



Universidad
Rey Juan Carlos

Dpto. de Estadística e
Investigación Operativa

Tesis Doctoral

**An Integrated Framework
for the Representation
and Solution of
Stochastic Energy
Optimization Problems**

Emilio López Cano

Directores
Javier Martínez Moguerza
Yurii Yermoliev

Febrero 2014



Don **Javier Martínez Moguerza**, Profesor Titular de Universidad del Departamento de Estadística e Investigación Operativa de la Universidad Rey Juan Carlos y Don **Yurii Yermoliev**, *Institute Scholar* en International Institute for Advanced Systems Analysis (IIASA), Laxenburg, Austria.

AUTORIZAN:

La presentación de la Tesis Doctoral titulada

**AN INTEGRATED FRAMEWORK FOR THE
REPRESENTATION AND SOLUTION OF STOCHASTIC
ENERGY OPTIMISATION PROBLEMS**

Realizada por Don Emilio López Cano bajo su inmediata dirección y supervisión y que presenta para la obtención del grado de Doctor por la Universidad Rey Juan Carlos.

En Móstoles, a 28 de Febrero de 2014.

Fdo.: D. Javier Martínez Moguerza

Fdo.: D. Yurii Yermoliev



Don **Antonio Alonso Ayuso**, Profesor Titular de Universidad y Director del Departamento de Estadística e Investigación Operativa de la Universidad Rey Juan Carlos.

INFORMA:

Que la presente Tesis Doctoral titulada

**AN INTEGRATED FRAMEWORK FOR THE
REPRESENTATION AND SOLUTION OF STOCHASTIC
ENERGY OPTIMISATION PROBLEMS**

Ha sido realizada por Don Emilio López Cano bajo la dirección y supervisión de Don Javier Martínez Moguerza y Don Yurii Yermoliev, y que el Departamento de Estadística e Investigación Operativa ha dado su conformidad para que sea presentada ante la Comisión de Doctorado.

En Móstoles, a 28 de Febrero de 2014.

Fdo.: D. Antonio Alonso Ayuso

An Integrated Framework for the Representation and Solution of Stochastic Energy Optimization Problems

Doctoral Thesis

Emilio López Cano

Supervisors:

Javier M. Moguerza and Yurii Yermoliev

February 2014

Resumen

Antecedentes

La importancia de la optimización de sistemas energéticos es creciente debido a la liberalización del sector energético, así como del establecimiento de objetivos como el 20-20-20 de la Unión Europea, que a su vez desembocan en nuevas regulaciones. Esta situación lleva a nuevos modelos estocásticos dinámicos incluyendo decisiones estratégicas (en el largo plazo) que deben ser tomadas teniendo en cuenta el desempeño de los sistemas y las decisiones operacionales (en el corto plazo). Estas decisiones incluyen tanto inversión en tecnologías como opciones de mercado y financieras. Así, los propietarios y gestores de edificios tienen el reto de tomar decisiones para obtener el conjunto óptimo de tecnologías y contratos, y adoptar un papel activo en los mercados energéticos. Además, estas decisiones se tienen que tomar bajo condiciones inherentes de incertidumbre.

Los procesos de toma de decisiones como el abordado en esta tesis, a menudo se realizan bajo condiciones de incertidumbre. En estos procesos de decisión, se buscará lógicamente la mejor decisión posible, es decir, se tratará de optimizar un valor determinado, por ejemplo minimizar el coste total para un determinado sistema. Estos problemas de optimización se pueden formular mediante modelos de programación matemática, donde una serie de parámetros son fijos, y ciertas variables de decisión son desconocidas. Dichos parámetros fijos suelen estar sujetos a incertidumbre, es decir, no se conoce su valor exacto. Una forma de tratar esta incertidumbre es estimar los parámetros y resolver el problema para los valores medios. Esta forma de resolver el problema resulta en la mejor solución para el escenario promedio. Pero la mayoría de las veces, esta solución no es óptima para el valor real de los parámetros una vez se desvela la incertidumbre, especialmente en modelos estratégicos a largo plazo. Incluso esa solución óptima para los valores promedio puede ser realmente mala para el escenario real, o peor aún, la solución puede ser no factible para el escenario que finalmente tiene lugar, incurriendo así en determinados riesgos. Téngase en cuenta además que

el escenario promedio normalmente nunca ocurre. Una forma más efectiva de tratar la incertidumbre en problemas de optimización es la Programación Estocástica. Mediante Programación Estocástica se obtiene la solución óptima para el problema de optimización teniendo en cuenta la distribución de probabilidad de los parámetros estocásticos, es decir, sujetos a incertidumbre, y no sólo sus valores medios sino también su variabilidad. De hecho, la solución no es óptima para ningún escenario en particular, sino una solución robusta frente a todos los escenarios posibles. Esto es especialmente importante en modelos estratégicos a largo plazo en general, y en planificación energética en particular. En esta tesis se desarrollan modelos novedosos de optimización estocástica capaces de tratar la incertidumbre a largo plazo.

Para abordar el proceso de toma de decisiones descrito, el uso de Sistemas de Ayuda a la Decisión (SAD) es inevitable. Esta tesis presenta una visión de los SAD más allá del puro programa informático, dando especial relevancia al modelo. El modelo es la base para la toma de decisiones basada en el Análisis, y por tanto en la evidencia científica. Otro aspecto importante del marco de trabajo propuesto es la importancia del diálogo entre *stakeholders* (actores involucrados), en el que el modelo vuelve a tener un papel importante, pero también otros aspectos que facilitan este diálogo, como la investigación reproducible. Así, dentro del marco de trabajo se integran el modelo mediante el llamado SMS (Symbolic Model Specification, especificación del modelo simbólico); los datos, mediante la definición de clases y estructuras lógicas; y el diálogo entre *stakeholders* mediante interfaces adecuados. En esta tesis se propone un marco de trabajo integrado capaz de tratar el problema de decisión abordado, y se realiza una implementación del mismo con los modelos desarrollados y con datos de edificios reales. La figura 2.5 resume el marco de trabajo propuesto.

Objetivos

La presente tesis se propone contestar a las siguientes preguntas:

- Cómo modelar los flujos de energía conjuntamente con estrategias de planificación;
- Cómo modelar la incertidumbre para conseguir modelos robustos de toma de decisiones en planificación de sistemas energéticos;
- Cómo gestionar el riesgo en estos modelos;
- Cómo debe ser un SAD efectivo para tratar con modelos, datos y soluciones;

Por tanto, los principales objetivos de la tesis son, por una parte el desarrollo de modelos de optimización estocástica capaces de tratar efectivamente la incertidumbre inherente al proceso de decisión, los flujos de energía y el riesgo, y por otra el desarrollo de un marco de trabajo para SAD que favorezca el necesario diálogo entre *stakeholders*.

Metodología

Para el desarrollo de los modelos se ha seguido una estrategia secuencial, escalando desde modelos sencillos deterministas hasta modelos estocásticos multi-etápicos bajo una visión sistémica. Así, un primer modelo determinista con decisiones operacionales y estratégicas y restricciones técnicas presenta el problema de forma fácilmente comprensible. La función objetivo será el coste total a lo largo del horizonte de decisión, la cual se pretende minimizar. Este primer modelo, que se encuentra en el Capítulo 3, se ilustra con un primer ejemplo numérico. En el Capítulo 4 el modelo determinista se extiende a la versión estocástica en dos etapas. Manteniendo la formulación sencilla, con pocas variables y parámetros, el modelo bi-etápico es desarrollado proporcionando las primeras pinceladas del valor de estos modelos a través del VSS (*Value of Stochastic Solution*, valor de la solución estocástica) y de la gestión del riesgo a través del CVaR (*Conditional Value at Risk*, valor a riesgo condicionado). El ejemplo determinista es igualmente extendido a su versión estocástica, mostrando la utilidad de las soluciones obtenidas.

En el siguiente paso se vuelve al modelo determinista para desarrollarlo y ampliarlo incluyendo características como la obsolescencia de las tecnologías, selección de contratos, emisiones o eficiencia, así como el balance de energía de forma global en el edificio. Este modelo determinista evolucionado se encuentra en el Capítulo 5. A continuación, el modelo con todas las características se extiende de nuevo a su versión estocástica en el Capítulo 6. En este caso el modelo finalmente utilizado es multi-etápico, resultando en una formulación compacta por nodos, en vez de aquella por escenarios que fue utilizada en el modelo bi-etápico. Estos nuevos modelos son también ilustrados mediante ejemplos numéricos. Finalmente, en el Capítulo 7 se añade al modelo multi-etápico la gestión del riesgo a través del CVaR, incluyendo en la función objetivo un término de riesgo que puede ser modelado de acuerdo a las preferencias del decisor: neutro al riesgo, aversión al riesgo, etc. Además, esta formulación final permite seleccionar el objetivo entre tres posibilidades: minimizar el coste, las emisiones, o el consumo de energía, expandiendo así el uso de CVaR más allá de la clásica selección de carteras óptimas, y aportando la posibilidad de incluir como objetivo la minimización del riesgo de altas emisiones o alto consumo energético.

En cuanto al marco de trabajo para SAD, éste se desarrolla bajo el paradigma de la investigación reproducible en el Capítulo 8 con la vista puesta en la principal misión del SAD: el diálogo entre *stakeholders*. Se muestran estructuras de datos basadas en técnicas de Programación Orientada a Objetos, conteniendo tanto el modelo como los datos del problema concreto o instancia. Se proporcionan directrices para la implementación del marco de trabajo, y se realiza una implementación del mismo utilizando el software estadístico y lenguaje de programación R y el software de modelización y optimización GAMS. A tal efecto se ha desarrollado una librería llamada `optimr`. Un ejemplo integral ilustra todas las fases descritas en la tesis.

Conclusiones

La investigación contenida en esta tesis se ha llevado a cabo en el contexto del proyecto EnRiMa (Energy Efficiency and Risk Management in Public Buildings, eficiencia energética y gestión del riesgo en edificios públicos) financiado por el Séptimo Programa Marco de la Unión Europea. Los resultados de esta investigación, tanto los modelos como el marco de trabajo y su filosofía, han sido aplicados con éxito en dicho proyecto y sus edificios de prueba. Las principales aportaciones de esta tesis son:

- La modelización de flujos de energía a nivel de edificio de forma global, conjuntamente con las características y evolución del equipamiento en el edificio tales como obsolescencia, o renovación, en un mismo modelo;
- Tanto los modelos deterministas como los estocásticos han sido probados en edificios reales obteniendo resultados coherentes con el desempeño real de los mismos;
- La modelización de la incertidumbre ha demostrado la utilidad de la programación estocástica en optimización de sistemas energéticos y cómo modelos deterministas nos llevan a escenarios no factibles;
- Se ha propuesto una nueva aplicación del gestión del riesgo más allá de riesgos económicos, considerando riesgos ambientales y sociales;
- La aplicación de métodos de investigación reproducible en implementaciones reales de optimización es también novedosa;
- Este enfoque de investigación reproducible junto con el diálogo entre *stakeholders* proporciona a los SAD que adopten el marco de trabajo propuesto grandes posibilidades de éxito;
- La librería desarrollada es un resultado en sí mismo, aunque el marco de trabajo puede implementarse utilizando otras tecnologías;

- El ejemplo integral demuestra un ciclo completo de investigación reproducible haciendo uso de los modelos, los datos y el marco de trabajo desarrollados en esta tesis.

Futuras líneas de trabajo del autor incluirán las siguientes:

- Estudio e implementación de métodos de generación de escenarios, especialmente para determinar distribuciones de probabilidad a largo plazo y eventos extremos;
- La comparación de los modelos bi-etápicas y multi-etápicas desde distintos puntos de vista: eficiencia computacional, mejora obtenida con modelos más complejos frente a otros más sencillos, o validación de las asunciones realizadas en la generación de escenarios;
- En cuanto al marco de trabajo y la librería desarrollada, la extensión a otros formatos más allá de los utilizados en la primera implementación (L^AT_EX y GAMS), la mejora del código y la documentación, así como la publicación en repositorios públicos, son tareas a abordar en el futuro.

To my wife and daughters

Abstract

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 optimum portfolio and they are encouraged to adopt an active role in energy markets. Moreover, those decisions must be made under inherently uncertain conditions. Scaling from simple deterministic models to complex stochastic models under a systemic view, an integrated framework for the representation and solution of such energy systems optimization problems is proposed. Such models are to be implemented in Decision Support Systems (DSSs) for robust decision making at the building level, as well as to assess current and potential policies to face rising economic and environmental global challenges. As the combination of operational and strategic decisions in the same model induces risk aversion in strategic decisions, it is needed to include quantile-based measures such as Conditional Value at Risk (CVaR). Furthermore, two different approaches are developed: a two-stage model where first-stage decisions are the strategic, long-term decisions and a multi-stage model where first-stage decisions are both operational and strategic decision for the first long-term period.

Such complex energy systems need to be accurately described in a condensed way representing the huge amount of variables, parameters and constraints. Thus, a systemic view on all interactions of energy sub-systems, their interdependencies, possible systemic risks, which may propagate through the system, magnify and cause system's collapse, is followed. A comprehensive Symbolic Model Specification (SMS) development is part of the research work, encompassing a number of features covering all the energy types used in a building, as well as investments, renovation, technologies' ob-

solescence, emissions, efficiency, and even operations in a unified model. An integrated framework is proposed to cover the needs of the whole decision making problem, ranging from data analysis and estimation to effective representation of models and decisions to be consumed for both humans and machines. Furthermore, such a framework shall allow to communicate with different types of optimization software. The proposed framework is suitable to be implemented in DSSs. An implementation of such framework is also part of this thesis, entailing a novel approach based on the *stakeholders dialog* and Reproducible Research philosophy.

The research has 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. The results of this work have been applied to real test sites within the project.

Acknowledgements

Great thanks to my supervisor Javier M. Moguerza who relied on me from the very beginning. I am also grateful with the rest of colleagues from the Department of Statistics and Operations Research at Rey Juan Carlos University, especially with David Ríos Insua. Special thanks to Yurii Yermoliev and Tatiana Ermolieva, who supervised my work at the IIASA Young Scientists Summer Program (YSSP 2013), where I performed part of my research at the Advanced Systems Analysis (ASA) program. I acknowledge also the partners of the EnRima project (EC FP7 project number 260041): Stockholm University (SU); University College London (UCL); International Institute for Advanced Systems Analysis (IIASA); Center for Energy and Innovative Technologies (CET); Minerva Consulting and Communication (MC&C); SINTEF Group; Tecnalia Research and Innovation; Hidrocantábrico Energía (HCE). Part of this thesis is based in the work developed within that project, and all the team members were kind and valuable colleagues. I would like to notably recognize the great effort done by the project coordinator, Afzal Siddiqui. Last but not least, thanks to Ángela, Lucía and Sonia, who had to put up with my commitments.

Contents

List of Acronyms	xxi
List of Tables	xxiii
List of Figures	xxvi
I. Basics	1
1. Introduction	3
1.1. Global Changes, Local Challenges	3
1.1.1. Overview	3
1.1.2. Relevant Policies	4
1.2. Energy Systems	7
1.2.1. Scope	7
1.2.2. Energy Systems Dynamics	8
1.2.3. Strategic and Operational Decisions	9
1.3. Decision Making Under Uncertainty	11
1.3.1. The Role of Uncertainty	11
1.3.2. Decision Support Systems (DSSs)	12
1.4. Dissertation Outline	14
1.4.1. Outlook	14
1.4.2. A Baseline Example	15
1.4.3. Main Questions Addressed	17
1.4.4. Conventions and Structure	18
2. Background	21
2.1. The Energy Efficiency and Risk Management in Public Buildings (EnRiMa) Project	21

Contents

- 2.2. Methodology 26
 - 2.2.1. Operations Research (OR) 26
 - 2.2.2. Mathematical Programming 27
 - 2.2.3. Stochastic Optimization (STO) 30
 - 2.2.4. Uncertainty Modeling 35
- 2.3. Decision Support Systems (DSSs) for Optimization . 41
 - 2.3.1. General Framework 41
 - 2.3.2. The Model 44
 - 2.3.3. The Data 46
 - 2.3.4. The System 47
- 2.4. State of the Art 49

II. Modeling 53

- 3. The Basic Problem 55**
 - 3.1. Introduction 55
 - 3.1.1. Scope 55
 - 3.1.2. Model Overview 55
 - 3.2. Building the Model 56
 - 3.2.1. The Simplest Case 56
 - 3.2.2. Adding Features 58
 - 3.3. Numerical Example 61
- 4. Adding Uncertainty: Two-stage Problems 67**
 - 4.1. Introduction 67
 - 4.2. A Basic Two-stage Dynamic Model 68
 - 4.3. Numerical Methods 72
 - 4.4. Value of Stochastic Solution 75
- 5. A Comprehensive Deterministic Problem 81**
 - 5.1. Introduction 81
 - 5.2. Time Modeling 83
 - 5.3. Technologies Modeling 84
 - 5.4. Strategic modeling 86
 - 5.5. Operational Constraints 89

5.6. Objective Function	92
5.7. Numerical Example	94
5.8. Model Evolution	98
6. Stochastic Multi-stage Model	107
6.1. Evolution of the Symbolic Model Specification (SMS)	107
6.2. Improvements in Systems Modeling	108
6.3. Scenario Tree Representation	111
6.4. Symbolic Model Specification	113
6.4.1. Sets	114
6.4.2. Constants	118
6.4.3. Parameters	118
6.4.4. Variables	123
6.4.5. Equations	127
6.5. Numerical Example	139
6.5.1. Scenario Tree	140
6.5.2. Sets	140
6.5.3. Parameter Values	141
6.5.4. Solution	142
7. Risk Management	151
7.1. Introduction	151
7.2. Risk Measures	151
7.3. Conditional Value at Risk (CVaR)	153
7.4. CVaR in the Two-stage Model	154
7.5. Risk Measures in the Multi-stage Model	156
III. Implementation and Outcomes	159
8. An Integrated Framework	161
8.1. Introduction	161
8.2. A Reproducible Research Approach	164
8.3. The <code>optimr</code> R Package	167
8.3.1. Overview	167
8.3.2. Details	169

Contents

8.4. Illustrative Example	171
8.4.1. The Model	171
8.4.2. The Instance	173
8.4.3. The Solution	174
8.4.4. Extending the Instance	174
8.4.5. Comprehensive Reproducible Example	177
9. Conclusions	187
9.1. Results	187
9.2. Discussion	189
9.2.1. Stochastic Energy Systems Planning	189
9.2.2. Deterministic Modeling	190
9.2.3. Stochastic Multi-stage Modeling	190
9.2.4. A Framework for DSS	191
9.3. Further Research	193
References	195
Appendices	213
A. Glossary	213
B. Code for example in Chapter 8	217
B.1. File <code>comprehensiveExample.Rnw</code>	217
B.2. File <code>model1SMS.R</code>	221
B.3. File <code>model2SMS.R</code>	226
B.4. File <code>model2Instance2.R</code>	228
B.5. File <code>comprehensiveExample.Rnw</code>	231
C. Session Info	237

List of Acronyms

- AIMMS** Advanced Interactive Multidimensional Modeling System
- AML** Algebraic Modeling Language
- AMPL** A Mathematical Programming Language
- CET** Center for Energy and innovative Technologies
- CHP** Combined Heat and Power
- CMPL** COIN Mathematical Programming Language
- COIN-OR** Computational Infrastructure for Operations Research
- Coopr** A Common Optimization Python Repository
- CVaR** Conditional Value at Risk
- DEP** Deterministic Equivalent Program
- DER** Distributed Energy Resources
- DG** Distributed Generation
- DHW** Domestic Hot Water
- DSS** Decision Support System
- DV** Decision Variable
- EC** European Commission
- EnRiMa** Energy Efficiency and Risk Management in Public Buildings
- EU** European Union
- EVPI** Expected Value of Perfect Information
- FASAD** Fundación asturiana de atención y protección a personas con discapacidades y/o dependencias

Contents

- FP7** 7th Framework Program
- GAMS** General Algebraic Modeling System
- GLPK** GNU Linear Programming Kit
- GUI** Graphical User Interface
- HCE** Hidrocarburo Energía
- ICT** Information and Communication Technologies
- IIASA** International Institute for Advanced Systems Analysis
- IP** Integer (Linear) Programming
- INFORMS** Institute for Operations Research and the Management Sciences
- LP** Linear Programming
- FMATHL** Formal Mathematical Language
- MCC** Minerva Consulting and Communication
- MIP** Mixed Integer Programming
- MPS** Mathematical Programming System
- NG** Natural Gas
- NLP** Non Linear Programming
- NZEB** Net-Zero Energy Building
- OM** Operation and Maintenance
- OOP** Object Oriented Programming
- OR** Operations Research
- OS** Optimization Services
- Pyomo** Python Optimization Modeling Objects

- PV** Photovoltaic
- rhs** right hand side
- SINTEF** Stiftelsen for Industriell og Teknisk Forskning
- SP** Stochastic Programming
- s.t.** subject to
- STO** Stochastic Optimization
- SMS** Symbolic Model Specification
- SU** Stockholms Universitet
- TECNALIA** Tecnalía Research & Innovation
- TEP** Transmission Expansion Planning
- ToU** Time of Use
- UCL** University College London
- URJC** Universidad Rey Juan Carlos
- VaR** Value at Risk
- VSM** Value of Stochastic Modeling
- VSS** Value of Stochastic Solution
- WG** Wind Generator
- XML** eXtended Markup Language

List of Tables

3.1.	Electricity demand level (kW)	63
3.2.	PV technology availability (ratio)	65
3.3.	Total technologies investment (EUR)	65
4.1.	Annual incremental ratios for each scenario	74
4.2.	Objective function assessment for each scenario (EUR).	74
4.3.	Objective value for the models described.	79
5.1.	Conversion factors for energy-generation technologies.	99
5.2.	Example: Availability factor for WG	101
5.3.	Example: Availability factor for PV	102
5.4.	Example: technologies installation	102
5.5.	Example: Technologies decommissioning	102
6.1.	\mathcal{N}_{Pur}^k multidimensional set	143
6.2.	Energy input/output multidimensional sets	143
6.3.	\mathcal{V}_{Pa}^v multidimensional set	143
6.4.	$\mathcal{A}_{Ages}^{i,v}$ multidimensional set	144
6.5.	\mathcal{T}_{Tm}^m multidimensional set	144
6.6.	PT^v parameter	144
6.7.	PR^v parameter	145
6.8.	CI_i^v parameter	145
6.9.	SU_i^v parameter	145
6.10.	xi_i^v variable	148
8.1.	Base parameter values an uncertain evolution	176
9.1.	Optimization results for the complete risk model.	188

List of Figures

1.1. Sankey diagram	10
1.2. DSS diagram	14
2.1. EnRiMa DSS functionality	23
2.2. EnRiMa Solver Manager structure	24
2.3. EnRiMa DSS Architecture	24
2.4. EnRiMa project DSS Schema	26
2.5. DSS updated diagram	43
3.1. Temporal resolutions of the strategic planning model.	57
3.2. Energy dispatch for Example 4	64
4.1. Strategic decisions in Example 6	75
5.1. EnRiMa project DSS Schema.	82
5.2. Deterministic example building systems	100
5.3. Deterministic example: energy demand	100
5.4. Deterministic example: Energy prices	101
5.5. Example: technologies capacity	103
5.6. Deterministic example: technologies deployment chart	103
5.7. Example: energy purchasing decisions	104
5.8. Example: energy sales decisions	104
5.9. Example: energy input decisions	105
5.10. Example: energy output decisions	105
6.1. Scenario tree symbolic representation.	113
6.2. Scenario tree multi-stage example	140
6.3. $AF_{PV}^{v,m,t}$ parameter by scenario	146
6.4. $D_k^{v,m,t}$ parameter by node	146

List of Figures

6.5.	$PP_{k,n}^{p,m,t}$ parameter by node	147
6.6.	x_i^v variable	148
6.7.	xc_i^v variable	149
6.8.	$u_{k,n}^{v,m,t}$ variable	149
6.9.	$z_{i,k}^{v,m,t}$ variable	150
8.1.	Elements in an optimization decision making process.	162
8.2.	The optimr package structure	167
8.3.	Example: visualization of operational decisions . . .	175
8.4.	Example: Visualization of strategic decisions	175
8.5.	Example: Integrated graphics	178
8.6.	Page 1 of the report generated through the example.	184
8.7.	Page 2 of the report generated through the example.	185
8.8.	Page 3 of the report generated through the example.	186
9.1.	Example: energy balance in the test site	189

Part I.

Basics

1. Introduction

1.1. Global Changes, Local Challenges

1.1.1. Overview

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 that 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 those global changes must be tackled at a regional or local scale. Thus, utilities and fuel producers, yet global, must fulfill localized 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

1. Introduction

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 tariffs alternatives, as well as new technologies and refurbishment options are available and continuously evolving, widening the range of choices for decision makers.

1.1.2. Relevant Policies

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

- 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 EU **Energy Efficiency Plan 2011**² was adopted by the EC for saving more energy through concrete measures. It

¹http://ec.europa.eu/clima/policies/package/index_en.htm

²http://ec.europa.eu/energy/efficiency/action_plan/action_plan_en.htm

1.1. Global Changes, Local Challenges

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, whose aim is to achieve buildings with zero net energy consumption and zero carbon emissions annually. Hernandez and Kenny (2010) go beyond this concept from an economical ecologies perspective and introduce new concepts. A classification and description of NZEBs 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

³http://ec.europa.eu/energy/efficiency/eed/eed_en.htm

1. Introduction

grid. According to Hardesty (2013), US favors net metering while Europe and Japan feed-in tariffs.

1.2. Energy Systems

1.2.1. Scope

Energy is a wide topic which can be tackled from different scientific and technical disciplines, from biology to engineering. Energy systems are conceived in this work as the technologies and devices used to provide people with the energy needed for their everyday activities. From this standpoint, different types of energy systems can be found, namely:

- Appliances: the final elements of a net that consumes a given amount of energy to produce a given *service* to a user. For example, a bulb consumes electricity and provide light to a room in a household;
- Nets: The wires and passing devices that transport the energy throughout the whole system;
- Generation and transformation technologies. Energy can not be created, however *generation* is an accepted term when talking about energy transformation from one type to another. For example Photovoltaic (PV) panels *generate* solar irradiation into electricity; Combined Heat and Power (CHP) units transform Natural Gas (NG) into electricity and heat. Sometimes these technologies do not change the type of energy but are considered, for example, to reflect energy losses;
- Storage technologies. Some types of energy can be stored in storing devices such as batteries or reservoirs;
- Passive technologies. There are different types of technologies whose target is not to reduce the level of service received, but to mitigate the energy needs to achieve it. For example, demand response technologies allow to shift the moment in which the energy is consumed in order to avoid peak periods; Other innovations could be refurbishment or renovation.

1. Introduction

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. Examples of buildings under this conception are university campuses, sports centers, administrative buildings, hospitals, or airports. Therefore, primary energy generation by utility companies and distribution networks are not considered. Likewise, household buildings such as apartment blocks and isolated houses are not yet as mature as needed to adopt the models and approaches presented hereby. Nevertheless, the building sector can benefit of them when planning new buildings as, in spite of the focusing on large energy consumers, the models could also be applied at a domestic level. In summary, the target buildings are those that are managed by an identified party (individual or organization) that can make decisions regarding energy systems.

