
Long Term Electricity Demand & Peak Power Load Forecasting Variables Identification & Selection

Burak Omer Saracoglu

Independent Scholar, Istanbul, Turkey

Email address:

urakomersaracoglu@hotmail.com

To cite this article:

Burak Omer Saracoglu. Long Term Electricity Demand & Peak Power Load Forecasting Variables Identification & Selection. *Science Journal of Circuits, Systems and Signal Processing*. Vol. 6, No. 2, 2017, pp. 18-28. doi: 10.11648/j.cssp.20170602.13

Received: March 12, 2017; **Accepted:** March 29, 2017; **Published:** November 28, 2017

Abstract: Electricity demand (kilowatt hour: kWh) and peak power load (kilowatt: kW) forecasting is very important for not only expansion planning purposes (long term), but also for dispatching purposes (short term). Hence, from the long term forecasting perspective to the very short term forecasting perspective, the nature of electricity demand and the peak power load forecasting has to be studied and understood very well. At first, the problem has to be understood very well, then the solution of this problem has to be studied and solved. These activities are in the scope of this research, development, demonstration, & deployment (RD³) studies. The author thinks that the natural mechanisms of electricity demand and peak power load forecasting problem can be understood very well by finding, defining, identifying, and describing the factors (parameters, variables) that affect the electricity demand and peak power load. In this study, GATE is only used during corpus development as a backup check. R text mining package (Rtm) and TextSTAT are used as main text mining and analysis tools. 314 terms as candidate variable terms are found by this text analysis. Afterwards, all variables are studied and analyzed by a grey based natural reasoning with simple weighted average approach (WA) (only for long term factors as preliminary in this application) (on way of simple additive weighting method: SAW). Finally, 43 terms (e. g. population, weather, climate, economy, price) for variables are found for infant and mature RD³ studies of 100% renewable energy (RE) worldwide grid (Global Grid). Findings of this study can also be used in other grid types. It is believed that a specific dictionary and encyclopedia in this particular subject should be developed for researchers common sense which will also help building of the Global Grid Prediction Systems (G²PS).

Keywords: Grey, Forecast, Global Grid, Natural Reasoning, Power, Text Analysis

1. Introduction

Humankind lives with electricity today. It is consumed almost anywhere. It is generated with different methods. On one hand, there is non-renewable energy sources (NRES) group, which pollutes our World in every aspects and threatens our health and wildlife very much, on the other hand, there is renewable energy sources (RES) group, which is much cleaner than NRES group and threatens our World much less. In any case, all of these technologies (NRES or RES) change the environment. The question is about the amount and the effects of these changes. The main cause of change in the environment is humankind. None of the human activities is perfect. Human activities and decisions weren't good in past, they aren't good now, and they shall never be good enough in future.

Author of this study thinks that RES group should be preferred and 100% RES based grids should be operated. This research study is based on this main idea. Facts show us that 100% RES based grids may be operational in the long term. This utopia can't be possible by separated and distributed grids in our World according to this study's approach. Hence, subcontinents, continents and World should be connected for clean electricity generation. Interconnections can possibly be with some futuristic wire or wireless technologies (for wireless transmission correspond also to Dr. Prabhat Ranjan Tripathi). There are already some futuristic ideas announced and published for multi-continental and international electricity grids such as European Supergrid [1], Supergrid Concept for America [2], DESERTEC [3], Gobitec [4, 5], Asian Super Grid [4, 5], and Global Grid [6].

This research study focuses only on the Global Grid Concept. It is described as *"a grid spanning the whole planet and connecting most of the large power plants in the world"* [6]. The Global Grid Concept may be operational in the long term, so that research, development, demonstration, and deployment (RD³) studies in several topics such as modeling of the Global Grid should be conducted and brought to scientific literature. One of the important steps of the Global Grid modeling is the electricity demand and peak power load prediction in the long run time horizon. Several methods should be studied and used for the electricity demand forecasting of the Global Grid Concept. Research findings will mainly affect the renewable power plants investments decisions, the Global Grid goals, and the international laws and agreements. Other important steps of the Global Grid modeling is the electricity demand prediction in the medium, short and very short run time horizons. The research findings of these studies will affect the whole Global Grid design. Some electricity demand and peak power load forecasting methods are based on variables (e. g. linear regression, multivariate regression, fuzzy inference system).

This research study aims to find, identify, and select the variables for the electricity demand and peak power load forecasting of the Global Grid (only for long term application and presentation due to research duration limitation).

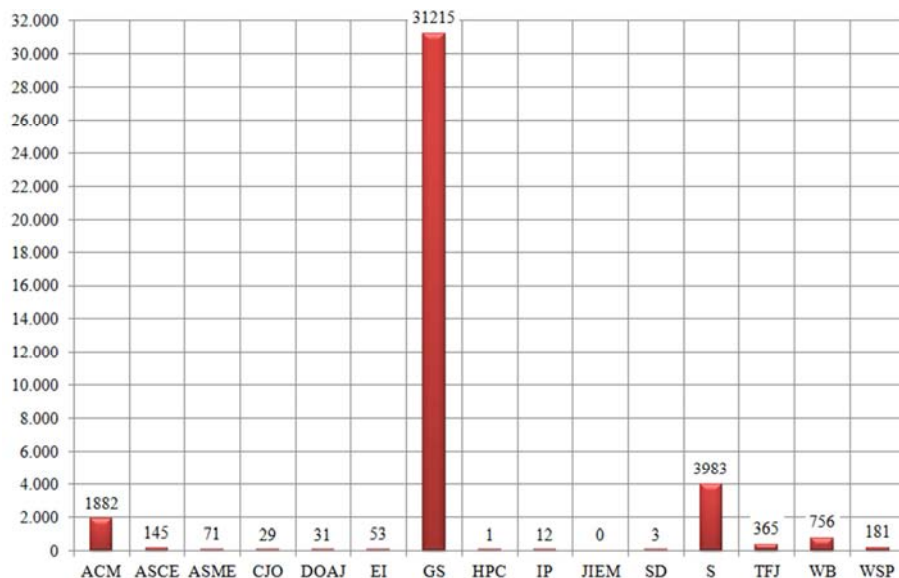
This paper consists of four sections. Second section presents the literature review. Third section presents how variables are identified and selected for the electricity demand and peak power load prediction in the Global Grid by text analysis and grey based natural reasoning with simple weighted average approach. This approach and some identified and selected variables can also be used in grey, linear and non-linear regression and other models in the electricity demand and peak power load forecasting. The concluding remarks and further research are presented in the last section.

