

The use of Enterprise Miner™ with large volumes of data, for forecasting in an automated batch process.

By

Matthew Glasson & Justine Eastman

London Electricity & OCS Consulting

25th September 2001

London Electricity Group





OCS Consulting



Agenda

- Overview
- Environment and data detail
- Enterprise Miner™ software
- Model selection
- Model refinement
- Automation
- Summary.

Overview

- The introduction of NETA has introduced new challenges to the electricity industry
- Forecasting supply is particularly testing
- Energy demand forecasting is subject to a variety of volatile parameters
- A solution is needed to be able to provide fast, accurate forecasts for easy inclusion into the existing systems and software
- SAS® software identified as providing the optimum solution for the overall system that London Electricity had designed.

Project Background

- London Electricity were already using Enterprise Miner™ software to develop models for forecasting
- Requirement to use Enterprise Miner™ software in a client/server environment
- Requirement to automate the forecasting element wherever possible
- Joint meeting with SAS Institute, London Electricity and OCS Consulting
- OCS produced proposal to undertake task to meet London Electricity requirements.

Environment and Data

- Oracle database within UNIX environment
- SAS v8.0 with SAS Enterprise Miner™ software version 4.0
- Client Server project created in Enterprise Miner™ software
- Three main areas of data utilised:
 - Demand data
 - Weather data
 - Calendar data.
- Prediction of demand at the half hourly level
- Data tables processed by Enterprise Miner™ software are up to 1GB, sometimes over a million records.

