally obtained data set; however, if deployed in practice, this could promote the use of intelligent classifiers without the burden of having to manually label fault exemplars, making it a more scalable option.

# High-resolution fault analysis and diagnosis: solution outline

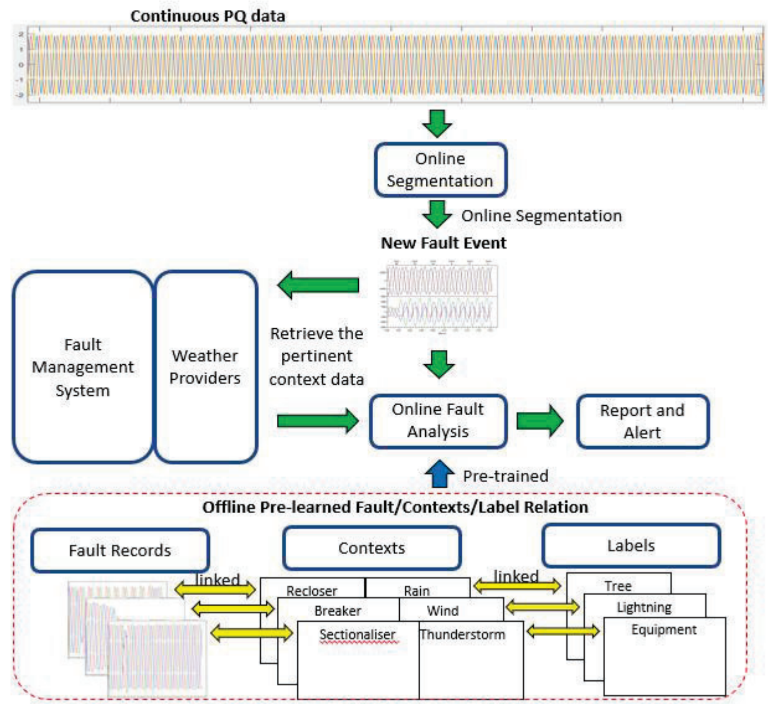
**Input**: PQ data and associated context
**Output**: Fault cause

### Online System:

- online signal segmentation using Bayesian Changepoint Detection
- online fault analysis function, which uses pre-trained K-Nearest Neighbour (KNN) with a novel similarity metric to diagnose fault cause from waveform data

### Offline System:

- pre-learned fault/contexts/label relations (NLP) for automated training of fault diagnostic model from free-text maintenance reports



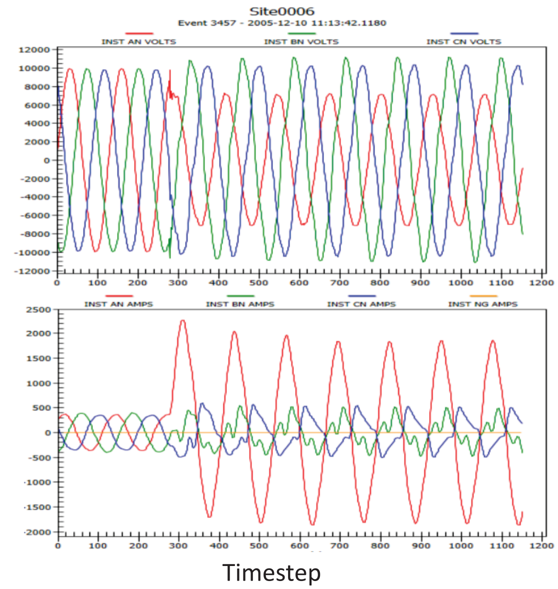
# Fault Manifestation

## PQ Fault Cause Signatures

# Fault Manifestation through waveform data

Particular faults resulting from equipment failure, weather or external disturbance manifest through waveform as:

- Cycle: number of affected and point on the cycle
- Changing harmonic content
- Number of phases affected
- Noise behaviour




Vehicle impact with pole

*-examples taken from EPRI/DoE Power Quality Disturbance library – 300+ examples taken from North American substations. See: K. M. Kittredge, J. P. Lennane and D. D. Sabin, "New publicly accessible online power quality monitoring databases," 2016 17th International Conference on Harmonics and Quality of Power (ICHQP), Belo Horizonte, 2016, pp. 137-141. doi: 10.1109/ICHQP.2016.7783379

X. Jiang, B. Stephen and S. McArthur, "Automated Distribution Network Fault Cause Identification With Advanced Similarity Metrics," in IEEE Transactions on Power Delivery, vol. 36, no. 2, pp. 785-793, April 2021, doi: 10.1109/TPWRD.2020.2993144.