2. Literature Review

The literature review had two main folds. The first one was for finding relevant previous research studies in the electricity demand and peak power load forecasting subject (studies also include fuzzy word). The second one was for finding the appropriate publications that focused on the identification and selection of variables amongst all studies found in the first review step.

In the first review step, some online academic publication websites (totally 15 academic websites) were searched in June 2015 (search period: 20 days). These websites were alphabetically ordered as ACM Digital Library-ACM [7], ASCE Online Research Library-ASCE [8], American Society of Mechanical Engineers-ASME [9], Cambridge Journals Online-CJO [10], Directory of Open Access Journals-DOAJ [11], Emerald Insight-EI [12], Google Scholar-GS [13], Hindawi Publishing Corporation-HPC [14], Inderscience Publishers-IP [15], Journal of Industrial Engineering and Management-JIEM [16], Science Direct-SD [17], Springer-S [18], Taylor & Francis Online Journals-TFJ [19], Wiley-Blackwell/Wiley Online Library-WB [20], and World Scientific Publishing-WSP [21]. It was believed that searching and reviewing on these online academic publication websites was a good start for this research study and its following ones (a drop in the ocean: <http://www.phrases.org.uk/> on web).

The main structure of search queries were based on <fuzzy> <inference> <rule> <control> <electricity> <demand> <forecast>. There were 24 different queries in this review such as "fuzzy inference system" and "electricity" and "demand"; and "fuzzy inference system" and "electricity" and "forecast". The highest documents hit was on the Google Scholar by 31215 search results. The second highest hit was on the Springer by 3983 search results. The third highest hit was on the ACM Digital Library by 1882 search results. Other online academic publication websites followed these three websites by the number of search results under 1000 ("Figure 1").



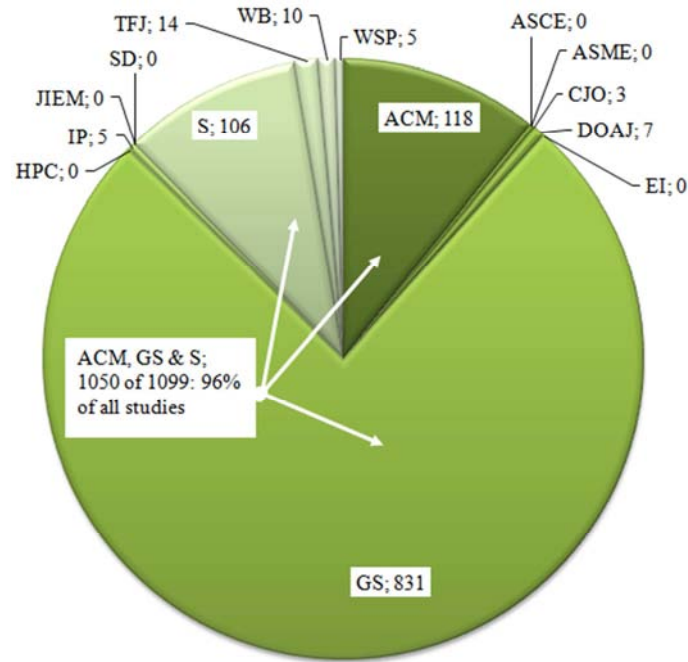


Figure 1. Literature review summary (ACM [7], ASCE [8], ASME [9], CJO [10], DOAJ [11], EI [12], GS [13], HPC [14], IP [15], JIEM [16], SD [17], S [18], TFJ [19], WB [20], WSP [21]), visualization generated by the column graph of Microsoft Office Excel 2007 (<http://www.microsoft.com/en-us/>) & Paint.NET (<http://www.getpaint.net/index.html>) (top: all search hits; bottom: close relevant studies).

Duplications, triplications, quadruplications, etc. were included in these counts. All documents under these search results were reviewed by their titles, abstract, and keywords. It was observed that there were documents in different languages such as Spanish, Chinese and Japanese. Sadly only English documents could be reviewed in the current study (not any Arabic: العربية اللغة, Chinese: 达伟/汉语 / 汉语/华语 / 華語 / 中文, Russian: Русский, etc.) (visit <http://www.freelang.net/expressions/languages.php> for languages).

The first documents elimination was performed in this review step. The irrelevant studies weren't downloaded and stored in their specific folders. The journal and conference papers were only taken into account. The studies, that were related with the electricity demand forecasting, were also analyzed based on the websites.

The first three online academic publication websites were the Google Scholar, the ACM Digital Library, and the Springer. When these three websites were searched 96% of all research studies in this subject could be found. The Google Scholar covered 76% of all research studies in this subject. Henceforth, it was concluded that reviewing only on the Google Scholar was almost enough in this subject. Above all, reviewing on the Google Scholar, the ACM Digital Library, and the Springer were representing almost all previous research studies in this subject.

In the second review step, the saved and stored documents in the specific folder of this study were reviewed to find publications on the identification of variables. It was observed that there weren't any studies in this subject in this current collection. However, there were a few literature review and survey papers, that could be visited as for the previous studies.

One of the interesting literature review conference papers was by Elakrmi and Abu Shikhah [22]. It was found on the Jordan Engineers Association Conferences Websites (<http://www.jeaconf.org/>) reached by the Google Scholar. They mentioned that weather, cultural and social factors were used in the models. They also presented the methods in the literature in this subject. These methods were statistical-based methods, regression methods (single variable linear, polynomial, selected-model function such as exponential, logarithmic, and multi-variable), time series methods (ARMA: autoregressive moving average, ARIMA: autoregressive integrated moving average or Box-Jenkins, ARIMAX: autoregressive integrated moving average with exogenous variables, FARMAX: fuzzy autoregressive moving average with exogenous input variables, exponential smoothing, PCA: principal component analysis, similar-day approach, econometric or causal approaches, simulation or end-use approaches), artificial intelligence (AI) based methods (neural networks: ANN or NN, expert systems, fuzzy logic systems, support vector machines: SVM, particle swarm optimization: PSO) in their paper. They only mentioned two variables [22].

