ePolicy

Engineering the Policy-making Life Cycle

Policy reasoning prototype evaluation V1

Document type:	Deliverable
Dissemination Level:	Public
Editor:	Michela Milano
Document Version:	0.1
Contributing Partners:	UNIBO, UNIFE
Contributing WPs:	WP3
Estimated P/M (if applicable):	5
Date of Completion:	30 September 2013
Date of Delivery to EC	30 September 2013
Number of pages:	13

ABSTRACT

The aim of this document is to evaluate the first version of the global optimization service producing a regional energy plan, described in the deliverable 3.2. The evaluation is performed by considering both functional and non functional requirements.

The project is co-funded by the European Community under the Information and Communication Technologies (ICT) theme of the Seventh Framework Programme (FP7/2007-2013). Grant Agreement n°288147.

Authors of this document:

Michela Milano[1]

[1]: DISI, University of Bologna
email: michela.milano@unibo.it

Contents

1 Introduction to Global Optimization Service

In the ePolicy project, one of the components of the Decision Support System is the global optimizer [4, 2], that encapsulates the regional perspective for the definition of the Regional Energy plan. There are mainly two users of the service: the **policy maker** and the **environmental (domain) expert**, each accessing the service with different access roles and different functionalities. The policy maker can access the system by providing three inputs:

1. the total energy produced by renewable energy sources expected in the plan;
2. the minimum and maximum amount of energy per energy source;
3. a number of objective functions (taken from a list of environmental receptors and other factors such as cost, energy produced).

Note that the minimum amount of energy per energy source can be set for obtaining energy source diversification while the maximum amount of energy should respect a constraint on the regional receptivity (stated by the environmental expert). In output, the policy maker obtains a number of different scenarios, i.e., a number of regional plans that satisfy the given constraints and that are optimal (in the Pareto sense) with respect to the given objective functions.

The environmental expert instead has all the functionalities of the policy maker, but in addition he/she can configure the system by providing data on the policy domain. The environmental expert can provide:

1. matrices for performing the environmental assessment of plans;
2. the definition of primary activities for the plan, the definition of secondary activities and their relations thereof;
3. the costs related to primary and/or secondary activities;
4. maximum amount of energy per energy source that can be produced in the considered region; these numbers represent an upper limit to the maximum amount of energy that can be set by the policy maker.

The environmental expert can run the optimizer with the main goal of evaluating if the introduced numbers and matrices are indeed reasonable or should be adjusted.

In the following we report the UML diagrams representing the use case of the global optimizer utilization. The service in this case is considered as a stand-alone application, i.e., not integrated with other services.

In the requirement document, we have reported a detailed description of all these activities.

1.1 Functional Requirements

Concerning functional requirements, we have identified a number of functionalities of the system concerning its configuration and its usage.

We test the functional requirements by considering two lists of requirements: one for the functionalities required by the environmental expert and one for the functionalities