# Fault Exemplars: EPRI/DoE PQ Library

- DoE data library includes 334 incipient faults and permanent faults
- This is a 'gold standard' data set – very useful but operationally not available! Or…
- Library records kHz sampling rate voltage and current signatures?
  - As would PQ monitoring equipment
- Protection operation?
  - Fault recorder would provide
- Weather at the time is recorded?
  - As would weather station
- Events are labelled by experts?
  - No…
- Field crew reports are also included?

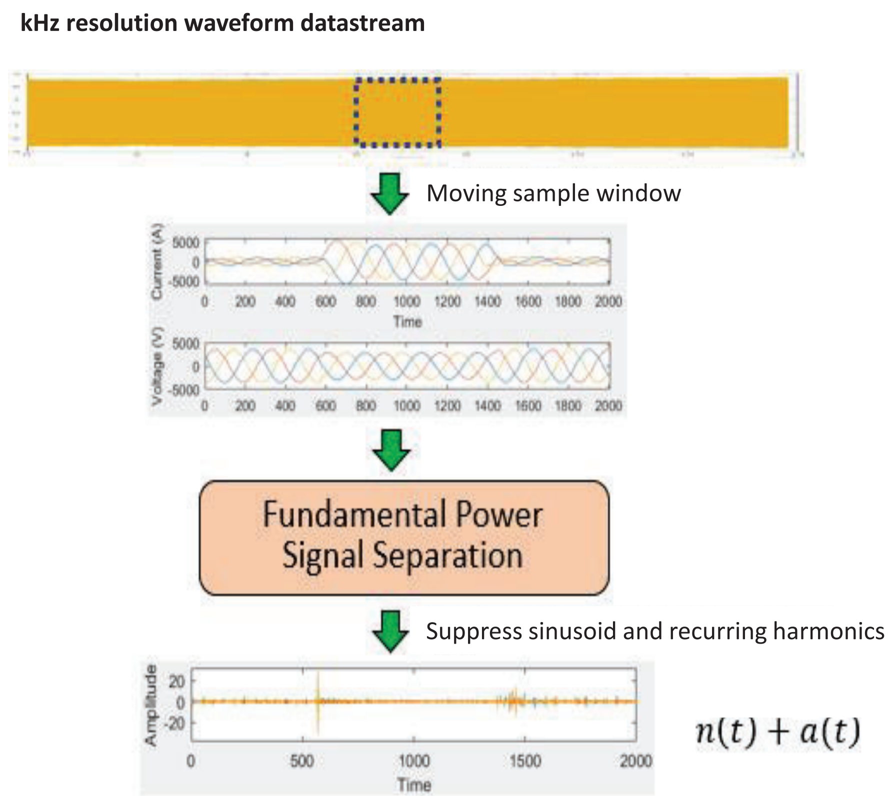
| EventId | Phase | Cause | Weather | Details (free text) |
|---|---|---|---|---|
| 0001 | 2 | Tree | Clear Weather | Fault caused line recloser lockout. Tree Outside Right of Way (Fall/Lean On Primary) |
| 0004 | 2 | Tree | Clear Weather | Fault caused line recloser lockout. Tree Outside Right of Way (Fall/Lean On Primary) |
| 0005 | 2 | Tree | Clear Weather | Fault caused line recloser lockout. Tree Outside Right of Way (Fall/Lean On Primary) |
| 0007 | 2 | Tree | Clear Weather | Fault caused line recloser lockout. Tree Outside Right of Way (Fall/Lean On Primary) |
| 3042 | 4 | Equipment | Unknown | Equipment, Device UG, Damaged. |
| 0021 | 1 | Equipment | Clear Weather | Overhead Insulator Failure. BROKEN INSULATOR |
| 0022 | 1 | Equipment | Clear Weather | Overhead Insulator Failure. BROKEN INSULATOR |
| 0062 | 4 | Undetermined | Raining | storm |
| 0064 | 4 | Undetermined | Raining | storm |
| 0067 | 4 | Tree | Thunderstorm | Tree/Limb Growth |
| 0065 | 4 | Tree | Thunderstorm | Tree/Limb Growth |
| 0068 | 2 | Tree | Clear Weather | VINES ON TRANSFORMER |
| 2760 | 1 | Unknown | Unknown | Short duration variation. No outage information found. |
| 3048 | 3 | Equipment | Unknown | Equipment, Capacitor Station, Damaged. |