Another important study was by Hahn et. al. [23]. It was found on the European Journal of Operational Research journal website reached by the Google Scholar and the Science Direct. Their review was similar to Elakrmi and Abu Shikhah's approach. Other review studies had also similar review approaches.

This literature review was finalized in 29/08/2015 (search period in June 2015). It was understood that the researchers had never studied the identification and selection of variables for the electricity demand and peak power load forecasting of the Global Grid until June 2015.

This research study would hopefully give a good and interesting start for the scientific publications in this research topic.

The application in this study was only performed on the long term time horizon publications due to the limited time budget. The very-short, short, and medium term time horizon publications were left for the following research studies.

3. Identification & Selection of Variables

This study is performed and finalized under some

limitations and preferences, such as the personal computer capacity limitation (Windows 10 Pro, Intel(R) Core(TM) i5 CPU 650 @ 3.20 GHZ, 6, 00 GB RAM), the internet connection speed limitation, the free software license or free and open-source software license applications and tools (offline or online) usage preference (academic study approach), and finally duration and time (no major coding, scripting, thousands of lines of code: KLOC) limitation. The simplified overall overview of the text analysis is presented in “Figure 2”.

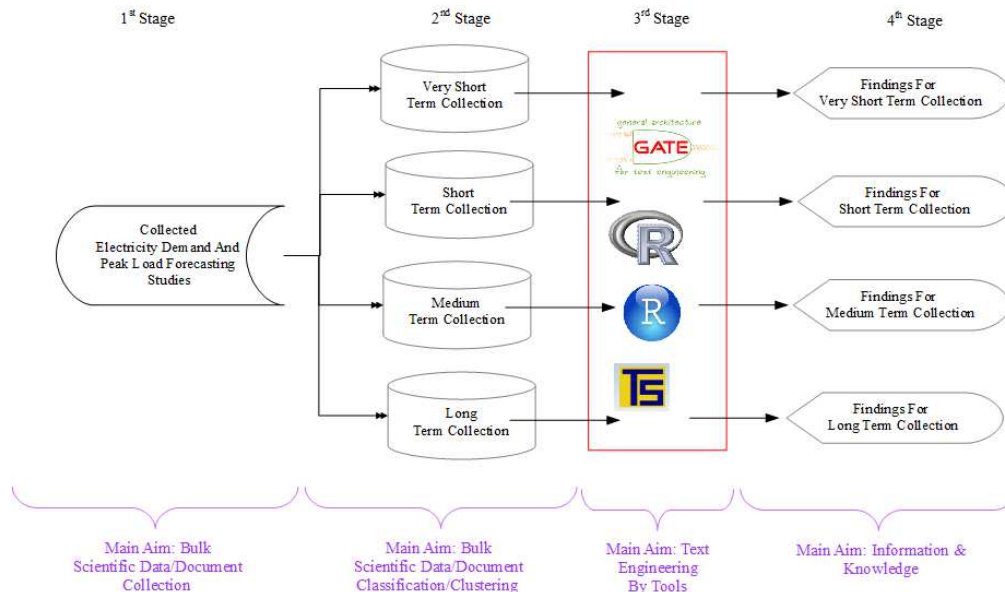


Figure 2. Simplified general overview of text analysis, visualization generated by Apache OpenOffice Draw 4.1.2 (<https://www.openoffice.org/product/draw.html>) & Paint.NET.

The files are saved in an organized manner into four classified folders (long, medium, short and very short term collections). Only some of the documents (totally 220) can be classified and analyzed in this study. First, all files are converted into *. pdf format. Second, these files are merged by the PDF Split and Merge basic (<http://www.pdfsam.org/>). Third, this merged file for each classified folder is saved to its special folder. Fourth, these files are converted into *. txt and

*. docx format by the PDF to Text (<http://pdftotext.com/>) and the PDF to DOC converter (<http://pdf2doc.com/>). Fifth, *. txt and *. docx files are compared and the most appropriate one is selected for the text analysis. These activities can be performed without any major trouble, because of having rather small number of files. When the number of files increases, the activities will take much longer. These data file preprocessing steps are presented in “Table 1”.

Table 1. File preprocessing of this study (01/09/2015- 05/10/2015).

Properties	Collections	Long Term	Medium Term	Short Term	Very Short Term
number of total *. pdf files		43	29	129	17
total size of *. pdf files (MB)		20, 3	20, 6	61, 6	19, 9
time of merging (sec)		6	6	11	1
size of merged *. pdf file (MB)		19, 9	20, 5	61, 4	20, 2
number of A4 pages		323	322	1210	285
time of conversion *. pdf to *. txt		180	180	480*	180
size of merged *. txt file (MB)		1, 1	0, 9	3, 2	0, 7
time of conversion *. pdf to *. docx		360	360	1140**	240
size of merged *. docx file (MB)		9, 9	19, 4	55, 9	26, 3
selected file for analysis		*txt	*txt	*txt	*txt

* 27 MB and 35 MB, 2 pdf files conversion duration 180 sec and 300 sec, ** 27 MB and 35 MB, 2 pdf files conversion duration 540 sec and 600 sec, MB: megabyte, sec: second

The conversion of *. pdf to *. txt is also performed by GATE Developer 8.0 (<https://gate.ac.uk/>) (language resources

GATE document). The converted *. txt files are compared and the most appropriate one according to the author's view is

selected as the corpus of this research (Corpus: "a collection of writing" [24-31]).

The corpus of this research is the documents that are collected and classified according to their forecasting period (long term collection). For instance, the *.txt file of the whole documents in the long term collection is the long term collection corpus (see in online electronic supplementary material files: ESM).