The models developed in chapters 3 to 7 are based on the Energy Efficiency and Risk Management in Public Buildings (EnRiMa) project test sites. The EnRiMa project is described in more detail in Section 2.1. Its test sites are:

- The FASAD building. It is an integration center for handicapped people in Siero (Asturias, Spain) that belongs to the Asturian Foundation for Attending Handicapped People (Fundación asturiana de atención y protección a personas con discapacidades y/o dependencias, FASAD);
- The Pinkafeld campus. It is a university campus of the University of Applied Sciences in Pinkafeld (Burgenland, Austria).

1.2.2. Energy Systems Dynamics

Buildings' energy flows can be represented by Sankey diagrams (see Schmidt 2008) providing a straightforward way of visualizing the building energy systems' dynamics. Figure 1.1 shows an actual

Sankey diagram for the Pinkafeld campus EnRiMa test site. The energy flows from the supply side (left) to the demand side (right) throughout technologies (boxes). On the supply side there may be markets, such as the electricity macro grid, and renewable sources, such as solar irradiation. Different types of energy are transformed into others to meet the users demand.

1.2.3. Strategic and Operational Decisions

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 to buy from the grid at a given hour; how much NG to input into a CHP generator. Note that both types of decisions are interdependent as only systems that are available can be used, and decisions on investing on new equipment or renovation depend on how they can be used to meet the overall requirements.

In the Sankey diagram shown in Figure 1.1, the numbers in the white boxes over the arrows and technologies are operational decisions. Some of them are the result of other *real* decisions. For example, the energy output from a technology might be determined by the energy input (real decision) and the conversion ratio. The strategic decisions are the cinnamon-colored boxes: which technologies are available in the building and their capacity.

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

1. Introduction

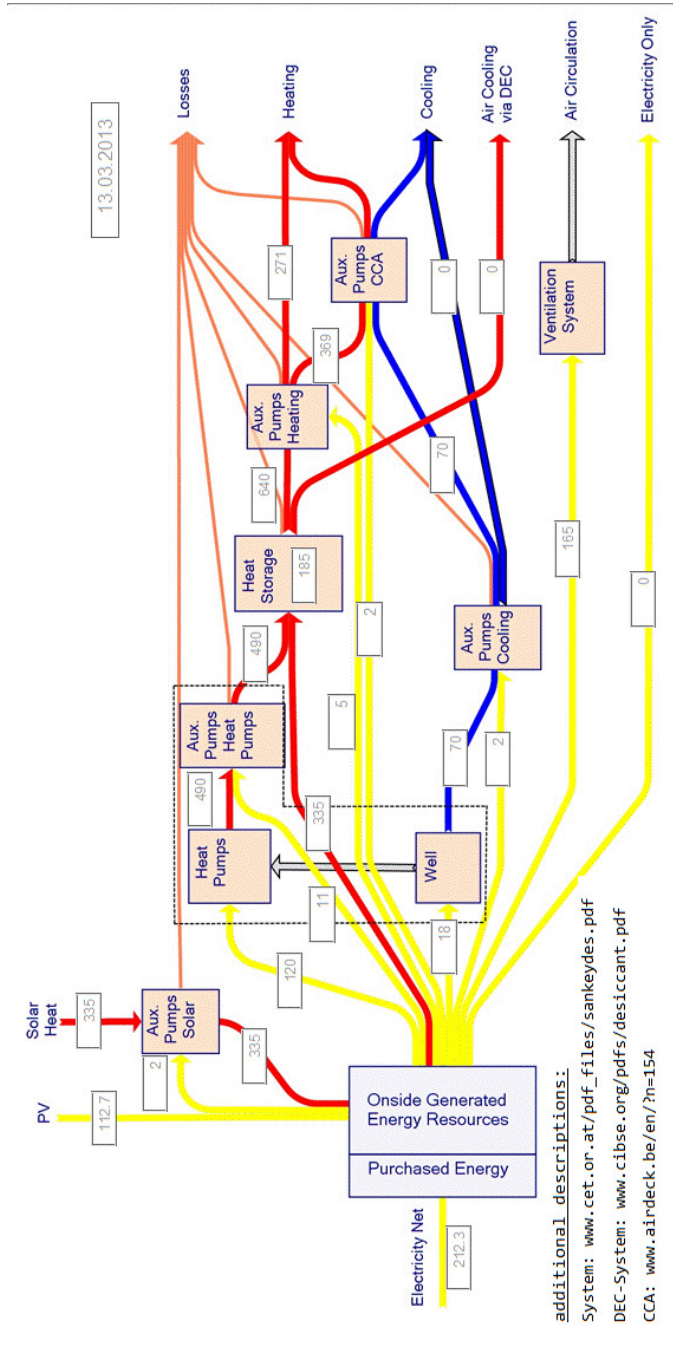


Figure 1.1.: Sankey diagram example for the Pinkafeld Campus EnRiMa test site.

1.3. Decision Making Under Uncertainty

1.3.1. The Role of Uncertainty

Rational decision making must be based on evidence. Some decisions are made under *perfect information*, i.e., knowing all the outcomes and relevant facts affecting such decisions. For example, one can decide whether to vent a room or not knowing the inside and outside temperatures and one's desired comfort level. All the data are known. 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 volatile, 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.

1. Introduction

1.3.2. Decision Support Systems (DSSs)

Several scientific and technical disciplines deal with decision making. Under several denominations, e.g. decision science, decision engineering, or decision analytics, the underlying foundations are those from **Decision Analysis**. 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 may be too rigid for changing environments. Decisions based on analysis require the adoption of a **model** that summarizes and simplifies the problem in order to understand the real problem and find a way to solve it.

Decision making problems can be classified according to several criteria. Bell et al. (1988) proposed a taxonomy based on three categories: descriptive, normative and prescriptive. According to the level at which decisions are made within an organization, decisions can be strategic, tactical, or operational. In Klein et al. (1993) and French et al. (2009) an *instinctive* level is added at the bottom of this pyramidal classification. Note that strategic and operational levels are also linked to the long- and short-term scope of decisions described in Section 1.2. The nature and number of stakeholders, the objectives and the availability of information are other classification criteria.

In order to identify the most 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?

1.3. Decision Making Under Uncertainty

- How does time affect the process?
- Which are the requirements of the system?
- Who is (are) affected by decisions?

In the case at hand, some of the questions have been already answered: 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 on 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, e.g., emissions minimization or efficiency maximization. On the other hand, there are some systems requirements and limitations, such as the energy balance between demand and supply, and the capacity of systems and markets.

The problem described 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. The strategic models developed in Part II are linear, i.e., both the objective and the constraints are linear combinations of the decision variables. Therefore, Linear Programming (LP) is applicable. Moreover, as uncertainty is a key part of the problem, Stochastic Programming (SP) will be used. These techniques are encompassed within the Operations Research (OR) discipline. Other approaches can be followed, and some of them are remarked in Chapter 2.

Considering the complexity of the problem, the use of a **Decision Support System (DSS)** is unavoidable. Usually defined as *an information system that supports decision making* with more or less detail, the 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

1. Introduction

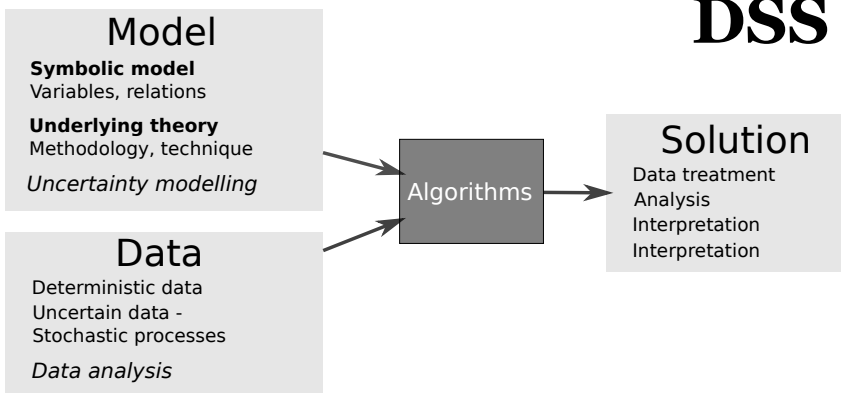


Figure 1.2.: Decision Support System (DSS) diagram.

essence in the first sentence of this section. In this regard, French (2010) proposes “a typology of DSSs in relation to their context of use and the support that they offer.”

Under that paradigm, followed in the framework proposed, the model plays an important role in a DSS. This can be seen in Figure 1.2. Both the model and the data are the basis for the decisions. Appropriate algorithms are applied once the model is defined and the data is available. The DSS should be also capable of preparing the data in a model-suitable way. The model must be based on strong scientific knowledge. Decisions obtained by the DSS, regardless their category (descriptive, normative, or prescriptive), should include interpretation and analysis, probably requiring some posterior data analysis.

1.4. Dissertation Outline

1.4.1. Outlook

The framework is presented in a similar flow of that in Figure 1.2. Starting from the model, a bottom-up approach is followed so as to

reach a good understanding of the models before extending them up to the realistic final problem. The successive models are illustrated with simplified numerical examples. Then, before analyzing solutions, some DSS *enveloping* matters are explained. The framework is eventually used to solve realistic model instances.

In spite of focusing on energy systems modeling, some of the present work can be easily extended to further problems. In particular, the stochastic models are suitable for problems where strategic and operational decisions appear at the same time. Furthermore, as a framework, the features in Chapter 8 are valid for any optimization problem.

There is a short reference glossary in Appendix A. It may be advisable to check it at this point, before reading the whole document.

1.4.2. A Baseline Example

To illustrate the problem, a simple example will be used. It is inspired by the classical *news vendor* problem used in many textbooks, see for example Birge and Louveaux (2011).

Example 1. *Toy example:* 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 . Usually there is a limit \bar{x} , $0 \leq x \leq \bar{x}$ for the capacity in the building, for example due to physical limitations. 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, the cost function for the building energy procurement is:

1. Introduction

$$cx + d^+ y^+(\xi) - d^- y^-(\xi). \quad (1.1)$$

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 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 to be minimized, i.e., Equation (1.1) can be expressed as:

$$C = cx + \mathbb{E}_\xi [d^+ y^+(\xi) - d^- y^-(\xi)] = cx + \mathbb{E}_\xi [d^+ \max\{0, \xi - x\} - d^- \max\{0, x - \xi\}], \quad (1.2)$$

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

$$C'(x) = \frac{\partial C}{\partial x} = 0, \quad (1.3)$$

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

$$\mathbb{P}[\xi < x] = \frac{d^+ - c}{d^+ - d^-}, \quad (1.4)$$

where $\mathbb{P}[\cdot]$ is the probability function. So, the probability of the demand being higher than the strategic decision is fixed by the data. Given that $d^+ > c > d^-$, Equation (1.4) 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, returns the solution 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 the example presented, both first- and second-stage decisions are made within comparable periods of time. In reality, the operational decisions are made within the strategic periods. The challenge arises when having strategic (long-term) and operational (short-term) decisions within the same model. Due to the own structure of the problem, operational decisions induce risk aversion on strategic decisions. On the other hand, as the operational periods are embedded into the strategic ones, the size of the problem tremendously increases as elements are added to the model. Several modeling approaches to deal with these issues are detailed and compared in the following chapters.

1.4.3. Main Questions Addressed

The main questions that will be addressed in the following chapters and, hopefully, answered by the results are:

- How to model the energy flows jointly with strategic issues;
- How to model uncertainty in order to make the models robust;
- How to manage risks within the model;
- What should be a DSS like to better deal with the model, the data and the solution;
- What are the advantages, if any, of using more detailed models instead of simpler, aggregated ones.

1. Introduction

1.4.4. Conventions and Structure

The following conventions regarding the notation apply throughout this work, unless another interpretation is specified:

- Variables are identified by small Latin letters, usually using the last letters of the alphabet, e.g., x, y, z . More than one letter can be used;
- Parameters are identified almost always by capital Latin letters, usually using the first letters of the alphabet, e.g., A, D . An exception applies for well-known generic model formulation, e.g., right-hand-side vector \mathbf{b} in LP;
- For both variables and parameters, when the symbol appears in bold face refers to a vector or matrix, e.g., \mathbf{A} ;
- Examples are identified by a vertical gray line to the left of the text from the beginning to the end of the example;
- Uncertain values or vectors are represented by Greek letters, e.g., ξ ;
- The special functions for the expectation of a random variable and the probability function are represented by \mathbb{E} and \mathbb{P} symbols respectively;
- Acronyms are shown in full form, i.e., including both written out form and acronym, at least the first time they appear in the text. The complete list of acronyms can be consulted just after the table of contents;
- Code is shown in `teletype` font with a light-grey background. Input code may be shown with colored syntax. Output code lines are always preceded by two comment symbols, i.e., `##`;
- The names of software applications and programming languages are printed in sans-serif font, e.g., `R`, `Python`. The names of functions, objects, and classes are printed within the text in teletype font, e.g., `optimSMS`.

1.4. Dissertation Outline

After this introductory chapter providing an overview of the problem, the following structure completes the dissertation:

- Chapter 2 provides the scientific and technical background that applies, including a review of the literature and the state of the art;
- In Chapter 3, the simplest model is presented. Few variables and parameters under a deterministic perspective allow to understand the essence of the problem and the need for apply more sophisticated techniques;
- Chapter 4 extends the deterministic model to a stochastic two-stage model;
- Chapter 5 broaden the deterministic model in Chapter 3 in order to include more options reflecting the reality of a building;
- Finally, Chapter 6 contains the multi-stage stochastic model gathering all the features;
- Chapter 7 implements risk management in the model. In addition to the treatment of uncertainty inherent to SP, risk terms and weights are added to the model in order to deal with risk aversion;
- Chapter 8 presents a framework for DSSs based on a reproducible research approach, and its implementation;
- Chapter 9 summarizes the findings and present conclusions and further research;
- Finally, three appendices are available for reference: a glossary of terms, the code used in Chapter 8, and the session information where the code was run.

2. Background

2.1. The EnRiMa Project

The framework proposed in this work has been applied to the EnRiMa project¹. EnRiMa (EnRiMa project consortium 2010-2014) is a 7th Framework Program (FP7) research project funded by the EC, whose overall objective is to develop a 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.

¹<http://www.enrima-project.eu>

2. Background

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 Cano et al. (2012b) and Groissböck et al. (2013b). 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 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, which provides the DSS that would aid the operators in integrated management of conflicting goals such as cost reduction, meeting energy, efficiency, and CO₂ 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 2.1.

This dissertation is, in part, the result of the contributions to the project, and the knowledge acquired from it. The models developed in Chapters 3 to 7 are the result of fruitful discussions and cooperation with URJC colleagues and with other project partners' teams. Likewise, the spirit of the framework described in Chapter 8 has been applied while developing the Solver Manager, an important component of the EnRiMa DSS. Furthermore, such a Solver Manager uses the `optimr` library developed, whose functionality is explained in Chapter 8. The EnRiMa Solver Manager structure can be seen in Figure 2.2. The Solver Manager is part of the EnRiMa DSS Engine, along with the Scenario Generator and the DSS Kernel. A diagram of the EnRiMa DSS architecture and the relationship between the Solver Manager and the rest of the modules can be seen in Figure 2.3.

2.1. The EnRiMa Project

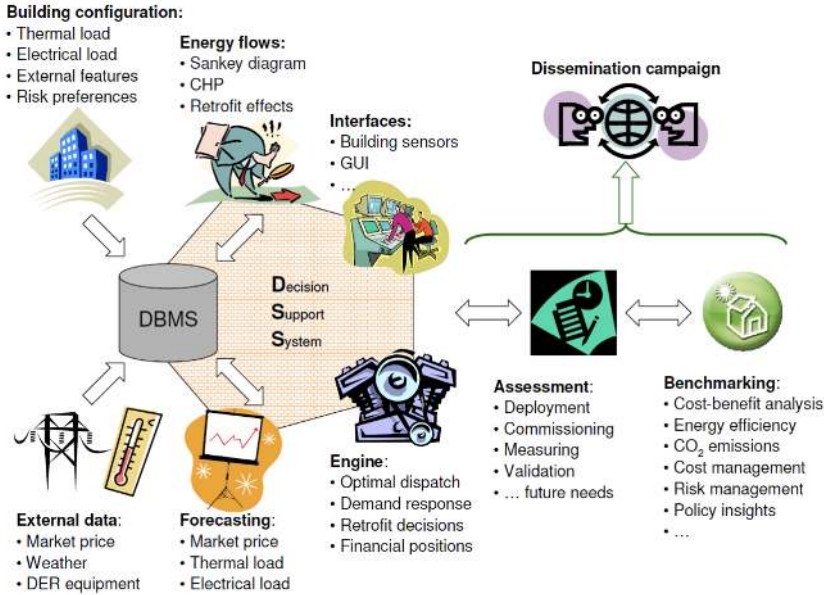


Figure 2.1.: EnRiMa DSS functionality.

One of the EnRiMa project's innovations is the capability to optimize adoption and deployment of energy-efficient equipment taking into account long- and short-term decisions. Thus, two different models have been developed: strategic and operational. Moreover, strategic decisions are influenced by the operational performance, while operations depend on historic strategic decisions. Operational decisions are those involving the dispatch of installed technologies in the short term, whereas strategic decisions concern, in the long term, to which technologies to install and (or) decommission, or which contracts to select from the market. In an attempt to tackle short- and long-term decisions as a whole, the strategic model includes a simplified version of operational energy-balance constraints. The operational model, in turn, includes the realization of the strategic decisions as parameters. Figure 2.4 shows the interplays between both models and their Decision Variables (DVs) and constraints.

2. Background

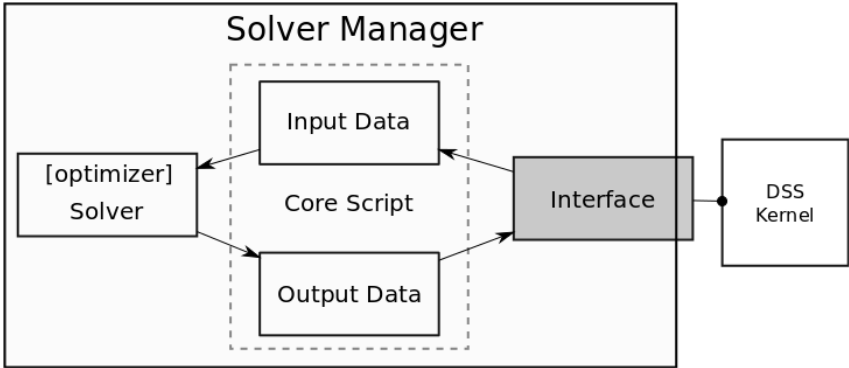


Figure 2.2.: EnRiMa Solver Manager structure.

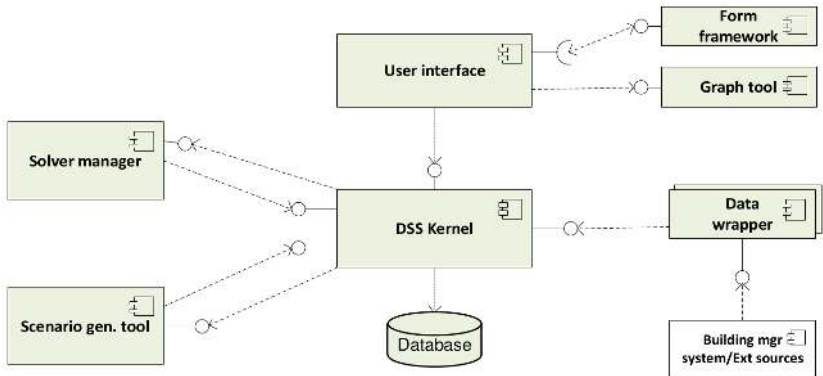


Figure 2.3.: EnRiMa DSS Architecture.

2.1. *The EnRiMa Project*

Thus, two different models have been developed for the EnRiMa DSS, namely: a strategic model and an operational model. A third extension of both models is the upper-level operational model, implementing the upper-level constraints, decisions and parameters to deal with several energy types in the short term. In this work, only the strategic model, which is importantly affected by uncertainties, is taken into account.

2. Background

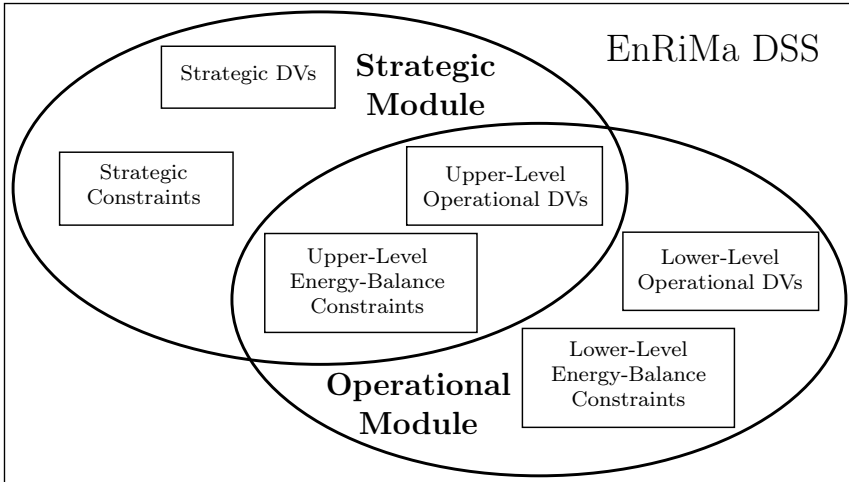


Figure 2.4.: EnRiMa project DSS Schema.

2.2. Methodology

2.2.1. Operations Research (OR)

In the introductory chapter, the problem tackled in this work was framed within the OR discipline. The Institute for Operations Research and the Management Sciences (INFORMS) defines OR as “the discipline of applying advanced analytical methods to help make better decisions”². Thus, DSSs fit perfectly into OR. A number of techniques are used within OR in order to apply Science to decision making. Some of the more relevant are:

- Simulation;
- Optimization;
- Probability and Statistics.

Other underlying tools used in OR are network theory, queuing theory, inventory theory, or scheduling, among others. Historically,

²<http://www.scienceofbetter.org/>

OR originated during the World War II to describe *operative* aspects, in contrast to technical aspects, and the term Operations Research was coined by the Bawdsey Research Station superintendent A.P. Rowe in the UK in 1937.

OR requires mainly two skills: modeling and optimizing. The former allows to obtain an abstraction of a real system in terms of mathematical entities (variables, constants, parameters, etc.) and their relations. The latter applies algorithms to understand how the system works and unveil how the decision maker can take action over it to get an optimal performance under certain goals.

A field related to OR from which decision making problems similar to the ones studied here can be tackled is Game Theory. It is more suitable when different decision makers interact with the system, usually with conflicting goals and where different strategies are available, and a sort of *equilibrium* is pursued. The foundations for Game Theory were set by Von Neumann and Morgenstern (1953). A comprehensive OR methods compendium can be found in Taha (2011).

2.2.2. Mathematical Programming

Mathematical Programming, also known as Mathematical Optimization, is one of the optimization techniques used in OR, see Shapiro (1979). It consists on the selection of the best option among several available alternatives. This set of alternatives, namely the feasible set, is usually constrained by certain restrictions. Thus, a mathematical programming problem can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}), & \quad (2.1) \\ \text{s.t. } g_i(\mathbf{x}) \leq 0, & \quad i = 1, \dots, n, \\ h_j(\mathbf{x}) = 0, & \quad j = 1, \dots, m, \end{aligned}$$

where the goal is to get the values for the vector \mathbf{x} that minimize the objective function $f(\mathbf{x})$, subject to the sets of constraints $g_i(\mathbf{x})$ and $h_j(\mathbf{x})$. Such \mathbf{x} vector contains the decisions that can be made and

2. Background

is usually referred to as *decision variables*. In general, the problem might result in the following situations:

1. The feasible set is empty, and therefore there is no solution to the problem;
2. The problem is unbounded, and therefore there is not an optimal solution;
3. There is a subset of solutions that are better than the rest of the feasible set, but none of them is an optimum;
4. There is an optimal solution \mathbf{x}^* whose objective function value is better than any other from the feasible set.

Linear Programming (LP)

A particular case of mathematical optimization is when both the objective function and the constraints are linear functions, i.e., they are linear combinations of the decision variables' vector \mathbf{x} . In this case, the mathematical method to be used is Linear Programming (LP). It was first formulated by G.B. Dantzig in 1947, who proposed the *simplex algorithm* for its solution, see Dantzig (1965). In general, a LP problem can be formulated in the canonical form as follows:

$$\begin{aligned} \min_{\mathbf{x}} \mathbf{c}^T \mathbf{x}, & \quad (2.2) \\ \text{subject to : } \mathbf{A}\mathbf{x} \leq \mathbf{b}, & \\ \mathbf{x} \geq 0, & \end{aligned}$$

where \mathbf{x} is the vector of unknown decisions, \mathbf{c} is the vector of known costs, \mathbf{A} is the known matrix of coefficients, and \mathbf{b} is the so-called vector of *right-hand-side (rhs)* coefficients, which are also known. Every linear programming problem, namely *primal problem*, can be converted into its *dual problem*. The dual problem provides an upper bound to the optimal value of the primal problem. Thus, the dual of problem (2.2) is:

2.2. Methodology

$$\begin{aligned} \min_{\mathbf{y}} \mathbf{b}^T \mathbf{y}, & \quad (2.3) \\ \text{subject to : } \mathbf{A}^T \mathbf{y} \geq \mathbf{c}, & \\ \mathbf{y} \geq 0. & \end{aligned}$$

There is a strong relationship between the objective values of the primal and dual problems which is useful in order to analyze the model and the solution.

In order to apply the simplex algorithm, the problem must be formulated in *augmented form*, i.e.:

$$\begin{aligned} \min_{\mathbf{x}} \mathbf{c}^T \mathbf{x}, & \quad (2.4) \\ \text{subject to : } \mathbf{A}\mathbf{x} = \mathbf{b}, & \\ \mathbf{x} \geq 0. & \end{aligned}$$

Note that regardless the original linear problem form, it can always be formulated in canonical or augmented form by reshaping the model without changing the interpretation of the results:

- Maximize an objective function is equivalent to minimize the objective function multiplied by -1;
- Likewise, the direction of an inequality can be changed by multiplying by -1 both sides;
- An inequality can be converted into an equality by adding *slack* variables;
- If a variable is unrestricted in sign, or its upper bound is negative, it can be easily redefined.

Optimization software usually transforms the model automatically, so it is not needed to do any changes before calling the solver. Slack variables can provide additional information about the decisions.

2. Background

Problem (2.2) assures that the feasible set is convex. Moreover, the objective function is convex and concave, as it is linear. Therefore, a local minimum is also a global minimum. This is the basis of the simplex method, which basically consists on moving through the boundaries (frontier) of the feasible region from one vertex, i.e., extreme point, to the next. Other important algorithm, actually family of algorithms as both have evolved throughout time, is the interior-point algorithm, see Karmarkar (1984), Gill et al. (1986), or the seminal book by Fiacco and McCormick (1968). Also based on convex optimization, it improves a feasible interior solution point by steps through the interior, rather than around the frontier. In depth explanations of LP algorithms and applications can be found in Dantzig and Thapa (1997) and Dantzig and Thapa (2003).

