RAIN – Risk Analysis of Infrastructure Networks in Response to Extreme Weather

Project Reference: 608166

FP7-SEC-2013-1 Impact of extreme weather on critical infrastructure

Project Duration: 1 May 2014 – 30 April 2017

Security Sensitivity Committee Deliverable Evaluation

Deliverable Reference	**D 5.3**
Deliverable Name	**Web Based Tool for network usage forecasting**
Contributing Partners	**AIA/IPTO**
Date of Submission	**January 2017**

The evaluation is:

- **The content is not related to general project management**
- **The content is not related to general outcomes as dissemination and communication**
- **The content is related to critical infrastructure vulnerability or sensitivity**
- **The content is not publicly available or commonly known**
- **The content does not add new information that might be misused by possible criminal offenders to exploit vulnerabilities.**
- **The content could not cause any harm to the essential interests of EU or one or more Member States.**
- **The content could not cause societal anxiety or social unrest.**
- **There are not any uncertainties that might need to contact the National Security Authority**
- **The content does not potentially harm scientific or commercial interests of consortium partners**

Diagram path 1-2-3-4-6-7-8-9-10. Therefore the evaluation is Public.

Decision of Evaluation	**Public**	**~~Confidential~~**
	~~Restricted~~	

Evaluator Name	**P.L. Prak, MSSM**
Evaluator Signature	**Signed by the chairman of the SSC**
Date of Evaluation	**2017-02-08**

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608166

This project is funded by the European Union

D5.3

Web based tool for network usage forecasting

Authors

*** Milenko Halat, Xavier Clotet, Vicens Gaitán (Grupo AIA),**

****Dimitrios Bechrakis (IPTO)**

*** Corresponding author: Aplicaciones en Informática Avanzada SL, Av. De la Torre Blanca 57, Edificio ESADECREAPOLIS, 08172 Sant Cugat del Vallès, Spain. E-mail: halatm@aia.es. Phone: +34 935 044 900.**

**** Corresponding author: IPTO, Dyrachiou 89 & Kifisou 10443, Athens, Greece. Email:dmpexrak@admie.gr. Phone: +30 210 5192669.**

Date: 02/01/2017

Dissemination level: (PU, PP, RE, CO): CO

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608166

This project is funded by the European Union

DOCUMENT HISTORY

Index	Date	Author(s)	Main modifications
0.0	October 30th 2016	Milenko Halat, Xavier Clotet, Dimitrios Bechrakis	First version
1.0	January 2st 2017	Milenko Halat, Xavier Clotet	First Draft
1.1	January 24th 2017	Milenko Halat, Xavier Clotet	Final version , updated according to internal review (TUDelft and Hellenberg International)

Document Name: Web based tool for network usage forecasting

Work Package: 5

Task: 5.4

Deliverable: D5.3

Deliverable scheduled date: October 2016

Responsible Partner: AIA

Table of Contents

1. Executive summary

This document corresponds to the description of deliverable **D5.3** *Web based tool for network usage forecasting* which is a actually a software module. This report references to the documents were the design of forecasting is made:

- **Deliverable D4.4 v2** *Forecast modules definition*, and
- **Milestone MS8** *Modelling of Forecasting Modules*.

This software tool generates the forecasting of the network usage (i.e. stress levels of the power grid, specifically in each transmission line segment) given some geographically distributed consumption and generation profiles. These profiles are obtained taking into account the hour of the day and the weather conditions, among others.

The developed tool provides a visual interface for the tuning the configuration parameters of the simulation for a number of possible scenarios. It also displays the results in different formats obtained and provides summary reports.

The document overviews the forecasting steps and explains how the modules are implemented, presenting some screenshots of the actual tool.

The software is made of several modules that interact sequentially, namely:

- **Meteorological event impact modules**, relating extreme weather events critical elements and their failure modes,
- Generation and consumption **forecast modules**, related to the context,
- Electrical **simulation modules**,
- Impact **evaluation** modules.

The emphasis is put in the **network usage**, in which power consumption forecasting is only one input. Located sources and sinks of power determines (through physical equations) the power flows through lines and components. The level of stress of each component determines then its failure probability.

The next steps, that is, the information on how to convert network usage and usage forecasting to probabilities of failure of the components, and to the consequences of failure of the components are part of the next deliverable D5.4,

2. Main modules

2.1 Meteorological event impact modules

The impact of extreme weather events is mainly assessed on the electrical grid since, as explained in Deliverable 4.3, direct damage due to extreme weather events on the telecommunications network are usually rarer than in the electric network.