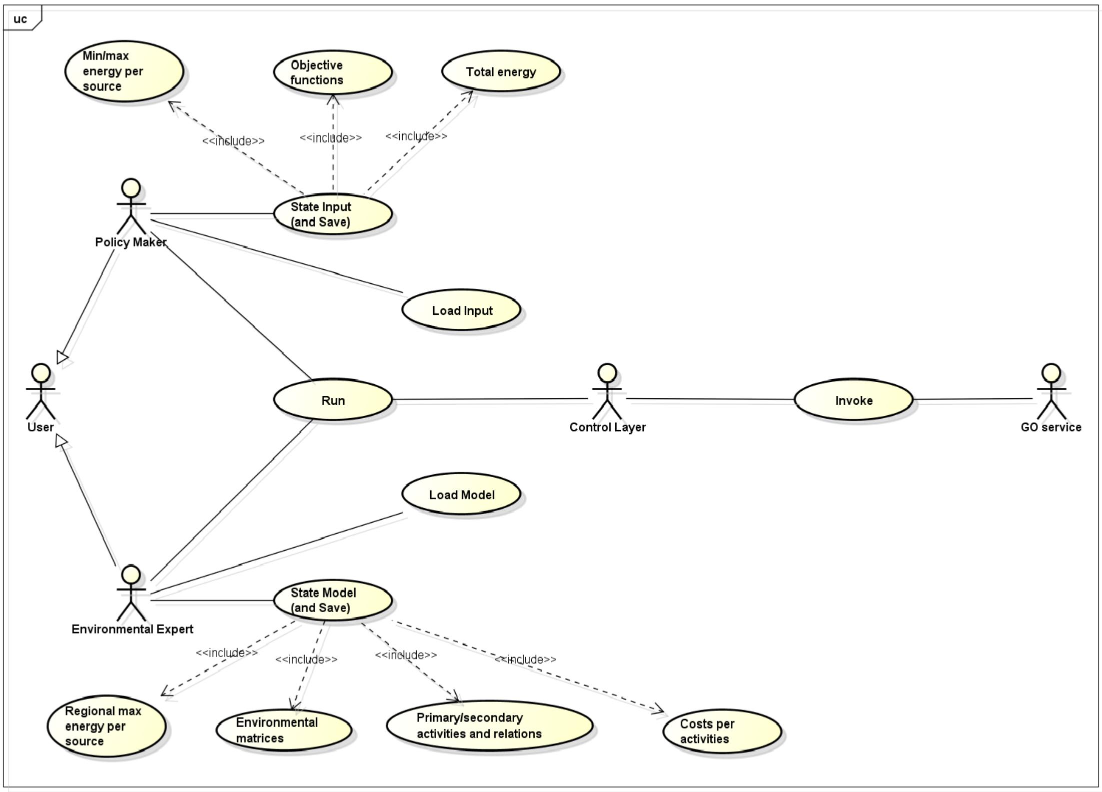

Figure 1: Use cases

required by the policy maker. By analysing the code and testing it on data from the RER Energy plan 2011-2013 plus other instances, we tick the implemented functionalities.

At the end of the test, the evaluation deliverable 2.3 [3] states that if the number of functional requirements implemented is lower or equal to the 60% of the requirements contained in the requirement document, then the component evaluation has to be considered unsatisfactory. If this number is between 60 and 80%, the evaluation has to be considered sufficient, while if the number is higher than 80% the evaluation is successful.

1.2 Non Functional Requirements

We report here the section on non functional requirements reported in the deliverable 2.3 on Means for Project Evaluations [3], which has been submitted to the commission at month 20.

> The nonfunctional requirements will be evaluated by running tests and by performing field tests involving users and asking them to evaluate the software through interviews and questionnaires. In particular, we will consider the following means for evaluation of non-functional requirements:
>
> - Scalability, performance and efficiency: we plan to run a battery of tests, both on the first version of the prototype (whose evaluation is due at month 24, D3.3) and on the final version (whose evaluation is due at month 33, D3.5) by using the data from the RER regional plan of 2011, the

new data of 2014 and also a set of synthetic instances generated by changing the matrix coefficients, increasing the number of objective functions, changing constraints, and increasing the number of activities. For these tests we will extract the running time, the number of explored nodes, the mean and the variance of the tests. We consider this test satisfactory if for running the system on a real plan, the system has a response time less than five minutes.
- Cost of development and time of development: these figures will be extracted by the project time sheets measuring the time and the cost of MM exposes on these activities. We consider the evaluation of these parameters succesful if the numbers are in line with what expected in the project proposal. In particular, if the variation of time and cost is lower than the 5% with respect to the data provided in the budget.
- Accuracy, precision and re-configurability: these figures will be measured by using questionnaires, directed to policy makers (that test the first two requirements) and to environmental experts (evaluating all of them). If more than the 70% of users are satisfied with these features (derived from questionnaires) the evaluation has to be considered successful.

