International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

YSSP Interim Report **IR-14-010**

Developing a Strategic Stochastic Optimization Model, Robust Solutions, and a Decision Support System for Energy-efficient Buildings

Emilio L. Cano (emilio.lopez@urjc.es)

Approved by

Yurii Yermoliev (ermoliev@iiasa.ac.at)
YSSP Supervisor

Elena Rovenskaya
Program Director, Advanced Systems Analysis

July 24, 2014

Acknowledgments

This work has been done during my stay at the Young Scientists Summer Program (YSSP) at the International Institute of Advanced Systems Analysis (IIASA), Laxenburg, Austria during the summer of 2013. My stay was funded by Universidad Rey Juan Carlos (URJC) and Santander Universidades under the program "Programa de estancias cortas predoctorales". Great thanks to my PhD. supervisor Javier M. Moguerza and my supervisors at the YSSP 2013 Yurii Yermoliev and Tatiana Ermolieva, as well as to the IIASA ASA program staff. I acknowledge the "Energy Efficiency and Risk Management" (EnRiMa) EU FP7 project (number 260041) in which both IIASA and URJC are partners along with seven more partners: Stockholm University (SU), University College London (UCL), Center for Energy and Innovative Technologies (CET), Minerva Consulting and Communication (MC&C), SINTEF Group, Tecnalia Research and Innovation, and Hidrocantábrico Energía (HCE). I also acknowledge the national projects OPTIMOS3 (MTM2012-36163-C06-06), RIESGOS-CM: code S2009/ESP-1685, HAUS: IPT-2011-1049-430000, EDUCALAB: IPT-2011-1071-430000, DEMOCRACY4ALL: IPT-2011-0869-430000, CORPORATE COMMUNITY: IPT-2011-0871-430000, and CONTENT & INTELIGENCE: IPT-2012-0912-430000.

About the Author

Emilio L. Cano is Researcher at Rey Juan Carlos University (Madrid, Spain). He arrived at the Department of Statistics and Operations Research after a period of 14 years working in several companies of varied sectors. He got a BSc. in Statistics (1994), a degree in Applied Statistics (2010) from Complutense University, and his MSc. in Decision Systems Engineering at Rey Juan Carlos University (2011). Currently, he is proceeding with his PhD research, which is framed in the EU FP7 project EnRiMa (Energy Efficiency and Risk Management). Besides, Emilio's research interests concern statistical quality control and Six Sigma methodologies, with a book (Six Sigma with R, Springer, 2012) and an R package contributed (SixSigma). Emilio participates in the technical subcommittee of statistical methods at the Spanish member of ISO (International Standards Organization) and is also adjunct lecturer at the University of Castilla-La Mancha.

Abstract

This research is being carried out in the context of the EnRiMa project (Energy Efficiency and Risk Management in Public Buildings), funded by the European Commission (EC) within the Seventh Framework Program. Energy Systems Optimization is increasing its importance due to regulations and de-regulations of the energy sector and the setting of targets such as the European Union's 20/20/20. This raises new types of dynamic stochastic energy models incorporating both strategic and operational decisions (short-term decisions have to be made from long-term perspectives) involving standard technological as well as market-oriented financial options. Thus, buildings managers are challenged by decision making processes to achieve robust optimal energy supply portfolio and they are encouraged to adopt an active role in energy markets. Moreover, those decisions must be made under inherently uncertain conditions. The goal of this paper is to develop an integrated framework for the representation and solution of such energy systems optimization problems, to be implemented in Decision Support Systems (DSSs) for robust decision making at the building level to face rising systemic economic and environmental global challenges. As the combination of operational and strategic decisions in the same model induces risk aversion in strategic decisions, the developed approach allows easy to include quantile-based measures such as Conditional Value at Risk (CVaR). Such complex energy systems need to be accurately described in a condensed way representing a large amount of variables, parameters and constraints reflecting endogenous and exogenous interdependencies, sustainability requirements and threats. Therefore, a comprehensive Symbolic Model Specification (SMS) development is a part of the research work. Using the R statistical software and programming language, an integrated framework is proposed to cover the needs of the whole decision making process, ranging from data analysis and estimation to effective representation of models and decisions to be used by both humans and machines. Such a framework provides an environment for enforcing the necessary stakeholders dialog. Furthermore, the framework allows communicating with different types of optimization software.

Contents

Developing a Strategic Stochastic Optimization Model, Robust Solutions, and a Decision Support System for Energy-efficient Buildings

Emilio L. Cano (emilio.lopez@urjc.es)[*]

1 Introduction

1.1 Motivation

Energy systems optimization is increasing its importance due to deregulations in energy markets and the setting of targets such as the European Union (EU) 20-20-20. In turn, those targets usually embody policies which motivate new regulations aimed at the achievement of such objectives. For example, emissions trading schemes, renewable energy and/or efficient generators subsidies, or efficiency requirements such as buildings labeling, among others. This new situation is motivated by several concerns of the post-industrial era, namely:

- Global warming;
- Economy globalization;
- Resources scarcity;
- Awareness for sustainability.

In spite of the above-mentioned globalization, usually global changes must be tackled at a regional or local scale. Thus, utilities and fuel producers, yet global, must fulfill local market requirements, e.g., enough amount of electricity for a given city. Moreover, final users of energy have their own requirements which satisfaction depends on decisions made at the *shop-floor* stage. Users' comfort, security, and energy availability are challenges for decision makers at the building level, who have to deal with limited budgets in addition to the regulations regardless their global, regional or local scope. Furthermore, new technologies and refurbishment options are available and continuously evolving, widening the range of options for decision makers.

1.2 Relevant policies

In the last decades several regulatory and market changes have altered the way energy is being used. Those changes in Europe were mainly focused on electricity markets (Jamasb and Pollitt 2005). Nevertheless, more recent regulations try to deal with energy as a whole. In the following, some of the more relevant policies are outlined. Even though they refer to Europe, similar schemes are being adopted worldwide.

[*]Universidad Rey Juan Carlos (Spain); Advanced Systems Analysis (ASA) program YSSP 2013

- The EU **climate and energy package**[1] aims to ensure the European Union meets its ambitious climate and energy targets for 2020. These targets are known as the 20-20-20 targets, namely:
  - A 20% reduction in EU greenhouse gas emissions from 1990 levels;
  - Raising the share of EU energy consumption produced from renewable resources to 20%;
  - A 20% improvement in the EU's energy efficiency.

  The targets were set in March 2007 and were enacted through the climate and energy package in 2009. Afterwards, the European Commission (EC) analyzed options to move beyond 20% greenhouse gas emissions through the Commission Communication SEC (2010) 650.

- The **Energy Efficiency Plan 2011**[2] was adopted by the EC for saving more energy through concrete measures. It included measures for a wide range of sectors, including building, transportation, or manufacturing, among others. Some of the measures included in this plan are the Energy Performance of Buildings Directive, the Labeling Directive, and the Energy End-Use Efficiency and Energy Services. More recently, the **Energy Efficiency Directive**[3] 2012/27/EU has been adopted by the EU, establishing a common framework of measures for the promotion of energy efficiency within the Union in order to reach the efficiency target in the climate and energy package.

- As for the **liberalization of energy markets**, the first liberalization directives were adopted in 1996 (electricity) and 1998 (gas), and the second ones in 2003. The third liberalization package includes new legislative proposals to strengthen competition in electricity and gas markets, based on the Commission's energy package as of 2007.

- Regarding **renewable sources**, the Directive 2009/28/EC of the European Parliament and of the Council on the promotion of the use of energy from renewable sources established a common framework for the production and promotion of energy from renewable sources. The Directive takes also into account energy from biofuels and bioliquids. Some systems-related topics stemmed from these new regulations: Net-Zero Energy Building (NZEB) strategies, which aim is to achieve buildings with zero net energy consumption and zero carbon emissions annually. Some authors go beyond this concept from an economical ecologics perspective and introduce new concepts (Hernandez and Kenny 2010). A classification and description of can be found in Pless and Torcellini (2010); Net metering is a policy for consumers who own renewable energy facilities which allows them to use the energy when it is needed through a sort of balance with the market. In contrast to net metering, Feed-In-Tariffs' policies foster the direct sale of energy to the grid. It seems that US favors net-metering while Europe and Japan feed-in-tariffs (Hardesty 2013).

1.3 The EnRiMa project

The framework proposed in this work has been applied to the EnRiMa project[4]. EnRiMa (Energy Efficiency and Risk Management in Public Buildings) is a 7^{th} Framework Program

[1] http://ec.europa.eu/clima/policies/package/index_en.htm
[2] http://ec.europa.eu/energy/efficiency/action_plan/action_plan_en.htm
[3] http://ec.europa.eu/energy/efficiency/eed/eed_en.htm
[4] http://www.enrima-project.eu

(FP7) research project funded by the EC, which overall objective is to develop a Decision Support System (DSS) for operators of energy-efficient buildings and spaces of public use. The consortium is formed by nine partners from six European countries:

- Stockholms Universitet (SU), Sweden;
- University College London (UCL), United Kingdom;
- International Institute for Advanced Systems Analysis (IIASA), Austria;
- Universidad Rey Juan Carlos (URJC), Spain;
- Center for Energy and innovative Technologies (CET), Austria;
- Minerva Consulting and Communication (MCC), Belgium;
- Stiftelsen for Industriell og Teknisk Forskning (SINTEF), Norway;
- Tecnalia Research & Innovation (TECNALIA), Spain;
- Hidrocantábrico Energía (HCE), Spain.

The project started in October 2010, with a duration of 42 months. At the time this is written all the planned milestones have been achieved and the project advances have been disseminated at both technical and non-technical levels, see for example Groissböck et al. (2013). The EnRiMa DSS would help managers of public buildings to find operational policies for controlling energy resources, such as energy purchases as well as small-scale, on-site Distributed Generation (DG) with Combined Heat and Power (CHP) applications for using recovered heat, and loads, which may be available for curtailment or shifting via storage technologies. The installation of renewable energy technologies based on biomass, biogas, and solar power is also considered whenever applicable. A key innovation of the project is to combine the proven methodology for modeling energy flows in buildings with recent advances in effective coping with uncertainty. This provides perspectives to create the DSS that would aid the operators in integrated management of conflicting goals such as cost reduction, meeting energy, efficiency, and CO_2 emissions targets while considering tolerance for comfort and risks, especially due to uncertainties in energy prices and loads, e.g., by the use of financial contracts to provide protection against adverse movements in energy prices and loads. The functionality of the EnRiMa DSS is summarized in Figure 1.

The framework presented in this report is a central part of the EnRiMa DSS, through the so-called Solver Manager module, whose structure can be seen in Figure 2. The Solver Manager is part of the Engine, along with the Scenario Generator and the Kernel. A diagram of the architecture and the relationship between the Solver Manager and the rest of the modules can be seen in Figure 3.

