Multivariate k-Nearest Neighbor Regression for Time Series data -
a novel Algorithm for Forecasting UK Electricity Demand

ISF 2013, Seoul, Korea

Fahad H. Al-Qahtani
Dr. Sven F. Crone

Management Science, Lancaster University

Lancaster Centre for Forecasting
www.forecasting-centre.com

Agenda

Multivariate KNN Regression for Time Series

1. Introduction
   - KNN for Classification
   - KNN for Regression
     - Formulation and algorithm Meta-parameters
     - KNN Univariate and Multivariate Models
2. KNN for Electricity Load Forecasting
   - Problem and Related work review
   - Experiment Setup
     - Data Description
     - Univariate Model
     - Multivariate Model with One Dummy Variable (WorkDay)
   - Result
3. Conclusions and Future Work

Introduction

KNN for Classification:

- Introduced by Fix and Hodges (1951) and later formalised by Cover and Hart (1967)

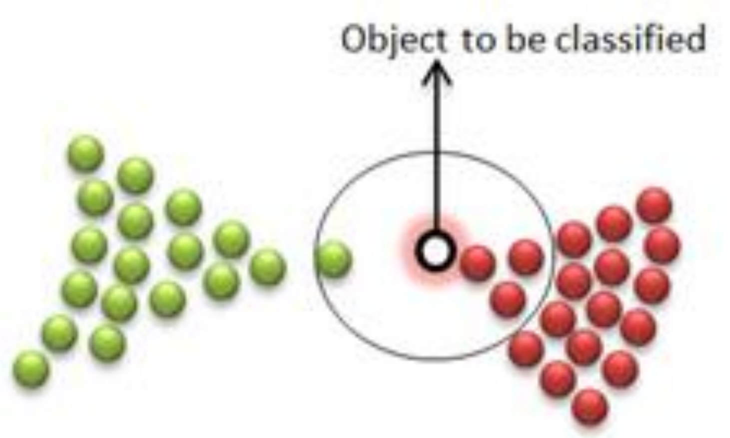
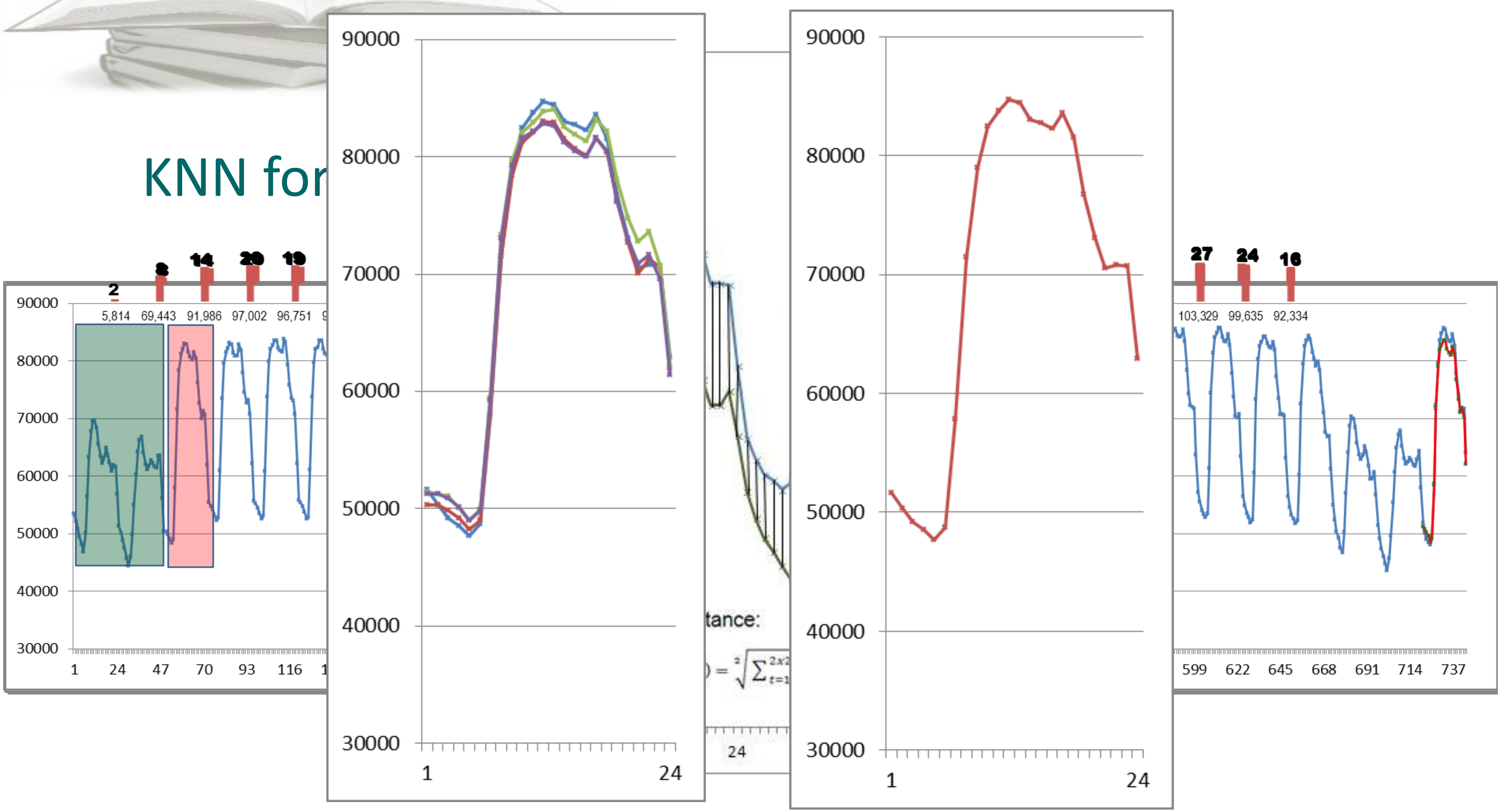