For each category of users we will evaluate the component by providing questionnaires customized for policy makers and for environmental experts that are devoted not to assess the visual interface of the component and its friendliness, but rather to assess the correctness of the results. We plan to involve policy makers of the RER, which is a partner of the project, and, after a training period, to use the global optimizer for the Regional Energy Plan 2014, therefore on a real plan. This will give us the possibility to evaluate the system on real setting, provide alternatives that could be evaluated on environmental, energy and economic aspects by the policy maker. We will report the experience of this test both on the web site and in a summary document, deliverable D3.5, due at month 33.

We also plan to make a trial by contacting policy makers from other regions contacted by the RER. The idea is to guide them in defining a plan by checking the functionalities they expect from a system like that. It is worth mentioning that the system requirements have been designed by looking at several plans, but they have been mainly tailored on the requirement stated by RER experts. Therefore, this test is devoted to understand how the functionalities of system is suitable for other regions, or if new requirements arise from this analysis. Again these policy makers will be appointed and their evaluation could be captured through a questionnaire and reported in the Exploitation Plan, due at month 28.

In parallel, we plan to measure the configurability of the system by asking several experts from ARPA Emilia-Romagna, which is partner of the Advisory and Dissemination Board of the project, to configure the system. During this activity, the environmental experts will also query the system for generating regional plans in order to check if the configuration is correct. Therefore the questionnaire of the environmental experts will be a superset of the one devoted to policy makers. Again this evaluation will be reported in the deliverable D3.5 due at month 33.

2 Prototype Evaluation V1

Before starting the evaluation, we remark again what written in the *"Means for Project Evaluations"* [3]:

> ... we will evaluate the component by providing questionnaires customized for policy makers and for environmental experts that are devoted not to assess the visual interface of the component and its friendliness, but rather to assess the correctness of the results.

In order to separate the evaluation of the Global Optimizer prototype from the one of the visual interface, contained in deliverable 7.2, we have implemented a web service wrapping the solver. The web service provides an interface for entering the data and a GUI that is devoted to ease the system use for a non IT expert. We are fully aware that this GUI is indeed not as user friendly as the interface described in deliverable 7.1, but it is a way to easily evaluate the correctness of the solutions provided by the system.

The graphic front-end has been described in deliverable 2.3 [3] and can be used by accessing the following link

`http://globalopt.epolicy-project.eu/GlobalOptClient/ePolicy.jsp`

The service available at this link is for testing purposes only. Notice that it does not represent the real Graphical User Interface (GUI) of the ePolicy prototype.

2.1 Functional Requirement Evaluation

The system has been evaluated as far as functional requirements are concerned. We have considered two tables: one concerning the features available for the environmental expert (Figure 2) and one concerning the tasks of the policy maker (Figure 3).

As we can see, the system implements 100% of the functional requirements identified in the requirement document. As far as the graphic front-end, which was not required nor described in the Description of Work of ePolicy [1], some configuration tasks have not been implemented.

As far as the policy maker is concerned, the system implements 100% of the functional requirements identified in the requirement document and these functionalities are also handled by the graphic front end.

For these reasons, we consider that the functional requirement evaluation has been successful.

Activity	System	GUI
Set regional max energy per source	✔	✖
Change environmental matrices	✔	✖
Change primary/ sec activity matrix	✔	✖
Def. cost per activity	✔	✖
Min and max energy	✔	✔
State objective functions	✔	✔
Set total energy	✔	✔
Run system	✔	✔
Obtain results	✔	✔

Figure 2: Environmental expert activity implementation

Activity	System	GUI
Min and max energy	✔	✔
State objective functions	✔	✔
Set total energy	✔	✔
Run system	✔	✔
Obtain results	✔	✔

Figure 3: Policy Maker activity implementation

2.2 Non functional Requirement Evaluation

As stated in Deliverable 2.3 [3], a large part of non functional requirements involve an evaluation from experts which is currently running and will be contained in the second evaluation document due at month 33.

Performance, efficiency and scalability. As far as performance and efficiency are concerned, we have run a battery of tests by using the data from the RER regional plan of 2011; for all these tests we have extracted the running time and for all of them, the running time is less than a second.

