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

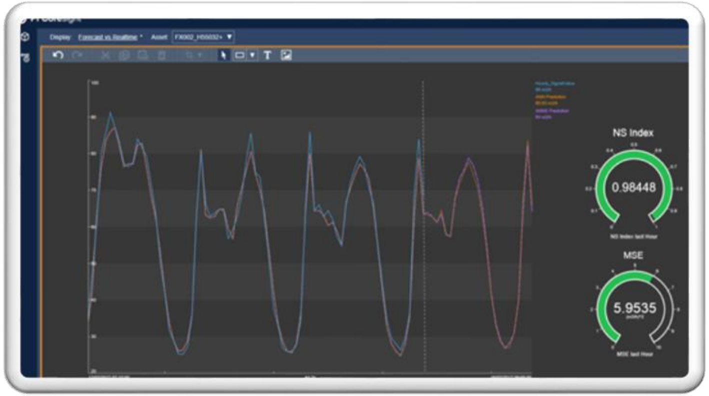
Presented by **Gary Wong & Kleanthis Mazarakis**

On behalf of **Kevin Woodward & Michele Romano**

United Utilities
helping life flow smoothly

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

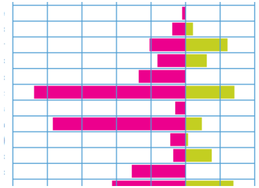


About United Utilities

Business Challenges & Drivers

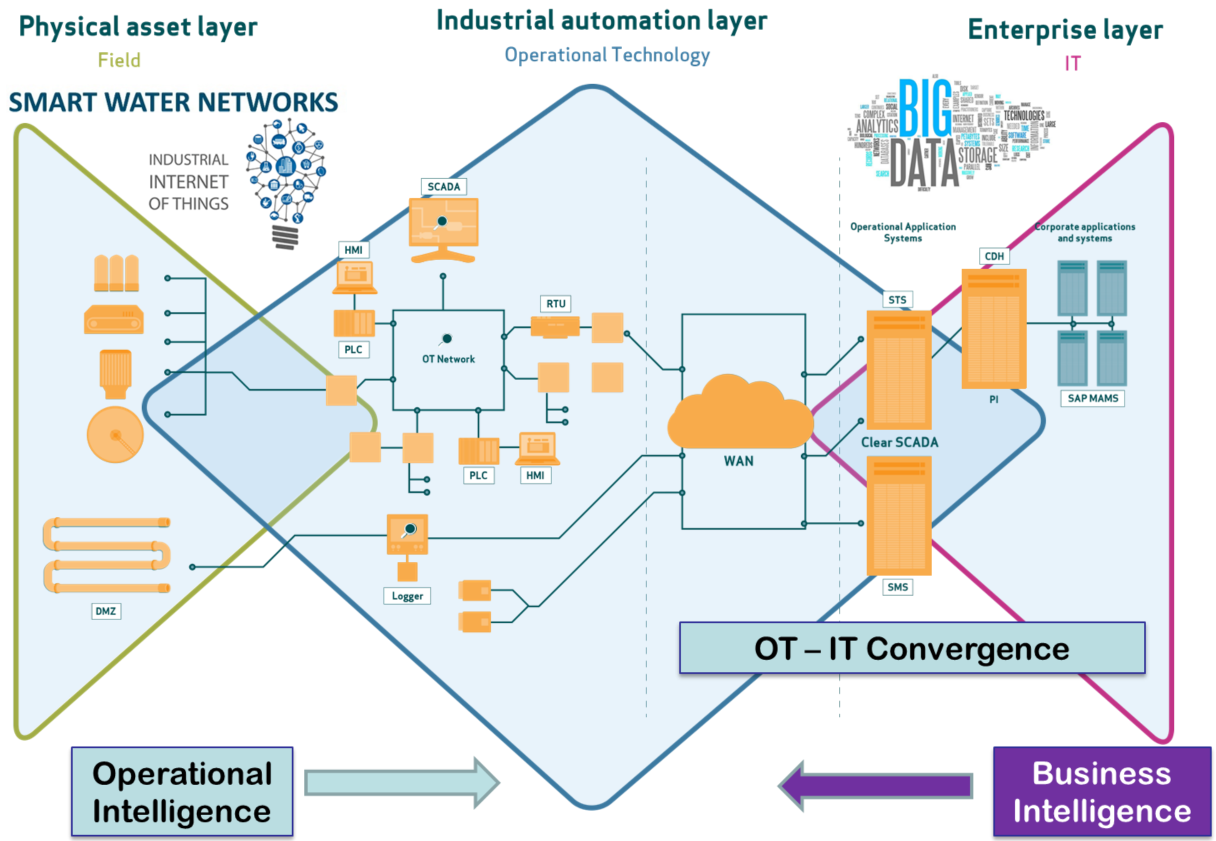
Ofwat

Towards Water 2020



- 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



Innovation and R&D

WATER INDUSTRY ACHIEVEMENT AWARDS 2014

STREAM
Shaping careers, delivering innovation
The Industrial Doctorate Centre for the Water Sector

Cranfield University · The University Of Sheffield. · University of Exeter · Imperial College London · Newcastle University · EPSRC

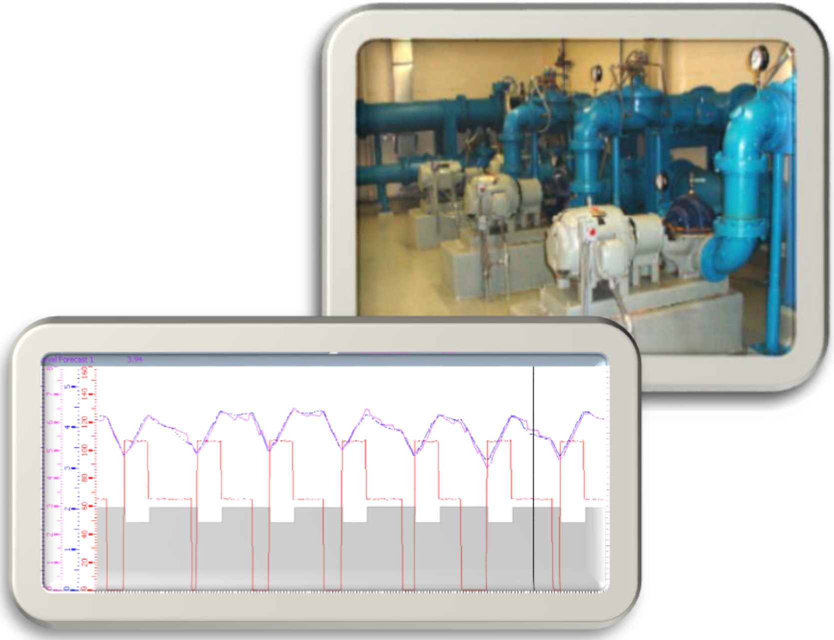
- Water Network Event Localisation
- Event Management for Intelligent Water Networks
- Wastewater Network Event Recognition
- Water Process Event Recognition