Figure 1: kNN algorithm with k=4 and Euclidian Distance

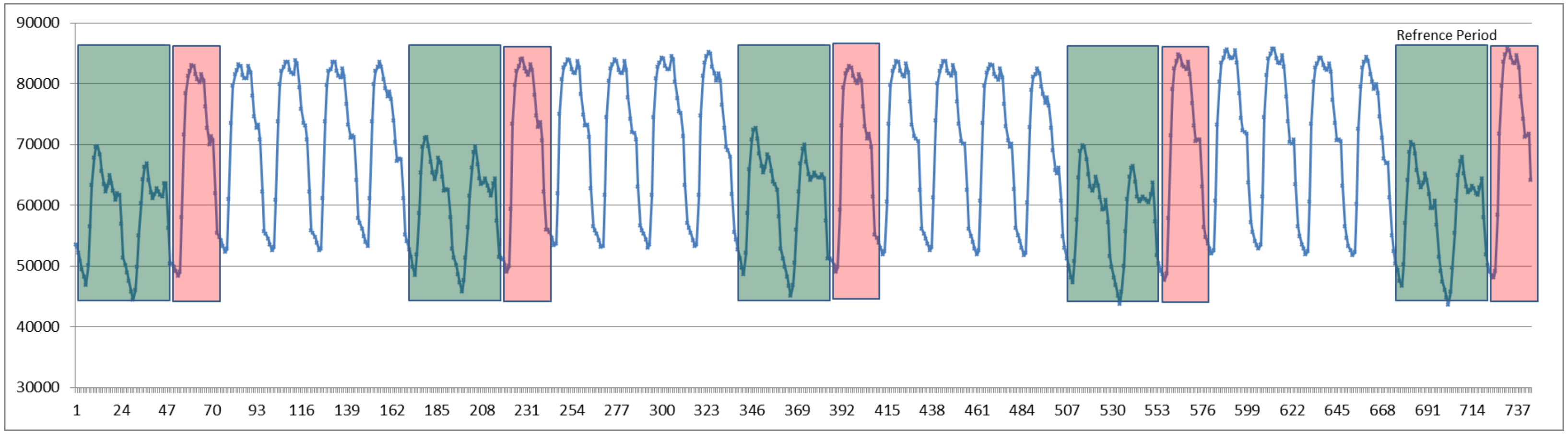
KNN for



90000
80000
70000
60000
50000
40000
30000
1
24
2
8
14
20
19
5,814
69,443
91,986
97,002
96,751
47
70
93
116
27
24
16
103,329
99,635
92,334
599
622
645
668
691
714
737
tance:

Introduction

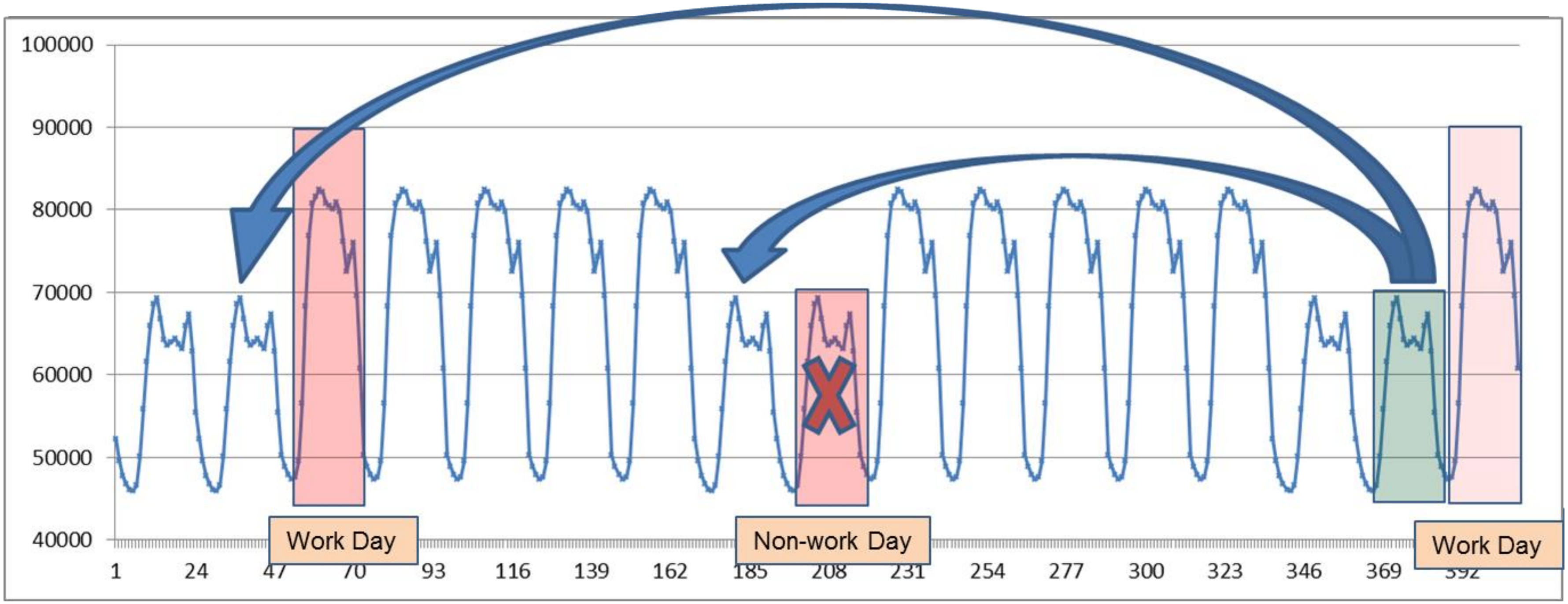
KNN for Regression:



- ❑Formulation: K, Distance Measure, Feature Vector (W) and an operator to combine selected neighbors to estimate forecasted result

Introduction

Multivariate Model :



Introduction

Multivariate Model :

Introducing a Multivariate Model Consisting of:

- Previous Load Observations
- Calendar Information about next day

Agenda

Multivariate KNN Regression for Time Series

1. Introduction
    - KNN for Classification
    - KNN for Regression
        - Formulation and algorithm Meta-parameters
        - KNN Univariate and Multivariate Models
2. KNN for Electricity Load Forecasting
    - Related work review
    - Experiment Setup
        - Data Description
        - Univariate Model
        - Multivariate Model with One Dummy Variable
    - Result
    - Extended Multivariate Model
3. Conclusions and Future Work

Electricity Load Forecasting Problem:

- Accurate load forecasting is essential for the planning and operations of utility companies

