Open source paradigm: a use case of the climate modelling for power engineering problems

E V Fedotova
Moscow Power Engineering Institute, Krasnokazarmennaya str.14, Moscow, 111250 Russia, www.mpei.ru
e.v.kasilova@gmail.com

Abstract. The open source development paradigm was employed to develop a computational tool to process daily air temperature data. The computational workflow included data preprocessing, clustering procedure, as well calculation and visualization of the energy-related climate characteristics. Applicability of the developed tool was demonstrated considering an analysis of the regional features related to the heating and cooling energy demand indicators. It has been found that the considered locations have distinctively different patterns of the heating and cooling degree-days, despite of the very close values of the annual average temperatures.

1. Introduction
An impact of the climate change on operation of the power equipment and the power systems is still evident around the world and will be only more pronounced during the whole twenty-first century [1]. These changes should be necessarily taken into consideration when solving engineering problems related to the systems that will last on the mid- and long-term time horizon. That is often the case for the power engineering where a proper match between the future climate conditions and the operation regimes of the power equipment should be ensured. For example, a climate-related decrease in the domestic heating demand may be essential for an efficient operation of the district heating plants and must be included into consideration when planning the district heating networks [2]-[3]. Another application is the assessment of climate factors governing renewable power generation, such as the wind speed distribution that is essential for estimation of the wind turbines capacity factors.

2. Open source energy modelling

2.1. State-of-the-art
An accurate assessment of the actual climate conditions is gaining importance for the modern power simulation workflow. Proper simulation tools are needed to satisfy this requirement. An extensive development of the open source paradigm provides a plenty of options to answer these challenges. The ideas of the code reuse and the distributed development collaboration has resulted in a dramatic availability rise of the most advanced implementation and simulation concepts. The open source data analysis tools are gradually getting a standard in climate research and often outperform the commercial products.

A plenty of the general-purpose computational approaches are available nowadays as open source products for the use and further development, including the most advanced modelling methods [4]-[5]. The development of the open source energy-specific modelling tools is impressive [6] but they are still
rather underrepresented as compared with the other branch-specific tools such as geospatial, financial or biological computational packages. Our work is aimed at contributing to fill this gap.

2.2. Practical problem under consideration

The aim of our work was to take an advantage of the open source paradigm and create a computation tool for quantitative assessment of the climate variables which influence operation of the power equipment. Applicability of the proposed approach was demonstrated on the most typical energy systems problem related to the climate change, namely, the climate-governed dynamics of the heating and cooling energy demand.

A shift of the domestic energy consumption is one of the most pronounced climate change effects observed in the energy systems located in moderate and severe climate zones. A noticeable heating demand decrease determines the overall positive effect of the climate change on the energy balance in these areas [7]. An increase in the cooling demand contributes to the summer maxima of the electricity consumption and belongs to the main urban vulnerability mechanisms towards the climate change [8].

The heating and cooling energy demand is usually quantified using a classic heating and cooling degree-days (HDD and CDD, respectively) concept. The HDD and CDD values are proportional to the heat amount which should be supplied to or released from a domestic space to maintain a comfort temperature inside. Analysis of the regional and temporal HDD and CDD dynamics is essential to develop adaptation strategies for the energy systems.

The HDD and CDD parameter calculations are based on the daily air temperatures that are the most essential meteorological records usually available as raw observation data which processing may be quite sophisticated due to data quality issues. The practical purpose of our work was to develop a computational approach which would allow us to process the original meteorological observation data and use them to quantify the regional heating and cooling demand patterns reflecting the real climate dynamics.

3. Methods

3.1. Dataset

Daily air temperature observations were used as the input data. The data were measured by the official terrestrial observation network of Russian meteorological Office (Roshydromet) [9]. The data archive contains the observations obtained on more than five hundred Russian meteorological stations with the earliest records corresponding to the second half of the nineteenth century.

3.2. Algorithm

The computational workflow was organized in three consecutive steps:

1) data preprocessing;
2) clustering of the meteorological time-series;
3) calculations on the local meteorological characteristics and results visualization.