1.4 Energy systems

Energy systems are conceived in this report as the technologies and devices used to provide people with the energy needed for their everyday activities. From this standpoint, we find different types of energy systems, namely:

- Appliances;
- Networks;
- Generation and transformation technologies;

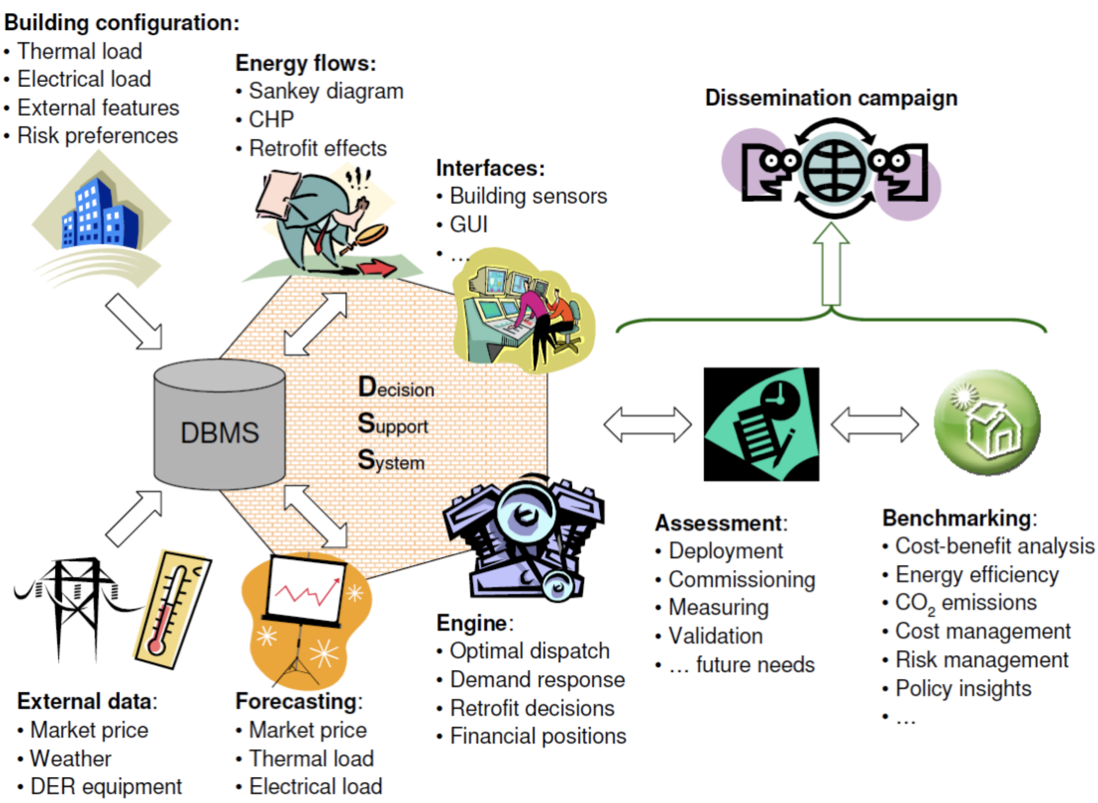

Figure 1: EnRiMa DSS functionality.

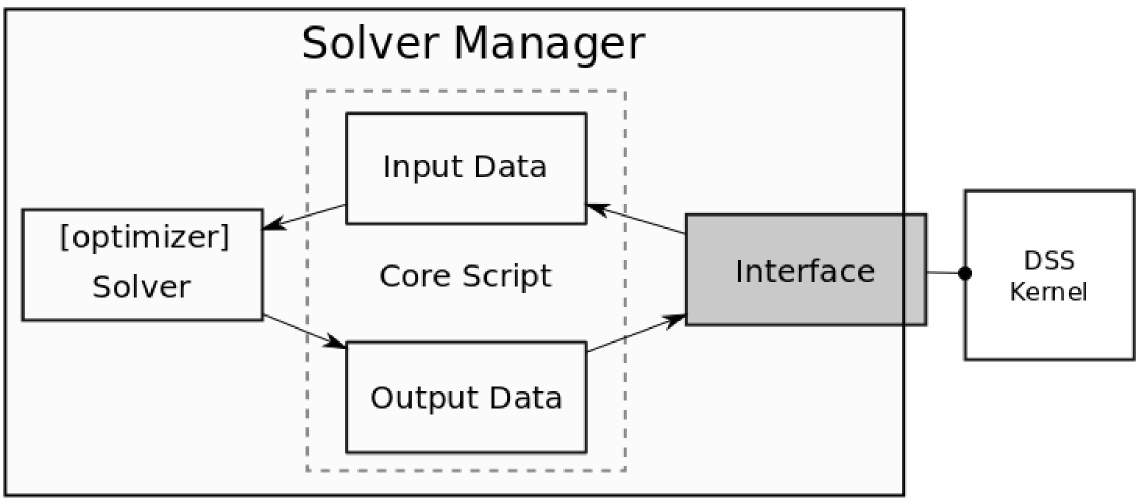

Figure 2: EnRiMa Solver Manager module

- Storage technologies;
- Passive technologies.

In what follows, the **building level** extent is assumed. Thus, the focus is on the consumer side. The meaning of building in this case can refer to different aggregation typologies, such as single buildings, set of buildings, or spaces of public or private use.

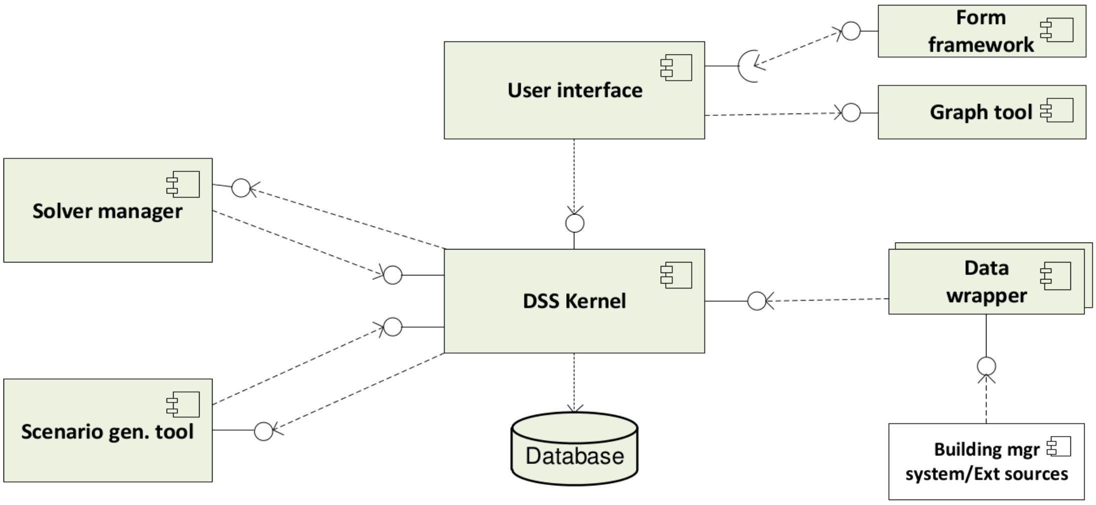

Figure 3: EnRiMa DSS architecture

Examples of buildings under this conception are university campus, sports centers, administrative buildings, hospitals, and airports. Therefore, the target buildings are those that are managed by an identified party (individual or organization) that can make decisions regarding energy systems. Within the EnRiMa project, two test sites have been used to create the models and the DSS:

- The FASAD building in Siero (Asturias, Spain).
- The Pinkafeld university campus in Pinkafeld (Burgenland, Austria).

Buildings' energy flows can be represented by Sankey diagrams, providing a straightforward way of visualizing the building energy systems' dynamics. Figure 4 shows an actual Sankey diagram for the Pinkafeld campus test site. The energy flows (arrows) from the supply side (left) to the demand side (right) throughout technologies (boxes). On the supply side we may have markets, such as the electricity grid, and renewable sources, such as solar irradiation. Different types of energy are transformed into others to meet the users demand.

Two types of decisions can be made regarding energy systems. On the one hand, there are decisions on which systems are available. These are **strategic decisions**. On the other hand, there are decisions on how to use the available systems. These are **operational decisions**. Strategic decisions are made in the long term (e.g., years) whereas operational decisions are made in the short term (e.g., hours). Examples of strategic decisions are: type of contract to sign with the grid; number of PV panels to install; renovation of building's envelope elements. Examples of operational decisions are: how much electricity buy from the grid at a given hour; how much energy input to a generator. Note that both types of decisions are interdependent as we can only use those systems that are available, and decisions on investing on new equipment or renovation depend on how they can be used to meet the overall requirements.

The proposed framework focuses on long-term strategic decisions. However, operational decisions are included in the models in order to take into account the short-term systems performance through **dynamic strategic models**.

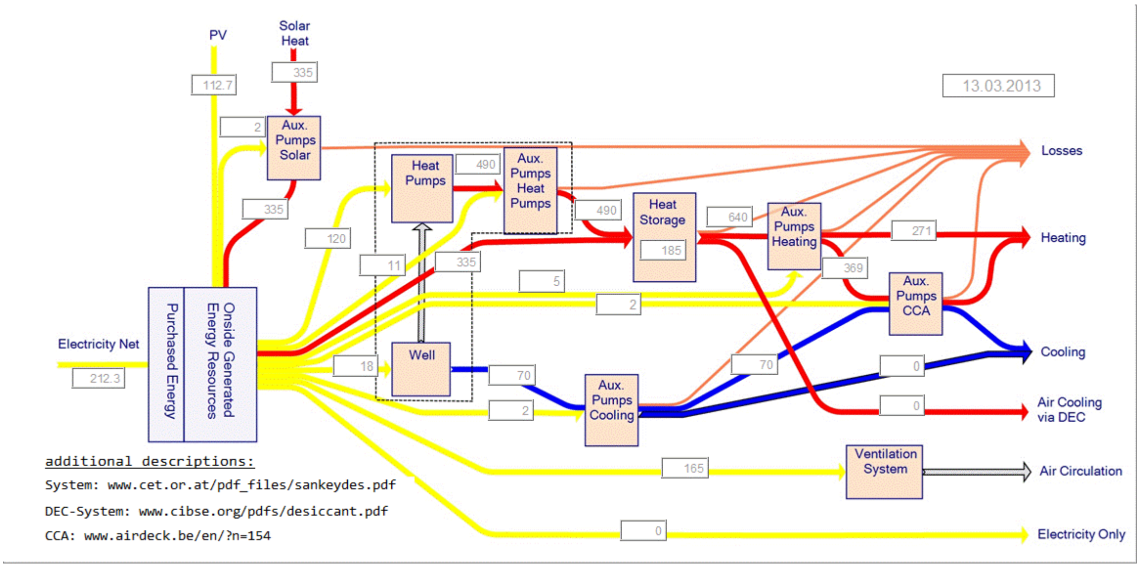

Figure 4: Sankey Diagram

1.5 Literature review

Regarding the strategic energy systems planning, different approaches can be found in the literature. Some of them deal with specific technologies (Siddiqui et al. 2005; Stadler et al. 2009). Other optimization models are designed from the production point of view (Hobbs 1995; El-Khattam et al. 2004; Heydari and Siddiqui 2010). Villumsen and Philpott (2012) apply Stochastic Optimization (STO) to capacity planning of electricity transmission networks with transmission switching, while Cai et al. (2008) focus on a regional perspective. Only recent papers tackle systems planning at the building level Salvador and Grieu 2012; Kumbaroğlu and Madlener 2011).

