Statistical Case Studies 2024/25 Semester 2
Modelling peak electricity demand in GB

1 Background

NESO (National Electricity System Operator) are the electricity system planner and operator in Great Britain. They ensure that electricity supply meets demand at all times and play a role in planning the future of the GB electricity system, for example, by advising where future generation and transmission should be built (https://www.neso.energy/).

In order to take decisions about the future of the GB electricity system, NESO must be able to estimate demand patterns over the long-term. New generation and transmission take a long time to build and are expensive, so NESO need to be able to estimate the uncertainty in peak demand many years ahead in order that they do not over- or under-build the electricity system.

Uncertainty in long-term electricity demand can be driven by many factors, for example:

- The weather and the climate,
- Economic trends (e.g. productivity),
- Energy efficiency policies and trends,
- Population changes,
- Technological advancements, e.g. take up of electric vehicles.

NESO use linear regression models based on historic data to help estimate future demand patterns. Historic data cannot be used directly as due to trends in population, the economy, climate etc., the past is unlikely to be representative of the future. To account for this, NESO use their regression models to rescale historic data using a 'year effect' statistic. Year effects for historic data can be estimated from these regression models and year effects for the future are set subjectively in discussion with stakeholders.

Your challenge in this project is to develop a new linear regression model of historic demand data that NESO could use to generate realistic traces of future daily peak demand for use in planning how much electricity generation to build to avoid electricity shortfalls (security of supply analysis). As the model is to be used for security of supply analysis, NESO is especially interested in making sure that the model fits well at high demands, as this is when shortfalls will occur. They will use the new model to estimate statistics such as the maximum annual demand for a future year or high quantiles (e.g the 95% quantile of peak daily demand over a year). As there is not much electricity storage currently installed on the system, they are not particularly concerned about clusters of high demand occurring close together in time.

For extra (optional) reading, see:

- NESO documentation on how they estimate historic year effects (known as the Average Cold Spell peak): `https://www.emrdeliverybody.com/Capacity%20Markets%20Document%20Library/NGESO%20ACS%20Methodology%202022.pdf`
- Details from NESO on how they develop subjective views of the future energy system: `https://www.neso.energy/publications/future-energy-scenarios-fes`

- NESO's annual electricity capacity report - this is one of the uses of this analysis: `https://nationalenergyso-emr.my.salesforce.com/sfc/p/#8d000002dUGC/a/J70000004CYD/cv3SY3Z5cLuiRsHHJuK5FZcNebxJDmgEeAqjo9ot1oo`
- Paper giving a previously developed model, but the temperature term is too variable `https://www.pure.ed.ac.uk/ws/portalfiles/portal/248522796/demand_forecasting_pmaps_submitted.pdf`
- Another paper on previous analysis which focuses on joint modelling of demand and wind (which you should not do), but contains some uncertainty modelling: `https://www.pure.ed.ac.uk/ws/portalfiles/portal/150700585/PMAPSwinddemand.pdf`

2 Task

NESO have asked you to help develop a new model based on linear regression that they can use to generate realistic traces of daily peak demand for use in their security of supply analysis. They want you to test and fit different models using the data sets provided and to recommend a model for their use. In addition to providing a recommended model, they also want you to answer the following questions:

- How does peak daily demand depend on temperature? NESO currently use a variable known as TE (described below and in your dataset) - is there a better alternative?
- How well does your model fit the historic data?
- How could the maximum annual demand have varied in the 2013-14 winter season if different weather conditions had occurred? For this you can investigate variation in the maximum annual demand in 2013/14 by inputting the weather conditions from previous winters (i.e. test weather in 1991/92, 1992/93, ...).
- What would be the limitations in using your model to generate traces of future peak demand over a 5-10 year timescale?

3 Data

You have been provided with two datasets. The main one is 'SCS_demand_modelling.csv' which contains most of the data that you will need for your analysis. You have also been given an additional dataset 'SCS_hourly_temp.csv' which may be useful when investigating the dependence of demand on temperature. It contains hourly temperature measurements covering the period over which your main dataset extends. The variables contained within 'SCS_demand_modelling.csv' are described below. You can assume that peak daily demand occurs at 6pm.

The main dataset you have been given covers the winter periods only. This is because NESO is interested in looking at periods of time where the probability of a shortfall of electricity is highest. In Great Britain, electricity demand is much higher in the winter due to electricity use for lighting and heat and shortfalls will not occur in summer.