WATER INFORMATICS SCIENCE & ENGINEERING
EPSRC CENTRE FOR DOCTORAL TRAINING

Cardiff University Prifysgol Caerdydd · University of Bath · University of Bristol · University of Exeter

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

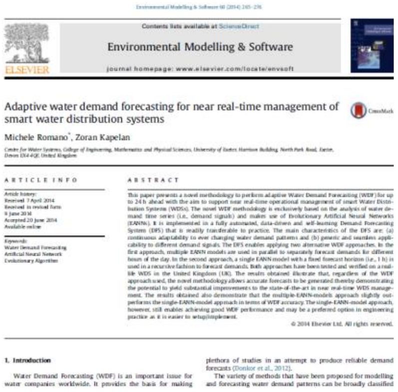
- Understand how advanced analytics can be developed in an agile way
- Enable ready access to time series data
- Link time series data and analytics platforms
- Enable data pre-processing
- Enable development and testing of ANN models
- Understand potential for productionisation
- Demonstrate a live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

Romano, M., and Kapelan, Z. (2014). "Adaptive Water Demand Forecasting for Near Real-Time Management of Smart Water Distribution Systems". Environmental Modelling & Software Journal, 60, 265–276. [Permalink: http://dx.doi.org/10.1016/j.envsoft.2014.06.016].


Environmental Modelling & Software 60 (2014) 265–276

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Adaptive water demand forecasting for near real-time management of smart water distribution systems

Michele Romano*, Zoran Kapelan

Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter, Devon EX4 4QF, United Kingdom

ARTICLE INFO

Article history:
Received 7 April 2014
Received in revised form
8 June 2014
Accepted 20 June 2014
Available online

Keywords:
Water Demand Forecasting
Artificial Neural Network
Evolutionary Algorithm

ABSTRACT

This paper presents a novel methodology to perform adaptive Water Demand Forecasting (WDF) for up to 24 h ahead with the aim to support near real-time operational management of smart Water Distribution Systems (WDSs). The novel WDF methodology is exclusively based on the analysis of water demand time series (i.e., demand signals) and makes use of Evolutionary Artificial Neural Networks (EANNs). It is implemented in a fully automated, data-driven and self-learning Demand Forecasting System (DFS) that is readily transferable to practice. The main characteristics of the DFS are: (a) continuous adaptability to ever changing water demand patterns and (b) generic and seamless applicability to different demand signals. The DFS enables applying two alternative WDF approaches. In the first approach, multiple EANN models are used in parallel to separately forecast demands for different hours of the day. In the second approach, a single EANN model with a fixed forecast horizon (i.e., 1 h) is used in a recursive fashion to forecast demands. Both approaches have been tested and verified on a real-life WDS in the United Kingdom (UK). The results obtained illustrate that, regardless of the WDF approach used, the novel methodology allows accurate forecasts to be generated thereby demonstrating the potential to yield substantial improvements to the state-of-the-art in near real-time WDS management. The results obtained also demonstrate that the multiple-EANN-models approach slightly outperforms the single-EANN-model approach in terms of WDF accuracy. The single-EANN-model approach, however, still enables achieving good WDF performance and may be a preferred option in engineering practice as it is easier to setup/implement.

1. Introduction

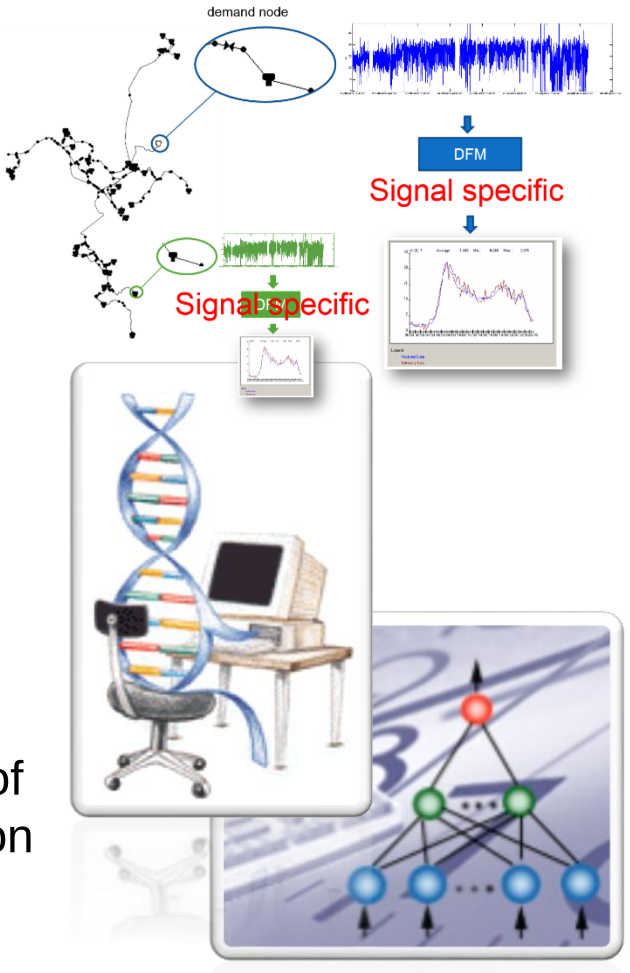
Water Demand Forecasting (WDF) is an important issue for water companies worldwide. It provides the basis for making plethora of studies in an attempt to produce reliable demand forecasts (Donkor et al., 2012).