In order to assess the scalability of the software, we performed a series of tests by randomly generating a set of data, including the co-axial matrices, the matrix relating primary and secondary activities, the activity costs, etc. In this way, we were able to stress-test the software with instances containing a number of activities, pressures and receptors larger than those in the actual data provided by ARPA.

The matrices currently used for the environmental assessment by ARPA contain 93 activities, 48 pressures and 23 receptors.

In the random instances, we used a parameter N that represents the size of the instance. Given a value N, we randomly generated N activities, N pressures and N receptors. Amongst the activities, a random number was selected as *primary*, and the rest are *secondary* activities. For each pair primary-secondary activity, we randomly decided weather that primary activity required the secondary with probability 0.5. If the primary required that secondary activity, the amount of secondary activity required for one unit of the primary was randomly generated in the interval 0-10.

Similarly, we generated the two matrices relating activities and pressures, and the one relating pressures and receptors. These matrices contain, in the real instance, qualitative values *high*, *medium* and *low* (or there can be no value, meaning that there is no relation between the given activity and pressure, or pressure and receptor). In the random instances, each cell of these matrices contains no value (meaning no relation) with probability $1/2$, and, in case the cell contains a non-null value, the values *high*, *medium* and *low* were selected with equal probability.

The experiments were performed on a laptop computer running Linux with a 8x Intel Core i7-3720QM CPU at 2.60GHz; note however that only one core was used in the experiments. The results are plotted in Figure 4, and in Figure 5 in log scale. Each point in the plot represents a test. The x-axis is the size of the instance, i.e., the N parameter mentioned earlier. The y-axis represents the running time of the instance, i.e., the time required to find the optimal solution.

We can see that the computing time, for sizes below 100, is always less than a second. Note that if $N = 100$, the matrix that relates activities and pressures has size $100 \times 100 = 10,000$, while in the real instance it is just $93 \times 48 = 4,464$; for the matrix linking pressures and receptors, the instance $N = 100$ again considers 100 pressures and 100 receptors, while in the real instance its size is $48 \times 23 = 1,104$.

We can also see that, with a size $N = 600$ we are below 5 minutes (300 seconds), and in this case the matrices are of size $600 \times 600 = 360,000$, more than 80 times the real instance.

From the log-scale plot (Figure 5), we can see that the computing time grows less than exponentially in the size N.

Since the system has a response time less than five minutes, according to the requirement given in [3], the test is considered satisfactory.

This test will be more significant on the second version of the prototype, where we will compute the Pareto Optimality curve. In this case, as the system should find a number of solutions, and these solutions should be non dominated by others, the system will take more time and it will be important to evaluate the performances of the system for an increasing number of objective functions and an increasing number of (uniformly distributed) points on the Pareto Curve.

Cost and time of development. As far as the cost of development and time of development are concerned, we are perfectly in line with what is stated in the DoW [1]. In fact we are at almost the 70% of the WP duration (as WP3 ends at month 33) and we have used around the 65% of the budget allocated.

Accuracy, precision and re-configurability. As far as accuracy, precision and re-configurability are concerned, we have already tested the system with two categories of users: a policy maker of the Emilia-Romagna region, providing positive feedback on the sys-