The *.txt files are used in this study, because the *.txt files can be loaded into the tools without any major effort (tools: TextSTAT: <http://neon.niederlandistik.fu-berlin.de/en/textstat/>, R tm: <https://cran.r-project.org/web/packages/tm/index.html>).

The simple text analysis is performed by the R tm and the TextSTAT. The word frequencies are calculated. The R needs to be scripted by help of installed R packages (developed and presented by R developers community) as presented in "Table 2", however the TextSTAT doesn't need any scripting at all. There are some filtrations in the R script (remove numbers, stopwords), but there isn't any filtration in the TextSTAT in this study. Some documents used during this study for the TextSTAT are [32, 33]. Some documents for the R are [34-37].

Table 2. Script for RStudio Version 0.99.491 Editor - load & matrices.

```
# Necessary packages that will be installed
Needed <- c("rJava", "RWeka", "RWekajars", "SnowballC", "tm",
"ggplot2")
# Install packages
install.packages(Needed, dependencies=TRUE)
# Stored files location on the hard disk
long.term.collection<-file.path("C:/...../Literature Review EN Long
Term/txt")
long.term.collection
# Check to see that loaded texts
length(dir(long.term.collection))
dir(long.term.collection)
# Load the R packages for text mining
library(rJava)
library(RWeka)
library(RWekajars)
library(SnowballC)
library(tm)
getReaders()
getSources()
# Load texts into R
long.term.collection.docs <- Corpus(DirSource(long.term.collection))
# Check to see that loaded texts into R
class(long.term.collection.docs)
summary(long.term.collection.docs)
inspect(long.term.collection.docs)
# Perform simple transformation
getTransformations()
long.term.collection.docs<- tm_map(long.term.collection.docs, tolower)
long.term.collection.docs<- tm_map(long.term.collection.docs,
removePunctuation)
long.term.collection.docs <- tm_map(long.term.collection.docs,
removeNumbers)
length(stopwords("english"))
stopwords("english")
long.term.collection.docs<- tm_map(long.term.collection.docs,
removeWords, stopwords("english"))
inspect(long.term.collection.docs)
alphabet=c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j", "k", "l", "m", "n",
"o", "p", "q", "r", "s", "t", "u", "v", "w", "x", "y", "z")
```

```
long.term.collection.docs<- tm_map(long.term.collection.docs,
removeWords, alphabet)
inspect(long.term.collection.docs)
long.term.collection.docs<- tm_map(long.term.collection.docs,
stripWhitespace)
# Perform stemming
getStemLanguages()
long.term.collection.docs<- tm_map(long.term.collection.docs,
stemDocument)
# Perform plaintext conversion
long.term.collection.docs<- tm_map(long.term.collection.docs,
PlainTextDocument)
inspect(long.term.collection.docs)
# Create document term matrices
long.term.collection.document.term.matrix<-
DocumentTermMatrix(long.term.collection.docs)
inspect(long.term.collection.document.term.matrix)
long.term.collection.docs.term.document.matrix<-
TermDocumentMatrix(long.term.collection.docs)
inspect(long.term.collection.docs.term.document.matrix)
long.term.collection<-
as.matrix(long.term.collection.docs.term.document.matrix)
# Set working directory
setwd("C:/...../Literature Review EN Long Term")
getwd()
# Write and export csv file into working directory
write.csv(long.term.collection, file="R.long.term.collection.csv")
```

Note 1: Script for long. term. collection in R.

Note 2: For library documents and their references visit R packages.

Note 3: "C:/...../Literature Review EN Long Term" directory or folder of the documents

This R script in "Table 3" helps to find out the words used by the academics and researchers in their publications. The words are presented with their usage frequencies.

Table 3. Script for RStudio Version 0.99.491 Editor - frequency & word cloud.

```
# Set working directory
setwd("C:/...../Literature Review EN Long Term")
getwd()
# Read CSV files
long.term.collection.forecast.terms.R.TextSTAT <-
read.csv(file="long.term.collection.forecast.terms.R.TextSTAT.csv",
head=TRUE,sep=",")
long.term.collection.forecast.terms.R.TextSTAT.above100<-read.csv(file="
long.term.collection.forecast.terms.R.TextSTAT.above100.csv",
head=TRUE,sep=",")
long.term.collection.forecast.terms.R.TextSTAT.10to100<-
read.csv(file="long.term.collection.forecast.terms.R.TextSTAT.10to100.csv
", head=TRUE,sep=",")
long.term.collection.forecast.terms.R.TextSTAT.below10<-
read.csv(file="long.term.collection.forecast.terms.R.TextSTAT.below10.cs
v", head=TRUE,sep=",")
# Show data at CSV files
long.term.collection.forecast.terms.R.TextSTAT
long.term.collection.forecast.terms.R.TextSTAT.above100
long.term.collection.forecast.terms.R.TextSTAT.10to100
long.term.collection.forecast.terms.R.TextSTAT.below10
# Plot data
plot(long.term.collection.forecast.terms.R.TextSTAT[, "terms"],
long.term.collection.forecast.terms.R.TextSTAT[, "occurrences"])
plot(long.term.collection.forecast.terms.R.TextSTAT.above100[, "terms"],
long.term.collection.forecast.terms.R.TextSTAT.above100[, "occurrences"])
plot(long.term.collection.forecast.terms.R.TextSTAT.10to100[, "terms"],
long.term.collection.forecast.terms.R.TextSTAT.10to100[, "occurrences"])
```



```

plot(long.term.collection.forecast.terms.R.TextSTAT.below10[, "terms"],
long.term.collection.forecast.terms.R.TextSTAT.below10[, "occurrences"])
# Plot data by ggplot2
library(ggplot2)
ggplot(long.term.collection.forecast.terms.R.TextSTAT,aes(x=terms,y=occ
urrences))+geom_bar(stat="identity",fill="darkred",colour="darkgreen")+t
heme(axis.text.x=element_text(angle=90,hjust=1))
# Word Cloud
word.cloud<-file.path("C:/...../Literature Review EN Long
Term/word cloud txt")
length(dir(word.cloud))
dir(word.cloud)
library(tm)
word.cloud.terms.R.TextSTAT <- Corpus(DirSource(word.cloud))
summary(word.cloud.terms.R.TextSTAT)
inspect(word.cloud.terms.R.TextSTAT)
word.cloud.terms.R.TextSTAT.Document.Term <-
DocumentTermMatrix(word.cloud.terms.R.TextSTAT)
inspect(word.cloud.terms.R.TextSTAT.Document.Term)
word.cloud.terms.R.TextSTAT.Term.Document<-
TermDocumentMatrix(word.cloud.terms.R.TextSTAT)
inspect(word.cloud.terms.R.TextSTAT.Term.Document)
m<-as.matrix(word.cloud.terms.R.TextSTAT.Term.Document)
v<-sort(rowSums(m),decreasing=TRUE)
d<-data.frame(word = names(v),freq=v)
library(wordcloud)
set.seed(1234)
wordcloud(words=d$word,freq=d$freq,scale=c(4,0.5),min.freq=1,max.wor
ds=Inf,random.order=FALSE,rot.per=0.35,colors=brewer.pal(8,"Dark2"))