A different module is defined per each extreme weather event (EWE). They provide the **failure probabilities of the critical components** of the network as a consequence of such event. They may provide also secondary effects on elements of the network like changes in their performance. Failure probability functions (known as *fragility curves*) of each module are designed according to the relevant bibliography on the subject and / or with help of expert knowledge.

Different modules have in common the following steps in it characterization;

1. **identification** of the properties of the event and the most vulnerable elements of the network studied is required to analyse a specific weather event. Of special relevance are the parameters that characterize the weather event (including thresholds), main channels of failure, and the factors that modify their probability of failure.
2. **quantification**: to provide equations, tables, or recommendations for the estimation of the failure probabilities given the values of relevant variables identified.

These modules have been validated in relation to the available use cases. For the rest, they contain free parameters that have to be calibrated using real data in each context / scenario analysed.

The implemented (so far) meteorological modules are

1. **Windstorms**: damage to transmission lines either directly (due to wind forces) or indirectly, disconnection of the turbines when the wind speed exceeds a safety threshold.

2. **Heat waves**: increase of consumption levels and power factor due to massive use of air conditioning systems. Loss of efficiency on transmission lines and generation (nuclear and photovoltaic)

3. **Heavy precipitation:** river flooding flash flooding cause disconnection of flooded substations and pylons due to landslides.

More detailed explanations can be found in deliverable **D4.4 v2**.

2.2 Generation and consumption forecast modules

2.2.1 Why to forecast?

In an interconnected electrical grid , the power generation and consumption have **to match** at each instant. While in conventional power plants generation can be set at the desired value, the production of **renewable sources** depends on a number of conditions, the most relevant of which is weather. On the other hand, even when the consumption depend individual decisions (i.e. to put or not the washing machine), the aggregated power demand exhibits certain patterns and periodicity, which allows to apply forecasting techniques. Therefore, to match the production and consumption of electricity, **forecastings** for the power demand *and* renewable energy generators are required by the electrical system operators to plan the usage of the different power plants or sources available to them (*optimal daily scheduling of power generation*).

On renewable generation, as for consumption, patterns are usually observed. Nevertheless, the techniques suitable for forecasting are quite different. The reason is that the level of activity is not related to aggregated human behaviour, but directly depending on the local weather conditions (wind speed, solar radiation, etc.)

In the context of RAIN assessment tool, the forecasting of demand and renewable sources, combined with other data, allows to obtain an estimation of the electrical state of the grid. This is highly relevant for the **impact estimation** of any kind of contingency, since small perturbations can destabilize a stressed grid or have no effect in the same grid depending on other circumstances.

2.2.2 Electrical generation

Renewable electricity production is conditioned by two groups of variables:

- Variables related to the **technology** used for production (solar, wind, hydro…).
- Variables describing the **context** (irradiation, availability of water, wind levels...).

Each technology has its own particular properties: e.g. solar power only provides energy during daytime and wind power farms can work all day but they both depend on the weather conditions of the moment; hydropower plants, can work all day long and in some cases they can even work in *reverse mode* to store energy by pumping water up back to the dam. These characteristics make them more (or less) appropriate at different moments of the day depending on the demand and weather forecasts. The key element for electrical generation forecast is having a good context description, i.e. a good weather forecast. Electric system operators have a number of proprietary products available specialized in weather forecasting for renewable power plants. These products are mainly based on self-learning algorithms like neural networks or genetic algorithms. Researchers propose hybrid models combining machine learning algorithms with ARIMA-like and regression models.

Changes in production capacity due to weather conditions may lead to complete halt of production in **wind farms** when security wind speeds are reached, or the loss of stored energy in dams when heavy rains make operators open floodgates.

Since the ultimate generation profile depends strongly on operational decisions (which generators will be active, up to which capacity, etc.) it has no practical use to create generic modules to forecast the generation profile. Moreover, since the electric companies constantly compute generation forecasts by themselves or through consultancy services, here the generation profile is considered an input except for the operational changes mentioned above (i.e. wind speed security threshold).

2.2.3 Electrical consumption

Forecasting the power consumption (or *demand,* a term more commonly used in the engineering environment) is of capital importance for electric system operators. Analogously to the forecasting of power generation, power demand depends on two groups of variables:

- Variables describing the **type of client** (residential, services, industrial, or other groups).
- Variables describing the **context**: temperature, hour, season, business calendar, special events, etc.