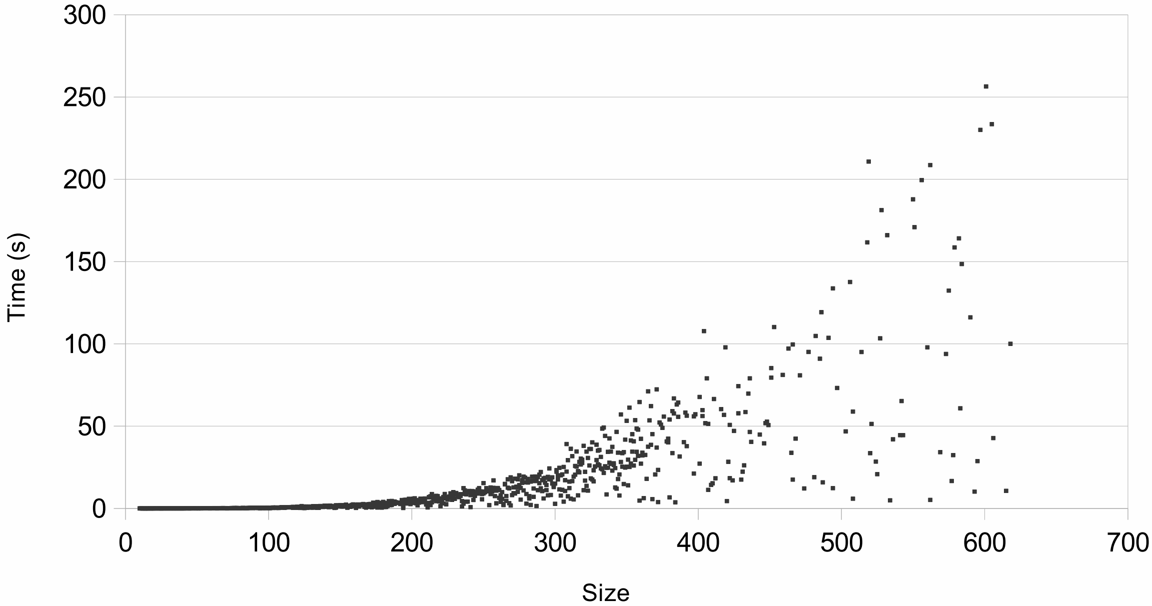

Figure 4: Stress test on randomly generated problems. The y axis plots the running time in seconds, while the x-axis is the size N of the problem, considering the same number N of activities, pressures and receptors.

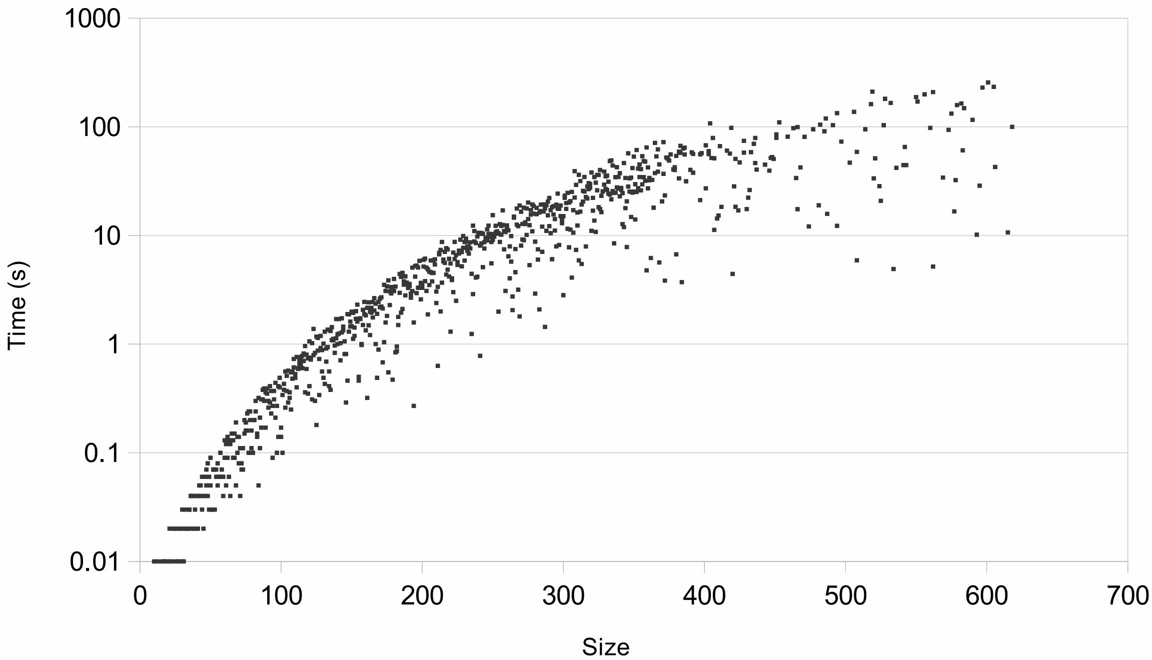

Figure 5: Stress test on randomly generated problems, log scale. The y axis plots the running time in seconds, while the x-axis is the size N of the problem, considering the same number N of activities, pressures and receptors.