In terms of ICT solutions for energy-efficient buildings and areas of public use, most of the existing analyses follow either a power systems engineering framework (Weinberg et al. 1991; Van Sambeek 2000, or follow a deterministic optimization approach (Hobbs 1995; Siddiqui et al. 2005; King and Morgan 2007; Marnay et al. 2008; Stadler et al. 2009) that is unable to provide robust decisions against inherent uncertainties (Ermoliev and Wets 1988). Even though STO has been applied for a long time to cope with uncertainties in other fields, there were not approaches based on the use of techniques in order to treat uncertainties for energy efficiency in buildings.

The solution of the stochastic problem involves adjusting operational decisions to hit long-term targets if additional information about prices, demand, weather is revealed in the future (Gritsevskii and Nakicenovic 2000; Gritsevskii and Ermoliev 2012). A key innovation of the stochastic EnRiMa DSS is a combination of the proven methodology for modeling energy flows in buildings (Siddiqui et al. 2005) with the advances in effective coping with uncertainty (Ermoliev and Wets 1988; Gritsevskii and Ermoliev 2012; Ermoliev et al. 2012).

2 Decision making under uncertainty

2.1 Sources of uncertainty

Some decisions are made under *perfect information*, i.e., knowing all the outcomes and relevant facts affecting such decision. For example, one can decide whether to vent a room or not knowing the inside and outside temperatures and one's desired comfort level. However, this is not always the case. In many decision making processes, there is uncertainty pertaining relevant facts and figures around the decision. In particular, decision making on energy systems is strongly affected by both short-term and long-term uncertainties. Some of these sources of uncertainty are:

- Energy demand (short-term). The amount of energy demanded depends on things like weather or building occupancy. Even though in the short term accurate estimations can be made, long-term perspectives, which are much more volatiles, are needed for strategic decision making.
- Energy costs (short-term). Even for long-term contracts, energy prices are subject to volatility throughout the time. Moreover, new price schemes are emerging such as Time of Use (ToU) or intra-day tariffs.
- Investment costs (long-term). Systems investment cost on the long term is uncertain as it depends on future market evolution and eventual policy changes. Subsidies can be considered part of this kind of uncertainty.
- Availability of new technologies (long-term). New types of energy systems (more efficient, cheaper, or cleaner) might appear throughout the decision horizon. This fact may change decisions made today.

2.2 Decision Support Systems (DSSs)

When making decisions, three approaches can be followed:

- Intuition;
- Rules;
- Analysis.

Decisions based on intuition are often inconsistent and biased. Decisions based on rules are clear and require less effort, but could be too rigid for changing environments. Decisions based on analysis require the adoption of a **model** that summarizes the most essential parts of the problem in order to understand the real problem and find a way to solve it. Decision making problems can be classified (Bell et al. 1988) as descriptive, normative and prescriptive. According to the level in which decisions are made within an organization, decisions can be strategic, tactical, or operational. An instinctive level can be added at the bottom of this pyramidal classification (French et al. 2009; Klein et al. 1993). Note that strategic and operational levels are also linked to the long- and short-term scope of decisions.

In order to identify the more appropriate technique for a decision making problem, the following questions are helpful:

- Who is/are the decision maker/s?

- Which are the objectives?
- What are the uncertainty sources?
- How does time affect the process?
- Which are the requirements of the system?
- Who is/are affected by decisions?

In the case at hand, building managers and operators decide; they are also affected by decisions, along with the building users and other stakeholders; two different time resolutions (short- and long-term) are involved in the problem; uncertainty sources have been identified mainly regarding prices and demand. As for the objective, the most frequent case is the minimization of costs. These costs are a function of the decisions made and the data available. Other possible objectives, likely conflicting, may also be used, i.e., emissions minimization or efficiency maximization. In addition, there are some systemic requirements and limitations, such as the energy balance between demand and supply, and the capacity of systems and markets.

The problem outlined so far, is suitable to be modeled as a Mathematical Programming problem, where the objective is to optimize (maximize or minimize) an objective function, subject to a set of constraints. Moreover, as uncertainty is a key part of the problem, Stochastic Optimization (STO) will be used.

Considering the complexity of the problem, the use of a DSS is unavoidable. Usually defined as an *information system that supports decision making* with more or less detail, this term has been often abused in Computer Science and in Management. Thus, any information system could claim to be a DSS. However, more specific boundaries are needed to capture the analysis approach mentioned above. Under that paradigm the model plays an important role in a DSS. Both the model and the data are the basis for the decisions. The DSS should be also capable of preparing the data in a model-suitable way. The model must be based on strong scientific knowledge. Appropriate algorithms are applied once the model is defined and the data is available. Decisions obtained by the DSS, regardless their category (descriptive, normative, or prescriptive), should include interpretation and analysis, probably requiring some post-data analysis.

It is important that, an effective DSS must be able to provide an environment for **stakeholders dialog**. Figure 5 reflects the whole structure of the framework. Usually decision making is not a static action, but rather an iterative process, regardless the time cycle duration. Therefore, the outcomes of the process provide endogenous feedback to the DSS structure. Useful analyses at different levels must be provided by the DSS that enforce the necessary stakeholders dialog. The purpose of this dialog is twofold: On the one hand, a dialog between the stakeholders and the DSS; on the other hand, between the stakeholders, likely with different motivations and targets. Examples of the former are:

- Comprehensive and understandable output reporting. It should be self-contained, pointing to the details for different stakeholders.
- The output should tell the decision maker and other stakeholders, in addition to the recommended optimal decisions, about the interpretation of the results, consequences, implementation, and usefulness.
- Some examples of stakeholders are decision makers, consultants, modelers, and data managers. All of them interact with the DSS in a continuous base through the feedback received in form of new inputs.

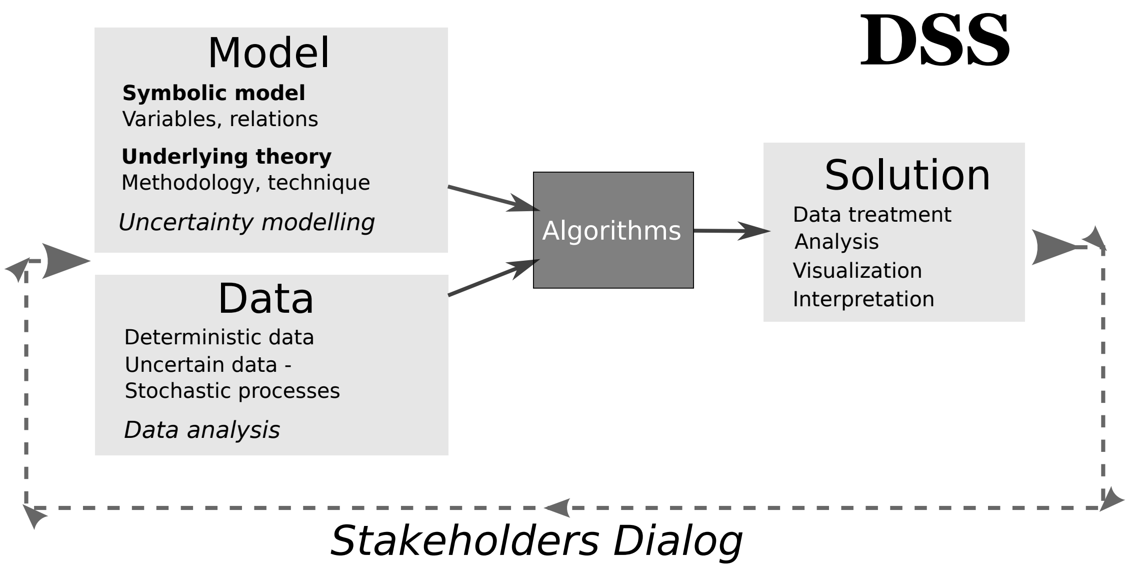

Figure 5: Decision Support System (DSS) diagram

- Sometimes the output of the DSS is to be used externally, e.g., by policy makers or mass media. This requires a different type of dialog, but consistent with the inner one.

Examples of the dialog between stakeholders are:

- Often data managers or operators are different to modelers, and a fluid communication between them is crucial for the accurateness of the inputs. Not for nothing the subsequent results will be based on these inputs.

- Decision based on analysis relies on the abstraction of the reality using models. Sometimes, for varied reasons, the models do not appropriately fit into the reality, and dialog between modelers and, let us say, *process owners*, is decisive. Depending on the problem at hand, such process owner can be the own decision maker, or any other recipient of the output, e.g., advisors, operators, technicians, or managers.

- Dialog with external stakeholders may be necessary at any point of the decision making process, e.g., between data managers and data providers, between managers and policy makers, mass media, shareholders, etc. Notice that this dialog may also provide exogenous feedback to the process

Considering the premises outlined above, some desirable features for a framework would be:

- Data analysis capabilities, including: statistical analysis, data cleaning;

- Data visualization capabilities;

- Generation of information both human and machine readable;

- Reporting capabilities;

- Implementable in user interfaces, including web interfaces;

- Interfaces to data sources;
- Interfaces to specialized optimization software;
- Flexibility for different representation systems;
- Flexibility for different algorithms and solvers;
- Adaptability to changes.

Different approaches to DSSs can be found in the literature. Some of them focus on the model as a way to provide decision support, other focus on the infrastructure of the DSS, or on a particular application (Salewicz and Nakayama 2004). Tanaka et al. (1995) proposed a DSS for multicriteria decision making which includes decision maker interaction. González et al. (2009) presented a generic core to build optimization-based DSSs and defined a framework for developing a DSS with web services. A brief history of DSS can be found in Power (2007). Some of the topics discussed in Shim et al. (2002) are tackled in the framework proposed in this report. The framework includes innovative features and provides a flexible framework fulfilling all the features enumerated above, and can be implemented in usable DSSs through the appropriate interfaces. The use of both human and machine readable formats through the use of Algebraic Modeling Languages (AMLs) boosts the dialog between stakeholders remarked in this subsection. An important implementations of machine-readable models is the Structure Modeling Language (SML) (Geoffrion 1992a; Geoffrion 1992b). On the other hand, the Reproducible Research approach (Leisch 2002) adopted in the following allows to record and track consistent updates throughout the time, and to provide a sort of balance scorecard to stakeholders consistent with all the components of the DSS. Furthermore, the results are reproducible for any of the stakeholders, which increase the efficiency in multi-disciplinary and changing environments, and the quality of the communication processes. Though a cutting-edge topic, only very recent works deal extensively with Reproducible Research (Stodden et al. 2013).