Regarding the type of consumers, one of the most important variable is their **mean consumption values**. Except for new clients (in which case is estimated using other variables) these data is available to electricity providers. For instance, the open GIS database of Openstreetmap (http://www.openstreetmap.org/) allows to obtain (through scripting) the type of buildings (residential, schools, hospital, company, etc.) and thus relate electrical infrastructures to specific types of client. On the other hand, there are public databases like the European Eurostat that periodically reports on the economical indicators, that include the aggregate value of power consumption by sector. These average values integrated over all the consumers on the grid provide the first approximation for the total electrical demand. Refinements are obtained by focusing on the context variables: the season, day of the week, hour of the day, and other variables (for instance, for domestic consumption the temperature is very relevant, since an important part of the consumption comes from heating, ventilating, and air conditioning systems).

The most relevant context variables are hour of the day, day of the week (mainly working days versus holidays or weekends), and season. An example of the effect of the hour and the season is shown D4.4. For this reason, these context variables are part of the configurable parameters for simulation on the User Interface. All this information is fed into **regression** or **machine learning algorithms** to obtain more accurate demand forecasts, as opposed to use *forecasts* based on averaged historical data only. Among these algorithms, the possibilities explored include

- generalized linear models,
- gradient boosting algorithms,

with Tikhonov or Lasso regularizations considered.

An example of a generalized linear model (a Machine Learning algorithm) applied to real consumption data is provided in next figure. An industrial average hourly consumption per day profile is plotted as a function of the day for almost an year period together with its one-day forecasted value. This model (and these kind of models in general) allows capturing the unusual behaviours[1] like it could be the persistent pronounced lower demand occurred during January and part of February in the data shown.

