Adaptive Water Demand Forecasting for Near Real-time Management of Smart Water Distribution Systems

Presented by Gary Wong & Kleanthis Mazarakis
On behalf of Kevin Woodward & Michele Romano
Overview

- About United Utilities
- Business Challenges & Drivers
- Operational Technology & Innovation
- Why do Demand Forecasting?
- Why do a Proof-of-Concept (POC)?
- Demand Forecasting Methodology
- POC Components & Architecture
- Case Study & Results
- Summary
About United Utilities

WHERE WE OPERATE

[Map showing locations such as Manchester, Liverpool, Blackpool, Bolton, Preston, Lancaster, Carlisle, Chester, Warrington, and around 5,000 skilled employees.]

This is what it takes to serve 7 million customers every day...

- 57,000 hectares of catchment land
- 72,000 km of sewers
- 7 million customers
- 96 water treatment works
- 104 reservoirs
- 575 wastewater treatment works
- 1,400 km of aqueducts
- And around 5,000 skilled employees.
Business Challenges & Drivers

- Regulatory and customer commitments
- Changing environment
- ODIs, Leakage, Blockages
- Reduce Totex – Capex and Opex
- Apply intelligence and innovation
- Operational Technology providing analytics capability
Operational Technology (OT) at United Utilities

Operational Intelligence

OT – IT Convergence

Business Intelligence

Physical asset layer

SMART WATER NETWORKS

Industrial automation layer

Enterprise layer

Operational Technology

IT
Innovation and R&D

• Water Network Event Localisation
• Event Management for Intelligent Water Networks
• Wastewater Network Event Recognition
• Water Process Event Recognition
Why do Demand Forecasting (DF)?

• DF provides indications of what the future customer demand will look like

• This has several applications in United Utilities:
  – Pump Scheduling
  – Production Planning
  – Event Recognition
  – Dynamic Hydraulic Modelling
    … Many Others

• With associated business benefits:
  – Energy Savings
  – More efficient Operations
  – Better Customer Service
    … Many Others
Why do a Proof-of Concept? - Aims

• Understand how advanced analytics can be developed in an agile way
• Enable ready access to time series data
• Link time series data and analytics platforms
• Enable data pre-processing
• Enable development and testing of ANN models
• Understand potential for productionisation
• Demonstrate a live demand forecasting system
• Demonstrate accuracy and robustness of system
DF Methodology

Predicts water demands for the scheduling horizon (i.e., next 24 hours) to support near real-time operational management of smart Water Distribution Systems (WDSs)

- data-driven and continuously adaptive
- embedded in a piece of software developed in MATLAB
- use Evolutionary Artificial Neural Networks (EANNs)


Why EANNs?

• self-learning ability
• continuous adaptability to ever changing water demand patterns
• robustness of the DF models building process i.e. no arbitrary selection and use of various explanatory variables and/or different lags of the demand variable
• generic and seamless applicability to different demand signals
• dramatic reduction of the efforts required from a human expert to design an ANN model for a given problem whilst enabling replicating or outperforming the quality of the results achievable through human expert intervention

Practicality of the methodology operationalisation
DF Approaches

**ensemble EANNs (eEANNs)**

24 ANN models
- The models run in parallel
- ANNs with different forecasting horizon (i.e., 1 to 24 hours ahead)
- ANN models tailored to the particular forecasting horizon

**recursive EANNs (rEANNs)**

1 ANN model
- The runs are performed sequentially
- Fixed forecasting horizon (i.e., 1 hour ahead)
- Accuracy can decrease due to error accumulation
DF Methodology Components & Modes of Operation

Requirement for online DF:
• System must be continuously adaptive

Three main modes of operation:

- **Set up**: Runs when the module is initialised and when operational changes occur.
- **Update**: Runs every $n$ days (e.g., every 7 days).
- **Forecast**: Runs every $p$ hours (e.g., every hour).
Data Pre-processing