The variety of methods that have been proposed for modelling and forecasting water demand patterns can be broadly classified

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

<u>Practicality of the methodology operationalisation</u>



DF Approaches

ensemble EANNs (eEANNs)

<u>24 ANN models</u>

- 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)

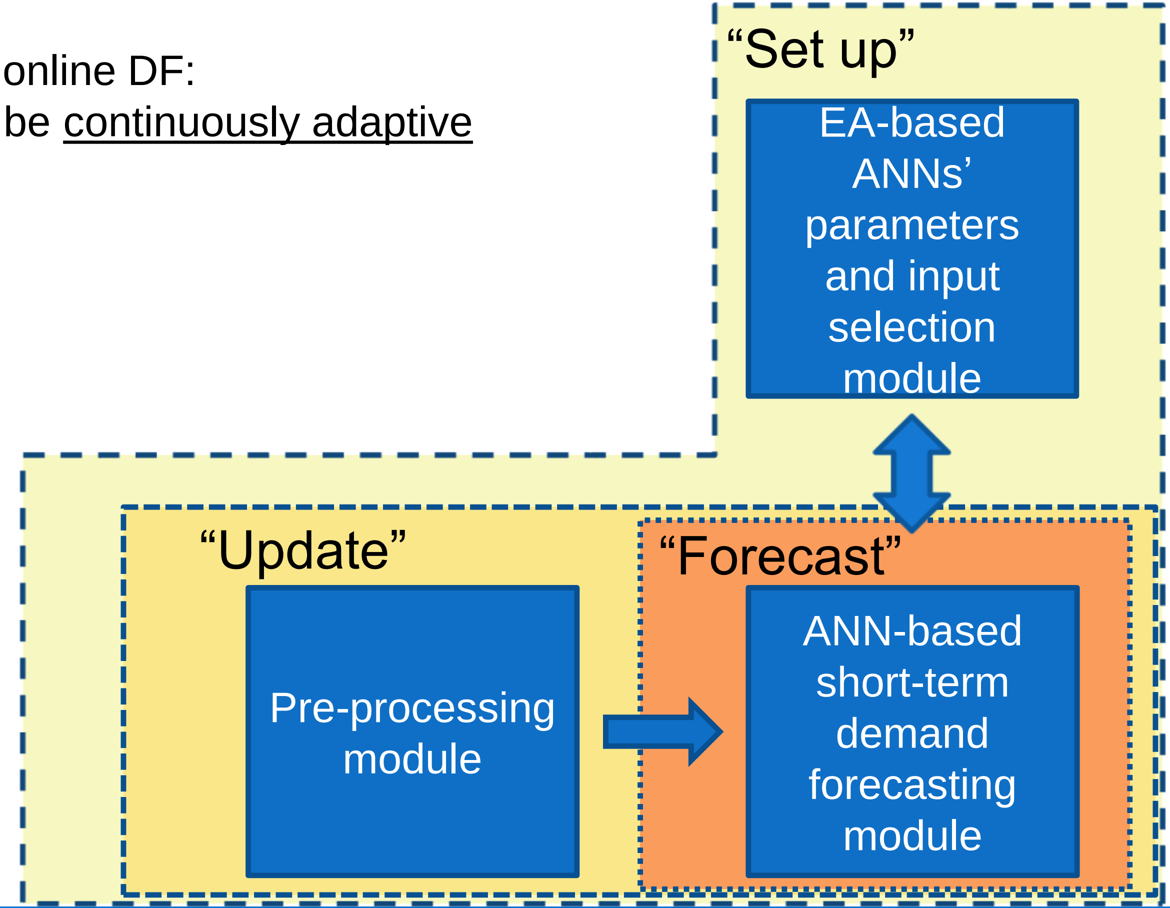
<u>1 ANN model</u>

- 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 <u>continuously adaptive</u>



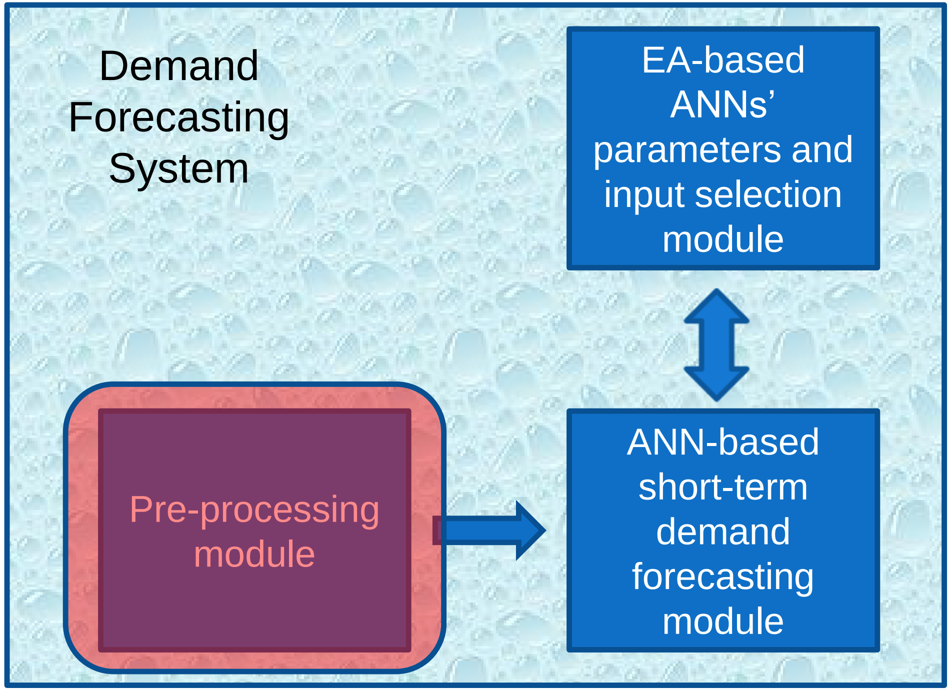
Three main modes of operation:

Set up:
Runs when the module is initialised and when operational changes occur

Update:
Runs every *n* days (e.g., every 7 days)