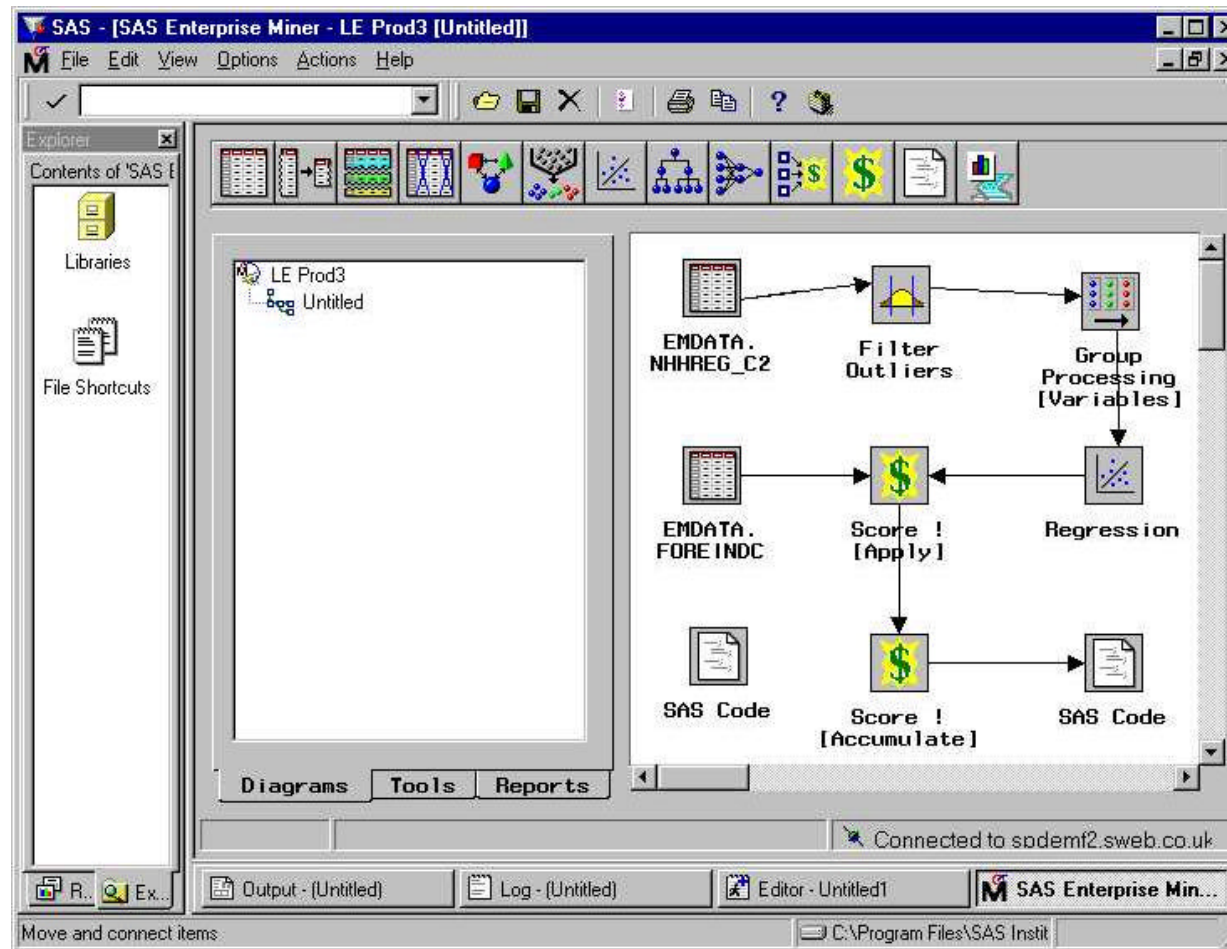
Enterprise Miner™ software

- Identified because of the ability to control the modelling process
- Ease of use and model building
- SEMMA methodology
- Ease of applying the model code to future data.

Data Mining Diagram



OCS Consulting



SEMMA Methodology

- **Sample**
 - **Explore**
 - **Modify**
 - **Model**
 - **Assess**
-
- **Not all steps are necessarily used - the methodology is completely flexible**

Statistical Modelling

- **Combination of prior expertise / business knowledge and understanding of regression techniques were important**
- **Regression is the best overall model because of the balance of accuracy and interpretability**
- **Started simple - using basic nodes**
- **Compared further models to substantially improve the initial model.**