```

Note 1: "C: / / Literature Review EN Long Term" directory or folder of the documents

In this study, one *. csv file for each collection is created as presented (see ESM). These files present the possible meaningful words, terms, phrases, verbs etc. for forecasting in the literature. One *. csv file for each collection is presented (ESM). The same activities are performed by the TextSTAT and the files are presented (see ESM). The counted words, that are found during this text mining/analysis, are presented in "Table 4".

Table 4. Word or term counts (see ESM).

Tools	Collections	Long Term	Online Electronic Supplementary Files
R		1545	R. long. term. collection. csv
TextSTAT		17116	TextSTAT. long. term. collection. csv
R. forecast. terms		249	R. forecast. terms. long. term. collection. csv
TextSTAT. forecast. terms		270	TextSTAT. forecast. long. term. collection. csv
R. TextSTAT. forecast. terms		315*	long. term. collection. forecast. terms. R. TextSTAT. csv

After finalizing the R and the TextSTAT studies, the findings as in files are combined together and only one file is generated. The possible terms are counted by the summation of counted terms in the R and the TextSTAT files. Moreover, some of the terms are eliminated during this activity (manual study). The visualizations are presented by "Figure 3" to "Figure 8".

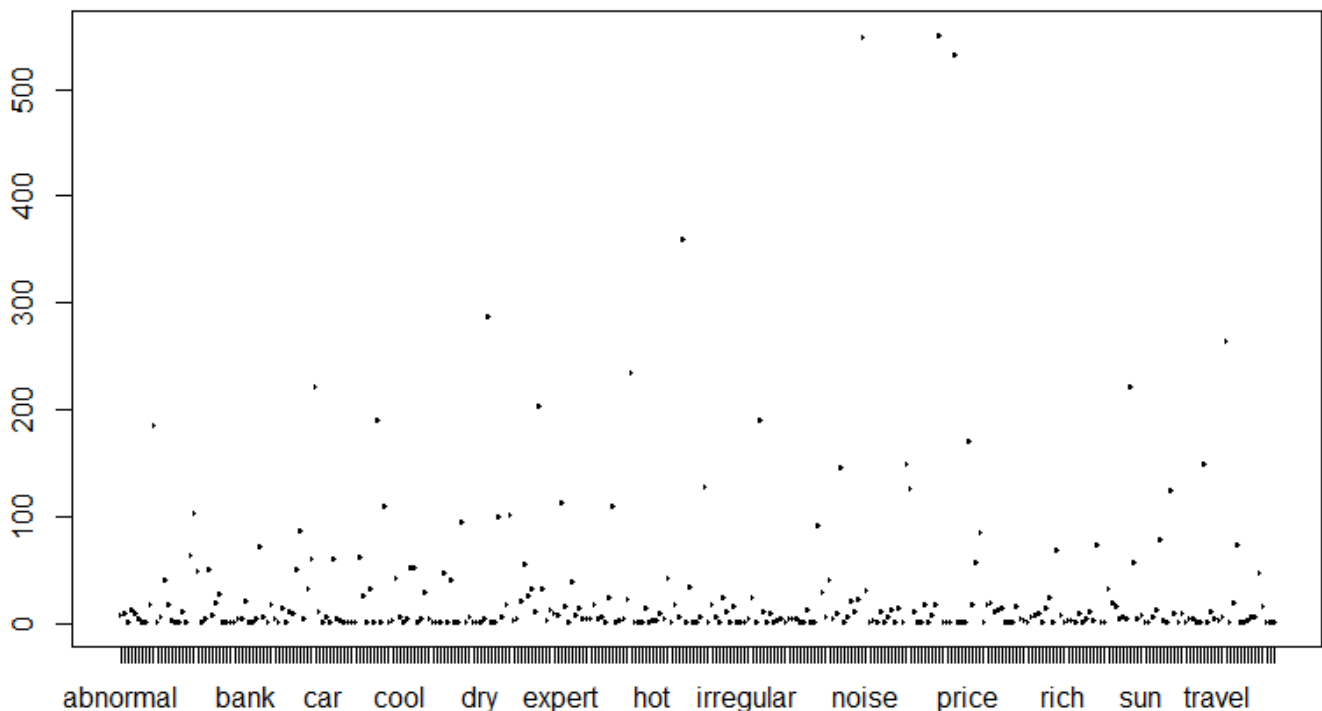


Figure 3. Long term collection forecast terms, whole list, visualization generated by plot of RStudio Version 0.99.491 & Paint. NET.