The data preprocessing included quality checks and preparation of a data chunk for further calculations. The target of this step was to obtain the continuous time series of the considered parameter for all the observation stations that provide enough data to reconstruct the missed values in a proper way. The linear interpolation was used to fill the missed data chunks if the number of the consequently missed days was less than a predefined threshold. Another option used to expand an amount of the data available for analysis was to estimate a daily temperature as a mean of the daily minimum and maximum values which were available for more observation locations as compared with the daily mean values data.

The resulted array of the daily temperature time-series was used as an input for the clustering procedure. The hierarchical clustering was used as a computational approach. A distance metric and a clustering algorithm were a subject of optimization which had a purpose to obtain a physically meaningful result. The Euclidean distance and the Ward agglomerating clustering method [10] were
found to give both an appropriate spatial distribution of the station in each cluster and distinctively different average time-series of each cluster temperature profile.

Evaluation of the energy-related climate parameters was the final workflow step. We have chosen the seasonal temperatures and the HDD and CDD values to address the heating and cooling demand dynamics in the considered regions. Both the mean values and the time-series patterns were calculated to demonstrate the applicability of the developed approach to detailed analysis of the energy specific climate parameters.

3.3. Implementation
The R programming language [11] was used for implementation of the described computational workflow. During the last ten years R along with Python became a standard de-facto of the data science. R was designed for statistical calculations and is intended to facilitate implementation of the ideas in the program code for the developers without a professional programming background. Such features make R a perfect tool for scientific and engineering calculations.

We were following the code reuse principle trying to use the existing programming packages whenever possible. Particularly, application of the zoo package [12] has facilitated the data preprocessing step that is the most demanding stage in the majority of the data science problems. The zoo was developed to process time-series data and was employed to select contiguous data chunks as well as to check and fill the missed data.

Another data science concept which we have been following was the tidy data paradigm [13]. The essence of this approach is organizing of the multi-dimensional variables to a special structure which allows for a straightforward application of the statistic calculation procedures. The tydiverse package provides functionality to easily implement this concept as a program code [14].

4. Case Study
The object of our analysis was to assess the regional peculiarities of the heating and cooling energy demand variability and change. The developed computational approach was applied to quantify the differences between the locations which have close integral climate parameters, but represent different climate patterns.

4.1. Selection of the locations to study
The developed computational toolkit was used to cluster all the daily temperature data available between 1950 and 2017. The base years for the mean calculations were 1968-2017.

![Figure 1. Location of the stations selected for the case study.](image)
| Region                | Koinas       | Ivdel        | Mangut       |
|-----------------------|--------------|--------------|--------------|
|站长群               | Archangelsk region | Severdlovsk region | Zabaykalski region |
|年度温度, °C          | 0.29         | 0.29         | 0.32         |
|冬季温度, °C          | -14.6        | -17          | -19.2        |
|夏季温度, °C          | 13.7         | 15.3         | 16.9         |

Three meteorological stations selected for the case study belong to three different clusters but have very close annual temperature values (table 1). The Koinas station is located not far from the White Sea coast (figure 1) where the Arctic amplification of climate warming is strongly manifested. The Ivdel station belongs to a Southern-Ural cluster which is characterized by stabilization of the winter temperatures since 1970 after quite a rapid rise and by a continuous increase in the annual temperature since 1980. The Mangut station is located in the vicinity of the Baikal Lake and belongs to a temperature cluster which demonstrates stabilization of both winter and annual temperatures.

4.2. Results and discussion
The HDD and CDD time series evaluated by the developed calculation toolset for each considered location are presented in the figures 2-3.

Figure 2. Evolution of the heating degree-days in the considered locations. Smoothing with the centered 11-years moving average was applied.
Generally, a decreasing HDD trend and a tendency towards a CDD increase are evident for all the considered locations on the multidecadal scale. The highest decrease in the heating demand is observed for Koinas that could be expected with the well-known Arctic amplification effect. The strong rise of the CDD value for the Mangut station may be explained with the generally higher summer temperatures as compared with other considered stations.

Figure 3. Evolution of the cooling degree-days in the considered locations. Smoothing with the centered 11-years moving average was applied.