3.1 Variables

- Date
- Wind - estimated capacity factor of wind generation at 6pm based on wind speeds on the given date, with installed wind generation as at 1 January 2015. The capacity factor is the wind generation divided by the installed wind capacity. The wind generation is estimated using a physical model of installed wind generators across GB combined with historical wind speed data. The wind speeds used are generated from the NASA MERRA1 reanalysis dataset.

- Solar_S - estimated capacity factor of solar generation based on solar output on the given date at 6pm, with installed solar generation as at 1 January 2015. The capacity factor is the solar generation divided by the installed solar capacity. The solar generation is estimated using a physical model of installed solar generators across GB combined with historical solar output data. The solar output time series is generated from CMSAF SARAH satellite images.
- Demand_gross (MW) - the electricity demand recorded for 6pm on the given date. Demand is for all of GB as reported by NESO. This is measured as supply from all major power stations minus exports minus pumping for hydro minus transformer load. Alternatively it is end-user consumption minus embedded generation. It has estimates of embedded renewables added back on (these are wind farms and solar farms whose output is not recorded by NESO).
- temp - British population-weighted average temperature from MERRA1 at 6pm on given date.
- wdayindex - day of week, beginning from Sunday (0).
- monthindex - month of year, beginning with January (0).
- year
- TO - average temperature from 3pm-6pm on given date.
- TE - average of TO at 6pm and TE at 6pm on the previous day.
- start_year - the year at the start of the winter period.
- DSN - the number of days since the 1st of November.

4 Format

You should submit a written report detailing your findings. Your report must contain no more than 4000 words (not including references, tables, figures and their captions). Please state the word count under the title of your report. You can structure your report as you like, but:

(a) It should contain an executive summary that could be presented in a business context to people without a background in mathematics. This should present and explain your results. This section of the report should be no longer than 500 words and should be readable without specialist statistical knowledge.

(b) The rest of your report should be aimed at analysts at NESO (e.g. those with a scientific background but without specific statistical expertise), and should explain exactly what you did, why you did it, and what the conclusions were, in a manner that would allow the analysis to be repeated. By this, I mean that the statistical structure of what you have done should be clear, so that it could be repeated using any statistical software the reader wanted to use. This means that models and results should be presented mathematically and with graphs or tables as necessary, and not using computer code and output.

(c) You should include a description of the data used along with any exploratory data analysis.

(d) You can assume that the reader of your report has read the executive summary.

(e) The report should include no computer commands and no raw output.

(f) You should also submit well-documented R code, either plain or as markdown so that your analysis is reproducible.

The work must be completed in your group of 3, which you must have arranged and registered on Learn. If you have any difficulties with this or have not managed to find yourself a group, please email me or speak to me in the workshops. The first paragraph of your report must list your names with university user names. Contributions from different team members never end up completely equal, but you should aim for rough equality, with team members each making sure to 'pull their weight', as well as not unfairly dominating.

5 Mark Scheme

There is no single correct analysis for this type of project, so you will not be marked on the basis of how close you get to some particular model answer. The marks are not subdivided, but will be allocated on a combination of statistical approach and justification, interpretation of results in context and presentation.

- 80 – 100% A report that could be presented to the client or collaborator with little or no revision. Analysis is sound so that conclusions are well-supported statistically. Interpretation is reasonably mature. The project should demonstrate a clear overview of the work, without getting lost in details, and be free of all but minor statistical errors. The work is to a publishable standard.
- 70-79% A report that could be presented to the client or collaborator with little or no revision. Analysis is sound so that conclusions are well-supported statistically. Interpretation is reasonably mature. The project should demonstrate a clear overview of the work, without getting lost in details, and be free of all but minor statistical errors.
- 60 – 69% A project that could be presented after a round of revision, but without having to re-do much of the actual analysis. Some flaws in the analysis or presentation (or minor flaws in both), but basically sound. A good grasp of the statistics and context, so that interpretation is reasonable.
- 50 - 59% Major re-working required before the project could be presented, but containing some sound statistics demonstrating understanding of statistical modelling and its application. Reasonable presentation and organisation.
- 40 – 49% Major flaws in analysis and presentation, but demonstrating some understanding of statistics, and a reasonable attempt to present the results.
- Fail (below 40%) Flawed analysis demonstrating little or no understanding of statistics, and/or incomprehensible or very badly organised presentation.