# Segmentation

## Anomaly detection and duration estimate

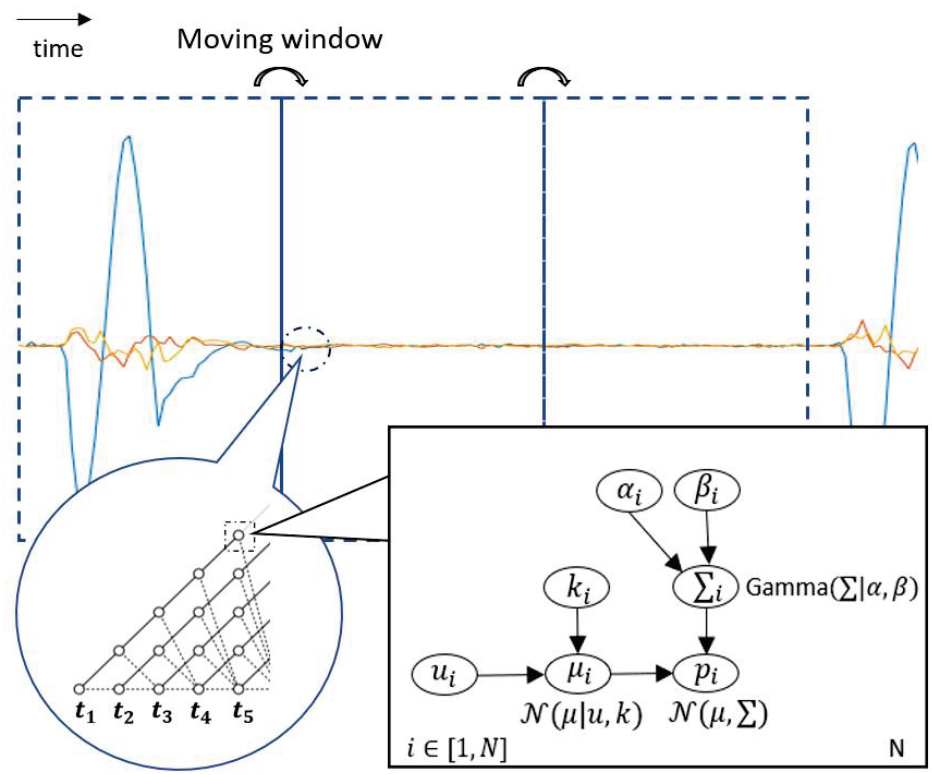
# Operational Sense from waveform data

- Generally know what normal behaviour looks like
  - Don't want to keep this
- Want to keep abnormal data
  - Need to identify when it starts…
  - …and when it finishes…
  - …and retain this segment
- Want to diagnose abnormal data in a segment
  - Fault or operational artefact?
- Need to segment first…how?



# Bayesian Changepoint Detection

- Segmentation of a high resolution stream of waveform data raises a number of considerations:
  - **Temporal dependency:** subsequent predictions to depend on the previous observations – fault signature not an isolated data point
  - **Weak delineation from normal behavior:** fault signature interleaved with normal operation – again – not isolated data point
  - **Online Operation**: no storage of data or time (or computational effort) spent training
  - **Adaptive thresholds:** need to set this in operational context – network anomalies take various forms (bigger/smaller/more variable/less variable) so a fixed threshold is not ideal
- Propose Bayesian Changepoint Power Quality Segmentation (BCPQS)
  - Model the evolving probability distributions of the PQ waveform noise at various points – not the signal
  - Capture low likelihood sequences of changes in a moving window




X. Jiang, B. Stephen and S. McArthur, "A Sequential Bayesian Approach to Online Power Quality Anomaly Segmentation," in IEEE Transactions on Industrial Informatics, vol. 17, no. 4, pp. 2675-2685, April 2021, doi: 10.1109/TII.2020.3003979.