Extensions of LP problems include those in which some or all the decision variables must be integer numbers. More specific cases include binary variables that can only be zero or one. Mixed Integer Programming (MIP) methods are then to be used. Some of those methods are extensions of LP methods. In other cases, approximate methods and heuristics are used. For pure Integer (Linear) Programming (IP) problems, exact algorithms like cutting plane or brunch and cut can be used. Non-convex, non-linear, and combinatorial optimization are other types of mathematical optimization problems that are out of the scope of this work.

2.2.3. Stochastic Optimization (STO)

In the methods described above, the model parameters are assumed to be known. However, in many cases this is not true and there is uncertainty about the real values of the data affecting the system. This is especially the case in dynamic models where parameters refer to future values of certain characteristics of the system or its environment. Moreover, strategic models include long-term perspectives, which are even more difficult to predict. Energy prices, demand or investment costs were identified in Chapter 1 as the more relevant uncertainties in the problem studied here. When

2.2. Methodology

uncertainty appears, optimization problems become Stochastic Optimization (STO) problems, and parameters are actually random variables, see Subsection 2.2.4 for a brief introduction to uncertainty modeling. The STO term may include further numerical methods and algorithms such as optimal search, among others. To make a distinction, it is common to use also the term Stochastic Programming (SP) for the stochastic extension of mathematical programming problems. Formally, instead of having functions of the vector of decisions \mathbf{x} , the model includes functions of both decisions \mathbf{x} and uncertain data represented by a vector of random variables $\boldsymbol{\xi}$, whose probability distribution is assumed to be known, e.g., through its cumulative distribution function. The random variables are defined by a given sample space $\omega \in \Omega$. Therefore, the real values of $\boldsymbol{\xi}$ are only known after the experiment or observation, $\boldsymbol{\xi} = \boldsymbol{\xi}(\omega)$. Thus, model (2.1) is formulated in the following general form:

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\xi}), & \quad (2.5) \\ \text{s.t. } g_i(\mathbf{x}, \boldsymbol{\xi}) \leq 0, & \\ h_j(\mathbf{x}, \boldsymbol{\xi}) = 0. & \end{aligned}$$

A simplistic way to deal with these types of problems is to estimate the parameter values through its expected value and solve the corresponding deterministic problem:

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}, \mathbb{E}_{\omega} [\boldsymbol{\xi}(\omega)]), & \quad (2.6) \\ \text{s. t. } g_i(\mathbf{x}, \mathbb{E}_{\omega} [\boldsymbol{\xi}(\omega)]) \leq 0, & \\ h_j(\mathbf{x}, \mathbb{E}_{\omega} [\boldsymbol{\xi}(\omega)]) = 0. & \end{aligned}$$

This approach may lead to wrong decisions for several reasons. First, it provides *degenerated* optimal solutions for the average scenario. Eventually, this solution will not be optimal for the actual situation, or even it could lead to a really bad outcome. Second, the average scenario may never occur. And, more importantly, the

2. Background

solution for the average scenario can be infeasible for the real scenario, resulting on certain risks, 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?

Example 2. *Toy example (cont.):* In the baseline example, the deterministic solution would be the following:

$$x^{*det} = \begin{cases} 0 & \text{if } \mathbb{E}[\xi] \leq 0 \\ \bar{x} & \text{if } \mathbb{E}[\xi] > \bar{x} \\ \mathbb{E}[\xi] & \text{otherwise} \end{cases} \quad (2.7)$$

This degenerated solution may result in shortfalls in case there are restrictions on the second stage decisions. Moreover, it minimizes the objective function for the expected value of the demand, but the real value will be different to the expected value, and the real cost may be much higher.

Instead of using expected values as known data, a SP problem is formulated as the optimization of the expected value of the objective function:

$$\min_{\mathbf{x}} C = \mathbb{E}_{\omega} [f(\mathbf{x}, \boldsymbol{\xi}(\omega))], \quad (2.8)$$

$$\text{s.t. } g_i(\mathbf{x}, \boldsymbol{\xi}(\omega)) \leq 0 \quad \forall \omega, \quad (2.9)$$

$$h_j(\mathbf{x}, \boldsymbol{\xi}(\omega)) = 0 \quad \forall \omega.$$

The solution of SP problems are in general not optimal for any possible scenario. In addition, the value of the objective function for the SP problem solution is always worse compared to the value of the objective function of the deterministic problem solution. Nevertheless, this solution is the best one considering all the plausible scenarios, and therefore it is a **robust solution**. In fact, the solution

2.2. Methodology

of the deterministic problem results always in a worse value of the objective function considering all the scenarios. Madansky (1960) demonstrated these inequalities, and later Birge (1982) defined the Value of Stochastic Solution (VSS) as the difference between the value of the expected objective function (2.8) using the deterministic problem solution \mathbf{x}^{*det} and the value of the expected objective function using the solution of the SP problem \mathbf{x}^{*sto} , that is:

$$C(\mathbf{x}^{*det}) - C(\mathbf{x}^{*sto}). \quad (2.10)$$

The solution of an SP problem depends on the probability distribution of the vector of uncertain parameters $\boldsymbol{\xi}$. It can be seen in the simple example introduced in Chapter 1.

Example 3. *Toy example (cont.):* The SP problem in example 1 can be formulated as follows:

$$\begin{aligned} \min_{x, y^+, y^-} \quad & cx + d^+ y^+(\xi) - d^- y^-(\xi), \quad (2.11) \\ \text{subject to:} \quad & x + y^+ = \xi, \\ & x \leq \bar{x}, \\ & y^- \leq x, \\ & x, y^+, y^- \geq 0. \end{aligned}$$

From Equations (1.2)–(1.3), the optimal solution of this simple example is:

$$x^* = \begin{cases} 0 & \text{if } \frac{d^+ - c}{d^+ - d^-} < F(0) \\ \bar{x} & \text{if } \frac{d^+ - c}{d^+ - d^-} > F(\bar{x}), \\ F^{-1}\left(\frac{d^+ - c}{d^+ - d^-}\right) & \text{otherwise} \end{cases}, \quad (2.12)$$

where $F^{-1}(\alpha)$ is the α -quantile of the cumulative distribution function $F(\xi)$. Note that $F(\xi) = \mathbb{P}[\xi \leq x] = 1 - \mathbb{P}[\xi > x]$, which

2. Background

expression was obtained in Example 1. Therefore, an accurate cumulative distribution function of the demand, i.e., $F(\xi)$, is needed in order to provide a concrete decision.

Examples 1–3 describe a basic *recourse program*. In these types of problems, introduced by Dantzig (1955), some of the decisions are made before uncertainty is resolved. These decisions are called first-stage decisions or *here-and-now* decisions. In two-stage stochastic problems, the rest of the decisions are second-stage decisions or *wait-and-see* decisions. They are made after new information arrives about the random variates. Formally, these types of problems are formulated as follows:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{y}(\omega)} \mathbf{c}^T \mathbf{x} + \mathbb{E}_{\xi} [\min_{\mathbf{y}(\omega)} \mathbf{q}(\omega)^T \mathbf{y}(\omega)], & \quad (2.13) \\ \text{s. t. } \mathbf{Ax} \leq \mathbf{b}, & \\ \mathbf{T}(\omega)\mathbf{x} + \mathbf{Wy}(\omega) = \mathbf{h}(\omega), & \\ \mathbf{x} \geq 0, \mathbf{y} \geq 0. & \end{aligned}$$

An alternative way of representing the model is the following:

$$\begin{aligned} \min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} + \mathbb{E}_{\xi} [Q(\mathbf{x}, \omega)], & \quad (2.14) \\ \text{s. t. } \mathbf{Ax} \leq \mathbf{b}, & \\ \mathbf{x} \geq 0, & \end{aligned}$$

where:

$$\begin{aligned} Q(x) = \min_{\mathbf{y}(\omega)} \mathbf{q}(\omega)^T \mathbf{y}(\omega), & \quad (2.15) \\ \text{s. t. } \mathbf{T}(\omega)\mathbf{x} + \mathbf{W}(\omega)\mathbf{y}(\omega) = \mathbf{h}(\omega), & \\ \mathbf{y}(\omega) \geq 0. & \end{aligned}$$

2.2. Methodology

For the first-stage decisions, vectors \mathbf{c} and \mathbf{b} , and matrix \mathbf{A} are known. In the second stage, a number of random events $\omega \in \Omega$ may occur. For each of these events, the second-stage vectors \mathbf{q} and \mathbf{h} , and matrices \mathbf{T} and \mathbf{W} become known. Again, the probability distribution of $\boldsymbol{\xi}$ is also known.

Through the characterization of uncertainty by means of scenarios, i.e., $\Omega = \{\omega_1, \dots, \omega_T\}$, SP problem (2.13)–(2.15) can be converted into its Deterministic Equivalent Program (DEP):

$$\min_{\mathbf{x}, \mathbf{y}(\omega)} z = \mathbf{c}^T \mathbf{x} + \sum_{\omega \in \Omega} \mathbb{P}(\omega) \mathbf{q}(\omega)^T \mathbf{y}(\omega), \quad (2.16)$$

subject to: $\mathbf{A}\mathbf{x} = \mathbf{b}$,

$$\mathbf{T}(\omega)\mathbf{x} + \mathbf{W}(\omega)\mathbf{y}(\omega) = \mathbf{h}(\omega) \quad \forall \omega \in \Omega,$$

$$\mathbf{x} \in X, \mathbf{y}(\omega) \in Y \quad \forall \omega \in \Omega, \quad (2.17)$$

where $\mathbb{P}[\omega]$ is the probability of scenario ω . In multi-stage SP problems, Constraints (2.17) must include nonanticipativity constraints, which guarantee that decisions at early stages are the same for all the scenarios. In a multi-stage SP problem with n stages, there are first-stage decisions, second-stage decisions, \dots , n -stage decisions. At each stage, new information about the random parameters arrives, and decisions are made. Two-stage SP problems are special cases of multi-stage problems. In practice, two types of notation can be used to represent and solve SP problems, namely: node-variable formulation and scenario-variable formulation. Node-variable formulation is also known as *compact form*, see for details Alonso Ayuso et al. (2009) or Conejo et al. (2010). It will be used in Chapter 6.

For certain problems where SP is not suitable for some reason, e.g., due to high dimensionality, other STO techniques can be used. In the books by Marti et al. (2006), Marti (2008), Marti et al. (2010), Ermoliev et al. (2012a) some of them can be found, including recent approaches.

2. Background

2.2.4. Uncertainty Modeling

Since the uncertain parameters affecting the system are actually random variables, some background on this topic is needed to solve SP problems. Several techniques for estimation, forecasting, scenario generation and scenario reduction, which are not included in this work, are available in, for example, Conejo et al. (2010) or Dupačová et al. (2003).

Theory of Probability and Statistics are the most important knowledge areas to deal with uncertainty. In what follows, a summary of the main concepts used in this work are outlined. Most of the concepts can be found in the international standard ISO 3534-1:2006 (ISO, 2006).

In probability theory, *event* is an essential concept. Roughly, an event is a possible outcome of a random experiment or observation. Formally, an event ω is a subset of the sample space Ω , which is the set of all possible outcomes. The uncertainty of those events are then modeled by means of a probability function \mathbb{P} . The following types of events can be defined:

- Event ω, ω_i : subset of the sample space Ω ;
- Complementary event ω^c is the sample space excluding the given event, $\{\Omega \setminus \omega\}$;
- Events are independent if they can not occur at the same time;
- Elementary event: contains only a single event of the sample space. Elementary events are independent and mutually exclusive;
- Empty event \emptyset is the event that none of the events in the sample space occur;
- Sure event is the sample space Ω event.

The probability function assigns to each event a probability of occurrence. A probability function is defined by the following axioms:

2.2. Methodology

1. $\mathbb{P}[\omega] \in \mathbb{R}, \mathbb{P}[\omega] > 0 \forall \omega \in \Omega$;
2. $\mathbb{P}[\Omega] = 1$;
3. $\mathbb{P}[\omega_1 \cup \omega_2 \cup \dots] = \sum_{i \in \mathcal{I}} \mathbb{P}[\omega_i]$, where \mathcal{I} is a set of independent events in the sample space Ω .

Some consequences:

1. $\mathbb{P}[\omega_i] \leq \mathbb{P}[\omega_j] \forall \omega_i \subseteq \omega_j$;
2. $\mathbb{P}[\emptyset] = 0$;
3. $0 \leq \mathbb{P}[\omega] \leq 1 \forall \omega \in \Omega$;
4. $\mathbb{P}[\omega_i \cup \omega_j] = \mathbb{P}[\omega_i] + \mathbb{P}[\omega_j] - \mathbb{P}[\omega_i \cap \omega_j]$;
5. $\mathbb{P}[\omega^c] = 1 - \mathbb{P}[\omega]$;
6. $\mathbb{P}[\omega_i \cap \omega_j] = 0 \forall i, j \in \mathcal{I}, i \neq j$, where \mathcal{I} is a set of independent events in the sample space Ω .

Another important concept for probability calculus is the conditional probability. The probability of the event ω_i given that the event ω_j occurs is:

$$\mathbb{P}(\omega_i|\omega_j) = \frac{\mathbb{P}[\omega_i \cap \omega_j]}{\mathbb{P}[\omega_j]}.$$

Finally, the following important theorems are widely used when modeling uncertainty:

Theorem 1. *Law of total probability: If ω_j are n independent events and $\bigcup_{j=1}^n \omega_j = \Omega$, then for any event $\omega_i \in \Omega$:*

$$\mathbb{P}[\omega_i] = \sum_{j=1}^n \mathbb{P}[\omega_i|\omega_j] \cdot \mathbb{P}[\omega_j].$$

Theorem 2. *Bayes' theorem:*

$$\mathbb{P}[\omega_i|\omega_j] = \frac{\mathbb{P}[\omega_j|\omega_i] \cdot \mathbb{P}[\omega_i]}{\mathbb{P}[\omega_j]}.$$

2. Background

Random variables

Formally, a random variable ξ is a function defined on a sample space Ω where the values of the function are ordered k-tuplets of real numbers. In short, a random variable assigns a number to an event of the sample space Ω . A random variable can be continuous or discrete.

Probability mass function For a discrete random variable, the probability mass function gives the probability that the random variable equals a given value:

$$f_{\xi}(x) = \mathbb{P}[\xi = x],$$

where $f_{\xi}(x)$ is defined for all the possible values of the random variable ξ .

Probability density function For continuous variables, the probability defined above has no sense, as the probability of an exact point in the real numbers is equal to zero. Instead, the probability density function is defined as a function over a random variable with the following characteristics:

1. $f_{\xi}(x) \geq 0 \forall x$;
2. $\int_{-\infty}^{\infty} f_{\xi}(x)dx = 1$.

Thus, in continuous functions the probability of the random variable to be between two values can be computed as follows:

$$\mathbb{P}[a \leq \xi \leq b] = \int_a^b f_{\xi}(x)dx.$$

Cumulative distribution function The cumulative distribution function is often referred simply as distribution function, and by definition it is:

$$F_{\xi}(x) = \mathbb{P}(\xi \leq x).$$

Therefore, for discrete random variables:

$$F_{\xi}(x) = \sum_{x_i \leq x} f_{\xi}(x_i),$$

and for continuous distributions:

$$F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(x) dx. \quad (2.18)$$

There is a direct and important relationship between the cumulative distribution function and the density function of a continuous variable:

$$F'_{\xi}(x) = \frac{dF_{\xi}(x)}{dx} = f_{\xi}(x). \quad (2.19)$$

The inverse function of the cumulative distribution function is the α -quantile, defined as the value of the random variable ξ equal to the infimum of all $x \in \xi$ such that the cumulative distribution function is greater than or equal to α :

$$F_{\xi}^{-1}(\alpha) = \inf_{x \in \xi} \{\mathbb{P}[\xi \leq x] \geq \alpha\}.$$

Mathematical expectation Expectation is defined for functions of random variables as:

$$\mathbb{E}[g_{\xi}(x)] = \int_{\xi} g_{\xi}(x) dF(x).$$

2. Background

In the case of discrete variables, a summation applies instead of an integral:

$$\mathbb{E}[g_\xi(x)] = \sum_{x_i \in \xi} g_\xi(x_i) dF(x).$$

When the function used in the expectation is just a power of order r of the random variable, the expectation is the moment of order r . The mean μ of the random variable is the moment of order 1, for continuous distributions:

$$\mu = \mathbb{E}[\xi] = \int_{\xi} x dF(x), \quad (2.20)$$

and for discrete distributions:

$$\mu = \mathbb{E}[\xi] = \sum_{x_i \in \xi} x_i \mathbb{P}[x_i].$$

Some useful properties of the expectation are:

$$\mathbb{E}[\xi + a] = \mathbb{E}[\xi] + a, \quad (2.21)$$

$$\mathbb{E}[\xi_1 + \xi_2] = \mathbb{E}[\xi_1] + \mathbb{E}[\xi_2], \quad (2.22)$$

$$\mathbb{E}[a\xi] = a\mathbb{E}[\xi]. \quad (2.23)$$

Variance The variance σ^2 is a measure of variability of the random variable, and it is defined by the 2^{nd} central moment, i.e., about the mean, of the random variable:

$$\sigma^2 = V[\xi] = \mathbb{E}_{\xi} [(x - \mu)^2].$$

2.3. Decision Support Systems (DSSs) for Optimization

Therefore, for continuous variables:

$$V[\xi] = \int_{\xi} (x - \mu)^2 dF(x),$$

and for discrete variables:

$$V[\xi] = \sum_{x_i \in \xi} (x_i - \mu)^2 \mathbb{P}(x_i).$$

The standard deviation σ is the positive square root of the variance. The variance can be also computed with the following alternative formula:

$$V[\xi] = \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2.$$

Other important properties of the variance are:

- $V[\xi + a] = V[\xi]$;
- $V[a\xi] = a^2V[\xi]$;
- $V[a\xi_1 + b\xi_2] = a^2V[\xi_1] + b^2V[\xi_2] + 2abCov[\xi_1, \xi_2]$,

where $Cov[\xi_1, \xi_2]$ is the covariance between two random variables, defined by:

$$Cov[\xi_1, \xi_2] = \mathbb{E}[(\xi_1 - \mu_1)(\xi_2 - \mu_2)].$$

2.3. Decision Support Systems (DSSs) for Optimization

2.3.1. General Framework

In Chapter 1, the main ideas about DSS guiding the framework proposed were outlined. Under this approach, an authentic DSS

2. Background

should be based not only on data and information technology, but also on models. One step beyond, an effective DSS must be able to provide an environment for *stakeholders dialog*. Figure 2.5 updates the DSS diagram in Chapter 1. Usually decision making is not a static action without continuity, but rather an iterative process, regardless the time cycle duration. Therefore, the outcomes of the process provide endogenous feedback to the own DSS. Moreover, this feedback is not only quantitative in order to improve the accuracy of the estimation of data or distributions. 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 DSS output should tell the decision maker and other stakeholders not only the recommended optimal decisions, but also about the interpretation of the results, consequences, implementation, and usefulness;
- Some examples of stakeholders are decision makers, consultants, modelers, or data managers. All of them interact with the DSS in a continuous base through the feedback received in form of new inputs;
- 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

2.3. Decision Support Systems (DSSs) for Optimization

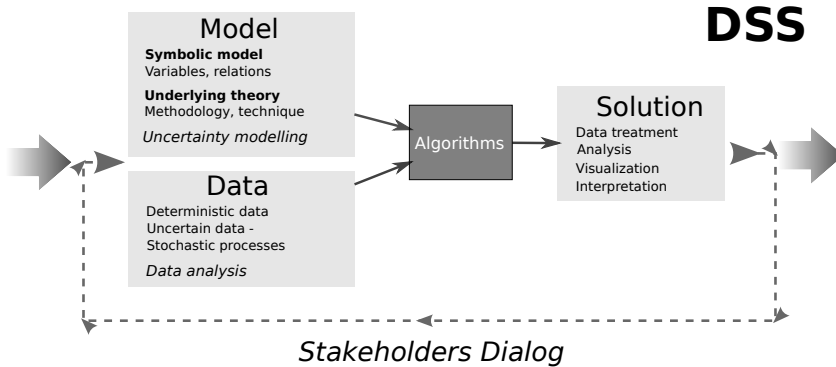


Figure 2.5.: DSS updated diagram: stakeholders dialog.

accurateness of the inputs. Not for nothing the subsequent results will be based on these inputs;

- Decision making 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 DSS output, e.g., advisors, operators, technicians, or managers;
- Dialog with external stakeholders may be necessary at some 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 DSS framework would be:

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

2. Background

- 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.

The developed framework includes innovative features and provides a flexible environment 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. On the other hand, the *reproducible research* approach (see Leisch 2002) adopted in Chapter 8 allows to record and track consistent updates throughout the time, and to provide a sort of *balance score-card* to stakeholders, consistently with all the components of the DSS. Furthermore, the results are made reproducible for any of the stakeholders, which increase the efficiency in multi-disciplinary and changing environments, and the quality of the communication processes.

2.3.2. The Model

Within the structure of the DSS, the model component is represented by the Symbolic Model Specification (SMS). 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

2.3. Decision Support Systems (DSSs) for Optimization

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 will be developed in the forthcoming chapters, from the baseline example in Chapters 1 and 2. In order to generically represent the SMS within the DSS, in Chapter 8 specific data structures are proposed and the framework is implemented in R .

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, see for example Fourer (1983) or 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 has 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.

According to Kallrath (2012c), AMLs are “declarative languages for implementing optimization problems”. 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 allow 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” expressed by Kallrath (2012b). The following list enumerates some of the most important AMLs. A reference with the initial steps of them is provided:

- **GAMS** (General Algebraic Modeling System, Bisschop and Meeraus 1982) has its origin in the World Bank during the

2. Background

late 1970s;

- AMPL (A Mathematical Programming Language, Fourer et al. 1990);
- AIMMS (Advanced Interactive Multidimensional Modeling System, Bisschop 2006) is historically an evolution from GAMS;
- CMPL (COIN Mathematical Programming Language, Schleiff and Steglich 2013) 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, Hart et al. 2012) is a Python package, part of the A Common Optimization Python Repository (Coopr) software library, which includes modeling capabilities in a high-level language.

Despite AMLs have been selected to build the framework, it is important to remark that other structured formats, e.g., markup languages, can be used as far as they are useful to accomplish the DSS main mission, i.e., the stakeholders dialog. For example, Optimization Services (OS)³ is a COIN-OR⁴ project that uses the XML format to represent optimization problems and that is suitable to effectively communicate within an eventual DSS.

2.3.3. The Data

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

³<https://projects.coin-or.org/OS>

⁴<http://www.coin-or.org>

2.3. Decision Support Systems (DSSs) for Optimization

wide range of options both commercial and open source. A non-exhaustive list would include:

- STATA⁵: widely used in econometrics;
- SPSS⁶: time-honored statistical software, recently acquired by IBM;
- SAS⁷: The leader in business analytics;
- Minitab⁸: Well-known statistical software for quality control and improvement;
- R⁹: 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 SP, scenario generators are also needed to combine the data and the uncertainty modeling in order to provide the DSS with the appropriate inputs.

2.3.4. The System

There is a component of the DSS in charge of running the optimization. Usually it is a piece of software containing the algorithms to solve the optimization problem, and it is in general named *the*

⁵<http://www.stata.com/>

⁶<http://www-01.ibm.com/software/analytics/spss/>

⁷<http://www.sas.com/>

⁸<http://www.minitab.com/>

⁹<http://www.r-project.org/>

2. Background

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., LP, Non Linear Programming (NLP), or for different types of problems. The following is a non-exhaustive list of commonly used solvers:

- CPLEX¹⁰: For linear and quadratic problems;
- `lp_solve`¹¹ is an open source solver for MIP problems;
- CLP¹², the LP 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: The authors of this successful collection of solvers claim that they were developed from the ground.

More open source solvers can be found in the COIN-OR projects' website¹³. In addition to solvers' projects, developer tools and interfaces can also be found. For example, the OS project¹⁴, 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".

¹⁰<http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html>

¹¹<http://lpsolve.sourceforge.net/5.5/>

¹²<http://www.coin-or.org/projects/Clp.xml>

¹³<http://www.coin-or.org/projects/>

¹⁴<https://projects.coin-or.org/OS>

2.4. State of the Art

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 from 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 Matlab¹⁵, SciLab¹⁶, or Mathematica¹⁷, among others, include modules to solve mathematical programming problems, or to call further solvers. OpenOpt¹⁸ 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 the components of a DSS disseminated and, more importantly, disconnected between them. A heterogeneous set of tools is often being used for similar tasks. Definitely, this fact blocks stakeholders dialog. In contrast, the proposed framework, whose implementation using R is in Chapter 8, propounds an integration of all the components under the Reproducible Research paradigm. Nevertheless, the framework can be implemented using different technologies according to the stakeholders needs, as far as their dialog is assured.

2.4. State of the Art

Different approaches to DSS 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, e.g., Salewicz and Nakayama (2004) propose a web-based

¹⁵<http://www.mathworks.de/products/matlab/>

¹⁶<https://www.scilab.org>

¹⁷<http://www.wolfram.com/mathematica/>

¹⁸<http://openopt.org/>

2. Background

DSS for managing large international rivers. 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 Boza et al. (2010) 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 work. An important implementations of machine-readable models is the Structure Modeling Language (SML) (Geoffrion 1992a; Geoffrion 1992b). Though a cutting-edge topic, only very recent works deal extensively with Reproducible Research, such as the book edited by Stodden et al. (2013).

Regarding the operational module (see Figure 2.4), at the lower level the operational module models building physics and system thermodynamics in order to make certain end use energy demands (such as those for space heat and cooling) endogenous. The operational module takes installed equipment, the building envelope, and financial positions as given. In modeling the lower-level energy flows, the operational low-level model relies upon the literature on building physics (Engdahl and Johansson 2004; Xu et al. 2008; Platt et al. 2010) or on combining optimization techniques with an accurate representation of the thermodynamics and physics of the building as in Liang et al. (2011). The detailed operational model can be found in Groissböck et al. (2013a).

As far as the strategic module is concerned, different approaches can be found in the literature about energy systems planning. Some of them deal with specific technologies, for example Siddiqui et al. (2005), or Stadler et al. (2009), focus on Distributed Energy Resources (DER) technologies, while Lund (2005) focus on wind technologies. Other optimization models are designed from the production point of view, like Hobbs (1995) or more recently El-Khattam et al. (2004) and Heydari and Siddiqui (2010). Villumsen and Philpott (2012) apply STO to capacity planning of electricity transmission networks with transmission switching. Cai et al. (2008) focus on a regional perspective. Only recent papers tackle systems

planning at the building level. Salvador and Grieu (2012) evaluate the energy impact of buildings equipped with energy production and storage systems on the electricity grid. Kumbaroğlu and Madlener (2011) rely on aggregated demand for optimal retrofit investment options.

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, e.g., Weinberg et al. (1991), or Van Sambeek (2000). Other works, such as Hobbs (1995), Siddiqui et al. (2005), King and Morgan (2007), Marnay et al. (2008), and Stadler et al. (2009) follow a deterministic optimization approach that is unable to provide robust decisions against inherent uncertainties, as stated by Ermoliev and Wets (1988). Even though stochastic optimization has been applied for a long time to cope with uncertainties in other fields, there were not approaches based on the use of STO techniques in order to treat uncertainties for energy efficiency in buildings. This work focuses on the use of SP (Birge and Louveaux 2011) for the development of models suitable to be used within strategic decision making.