tem accuracy and precision; also we have involved a member of our ABD, namely Paolo Cagnoli from ARPA. He has tested the system configuration and use. As far as configuration is concerned, the re-configurability of the system could be improved by exposing this feature in the graphic front. However, working with an IT expert, it was possible to change the matrices of the environmental receptors and pressures, the energy constraints and the primary/secondary activity matrix. Indeed, by using the system the environmental expert has noted some wrong parameters in the matrices and has corrected them. As far as the system usability accuracy and precision are concerned, the ARPA expert has provided a very positive feedback on the system.

We are aware that we need some more extensive test and for this reason we are involving at the moment the Region as well as other policy makers and environmental experts. More extensive test results will be presented in the Evaluation document (deliverable 3.5) at month 33.

2.3 Future tests due at month 33

We are planning an evaluation of the developed tool which is due at month 33. For each category of users we will evaluate the component by providing questionnaires customized for policy makers and for environmental experts that are devoted not to assess the visual interface of the component and its friendliness, but rather to assess the correctness of the results. We plan to involve policy makers of the RER, which is a partner of the project, and, after a training period, to use the global optimizer for the Regional Energy Plan 2014, therefore on a real plan. This will give us the possibility to evaluate the system on a real setting, provide alternatives that could be evaluated on environmental, energy and economic aspects by the policy maker. We will report the experience of this test both on the web site and in a summary document, deliverable D3.5, due at month 33.

We also plan to make a trial by contacting policy makers from other regions contacted by the RER. The idea is to guide them in defining a plan by checking the functionalities they expect from a system like that. It is worth mentioning that the system requirements have been designed by looking at several plans, but they have been mainly tailored on the requirements stated by RER experts. Therefore, this test is devoted to understand how the functionalities of the system is suitable for other regions, or if new requirements arise from this analysis. Again these policy makers will be appointed and their evaluation could be captured through a questionnaire and reported in the Exploitation Plan, due at month 28.

In parallel, we plan to measure the configurability of the system by asking several experts from ARPA Emilia-Romagna, which is partner of the Advisory and Dissemination Board of the project, to configure the system. During this activity, the environmental experts will also query the system for generating regional plans in order to check if the configuration is correct. Therefore the questionnaire of the environmental experts will be a superset of the one devoted to policy makers. Again this evaluation will be reported in the deliverable D3.5 due at month 33.

References

[1] Description of work. ePolicy project, June 2011.

[2] Marco Gavanelli, Michela Milano, Federico Chesani, and Elisa Marengo. Prototype of the global level policy reasoning system V1. In Marco Gavanelli, editor, *Deliverable of the ePolicy project*, number 3.2, pages 1–31. ePolicy project, May 2013.

[3] Michela Milano, Federico Chesani, Stefano Bragaglia, Tina Balke, Nigel Gilbert, Lars Kotthoff, Barry O'Sullivan, Luís Torgo, Tobias Ruppert, Sabrina Franceschini, Daniele Sangiorgi, Tony Woods, and Marco Gavanelli. Means of project evaluation. In Michela Milano, editor, *Deliverable of the ePolicy project*, number 2.3, pages 1–19. ePolicy project, May 2013.

[4] Michela Milano, Attilio Raimondi, Marco Gavanelli, and Fabrizio Riguzzi. Design of the global regional model and its instantiation on the energy plan. In Michela Milano, editor, *Deliverable of the ePolicy project*, number 3.1, pages 1–42. ePolicy project, October 2013.