2.3 DSS model component

Within the structure of the DSS, the model component is represented by the Symbolic Model Specification (SMS), e.g., see Subsection 3.2. The SMS defines the mathematical representation of the optimization model, including all relevant subsystems and their interactions. This mathematical representation is composed of variables, parameters, and relations between them. Such relations are, in turn, represented by equations and inequations. Sets are used to represent parameters and variables membership, as well as domains and conditions within equations. The models applied to the specific problem of energy systems optimization at the building level are developed in Section 3. In order to generically represent the SMS within the DSS, specific data structures developed in R (R Core Team 2013) are proposed.

The proposed framework relies on the use of Algebraic Modeling Languages (AMLs), in contrast to the use of whole matrices to represent the optimization problems. The advantages of AMLs versus matrix-like systems have been largely discussed (Fourer 1983; Kuip 1993). Recent advances on AMLs can be found in Kallrath (2012a). Nevertheless, usually optimization software accepts matrix files with the model coefficients and actually modeling software generates the matrix from the algebraic language. The process however is usually more straightforward and less prone-error when using AMLs, as the modeler have just to write the model, and the coefficients are generated combining the data and the model. MPL and LP are the most used file formats for matrix data models.

AMLs are "declarative languages for implementing optimization problems" (Kallrath 2012c). They are able to include the elements of optimization problems in a similar way they are formulated mathematically using a given syntax that can be interpreted by the modeling software. This approach is essential for representing the models not only for machines, but also for humans, and allows to organize the stakeholders dialog. One of the capabilities of the framework is to represent the models in LaTeX format, which is one of the "Practioner's Wish List Towards Algebraic Modeling Systems" (Kallrath 2012b). The following list enumerates some of the most important AMLs:

- GAMS (General Algebraic Modeling System);
- AMPL (A Modeling Language for Mathematical Programming);
- AIMMS (Advanced Interactive Multidimensional Modeling System);
- CMPL (COIN Mathematical Programming Language) is an AML within the Computational Infrastructure for Operations Research (COIN-OR) project;
- MathProg is the algebraic language used by GLPK (GNU Linear Programming Kit), implements a subset of AMPL;
- Pyomo (Python Optimization Modeling Objects) is a Python package, part of the Coopr software library, which includes modeling capabilities in a high-level language.

2.4 DSS data component

Some of the AMLs described above and some of the software packages in the next subsection include data import and export capabilities, and even some analysis functionality. However, it is common that analysts and modelers use specific data analysis software to make the data available for the DSS. In this regard, there are a wide range of options both commercial and open source. A non-exhaustive list would include:

- Stata (http://www.stata.com/): widely used in econometrics;
- SPSS (http://www-01.ibm.com/software/analytics/spss/): time-honored statistical software, recently acquired by IBM;
- SAS (http://www.sas.com/): The leader in business analytics;
- Minitab (http://www.minitab.com/): Well-known statistical software for quality control and improvement;
- R (http://www.r-project.org/): The R statistical software and programming language is a free, open source software that is increasing its use as data analysis and visualization software in academics, governmental agencies, and companies.

The data component of a DSS can be also developed using general-purpose programming languages such as C++ or Java, or specific programming language libraries, e.g., pandas and matplotlib for Python. Moreover, interfaces to diverse data sources may be needed in order to import and export data from/to the existing data sources. For Stochastic Programming (SP), scenario generators are also needed to combine the data and the uncertainty modeling in order to provide the DSS with the appropriate inputs.

2.5 The system

There is a component of the DSS in charge of running the optimization, that in our case is given by the strategic-operational (two-stage) model of Section 3.2. Usually it is a piece of software containing the algorithms to solve optimization problems, and it is in general named the solver. Solvers are usually available as standalone, low-level applications that can be embedded in high-level applications, i.e., with a user interface. Solvers may be specific for a given optimization type of problem, e.g., Linear Programming (LP), Non Linear Programming (NLP), or for different types of problems. The following is a non-exhaustive list of commonly used solvers:

- CPLEX (http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html): For linear and quadratic problems;
- lp_solve (http://lpsolve.sourceforge.net/5.5/) is an open source solver for problems;
- CLP (http://www.coin-or.org/projects/Clp.xml), the solver of the COIN-OR project;
- BARON: For non-convex, non-linear problems;
- MINOS: For NLP, developed by the Systems Optimization Laboratory at Stanford University;
- CONOPT: Another non-linear solver;
- IPOPT: for large scale nonlinear optimization of continuous systems, is part of the COIN-OR project;
- GUROBI: This successful collection of solvers was developed from the ground.

More open source solvers can be found in the COIN-OR projects' website (http://www.coin-or.org/projects/). In addition to solvers' projects, developer tools and interfaces can also be found. For example, the OS project (https://projects.coin-or.org/OS), whose objective is "to provide a set of standards for representing optimization instances, results, solver options, and communication between clients and solvers in a distributed environment using Web Services".

The AMLs explained above, as well as other optimization software, contain solvers that are called once the model and the data are available. The list of solvers available for each optimization software is provided in the documentation of each system. For example, the list of solvers supported by GAMS can be consulted at http://www.gams.com/solvers/. The use of some commercial solvers may require additional licenses.

In addition to AMLs, other software packages can be used for optimization. For example, scientific software such as Matlab (http://www.mathworks.de/products/matlab/), SciLab (https://www.scilab.org), or Mathematica (http://www.wolfram.com/mathematica/), among others, include modules to solve mathematical programming problems, or to call further solvers. OpenOpt (http://openopt.org/) is an open source option. Last but not least, spreadsheets such as Microsoft Excel or LibreOffice Calc can solve optimization problems.

In summary, it is common to find different components of a DSS disconnected between them. A heterogeneous set of tools is often being used for similar tasks that unfortunately blocks stakeholders dialog. In contrast, the proposed framework uses R for all tasks including data analysis, visualization and representation tasks, allowing communicating to different optimization software through inner interfaces. Data cleaning and management can also be done easily with R and access through interfaces can be easily provided, both through other technologies such as php[5], or .NET[6], or through libraries devoted to user interfaces, such as shiny[7].

3 DSS Models

To illustrate the problem, a simple example will be used. It is inspired by the classical news vendor problem used in many textbooks (e.g., Ermoliev and Wets 1988; Birge and Louveaux 2011). Suppose a building manager can decide each year the energy capacity x of the building. For simplicity in the exposition, aggregated values and decisions are assumed. The price of each unit of capacity, e.g., kW, is c. During the year, the energy demand varies following a probability distribution described by a random variable ξ. If the demand is higher than the capacity, i.e., $\xi > x$, then the building manager has to increase the capacity in order to fulfill the demand, but at a higher cost $d^+ > c$. If the demand is lower than the capacity x, i.e., $\xi < x$ then the building manager can sell energy at a lower price $d^- < c$. Let y^- (y^+) be such excess (shortage) of capacity. Then, for a given ξ, the cost function for the building energy procurement is:

$$cx + d^+y^+(\xi) - d^-y^-(\xi). \tag{1}$$

Note that in these types of problems, there are strategic first-stage decisions x that are to be made before uncertainty ξ is resolved and operational second-stage decisions y that are made once uncertainty is resolved. As we have seen above, the optimal value of the second stage decision depends on both the random variable ξ and the first-stage decision x: ${y^+}^* = \max\{0, \xi - x\}$ and ${y^-}^* = \max\{0, x - \xi\}$. Therefore, the expected value of the cost function we want to minimize can be expressed as:

$$\begin{aligned} C(x) = cx + \mathbb{E}_\xi \left[d^+y^+(\xi) - d^-y^-(\xi)\right] = \\ cx + \mathbb{E}_\xi \left[d^+ \max\{0, \xi - x\} - d^- \max\{0, x - \xi\}\right], \end{aligned} \tag{2}$$

where $\mathbb{E}[\cdot]$ is the mathematical expectation function. Developing the following optimality condition under optimal $x > 0$:

$$C'(x) = \frac{\partial C}{\partial x} = 0, \tag{3}$$

where $C'(x)$ denotes the first order derivative of $C(x)$ evaluated at x, yields the following expression:

$$\mathbb{P}\left[\xi < x\right] = \frac{d^+ - c}{d^+ - d^-}. \tag{4}$$

[5] http://php.net/
[6] http://www.asp.net/
[7] http://www.rstudio.com/shiny/

where $\mathbb{P}[\cdot]$ is the probability function. So, the probability of the demand being lower than the strategic decision is fixed by the data. Given that $d^+ > c > d^-$, Equation (4) assures a level of security for the solution. This solution, in turn, depends on the probability distribution of ξ. Therefore, the solution of two-stage stochastic problems ends up in the fulfillment of some security level. Such solutions are optimal for all the scenarios at a time, thereby providing **robust solutions** for strategic decisions. In contrast, the solution of the deterministic problem, i.e., substituting the uncertain parameters ξ by its expectation $\mathbb{E}[\xi]$ and solving the optimization problem for the average scenario, which might never occur. Likewise, solving the *worst case* scenario, i.e., using $\max\{\xi\}$ as fixed, would be too conservative and unrealistic, consequently leading to very high costs.

In this example both first- and second-stage decisions are represented within a given time horizon. Due to the own structure of the problem, operational decisions induce risk aversion on strategic decisions. As the operational periods are embedded into the strategic ones, the size of the model tremendously increases. Several modeling approaches to deal with these issues are detailed and compared in Section 3.3.

3.1 The deterministic approach

A simplistic way to deal with optimization problems under uncertainty is to estimate the parameter values through its expected value and solve the corresponding deterministic problem. This approach may lead to wrong decisions for several reasons. First of all, it provides degenerated optimal solutions for the average scenario that may never occur. And, more importantly, the solution for the average scenario can be infeasible for real scenarios, providing, e.g., not enough capacity to fulfill the real demand of energy. As a straightforward metaphor, would anyone go to a hospital whose patients receive treatment according to the average body temperature of all of them?

In the baseline example, the deterministic solution would be the following:

$$x^{*det} = \mathbb{E}[\xi] \tag{5}$$

That is to say, the capacity to be installed in the building is exactly the expected demand. This degenerated solution may result in shortfalls as there are no second stage decisions.