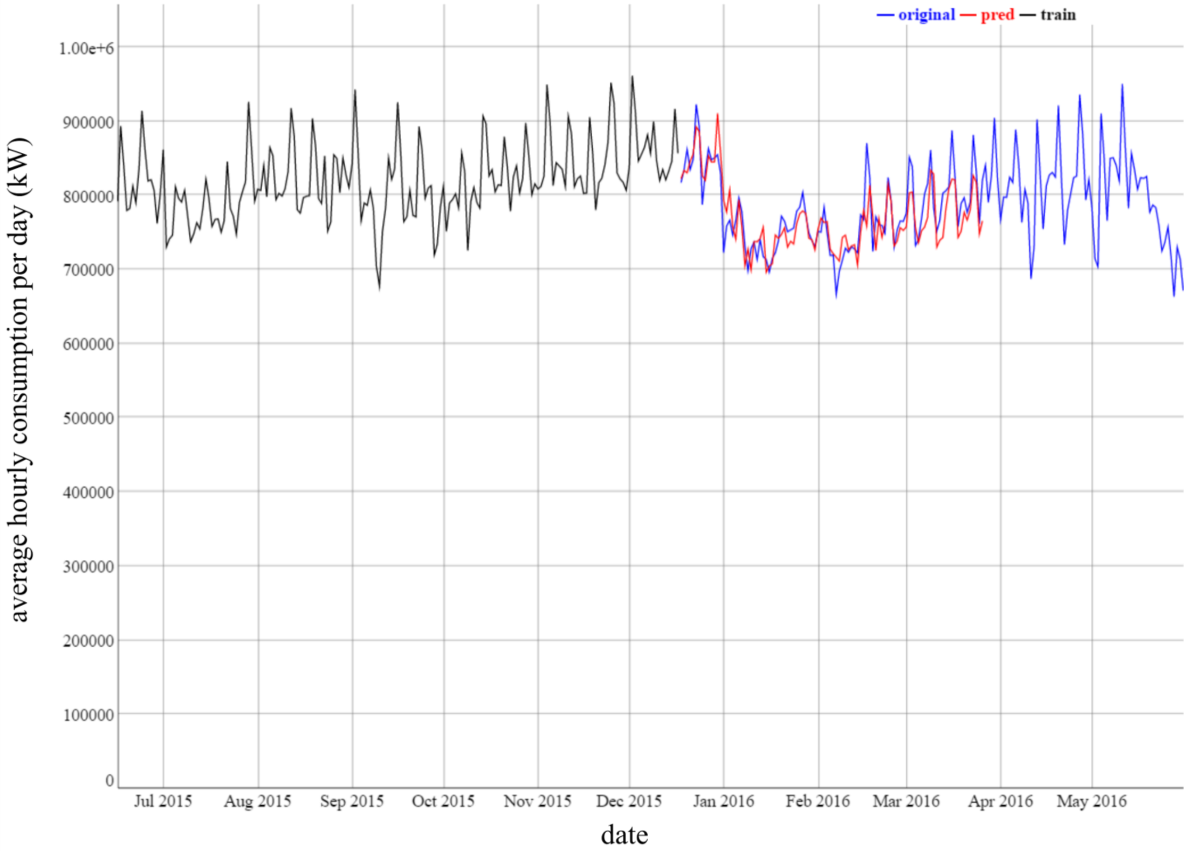

Figure 2-1. Electrical industrial load profile of Greece: average hourly consumption per day. Black: subset of the data used as training set for the generalized linear model (with Lasso regularization) applied. Red: predicted consumptions using the model and data up to the previous day forecasted. Blue: corresponding actual consumption values. Data provided by IPTO.

Another approach for demand forecasting is based on capturing the status of the system: the so call **symbolic forecasting**[2]. In this method, the analysis of historical data and available context information are used to define *symbols* (variables) that best describe the system. Symbols may include direct context variables, such as day of the week, or a (complex) combination of a number of them.

Each one of these symbols contributes to the prediction. Given a context at time t_A, the forecasted value (x_{t1}) is obtained by the summation of the value of each symbol times its weight:

$$x_{tA} = p_{symb1}(t_A)s_{symb1}(x_{t<tA}) + p_{symb2}(t_A)s_{symb2}(x_{t<tA}) + \dots + p_{symbN}(t_A)s_{symbN}(x_{t<tA})$$

[1] Unusual compared to historical data available.
[2] Internal terminology

Where $p_{symb1}(t_A)$ is the weight of symbol 1 at time t_A that could be binary (0/1) in the case of days of the week (for symbol *Tuesday* the weight is 1 if t_A is Tuesday or 0 otherwise), or it can be continuous between 0 and 1 (for instance temperature exceedance with respect a know value). $s_{symb1}(x_{t<tA})$ is the value of the symbol at time t_A, that depends on the historical data $t < t_A$. N different symbols are used in this example. A base contribution is usually defined from which the contributions by the different symbols are added or subtracted.

Then, the method implements the Holt-Winters exponential smoothing to obtain the values of the symbols $s_{symbn}(x_{t<tA})$: the algorithm learns from the errors made in the past with a given learning rate. Different learning rates can be defined per symbol, so that the contribution of the symbol can better follow the periodicity/seasonality corresponding to that variable.

An example of symbolic forecasting is shown in for industrial gas consumption (rescaled and properly anonymized). Top panel shows the historical data together with its predicted value, and the contribution of the base symbol to the forecast. Bottom panel shows the contribution of each of the other symbols (either positive or negative) with respect the base symbol. It is clearly shown the meaning and influence of each symbol:

- **Working days** contribute positively, increasing gas consumption,
- **Working days after or before weekend** or holiday contribute positively but a smaller amount,
- **Non-working days**, but non holidays nor Sunday, also contribute positively but a fraction of previous day.
- Finally, **Sundays and holidays** contribute strongly negatively to the forecast.