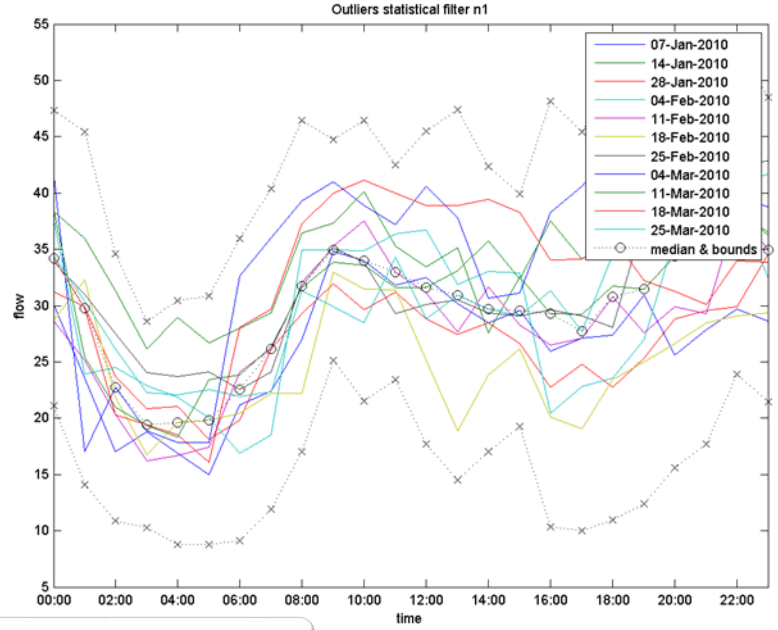
Forecast:
Runs every *p* hours (e.g., every hour)

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

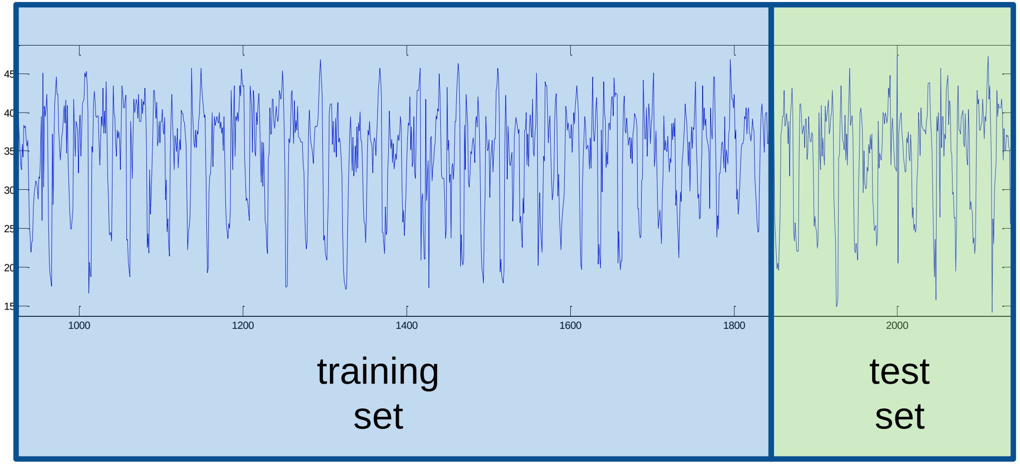


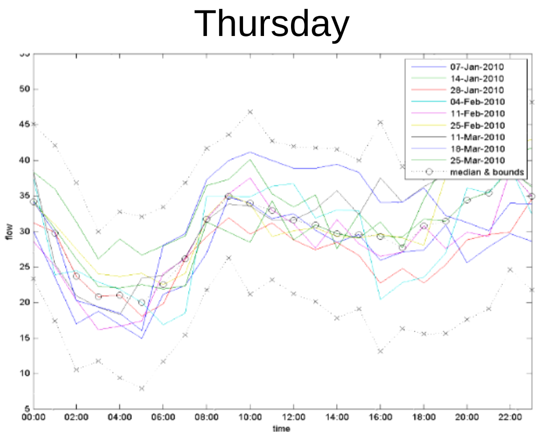
- The retrieved data are rearranged in *m* vectors (i.e., one for each day)
- Missing data in each vector are filled in
- Vectors are **grouped according to their day of week** (e.g., Monday, Tuesday, etc.)
- For each day of the week, a ‘**median day**’ vector and a vector of daily standard deviations 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”

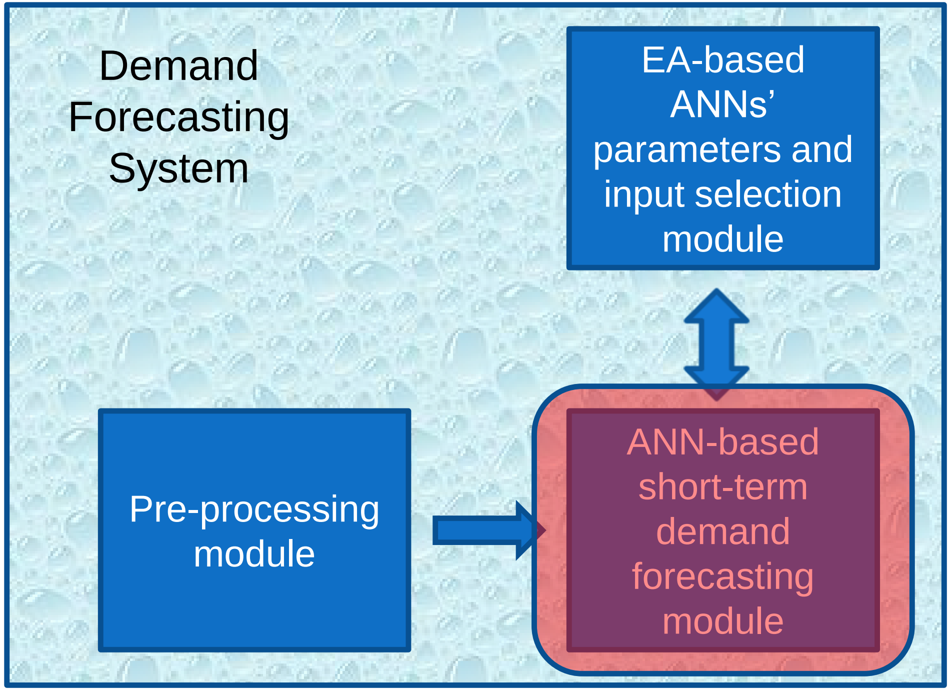
Assembling Good ANN Training and Test Sets & DF Robustness

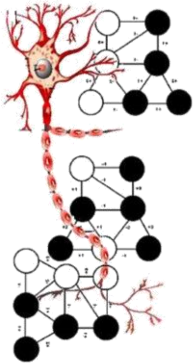
- A set of “good” flow values is assembled. This set will be used for training and testing the ANN models
- A “median day” vector, for each day of the week, is computed. This will be used as surrogate prediction when the ANN models cannot return an output (i.e., lack of incoming real-time data)