Instead of using expected values of data $\xi(\omega)$, a Stochastic Programming (SP) model is formulated as the optimization of the expected value of the objective function of the type (2):

$$\min_{\boldsymbol{x}} C(x) = \mathbb{E}_\omega \left[f\left(\boldsymbol{x}, \boldsymbol{\xi}(\omega)\right)\right], \tag{6}$$

$$\text{s.t. } x = (x_1, \ldots, x_n) \geq 0 \tag{7}$$

Solutions of SP problems are in general not optimal for any possible scenario. Nevertheless, this solution is the best one considering all the plausible scenarios, and therefore it is a robust solution. In fact, the solution of the deterministic problem results always in a worse value of the objective function considering all potential scenarios. These inequalities were demonstrated by Madansky (1960), and later on Birge (1982) defined the value of the stochastic solution as the difference between the value of the expected objective function using the deterministic problem solution x^{*det} and the value of the expected objective function using the solution of the SP problem x^{*sto}:

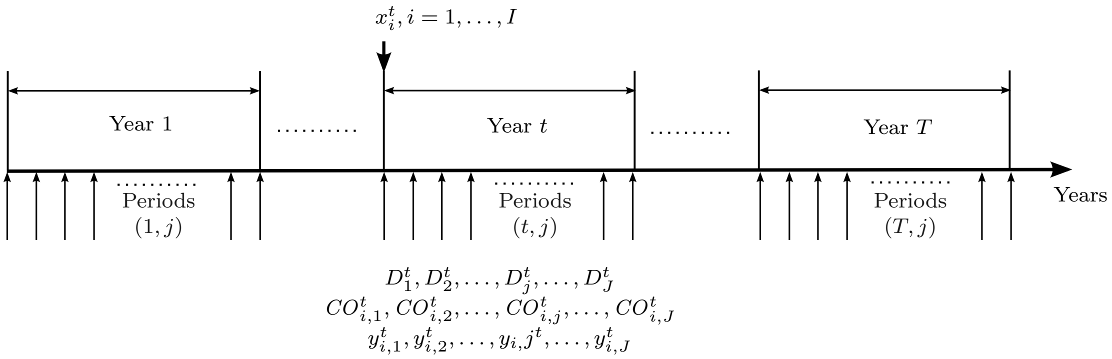

Figure 6: Temporal resolutions of the strategic planning model.

$$VSS = C(\boldsymbol{x}^{*det}) - C\left(\boldsymbol{x}^{*sto}\right). \tag{8}$$

3.2 The dynamic two-stage model

Strategic and operational decisions concern demand and supply sides of different energy loads and resources (electricity, gas, heat, etc.). The demand side is affected by old and new equipment and activities including such end uses as electricity only, heating, cooling, cooking, new types of windows and shells, and energy-saving technologies, etc. For example, new activities may change peak loads. Accumulators such as batteries may considerably smooth energy demand -supply processes.

The supply side is affected by decisions on new technologies. The notion of technology must be understood in a rather broad sense. This may be either direct generation of electricity and heat, or the purchase of certain amounts of, e.g., electricity from a market, i.e., the market can also be viewed as energy generating technology with specific cost functions. Independently of the content, different options i are available at time t to satisfy energy demand, $i \in \mathcal{I} = \{1, \ldots, I\}, t \in \mathcal{T} = \{1, \ldots, T\}$. For each case study, feasible options at time t have to be characterized explicitly.

The model is dynamic and the planning horizon comprises T years. Uncertainties pertaining to demands, fuel prices, operational costs, and the lifetime of technologies are considered. Demand may be affected by weather conditions. It may also substantially differ by the time of the day and the day of a week. However instead of considering 8760 hourly values, demands and prices are aggregated into J periods representatively describing the behavior of the system within a year. Similar approaches can be found in the literature (Conejo et al. 2007).

The demand profile within each year t, can be adequately characterized by the demand within representative periods j, $j \in \mathcal{J} = \{1, \ldots, J\}$. This time structure is represented in Figure 6, where D_j^t, $CO_{i,j}^t$ denote the energy demand and costs of technology i in period j of year t, and $y_{i,j}^t$ are operational decisions for technology i in period j of year t. The goal of the strategic model is to find technologies i and their capacities x_i^t, installed at the beginning of year t in order to satisfy demands D_j^t, in each period j.

Formally, assume planning time horizon of T years. Let x_i^t be the additional capacity of technology i installed in year t, and s_i^t the total capacity by i available in t. Then

$$x_i^t \geq 0 \qquad \forall i \in \mathcal{I}, t \in \mathcal{T}, \tag{9}$$
$$s_i^t = s_i^{t-1} + x_i^t - x_i^{t-LT_i} \qquad \forall\, i \in \mathcal{I},\ t \in \mathcal{T}, \tag{10}$$

where LT_i is the lifetime of technology i and s_i^0 is initial capacity of i existent before $t = 1$.

In addition to operational costs $CO_{i,j}^t$, investment costs CI_i^t are considered. In general, the operational and investment costs, as well as energy demand D_j^t, are uncertain. Strategic first stage investment decisions x_i^t, are made at the beginning of the planning horizon $t = 1$ using a perception of potential future scenarios $CI_i^t(\omega)$, $CO_{i,j}^t(\omega)$, $D_j^t(\omega)$ of costs and energy demands dependent on the stochastic parameter ω. Here ω is used to denote a sequence $\omega = (\omega_1, \omega_2, \ldots, \omega_t, \ldots, \omega_T)$ of uncertain vectors ω_t of in general interdependent parameters which may affect outcomes of the strategic model, e.g., market prices or weather conditions. In general, there are different components of ω_t, e.g., components ω_t^{dem} characterizing the variability of the demand, and other components ω_t^{str}, ω_t^{ope} characterizing uncertainties associated with strategic and operational costs. Therefore, functions $CI_i^t(\omega)$, $CO_{i,j}^t(\omega)$, $D_j^t(\omega)$ depend in general only on some components of ω_t, although dependence on ω is indicated for simplicity of notation. Second stage adaptive operational decisions $y_{i,j}^t$ are made after observing real demands and costs. They depend on observable scenario ω, i.e., $y_{i,j}^t = y_{i,j}^t(\omega)$. Therefore, any choice of investments decisions $x = x_i^t$, may not yield feasible second stage solutions $y(\omega) = y_{i,j}^t(\omega)$ satisfying the following equations for all ω:

$$\sum_{i \in \mathcal{I}} y_{i,j}^t(\omega) = D_j^t(\omega) \qquad \forall\, j \in \mathcal{J},\ t \in \mathcal{T}, \tag{11}$$
$$y_{i,j}^t(\omega) \leq G_{i,j}^t \cdot s_i^t \qquad \forall\, i \in \mathcal{I},\ j \in \mathcal{J},\ t \in \mathcal{T}, \tag{12}$$
$$y_{i,j}^t(\omega) \geq 0 \qquad \forall\, i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T}, \tag{13}$$

where $G_{i,j}^t$ may be interpreted as the availability factor corresponding to the technology operating in period j in t ($G_{i,j}^t = 0$ for not yet existing technologies).

The feasibility of constraints (11)–(13) for any scenario ω can be guaranteed by assuming the existence of a back-stop technology with high operating costs that can also be viewed as purchasing without delay but at high price. In particular, it can be viewed as a contingent credit or a catastrophe (black-out) bond, similar as in Ermoliev et al. 2012. Without loosing generality it can be assumed that for any period j and time t it is the same technology $i = 1$. Then the basic dynamic stochastic two-stage model is formulated as the minimization of the expected total cost function:

$$\mathbf{F}(x) = \mathbb{E}_\omega \left[\min_{y(\omega)} \sum_{i \in \mathcal{I}, t \in \mathcal{T}} \left(CI_i^t(\omega) \cdot x_i^t + \sum_{j \in \mathcal{J}} CO_{i,j}^t(\omega) \cdot DT_j^t \cdot y_{i,j}^t(\omega) \right) \right] = \tag{14}$$
$$= \sum_{i \in \mathcal{I}, t \in \mathcal{T}} \left(CI_i^t \cdot x_i^t + \mathbb{E}_\omega \left[\min_{y(\omega)} \sum_{j \in \mathcal{J}} CO_{i,j}^t(\omega) \cdot DT_j^t \cdot y_{i,j}^t(\omega) \right] \right),$$

where $\mathbb{E}[\cdot]$ is the expectation function. This SP problem can be easily extended in order to deal with advanced energy systems features such as efficiency, emissions, or storage (**?**).

3.3 Numerical methods: learning by doing and rolling time horizons

The model (9)–(14) is formulated in the space of variables

$$\left(x_i^t, y_{i,j}^t(\omega), i \in \mathcal{I}, t \in \mathcal{T}, \omega \in \Omega\right),$$

where the set of scenarios Ω may include a finite number of implicitly given scenarios, e.g., by scenario trees (Kaut et al. 2013). A realistic practical model (9)–(14) excludes analytically tractable solutions, although the model has an important block-structure that is usually utilized for most effective numerical solutions in DSS.

In a rather general case, Ω contains or can be approximated by scenarios ω_s, $s \in \mathcal{S}$, characterized by probabilities p_s, $s \in \mathcal{S}$. Then the model (9)–(14) is formulated as the minimization of the function:

$$\sum_{s \in \mathcal{S}} p_s \left[\sum_{i \in \mathcal{I}, t \in \mathcal{T}} \left(CI_i^t(\omega_s) \cdot x_i^t + \sum_{j \in \mathcal{J}} CO_{i,j}^t(\omega_s) \cdot DT_j^t \cdot y_{i,j}^t(\omega_s) \right) \right], \tag{15}$$

subject to:

$$\sum_{i \in \mathcal{I}} y_{i,j}^t(\omega_s) = D_j^t(\omega_s) \; \forall \, j \in \mathcal{J}, \; t \in \mathcal{T}, \; s \in \mathcal{S}, \tag{16}$$

$$y_{i,j}^t(\omega_s) \le G_{i,j}^t \cdot s_i^t \; \forall \, i \in \mathcal{I}, \; j \in \mathcal{J}, \; t \in \mathcal{T}, \; s \in \mathcal{S}, \tag{17}$$

$$y_{i,j}^t(\omega_s) \ge 0 \; \forall \, j \in \mathcal{J}, \; t \in \mathcal{T}, \; s \in \mathcal{S}, \tag{18}$$

$$s_i^t = s_i^{t-1} + x_i^t - x_i^{t-LT_i} \; \forall \, i \in \mathcal{I}, \; t \in \mathcal{T}, \tag{19}$$

$$x_i^t \ge 0 \; \forall i \in \mathcal{I}, t \in \mathcal{T}. \tag{20}$$

Remark 1. *Learning-by-doing: models with rolling horizons.* The initial model (15)–(20) is focused on time horizon $[1, T]$. The robust strategic solution w.r.t. scenarios $\omega = (\omega_1, \ldots, \omega_T)$ can be written as:

$$\boldsymbol{x}^{[1,T]} = \left(x_i^{1,[1,T]}, \ldots, x_i^{T,[1,T]}\right), i \in \mathcal{I}.$$

Solutions $\left(x_i^{1,[1,T]}\right)$, $i \in \mathcal{I}$, are implemented at $t = 1$ that may reveal significant new information about future uncertainties. Let us denote scenario ω for interval $[1, T]$ as $\omega^{[1,T]}$. New information provides a basis for readjustments of scenarios $\omega^{[1,T]}$ perceived at the beginning of time horizon $[1, T]$. Then, new set of scenarios $\omega^{[2,T+1]}$ are evaluated, robust strategic solutions $\left(x_i^{2,[2,T+1]}\right)$, $i \in \mathcal{I}$ for $t = 2$, are obtained, and so on. Thus, initially a long-term strategic trajectory $x^{[1,T]}$ is evaluated, the first time interval solutions $\left(x_i^{1,[1,T]}\right)$, $i \in \mathcal{I}$, new data are received, new scenarios $\omega^{[2,T+1]}$ and solutions $x^{[2,T+1]}$ are adjusted, and so on. This approach introduces a new type of models incorporating endogenous scenario generation shaped by previous decisions, i.e., learning-by-doing procedures.