- Deals with the missing data
- Filters out outliers
- Assembles the ANN models’ training and test sets
- Provides a “median day” vector for each day of the week to be used as surrogate prediction when the ANN models cannot return an output (i.e., lack of incoming real-time data)
Data Cleansing

- The retrieved data are rearranged in \( m \) vectors (i.e., one for each day)
- Missing data in each vector are filled in
- Vectors are grouped according to their day of week (e.g., Monday, Tuesday, etc.)
- For each day of the week, a ‘median day’ vector and a vector of daily standard deviations are computed
- These vectors are then used by three statistical filters that aim to deal with outliers and other abnormal measurements

Flexibility in setting/finding a desired/most suitable working definition of “anomalous data”
Assembling Good ANN Training and Test Sets & DF Robustness

- A set of “good” flow values is assembled. This set will be used for training and testing the ANN models.
- A “median day” vector, for each day of the week, is computed. This will be used as surrogate prediction when the ANN models cannot return an output (i.e., lack of incoming real-time data).
ANN Short-term Demand Forecasting

• Trains and tests ANN models in the “Set up” and in the “Update” mode
• Uses the incoming flow data and the trained ANN models to predict the next 24 hour demand in the “Forecast” mode
ANN Structure

Input units

Hidden units

Output unit

Weights

Output layer

Hidden layer

Input layer
ANN Supervised Learning

- ANN is not programmed – it learns from the presented examples (training data)
- Iterative process
  - Input data are presented repeatedly to the ANN
  - With each presentation, the output of the ANN is compared to the desired output and an error is computed
  - This error is then fed back (back-propagated) to the ANN and used to adjust the weights such that the error decreases with each iteration and the ANN model becomes closer to producing the desired output
ANN Overfitting

The ANN learns the noise in the input as well as the common patterns, the result is poor performance on unseen examples (poor generalisation performance)

Overfitting occurs:
- when we train the network on a task for too long
- when we train the network on a task using too many hidden units
- when we train the network on a task using too many inputs

William of Occam, 1284-1347  “Do not multiply explanations unnecessarily”

Urban man, 1990  “Keep it simple, stupid”
EA Selection of the ANN Parameters and Inputs

- Finds the “best” set of parameters for training the ANN models
- Finds the “best” set of inputs to be used by the ANN models for accurate demand forecasting
EA-based Algorithm

The EA-based optimisation strategy automatically selects the input to the ANN models and their parameters

Characteristics:

• Single-objective (i.e., generalization error on the test set)
• Decision variables:
  o Lag size (i.e., number of past flow values in input to the ANN) - Range Of Variability (ROV): from 4 hours to 72 hours;
  o Time Of the Day - ROV: use/do not use;
  o Day Of the Week - ROV: use/do not use;
  o Number of hidden neurons - ROV: 10$\div$100;
  o Number of training cycles - ROV: 50$\div$500;
  o Value of the coefficient of Weight Decay Regularisation - ROV: $10^{-5} \div 10^3$;
POC Components

- PI Data Archive 2016 R2
- PI AF 2016 R2
- PI Integrator for BA 2016
- PI Interface for UFL
- PI to PI Interface

- PI Coresight 2016 R2
- PostgreSQL
- MATLAB
POC Architecture – can be customized to any database

Production Server → PI to PI [Historical & Real Time]

PI Data Archive → Contextualize

PII4BA → Text File

Text File → Any Database

PostgreSQL

MATLAB

ANN

PI Data Archive → PI UFL

Predictions Text File

Clean & Shape

Contextualize

POC Server
Prepare and Deliver Process Data

- Pull
- Cleanse
- Augment
- Shape
- Transmit

To any Visualization Tool or Analysis database on the ODBC standard

- OSIsoft
- Power BI
- HARMONY
- Tableau
- Qlik
- Cognos
- TIBCO Spotfire
- Oracle Business Intelligence
- memsql
- hadoop
Case Study