  >1% in forecast error can increase the operating cost of a power utility by £10 million

- Challenges:
  - Data with Triple Seasonality (Daily, Weekly and Annual)
  - Outliers, Bank Holidays and Exogenous drivers (Temperature, Economy, Special Events …)

- Models:

  From Conventional Statistical Models to Advanced Computational Models

Review of Related Work:

- kNN for time series Forecasting Application Areas:
  - Most applications are in the following areas:
    - Finance (Fernández-Rodríguez, Sosvilla-Rivero et al. 1999; Andrada-Félix, Fernadez-Rodriguez et al. 2003)
    - Hydrology and Earth Science (Jayawardena, Li et al. 2002; She and Yang 2010)
    - climatology (Dimri, Joshi et al. 2008)

Review of Related Work:

- Within the Electricity Demand Forecasting Application Area:
  - Four journal papers:

    (Lora 2006; Lora, Santos et al. 2007; Sorjamaa, Hao et al. 2007; Jursa and Rohrig 2008)

  - Eight conference contributions:

    (Tsakoumis, Vladov et al. 2002; Fidalgo and Matos 2007; Bhanu, Sudheer et al. 2008; El-Attar, Goulermas et al. 2009; Kang, Guo et al. 2009; Swief, Hegazy et al. 2009; Karatasou and Santamouris 2010; Zu, Bi et al. 2012)

    - No systemic way to set the kNN algorithm parameters
    - Exclude Bank holiday and weekends
    - Rely only on pervious observation

Experiment Setup

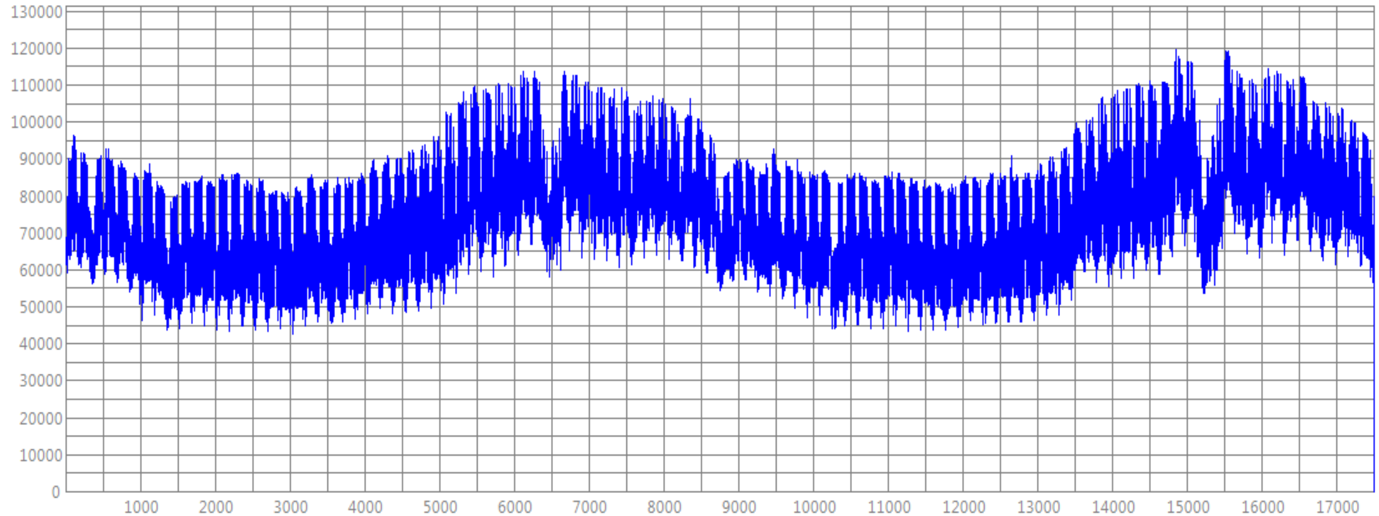
❑Objectives:

- Evaluate the influence of adding features to the KNN algorithm by comparing the accuracy and performance of the univariate and multivariate models (with only the workday feature)

- Set the parameters of the KNN algorithm for the univariate and multivariate models and produce forecast for the UK electricity data. Also, Evaluate the performance of both models against Statistical Benchmarks