3.4 Value of Stochastic Solution

In this subsection the value of stochastic optimization models, often termed (Birge 1982, Delage et al. 2012) as the Value of Stochastic Solution (VSS) is discussed in some details. This notion has a misleading character because the two-stage models incorporate both ex-ante deterministic first stage decisions chosen before observations of uncertain parameters (events) and ex-post stochastic adaptive decisions chosen when additional information becomes available. We have to emphasize that the VSS or may be better the Value of Stochastic Modeling (VSM) is different from the expected value of perfect information which is defined as the improvement of the objective function by learning perfect information about parameters of the true deterministic model. In other words, the advantage of using deterministic models with exact values of parameters which are the mean values of observable random variables. On the contrary, the VSS considers deterministic models as an approximation of real stochastic optimization models with inherently uncertain parameters which cannot be evaluated exactly by a single value. Disadvantages of using deterministic approximations of stochastic models were outlined in the baseline example. The advantages of using SP will be further demonstrated in the following section using real data.

The VSS is calculated by the non-negative difference:

$$\mathbf{F}(x^{*det}) - \mathbf{F}(x^{*sto}), \tag{21}$$

where $\mathbf{F}(x)$ is defined by (14), x^{*det} is the optimal solution of the deterministic version of model (9)–(14) used in the objective function of the stochastic model, i.e., Equation (14), and x^{*sto} is the optimal solution of this stochastic model. Non-negativity is due to the fact that the feasible set of the stochastic model includes the feasible set of the deterministic model. The solution x^{*det} is often calculated with the parameters, e.g., demand, which have been substituted by mean values combined with a sensitivity analysis. Results of such approaches are usually rather misleading because sensitivity analysis of the deterministic model with respect to variations of its parameters is focused on one only scenario (mean value) that may never occur in reality. The robust solution of the stochastic model depends on the whole probability distribution, see Equation (4), therefore variations in the mean values may be misleading especially for multimodal distributions. For example, in the case with two scenarios $-10, +10$ with probability 0.5, the mean value is even outside the set of feasible scenarios. In addition to the sensitivity analysis, the so-called scenario analysis is applied, i.e. a set of possible future "trajectories" of uncertain parameters is considered and for each of them optimal solutions of the deterministic model are calculated. This generates a set of degenerated deterministic solutions without identifying a solution that is equally good (robust) with respect to all potential scenarios.

4 DSS Data

4.1 Two-stage problem instance

In this section, real data from the EnRiMa project are used in order to demonstrate the modeling approach. In particular, historical data from the FASAD EnRiMa test site in Asturias (Spain) has been used (see Subsection 1.3). Let us consider the model defined by (15)–(20). Starting from base values, the future development of the parameter values have been modeled through expert opinions getting average values and standard deviations for annual variations, see Table 1.

Assuming normal distributions, a set of 100 scenarios ω_s have been simulated. Figure 7 shows a representation of this simulation, where the dark-red line indicates the average value of the parameter. For the sake of simplicity, only four representative periods (set $\mathcal{J}$) have been defined: winter, spring, summer and autumn. The input technologies (set $\mathcal{I}$) are

Table 1: Base parameter values an uncertain evolution

Paramter	Base value	Average variation	Variation Std.Dev.
CI_{RTE}	50.00	0.10	0.04
CI_{CHP}	795.99	-0.10	0.05
CI_{PV}	2204.26	-0.05	0.06
CO_{RTE}	0.13	0.10	0.04
CO_{RTG}	0.05	0.03	0.02
D	24.37	0.10	0.05

Regulated Tariff of Electricity (RTE), Photovoltaic (PV) and Combined Heat and Power (CHP). In this simple example with only electricity demand, it is assumed that the heat produced by the CHP technology is not used. Regarding the technologies availability, RTE and CHP are always available ($G^t_{i,j} = 1$), whereas PV availability depends on the season as shown in Table **??** (assuming the same values for all the years). A Sunmodule SW 245 by Solarworld has been considered (http://www.solarworld.de/en/home/). The availability factor has been computed using the on-line PGIS tool (Photovoltaic Geographical Information System) by the European Commission Joint Research Center – Institute for Energy and Transport, http://re.jrc.ec.europa.eu/pvgis/apps4/pvest.php.

```
## Error:  could not find function "thesisTable"
```

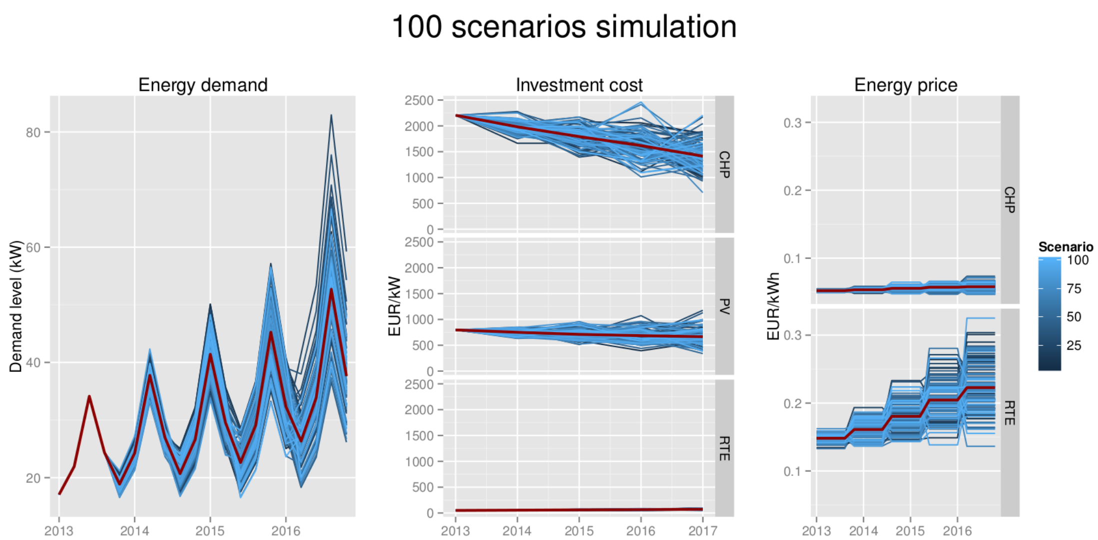

Figure 7: Scenarios simulation for two-stage model

As for investment costs CI^t_i, the price for the PV panels has been taken from the PREOC price database (http://www.preoc.es/retrieved2013-02-12), whilst the price for the CHP has been gathered from the on-line seller myTub (http://www.mytub.co.uk/product_information.php?product=465447, retrieved 2013-02-12). A 40% reduction has been applied to the investment costs in order to take into account available subsidies in the market (http://www.faen.es/nueva/Intranet/documentos/3577_Bases.pdf). This parameter also gathers a cost of contracting RTE of 50 EUR/kW, which increases at the same rate as the energy cost. For the operational costs $CO^t_{i,j}$, the base fuel

Table 2: Strategic solutions for the two-stage problem

i	t	value
RTE	2013	45.65
PV	2013	57.65
PV	2014	1.78

prices for electricity and natural gas are 0.134571 EUR/kWh and 0.05056 EUR/kWh for RTE and CHP respectively, based on the EnRiMa project deliverable D1.1 "Requirement Assessment", and no cost for PV. As a short horizon is considered, the lifetime parameter LT_i, which has been set to 20 years, has no influence on the result. Finally, the duration time is set to 91 days × 8 hours, considering 13 weeks each period.

Solving the SP problem the strategic decisions to be made are (see Table 2): contracting 45.65 kW to RTE and installing 57.65 kW of PV the first year, and extend the PV installation the second year in 1.77 kW. Note that the actual decisions to be made by the building manager are those for the first year.

The total cost stemming from those decisions is 68,595 EUR. If we assumed average values for the uncertain parameters, i.e., solve the deterministic problem using the mean values represented in Figure 7 as the dark-red line, we would get a total cost of 66.920 EUR and slightly different values for the decision variables. One could think that the deterministic solution is better than the stochastic one. But this is an illusion, because if we analyze the variability (robustness) of solutions using separately the 100 different scenarios, we realize that the solution returned by the deterministic optimization is infeasible for 56 of them. This means that more than half the times the capacity of the building will not be able to fulfill the requirements of energy. On the contrary, the solution returned by the SP problem is a robust solution against all the scenarios.

4.2 VSS computation

In order to compute the VSS, the first-stage decisions obtained in the deterministic problem are fixed in the SP problem (15)–(20), which is then solved. The solution of this problem is called the expected result of using the expected value problem (Birge 1982) and represented by $\mathbf{F}(x^{*det})$, while the solution of the SP problem is represented by $\mathbf{F}(x^{*sto})$. In this case, as $\mathbf{F}(x^{*det})$ is infeasible, it is considered infinite and therefore the VSS, see equation (21) above, is infinite:

$$\mathbf{F}(x^{*det}) - \mathbf{F}(x^{*sto}) = \infty - 68,595 = \infty.$$

It is important to remark that even if $\mathbf{F}(x^{*det})$ is feasible, the VSS is positive, and the magnitude will depend on the uncertainty structure. The value $\mathbf{F}(x^{*sto})$ is smaller than $\mathbf{F}(x^{*det})$ because the stochastic model has a richer set of feasible solutions, i.e., the deterministic solution x^{*det} is a degenerated version of x^{*sto}.

5 DSS Framework

5.1 A reproducible research approach

Against the "copy-paste" approach frequently used to reach the final outcome of a decision making problem, the reproducible research one adopted in the framework developed has a series of advantages worthy to consider, namely:

- When coming back to the research in the future, the results can be easily obtained again.
- In case other researchers have to contribute to the work, all the process is at hand.
- Changes on any step of the process (e.g. a new index in the mathematical model) are made seamlessly just changing the appropriate data object. The whole analysis is made again with the new information, and the changes are automatically reflected in the output results.
- The results can be verified by independent reviewers. This is particularly important in health research and other disciplines where security is an issue. A paradigmatic example to realize the importance of reproducible research is the scandal of the Duke cancer trials (CBS 2012; The New York Times 2011). For an example on energy issues see Jelliffe (2010).

In order to fulfill the requirements for a DSS detailed in Section 2 under the reproducible research approach, an R library has been developed. The R Project for Statistical Computing is becoming the "de-facto standard for data analysis", according to more and more authors from a variety of disciplines, from Ecology to Econometrics (Cano et al. 2012). "R is a system for statistical computation and graphics. It consists of a language plus a run-time environment with graphics, a debugger, access to certain system functions, and the ability to run programs stored in script files" (Theussl and Hornik 2013). As mentioned above, decision making needs statistical software in order to prepare, analyze, and present data. Some of the advantages of choosing R as the statistical software for DSS are:

- It is Open Source.
- It has Reproducible Research and Literate Programming capabilities (Leisch 2002).
- It can be used as an integrated framework for models, data and solvers.
- It supports advanced data analysis (pre- and post-), graphics and reporting.
- Interfacing with other languages, as C or Fortran is possible, as well as wrapping other programs within R scripts.