- 5 demand zones (DZs) with different sizes and characteristics e.g. light industrial, urban and rural
- Data for the period from 01/11/2016 to 15/03/2017
- Demand forecasting using the two alternative approaches (i.e. 1 ANN model & 24 ANN models)

<table>
<thead>
<tr>
<th>Demand Zone Name</th>
<th>Commercial Customers</th>
<th>Residential Customers</th>
<th>Mains Length (Km)</th>
<th>Average Demand (m3/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DZ1</td>
<td>8</td>
<td>327</td>
<td>5.1</td>
<td>5.4</td>
</tr>
<tr>
<td>DZ2</td>
<td>121</td>
<td>1751</td>
<td>10.7</td>
<td>30.8</td>
</tr>
<tr>
<td>DZ3</td>
<td>129</td>
<td>2078</td>
<td>16.4</td>
<td>57.1</td>
</tr>
<tr>
<td>DZ4</td>
<td>258</td>
<td>4156</td>
<td>70.2</td>
<td>122.3</td>
</tr>
<tr>
<td>DZ5</td>
<td>516</td>
<td>8312</td>
<td>118.5</td>
<td>215.6</td>
</tr>
</tbody>
</table>
Results - Excellent Trained & Tested ANN Models

<table>
<thead>
<tr>
<th>rEANNS</th>
<th>eEANNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nash-Sutcliffe index Training Set</td>
<td>0.98193</td>
</tr>
<tr>
<td>Nash-Sutcliffe index Test Set</td>
<td>0.98132</td>
</tr>
</tbody>
</table>

Performance Indicator: Nash-Sutcliffe index

- Normalized statistic
- Determines the relative magnitude of the residual variance compared to the measured data variance
- Ranges between $-\infty$ and 1
- Robust in terms of applicability to various models

Performance levels usually categorised as:

| > 0.65 | EXCELLENT |
| 0.65-0.5 | VERY GOOD |
| 0.5-0.2 | GOOD |
| < 0.2 | POOR |

Results - Excellent Quality Forecasts
Results v Aims

✓ Understand how advanced analytics can be developed in an agile way
✓ Enable ready access to time series data
✓ Link time series data and analytics platforms
✓ Enable data pre-processing
✓ Enable development and testing of ANN models
✓ Understand potential for productionisation
✓ Demonstrate a live demand forecasting system
✓ Demonstrate accuracy and robustness of system

• POC will enable further development and application
## Summary – A Successful Proof of Concept

### COMPANY and GOAL
United Utilities vision is to be the best UK water and wastewater company, providing great service to our customers. The company seeks innovative solutions to support this vision.

### CHALLENGE
Agile development and adoption of advanced analytics is difficult
- Access to real time data
- Integration of real time data to analytics platforms
- Development and assessment of algorithm performance
- Automation and Productionisation of algorithms

### SOLUTION
Integrated real-time data to advanced analytics platform via PII4BA to create a live demand forecasting system
- Small scale POC (5 DZs)
- Fully automated
- Data pre-processing
- Optimised machine learning model

### RESULTS
Agile development of a live system producing accurate real-time demand forecasts
- Successful integration of real-time data to analytics platform
- Technology customisable to client needs
- Highly accurate short term demand forecasts
Contact Information

Gary Wong  
gwong@osisoft.com  
Industry Principal – Global Water  
OSIsoft LLC

Kleanthis Mazarakis  
kmazarakis@osisoft.com  
Customer Support Engineer  
OSIsoft UK Ltd

Kevin Woodward  
kevin.woodward@uuplc.co.uk  
OT Strategy Manager  
United Utilities

Michele Romano  
michele.romano@uuplc.co.uk  
OT Senior Engineer  
United Utilities
Questions

Please wait for the **microphone** before asking your questions

State your name & company

Please remember to...

Complete the Online Survey for this session

Download the Conference App for OSIsoft Users Conference 2017

- View the latest agenda and create your own
- Meet and connect with other attendees

Thank You

감사합니다

谢谢

Danke

Merci

Gracias

Спасибо

Obrigado

ありがとう