# How/where to test?

No fully observed distribution networks

- Too expensive
- No operational justification

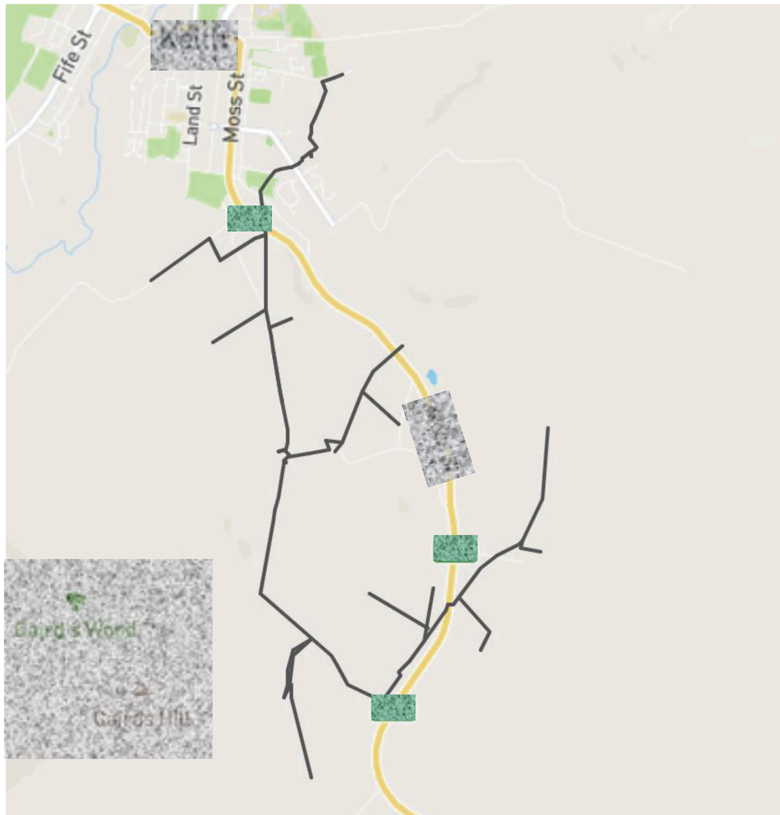
Start with simulation

- Entire (real) 11kV network with (real) substation loads
- Censor this to the point of realism

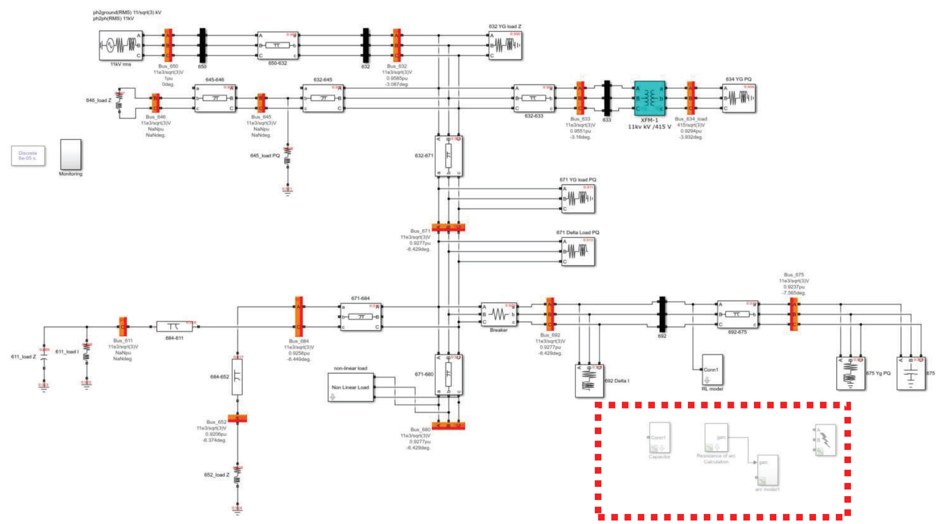
Simulate fault occurrence on network

- Injection of real fault data (e.g. tree contact)
- Physics based simulation (e.g. kizliky's arcing)