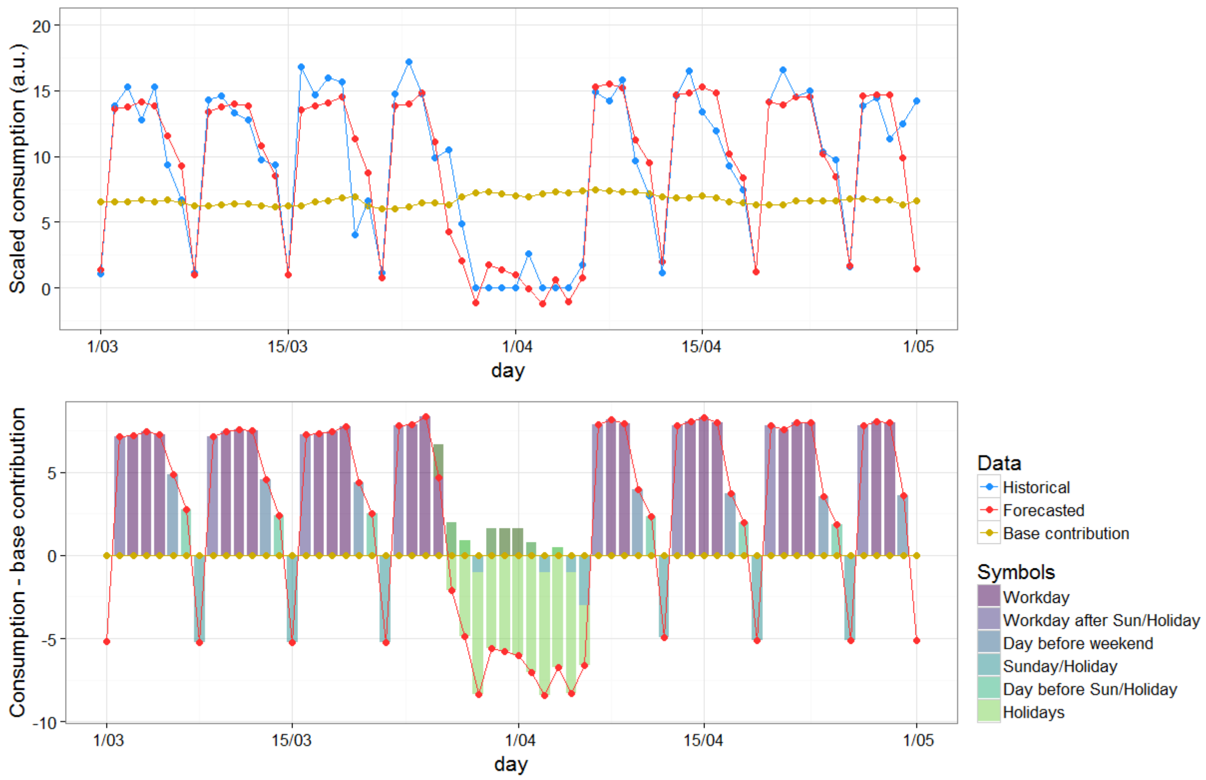

Figure 2-2 Top: scaled and properly anonymized industrial gas consumption as function of time in arbitrary units. Forecasted values are compared to historical (actual) data. The base contribution of the model is also shown. Bottom:

Additions or subtractions of each symbol to the base contribution are shown in different colors. In both plot, weekly periodicity is clearly observed. Also, the effect of Eastern holidays is explicit.

These algorithms can provide good consumption forecasts, however, they require historical data and expert knowledge to be applied, as well as a regular update of the models. Therefore, their implementation in a final version of the tool would need strong synergies among the interested end-users.

3. Implementation

This software tool is developed using R[3] for the main part of the logic and Shiny[4] for the web interface. The core of electrical computation is HELM[5] and a scripting layer written in JavaScript is used to perform the specific analysis for meteorological contingencies (Monte Carlo simulations, electrical violations, etc.).

- **Visualization**. It provides a view of the geographical situation of the networks and its elements. It also shows the intensity of the weather event in the area in the same month, the previous year.
- **Configuration**. It allows to select the extreme event, its features (such as intensity and effective area), and configure the status of the grid (for instance the right-of-way of each line). In the case of electrical grid, it is possible to select the usage level through typical consumption patterns (summer/winter and hour of the day).
- **Simulation**. This module performs the electrical simulation, using the provided inputs, and generates impact reports.
- **Reporting.** The results of electrical simulations are translated into several social and economical markers. Two kind of reports are automatically generated:
    - Electrical & Telco
    - Economic & Social

  The reports are interactive, based on html technology, and all data is exportable to csv (comma separated files), pdf or Excel files for custom analysis.

[3] https://www.r-project.org/
[4] shiny.rstudio.com/
[5] www.gridquant.com/solutions/helm-flow/