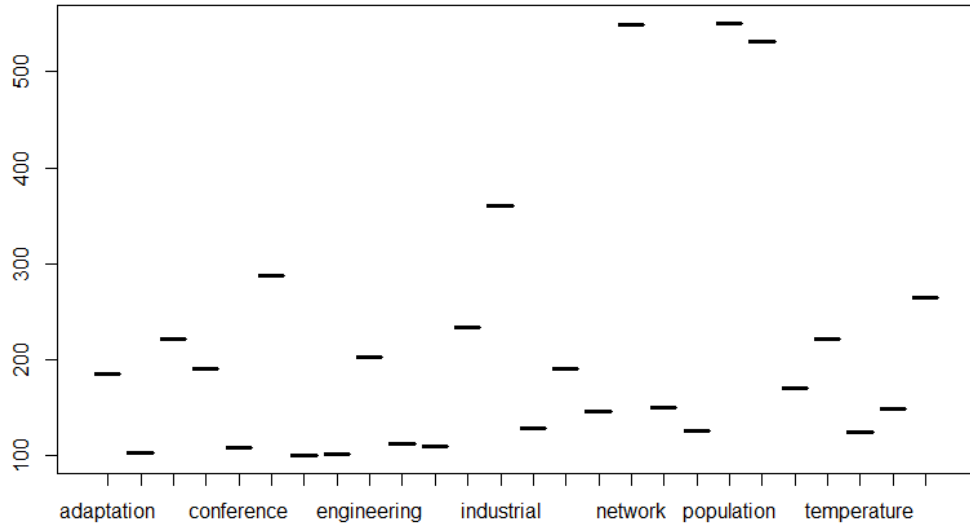


Figure 4. Long term collection forecast terms, most frequent ones: above 100, visualization generated by plot of RStudio Version 0.99.491 & Paint. NET.

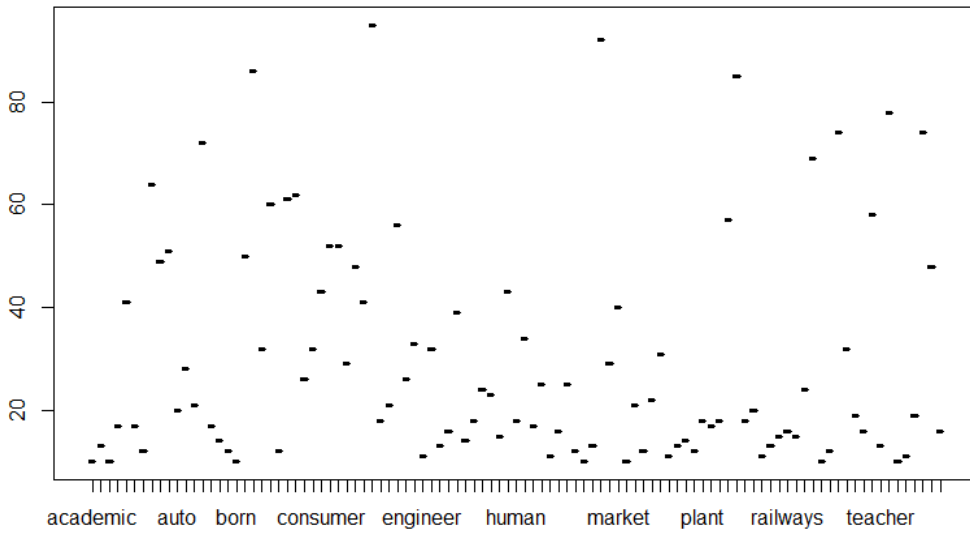


Figure 5. Long term collection forecast terms, mid frequent ones: above 10 and below 100, visualization generated by plot of RStudio Version 0.99.491 & Paint. NET.

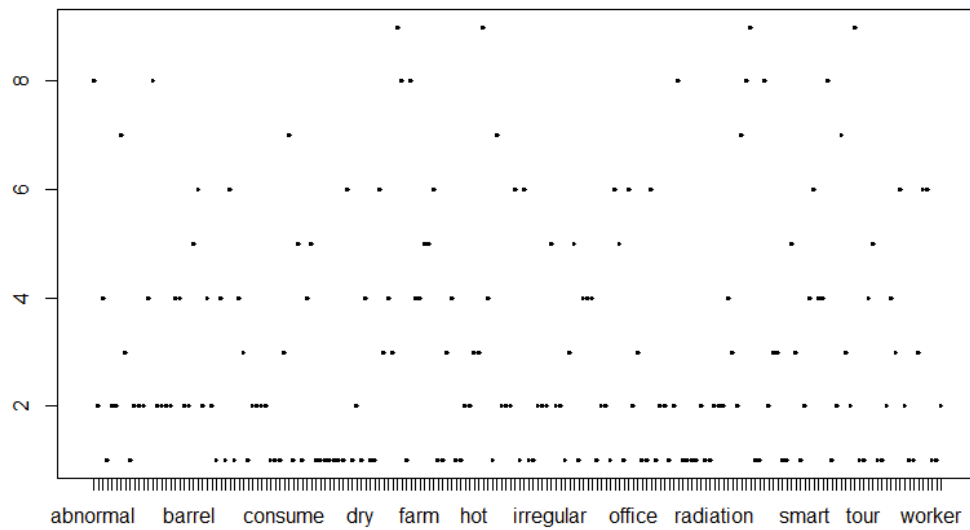


Figure 6. Long term collection forecast terms, least frequent ones: below 10 (bottom right), visualization generated by plot of RStudio Version 0.99.491 & Paint. NET.

Multiplication:

$$\otimes n_1 \times \otimes n_2 = (\min(n_{1a} \times n_{2a}, n_{1\gamma} \times n_{2\gamma}, n_{1a} \times n_{2\gamma}, n_{1\gamma} \times n_{2a}), \max(n_{1a} \times n_{2a}, n_{1\gamma} \times n_{2\gamma}, n_{1a} \times n_{2\gamma}, n_{1\gamma} \times n_{2a})) \quad (3)$$

Division:

$$\otimes n_1 \div \otimes n_2 = (\min(n_{1a} \div n_{2a}, n_{1\gamma} \div n_{2\gamma}, n_{1a} \div n_{2\gamma}, n_{1\gamma} \div n_{2a}), \max(n_{1a} \div n_{2a}, n_{1\gamma} \div n_{2\gamma}, n_{1a} \div n_{2\gamma}, n_{1\gamma} \div n_{2a})) \quad (4)$$

Multiplication by real numbers:

$$k \times (\otimes n_l) = (kn_{1a}, kn_{1\gamma}) \quad (5)$$

Whitenization of grey numbers:

$$\otimes x = [a, \gamma] \text{ then } \widetilde{\otimes} x = \otimes x \alpha + (1 - \otimes) \times \gamma, \otimes \in [0, 1] \quad (6)$$

equal weight mean whitenization: $\otimes = \frac{1}{2}$ [46-48]

In this study, the equations (1), (2), (5) and (6) are applied directly with equal weight mean whitenization, because the weights of candidate variables are not modeled by grey approach (only positive real numbers) as such in “Table 5”:

Table 5. Grey based natural reasoning with WA for future SAW.