# 11kV Feeder

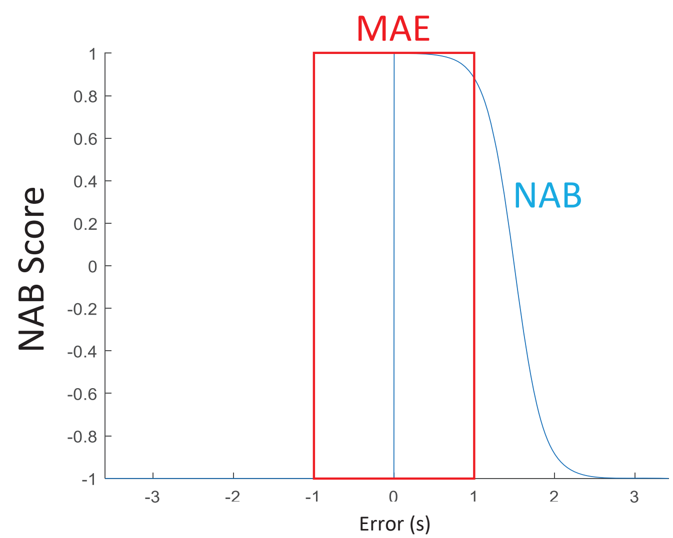



## POWER SYSTEM MODEL



# Segmentation Results

- Run network model with 271 randomly injected faults
  - And switching operations
- Test against a simple threshold based model
  - Differential RMS
- MSE tests how early/late the detection is
  - Low is good
- Numenta Anomaly Benchmark (NAB) penalizes early detection (as well as very late detection)
  - High is good




| Detector | MSE (s2) | NAB |
|---|---|---|
| | | |
| Differential RMS | 0.0697 | 71.67 |
| BCPQS | 0.0235 | 91.052 |

# Diagnosis

## Exemplar Fault Similarity

# Novel Similarity Measure

Anomalies are useful for bringing operator attention to network area

- They don't help with fault resolution

Two inputs available:

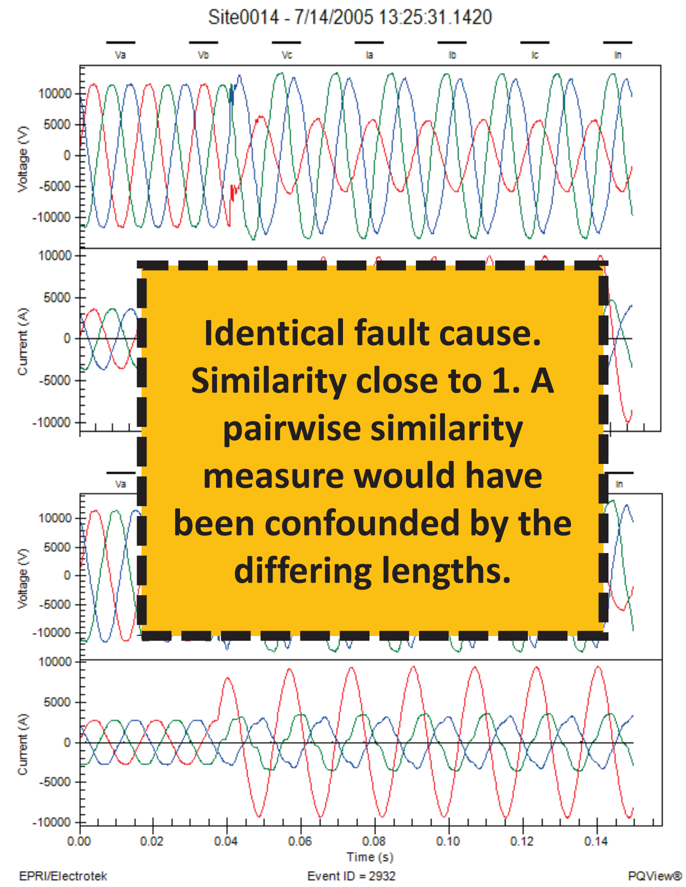
- Waveform
- Context (weather, datetime)