Experiment Setup

❑UK Electricity Demand Data



Hourly Electrical Load Time Series for 2 Years

- Data from 2001 to 2008

Experiment Setup

- UK Electricity Demand Data
  - Training Data set:

    All days in 2004

  - Testing Data set:

    All days in 2005

Experiment Setup

❑Univariate Model Tuning

k \ w	1	2	3	4	5	6	7	8	9	10
1	2.88	2.86	2.93	2.94	2.92	2.86	2.80	2.98	3.01	2.99
2	2.88	2.86	2.93	2.94	2.92	2.86	2.80	2.98	3.01	2.99
3	2.60	2.61	2.69	2.64	2.67	2.60	2.58	2.71	2.73	2.73
4	2.48	2.53	2.54	2.49	2.50	2.46	2.48	2.60	2.67	2.67
5	2.46	2.47	2.42	2.45	2.43	2.41	2.40	2.53	2.60	2.60
6	2.43	2.50	2.41	2.41	2.41	2.41	2.37	2.50	2.58	2.56
7	2.46	2.52	2.43	2.41	2.41	2.40	$\boxed{2.37}$	2.49	2.57	2.54
8	2.48	2.52	2.45	2.43	2.41	2.40	2.37	2.49	2.57	2.56
9	2.50	2.53	2.46	2.45	2.41	2.42	2.39	2.50	2.57	2.55
10	2.52	2.53	2.49	2.47	2.42	2.43	2.40	2.50	2.56	2.55

Table 1: Tuning the w, k parameters for the kNN model using the 2004 data

Experiment Setup

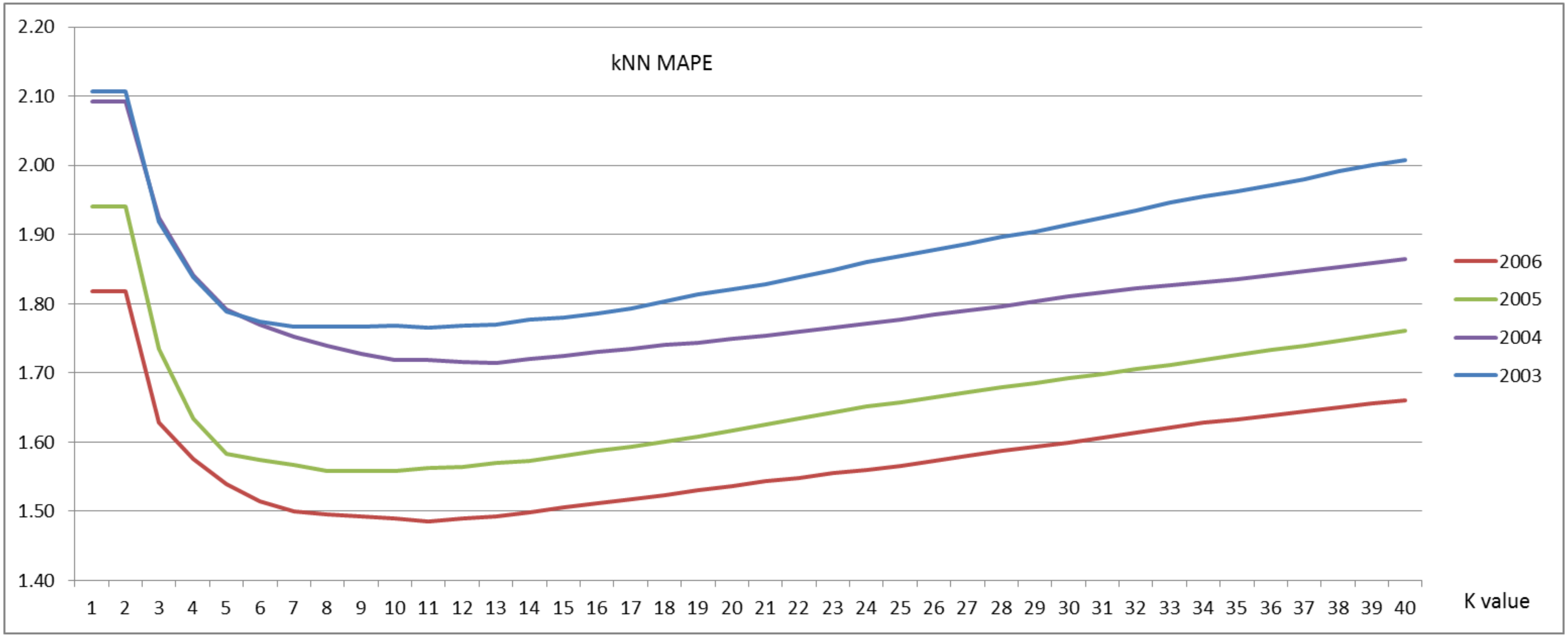
❑Multivariate Model with the WorkDay feature