4. Summary of current functionalities

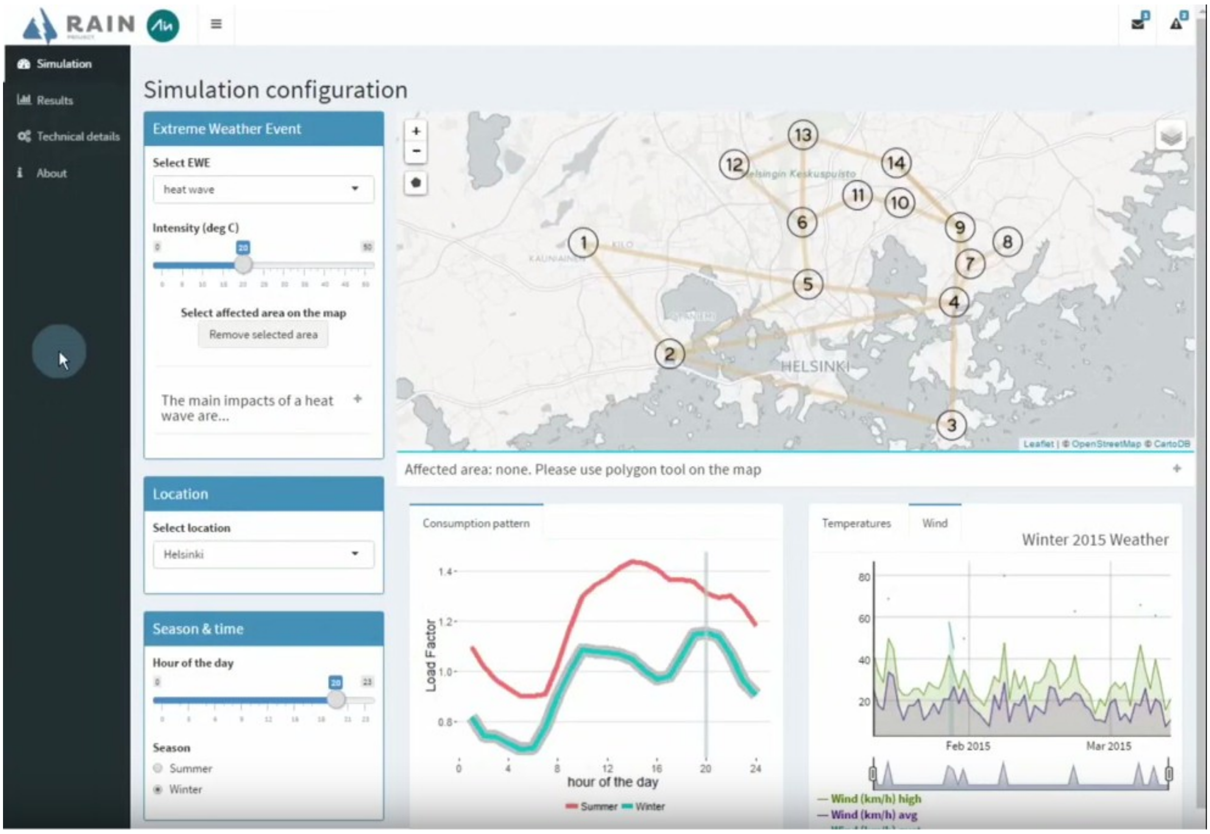

Figure 4-1. Screenshot of the main screen of the tool. See the demo video at https://youtu.be/mXBftjvyXQs

This software tool has been developed using R[6] for the main part of the logic and Shiny[7] for the web interface. The core of electrical computation is HELM[8], and on top of it, scripting in Javascript to perform the specific analysis for meteorological contingencies (Monte Carlo simulations, electrical violations, etc.).

- **Visualization**. It provides a view of geographical situation of the networks and its elements. It also shows the intensity of weather event in the area in the same month, the previous year.

- **Configuration**. It allows to select the extreme event, its features (as intensity and effective area) configure the status of the grid (for instance the right-of-way of each line). In the case of electrical grid, it is possible to select the usage level through typical consumption patterns (summer/winter and hour of the day).

- **Simulation**. This module performs the electrical simulation, using the provided inputs, and generate impact reports.

---

[6] https://www.r-project.org/
[7] shiny.rstudio.com/
[8] www.gridquant.com/solutions/helm-flow/

- **Reports.** The results of electrical simulations are translated in several social and economical markers. Two kind of reports are automatically generated.
    - Electrical & Telco
    - Economic & Social

  The reports are interactive, based on html technology, and all data is exportable to csv (comma separated files), pdf or Excel files for custom analysis