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, see Gritsevskii and Ermoliev (1999) and Gritsevskii and Nakicenovic (2000)). 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. 2012b).

It is important to remark the difference between the models presented here and the classical capacity expansion models (Buehring et al. 1984). The traditional Operations Research approach to capacity expansion in the energy sector (Hobbs 1995) mentions MIP problems for resource planning. Capacity expansion models have a utility focused approach, while the models described in this work follow a customer perspective. Discussions on the difficulties of

2. Background

adapting capacity expansion models to the building level, e.g., demand charges, storage, CHP, etc. can be consulted in the literature, see for example Siddiqui et al. (2007). In fact, the management of a building behaves in a different way to the energy management of a whole energy grid. For instance, capacity expansion models do not deal with passive measures on the buildings side, and only cost minimization is normally used. In the same vein, the structure of some Transmission Expansion Planning (TEP) models is similar to the optimization models presented in this work. However, that kinds of models are applied to a different problem far from the building level. Moreover, classical TEP models may not be suitable for deregulated markets (Hobbs 1995), including more recent proposals, such as Garcés et al. (2009), game-theoretic aspects.

Part II.

Modeling

3. The Basic Problem

3.1. Introduction

3.1.1. Scope

In this chapter, a basic deterministic model for strategic decision making concerning planning of energy systems in buildings is developed. As outlined in Chapter 2, the deterministic model relies on average values of stochastic parameters. Moreover, it includes few options in order to better understand the modeling at this stage. Nevertheless, the model will evolve throughout the rest of the chapters of Part II of the dissertation. The model is illustrated with a numerical example, which will be also extended in the next chapter, see also Cano et al. (2014). The model presented in this chapter approaches the real world from the baseline example in Chapters 1 and 2. Now, long- and short-term periods are included in the model, different technologies and types of energy are considered, and other features as efficiency and pollution are added.

3.1.2. Model Overview

Investment 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.

3. The Basic Problem

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, $i \in \mathcal{I} = \{1, \dots, I\}, t \in \mathcal{T} = \{1, \dots, T\}$. 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. For example, Conejo et al. (2007) aggregate hourly energy pool prices and demands for electricity in 72 prices/demands periods for annual planning of energy provision, considering, per month, Monday peak and off-peak values, working day (other than Monday) peak and off-peak values, and Saturday and Sunday values.

3.2. Building the Model

3.2.1. The Simplest Case

The demand profile within each year t , can be adequately characterized by the demand within representative periods j , $j \in \mathcal{J} = \{1, \dots, J\}$. This time structure is represented in Figure 3.1, 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 .

3. The Basic Problem

$$\sum_{t \in \mathcal{T}} \left(\sum_{i \in \mathcal{I}} CI_i^t \cdot x_i^t + \sum_{i \in \mathcal{I}, j \in \mathcal{J}} CO_{i,j}^t \cdot DT_j^t \cdot y_{i,j}^t \right), \quad (3.6)$$

where CI_i^t is the unit investment cost for technology i in year t , $CO_{i,j}^t$ is the unit production cost (including fuel costs) for i in period j in t , and DT_j^t is the duration of period j in t .

3.2.2. Adding Features

Efficiency

The energy efficiency and aging processes may be introduced in equation (3.2) as

$$s_i^t = E_i^t \cdot s_i^{t-1} + x_i^t - x_i^{t-LT_i} \quad \forall i \in \mathcal{I}, t \in \mathcal{T}, \quad (3.7)$$

where E_i^t denotes “depreciation” (aging) rate in t . Aging can be also modeled as a dynamic process in a detailed technology-based manner. This will be done in Chapter 5, where the deterministic model will be extended with further features.

Budget

An investment constraint may also be introduced for each t as:

$$\sum_{i \in \mathcal{I}} CI_i^t \cdot x_i^t \leq IL^t \quad \forall t \in \mathcal{T}. \quad (3.8)$$

Energy Types

In general cases, there is a set \mathcal{K} of available energy types (electricity, gas, heat, possibly also solar radiation). Each technology i is then characterised by its set of input-output conversion coefficients $A_{i,l,k}^t$, $l, k \in \mathcal{K}$, which convert the unit input energy type l into $A_{i,l,k}^t$ units

3.2. Building the Model

of energy type k . In this case, i denotes different energy technology $i \in \mathcal{I}$ to satisfy demand $D_{j,k}^t$ of energy type k in period j in t . Let us define nonnegative decisions $y_{i,j,k}^t$ operating in period j of energy type k . Therefore,

$$\sum_{i \in \mathcal{I}, l \in \mathcal{L}} A_{i,l,k}^t \cdot y_{i,j,k}^t = D_{j,k}^t \quad \forall j \in \mathcal{J}, t \in \mathcal{T}, k \in \mathcal{K}, \quad (3.9)$$

$$y_{i,j,k}^t \leq G_{i,j}^t \cdot s_i^t \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T}, k \in \mathcal{K}. \quad (3.10)$$

Thus, $\sum_{i,l} A_{i,l,k}^t \cdot y_{i,j,k}^t$ is the input of energy type l required to satisfy demand in energy type k by production plan $y_{i,j,k}^t$ for technology i . The operational costs of technology i are spread over technology life LT_i . The total operational cost $CO_{i,j,l,k}^t$ includes now fuel costs and conversion costs from l to k , and the objective function is to be modified accordingly.

Remark 1. (*Purchasing, energy efficiency and availability factors*).

The purchased electricity can be modeled as a technology, with electricity on both input and output and the purchase cost on the input. The total available electricity would be a sum of the electricity generated by technologies and the electricity purchased from the market. The availability factor $G_{i,j}^t$ defines if the whole ($G_{i,j}^t = 1$) or only a certain portion of the capacity/output s_i^t can be used during time t and period j . This may apply, e.g. for wind-power plants and heating devices (Karlsson et al. 1995). If equation (3.10) is turned into an equality, the portion defined by $G_{i,j}^t$ has to be used during each time period. This may describe/introduce efficiency improvements (energy conservation). For wall insulation (Henning 1997), which has a certain time-dependent influence on the heat demand, $D_{j,k}^t$ is large in the winter but small in the summer.

Pollution Constraints

Besides cost minimization and investment constraints, there may exist environmental constraints, e.g., on feasible CO_2 and other pollutants $h \in \mathcal{H}$ emissions:

3. The Basic Problem

$$\sum_{i,j,l,k} PH_{i,l,k,h} \cdot A_{i,l,k}^t \cdot y_{i,j,k}^t \leq PL_h^t \quad \forall t \in \mathcal{T}, h \in \mathcal{H}, \quad (3.11)$$

where $PH_{i,l,k,h}$ is the amount of pollutant from technologies, and PL_h^t is the emissions limit. Note that the left hand side of equation (3.11) could become an objective function if the decision maker's objective is emissions minimization.

Batteries

Storage processes can be explicitly represented in operational models. In the case of a strategic model, the load profiles are approximated by a proper number of periods j . In this case, it is possible to represent the accumulators by additional decision variables $w_{j,k}^t$, $u_{j,k}^t$ on energy storage and its use at t and j . Equation (3.9) then can be modified as:

$$\sum_{i \in \mathcal{I}, l \in \mathcal{L}} A_{i,l,k}^t \cdot y_{i,j,k}^t \geq D_{j,k}^t + w_{j,k}^t - M_{j,k}^t \cdot u_{j,k}^{t-1} \quad (3.12)$$

$$\forall j \in \mathcal{J}, t \in \mathcal{T}, k \in \mathcal{K},$$

and the following constraints added:

$$Q_k^t \geq \sum_{j \in \mathcal{J}} w_{j,k}^t \quad \forall t \in \mathcal{T}, k \in \mathcal{K}, \quad (3.13)$$

$$\sum_{j \in \mathcal{J}} w_{j,k}^t \geq \sum_{j \in \mathcal{J}} u_{j,k}^t \quad \forall t \in \mathcal{T}, k \in \mathcal{K}, \quad (3.14)$$

$$w_{j,k}^t \geq 0 \quad \forall j \in \mathcal{J}, t \in \mathcal{T}, k \in \mathcal{K}, \quad (3.15)$$

$$u_{j,k}^t \geq 0 \quad \forall j \in \mathcal{J}, t \in \mathcal{T}, k \in \mathcal{K}, \quad (3.16)$$

where Q_k^t is the capacity of accumulators for energy k at time t , $M_{j,k}^t$ loss per unit stored energy, $w_{i,j}^t$ energy storage, and $u_{i,j}^t$ is storage

3.3. Numerical Example

use. Note that parameters Q_k^t and $M_{j,k}^t$ should reflect the effect of aging (degradation) of batteries throughout the decision time horizon. These equations in an aggregate manner redistribute loads among periods. With a shorter time scale, it may be reasonable to introduce shifts of demands in time by substituting $u_{j,k}^t$ in (3.14) with $u_{j,k}^{t-1}$.

3.3. Numerical Example

The model presented in this chapter evolves throughout Chapters 4 to 7. In the same vein, a first simple numerical example, which will also evolve with the models, is constructed in this section. Intentionally, it does not include all the features described in previous sections in order to make it easily understandable. The example has been implemented using the framework described in Chapter 8.

Example 4. *Basic deterministic model.* Let us consider the simplest model defined by (3.1)–(3.6):

$$\begin{aligned} \min \sum_{t \in \mathcal{T}} & \left(\sum_{i \in \mathcal{I}} CI_i^t \cdot x_i^t + \sum_{i \in \mathcal{I}, j \in \mathcal{J}} CO_{i,j}^t \cdot DT_j^t \cdot y_{i,j}^t \right) \\ \text{s.t. :} & \quad s_i^t = s_i^{t-1} + x_i^t - x_i^{t-LT_i} \quad \forall i \in \mathcal{I}, t \in \mathcal{T} \\ & \quad \sum_{i \in \mathcal{I}} y_{i,j}^t = D_j^t \quad \forall j \in \mathcal{J}, t \in \mathcal{T} \\ & \quad y_{i,j}^t \leq G_{i,j}^t \cdot s_i^t \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T} \end{aligned}$$

The demand D_j^t in Table 3.1 is based on the FASAD EnRiMa test site in Asturias (Spain), see Section 2.1. Starting from an annual demand of 213.50 MWh, projections on the demand level have been simulated for all the periods.

For the sake of simplicity, only four representative periods (set \mathcal{J}) have been defined: winter, spring, summer and autumn.

3. The Basic Problem

The input technologies (set \mathcal{I}) are 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 ($G_{i,j}^t = 1$), whereas PV availability depends on the season as shown in Table 3.2 (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>.

As for investment costs CI_i^t , a yearly cost reduction for generating technologies (5% for PV, 10% for CHP) from the current prices (1,326.65 and 3,673.77 EUR/kW respectively) is assumed. The price for the PV panels has been taken from the PREOC price database (<http://www.preoc.es/> retrieved 2013-02-12), 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 $CO_{i,j}^t$, the fuel 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”, considering an increase of 10% and 3% per year for electricity and NG respectively) 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,

3.3. Numerical Example

the duration time is set to 91 days \times 8 hours, considering 13 weeks each period.

The total cost for the optimal solution amounts to 65,213 EUR. The optimal values for the operational decisions are displayed in Figure 3.2. Table 3.3 shows how technologies are incorporated to the building as costs evolve.

Table 3.1.: Electricity demand level (kW)

j	t	D_j^t
winter	2013	17.06
spring	2013	21.93
summer	2013	34.12
autumn	2013	24.37
winter	2014	18.77
spring	2014	24.13
summer	2014	37.53
autumn	2014	26.81
winter	2015	20.64
spring	2015	26.54
summer	2015	41.29
autumn	2015	29.49
winter	2016	22.71
spring	2016	29.20
summer	2016	45.42
autumn	2016	32.44
winter	2017	24.98
spring	2017	32.11
summer	2017	49.96
autumn	2017	35.68

3. The Basic Problem

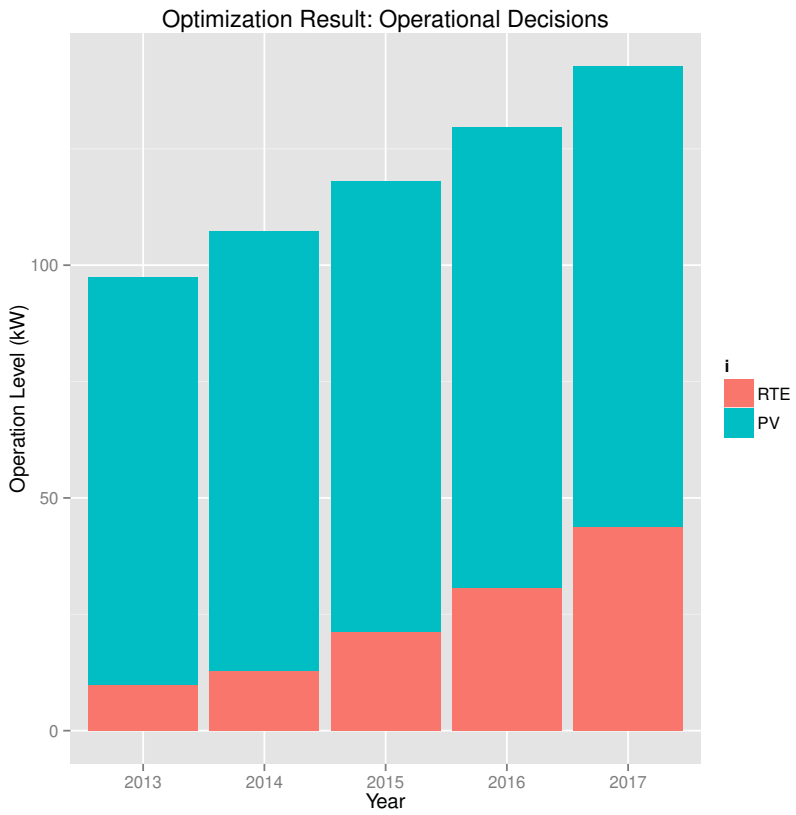


Figure 3.2.: Energy dispatch for Example 4.

3.3. Numerical Example

Table 3.2.: PV technology availability (ratio)

i	j	t	$G_{i,j}^t$
PV	winter	2013-2017	0.30
PV	spring	2013-2017	0.48
PV	summer	2013-2017	0.63
PV	autumn	2013-2017	0.25

Table 3.3.: Total technologies investment (EUR)

i	t	$CI_i^t \times x_i^t$
PV	2013	45891.10
PV	2014	1630.44
RTE	2013	1031.47

4. Adding Uncertainty: Two-stage Problems

4.1. Introduction

Since uncertainties may seriously affect long-term decisions, the model should be formulated as a STO problem. Among the main stochastic parameters are fuel prices and demands. In the presence of uncertainties, there is a dilemma to choose an efficient technological portfolio in real time while pursuing long-term goals. Therefore, the solution of the problem involves the so-called two-stage dynamic stochastic optimization models with a rolling horizon. A two-stage stochastic model is proposed, where some decisions (first-stage decisions) including investments into new energy technologies have to be taken before uncertainties are resolved and some others (second-stage decisions) will be taken once values for uncertain parameters become known, thereby providing a trade-off between long- and short-term decisions.

The main goal of the dynamic stochastic two-stage strategic planning model is to find such a combination of technologies installed at the beginning of each year t , so that the mixture of these technologies operating in each period j would ensure a safe energy “provision” plan minimizing investment costs, possibly deterministic equipment maintenance costs and stochastic operational costs (which may include stochastic fuel prices). Schematically, the structure of the modeling framework is similar to Figure 3.1.

4.2. A Basic Two-stage Dynamic Model

The basic dynamic deterministic model presented in Chapter 3 can be generalized to formulate a stochastic version in a similar manner. For the simplicity of notation only the case in which the evolution of costs CI_i^t , $CO_{i,j}^t$, and demands D_j^t is 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 $\omega = (\omega_1, \omega_2, \dots, \omega_t, \dots, \omega_T)$ 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. There are essential differences between the basic deterministic model (3.1)–(3.6) and its stochastic version. Because strategic decisions x_i^t depend only on t , equations (3.1) and (3.2) remain the same. 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 $\mathbf{x} = x_i^t$, may not yield feasible second stage solutions $\mathbf{y}(\omega) = y_{i,j}^t(\omega)$ satisfying the following equations for all ω :

$$\sum_{i \in \mathcal{I}} y_{i,j}^t(\omega) = D_j^t(\omega), \quad \forall j \in \mathcal{J}, t \in \mathcal{T}, \quad (4.1)$$

$$y_{i,j}^t(\omega) \leq G_{i,j}^t \cdot s_i^t, \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T}, \quad (4.2)$$

$$y_{i,j}^t(\omega) \geq 0, \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T}, \quad (4.3)$$

where

4.2. A Basic Two-stage Dynamic Model

$$s_i^t = s_i^{t-1} + x_i^t - x_i^{t-LT_i}, \quad \forall i \in \mathcal{I}, t \in \mathcal{T}, \quad (4.4)$$

$$x_i^t \geq 0, \quad \forall i \in \mathcal{I}, t \in \mathcal{T}. \quad (4.5)$$

The feasibility of constraints (4.1)–(4.3) 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. Without losing 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:

$$\begin{aligned} \mathbf{F}(x) &= \mathbb{E}_\omega \left[\min_{y(\omega)} \sum_{i \in \mathcal{I}, t \in \mathcal{T}} \left(CI_i^t(\omega) \cdot x_i^t + \right. \right. & (4.6) \\ & \left. \left. \sum_{j \in \mathcal{J}} CO_{i,j}^t(\omega) \cdot DT_j^t \cdot y_{i,j}^t(\omega) \right) \right] = \\ &= \sum_{i \in \mathcal{I}, t \in \mathcal{T}} \left(CI_i^t \cdot x_i^t + \right. \\ & \left. \mathbb{E}_\omega \left[\min_{y(\omega)} \sum_{j \in \mathcal{J}} CO_{i,j}^t(\omega) \cdot DT_j^t \cdot y_{i,j}^t(\omega) \right] \right), \end{aligned}$$

where $\mathbb{E}[\cdot]$ is the expectation function. Thus, CI_i^t is the expected investment cost, $CI_i^t = \mathbb{E}_\omega [CI_i^t(\omega)]$. Because the back-stop technology is available in unlimited amount, $CI_1^t = 0$, $t \in \mathcal{T}$; $CO_{1,j}^t = C$ for all j where C is a large enough positive number. If the optimal $y_{1,j}^t$ is positive for some scenarios ω , then this indicates that there is a risk of installed technologies (excluding $i = 1$) not being able to satisfy demands. Example 5 illustrates this risk and the advantages of stochastic models.

4. Adding Uncertainty: Two-stage Problems

Example 5. (*Back-stop technology and induced risks*). As this simple example shows, the back-stop technology $i = 1$ induces critically important safety constraints connected with a type of Value at Risk (VaR) measures of risks.

It is assumed that $\mathcal{T} = \{1\}$, $\mathcal{J} = \{1\}$, costs $CO_{i,j}^t$ do not depend on ω , costs $CO_{i,j}^t$ include DT_j^t , i.e., $CO_{i,j}^t \rightarrow CO_{i,j}^t \cdot DT_j^t$, and there are only two electricity supplying technologies $\mathcal{I} = \{1, 2\}$, where $i = 1$ is the back-stop technology enabling instantaneous “purchasing” electricity at a high price $CO_{1,1}^1$; and $i = 2$ is a traditional technology that may require investments with expected unit costs $CI_2^1 > 0$, whereas $CI_1^1 = 0$. Demand D_1^1 depends on ω . In this case the model (4.1)–(4.6) is formulated as the minimization of the function

$$\mathbf{F}(x) = CI_2^1 \cdot x_2^1 + \mathbb{E}_\omega \left[\min_{y(\omega)} \left\{ CO_{1,1}^1 \cdot y_{1,1}^1(\omega) + CO_{2,1}^1 \cdot y_{2,1}^1(\omega) \right\} \right],$$

subject to:

$$\begin{aligned} y_{1,1}^1(\omega) + y_{2,1}^1(\omega) &= D_1^1(\omega), \\ s_2^1 &= x_2^1, \\ y_{2,1}^1(\omega) &\leq s_2^1, \\ x_2^1 &\geq 0, \quad y_{1,1}^1(\omega) \geq 0, \quad y_{2,1}^1(\omega) \geq 0, \end{aligned}$$

where $s_2^0 = 0$, $G_{2,1}^1 = 1$. The capacity of the back-stop technology $i = 1$ is assumed to be unlimited, and $CO_{1,1}^1 > CI_2^1 > CO_{2,1}^1$. Therefore, the back-stop technology $i = 1$ ensures the satisfaction of the constraints (4.2) at high operational costs $CO_{1,1}^1 \cdot y_{1,1}^1$.

For any strategic decision x_2^1 and scenario ω there may be situations: $x_2^1 \geq D_1^1(\omega)$ or $x_2^1 < D_1^1(\omega)$. From the structure of costs it follows that the corresponding optimal operational decisions

4.2. A Basic Two-stage Dynamic Model

are $y_{2,1}^{1*}(\omega) = D_1^1(\omega) = x_2^1 - \max\{0, x_2^1 - D_1^1\}$, $y_{1,1}^{1*}(\omega) = 0$ or $y_{2,1}^{1*}(\omega) = x_2^1$, $y_{1,1}^{1*}(\omega) = D_1^1(\omega) - x_2^1 = \max\{0, D_1^1(\omega) - x_2^1\}$. Therefore,

$$\begin{aligned} \mathbf{F}(x) = & CI_2^1 x_2^1 + CO_{2,1}^1 x_2^1 - \\ & CO_{2,1}^1 \mathbb{E} [\max\{0, x_2^1 - D_1^1(\omega)\}] + \\ & CO_{1,1}^1 \mathbb{E} [\max\{0, D_1^1(\omega) - x_2^1\}]. \end{aligned}$$

Only the case in which ω has a continuous probability density, i.e. function $\mathbf{F}(\cdot)$ is continuously differentiable is considered. Otherwise a more complicated analysis must be used. It is clear that an optimal strategic decision x_2^1 is positive, therefore, the optimality condition $\mathbf{F}'(x_2^1) = 0$ yields the following important equations:

$$\begin{aligned} CI_2^1 + CO_{2,1}^1 - CO_{2,1}^1 \cdot \mathbb{P} [D_1^1(\omega) < x_2^1] - \\ CO_{1,1}^1 \cdot \mathbb{P} [D_1^1(\omega) \geq x_2^1] = 0 \Leftrightarrow \end{aligned}$$

$$CI_2^1 + (CO_{2,1}^1 - CO_{1,1}^1) \cdot \mathbb{P} [D_1^1(\omega) \geq x_2^1] = 0 \Leftrightarrow$$

$$\mathbb{P} [D_1^1(\omega) \geq x_2^1] = \frac{CI_2^1}{(CO_{1,1}^1 - CO_{2,1}^1)}, \quad (4.7)$$

where $\mathbb{P}[\cdot]$ is the probability function. Under large enough back-stop unit costs $CO_{1,1}^1$, the value $\frac{CI_2^1}{(CO_{1,1}^1 - CO_{2,1}^1)} < 1$, i.e. the optimal solution satisfies important safety constraints (4.7). Therefore, the use of back-stop technologies induces the safety constraints (4.7). These types of constraints are used in reliability theory, insurance industry, and financial applications. The safety level, i.e., right hand side of equation (4.7), can be regulated by the back-stop unit cost $CO_{1,1}^1$. The simplicity of this example easily illustrates the advantages of using stochastic models. If $D_1^1(\omega)$ is substituted by deterministic average value D_1^1 ,

4. Adding Uncertainty: Two-stage Problems

the optimal solution is $x_2^1 = D_1^1$ and the value of the deterministic objective function is $(CI_2^1 + CO_{2,1}^1)D_1^1$. In fact, its real value is $\mathbf{F}(D_1^1)$ taking into account the variability $D_1^1(\omega)$ of the demand. Definitely, $\mathbf{F}(D_1^1) > \mathbf{F}(x_2^{1*})$, where x_2^{1*} is an optimal solution of the stochastic model, see section 4.4.

4.3. Numerical Methods

The model (4.1)–(4.6) is formulated in the space of variables

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

where the set of scenarios Ω may include a finite number of implicitly given scenarios, e.g., by scenario trees. This often induces models of extremely large size. In this regard, Kaut et al. (2013) propose a method that has been applied in the EnRiMa project to deal with multi-horizon trees. A realistic practical model (4.1)–(4.6) excludes analytically tractable solutions, although the model has an important block-structure that is usually utilized for most effective numerical solutions. In particular, the second stage submodels often have simple solutions as in Example 5. In this case the two stage model is reduced to the optimization in the space of only strategic solutions x_i^t , similar to the optimization of function $\mathbf{F}(\cdot)$ in that Example.

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 (4.1)–(4.6) is formulated as the minimization of the function:

$$\sum_{s \in \mathcal{S}} p_s \left[\sum_{i \in \mathcal{I}, t \in \mathcal{T}} \left(CI_i^t(\omega_s) \cdot x_i^t + \sum_{j \in \mathcal{J}} CO_{i,j}^t(\omega_s) \cdot DT_j^t \cdot y_{i,j}^t(\omega_s) \right) \right], \quad (4.8)$$

subject to:

$$\sum_{i \in \mathcal{I}} y_{i,j}^t(\omega_s) = D_j^t(\omega_s) \quad \forall j \in \mathcal{J}, t \in \mathcal{T}, s \in \mathcal{S}, \quad (4.9)$$

$$y_{i,j}^t(\omega_s) \leq G_{i,j}^t \cdot s_i^t \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T}, s \in \mathcal{S}, \quad (4.10)$$

$$y_{i,j}^t(\omega_s) \geq 0 \quad \forall j \in \mathcal{J}, t \in \mathcal{T}, s \in \mathcal{S}, \quad (4.11)$$

$$s_i^t = s_i^{t-1} + x_i^t - x_i^{t-LT_i} \quad \forall i \in \mathcal{I}, t \in \mathcal{T}, \quad (4.12)$$

$$x_i^t \geq 0 \quad \forall i \in \mathcal{I}, t \in \mathcal{T}. \quad (4.13)$$

Example 6. *Two-stage extension of the basic deterministic model example.* In example 4, it was assumed that in the next years the fuel prices would rise 10% and 3% per year for electricity and natural gas respectively, the cost of technologies would decrease 10% and 5% per year for CHP and PV respectively, and the demand would increase 10% per year. New information about the possible evolution of the market and the demand of the building is available and summarized in Table 4.1. The stochastic optimization problem (4.8)–(4.13) is now solved for all the scenarios ω_s . As a result, the value of the objective function is 62,701 EUR. The optimal values for the strategic first-stage decision variables are displayed in Figure 4.1.