These capabilities allow the researcher to apply innovative methods and coherent results increasing the productivity and reducing errors and unproductive time. Some of the strengths of the R project are:

- The system runs in almost any system and configuration and the installation is easy.
- There are thousands[8] of contributed packages for a wide range of applications of R, covering statistics, econometrics, optimization, simulation, data mining, graphics, and many other topics. The packages are freely available in repositories as The Comprehensive R Archive Network (CRAN, http://cran.r-project.org).
- The system can be extended with new libraries and functions, either public or private, to fulfill any requirement, for example: customization, deployment of new methods, integration with existing systems and data bases, etc.

[8]4847 at the Comprehensive R Archive Network (CRAN) on September 23, 2013. Other repositories are bioconductor, omegahat, r-forge, and github.

- The system can be adapted to the needs of any user. If there is a function that a user would like to perform in a different way, they can modify it accordingly to their needs. Moreover, it is easier to detect bugs and errors as one can dig into the code.
- The active R-Core development team jointly with the huge community of users provide an incredible support level (without warranty, skeptics would say), difficult to surpass by other support schemes.
- New methods, tools or algorithms can be deployed very fast. A company or organization can develop and deploy an innovative method from its R&D department, or from the result of other published research.

5.2 The optimr R package

An R package called optimr has been developed as part of the DSS described in this paper to deal with the model, the data, and the solutions. The optimr library revolves around two classes of objects: `optimSMS` and `optimInstance`. The former contains the Symbolic Model Specification (SMS), i.e., the mathematical model including all the entities such as parameters and variables and their interrelations. The latter contains the data of the particular instance of the problem to be solved. Figure 8 shows an outline of the package structure.

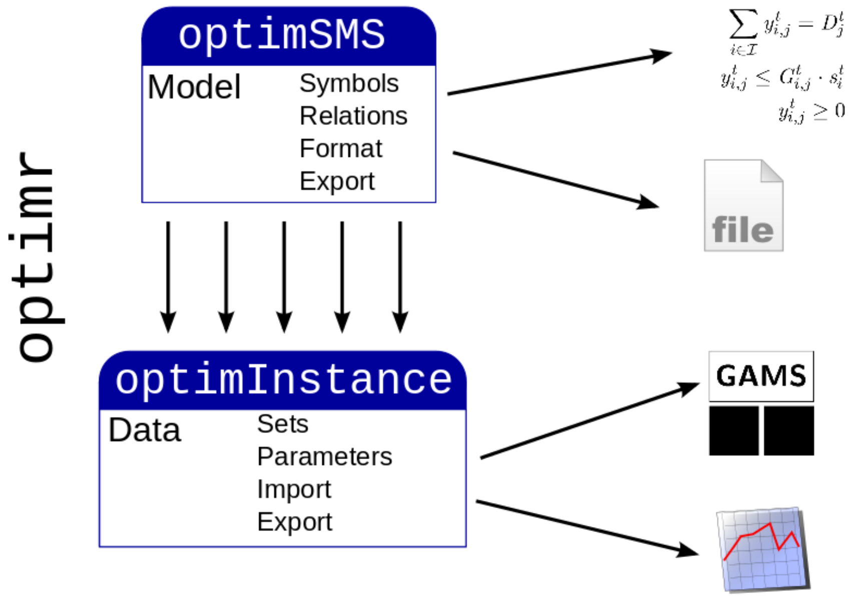

Figure 8: The optimr package structure

The model can be represented in both human and machine readable formats through standard data structures such as R `data.frames`. The `optimSMS` class is composed by the following slots (members):

- Descriptive characters: `name`, `sDes`, and`lDes`;
- Model entities: `consts`, `sets`, `vars`, and `pars` for constants (scalars), sets, decision variables and model parameters respectively.

- Relations: `eqs` and `terms` contain the equations and the terms respectively, using a tree structure.

It also has a bunch of methods to get and represent the SMS, some of the more relevant are:

- To get expressions: `getAliases`, `getConsts`, `getEq`, `getEqs`, `getExpr`, `getModel`, `getMultiSets`, `getPars`, `getSets`, `getSubsets`, `getSymbol`, `getVars`;
- To get R `data.frames`: `SMSconsts`, `SMSeqs`, `SMSpars`, `SMSsets`, `SMSterms`, `SMSvars`;

The creation and addition of elements in a SMS is made through the specific functions `newSMS`, `newSMSconst`, `newSMSeq`, `newSMSpar`, `newSMSset`, and `newSMSvar`. Thus, to create the deterministic model in the previous section we need the following code[9]:

```
model1SMS <- newSMS("Deterministic1",
    "A Basic Case",
    "The simplest model only with electricity")
addItem(model1SMS, "sets") <- list(symbol = "i",
sDes = "Technology", setType = "set")
... ...
addItem(model1SMS, "vars") <- list(
    symbol = "x",
    sDes = "Capacity to be installed",
    units = "kW",
    positive = TRUE,
    ind = as.array(list(c(1,3))))
... ...
addItem(model1SMS, "pars") <- list(
    symbol = "D",
    sDes = "Demand Level",
    units = "kW",
    ind = as.array(list(c(2,3))))
... ...
addItem(model1SMS, "eqs") <- list(
    symbol = "eqDemand",
    sDes = "Production plan for demand",
    relation = "eq",
    nature = "constraint",
    domain = as.array(list(c(2,3))))
addItem(model1SMS, "terms") <- list(
    eq = 4,
    side = "l",
    nature = "vars",
    setSums = as.array(list(c(1))),
    item = 2)
addItem(model1SMS, "terms") <- list(
    eq = 4,
    side = "r",
```

[9]For the sake of space, only one example of each entity is included.

```
    nature = "pars",
    item = 2)
```

Once the SMS is in an `optimSMS` object, any expression can be obtained easily, for example the equation defined above can be obtained in GAMS format as follows:

```
> getEq(object = model1SMS, getid = 4, format = "gams")
[1] "eqDemand(j,t) ..\n\t Sum((i), y(i,j,t)) =e= D(j,t)  \n;\n"
```

Combining different expressions and working with text in R complex representations of the models can be produced.

As for the instance, i.e., the concrete model to be solved using specific data, it is stored in `optimInstance` class objects. An instance is always referred to a model, and therefore to create an `optimInstance` object it is needed an `optimSMS` object. Once created, elements (actual sets, parameter values and equations to include) are added to the instance, related to its SMS[10]:

```
model1Instance1 <- newInstance(model1SMS, name = "model1Instance1")

newInstanceSet(model1Instance1, "i", c("RTE", "PV", "CHP"))
... ...
newInstancePar(model1Instance1, "DT", data.frame(
        j = rep(model1Instance1@sets[["j"]][,2], each = 5),
        t = model1Instance1@sets[["t"]][,2],
        value = 91*8))
... ...
defInstanceEqs(model1Instance1, constEqs = c(3, 4, 5), objEqs = 6)
```

The slots (members) of an instance can be also accessed easily using self-explained functions: `instanceSets`, `instancePars`, and `instanceVars`. Finally, the optimization problem can be written in the appropriate format and be solved as follows:

```
wProblem(model1Instance1,
    filename = "./data/model1Instance1.gms",
    format = "gams",
    solver = "LP")
gams("./data/model1Instance1.gms --outfile=./data/model1Instance1.gdx")
importGams(model1Instance1) <- "./data/model1Instance1.gdx"
```

The last command imports the solution to the `optimInstance` object. Note that at any point advanced data analysis and data visualization can be straightforwardly performed over the data, as they are stored in homogeneous and consistent data structures. For example, Figure 9 and Figure 10 show possible visualization of the model presented above using a common framework.

[10] Again, only an example of each type is printed.

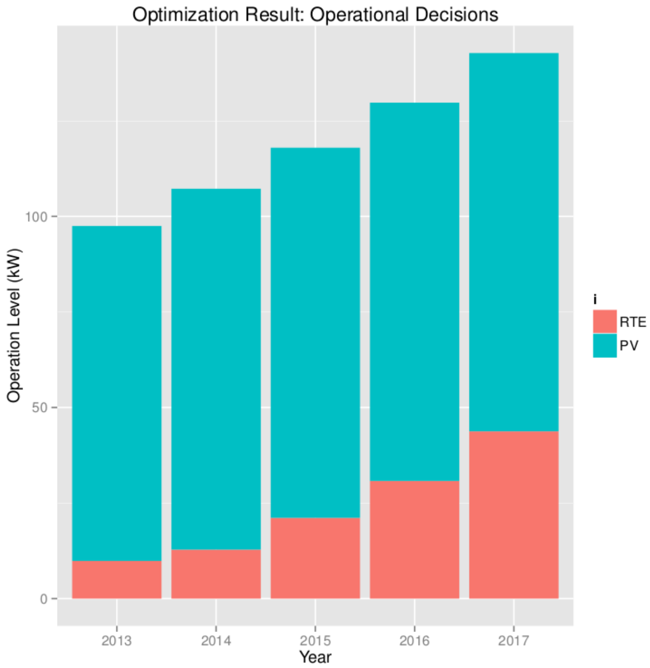

Figure 9: Example of visualization of operational decisions

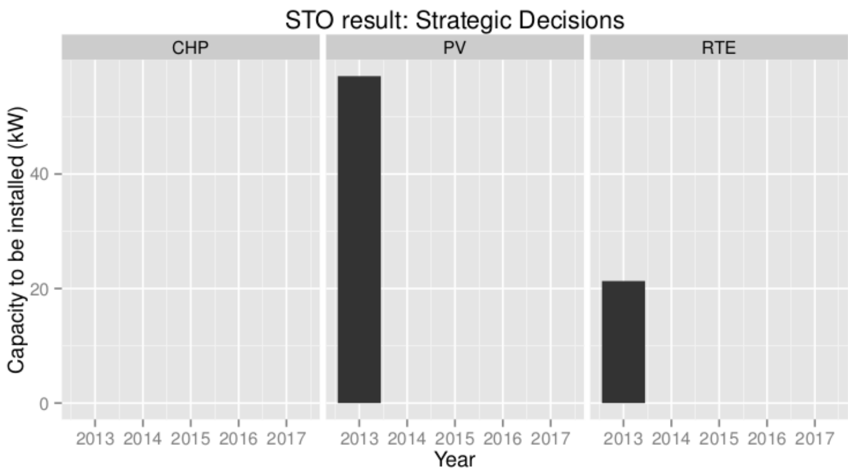

Figure 10: Example of visualization of strategic decisions

6 Concluding remarks

The model and DSS presented in this report have been tested using real data from the EnRiMa project. Results demonstrate the importance of using stochastic strategic models improving the outcomes of deterministic models including hedging against risks. In summary, providing robust solutions for long-term energy supply planning under uncertainty. In particular, using average values, deterministic models provide degenerated solutions violating simplest energy supply security requirements and even being infeasible for all real scenarios.