Candidate variable preference	Grey Model	Weighted Average
Very Low (VL)	$\otimes_{VL} = [0, 00; 0, 25]$	$W_{\text{impact}}: 0, 5$
Low (L)	$\otimes_L = [0, 20; 0, 45]$	$W_{\text{data}}: 0, 5$
Medium (M)	$\otimes_M = [0, 40; 0, 65]$	Whitenization
High (H)	$\otimes_H = [0, 60; 0, 85]$	$\otimes = 0, 5$
Very High (VH)	$\otimes_{VH} = [0, 80; 1, 00]$	

In the current natural reasoning, not only the relation of variables with electricity demand and peak power, and the impact of variables on models and consequences are analyzed, but also the data availability and possibility, the mindfulness and uniqueness of variables are studied according to the expert decision maker perception.

A simple natural reasoning model is built in this study as

following (“Figure 9” and “Table 5”).

IF term's impact AND data possibility is Very Low THEN Eliminate term directly,

IF term's impact is Very High AND data possibility is Above Low THEN Choose term directly,

ELSE Execute the current grey based WA approach (see ESM and “Figure 9”).

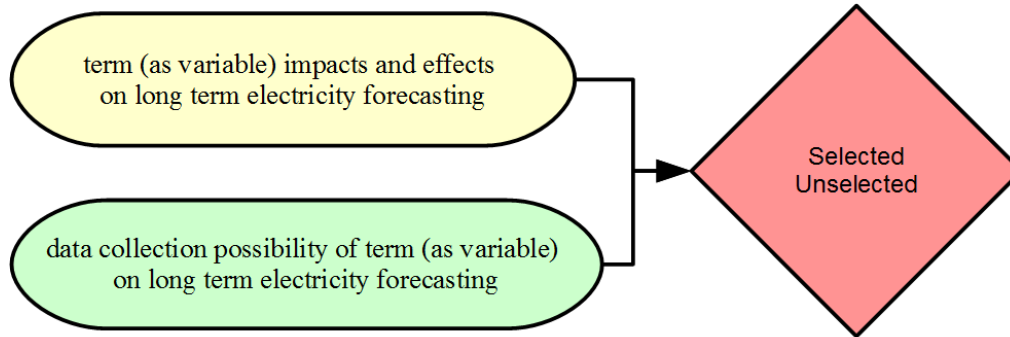


Figure 9. Grey based natural reasoning with WA approach, visualization generated by Apache OpenOffice Draw 4.1.2 & Paint. NET.

There are 21 chosen terms (population, growth, temperature, economy, development, production, weather, climate, wind, employee, air temperature, anomaly, irradiance, irradiation, labor, radiation days, radiation, developing countries, domestic production, economic development, economic growth). There are 22 selected terms (economic, computer, journal, price, employment, bank, publication, tariff, climatic, democratic, earthquake, rainfall, cloudiness, coastal, industrialization, inequality, rich, deserts, internet, power-system, refrigerator, sunshine). Some of these terms can be used directly and some of them can be used with one another. Some of them have similar meanings can be eliminated in further analysis. The ones like population (551), weather (74), climate (61), economy (100), price (170) are most likely useable ones according to the author.

4. Conclusions, Future Applications and Research

This study aims to find the variables for the electricity demand and power load forecasting models (e. g. fuzzy, grey, fuzzy inference). There is only one researcher in all parts of this study. Actually, the number of researchers has to be much higher. In following studies, the number of documents has to be increased in their specific folders. The text analysis activities and the grey based natural reasoning with WA approach have to be reviewed and improved in more detailed analysis. A simple additive weighting method (SAW) should be built for calculating the ranks of terms that represent the

usage preference of terms as the variables. Moreover, a specific dictionary and encyclopedia in this particular subject have to be developed and presented on open access world wide web sites. It is believed that findings of these kinds of studies will help developing the Global Grid Prediction Systems (visit [47]).

Acknowledgements

The author would sincerely like to express his deepest thankfulness to The Absolute Ruler, The King and to Prabhat Ranjan Tripathi, Gina Cook Chiodo, Angus Roberts, Diana Maynard, Ian Fellows, Matthias Hüning, Bernadetta Kwintiana Ane and Riswan Efendi for guidance and help