Note that, even though there is no much difference in terms of total cost between the deterministic solution of the first scenario in Example 4, which is the most probable, the strategic decisions, (whose effect in terms of risk might be crucial) that are to be made are different. Moreover, that difference could considerably increase if other scenarios eventually occur, as shown in Table 4.2. In this table, the solution for the most probable scenario (s1) and the stochastic solution are evaluated for each one of the scenarios (columns “Deterministic cost“ and “Stochastic cost”). It is clear that the cost of the stochastic solution for the most probable scenario (s1) is slightly higher than the cost of the deterministic solution. However, if either scenario s2 or s4 (whose probabilities of occurrence are still high) correspond

4. Adding Uncertainty: Two-stage Problems

Table 4.1.: Annual incremental ratios for each scenario

ω_s	p_s	ΔCI_{CHP}	ΔCI_{PV}	ΔCO_{RTE}	ΔCO_{NG}	ΔD
s1	0.24	-0.10	-0.05	0.10	0.03	0.10
s2	0.19	-0.15	-0.10	0.10	0.05	0.05
s3	0.18	-0.15	-0.08	0.12	0.03	0.10
s4	0.20	-0.10	-0.05	0.06	0.03	0.05
s5	0.19	-0.15	-0.08	0.12	0.05	0.10

Table 4.2.: Objective function assessment for each scenario (EUR).

Scenario	Deterministic cost	Stochastic cost	Δ Stochastic
s1	65,212.56	65,352.15	139.59
s2	65,126.75	57,995.54	-7,131.21
s3	66,370.24	66,714.09	343.85
s4	63,000.01	56,567.34	-6,432.66
s5	66,370.24	66,714.09	343.85

to the real situation, the deterministic solution results in much higher costs.

Remark 2. *Two-stage models with rolling horizons.* The initial model is focused on time horizon $[1, T]$. The robust strategic solution can be written as:

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

Solutions $\left(x_i^{1,[1,T]} \right)$, $i \in \mathcal{I}$, are implemented at $t = 1$. This provides a basis for readjustments of scenarios $\omega^{[1,T]}$ perceived at the beginning of time horizon $[1, T]$. Then, a new set of scenarios $\omega^{[2,T]}$ is evaluated, new robust strategic solutions $\left(x_i^{2,[1,T]} \right)$, $i \in \mathcal{I}$, are obtained, and so on. Thus, initially a long-term strategic trajectory

4.4. Value of Stochastic Solution

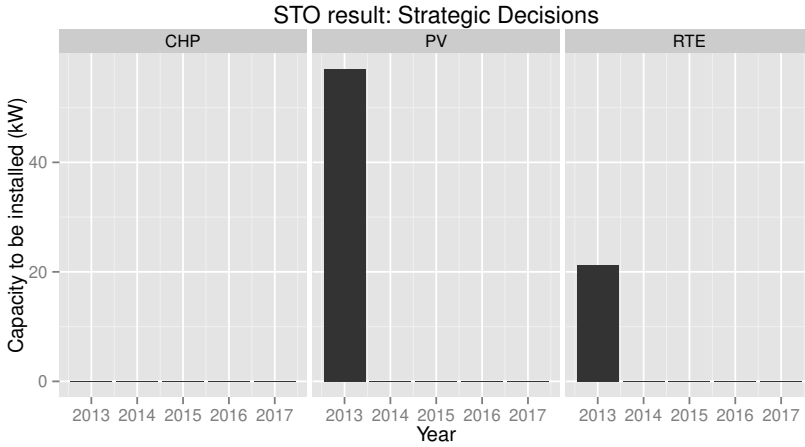


Figure 4.1.: Capacity to be installed for Example 6.

$x^{[1,T]}$ is evaluated, the first time interval solutions $(x_i^{1,[1,T]})$, $i \in \mathcal{I}$, are implemented, new data are received, new scenarios $\omega^{[2,T+1]}$ are adjusted, and so on.

4.4. Value of Stochastic Solution

In this section the value of stochastic optimization models, already introduced in Chapter 2, Subsection 2.2.3, is discussed. It is often termed as the Value of Stochastic Solution (VSS), see for example Birge (1982) or Delage et al. (2012). 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. It has to be emphasized that the VSS or may be better known as the Value of Stochastic Modeling (VSM) is different from the Expected Value of Perfect Information (EVPI), which is defined as the improvement of the objective function by learning perfect information

4. Adding Uncertainty: Two-stage Problems

about parameters of the true deterministic model. In other words, the EVPI is 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 by using real observations. Disadvantages of using deterministic approximations of stochastic models become clear from Example 5. In this example only demand $D_1^1(\omega)$ is uncertain because the stock of back-stop technology is unbounded. In the deterministic version of the model $D_1^1(\omega)$ is replaced by the mean value $D_1^1 = \mathbb{E} [D_1^1(\omega)]$. Accordingly, the stochastic strategic planning model is reduced to the minimization of the cost function

$$CI_2^1 \cdot x_2^1 + CO_{1,1}^1 \cdot y_{1,1}^1 + CO_{2,1}^1 \cdot y_{2,1}^1,$$

subject to:

$$\begin{aligned} y_{1,1}^1 + y_{2,1}^1 &= D_1^1, \\ s_2^1 &= x_2^1, \\ y_2^1 &\leq s_2^1, \\ x_2^1 \geq 0, y_{1,1}^1 \geq 0, y_{2,1}^1 &\geq 0, \end{aligned}$$

that clearly has trivial degenerated solution $x_2^1 = D_1^1$, $y_{1,1}^1 = 0$, $y_{2,1}^1 = D_1^1$. This solution assumes that the future demand is exactly known, therefore the operational decision coincides with the strategic decision. Definitely, this is an unrealistic solution because the average energy demand D_1^1 never occurs in reality. The strategic robust solution of stochastic energy supply model may be significantly different from the mean value as it is indicated by equation (4.7).

In this case, the VSS is calculated by the non-negative difference:

$$\mathbf{F}(x^{*det}) - \mathbf{F}(x^{*sto}), \quad (4.14)$$

4.4. Value of Stochastic Solution

where $\mathbf{F}(x)$ is defined by (4.6), x^{*det} is the optimal solution of the deterministic model (3.1)–(3.6) used in the objective function of the stochastic model, i.e., Equation 4.6, 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 combined with the parameters, e.g., demand, which have been substituted by mean values. This analysis may be rather misleading because sensitivity analysis of the deterministic model with respect to variations of the deterministic model is focused on one only scenario (mean value) that may never occur in reality. As Equation (4.7) demonstrates, the robust solution of the stochastic model depends on the whole probability distribution, 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 good enough (robust) with respect to all potential scenarios. Again, Example 5 nicely illustrates this: any scenario ω_s , $s \in \mathcal{S}$, of the deterministic model has trivial solution $x_2^1 = D_1^1(\omega_s)$, $y_{1,1}^1 = 0$, $y_{2,1}^1 = D_1^1(\omega_s)$, $s \in \mathcal{S}$. The analysis ignores the essential specifics of strategic solutions which have to be made before scenarios ω_s , $s \in \mathcal{S}$, become known. The strategic stochastic optimization (programming) model aims to find solutions robust with respect to all potential scenarios. Equation (4.7) shows that solutions of such models depend on the whole probability distribution, i.e. on all scenarios.

Example 7. *VSS in example 6.* In order to compute the VSS in example 6, expected values of the random variables defined for each stochastic parameter in Table 4.1 have been calculated. Table 4.3 shows the total cost for the solution of the following

4. Adding Uncertainty: Two-stage Problems

models:

- Model 1 is the particular instance of the deterministic optimization problem defined by (3.1)–(3.6) using the expected values described above;
- Model 2 is the deterministic optimization in Example 4 for scenario 1;
- Model 3 is the stochastic optimization in Example 6;
- Model 4 shows the cost of the stochastic optimization, fixing as first-stage decisions those obtained from model 2;
- Model 5, whose solution is infeasible, would correspond to the cost of the stochastic optimization fixing as first-stage decisions those obtained from model 1.

Notice that model 5, i.e., the stochastic optimization problem defined by (4.8)–(4.13) using as first-stage decisions those obtained from the solution of model 1 is infeasible, due to the fact that the solution of model 1 is not feasible for all the scenarios. This result demonstrates by itself the usefulness of the stochastic optimization approach on building energy systems planning. In this case, as $\mathbf{F}(x^{*det})$ is infeasible, it is considered infinite (Birge 1982) and therefore the VSS is infinite:

$$\mathbf{F}(x^{*det}) - \mathbf{F}(x^{*sto}) = \infty - 62,701 = \infty.$$

It is important to remark that even if $\mathbf{F}(x^{*det})$ is feasible, the VSS is positive, and the magnitude will depend of the uncertainty structure. As an example, this difference can be calculated using the deterministic solution obtained for model 2 (the model based on the scenario with the highest probability in Example 4) in order to fix the first-stage decisions. Therefore, following the notation in (4.14), $\mathbf{F}(x^{*det})$ is the cost for model 4 and $\mathbf{F}(x^{*sto})$ is the cost for model 3:

4.4. Value of Stochastic Solution

Table 4.3.: Objective value for the models described.

Model	Description	Total Cost
1	Expected Values Optimization	62,124
2	Deterministic Optimization	65,213
3	Stochastic Optimization (SO)	62,701
4	SO given deterministic solution	62,743
5	SO given expected values	Infeasible

$$\mathbf{F}(x^{*det}) - \mathbf{F}(x^{*sto}) = 62,743 - 62,701 = 42 \text{ EUR.}$$

Definitely, the value $\mathbf{F}(x^{*sto})$ is smaller than $\mathbf{F}(x^{*det})$ because, as already mentioned, the stochastic model has a richer set of feasible solutions, i.e., the deterministic solution x^{*det} is a de-generated version of x^{*sto} .

5. A Comprehensive Deterministic Problem

5.1. Introduction

In this Chapter, a comprehensive deterministic model including more features both strategic and operational is presented. The strategic model has been designed in order to make strategic decisions concerning which technologies to install and/or decommission in the long term, that is, the energy technologies portfolio planning. Besides technologies, this planning includes forward contracts in energy markets. The operational model jointly developed within the EnRiMa project, deals with decisions involving the dispatching of energy through the installed technologies in the short term or through purchases in spot markets, that is, the energy portfolio selection.

In an attempt to tackle short- and long-term decisions simultaneously, the strategic model includes a simplified version of operational energy-balance constraints. The operational model, in turn, includes the realization of the strategic decisions as parameters. Figure 5.1 summarizes the relationship and interaction between both approaches. This thesis focuses on the strategic module, where the optimal values of strategic DVs are the target of the decision maker, that is, decisions on investment on new energy technologies such as DER, renewable energy generators, or batteries among others, as well as long-term energy contracts. These strategic DVs are constrained by strategic constraints, e.g., physical limit for technologies, emissions limits, or budget limits. At that stage, it can be found in the literature strategic optimization models whose energy-balance con-

5. A Comprehensive Deterministic Problem

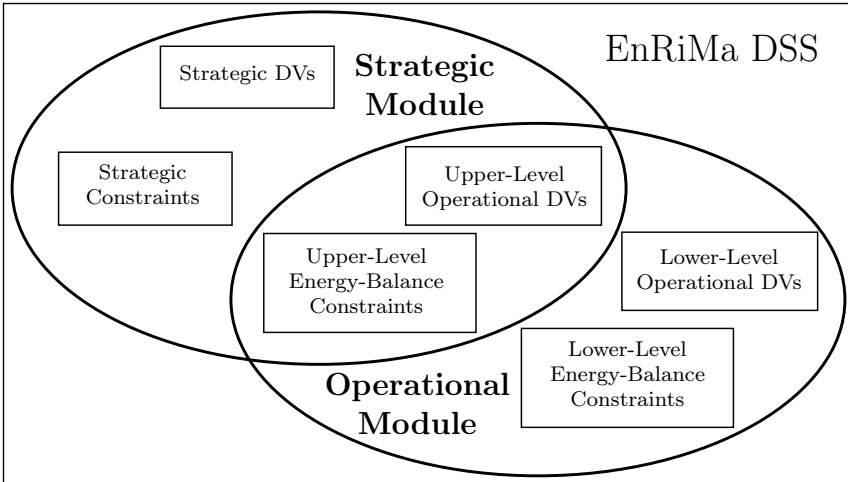


Figure 5.1.: EnRiMa project DSS Schema.

straints (which assures covering the building energy demand) are aggregated in the long term, for example the tool by International Atomic Energy Agency (2001). This approach does not allow the decision maker to take into account the performance of the installed technologies in the short term, leading to optimal solutions that may result in unrealistic implementations. In order to avoid this drawback, the strategic module includes operational DVs that manage the energy flows from inputs to outputs through technologies. Moreover, a simplified version of the energy-balance constraints models the relations among the energy subsystems at an upper level.

Real complex systems such as the ones modeled in the EnRiMa project should also take into account as much relevant information as possible. In this regard, the models to be implemented contain additional information of the systems performance beyond energy dispatching and systems installation, such as technologies decommissioning, technologies aging, energy storing, or passive technologies, i.e., those that save energy. Furthermore, the deterministic model presented in this chapter has been extended to a stochas-

tic optimization formulation, see Chapter 6. Hence, uncertainties about the stochastic parameters, mainly prices and demands, are taken into account, becoming a risk management tool for building managers. All in all, resulting in a holistic DSS for energy efficient building management.

This chapter focuses on the strategic deterministic part, explaining how some of the interrelations described above are tackled, and the time scaling between different time spans. In particular, technologies aging modeling and accounting for operational performance, are novel contributions to the state-of-the-art. Thus, ignoring the aging and therefore the decay of equipment can have consequences for buildings mid- and long-term management. Likewise, this model allows the inclusion of operational parameters in addition to the strategic modeling, which is something new regarding existing proposals. Moreover, the model has been designed so that it is straightforwardly used within a stochastic programming approach in Chapter 6. Finally, a numerical example using data from an EnRiMa project test site and other simulated or publicly available data illustrates the model.

5.2. Time Modeling

In the strategic model, where decisions are made in the long term, e.g., years, operational decisions and constraints are embedded in order to take into account the energy systems performance and decisions in the short-term, e.g., hours. To achieve this goal, instead of including all the possible hours for each year, i.e., 8760, which will likely result in unacceptable computational time, a given number of representative “in-between” mid-term periods are considered, similarly to Conejo et al. (2007). In this way, different parameter values for different performance scenarios can be consider concurrently, e.g., day/night hours, hot/cold seasons, and so on.

The SMS tackles this multiple time resolution including three time sets: \mathcal{P} for long-term periods (usually years), \mathcal{M} for mid-term representative periods, and \mathcal{T} for short-term periods (usually

5. A Comprehensive Deterministic Problem

hours). Decision variables and parameters including the counterpart indices p , m , and t gather this logic. In order to scale between different time resolution terms, the parameters DT and $DM^{p,m}$ contain the time duration for each time span. The value of DT is supposed to be the same during all the decision time horizon. In general cases, operational decisions are made each hour. Then, $DT = 1$. It is used, for example, to compute the maximum amount of energy that a technology can provide during each short-term period. On the other hand, the value of $DM^{p,m}$ can be different for each long- and mid-term periods. For example, a representative mid-term period for $m = \text{'hot season'}$ might have a different number of days one year and another because of weather forecasts and climate predictions. In addition, every mid-term period m may be different to each other. For example a representative mid-term period for $m = \text{'cold season'}$ is likely to have a different number of days than the 'hot season' one.

It is important to remark that the decisions to be actually made after solving the model are only the strategic decisions for the first long-term period ($p = 1$). Nevertheless, a long-term decision horizon is needed in the model in order to consider the systems performance throughout the time, and to allow a long-term objective optimization, e.g. minimize global cost during the next, for example, 25 years.

The set $\mathcal{A} = \{0, \dots, |\mathcal{P}| - 1\}$ is used to model technologies' aging. It is used to allow a variable efficiency of technologies throughout their lifetime (the older, the less effective) and a variable Operation and Maintenance (OM) cost (the older, the more expensive OM cost).

5.3. Technologies Modeling

In the strategic model proposed, both installation and decommissioning of energy systems are considered. As decisions are made throughout a long-term horizon, within each long-term period there is a variable capacity available for generating or storing energy for

5.3. Technologies Modeling

each technology. Technologies are modeled through the \mathcal{I} set, for energy-generation technologies such as CHP, or PV; and the \mathcal{J} set, for storage technologies, such as batteries. Thus, variables represented by si_i^p and xi_j^p are the actual decisions for the decision maker about which technologies to install; sd_i^p , xd_j^p are the decisions for regarding technologies to be decommissioned; and variables represented by s_i^p and x_j^p are used to dynamically calculate the available capacity of each technology $i \in \mathcal{I}$ and $j \in \mathcal{J}$ throughout the decision horizon $p \in \mathcal{P}$:

$$s_i^p = G_i \sum_{0 \leq a' \leq p} AG_i^{p-a'} \left(si_i^{a'} - \sum_{a' < a'' \leq p} sd_i^{a', a''} \right), \quad (5.1)$$

$$x_j^p = GS_j \sum_{0 \leq a' \leq p} AS_j^{p-a'} \left(xi_j^{a'} - \sum_{a' < a'' \leq p} xd_j^{a', a''} \right). \quad (5.2)$$

Equations (5.1) and (5.2) compute the capacity of each energy technology during a long-term period. Note that while the installation variables have only one long-term period index, the decommissioning ones have two, being the former for the installation period, and the latter for the decommissioning period. For example, $sd_{PV}^{2,15}$ means ‘number of PV panels to be decommissioned during the 15th year, from the ones that were installed during the 2nd year’. In addition to the installation-decommissioning flow, these equations reflect another fact influencing the amount of available systems capacity: the effect of aging on the systems capacity through the parameters AG_i^a and AS_j^a .

In general, technologies are considered as devices which can be installed in the building. Each device has a nominal capacity, represented in the SMS with the parameters G_i and GS_j . For example, the building manager can decide how many PV panels to install at the building site and the total capacity of the PV technology is obtained by multiplying the total devices by the nominal capacity of the device. Therefore, the variables si_i^p , $sd_i^{p,q}$, xi_j^p and $xd_j^{p,q}$ are, in general, integer variables, while s_i^p and x_j^p are continuous

5. A Comprehensive Deterministic Problem

variables. In order to reduce the number of integer variables (and so the solution time), technologies that can be considered continuous can be distinguished, i.e., the capacity (kW) to be installed and/or decommissioned can be decided directly, by splitting the set \mathcal{I} into subsets \mathcal{I}_C and \mathcal{I}_D for continuously-sized and discretely-sized energy-generation technologies respectively. Likewise, the set of storage technologies can be partitioned so that $\mathcal{J} = \mathcal{J}_C \cup \mathcal{J}_D$.

Usually, a physical limit must be considered for the number of devices that can be installed as a function of the building characteristics, e.g., the total surface suitable to allocating PV panels. Equations (5.3) and (5.4) model that limit:

$$s_i^p \leq GL_i^p, \quad (5.3)$$

$$x_j^p \leq SL_j^p, \quad (5.4)$$

where the parameters GL_i^p and SL_j^p are a function of the building and each technology physical characteristics. For example, ‘PV panel surface’ vs. ‘building suitable surface for PV (ground + roof)’. Note that the capacity may be different each year.

As far as technologies decommissioning is concerned, there is an obvious constraint that must be set: devices must had been installed before decommissioning them. Equations (5.5) and (5.6) for all i, j, p are in charge of this control:

$$\sum_{q>p} sd_i^{p,q} \leq si_i^p, \quad (5.5)$$

$$\sum_{q>p} xd_j^{p,q} \leq xi_j^p. \quad (5.6)$$

5.4. Strategic modeling

Further strategic constraints to the model are added in order to deal with certain policy or economic restrictions:

5.4. Strategic modeling

$$\sum_{m \in \mathcal{M}} DM^{p,m} \cdot \left(\sum_{i \in \mathcal{I}, t \in \mathcal{T}} \sum_{k \in \mathcal{K}_I} LH_{i,k,l} \cdot y_{i,k}^{p,m,t} + \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}_{BP}, m' \in \mathcal{M}_B} LC_{k,l,n} \cdot u_{k,n}^{p,m,t,m'} \right) \leq PL_l^p. \quad (5.7)$$

Equation (5.7) sets a limit by pollutant $l \in \mathcal{L}$ for the total emissions during each long-term period p . Both on-site and off-site energy production generate pollution. In-site emissions are calculated through the energy used as input to technologies $y_{i,k}^{p,m,t}$ using the parameter $LH_{i,k,l}$, which only applies for the input type of energy (\mathcal{K}_I) and off-site emissions through the energy purchased in the market $u_{k,n}^{p,m,t,m'}$ and the parameter $LC_{k,l,n}$. Note that this constraint and the following one are classified as “strategic” due to their nature and the fact that they apply to long term. However, the decision variables involved are “operational”, i.e., in the short term. The $y_{i,k}^{p,m,t}$ decision variable is the amount of a type of energy that is used as input in an energy-generation technology during each short-term period. The $u_{k,n}^{p,m,t,m'}$ variable is the amount of a type of energy to be purchased in the market during each short-term period. This variable and the $w_{k,n}^{p,m,t,m'}$ one in constraint (5.9) below include two indices for the mid-term period set, namely: m and m' . This feature allows energy trading in forward markets, where energy can be purchased (sold) during mid-term period m' to be delivered during mid-term period m . Note that for spot markets $m' = m$. See Section 5.8 for a discussion on this aspect of the model.

Efficiency constraints are also included in the model:

$$e^{p,m,t} = \sum_{k \in \mathcal{K}, m' \in \mathcal{M}_B} \left(\sum_{n \in \mathcal{N}_{BP}} B_{k,n} \cdot u_{k,n}^{p,m,t,m'} + \sum_{n \in \mathcal{N}_{GNF}} u_{k,n}^{p,m,t,m'} \right), \quad (5.8)$$

5. A Comprehensive Deterministic Problem

$$\sum_{k \in \mathcal{K}, p \in \mathcal{P}, m \in \mathcal{M}, t \in \mathcal{T}} \left(D_k^{p,m,t} + \sum_{n \in \mathcal{N}_S, m' \in \mathcal{M}_S} w_{k,n}^{p,m,t,m'} \right) \geq EF \cdot \sum_{p \in \mathcal{P}, m \in \mathcal{M}, t \in \mathcal{T}} e^{p,m,t}. \quad (5.9)$$

The total primary energy not from a fictitious market, e.g., renewables, consumed by the building during each short-term period is computed by Equation (5.8) which is the sum of the processed energy, $n \in \mathcal{N}_{BP}$, of each type and the one that is used as an input fuel on site, $n \in \mathcal{N}_{GNF}$. Then, Equation (5.9) constraints the global building efficiency during the decision time horizon by setting a minimum efficiency factor required by the building EF . On the left hand side of Equation (5.9), the energy consumed ($D_k^{p,m,t}$) and the energy sold ($w_{k,n}^{p,m,t}$) in spot markets \mathcal{N}_S are added up.

Also budget constraints are considered for the available economic resources for each long-term period:

$$\begin{aligned} & \sum_{i \in \mathcal{I}} \left(\sum_{0 \leq a' \leq p} CI_i^{p,p-a'} \cdot G_i \cdot \left(si_i^p - \sum_{a' < a'' \leq p} sd_i^{a',a''} \right) \right. \\ & \quad \left. + \sum_{0 \leq a' < p} CD_i^{p,p-a'} \cdot sd_i^{a',p} \right) \\ & + \sum_{j \in \mathcal{J}} \left(\sum_{0 \leq a' \leq p} CIS_j^{p,p-a'} \cdot GS_j \cdot \left(xi_j^p - \sum_{a' < a'' \leq p} xd_j^{a',a''} \right) \right. \\ & \quad \left. + \sum_{0 \leq a' < p} CDS_j^{p,p-a'} \cdot xd_j^{a',p} \right) \leq IL. \end{aligned} \quad (5.10)$$

Note that the left hand side of Equation (5.10) is part of the total cost described in Section 5.6. For energy-generation technologies, the investment cost depends on the technology capacity. Thus, $CI_i^{p,0}$ represents the investment cost per kW . For $a > 0$, OM costs

5.5. Operational Constraints

by kWh are also allowed, in addition to OM costs depending on energy production (per kWh) included in operational terms of the objective function, see Section 5.6. Similarly, $CIS_j^{p,0}$ is for the investment costs for storage technologies (in this case, by kWh). On the other hand, decommissioning costs are also considered through the parameters $CD_i^{p,a}$ and $CDS_j^{p,a}$. Note that it is possible to have negative costs if the decommissioned technology can be sold.

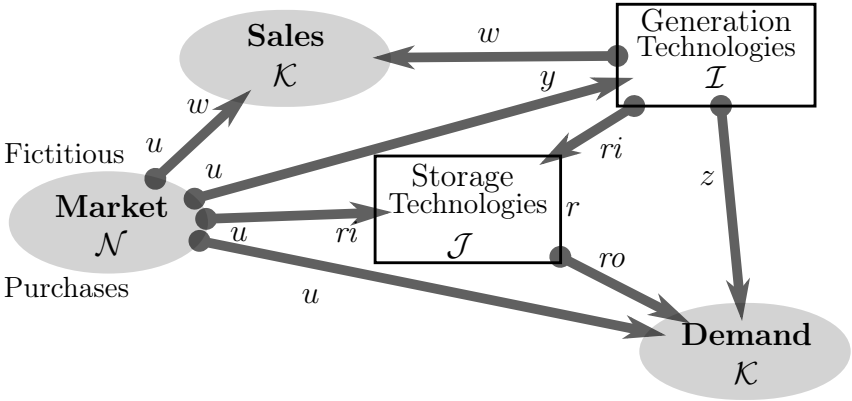
5.5. Operational Constraints

Energy Balance The main equation that models the energy flows is the “energy balance” constraint. This constraint models the energy flow by decision variables that, for each short-term period, amount the energy flowing from markets to loads, through technologies:

$$\begin{aligned}
 \sum_{i \in \mathcal{I}} z_{i,k}^{p,m,t} &+ \sum_{n \in \mathcal{N}_B(k), m' \in \mathcal{M}_B} w_{k,n}^{p,m,t,m'} - \sum_{i \in \mathcal{I}} y_{i,k}^{p,m,t} \\
 &- \sum_{n \in \mathcal{N}_S(k), m' \in \mathcal{M}_S} w_{k,n}^{p,m,t,m'} - \sum_{j \in \mathcal{J}_{Sto}} r_i^{p,m,t} \\
 &= D_k^{p,m,t} - \sum_{j \in \mathcal{J}_{Sto}} rO_{j,k}^{p,m,t}. \tag{5.11}
 \end{aligned}$$

Figure 5.5 represents this energy flow modeling. Variable u decides on how much energy needs to be acquired from the market. In order to take into account renewables, fictitious markets are allowed, e.g., solar radiation, wind. This energy can be directly supplied to the demand side, used to generate energy with energy-generation technologies, stored in storage technologies, or sold again in the market. Conversion rates are applied using the appropriate parameters to model the flows. Thus, y is the energy that actually inputs the energy generator, z is the generated energy to supply the demand, w is the energy sold to the market, and ri and ro decide the input and output energy of the storage respectively. The “inventory” variable r calculates the energy stored during each period.

5. A Comprehensive Deterministic Problem



Equation (5.11) balances the energy flow that meets the demand load. On the left side, the terms add up the energy available for the demand side. This total amount of energy is composed by the total energy generated for the loads z plus the total energy purchased u , minus the energy used by the energy-generation technologies y , the energy sold in the market w , and the energy added to the storage. On the right side, the energy released from the storage is subtracted from the total energy demanded. The equation allows different markets for buying ($\mathcal{N}_B(k)$) and selling ($\mathcal{N}_S(k)$) energy. When only spot markets are considered, $m' = m$. Otherwise, the subsets \mathcal{M}_B and \mathcal{M}_S are to be explicitly defined to reflect the appropriate periods when energy can be bought or sold in advance respectively, see Section 5.8.