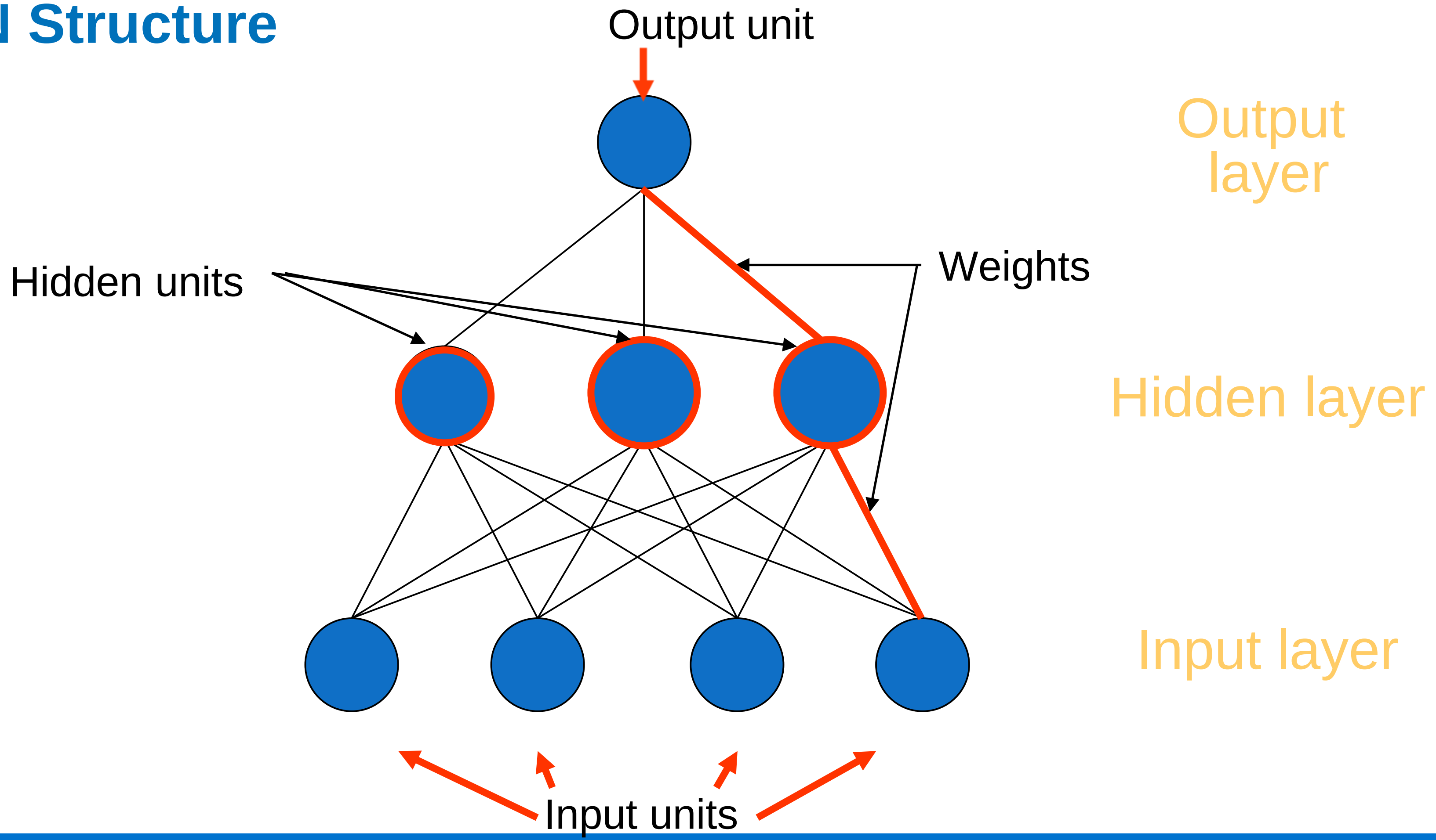
ANN Short-term Demand Forecasting





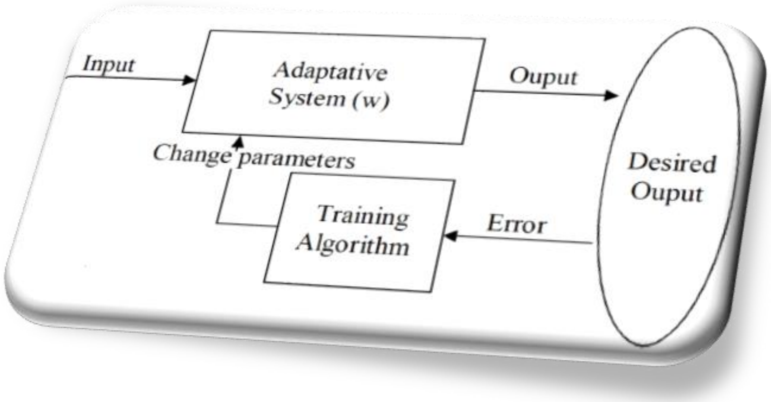
- Trains and tests ANN models in the “Set up” and in the “Update” mode
- Uses the incoming flow data and the trained ANN models to predict the next 24 hour demand in the “Forecast” mode

ANN Structure



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

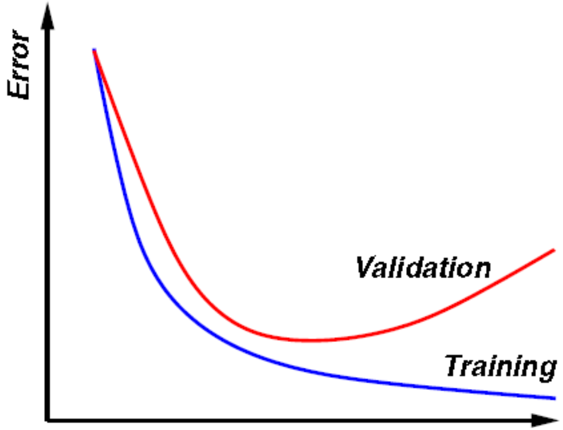


ANN Overfitting

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 <u>for too long</u>
- when we train the network on a task using <u>too many hidden units</u>
- when we train the network on a task using <u>too many inputs</u>