Decision support is not a static action, but rather an iterative process that requires stakeholders dialog. Moreover, strategic decisions under uncertainty require the application of advanced models that provide robust solutions against all the possible scenarios under security requirements. Applications of inadequate DSS (regarding data treatment, models' structure, analysis of results, etc.) generates serious risks of adopting wrong policies and irreversible developments.

The framework proposed deals with those requisites in a flexible and extensible way. Reproducible research techniques can be applied over different decision problems and environments taking advantage of a common structure and acquired knowledge.

Future work will include the final implementation in EnRiMa, whose DSS now is in prototype version. As far as the R library is concerned, some improvement and optimization over the code must be done before publishing it in public repositories. Currently it is available at the author's personal webpage (`http://www.proyectum.es`). A version of this report is being prepared for submission to a scientific journal.

Further research over these results will be the in-depth analysis of global policies and long-term uncertainty modeling, as well as the benchmarking of the strategic two-stage dynamic model against multi-stage models. Definitely, the proposed idea of learning-by-doing based on the rolling time horizon (Section 3.3) provided a way to escape from irreversible predetermined in advance (at $t = 0$) decisions by using adaptive endogenous scenario generators.

References

Bell, D.; Raiffa, H., and Tversky, A. *Decision Making: Descriptive, Normative, and Prescriptive Interactions*. Cambridge University Press, 1988. ISBN 9780521368513.

Birge, J. R. The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming*, (24):314–325, 1982.

Birge, J. and Louveaux, F. *Introduction to Stochastic Programming*. Springer series in operations research and financial engineering. Springer New York, 2011.

Cai, Y. P.; Huang, G. H.; Yang, Z. F.; Lin, Q. G.; Bass, B., and Tan, Q. Development of an optimization model for energy systems planning in the region of waterloo. *International Journal of Energy Research*, 32(11):988–1005, 2008. doi: 10.1002/er.1407.

Cano, E. L.; Moguerza, J. M., and Redchuk, A. *Six Sigma with R. Statistical Engineering for Process Improvement*, volume 36 of *Use R!* Springer, New York, 2012. ISBN 978-1-4614-3651-5. URL http://www.springer.com/statistics/book/978-1-4614-3651-5.

CBS. Deception at Duke. TV Report, February 2012. URL http://www.cbsnews.com/video/watch/?id=7398476n. [Retrieved 20120626].

Conejo, A. J.; Carrión, M., and García-Bertrand, R. Medium-term electricity trading strategies for producers, consumers and retailers. *International Journal of Electronic Business Management*, 5(3):239–252, 2007.

Delage, E.; Arroyo, S., and Ye, Y. The value of stochastic modeling in two-stage stochastic programs with cost uncertainty. Technical Report G-2012-05, GERAD (HEC Montreal), 2012.

El-Khattam, W.; Bhattacharya, K.; Hegazy, Y., and Salama, M. Optimal investment planning for distributed generation in a competitive electricity market. *Power Systems, IEEE Transactions on*, 19(3):1674–1684, aug. 2004.

Ermoliev, Y. and Wets, R., editors. *Numerical techniques for stochastic optimization*. Springer series in Computational Mathematics. Springer-Verlag, 1988.

Ermoliev, Y.; Makowski, M., and Marti, K., editors. *Managing Safety of Heterogeneous Systems: Decisions under Uncertainties and Risks*. Lecture notes in economics and mathematical systems. Springer, 2012. ISBN 9783642228841.

Fourer, R. Modeling languages versus matrix generators for linear programming. *ACM Trans. Math. Softw.*, 9(2):143–183, 1983.

French, S.; Maule, J., and Papamichail, N. *Decision Behaviour, Analysis and Support*. Cambridge University Press, 2009. ISBN 9780521883344.

Geoffrion, A. M. The SML language for structured modeling: Levels 1 and 2. *Operations Research*, 40(1):38–57, 1992a.

Geoffrion, A. M. The SML language for structured modeling: Levels 3 and 4. *Operations Research*, 40(1):58–75, 1992b.

González, J. R.; Pelta, D. A., and Masegosa, A. D. A framework for developing optimization-based decision support systems. *Expert Systems with Applications*, 36(3, Part 1):4581 – 4588, 2009.

Gritsevskii, A. and Ermoliev, Y. *Managing Safety of Heterogeneous Systems*, chapter Modeling technological change under increasing returns and uncertainty, pages 109–136. Springer-Verlag, Heidelberg, Germany, February 2012.

Gritsevskii, A. and Nakicenovic, N. Modeling uncertainty of induced technological change. *Energy Policy*, 28(13):907–921, 2000.

Groissböck, M.; López, E.; Perea, E.; Siddiqui, A., and Werner, A. Improving energy efficiency and risk management in EU public buildings. *IAEE Energy Forum*, 22:17–20, 2013. URL http://www.iaee.org/en/publications/newsletterdl.aspx?id=193.

Hardesty, L. US favors Net-Metering while Europe, Japan like feed-in-tariffs. online [accessed 2013-07-25], May 2013. URL http://goo.gl/xCNiAf.

Hernandez, P. and Kenny, P. From net energy to zero energy buildings: Defining life cycle zero energy buildings (LC-ZEB). *Energy and Buildings*, 42(6):815 – 821, 2010.

Heydari, S. and Siddiqui, A. Valuing a gas-fired power plant: A comparison of ordinary linear models, regime-switching approaches, and models with stochastic volatility. *Energy Economics*, 32(3):709 – 725, 2010.

Hobbs, B. Optimization methods for electric utility resource planning. *European Journal of Operational Research*, 83(1):1–20, 1995.

Jamasb, T. and Pollitt, M. Electricity market reform in the european union: Review of progress toward liberalization & integration. *The Energy Journal*, 0(Special I):11–42, 2005. URL http://ideas.repec.org/a/aen/journl/2005se-a02.html.

Jelliffe, R. Climate wars: Global warming, climategate, web 2.0 and grey power. Blog post, March 2010. URL http://broadcast.oreilly.com/2010/03/climate-wars-global-warming-cl.html. [retrieved: 2013/01/31].

Kallrath, J., editor. *Algebraic Modeling Systems: Modeling and Solving Real World Optimization Problems*. Applied Optimization. Springer Berlin Heidelberg, 2012a.

Kallrath, J. A practioner's wish list towards algebraic modeling systems. In Kallrath, J., editor, *Algebraic Modeling Systems*, volume 104 of *Applied Optimization*, pages 213–222. Springer Berlin Heidelberg, 2012b.

Kallrath, J. Algebraic modeling languages: Introduction and overview. In Kallrath, J., editor, *Algebraic Modeling Systems*, volume 104 of *Applied Optimization*, pages 3–10. Springer Berlin Heidelberg, 2012c.

Kaut, M.; Midthun, K.; Werner, A.; Tomasgard, A.; Hellemo, L., and Fodstad, M. Multi-horizon stochastic programming. *Computational Management Science*, 2013. doi: 10.1007/s10287-013-0182-6.

King, D. and Morgan, M. Adaptive-focused assessment of electric power microgrids. *Journal of Energy Engineering*, 133(3):150–164, 2007.

Klein, G.; Orasanu, J., and Calderwood, R. *Decision Making in Action: Models and Methods*. Cognition and Literacy. Ablex Publishing Corporation, 1993. ISBN 9780893919436.

Kuip, C. Algebraic languages for mathematical programming. *European Journal of Operational Research*, 67(1):25 – 51, 1993.

Kumbaroğlu, G. and Madlener, R. Evaluation of economically optimal retrofit investment options for energy savings in buildings. Working paper 14/2011, Institute for Future Energy Consumer Needs and Behavior (FCN), Aachen, September 2011.

Leisch, F. Sweave: Dynamic generation of statistical reports using literate data analysis. In Härdle, W. and Rönz, B., editors, *Compstat 2002 — Proceedings in Computational Statistics*, pages 575–580. Physica Verlag, Heidelberg, 2002. URL http://www.stat.uni-muenchen.de/~leisch/Sweave.

Madansky, A. Inequalities for stochastic linear programming problems. *Management Science*, (6):197–204, 1960.

Marnay, C.; Venkatarmanan, G.; Stadler, M.; Siddiqui, A.; Firestone, R., and Chandran, B. Optimal technology selection and operation of commercial-building microgrids. *IEEE Transactions on Power Systems*, 23(3):975–982, 2008.

Pless, S. and Torcellini, P. Net-zero energy buildings: A classification system based on renewable energy supply options. Technical Report NREL/TP-550-44586, National Renewable Energy Laboratory (NREL), Golden, CO, USA, June 2010. URL http://www.nrel.gov/sustainable_nrel/pdfs/44586.pdf.

Power, D. A brief history of decision support systems. DSSResources.COM, World Wide Web, 2007. URL http://DSSResources.COM/history/dsshistory.html. version 4.0 [accessed 2014-01-31].

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/. [software version 3.0.2].

Salewicz, K. A. and Nakayama, M. Development of a web-based decision support system (DSS) for managing large international rivers. *Global Environmental Change*, 14, Supplement(0):25 – 37, 2004.

Salvador, M. and Grieu, S. Methodology for the design of energy production and storage systems in buildings: Minimization of the energy impact on the electricity grid. *Energy and Buildings*, 47(0):659 – 673, 2012.

Shim, J.; Warkentin, M.; Courtney, J. F.; Power, D. J.; Sharda, R., and Carlsson, C. Past, present, and future of decision support technology. *Decision Support Systems*, 33(2): 111 – 126, 2002.

Siddiqui, A.; Marnay, C.; Edwards, J.; Firestone, R.; Ghosh, S., and Stadler, M. Effects of carbon tax on combined heat and power adoption. *Journal of Energy Engineering*, 131(1):2–25, 2005.

Stadler, M.; Siddiqui, A. S.; Marnay, C.; Aki, H., and Lai, J. Control of greenhouse gas emissions by optimal DER technology investment and energy management in zero-net-energy buildings. *European Transactions on Electrical Power*, 21(2):27, 2009.

Stodden, V.; Leisch, F., and Peng, R. D., editors. *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2013.

Tanaka, M.; Watanabe, H.; Furukawa, Y., and Tanino, T. GA-based decision support system for multicriteria optimization. In *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, volume 2, pages 1556–1561 vol.2, 1995.

The New York Times. How Bright Promise in Cancer Testing Fell Apart. Newspaper, 7 2011. URL http://www.nytimes.com/2011/07/08/health/research/08genes.html. [retrieved 2014-01-31].

Theussl, S. and Hornik, K. *Rglpk: R/GNU Linear Programming Kit Interface*, 2013. URL http://CRAN.R-project.org/package=Rglpk. R package version 0.5-2.

Van Sambeek, E. Distributed generation in competitive electricity markets. Working Paper no. 00-S4, 2000. Center for Energy and Environmental Policy, Newark, DE, USA.

Villumsen, J. and Philpott, A. Investment in electricity networks with transmission switching. *European Journal of Operational Research*, 222(2):377 – 385, 2012.

Weinberg, C.; Iannucci, J., and Reading, M. The distributed utility: Technology, customer, and public policy changes shaping the electrical utility of tomorrow. *Energy Systems Policy*, 15(4):307–322, 1991.