Generation Technologies Constraints The energy produced by an energy-generation technology i is determined by the conversion coefficient EC :

$$z_{i,k'}^{p,m,t} = \sum_{k \in \mathcal{K}_{I(i)}} EC_{i,k,k'} \cdot y_{i,k}^{p,m,t}, \quad (5.12)$$

where there are two indexes for the set \mathcal{K} ; k is for the input type of energy, k' is for the output type of energy, and the input type of

5.5. Operational Constraints

energy is obtained through the subset \mathcal{K}_I . Note that usually only one type of energy is possible and the summation can be ignored. Equation (5.12) computes the energy output of each technology and adds up all of the output energy of each energy type during each period.

The energy output of each technology is limited by the capacity installed of each technology:

$$z_{i,k}^{p,m,t} \leq DT \cdot AF_i^{p,m,t} \cdot s_i^p. \quad (5.13)$$

Note that Equation (5.13) is the link between the strategic decisions and the operational systems performance. The availability factor $AF_i^{p,m,t}$ models situations such as intra-day technologies availability (e.g., solar radiation is not available during the whole day) or any other availability issue.

Storage Technologies Constraints In the energy balance equation (5.11) the demand side was diminished through the use of storage technologies. Energy storage is modeled as follows:

$$r_{j,k}^{p,m,t} = OS_{j,k} \cdot r_{j,k}^{p,m,t-1} + OI_{j,k} \cdot ri_{j,k}^{p,m,t-1} - OO_{j,k} \cdot ro_{j,k}^{p,m,t-1}, \quad (5.14)$$

$$ro_{j,k}^{p,m,t} \leq OX_{j,k} \cdot r_{j,k}^{p,m,t}. \quad (5.15)$$

Equation (5.14) computes the energy stored at each period. Between two consecutive periods, the energy stored is affected by the storing ratio OS , representing the portion of energy that remains in the storage between two periods. In addition, the inflows and outflows are calculated by adjusting the energy added to the storage, ri , and the energy released from the storage, ro , using the charging and discharging ratios respectively, OI and OO . Once the amount of energy stored at each short-term period is calculated, Equation

5. A Comprehensive Deterministic Problem

(5.15) adds the constraint for the decision variable ro , i.e., the energy that is released from the storage to supply the demand. This amount of energy is also affected by losses which are undertaken by the parameter OX .

Finally, the link between the strategic decisions and the operational systems performance is modeled. The amount of stored energy is constrained by the available capacity during each long-term period calculated in Equation (5.2). Storage technologies usually need a minimum amount of residual energy stored in order to avoid damages to the battery cells. Similarly, it is advisable not to store the entire capacity of the storage technology for a better systems performance. Thus, in Equation (5.16):

$$OA_{j,k} \cdot x_j^p \leq r_{j,k}^{p,m,t} \leq OB_{j,k} \cdot x_j^p, \quad (5.16)$$

OA and OB adjust the total energy that can be stored, r , within the total capacity installed, x .

5.6. Objective Function

In principle, let us consider that the optimization objective is to minimize the total cost of the energy subsystems throughout the long-term decision horizon. Such a total cost is composed by: (i) the operational cost as a result of the energy dispatching in the short term, i.e., the purchases of primary fuels in the market, and the OM costs; and (ii) the strategic costs, regarding the investment costs in new technologies. The following objective function gathers all the costs stemmed from the energy systems model described above:

$$\sum_{p \in \mathcal{P}} \left(\sum_{i \in \mathcal{I}} \left(\sum_{0 \leq a' \leq p} CI_i^{p,p-a'} \cdot G_i \cdot \left(si_i^p - \sum_{a' < a'' \leq p} sd_i^{a',a''} \right) + \sum_{0 \leq a' < p} CD_i^{p,p-a'} \cdot sd_i^{a',p} \right) \right) \quad (5.17)$$

5.6. Objective Function

$$+ \sum_{j \in \mathcal{J}} \left(\sum_{0 \leq a' \leq p} CIS_j^{p,p-a'} \cdot GS_j \cdot \left(x_j^p - \sum_{a' < a'' \leq p} xd_j^{a',a''} \right) + \sum_{0 \leq a' < p} CDS_j^{p,p-a'} \cdot xd_j^{a',p} \right) \quad (5.18)$$

$$+ \sum_{m \in \mathcal{M}} DM^{p,m} \cdot \sum_{i \in \mathcal{I}, k \in \mathcal{K}, t \in \mathcal{T}} CO_{i,k}^{p,m,t} \cdot z_{i,k}^{p,m,t} \quad (5.19)$$

$$+ \sum_{m \in \mathcal{M}} DM^{p,m} \cdot \sum_{j \in \mathcal{J}, k \in \mathcal{K}, t \in \mathcal{T}} COS_{j,k}^{p,m,t} \cdot r_{j,k}^{p,m,t} \quad (5.20)$$

$$+ \sum_{m \in \mathcal{M}} DM^{p,m} \cdot \sum_{k \in \mathcal{K}, t \in \mathcal{T}} \sum_{n \in \mathcal{N}_B, m' \in \mathcal{M}_B} PP_{k,n}^{p,m,t,m'} \cdot u_{k,n}^{p,m,t,m'} \quad (5.21)$$

$$- \sum_{m \in \mathcal{M}} DM^{p,m} \cdot \sum_{k \in \mathcal{K}, t \in \mathcal{T}} \sum_{n \in \mathcal{N}_S, m' \in \mathcal{M}_S} SP_{k,n}^{p,m,t,m'} \cdot w_{k,n}^{p,m,t,m'} \quad (5.22)$$

$$- \sum_{i \in \mathcal{I}} SU_i^p \cdot G_i \cdot si_i^p \Big). \quad (5.23)$$

Terms (5.17) and (5.18) gather investment costs ($a = 0$) and “fixed” OM costs ($a > 0$) for both types of technologies, energy-generation technologies and energy-storage technologies. Investment costs are usually expressed as unitary costs by kW or kWh installed. So the number of installed devices is multiplied by the nominal capacity and the investment cost by unit. Moreover, lump-sum capital costs are used regardless how investment is paid. Decommissioning costs are also considered. Terms (5.19) and (5.20) compute the operation cost for energy-generation and storage technologies respectively. The more time a technology is operating, the higher the cost. So, the unitary costs are by kWh , and the total cost is obtained by multiplying the unitary cost by the energy generated or stored during each short term period, and scaled to a whole long-term period through the mid-term duration parameter DM . Similarly, term (5.21) computes the total cost for energy purchases, and term (5.22) the total incomes from energy sales in the market. Finally, if there are any subsidies in the market where the DSS is used, term (5.23) subtract the effect from the total cost,

5. A Comprehensive Deterministic Problem

where the parameter SU indicates the amount received by each kW installed.

At the beginning of this section it was remarked that *in principle* the objective is to minimize the total cost. This is the usual situation when the decision maker is a building manager with limited resources and lacking other objectives. More complex decision making processes may prioritize other objectives, such as maximizing the building energy efficiency or minimizing pollutant's emissions. In such a case, it is apparent that it would be enough telling the solver that the objective is an appropriate version of Equation (5.7) or (5.9) when generating the instance from the SMS and the real data. Moreover, multiobjective algorithms can be used in case the decision maker needs a trade-off solution between adverse objectives. Risk mitigation can be also be considered as discussed in Chapter 7 through stochastic programming.

5.7. Numerical Example

In this section, a problem instance is built following the logic of the chapter, defining the sets and parameters of a hypothetical implementation, and eventually solving the problem. For the sake of simplicity no equipment at time $p = 0$ is considered, though the model can be easily extended to allow this fact. The example is based on a real building, specifically the FASAD building in Asturias (Spain). It is a test site for the EnRiMa project whose characteristics and assessment data are defined on the EnRiMa project deliverables D1.1 (HCE et al. 2011) and D2.2 (UCL et al. 2012). Some simplification assumptions have been made for the example clarity.

Example 8. *Comprehensive deterministic example.* Notice that all the referred tables and figures are at the end of the chapter.

Time Let us consider the decision making process of a building manager that, at the end of year 0 (e.g., 2013) has to plan the investment strategy on energy systems for year 1 (e.g., 2014).

5.7. Numerical Example

The decision time horizon considered is 25 years, $\mathcal{P} = \{2014, \dots, 2038\}$. During each year, every single day can be classified into one of the four representative periods $\mathcal{M} = \{\text{winter, spring, summer, fall}\}$. To compute the DM parameter, it is assumed that, in that region, the extreme seasons (i.e., Winter, Summer) last 80 days, and the warm seasons (i.e., Spring, Fall) last 100 days: $DM^{p,\text{spring}} = DM^{p,\text{fall}} = 100$, $DM^{p,\text{winter}} = DM^{p,\text{summer}} = 80$. For short-term periods, a whole day is divided into six periods of four hours ($\mathcal{T} = \{0-4, 4-8, 8-12, 12-16, 16-20, 20-24\}$). Therefore, $DT = 4$ hours.

Technologies Three energy-generation technologies are considered, $\mathcal{I} = \{\text{CHP, PV, WG}\}$, Combined Heat and Power (Dachs 5.5 http://www.baxi-senertec.co.uk/html/baxi_senertec_dachs.htm), Photovoltaic (Solarworld Sunmodule SW 245 <http://www.solarworld.de/en/products/products/solarmodule/overview/>), and Wind Generators (Fortis Passat wind turbine <http://www.fortiswindenergy.com/products/wind-turbines/passaat>), with capacities $G_i = \{5.5, 0.245, 1.4\}$ kW. For all technologies, the aging factors AG_i^a are assumed to be equal to 1 for $a = 0$ and to 0.8 for $a = 24$, at a constant decreasing rate. Emissions and efficiency constraints are neither considered. A physical limit for the number of devices to be installed GL_i^p is set to 4, 1500, and 10 for CHP, PV, and WG respectively.

As for the costs of technologies, investment costs are $CI_i^{1,0} = \{3710, 1327, 5467\}$ EUR/kW (Sources: http://www.mytub.co.uk/product_information.php?product=465447, <http://www.preoc.es/>, and <http://www.allsmallwindturbines.com/>), with a constant decrease of 5% per year. For the example, fixed OM costs are 5% of investment costs per year and decommissioning costs are not taken into account.

Energy Types and Markets The real test site building uses some natural gas-fired boilers with and without heat recovery to

5. A Comprehensive Deterministic Problem

generate different types of heat, namely: space heating, hot water, pool heating, Domestic Hot Water (DHW), and hot water. For the sake of simplicity, aggregated heating demand is considered, that can be met through the CHP unit or directly with a sort of *NG energy*. Electricity is the other type of energy from the demand side. In addition, radiation and wind speed are the input energy types for PV and Wind Generator (WG) respectively. Therefore, $\mathcal{K} = \{\text{electricity, heat, NG, radiation, windspeed}\}$.

Only spot markets are considered (see Section 5.8 for a discussion on this topic). The markets needed for the technologies and types of energy used are $\mathcal{N} = \{\text{RTE, RTG, Sun, Wind}\}$, where the first two are regulated tariffs of electricity and gas, and the last two are fictitious markets. Figure 5.2 represents this simplified building's energy flow. The Electricity bought in the market is used for supplying the electricity demand; let's assume that NG can also supply the building heat demand. Thus, $\mathcal{N}_B(\text{heat}) = \{\text{RTG}\}$ and $\mathcal{N}_B(\text{electricity}) = \{\text{RTE}\}$. Electricity can be also be sold to the grid: $\mathcal{N}_S(\text{electricity}) = \{\text{RTE}\}$.

Operation To set the demand parameter values $D_k^{p,m,t}$, the annual demand of the test site in HCE et al. (2011) is used: 213.50 *MWh* for electricity and 1,553 *MWh* for total heat. To simulate the demand for all the periods, the different representative profiles and operational periods are weighed and add random noise is added. The resulting demand data for the first year is represented in Figure 5.3. For the subsequent years, an increase of 5% and 1% is assumed for electricity and heat demand respectively. The conversion factors $EC_{i,k,k'}$ for each input/output energy of technologies are those in Table 5.1.

As for the availability factor $AF_i^{p,m,t}$, different availability for PV and WG for different short and long-term periods are simulated. Table 5.2 shows estimated availability factors for WG based on its annual yield at 6 m/s (3.2 *MWh*) and the historical average wind speed at the

5.7. Numerical Example

site location (Sources: <http://www.fortiswindenergy.com/products/wind-turbines/passaat>, <http://datosclima.es/Vientostad.php>). Note that more complex analysis should be made in order to better estimate the availability factors for real applications. A similar reasoning (Sources: <http://re.jrc.ec.europa.eu/pvgis/>, <http://www.solarworld.de/>) is applied to compute the availability factor for PV in Table 5.3. More accurate estimations can be made using sophisticated combinations of mid- and short-term periods, or even taking into account predictions in the long term, e.g., climate change.

Objective The objective function defined in section 5.6 is to be optimized. Technologies costs (Investment, OM) have been set above. Regarding energy costs, they are simulated taking as baseline the information in HCE et al. (2011). Note the m' index disappears as only spot markets are considered. Subsidies are neither being considered for the example:

- $PP_{electricity,RTE}^{1,m,t} = 0.1436 \text{ EUR}/kWh,$
- $PP_{NG,RTG}^{1,m,t} = 0.0506 \text{ EUR}/kWh,$
- $SP_{electricity,RTE}^{1,m,t} = 0.1634 \text{ EUR}/kWh,$

and then apply an increase of 5% each year for purchasing prices. As for selling prices, they reduce during the first five years up to meeting the purchasing prices, and then a similar increase to the one for purchases is applied. Random noise is added to those variation coefficients in order to simulate the market prices volatility. Figure 5.4 shows the evolution of these parameters.

Solution The model instance resulting of the above building and market configuration has been solved using the GAMS (<http://www.gams.com>) software version 23.9.5 and the OSIC-PLEX (<http://www.coin-or.org/projects/Osi.xml>) solver. It was composed of 13,429 equations and 37,129 variables and

5. A Comprehensive Deterministic Problem

the execution time was 0.163 seconds in a 64 bits Linux laptop. The R statistical software and programming language (R Core Team 2013a) was used for data preparation, simulation and visualization. Tables 5.4 and 5.5 show the strategic decisions to be made by the building manager regarding the building's energy systems equipment. Note that the decisions that can actually be made are those for $p = 1$. Nevertheless, the whole solution provides more information about the envisaged performance and evolution of the energy systems. Thus, CHP and PV technologies should be installed the first year. The market evolution will prompt PV equipment renovation during the decision time span. Moreover, later on WG technologies deployment will be affordable. Figure 5.5 shows the building's energy systems capacity throughout the whole decision horizon whilst Figure 5.6 represents the installation and decommissioning decisions. This kind of output from a DSS can eventually help building managers to plan their financial positions. As discussed in previous sections, the strategic decisions are supported by the short-term performance of the building energy systems, whose expected outcomes can be visualized in Figure 5.7 (energy purchases), Figure 5.8 (energy sales), Figure 5.9 (energy input to technologies), and Figure 5.10 (energy output from technologies).

5.8. Model Evolution

The model will be extended to a SP framework in Chapter 6. In addition to the scenario tree needed changes, some other changes affecting the energy systems modeling are made in the formulation, which are explained therein. Furthermore, while generating problem instances and numerical examples for the deterministic model, some needed improvements emerged, which are briefly outlined in this subsection.

At the beginning, energy-generation and energy-absorbing technologies (both energy storage and passive measures) were separated

5.8. Model Evolution

in two different sets (\mathcal{I} and \mathcal{J}) because of their differentiated nature: the former transforms one type of energy into another one, the latter do not. Nevertheless, it seems that both types of technologies can be merged into one set, reducing the number of symbols in the model and, therefore, having a simpler deterministic model, easier to be reformulated in the stochastic notation.

Despite the fact that the current configuration of the test sites described in HCE et al. (2011) does not include forward contracts, both UCL et al. (2012) and URJC et al. (2012) include indices and sets in the energy-balance equations and in the SMS which could allow building managers to make decisions for other contracts rather than the ones in spot markets. Even though the equations work under certain conditions, it is difficult to make these conditions compatible with the profiles approach of the mid-term periods, due to the fact that under this approach the \mathcal{M} set is not ordered, and usually not compatible with real trading periods. A practical use of the model with such indices could be applying the strategic model at mid-term time resolution, turning the decision variables concerning technologies availability into parameters with the actual configuration of the building, and filling the \mathcal{M} set with realistic contracting options. This approach could eventually be useful for next-day decisions, such as those regarding ToU tariffs. Otherwise, for strategic long-term decision making a single index will be used as it can be seen in the following chapter.

Table 5.1.: Conversion factors for energy-generation technologies.

i	k	k'	$EC_{i,k,k'}$
CHP	NG	Heat	0.61
CHP	NG	electricity	0.27
PV	Radiation	electricity	1.00
WG	windspeed	electricity	1.00

5. A Comprehensive Deterministic Problem

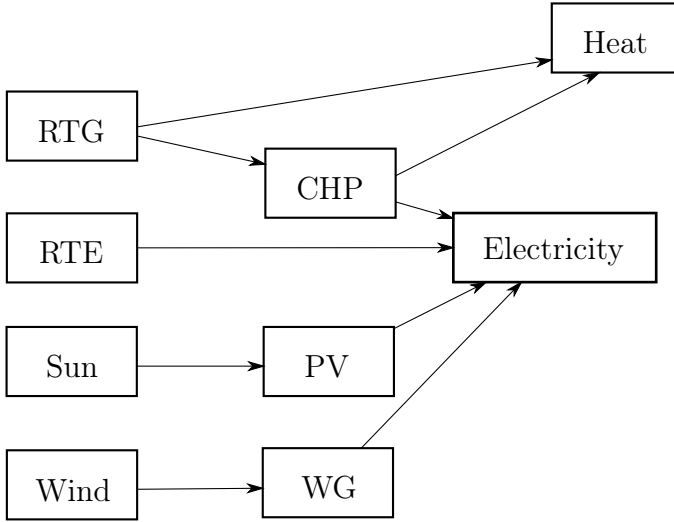


Figure 5.2.: Simplified building systems for the deterministic example.

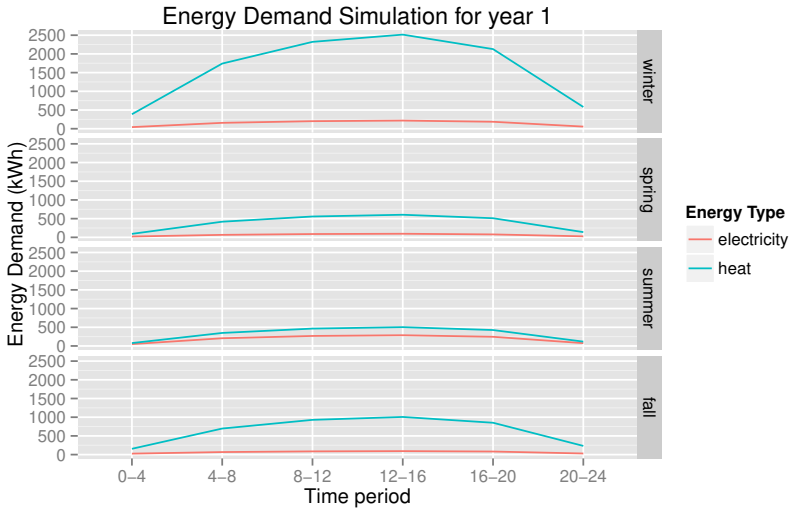


Figure 5.3.: Energy demand.

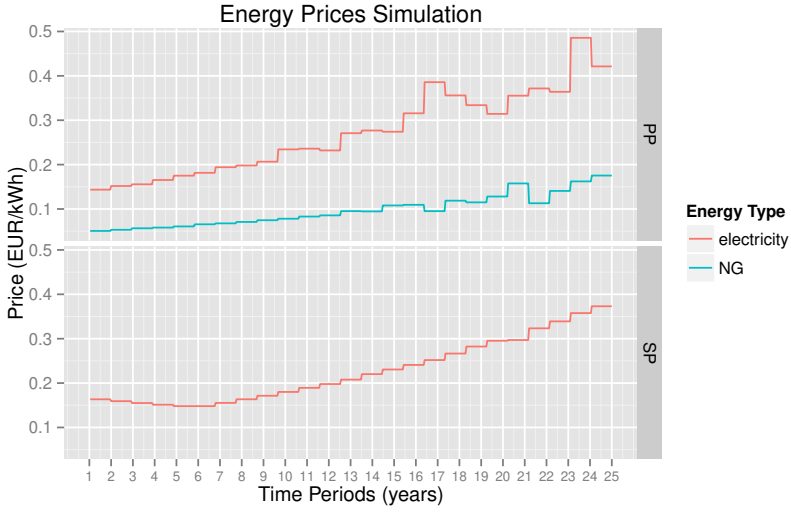


Figure 5.4.: Energy prices.

Table 5.2.: Availability factor for WG technology

t	Avg. Wind speed (m/s)	$AF_{WG}^{p,m,t}$
winter	4.7	0.20
spring	3.9	0.17
summer	5	0.22
fall	7.8	0.34

5. A Comprehensive Deterministic Problem

Table 5.3.: Availability factor for PV technology

t	winter	spring	summer	fall
0-4	0.00	0.00	0.00	0.00
4-8	0.02	0.09	0.08	0.01
8-12	0.42	0.50	0.52	0.37
12-16	0.47	0.53	0.56	0.41
16-20	0.05	0.13	0.12	0.03
20-24	0.00	0.00	0.00	0.00

Table 5.4.: Technologies to install

i	p	units
CHP	1	4
PV	1	1500
PV	10	1500
PV	17	1500
PV	22	1500
WG	11	10.00

Table 5.5.: Technologies decommissioning

i	p	q	units
PV	1	10	1500
PV	10	17	1500
PV	17	22	1500

5.8. Model Evolution

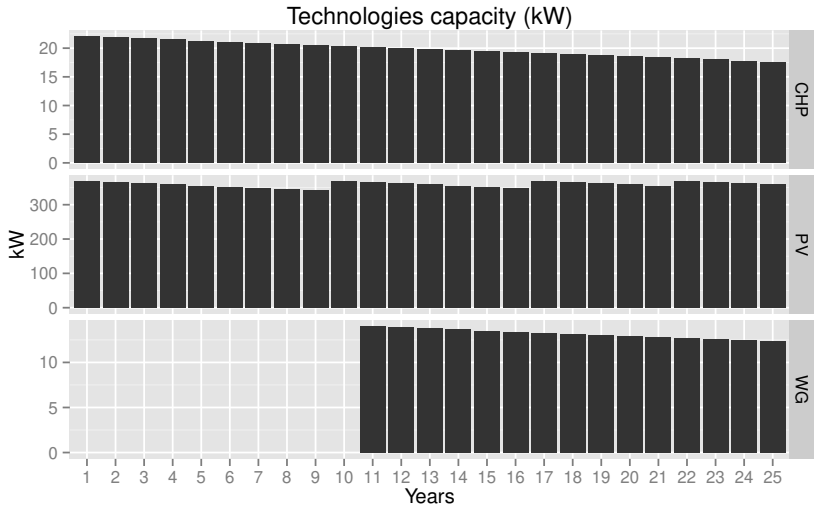


Figure 5.5.: Technologies capacity

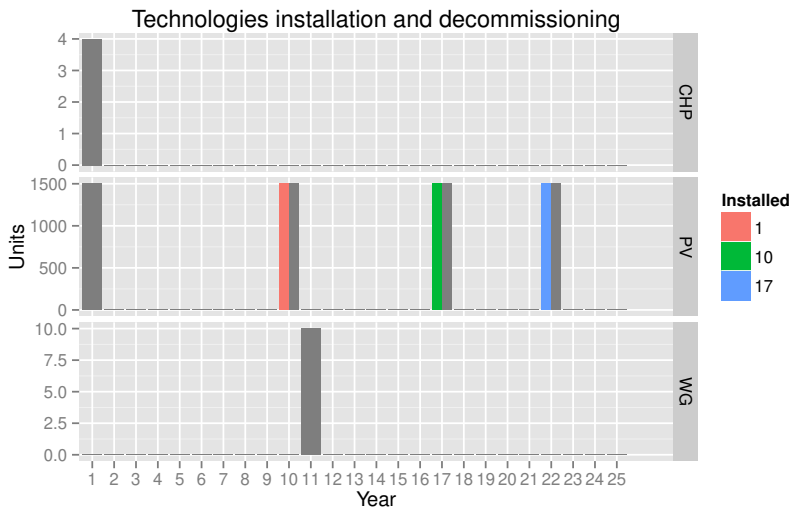


Figure 5.6.: Technologies deployment.

5. A Comprehensive Deterministic Problem

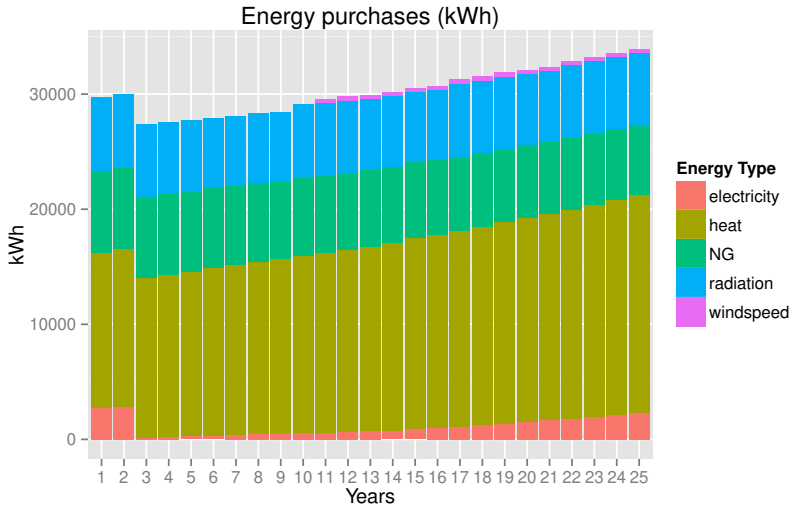


Figure 5.7.: Energy purchasing decisions.

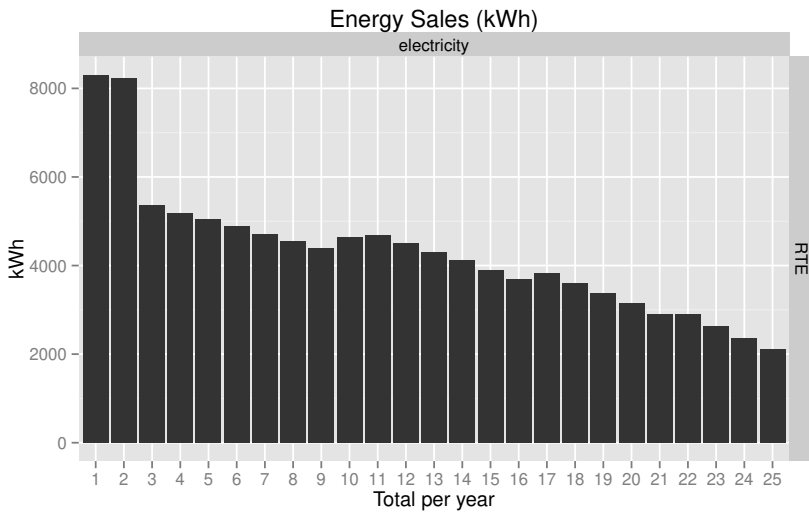


Figure 5.8.: Energy sales decisions.