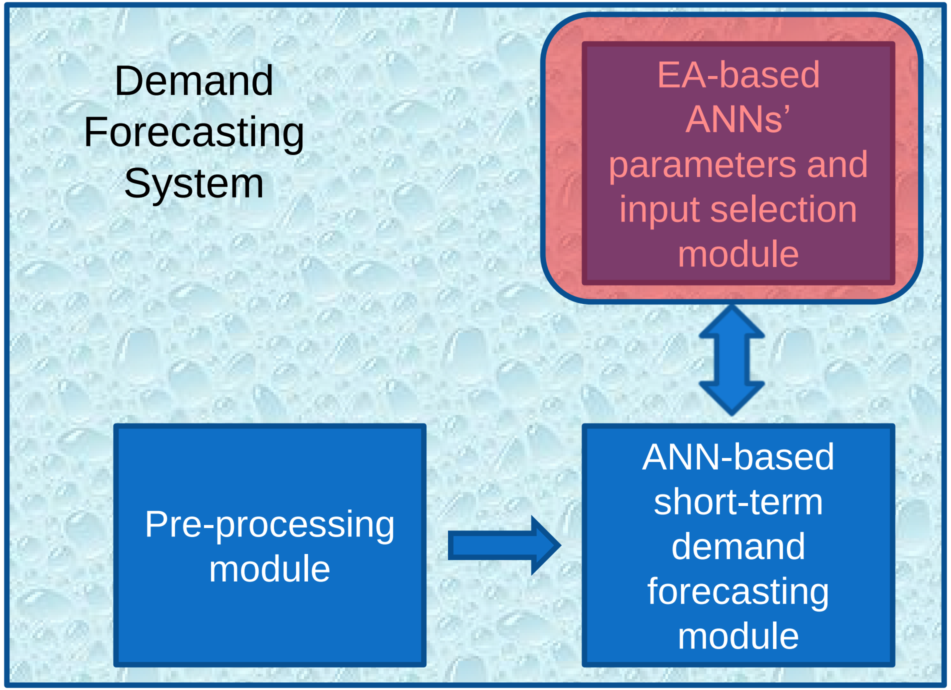


Number of parameters (e.g., hidden units, training cycles, inputs)

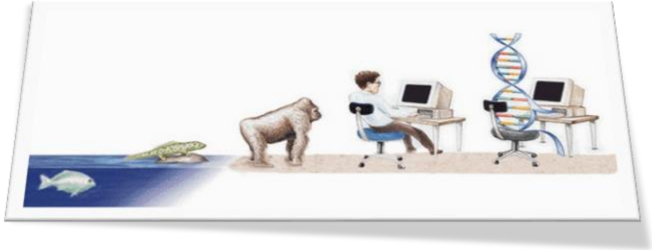
William of Occam, 1284-1347 *"Do not multiply explanations unnecessarily"*

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

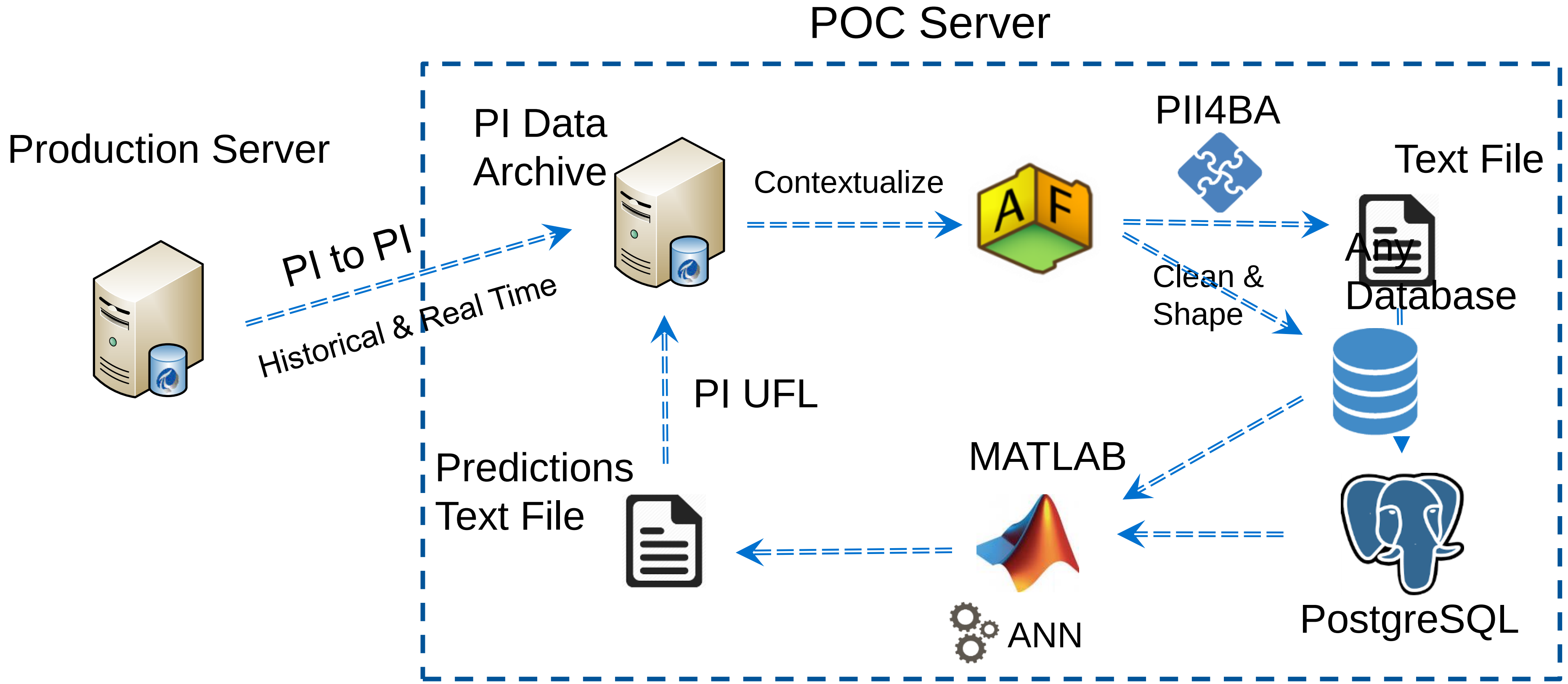
Characteristics:

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

POC Components

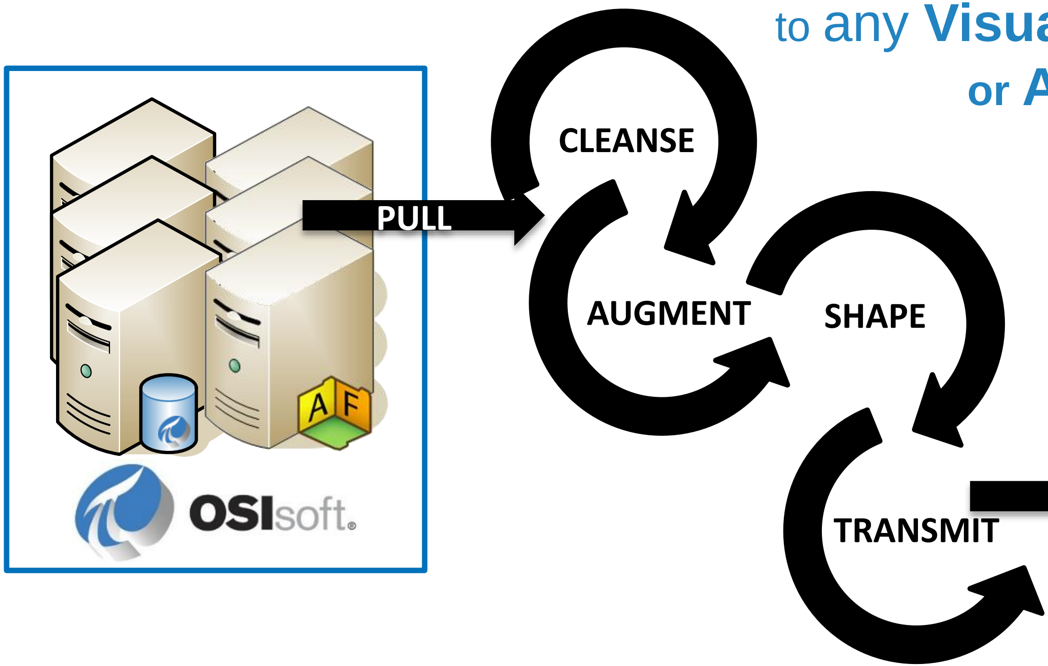
- PI Data Archive 2016 R2
- PI AF 2016 R2
- PI Integrator for BA 2016
- PI Interface for UFL
- PI to PI Interface
- PI Coresight 2016 R2
- PostgreSQL
- MATLAB

POC Architecture – can be customized to any database



Prepare and Deliver Process Data

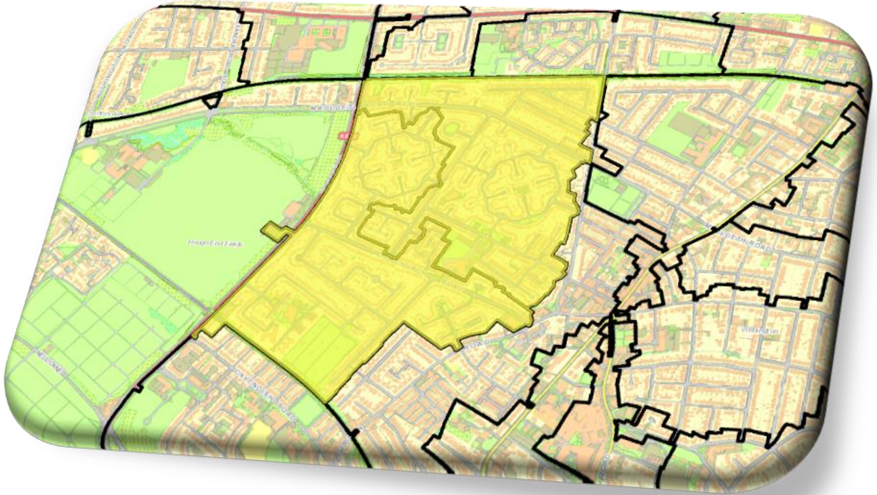
to any **Visualization Tool**
or **Analysis** database
on the ODBC standard



Power BI
IHS HARMONY
tableau
sas
Qlik
Microsoft SQL Server Reporting Services
COGNOS
memsql
TIBCO Spotfire
hadoop
ORACLE BUSINESS INTELLIGENCE

Case Study

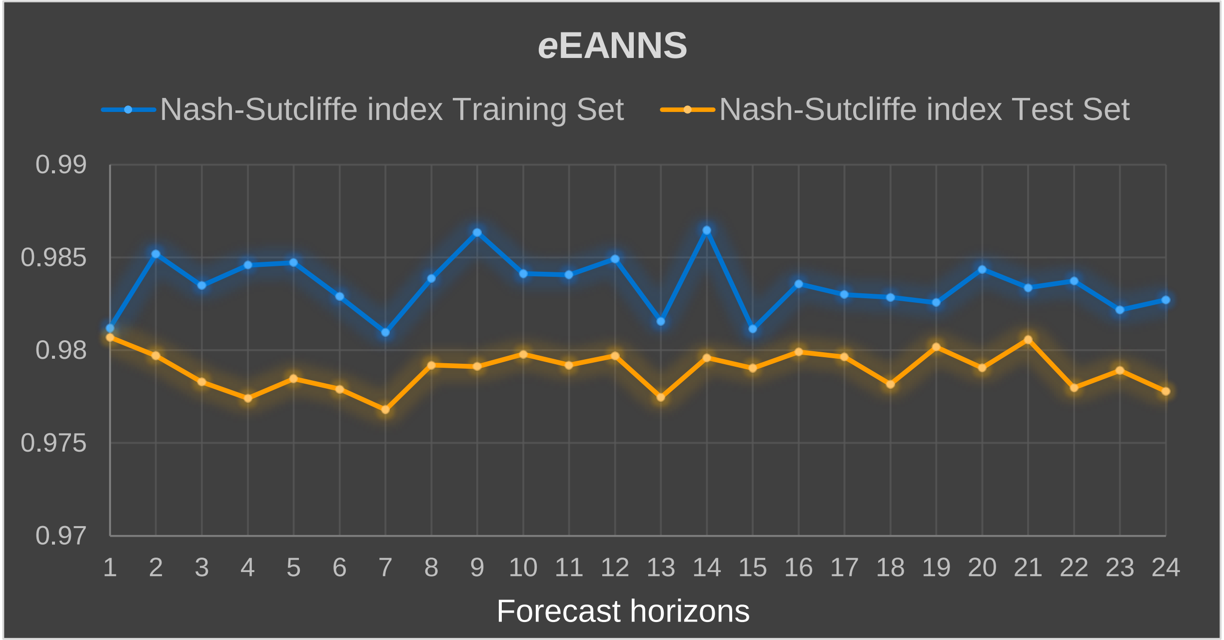
- 5 demand zones (DZs) with different sizes and characteristics e.g. light industrial, urban and rural
- Data for the period from 01/11/2016 to 15/03/2017
- Demand forecasting using the two alternative approaches (i.e. 1 ANN model & 24 ANN models)