References

- [1] Friends of the Supergrid "Roadmap to the Supergrid Technologies Final Report".
http://www.cesi.it/news_ideas/ideas/Documents/FOSG%20%20WG2%20Final-report.pdf (2012).
- [2] Overbye, T. J., Starr, C., Grant, P. M., Schneider, T. R. "National Energy Supergrid Workshop Report". Crowne Plaza Cabana Palo Alto Hotel, Palo Alto, California. www.supergrid.illinois.edu/sg1/SuperGridReportFinal.pdf. (2002).
- [3] DESERTEC Foundation, <http://www.desertec.org/> accessed by 11/06/2015. (2015).
- [4] Mano, S., Ovgor, B., Samadov, Z., Pudlik, M., Jülch, V., Sokolov, D., Yoon, J. Y. "Gobitec and Asian Super Grid for renewable energies in Northeast Asia". Spotinov print Ltd. (2014).
- [5] Seliger, B., Kim, G. E. "Tackling climate change, increasing energy security, engaging North Korea and moving forward Northeast Asian integration – "Green Growth" in Korea and the Gobitec project". Gobitec Outline Paper 1-03 10112009. (2009).
- [6] Chatzivasilieiadis, S., Ernst, D., Andersson, G. "The Global Grid", Renewable Energy, Vol: 57, pp.372–383. doi: 10.1016/j.renene.2013.01.032. (2013).
- [7] ACM Digital Library, <http://dl.acm.org/>.
- [8] ASCE Online Research Library, <http://ascelibrary.org/>.
- [9] American Society of Mechanical Engineers, <http://asmedigitalcollection.asme.org/>.
- [10] Cambridge Journals Online, <http://journals.cambridge.org>.
- [11] Directory of Open Access Journals, <http://doaj.org>.
- [12] Emerald Insight, <http://www.emeraldinsight.com/>.
- [13] Google Scholar, <http://scholar.google.com.tr/>.
- [14] Hindawi Publishing Corporation, <http://www.hindawi.com/>.
- [15] Inderscience Publishers, <http://www.inderscience.com/>.
- [16] Journal of Industrial Engineering and Management, <http://www.jiem.org/index.php/jiem>.
- [17] Science Direct, <http://www.sciencedirect.com>.
- [18] Springer, <http://www.springer.com/gp/>.
- [19] Taylor & Francis Online/Journals, <http://www.tandfonline.com/>.
- [20] Wiley-Blackwell/Wiley Online Library, <http://onlinelibrary.wiley.com/>.
- [21] World Scientific Publishing, <http://www.worldscientific.com/>.
- [22] Elakrmi, F., Shikhah, N. A. "Electricity Load Forecasting Science and Practices". (2013)
<http://www.jeaconf.org/UploadedFiles/Document/12b4c17b-6c84-4075-a638-7b34a74afde7.pdf> (accessed in 29/08/2015).
- [23] Hahn, H., Meyer-Nieberg, S., Pickl, S. "Electric load forecasting methods: Tools for decision making". European Journal of Operational Research, Vol: 199, No: 3, pp.902-907. (2009).
- [24] Macmillan Dictionary: corpus
<http://www.macmillandictionary.com/dictionary/british/corpus> accessed on 20/01/2016.
- [25] Oxford Dictionary: corpus
<http://www.oxforddictionaries.com/definition/english/corpus> accessed on 20/01/2016.
- [26] Longman Dictionaries: corpus
<http://www.ldoceonline.com/dictionary/corpus> accessed on 20/01/2016.
- [27] Merriam-Webster: corpus
<http://www.merriam-webster.com/dictionary/corpus> accessed on 20/01/2016.
- [28] Dictionary.com: corpus
<http://dictionary.reference.com/browse/corpus?s=t> on 20/01/2016.
- [29] Cambridge Dictionaries Online: corpus
<http://dictionary.cambridge.org/dictionary/english/corpus> accessed on 20/01/2016.
- [30] Collins Dictionaries: corpus
<http://www.collinsdictionary.com/dictionary/english/corpus> accessed on 20/01/2016.
- [31] Manca, E.: Context and Language eISBN 978-88-8305-092-3 (2012).
- [32] YouTube: sprachtheater, TextSTAT - Tutorial
<https://www.youtube.com/watch?v=juVal2nMWOE&feature=youtu.be> accessed on 20/01/2016.
- [33] Hüning, M.: TextSTAT - Simple Text Analysis Tool. (2001)
<http://neon.niederlandistik.fu-berlin.de/static/textstat/TextSTAT-Doku-EN.html> accessed on 20/01/2016.
- [34] R Core Team: R Language Definition (2015).
- [35] R Pubs by RStudio: Basic Text Mining in R
https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html accessed on 20/01/2016
- [36] Williams, G. "Hands-On Data Science with R Text Mining". (2014).
- [37] Feinerer, I. "Introduction to the tm Package Text Mining in R". (2015).

- [38] Grimmer, J., Stewart, B. M. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts". *Political Analysis*, mps028, pp: 1-31. (2013).
- [39] Loughran, T., McDonald, B. "When is a liability not a liability? Textual analysis, dictionaries, and 10 - Ks". *The Journal of Finance*, Vol.66, Issue: 1, pp.35-65. (2011).
- [40] O'Connor, B. "Artificial Intelligence (AI) and Social Science, Be careful with dictionary-based text analysis", <https://brenocon.com/blog/2011/10/be-careful-with-dictionary-based-text-analysis/> accessed on 21/01/2016.
- [41] Karttunen, L "The logic of infinitival complement constructions" <http://web.stanford.edu/~laurik/presentations/Shonan.pdf> accessed on 23/01/2016.
- [42] Karttunen, L. "From Natural Logic to Natural Reasoning" <http://web.stanford.edu/~laurik/presentations/CICLing.pdf> accessed on 23/01/2016.
- [43] Terr, David: Weighted Mean From Math World A Wolfram Web Resource, created by Eric W. Weisstein. <http://mathworld.wolfram.com/WeightedMean.html> accessed on 22/01/2016.
- [44] Grossman, J., Grossman, M., Katz, R. "The first systems of weighted differential and integral calculus". Archimedes Foundation Box 240, Rockport Massachusetts. (1980).
- [45] Price, G. R. "Extension of covariance selection mathematics". *Annals of human genetics*, Vol: 35, Issue: 4, pp.485-490. (1972).
- [46] Liu, S., Fang, Z., Yang, Y., Forrest, J. "General grey numbers and their operations". *Grey Systems: Theory and Application*, Vol: 2, Issue: 3, pp: 341–349. (2012).
- [47] Liu, S., Forrest, J., Yang, Y. "Advances in grey systems research". *Journal of Grey System*, Vol: 25, Iss: 2, pp.1-18. (2013).
- [48] Liu, S., Lin, Y. "Grey information: theory and practical applications". Springer Science & Business Media. (2006).
- [49] Saracoglu, B. O. "Global Grid Prediction Systems", DOI: 10.13140/RG.2.1.3575.3040 https://www.researchgate.net/publication/289813050_Global_Grid_Prediction_Systems accessed on 22/01/2016.