Get context vector similarity from Hamming distance

Get waveform similarity from Dynamic Time Warped series

- Handles differing lengths

Combine as normalised product



# Fault Diagnosis Accuracy

- New similarity measure using combination of waveform and context performs better than constituent measures on most fault types
- No feature selection used
- No prior knowledge required
- Faults relating to living things still pose problems

| Fault Cause | Waveform | Context | Combined |
|---|---|---|---|
| | | | |
| Tree | 75% | 69% | 89% |
| Equipment | 65% | 85% | 90% |
| Vehicle | 33% | 76% | 75% |
| Animal | 52% | 90% | 88% |
| Lightning | 61% | 65% | 94% |

Automated Distribution Network Fault Cause Identification with Advanced Similarity Metrics, X. Jiang, B. Stephen, S. D. J. McArthur, IEEE Transactions on Power Delivery, vol. 35, no. 3, pp. 1610-1613, June 2020, doi: 10.1109/TPWRD.2019.2947784.

# Automation

## Labelling exemplars for automated training

# Drawbacks…?

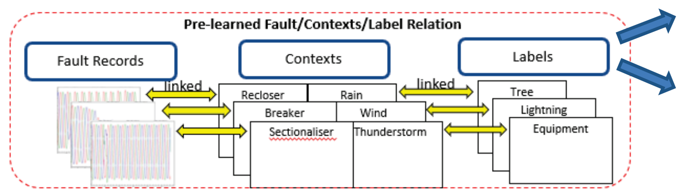
Reality is, to make a diagnostic classifier, need to learn from examples

Examples need to be labelled

- DNOs do not curate labelled data sets as a matter of operational practice

Could the work order/fault report be used?

- No: it's just notes – too ambiguous. Or is it…?




Expert labels?

Generated labels?

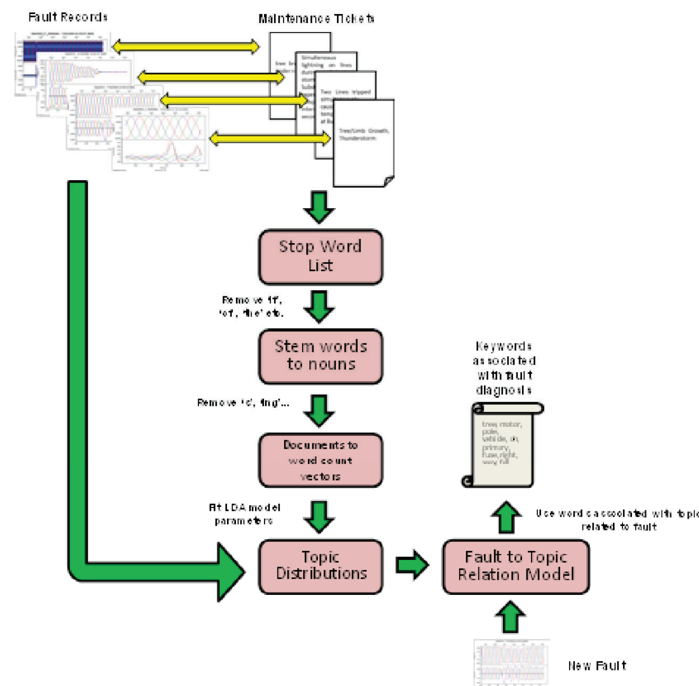
# Fault Cause Labelling with NLP

Automatic exemplars generation function, such as a Latent Dirichlet Allocation (LDA) model, to generate labelled data [Hindle11]

DoE data set already labelled

- Do the labels relate to the maintenance record ‘topic’?
- Learn the relation and find out…
- Get about 78% accuracy on a predictive model (topic->label) [Stephen21]

The performance of the automatic exemplars generation function will be evaluated through comparing the diagnosis result using two different resources





A. Hindle, M. W. Godfrey, and N. A. Ernst, “Automated Topic Naming to Support Cross-project Analysis of Software Maintenance Activities,” in *Proceedings of the 8th Working Conference on Mining Software Repositories*, 2011, pp. 163–172.

