

Project Description

Winter 2021

The project consists of forecasting electricity demand in an area operated by the Southwest Power Pool (SPP), an American regional transmission organization serving 17 states.

Information on the SPP can be found at <https://spp.org/about-us/>.

Data for 11 areas have been collated in the **R object** `sppdata`.

Data are stored in `sppdata.RData` available in Zone Cours.

The object `sppdata` has R class `timeSeries`.

See `names(sppdata)` in R after loading the object, and the column descriptions at <https://marketplace.spp.org/pages/hourly-load> to identify the 11 areas included in the object.¹

Each team will be assigned a unique area.

The starting point for all teams is the hourly demand data provided in `sppdata`.

Students' R code must read the `sppdata.RData` file.

Any data manipulation must be done in students' submitted R code. It is the only way to document changes and allow analyses to be replicated.

The data must be analyzed at the **daily** level.

The objective is the forecast of the **peak hourly load for day $t + 1$** , i.e. the largest hourly demand for day $t + 1$.

Forecasts are short-term, i.e. **for $h = 1$ from data available on day t , and this for each day of the year.**

At least one good candidate from naive, smoothing, linear regression, and time series methods must be included in the comparisons. As we will only cover nonparametric methods at the 11th lecture, they will not be used for the project.

Historical data for temperature and other possibly useful explanatory variables are available (free of charge) from other websites. See e.g. <https://www.ncdc.noaa.gov/cdo-web/search>.

Any forecasts produced by SPP or others cannot be used as entrants to students' models. Students' forecast models must use raw and free data.

Hourly load data are available for January 1, 2011 to December 31, 2020. It is not necessary to use all available data. Each team must properly argue its choice for the conditions in their area.

Data are provided in GMT, but *days* must be defined in local time. The latter depends on the area and some manipulation of the time is required before the peak hourly load for each day can be extracted.

¹The SPP provides the hourly load in areas of the pool in `.csv` format, one file per month. The file format has changed over the years, so have the areas included in the pool. Extracting the data is challenging. Usually, the project requires that the data be extracted from the organization's website. Exceptionally this year, the data are provided as an R object so that more of the students' time can be spent in other components of the project.

Part I of the written report must be submitted **no later than noon on February 12, 2021**. It covers:

- an introduction
- exploratory data analysis
- evaluation of naive methods
- description of possible explanatory variables, clearly indicating the source of the data, and providing some summary statistics

Included in the submission:

- Part I of the report, in **.pdf** format:
 - with AT MOST 12 PAGES, this includes all tables, figures, and any appendices;
 - in Times New Roman font, 12 pt, double-spaced, “letter” page format with minimum 1 inch margins.

3 marks (out of 30) are automatically deducted if the above instructions are not respected.

- the **.csv** files of historical data for explanatory variables (other than temperature),
- all R codes, in ASCII format as **.R** files, necessary to reproduce the results therein.

Part II of the written report must be submitted **no later than noon on March 19, 2021**. It covers:

- evaluation of smoothing methods
- evaluation of regression methods

Included in the submission:

- Part II (only) of the report, in **.pdf** format:
 - with AT MOST 10 PAGES, this includes all tables, figures, and any appendices;
 - in Times New Roman font, 12 pt, double-spaced, “letter” page format with minimum 1 inch margins.

4 marks (out of 35) are automatically deducted if the above instructions are not respected.

- all R codes, in ASCII format as **.R** files, necessary to reproduce the results therein.

Part III of the written report must be submitted **no later than noon on April 16, 2021**. It covers:

- evaluation of ARIMA/SARIMA models
- overall comparisons
- recommendation for users
- a conclusion

Included in the submission:

- Part III (only) of the report, in .pdf format:
 - with AT MOST SEVEN PAGES, this includes all tables, figures, and any appendices;
 - in Times New Roman font, 12 pt, double-spaced, “letter” page format with minimum 1 inch margins.

4 marks (out of 35) are automatically deducted if the above instructions are not respected.

- all R codes, in ASCII format as .R files, necessary to reproduce the results therein.

=====

IMPORTANT

R code submitted by students must:

- (1) contain no more than 70 characters per line;
- (2) have properly defined functions. All computations carried out in functions must only require objects that are passed as arguments of the function.

If code that does not satisfy these two requirements is submitted,

- it will not be debugged to assess the nature (minor/major, comprehension, coding, bad math, etc) of errors in solutions provided by students;
- any errors in the solutions provided by students will be assumed to be the result of a major issue with comprehension, and will be penalized as such.

Only R packages used in the R codes supplied by the teacher are allowed.

It is strictly prohibited to use all other R packages in the term project.

=====

The written part of the project is marked out of 100 points following:

Part I: 30 points

Part II: 35 points

Part III: 35 points

The written part of the project represents 30% of the final mark in the course.

=====

The presentation part of the project represents 10% of the final mark in the course.

Students present their projects on April 14, 2021.

Additional information about the presentation's requirements and the marking scheme will be supplied in a separate document later.

=====

Do your best work and good luck!