Visualization & Configuration

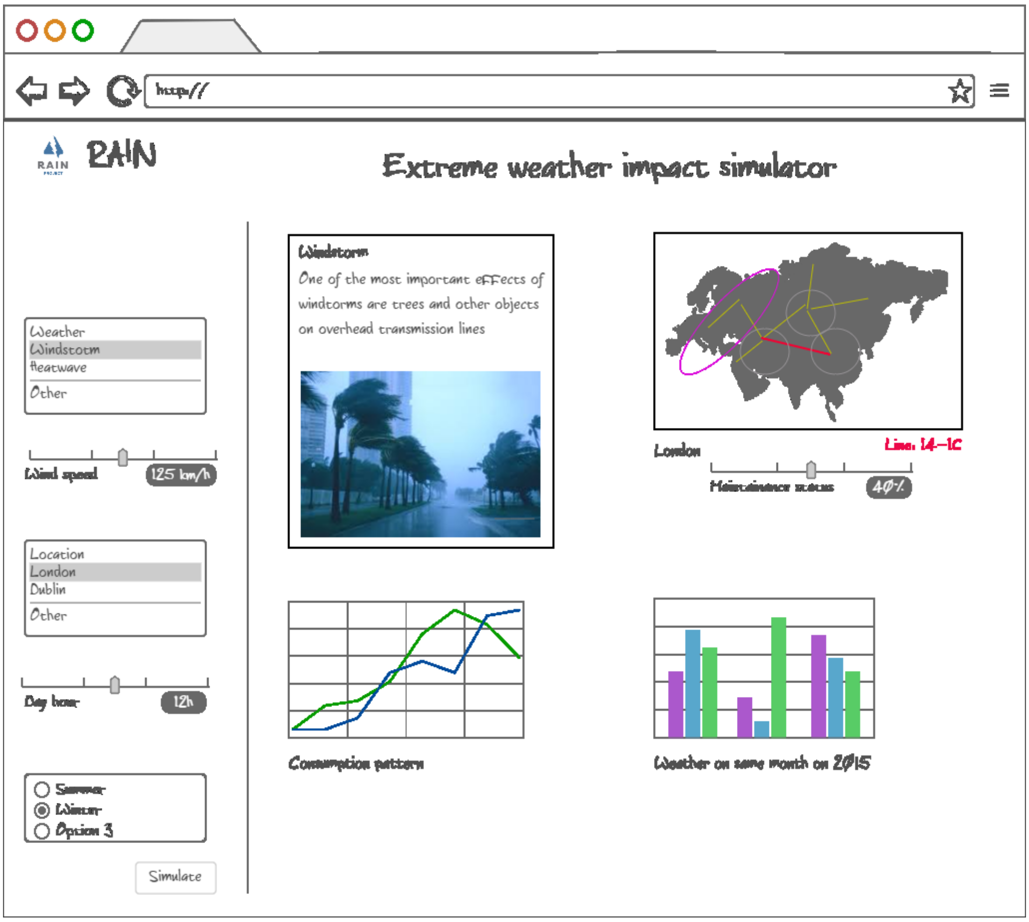

Figure 4-2.Wire frame design for the main screen: visualization and configuration.

The user interface is modular: different blocks provide different functionalities.
The configuration is made mainly on the left panel. It allows to select:

- **Type of extreme weather event**. So far windstorms and heat waves have been modelled.

- **Geographical area (city).** In the prototype it is necessary to select among the pre-configured cities. The reason is that electrical and telecommunication infrastructures have been prepared specifically for each city.

Season and hour of the day. The impact on the power grid strongly depends on its stress level, which depend on consumption patterns. The consumption patterns can be described roughly by the season, type of day and hour of the day. The user can select Winter/summer and the hour (0-23). One interesting type of analysis that the tool enables is to compare the effect of same event on the same infrastructure, but at different hours (for instance noon and midnight).

- **Brief explanation**. A panel reminds / summarizes the main effects of the selected event on the power and telco networks.

- **Weather reference**. This module connects with an online weather historical data provider and shows, in an interactive plot, the intensities of different weather events in the area on the same season the previous year.

- **Map**. The interactive map allows to visualize and explore the power grid topology. The substations are numbers and the power lines are labeled according to the substations they connect. For some type of events, the affected area can be set directly plotting a polygon over the map. The (maintenance) status of each line and pylon is set in a table right below the map.

4.1 Simulation

Once all parameters are set, the simulation can be started. The interface executes the prepared scripts to perform many electrical simulation calling the load flow engine.