Demand Zone Name	Commercial Customers	Residential Customers	Mains Length (Km)	Average Demand (m3/h)
DZ1	8	327	5.1	5.4
DZ2	121	1751	10.7	30.8
DZ3	129	2078	16.4	57.1
DZ4	258	4156	70.2	122.3
DZ5	516	8312	118.5	215.6

Results - Excellent Trained & Tested ANN Models

*r*EANNS	
Nash-Sutcliffe index Training Set	Nash-Sutcliffe index Test Set
0.98193	**0.98132**



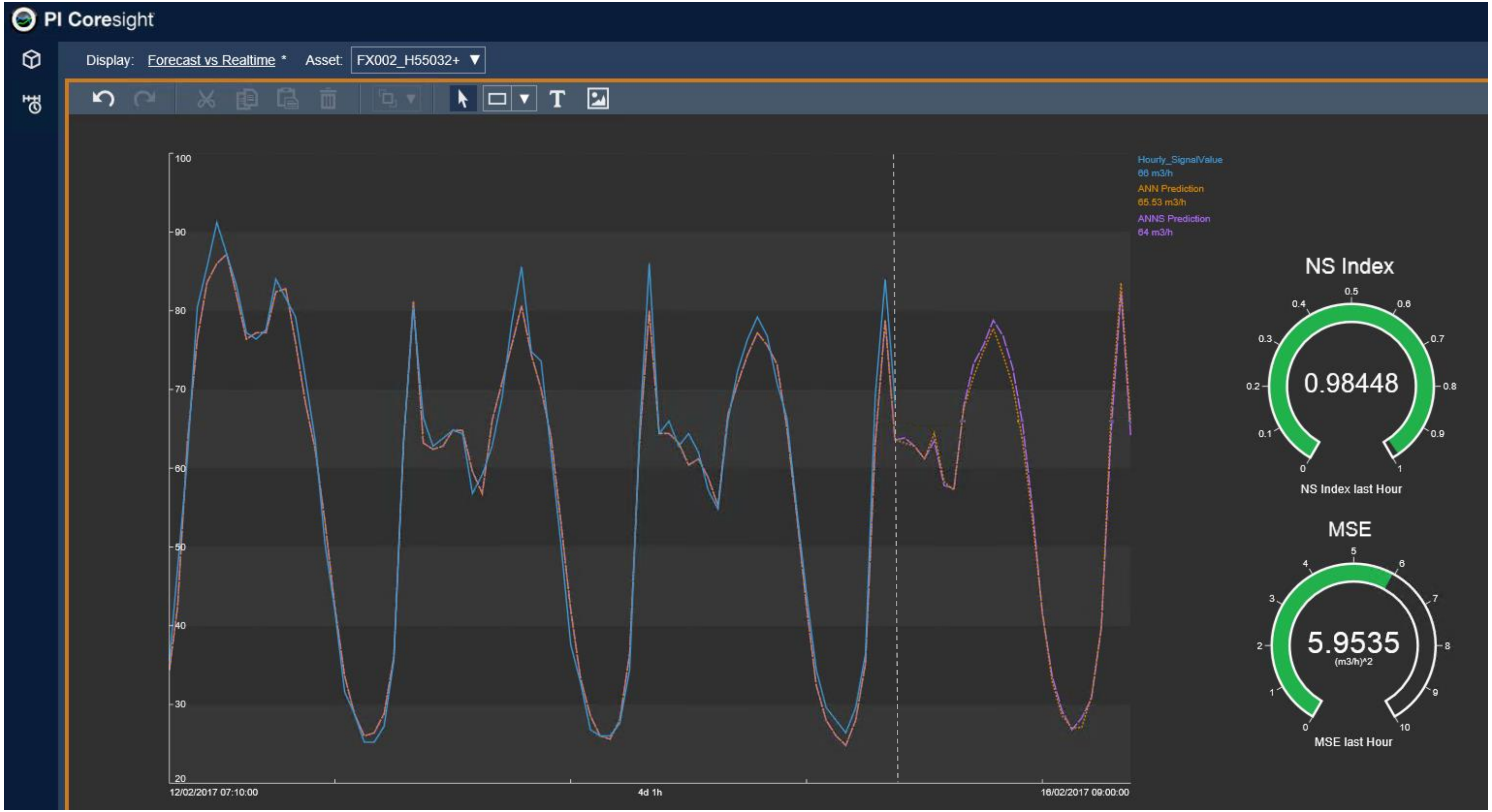
Performance Indicator: Nash-Sutcliffe index

- Normalized statistic
- Determines the relative magnitude of the residual variance compared to the measured data variance
- Ranges between −∞ 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

Nash, J. E. and Sutcliffe, J. V. (1970). “River flow forecasting through conceptual models, Part I - A discussion of principles”. Journal of Hydrology, vol. 10, pp. 282-290.

Results - Excellent Quality Forecasts

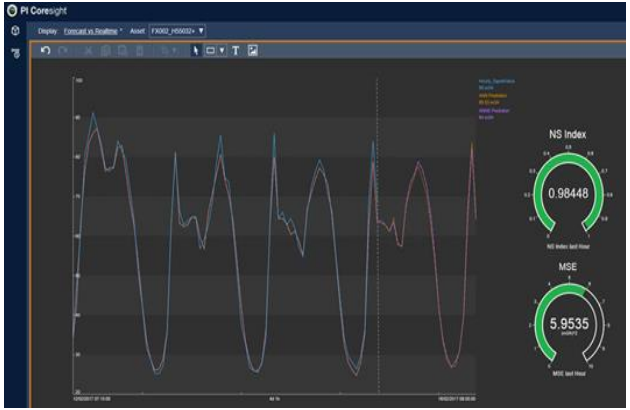


Results v Aims

- ✓ 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



United Utilities
helping life flow smoothly

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-leaning 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

http://bit.ly/uc2017-app

감사합니다

谢谢

Danke

Merci

Gracias

Thank You

ありがとう

Спасибо

Obrigado