5.8. Model Evolution

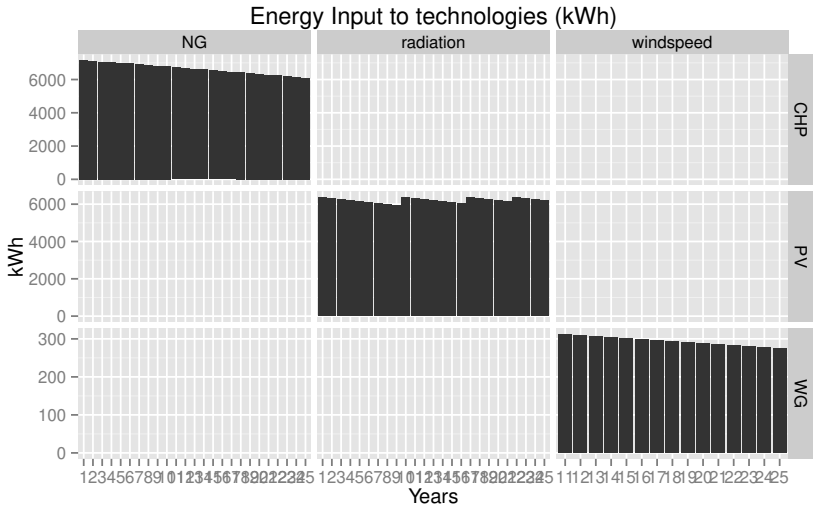


Figure 5.9.: Energy input decisions.

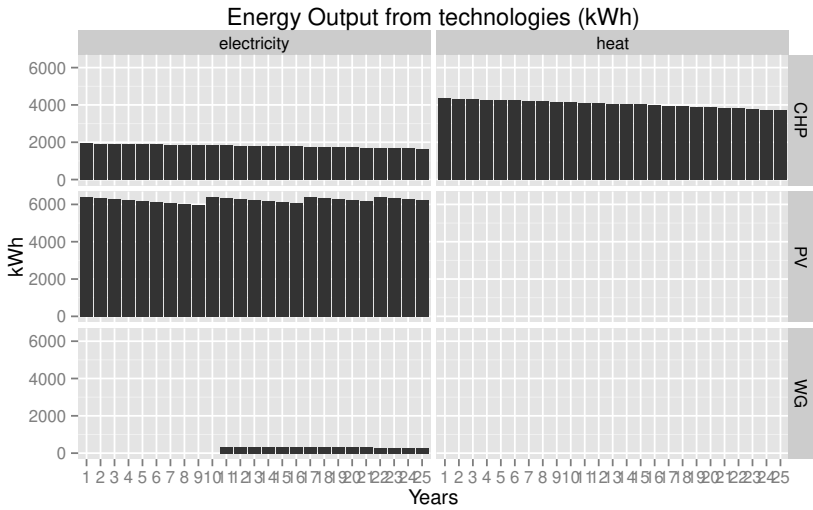


Figure 5.10.: Energy output decisions.

6. Stochastic Multi-stage Model

6.1. Evolution of the SMS

This chapter extends the deterministic model in Chapter 5 to a STO problem that deals with uncertainty, in order to overcome the drawbacks of a deterministic approach using average values for stochastic parameters as explained in Subsection 2.2.3, Chapter 2. Furthermore, a multi-stage approach was finally adopted within the project and uncertainty is modeled through the use of scenario trees gathering the uncertainty throughout the decision horizon. Moreover, a different mathematical notation is used for this complete model. Instead of using the *by-scenario* notation in Chapter 4, a *by-node* notation is followed. As the by-node notation is more compact, it is suitable for large model instances as the foreseen real ones. Moreover, while developing the stochastic model some improvements have been added to the systems modeling. Note that the computation of the VSS is also possible for multi-stage models, see for example Escudero et al. (2007). The new multi-stage approach entails essential differences with respect to the two-stage model developed in Chapter 4. In particular, the nature and interpretation of the first-stage decisions is distinct:

- Two-stage dynamic strategic model. Under this approach, all the strategic variables are first-stage decisions and are made before uncertainty is resolved. The optimal strategic decisions will provide a sort of **trajectory** throughout the long-term periods. All the operational decision variables are second-stage decisions, and are made once uncertainty is unveiled.

6. Stochastic Multi-stage Model

The optimal operational decisions will provide the **adjustments** to the trajectory within each long-term period. The model has a rolling horizon, i.e., at the end of each strategic period it is run again, getting a new trajectory ideally adjusted by the actual operational performance during the period;

- Multi-stage strategic model. Both strategic and operational Decisions to be made before the first tree branching are first-stage decisions. These include the decisions to be made right now, i.e., on the root node, both strategic and operational. Second-stage decisions are those to be made when new information arrives, from the branched node until a new branching, and so on. In this case, there is no trajectory for strategic decisions, but a sort of strategy based on *if-then* rules like in chess game.

In summary, the two-stage model provides a route which leads to the average objective. At each crossroads, i.e., strategic decision point, that route is re-calculated. The multi-stage model provides a roadmap with all the possibilities. This roadmap will be also evaluated again at each crossroads to get a new roadmap. A priori, the multi-stage model is more complex not only in terms of modeling and instance generation, but also in terms of interpretation. See Section 9.3 for future steps on this topic.

6.2. Improvements in Systems Modeling

The deterministic model presented in Chapter 5 contains all the features of a holistic energy system within a building. Nonetheless, during the tests run afterwards and the transformation into a stochastic model, some improvements have been implemented. In this section these improvements and changes are explained.

Changes in nomenclature Some symbols initially set can be incompatible with some reserved words used by the optimization software. Other symbols have been changed just to be a little more informative, without losing the general rules for the representation: small letters for variables and capital letters for parameters. Thus, the following changes apply:

- A new variable (h) has been added for contract selection. This feature substitutes the selection of forward markets through the profile set \mathcal{M} , see Section 9.2.2 in Chapter 9 for further details on this issue;
- Consistently with the previous point, the set \mathcal{N} stands for contracts instead of markets;
- Moreover, a new parameter (ME) has been added to set the maximum power that can be purchased or sold according to the contract. For fictitious contracts limit does not exist;
- A distinction between variables for generation technologies and energy-absorbing technologies is not necessary. Instead, a single variable symbol (x) is used for technologies, and the different types of technologies are defined through subsets. Therefore variables s , si , sd disappear. As a consequence, also parameters non-specific for storage technologies disappear;
- To appropriately compute the available capacity throughout the decision horizon, a new variable is used for the number of units installed of each technology and possible age. Thus, variable x is for the units installed and xc is for the available capacity;
- Availability factor A changes to AF to avoid using the same symbol that set \mathcal{A} ;
- Parameter C changes to LC ;
- Parameter E changes to EC ;

6. Stochastic Multi-stage Model

- Parameter H changes to LH ;
- A new parameter for physical limit of installed technologies has been added (LP), along with a related equation;
- The investment costs and the maintenance costs are now different parameters, which clarifies the notation. Thus, former CI parameter has been split into CI just for investment, and CM for maintenance;
- A new parameter (XZ) has been added in order to consider the existing devices in a building;
- Some of the parameters related to storage change their symbol or description for a better energy systems modeling. This also affects to the equation in charge of calculating the energy stored (inventory), and a new equation has been added to model the charging process. See the details below;
- It was also needed to add a limit for energy to be sold, limiting it to the amount of energy produced using generating technologies. Otherwise, the model became unsolvable for those particular situations in markets where the subsidies are higher than the gap between purchasing and sale energy price.

Another improvement in the formulation is the clarification of the sets' roles. In this regard, the following sets classification is followed, which is linked with the implementation in the optimization software easing the interpretation and reproducibility of the code (see Appendix A for definitions of these and other concepts):

- Normal set;
- Set alias;
- Subset;
- Conditional set.

6.3. Scenario Tree Representation

Conditional sets are used to map the elements of the sets and they are used both in equations' domains and as conditions in sums. The conditional sets are also named multidimensional sets, as they are related at least with two sets. Depending on the type of relationship, the resulting set may contain one single element or several. In one-to-many relationships, there may be several items for only one *input* set. For example, a technology i may have several types of energy output. On the contrary, one-to-one or many-to-one relationships result in single elements. For example, the principal output of a technology is unique (one-to-one). There may be also many-to-many relationships.

6.3. Scenario Tree Representation

As remarked above, a *by-node* notation is adopted, instead of the *by-scenario* notation followed in Chapter 4. Therefore, instead of adding an index for the scenarios, the index for the periods (p) had been substituted by a node index (v), and new data structures for the nodes have been introduced (parent node, period, and probability).

This section contains an explanation of the symbols used for the scenario tree as well as the relationship between the tree nodes and other entities in the model. Scenario trees are widely used in stochastic programming to discretize the huge, usually infinite, number of possible outcomes of the random variables in a stochastic model. Thus, a scenario tree gathers the most probable scenarios resulting from a combination of all random variables (stochastic parameters using the SMS language). Several size-reduction techniques can be used in order to make the problems computationally tractable. See Conejo et al. (2010) for a thorough explanation of scenario trees applied to energy markets.

Note that the stochastic model needs knowledge about the uncertainties, that is to say, the probability distribution of the stochastic parameters. Using this knowledge, a DSS for STO needs an appropriate *scenario generator* in order to generate as many scenarios as

6. Stochastic Multi-stage Model

the optimization software is capable to solve. Scenario generation is out of the scope of this work. Nevertheless, for scenario generation in the examples it has been followed a similar strategy to the approach used in EnRiMa. This strategy relies in the assumption that short-term stochastic parameters vary over the long-term, but can be modeled within each long-term period by means of profiles, using a multi-horizon approach. This scenario generation approach can be consulted in Kaut et al. (2013). Typical scenario trees are implicitly represented within the whole SMS. For the sake of clarity, the scenario trees used in the model are briefly explain next.

A scenario tree can be graphically represented as an acyclic graph consisting of nodes and arcs, where the root node has no parent (predecessor node), each node may have one or more children, and each node can only have one parent. The number of terminal nodes (leaves), which do not have children, determines the number of scenarios considered. Each scenario is a path from the root node to a leaf node. Nodes represent states of the system at a particular time, where decisions are made. The root node corresponds to the beginning of the planning horizon. Arcs represent the precedence relationship between nodes with an associated probability of occurrence. Therefore, in addition to the node identifier (v in the SMS), the following information is required:

- The parent node of each node. It is represented by the conditional set \mathcal{V}_{Pa}^v .
- The probability of each node. It is represented by the parameter PR^v .
- The time period of each node. It is represented by the parameter PT^v .

In addition, the conditional set $\mathcal{A}_{Ages}^{i,v}$ is defined as the possible technologies' ages at a given node for a given technology. For example, if period of node 3 is equal to 3, and at the beginning of the planning horizon a unit of CHP technology whose age was 5 years was already in the building, then the possible ages for that

6.4. Symbolic Model Specification

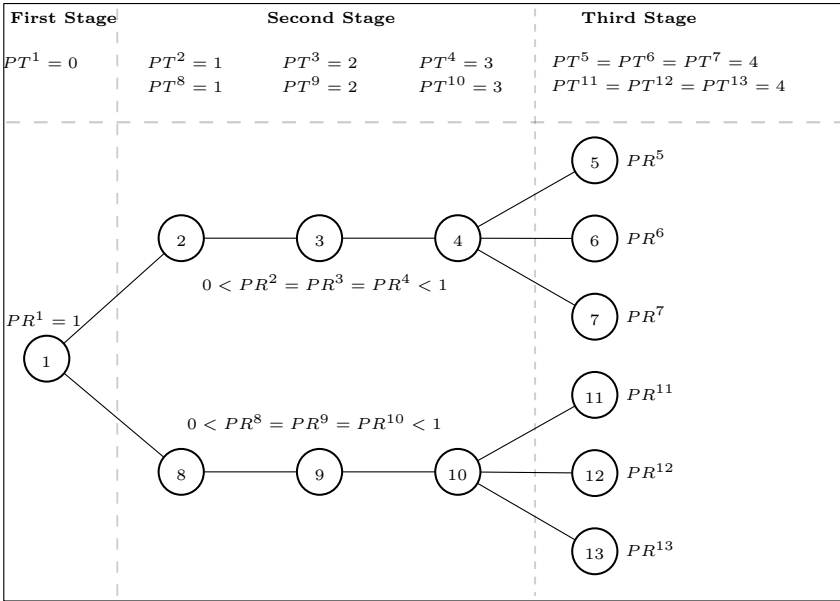


Figure 6.1.: Scenario tree symbolic representation.

technology at node 3 are $\{0,1,2,3,8\}$.

Figure 6.1 shows a simple scenario tree with all the symbols and expressions used in the SMS. Circles represent nodes with the node index (v) displayed inside. Nodes in the same column correspond to the same time period (PT^v). Each node is linked to its parent through the conditional set $\mathcal{V}_{P_a}^v$ and has a probability (PR^v) associated to its parent's branching. The represented tree corresponds to a three-stage stochastic problem, where new information arrives at periods 1 and 4.

6.4. Symbolic Model Specification

In this section the complete SMS of the stochastic multi-stage model is detailed. In the subsequent subsections the following information

6. Stochastic Multi-stage Model

about the mathematical representation of the entities in the model can be found:

- Sets, including standard sets, aliases, subsets, and conditional sets;
- Constants (scalars, in GAMS language);
- Decision variables (DV);
- Parameters;
- Equations, that model the relationship between energy systems.

6.4.1. Sets

The information presented for each set is:

- Symbol. Small Latin letter for sets and aliases, see below for subsets and conditional sets;
- Short description;
- Domain, representing the set by calligraphic capital letters;
- Long description.

Normal sets

a Technology age, $a \in \mathcal{A}$. This set is used to model the effect of aging on the capacity and the costs of technologies.

i Energy technology, $i \in \mathcal{I}$. Equipment available in the building, or suitable to be installed. This equipment can be: (1) Energy generator, (2) Energy storage, or (3) Energy saver. Each element of the set is a specific model of a type of technology (e.g., CHP), with different features.

k Energy type, $k \in \mathcal{K}$. Type of energy that will be used in the building.

6.4. Symbolic Model Specification

- l Type of pollutant, $l \in \mathcal{L}$. Energy generation and consumption generate emissions to the environment. The amount of a building's emissions of each pollutant depends on the emission ratios. The total emissions can be constrained by policy makers. Their minimization can also be an objective for certain decision makers.
- m Operational profile, $m \in \mathcal{M}$. This set gathers the representative profiles considered in the model to link the short- and long-term performance of the energy systems in the building: this short-term performance is scaled to the long term through a weight factor given as a parameter value.
- n Energy tariff, $n \in \mathcal{N}$. This set contains the tariffs available throughout the decision time span. It is possible that not all the tariffs are available at each scenario tree node.
- t Short-term period, $t \in \mathcal{T}$. These are the periods when operational decisions are made. Such decisions concern how much energy of each type must flow through the building energy systems, from markets to demand.
- v Tree node, $v \in \mathcal{V}$. This set contains the nodes in the scenario tree. For each node, its time period (cf., p index in the deterministic model), probability, and parent node must be specified

Sets aliases

- k' Output energy type, $k' \in \mathcal{K}$. This index is used in order to distinguish the input and output typeS of energy when using generators.
- t' Last short-term period, $t' \in \mathcal{T}$. This index is used to balance the energy stored between periods.
- v' Parent nodes, $v' \in \mathcal{V}$. This index is used to map the parent node of each node.

6. Stochastic Multi-stage Model

Subsets

Subsets are represented through the symbol of the set in which they are contained in calligraphic font and the symbol of the subset as subscript in *italics*.

\mathcal{A}_{New} Age = 0, $\mathcal{A}_{New} \subset \mathcal{A}$. This set contains only the element 0 from the age set.

\mathcal{A}_{Old} Age \neq 0, $\mathcal{A}_{Old} \subset \mathcal{A}$. This set contains all the elements except 0 from the age set.

\mathcal{I}_{Con} Continuously-sized technologies, $\mathcal{I}_{Con} \subset \mathcal{I}$. Technologies are continuously sized if they do not have a nominal capacity and the investment can be done by power units.

\mathcal{I}_{Ds} Discretely-sized technologies, $\mathcal{I}_{Ds} \subset \mathcal{I}$. Technologies are discretely sized if they have a nominal capacity and the investment has to be done by devices.

\mathcal{I}_{Gen} Energy-generation technologies, $\mathcal{I}_{Gen} \subset \mathcal{I}$. Technologies that receive energy as input and return other type(s) of energy as output.

\mathcal{I}_{PU} Passive technologies (unitary), $\mathcal{I}_{PU} \subset \mathcal{I}$. Passive technologies which have a multiplicative effect on the demand, that is, the higher the demand, the higher the savings. They entail savings over the use of the energy regardless of the building dimensions.

\mathcal{I}_{Sto} Storage technologies, $\mathcal{I}_{Sto} \subset \mathcal{I}$. Devices that store a type of energy from the market or the generation technologies and can release this energy to meet the demand. These technologies are subject to losses both at the input and at the output. The storage is also uncharged at a given ratio even if energy is not released to be consumed.

\mathcal{K}_{Epur} Energy to purchase, $\mathcal{K}_{Epur} \subset \mathcal{K}$. Types of energy that can be purchased.

6.4. Symbolic Model Specification

\mathcal{K}_{ES} Energy to be sold, $\mathcal{K}_{ES} \subset \mathcal{K}$. Types of energy that can be sold.

\mathcal{N}_{Tpur} Purchasing tariffs, $\mathcal{N}_{Tpur} \subset \mathcal{N}$. This subset contains the tariffs available to buy energy.

\mathcal{N}_{TS} Sales tariffs, $\mathcal{N}_{TS} \subset \mathcal{N}$. This subset contains the tariffs available to sell energy.

\mathcal{V}_{Fut} Future nodes, $\mathcal{V}_{Fut} \subset \mathcal{V}$. All the nodes that are not the root node.

\mathcal{V}_{Root} Root node, $\mathcal{V}_{Root} \subset \mathcal{V}$. This subset only contains the root node, and it is intended to identify states at time 0, for example, existing technologies.

Conditional sets

Conditional (or multidimensional) sets are represented as the symbol of the main set in calligraphic font, the symbol of the conditional set as subscript in *italics* font, and the index of the ‘input’ set(s) as superscript.

$\mathcal{A}_{Ages}^{i,v}$ Possible ages of a technology at a node, $i \in \mathcal{I}, v \in \mathcal{V}$. This conditional set provides all the possible ages that technologies may have at a given node.

\mathcal{K}_{In}^i Input energy types for a technology, $i \in \mathcal{I}_{Gen}$. Generation technologies can utilize different types of energy to generate the output.

\mathcal{K}_{Po}^i Principal energy of technologies, $i \in \mathcal{I}$. Each generation technology has a principal output type of energy (when more than one). For storage technologies, the input and output types of energy are the same. For passive measures, it is the type of energy which is saved.

\mathcal{K}_{Out}^i Output energy types for a technology, $i \in \mathcal{I}_{Gen}$. Generation technologies provide one or more output energy types.

6. Stochastic Multi-stage Model

- \mathcal{N}_{Pur}^k Purchase tariffs for each energy type, $k \in \mathcal{K}$. Conditional set to make purchase tariffs for energy types available.
- \mathcal{N}_{Sal}^k Sales tariffs for each energy type, $k \in \mathcal{K}$. Conditional set to make sales tariffs for energy types available.
- \mathcal{N}_{Tr}^k Contracts under which each type of energy can be actually traded, $k \in \mathcal{K}$. This conditional set is the union of \mathcal{N}_{Pur}^k and \mathcal{N}_{S}^k , and excluding fictitious markets.
- \mathcal{T}_{First}^m First short-term period in a profile, $m \in \mathcal{M}$.
- \mathcal{T}_{Tm}^m Short-term periods by profile, $m \in \mathcal{M}$. Each profile can contain several operational periods, whose duration is modeled through the DM parameter.
- \mathcal{T}_{Last}^m Last short-term period in a profile, $m \in \mathcal{M}$.
- \mathcal{V}_{Pa}^v Parent for each node, $v \in \mathcal{V}$. This conditional set contains the relationship between each node and its parent. Note that the parent node is represented as $Pa(v)$ when it is used as an index in an expression.

6.4.2. Constants

DR Discount rate, per year.

6.4.3. Parameters

For each parameter (applies also to variables), the following information is shown:

- Expression, formed by the parameter/variable symbol and the sets' indices that apply;
- Short description;
- Measurement units;
- Domain for the indices of the expression;

6.4. Symbolic Model Specification

- Long description.

In principle, all the parameters including an index for the node (v) are suitable to be stochastic. The eventual stochastic parameters are decided while generating scenarios, being the values of non-stochastic ones constant throughout the scenario tree and, therefore, the decision time horizon.

$AF_i^{v,m,t}$ Availability factor for a technology (kWh/kWh). $i \in \mathcal{I}$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{T_m}^m$. The capacity of a technology may be different throughout the optimization horizon. For example, photovoltaic panels do not have the same performance during the day and they even do not work during the night. The factor can also be used to model the availability of future technologies.

AG_i^a Technology aging factor (kW/kWh). $i \in \mathcal{I}$, $a \in \mathcal{A}$. This parameter adjusts the total capacity of a technology throughout its lifetime. The superindex is for the age. Thus, at age 0, a given technology (e.g., CHP Dachs 5.5) has factor 1, and it reduces at some rate each year.

$B_{k,n}$ Primary energy needed to produce final-use energy (kWh/kWh). $k \in \mathcal{K}$, $n \in \mathcal{N}_{P_{ur}}^k$. Units of primary energy required to produce one unit of a type of energy available from a market where processed energy can be bought.

$CD_i^{v,a}$ Technology decommissioning cost (EUR/kW). $i \in \mathcal{I}$, $v \in \mathcal{V}$, $a \in \mathcal{A}$. Decommissioning a technology may lead to a removal cost or revenue from selling the equipment (in such a case, the value of the parameter is negative). It can be related to the installation cost.

CI_i^v Technology installation cost (EUR/kW). $i \in \mathcal{I}$, $v \in \mathcal{V}$.

6. Stochastic Multi-stage Model

CM_i^v Technology maintenance cost (EUR/kW, EUR/kWh). $i \in \mathcal{I}$, $v \in \mathcal{V}$. This is a fixed cost per capacity installed. It may be linked to the installation cost.

$CO_{i,k}^v$ Technology operation cost (EUR/kWh). $i \in \mathcal{I}$, $k \in \mathcal{K}_{Out}^i$, $v \in \mathcal{V}$. This parameter is used when the supplier/maintainer's tariff is quoted per operated 'kWh'.

$D_k^{v,m,t}$ Energy demand (kWh). $k \in \mathcal{K}$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. Total energy load of the building for a type of energy, during each short-term (operational) period.

DM^m Weight (scaling factor) for the operational profile (days). $m \in \mathcal{M}$. This parameter is used to scale the operational system performance (energy, cost) to the strategic time resolution.

DT^m Duration of the short-term period within a given profile (hours). $m \in \mathcal{M}$. The sum over the durations of all the operational periods must correspond to a whole day. This parameter is used to convert energy to power or vice versa.

$EC_{i,k,k'}^v$ Output energy generated from one unit of input energy (kWh/kWh). $i \in \mathcal{I}_{Gen}$, $k \in \mathcal{K}_{In}^i$, $k' \in \mathcal{K}_{Out}^i$, $v \in \mathcal{V}$. This is a conversion factor. It is applied to the input energy of a technology, to compute the output energy of this technology. Both types of energy can be the same or different. There may also be several types of output and input energy (e.g., gas, biogas).

EF^v Required building energy efficiency (unitless). $v \in \mathcal{V}$.

6.4. Symbolic Model Specification

G_i Technology capacity (kW/Device). $i \in \mathcal{I}$. Nominal capacity of each device of a given technology. For continuous technologies, its value is 1.

IL^v Investment limit (EUR). $v \in \mathcal{V}$. This is needed when the building has a budget limit for investing on technologies.

$LC_{k,l,n}^v$ Pollution emissions by energy purchases (kg/kWh). $k \in \mathcal{K}$, $l \in \mathcal{L}$, $n \in \mathcal{N}_{Pur}^k$, $v \in \mathcal{V}$. Mean rate of emission of a pollutant from processed energy purchased in the market.

$LH_{k,l}^v$ Pollution emissions by generating technologies (kg/kWh). $k \in \mathcal{K}_{In}^i$, $l \in \mathcal{L}$, $v \in \mathcal{V}$. Amount of pollutant that is emitted by a generation technology during its operation, for each type of input energy.

LP_i^v Physical Limit (Devices/KW/kWh). $i \in \mathcal{I}$, $v \in \mathcal{V}$. Number of units or capacity of a technology that can be installed at the site at a time.

$ME_{k,n}$ Maximum purchase/sale of a type of energy under a given contract (kW). $k \in \mathcal{K}$, $n \in \mathcal{N}_{Tr}^k$.

$OA_{i,k}^v$ Fraction of storage lower limit (kWh/kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$. Minimum fraction of the capacity that must be charged in an energy-storage technology.

$OB_{i,k}^v$ Fraction of storage upper limit (kWh/kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$. Maximum fraction of the capacity that must be charged in an energy-storage technology.

6. Stochastic Multi-stage Model

$OD_{i,k}^v$ Energy demand reduction for a passive technology (kWh/kWh). $i \in \mathcal{I}_{PU}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$. For each unit of a passive technology, the total demand, is reduced by some value.

$OI_{i,k}^v$ Charging ratio to storage (kWh/kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$. Units of energy available for each unit sent to energy-storage technology.

$OO_{i,k}^v$ Discharging ratio from storage (kWh/kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$. Units of energy needed to be discharged from storage in order to deliver one unit of energy to the demand.

$OS_{i,k}$ Energy storage availability (kWh/kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$. This parameter models the loss of a storage technology over the time. It represents the units of energy available for each unit of energy stored after each operational period.

$OX_{i,k}^v$ Max. discharge rate (kW/kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$. Maximum energy discharge rate per unit of storage capacity.

$OY_{i,k}^v$ Max. charge rate (kW/kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$. Maximum energy charge rate per unit of storage capacity.

PL_l^v Pollution limit (kg). $l \in \mathcal{L}$, $v \in \mathcal{V}$. Maximum emissions allowed for the building, at each node.

$PP_{k,n}^{v,m,t}$ Energy purchasing cost (EUR/kWh). $k \in \mathcal{K}$, $n \in \mathcal{N}_{Pur}^k$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. This is the cost of the energy in markets where it can be bought. If there is no ToU tariff, the cost is equal for all operational periods within the same strategic period.

6.4. Symbolic Model Specification

PR^v Probability of the node (unitless). $v \in \mathcal{V}$.

PT^v Time period of the node (unitless). $v \in \mathcal{V}$.

$SP_{k,n}^{v,m,t}$ Energy sales price (EUR/kWh). $k \in \mathcal{K}$, $n \in \mathcal{N}_{Sal}^k$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. For the types of energy that can be sold, there is a price for each operational period.