The year-to-year variations of the HDD and CDD values are likely connected with the known climate natural variability modes. Particularly, an increase in the HDD value which was observed for all stations between 1950 and 1970 are likely to be attributed as manifestation of the Atlantic Multidecadal Oscillation.

The quantitative estimation of the HDD and CDD differences between the considered locations has shown that these differences are strongly dependent on the scale of the time-averaging. The relative HDD difference between the considered locations may be up to 15% during the single year, but is decreased to 4-6% for smoothing with the 5-11 years moving average and is only about 1-2% if a 30 years smoothing window is taken. However, the CDD values are distinctively different (30-50%) even on the multidecadal time scale. So, there are considerable regional dissimilarities of the cooling demand combined with quite similar heating demand in the considered areas. This means that the adaptation strategies in the considered regions should be distinctively different despite very close annual temperature values in these areas.
5. Summary
The work has demonstrated an application of the open source paradigm to the energy-related problems. The computational workflow was developed using the data science concepts created as a part of the R language infrastructure. The regional peculiarities of the domestic heating and cooling demand were assessed with the developed computational toolkit to demonstrate viability of the developed approach.

The hierarchical clustering was found to give meaningful results for the large-scale patterns of the air daily temperatures as well as the heating and cooling energy demand indicators under a condition of proper setting of the clustering parameters. Particularly, the Ward clustering method with Euclidian distance measure was found to give quite a satisfactory result.

Relevance of the proposed approach was demonstrated in a case study focused on the regional differences of the HDD and CDD patterns for the areas with the close values of the annual air temperature. It has been shown that the dynamics of the heating and cooling energy demand is clearly different for the considered regions despite almost the same integral climate parameters in all of them. The difference between the ten-years averages was shown to be up to 5% for HDD and as high as an order for CDD.

Acknowledgments
This study is supported by the Russian Science Foundation (project No 18-79-10255).

References
[1] IPCC 2013 Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change ed T F Stocker , D Qin, G-K Plattner, M Tignor, S K Allen, J Boschung, A Nauels, Y Xia, V Bex and P M Midgley (Cambridge: Cambridge University Press)
[2] Andric I, Fournier J, Lacarrriere B, Le Corre O, Ferrao P The impact of global warming and building renovation measures on district heating system techno-economic parameters 2018 Energy 150 926—937
[3] Andric I, Gomes N, Pina A, Ferrao P, Fournier J, Lacarrriere B, Le Corre O Modeling the long-term effect of climate change on building heat demand: Case study on a district level 2016 Energy and Buildings 126 77—93
[4] Chollet F and others 2015 Keras https://keras.io (accessed 1.12.2019)
[5] Martin A, Ashish A, Paul B et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015 tensorflow.org.
[6] Pfeffinger S, Hirth L, Schlech I et al Opening the black box of energy modelling: Strategies and lessons learned 2018 Energy Strategy Reviews 19 63—71
[7] Klimenko V, Fedotova E, Tereshin A Vulnerability of the Russian power industry to the climate change 2018 Energy 10 1010—22
[8] Demchenko P, Ginzburg A Influence of Feedbacks in the Climate–Energetics System on the Intensity of an Urban Heat Island 2018 Izvestiya, Atm and Oc Phys 54 4 313—321
[9] Bulygina O N, Razuvaev V N Daily Temperature and Precipitation Data for Russian Meteorological Stations meteo.ru
[10] Murtagh F, Legendre P Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? 2014 Journal of Classification 31 3 274–295
[11] R Core Team 2019 R: A language and environment for statistical computing. R Foundation for Statistical Computing (Vienna, Austria) https://www.R-project.org/ (accessed 1.12.2019)
[12] Achim Zeileis and Gabor Grothendieck zoo: S3 Infrastructure for Regular and Irregular Time Series. Jour of Stat Softw 2005 14 6 1—27
[13] Wickham H Tidy Data Jour of Stat Softw 2014 59 10
[14] Hadley Wickham 2017 tidyverse: Easily Install and Load the ‘Tidyverse’. R package version 1.2.1 https://CRAN.R-project.org/package=tidyverse (accessed 1.12.2019)