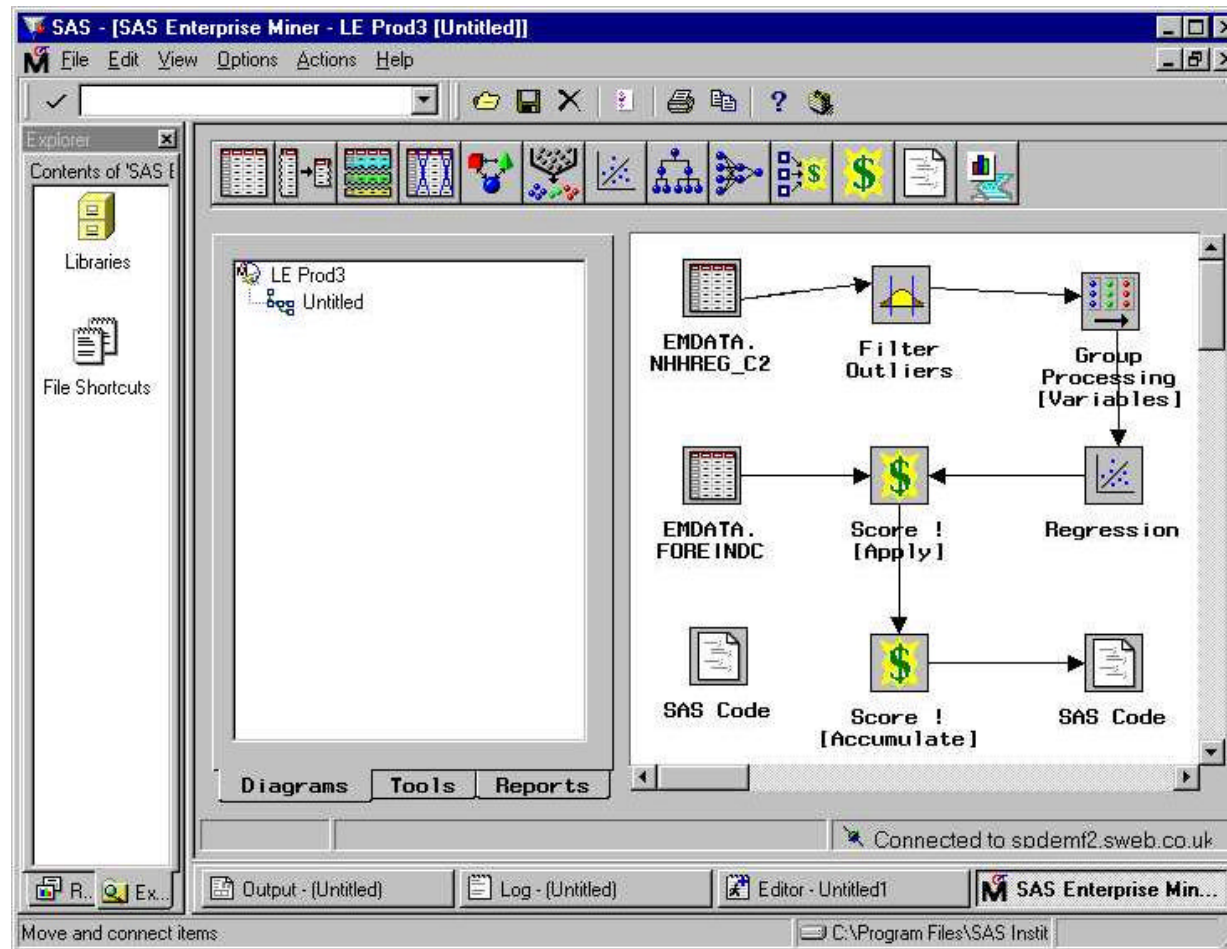
Perfecting the Model

- Aimed to improve the original model
- Refined the regression within the regression node
- Explored further nodes within Enterprise Miner™ software
- Steps added to the data mining diagram:
 - Filter outliers node
 - Group processing node
 - Score nodes.
- Pre-processing of input data.

Data Mining Diagram



OCS Consulting



Extraction of Model Code

- **Score node used to score the data**
- **Score code for scoring future data saved and exported**
- **Saved into SAS code file**
- **Incorporated into overnight scheduled environment**
 - generic UNIX script within the scheduler
 - controls the running of the correct model and associated parameters.

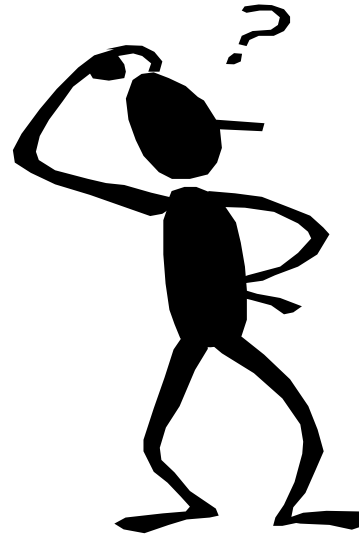
Future Developments?

- **Though regression proves to be a good model - could try other statistical models:**
 - **Neural Networks – further work could provide an insight into increased model accuracy**
 - **Revisit the modelling with Enterprise Miner™ software 4.1.**

Summary

- **Successful regression modelling now incorporated into the forecasting system solution**
- **Data mining process started simple and was refined by supplementing the approach with additional functionality of Enterprise Miner™ software**
- **A suite of SAS scripts are successfully being used within the production environment**
- **Demand successfully being predicted in the live environment, coinciding with the introduction of NETA in March 2001.**

Questions



For further details regarding the presentation:

- e-mail:
 - **info@ocs-consulting.com**
- Visit the OCS web-site: **www.ocs-consulting.com**