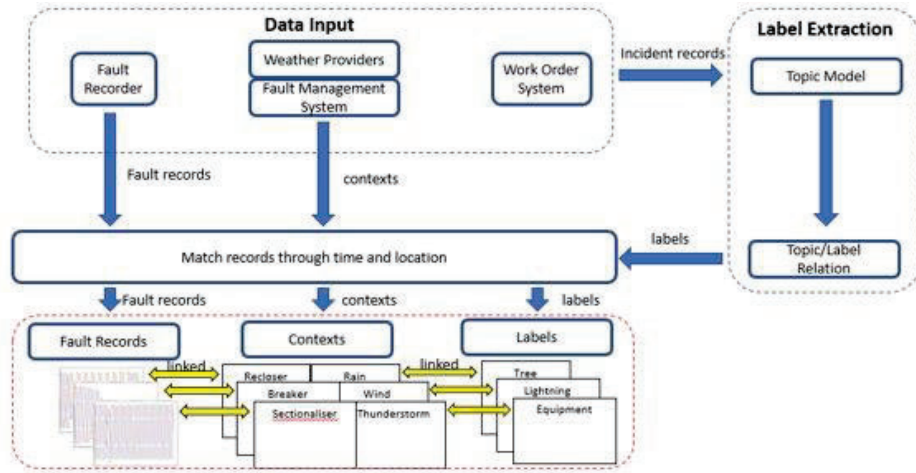
Extracting distribution network fault semantic labels from free text incident tickets, B. Stephen, X. Jiang, S. D. J. McArthur, IEEE Transactions on Power Delivery, vol. 36, no. 2, pp. 785-793, April 2021, doi: 10.1109/TPWRD.2020.2993144.

# End to End operation

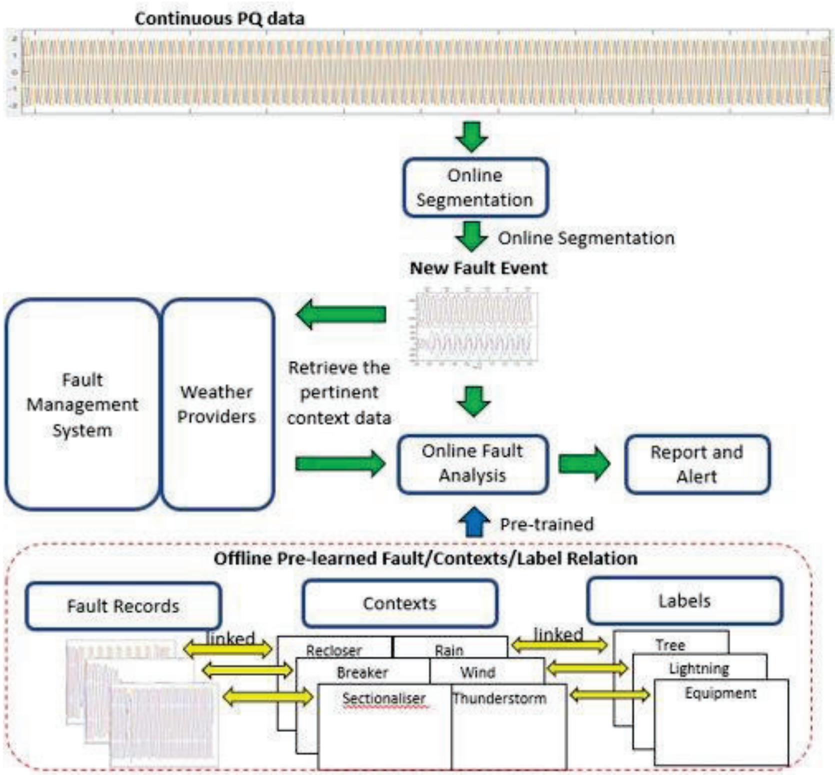
•**Input**: Free-text maintenance, waveform data and associated context data

•**Output**: Labeled faults

•**Offline System:** Waveform, associated context and labels are matched through time and location



# End to End performance




Automation of labelling results in a performance loss of ~5%

# Conclusion

- Analytics comprising an end to end PQ based fault diagnostic system have been developed, meaning:
  - In asset management terms:
    - Quantification of problem circuits
    - Informs quarterly spares budgets
  - In operational terms:
    - May be able to carry out pre-emptive switching (avoid loss of supply)
  - Longer term:
    - May justify investment in further monitoring
- Ultimately, improved service quality
- Next stage of development is implementation on hardware and deployment

# Thanks for listening!

IEEE PES Power & Energy Society®

IEEE

University of Strathclyde Glasgow