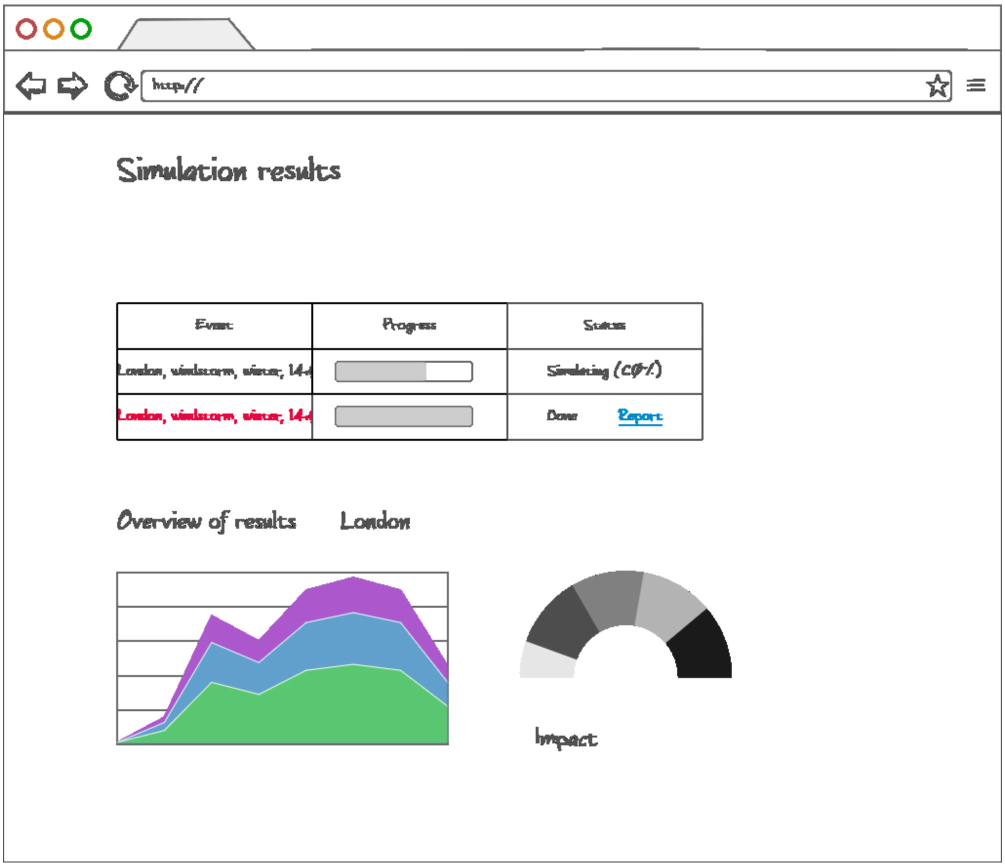

Figure 4-3. Wire frame design for the simulation screen.

In this section, a list of simulation is shown. Each simulation is described by:

- Unique id
- Geographical area
- Season and time
- Type of event
- Intensity
- Simulation status (finished or % performed)

When a simulation is selected, a brief summary of the consequences is displayed. It includes the evaluation (for instance the amount of power/energy not supplied) of the worst scenario, the most probable scenario and the mean values.

The simulation process can be summarized in the following way:

1. A triggering EWE is considered.
2. The various scenarios with failed elements according to the probabilities computed given a specific context (EWE, location, state of the system) are generated.
3. One scenario (with its failed critical elements) is selected.
4. Electrical simulation of the network without the affected elements is run.
5. Electrical protection/preventive measures are taken.

6. If some new elements in the network have to be modified / disconnected, their status is modified and points 4-6 are repeated). If no new elements are affected, electrical results (powers, voltages, status of the elements) are saved and points 3 to 6 are repeated as many times as different scenarios are left.

Once all simulations are finished and results stored, electrical consequences (probability of lost load, effects on the buses and generators, overload or disconnection of lines, etc.) are assessed.

5. Conclusions

The level of stress of the network, in the particular case of the electrical infrastructure, determines the order of magnitude of indirect and direct damage. The indirect damage is easier to understand: a midnight blackout will cause less damage to the economy (and people's activities) than one happening in the evening of a working day. Part of the *direct damage* to power grid's components depend on the electrical conditions. For instance, a disconnection of a transmission line can cause an overvoltage that can damage a transformer, if the grid is stressed (high power flows), that could not happen if the status of the grid in on low demand.

Therefore, assessing the impact of electrical contingencies it is necessary to estimate or compute it's the electrical status of the grid, which includes the generations levels and the power flows over each line. Since the power grid behavior is not trivial, simulation using the actual physical laws and forecasting (generation and demand) values are necessary to determine which zones and components are the most vulnerable ones. This is the role of the modules presented in this document. Specific conclusions are expected when the software will be integrated and tested over use cases.