K \ W	1	2	3	4	5	6	7	8	9	10
1	2.31	2.51	2.49	2.66	2.70	2.73	2.75	2.91	2.91	2.97
2	2.31	2.51	2.49	2.66	2.70	2.73	2.75	2.91	2.91	2.97
3	2.08	2.25	2.26	2.42	2.44	2.48	2.51	2.62	2.63	2.68
4	1.97	2.13	2.15	2.29	2.31	2.35	2.40	2.51	2.54	2.58
5	1.92	2.08	2.09	2.22	2.24	2.31	2.35	2.44	2.47	2.51
6	1.89	2.06	2.06	2.18	2.21	2.29	2.32	2.40	2.43	2.46
7	1.89	2.04	2.05	2.17	2.21	2.28	2.32	2.40	2.42	2.44
8	1.90	2.03	2.06	2.17	2.21	2.27	2.31	2.40	2.42	2.45
9	1.90	2.02	2.06	2.17	2.21	2.29	2.32	2.40	2.42	2.45
10	1.90	2.02	2.07	2.17	2.22	2.29	2.32	2.39	2.41	2.45

Table 2: Tuning the w, k parameters for the modified kNN model using 2004 data

Experiment Setup

❑Multivariate Model – Setting K:



Experiment Setup

❑Statistical Benchmarks

- 2 Seasonal Naïve Models (Random Walk):

  RW_{24} and RW_{168} ($RW_s : \hat{y}_{t+h} = y_{t-s+h}$)

- 2 Seasonal k Average Models:

  $MOVAV(7)_{24}$ and $MOVAV(7)_{168}$ ($MOVAV(k)_s : \hat{y}_{t+h} = \frac{1}{k}\sum_{i=1}^{k} y_{t-ks+h}$)

Experiment Setup

❑Result:

AVERAGE RESULT OF K-NN REGRESSION AND BENCHMARKS

	Average MAPE	Rank	Std. Dev MAPE	Rank
RW_{24}	5,4991 %	4	± 0,4865	3
RW_{168}	3,6812 %	3	± 1,3795	5
$MOVAV(7)_{24}$	6,6941 %	6	± 0,5653	2
$MOVAV(7)_{168}$	5,7835 %	5	± 2,8306	6
univariate k-NN	2,3824 %	2	± 0,8136	4
multivariate k-NN	1,8133 %	1	± 0,2970	1

Experiment Setup

- Result:

- Computation Cost:

  Univariate Model : 6.5 Minutes

  Multivariate Model : 2.7 Minutes

  59% Improvement

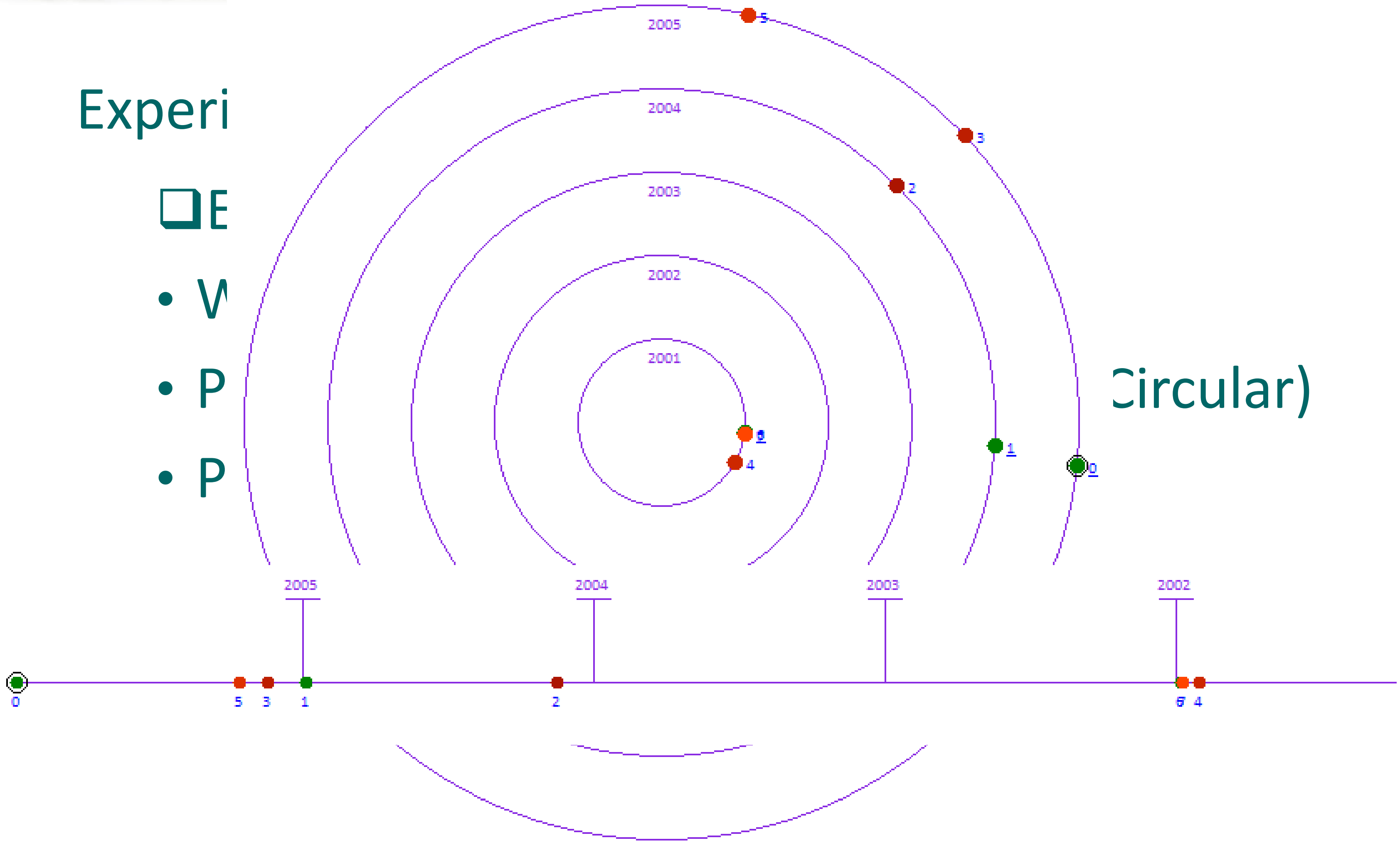
Experi

❑E

• W

• P … Circular)

• P



Agenda

Multivariate KNN Regression for Time Series

1. Introduction
    - KNN for Classification
    - KNN for Regression
        - Formulation and algorithm Meta-parameters
        - KNN Univariate and Multivariate Models
2. KNN for Electricity Load Forecasting
    - Related work review
    - Experiment Setup
        - Data Description
        - Univariate Model
        - Multivariate Model with One Dummy Variable
    - Result
    - Extended Multivariate Model
3. Conclusions and Future Work

Concluding Remarks:

- KNN algorithm is intuitive, easy to implement and can give reliable results for electricity demand forecasting when its parameters set correctly

- Including extra information about the day being predicted into the KNN algorithm can increase its accuracy and improve its performance.

Future Work

- Include exogenous variables such as:
  - Temperature
  - Humidity

- Improving KNN performance by Implementing an Active Learning Mechanism For Selecting The Most Informative Training Data.

- Integrate kNN with other forecasting frameworks such as NN and SVM

Lancaster University
Management School

Questions?

Fahad H. Al-Qahtani
Lancaster University Management School
Centre for Forecasting - Lancaster, LA1 4YX
email: alqahta2@exchange.lancs.ac.uk