SU_i^v Subsidies for a technology (EUR/kW). $i \in \mathcal{I}$, $v \in \mathcal{V}$. Policy makers can subsidize the investment of some efficient technologies. Usually an amount per kW is paid.

XZ_i^a Existing devices (Devices/kW/kWh). $i \in \mathcal{I}$, $a \in \mathcal{A}$. Number of existing devices of each technology at the start of the optimization horizon of with a given age.

6.4.4. Variables

Some aggregated computed variables are in the model for convenience. For such aggregated variables long descriptions are omitted as they are implicitly described in their counterpart split variables. In general, short-term decision variables are scaled to yearly aggregated, per node, variables, through the weight parameter. See equations below for a clarification on the role of each variable.

cn^v Total cost at a node (EUR). $v \in \mathcal{V}$. This is a computed variable (see equations).

dn_k^v Total demand at a node (kWh). $k \in \mathcal{K}$, $v \in \mathcal{V}$.

$e^{v,m,t}$ Primary energy consumed per operational period (kWh). $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. This is a computed variable for the

6. Stochastic Multi-stage Model

energy consumption of the building during each short-term period.

en^v Total energy consumed at a node (EUR). $v \in \mathcal{V}$.

ep^v Energy consumed and sold (kWh). $v \in \mathcal{V}$.

et Total averaged energy consumed (EUR). The minimization of this variable can be an objective.

$h_{k,n}^v$ Tariff choice (binary) (unitless). $k \in \mathcal{K}$, $n \in \mathcal{N}_{Tr}^k$, $v \in \mathcal{V}$. This is the decision for selecting among different tariffs. The choice is done for the subsequent period.

mn_i^v Fixed (maintenance) cost at a node (EUR). $i \in \mathcal{I}$, $v \in \mathcal{V}$.

p Total averaged pollutants emissions (EUR). The minimization of this variable can be an objective.

pn_l^v Total emissions at a node (EUR). $l \in \mathcal{L}$, $v \in \mathcal{V}$.

$r_{i,k}^{v,m,t}$ Energy stored (kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. This is an inventory of the amount of energy that is stored in the energy-storage technologies during each short-term period. It is calculated using the operational decisions and the technology parameters.

$ra_{i,k}^v$ Sum of energy stored at the beginning of short term periods (kWh). $i \in \mathcal{I}$, $k \in \mathcal{K}$, $v \in \mathcal{V}$. This aggregated value must be averaged using the total hours in a long-term period in order

6.4. Symbolic Model Specification

to be informative.

rc_i^v Total storage operational cost at a node (kWh). $i \in \mathcal{I}$, $v \in \mathcal{V}$.

$ri_{i,k}^{v,m,t}$ Energy input to storage (kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. Addition to energy storage for each type of energy during each short-term period.

$rn_{i,k}^v$ Total energy input to storage at a node (kWh). $i \in \mathcal{I}$, $k \in \mathcal{K}$, $v \in \mathcal{V}$.

$ro_{i,k}^{v,m,t}$ Energy output from storage (kWh). $i \in \mathcal{I}_{Sto}$, $k \in \mathcal{K}_{Po}^i$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. Release from each energy-storage technology of each type of energy during each operational period.

$rp_{i,k}^v$ Sum of energy output from storage at a node (kWh). $i \in \mathcal{I}$, $k \in \mathcal{K}$, $v \in \mathcal{V}$. This aggregated value must be averaged using the total hours in a long-term period in order to be informative.

sn_i^v Strategic (investment) cost at a node for a technology (EUR). $i \in \mathcal{I}$, $v \in \mathcal{V}$.

$u_{k,n}^{v,m,t}$ Energy to purchase, under a given tariff (kWh). $k \in \mathcal{K}$, $n \in \mathcal{N}_{Pur}^k$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. Energy purchased in the market for each type of energy, to be delivered during each operational period.

6. Stochastic Multi-stage Model

$uc_{k,n}^v$ Total energy purchases costs at a node for each energy type and contract (EUR). $k \in \mathcal{K}$, $n \in \mathcal{N}$, $v \in \mathcal{V}$.

$un_{k,n}^v$ Total energy purchases at a node for each energy type and contract (kWh). $k \in \mathcal{K}$, $n \in \mathcal{N}$, $v \in \mathcal{V}$.

$w_{k,n}^{v,m,t}$ Energy to sell, under a given tariff (kWh). $k \in \mathcal{K}$, $n \in \mathcal{N}_{Sal}^k$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. Energy of each type of energy to be sold in the market during each operational period.

$wc_{k,n}^v$ Total energy sales income at a node for each energy type and contract (EUR). $k \in \mathcal{K}$, $n \in \mathcal{N}$, $v \in \mathcal{V}$.

$wn_{k,n}^v$ Total energy sales at a node for each energy type and contract (kWh). $k \in \mathcal{K}$, $n \in \mathcal{N}$, $v \in \mathcal{V}$.

$x_i^{v,a}$ Installed units of a given age for each technology and node (Devices/kW/kWh). $i \in \mathcal{I}$, $v \in \mathcal{V}$, $a \in \mathcal{A}$. This is a computed variable.

xc_i^v Available capacity of a technology at each node (kW or kWh (storage)). $i \in \mathcal{I}$, $v \in \mathcal{V}$. This capacity is computed through the decisions and the parameters.

$xd_i^{v,a}$ Number of units of a technology to be decommissioned (Devices or kW). integer for $i \in \mathcal{I}_{Ds}$, $v \in \mathcal{V}$, $a \in \mathcal{A}_{Old}$; continuous for $i \in \mathcal{I}_{Con}$, $v \in \mathcal{V}$, $a \in \mathcal{A}_{Old}$. For continuously-sized technologies, this is the total capacity to be decommissioned. For discretely-sized technologies, it denotes number of devices to decommission.

6.4. Symbolic Model Specification

x_i^v Number of units of a technology to be installed (Devices or kW). integer for $i \in \mathcal{I}_{Ds}$, $v \in \mathcal{V}$; continuous for $i \in \mathcal{I}_{Con}$, $v \in \mathcal{V}$. For discretely-sized technologies, this is an integer variable, whilst for continuously-sized technologies, it is a continuous one.

$y_{i,k}^{v,m,t}$ Energy generator input (kWh). $i \in \mathcal{I}_{Gen}$, $k \in \mathcal{K}_{In}^i$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. Amount of energy used as input to an energy-creating technology, for each type of energy, operational profile and period.

$yn_{i,k}^v$ Total energy input at a node for each technology and type of energy (kWh). $i \in \mathcal{I}$, $k \in \mathcal{K}$, $v \in \mathcal{V}$.

$z_{i,k}^{v,m,t}$ Energy generator output (kWh). $i \in \mathcal{I}_{Gen}$, $k \in \mathcal{K}_{Out}^i$, $v \in \mathcal{V}$, $m \in \mathcal{M}$, $t \in \mathcal{T}_{Tm}^m$. Amount of energy as output from an energy-creating technology, for each type of energy, operational profile and period.

zc_i^v Total energy generation (operation) costs at a node for each technology (EUR). $i \in \mathcal{I}$, $v \in \mathcal{V}$.

$zn_{i,k}^v$ Total energy generated at a node for each technology and energy type (kWh). $i \in \mathcal{I}$, $k \in \mathcal{K}$, $v \in \mathcal{V}$.

6.4.5. Equations

The following equations model the relationship between the energy systems in the building along with the energy flows. They are classified into the following groups:

1. **Strategic.** These equations deal with strategic decisions, such as investments and contracts;

6. Stochastic Multi-stage Model

2. **Operational.** These equations deal with operational decisions, such as energy trade, generation, or storage;
3. **Strategic-operational links.** These equations link the operational performance with the strategic decisions or policies;
4. **Computations.** These equations are for computing auxiliary variables in order to simplify the largest equations as well as allowing the reuse of such auxiliary variables in different equations;
5. **Aggregations.** These equations aggregate values of operational decisions or variables per node scaling through the profile features. Those values are useful for visualization and post-analysis purposes;
6. **Objectives.** In this model, three magnitudes can be minimized: total cost, total emissions, or total energy consumed. Once selected the objective, the other two equations become constraints to compute the overall values.

Each equation is preceded by a short title and a brief description. Nonetheless this description is omitted for some computed variables for aggregation by node as the description is redundant.

(1) Strategic

Available new technologies (devices) at each node

The available new devices of a technology (age zero) are equal to the ones installed at each node:

$$x_i^{v,a} = x_i^{v} \quad \forall i \in \mathcal{I}, v \in \mathcal{V}, a \in \mathcal{A}_{New}. \quad (6.1)$$

Available old technologies (devices) at future nodes

The available devices whose age is not zero are equal to the ones available at the previous node, minus the number of decommissioned ones:

$$x_i^{v,a} = x_i^{v',a-1} - x d_i^{v,a} \quad (6.2)$$

$$\forall i \in \mathcal{I}, a \in \mathcal{A}_{Old}, v \in \mathcal{V}_{Fut}, a \in \mathcal{A}_{Ages}^{i,v}, v' \in \mathcal{V}_{Pa}^v.$$

6.4. Symbolic Model Specification

Available old technologies (devices) at root node

For technologies existing before the start of the optimization horizon, the number of devices available at the root node is equal to the number of existing devices minus those decommissioned at the beginning of the first period:

$$x_i^{v,a} = XZ_i^a - xd_i^{v,a} \quad (6.3)$$

$$\forall i \in \mathcal{I}, a \in \mathcal{A}_{Old}, v \in \mathcal{V}_{Root}, a \in \mathcal{A}_{Ages}^{i,v}.$$

Technology capacity calculation

The total capacity of a technology is the sum of the capacities of the installed devices at any age, corrected by the aging factors and nominal capacity:

$$xc_i^v = G_i \cdot \sum_{a \in \mathcal{A}_{Ages}^{i,v}} AG_i^a \cdot x_i^{v,a} \quad \forall i \in \mathcal{I}, v \in \mathcal{V}. \quad (6.4)$$

Investment limit

An upper limit may be imposed on the total installation, and decommissioning cost:

$$\sum_{i \in \mathcal{I}} sn_i^v \leq IL^v \quad \forall v \in \mathcal{V}. \quad (6.5)$$

Purchase tariff choice

Only one purchase tariff is allowed per energy type:

$$\sum_{n \in \mathcal{N}_{Pur}^k} h_{k,n}^v = 1 \quad \forall v \in \mathcal{V}, k \in \mathcal{K}_{Epur}. \quad (6.6)$$

Sales tariff choice

Only one sales tariff is allowed per energy type:

$$\sum_{n \in \mathcal{N}_{Sal}^k} h_{k,n}^v = 1 \quad \forall v \in \mathcal{V}, k \in \mathcal{K}_{ES}. \quad (6.7)$$

6. Stochastic Multi-stage Model

Physical limit

There is a limit for installing technologies, usually established by the space available in the site. :

$$\sum_{a \in \mathcal{A}_{Agés}^{i,v}} x_i^{v,a} \leq LP_i^v \quad \forall i \in \mathcal{I}, v \in \mathcal{V}. \quad (6.8)$$

(2) Operational

Storage available

The energy stored at the beginning of each period is the energy stored in the previous period, plus the energy sent to storage, minus the energy released from storage. All flows are corrected by their respective loss ratio parameter:

$$r_{i,k}^{v,m,t+1} = OS_{i,k} \cdot r_{i,k}^{v,m,t} + OI_{i,k}^v \cdot r_{i,k}^{v,m,t} - OO_{i,k}^v \cdot ro_{i,k}^{v,m,t} \quad (6.9)$$

$$\forall v \in \mathcal{V}, m \in \mathcal{M}, i \in \mathcal{I}_{Sto}, t \in \mathcal{T}_{Tm}^m, k \in \mathcal{K}_{Po}^i.$$

Storage level between periods

The storage level at the first short-term period must be equal to the level at the final period (in the same strategic period):

$$r_{i,k}^{v,m,t} = OS_{i,k} \cdot r_{i,k}^{v,m,t'} + OI_{i,k}^v \cdot r_{i,k}^{v,m,t'} - OO_{i,k}^v \cdot ro_{i,k}^{v,m,t'} \quad (6.10)$$

$$\forall v \in \mathcal{V}, m \in \mathcal{M}, i \in \mathcal{I}_{Sto}, k \in \mathcal{K}_{Po}^i, t \in \mathcal{T}_{First}^m, t' \in \mathcal{T}_{Last}^m.$$

Energy balance

The energy supplied must meet the energy demand minus the energy saved due to passive technologies. It is composed of the energy produced with energy-creating technologies plus the energy purchased in the market minus the energy sold, energy for storage and energy for production. On the demand side, the energy saved with passive technologies diminish the

6.4. Symbolic Model Specification

total demand:

$$\begin{aligned}
 & \sum_{i \in \mathcal{I}_{Gen}} z_{i,k}^{v,m,t} - \sum_{i \in \mathcal{I}_{Gen}} y_{i,k}^{v,m,t} \\
 & + \sum_{n \in \mathcal{N}_{Pur}^k} u_{k,n}^{v,m,t} - \sum_{n \in \mathcal{N}_{Sal}^k} w_{k,n}^{v,m,t} \\
 & + \sum_{i \in \mathcal{I}_{Sto}} \left(ro_{i,k}^{v,m,t} - ri_{i,k}^{v,m,t} \right) \\
 & = D_k^{v,m,t} \cdot \left(1 - \sum_{i \in \mathcal{I}_{PV}} OD_{i,k}^v \cdot xc_i^v \right) \\
 & \forall k \in \mathcal{K}, v \in \mathcal{V}, m \in \mathcal{M}, t \in \mathcal{T}_{Tm}^m.
 \end{aligned} \tag{6.11}$$

Sales limit by generation

The energy sold can not be greater than the energy produced:

$$\begin{aligned}
 \sum_{n \in \mathcal{N}_{Sal}^k} w_{k,n}^{v,m,t} & \leq \sum_{i \in \mathcal{I}_{Gen}} z_{i,k}^{v,m,t} \\
 \forall v \in \mathcal{V}, m \in \mathcal{M}, k \in \mathcal{K}_{ES}, t \in \mathcal{T}_{Tm}^m.
 \end{aligned} \tag{6.12}$$

(3) Strategic-operational links

Technology output limit

The energy that can be supplied by a technology is constrained by the availability of the technology and its capacity:

$$\begin{aligned}
 z_{i,k}^{v,m,t} & \leq DT^m \cdot AF_i^{v,m,t} \cdot xc_i^v \\
 \forall v \in \mathcal{V}, m \in \mathcal{M}, i \in \mathcal{I}_{Gen}, k \in \mathcal{K}_{Po}^i, t \in \mathcal{T}_{Tm}^m.
 \end{aligned} \tag{6.13}$$

Storage release limit

The amount of energy that can be discharged from any energy-storage technology is limited by the installed capacity and the maximum discharge rate:

$$\begin{aligned}
 ro_{i,k}^{v,m,t} & \leq OX_{i,k}^v \cdot DT^m \cdot xc_i^v \\
 \forall v \in \mathcal{V}, m \in \mathcal{M}, i \in \mathcal{I}_{Sto}, t \in \mathcal{T}_{Tm}^m, k \in \mathcal{K}_{Po}^i.
 \end{aligned} \tag{6.14}$$

6. Stochastic Multi-stage Model

Storage charge limit

The amount of energy that can be charged to any energy-storage technology is limited by the installed capacity and the maximum charge rate:

$$r_{i,k}^{v,m,t} \leq OY_{i,k}^v \cdot DT^m \cdot xc_i^v \quad (6.15)$$

$$\forall v \in \mathcal{V}, m \in \mathcal{M}, i \in \mathcal{I}_{Sto}, t \in \mathcal{T}_{Tm}^m, k \in \mathcal{K}_{Po}^i.$$

Lower storage limit

The amount of energy that can be stored in any energy-storage technology must be greater than the capacity installed, corrected by the minimum charge required:

$$r_{i,k}^{v,m,t} \geq OA_{i,k}^v \cdot xc_i^v \quad (6.16)$$

$$\forall v \in \mathcal{V}, m \in \mathcal{M}, i \in \mathcal{I}_{Sto}, t \in \mathcal{T}_{Tm}^m, k \in \mathcal{K}_{Po}^i.$$

Upper storage limit

The amount of energy that can be stored in any energy-storage technology must be lower than the capacity installed, corrected by the maximum charge allowed:

$$r_{i,k}^{v,m,t} \leq OB_{i,k}^v \cdot xc_i^v \quad (6.17)$$

$$\forall v \in \mathcal{V}, m \in \mathcal{M}, i \in \mathcal{I}_{Sto}, t \in \mathcal{T}_{Tm}^m, k \in \mathcal{K}_{Po}^i.$$

Purchasing limit by contract

The amount of energy that can be purchased at a given node must not exceed the amount stipulated in the previously signed contract:

$$u_{k,n}^{v,m,t} \leq h_{k,n}^v \cdot ME_{k,n} \cdot DT^m \quad (6.18)$$

$$\forall v \in \mathcal{V}, m \in \mathcal{M}, k \in \mathcal{K}_{Epur}, n \in \mathcal{N}_{Pur}^k, t \in \mathcal{T}_{Tm}^m.$$

Sales limit by contract

The amount of energy that can be sold at a given node must

6.4. Symbolic Model Specification

not exceed the amount stipulated in the previously signed contract:

$$w_{k,n}^{v,m,t} \leq h_{k,n}^v \cdot ME_{k,n} \cdot DT^m \quad (6.19)$$

$$\forall v \in \mathcal{V}, m \in \mathcal{M}, k \in \mathcal{K}_{ES}, n \in \mathcal{N}_{Sal}^k, t \in \mathcal{T}_{Tm}^m.$$

Required efficiency

The amount of energy consumed and sold must be larger than the amount of primary energy consumed corrected by the efficiency parameter:

$$ep^v \geq EF^v \cdot en^v \quad \forall v \in \mathcal{V}. \quad (6.20)$$

Emissions limit

The total emissions of each considered pollutant cannot exceed a specified limit:

$$pn_l^v \leq PL_l^v \quad \forall l \in \mathcal{L}, v \in \mathcal{V}. \quad (6.21)$$

(4) Computations

Strategic investment in a technology at a node

Strategic investments include investment costs reduced by the subsidies, and decommissioning costs:

$$sn_i^v = (CI_i^v - SU_i^v) \cdot G_i \cdot xi_i^v + \sum_{a \in \mathcal{A}_{Ages}^{i,v}} CD_i^{v,a} \cdot G_i \cdot xd_i^{v,a} \quad (6.22)$$

$$\forall i \in \mathcal{I}, v \in \mathcal{V}.$$

Output energy calculation

The amount of output energy is calculated from the input energy and the conversion factor:

$$z_{i,k'}^{v,m,t} = \sum_{k \in \mathcal{K}_{In}^i} EC_{i,k,k'}^v \cdot y_{i,k}^{v,m,t} \quad (6.23)$$

$$\forall v \in \mathcal{V}, m \in \mathcal{M}, i \in \mathcal{I}_{Gen}, k' \in \mathcal{K}_{Out}^i, t \in \mathcal{T}_{Tm}^m.$$

6. Stochastic Multi-stage Model

Primary energy calculation

The primary energy (not from a fictitious market) consumed is the sum of the processed energy of each type and the energy used as an input fuel:

$$e^{v,m,t} = \sum_{k \in \mathcal{K}, n \in \mathcal{N}_{Pur}^k} B_{k,n} \cdot u_{k,n}^{v,m,t} \quad (6.24)$$

$$\forall v \in \mathcal{V}, m \in \mathcal{M}, t \in \mathcal{T}_{Tm}^m.$$

Total cost at a node

Total cost includes investment and decommissioning cost, maintenance cost, and operational cost. The latter includes purchases, generation and storage operational cost, and is diminished by the income from energy sales:

$$cn^v = \sum_{i \in \mathcal{I}} sn_i^v + \sum_{i \in \mathcal{I}} mn_i^v \quad (6.25)$$

$$+ \sum_{k \in \mathcal{K}, n \in \mathcal{N}_{Pur}^k} uc_{k,n}^v - \sum_{k \in \mathcal{K}, n \in \mathcal{N}_{Sal}^k} wc_{k,n}^v$$

$$+ \sum_{i \in \mathcal{I}_{Gen}} zc_i^v + \sum_{i \in \mathcal{I}_{Sto}} rc_i^v \quad \forall v \in \mathcal{V}.$$

Energy consumed and sold calculation

This total output energy from the building, regardless its destination, is used to compute the efficiency of the building:

$$ep^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{k \in \mathcal{K}, t \in \mathcal{T}_{Tm}^m} \left(D_k^{v,m,t} + \sum_{n \in \mathcal{N}_{Sal}^k} w_{k,n}^{v,m,t} \right) \quad (6.26)$$

$$\forall v \in \mathcal{V}.$$

Emissions of a pollutant at a node calculation

Total emissions come from both energy input to generation

6.4. Symbolic Model Specification

technologies and energy purchases in the market:

$$pn_i^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in T_{Tm}^n} \left(\sum_{k \in \mathcal{K}_{In}^i} LH_{k,l}^v \cdot y_{i,k}^{v,m,t} \right. \\ \left. + \sum_{k \in \mathcal{K}, n \in \mathcal{N}_{Pur}^k} LC_{k,l,n}^v \cdot u_{k,n}^{v,m,t} \right) \quad (6.27) \\ \forall l \in \mathcal{L}, v \in \mathcal{V}.$$

(5) Aggregations

Aggregated maintenance cost at each node

$$mn_i^v = CM_i^v \cdot xc_i^v \quad \forall i \in \mathcal{I}, v \in \mathcal{V}. \quad (6.28)$$

Aggregated energy purchases for a type of energy at a node

$$un_{k,n}^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in T_{Tm}^n} u_{k,n}^{v,m,t} \quad (6.29) \\ \forall k \in \mathcal{K}, v \in \mathcal{V}, n \in \mathcal{N}_{Pur}^k.$$

Energy purchases aggregate cost at a node

$$uc_{k,n}^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in T_{Tm}^n} PP_{k,n}^{v,m,t} \cdot u_{k,n}^{v,m,t} \quad (6.30) \\ \forall k \in \mathcal{K}, v \in \mathcal{V}, n \in \mathcal{N}_{Pur}^k.$$

Aggregated energy sales for each type of energy at a node

6. Stochastic Multi-stage Model

$$wn_{k,n}^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} w_{k,n}^{v,m,t} \quad (6.31)$$

$$\forall k \in \mathcal{K}, v \in \mathcal{V}, n \in \mathcal{N}_{Sal}^k.$$

Energy sales aggregated income at a node

$$wc_{k,n}^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} SP_{k,n}^{v,m,t} \cdot w_{k,n}^{v,m,t} \quad (6.32)$$

$$\forall k \in \mathcal{K}, v \in \mathcal{V}, n \in \mathcal{N}_{Sal}^k.$$

Aggregated energy generated at a node

$$zn_{i,k}^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} z_{i,k}^{v,m,t} \quad (6.33)$$

$$\forall v \in \mathcal{V}, i \in \mathcal{I}_{Gen}, k \in \mathcal{K}_{Out}^i.$$

Energy generation aggregated cost at a node

$$zc_i^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} CO_{i,k}^v \cdot z_{i,k}^{v,m,t} \quad (6.34)$$

$$\forall v \in \mathcal{V}, i \in \mathcal{I}_{Gen}, k \in \mathcal{K}_{Po}^i.$$

Energy storage aggregated cost at a node

$$rc_i^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} CO_{i,k}^v \cdot ro_{i,k}^{v,m,t} \quad (6.35)$$

$$\forall v \in \mathcal{V}, i \in \mathcal{I}_{Sto}, k \in \mathcal{K}_{Po}^i.$$

6.4. Symbolic Model Specification

Aggregated input energy to each technology at a node

$$y_{i,k}^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} y_{i,k}^{v,m,t} \quad (6.36)$$

$$\forall v \in \mathcal{V}, i \in \mathcal{I}_{Gen}, k \in \mathcal{K}_{In}^i.$$

Aggregated storage (divide by 8760 to get ‘average use’)

$$ra_{i,k}^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} r_{i,k}^{v,m,t} \quad (6.37)$$

$$\forall v \in \mathcal{V}, i \in \mathcal{I}_{Sto}, k \in \mathcal{K}_{Po}^i.$$

Aggregated energy input to storage

$$rn_{i,k}^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} r_{i,k}^{v,m,t} \quad (6.38)$$

$$\forall v \in \mathcal{V}, i \in \mathcal{I}_{Sto}, k \in \mathcal{K}_{Po}^i.$$

Aggregated energy output from storage

$$rp_{i,k}^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} ro_{i,k}^{v,m,t} \quad (6.39)$$

$$\forall v \in \mathcal{V}, i \in \mathcal{I}_{Sto}, k \in \mathcal{K}_{Po}^i.$$

Aggregated demand at a node for each type of energy

6. Stochastic Multi-stage Model

$$dn_k^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} D_k^{v,m,t} \quad (6.40)$$

$\forall k \in \mathcal{K}, v \in \mathcal{V}.$

Aggregated primary energy consumed at a node

$$en^v = \sum_{m \in \mathcal{M}} DM^m \cdot \sum_{t \in \mathcal{T}_{T_m}^m} e^{v,m,t} \quad \forall v \in \mathcal{V}. \quad (6.41)$$

(6) Objectives

Total discounted expected cost

Mathematical expectation of the total cost throughout the decision horizon, using the value at each node and the node probability. Discount rate is used to get the present value:

$$c = \sum_{v \in \mathcal{V}} (1 + DR)^{-PT^v} \cdot PR^v \cdot cn^v. \quad (6.42)$$

Total averaged primary energy

Mathematical expectation of the total primary energy consumed throughout the decision horizon, using the value at each node and the node probability:

$$et = \sum_{v \in \mathcal{V}} PR^v \cdot en^v. \quad (6.43)$$

Total averaged emissions

Mathematical expectation of the total emissions throughout the decision horizon, using the value at each node and the node probability:

$$p = \sum_{v \in \mathcal{V}} PR^v \cdot \sum_{l \in \mathcal{L}} pn_l^v. \quad (6.44)$$

6.5. Numerical Example

This section contains a numerical example of the optimization problem defined in this chapter with a limited number of SMS entities. The following information is included:

1. The set elements in each SMS set used in the instance.
2. Parameter values used in the instance. When the table is large enough, the first and last rows are shown. A graphical representation for some parameters is shown.
3. The result of the optimization. Similarly to the parameters, values and charts are presented for the decision variables optimal values.

The following features apply to the example:

- Three long-term periods;
- Six short-term periods, four hours each;
- Four representative profiles;
- One generation technology, no existing technologies;
- Two possible contracts;
- Stochastic parameters: demand, energy prices, investment costs and subsidies;
- Emissions and efficiency have not been considered for this example;
- Only investment costs have been considered;
- The objective was to minimize total expected cost.

Both the model and the instance have been built using the framework described in Chapter 8. After building the model and instance, the solution has been obtained through the same framework

6. Stochastic Multi-stage Model

using GAMS 23.9 as optimization and modeling software, and OS-ICPLEX as solver. The solution time was 0.137 seconds. Tables and charts have been moved to the end of the chapter for the sake of clarity.

6.5.1. Scenario Tree

Figure 6.2 shows the representation of the scenario tree for this simple example with two equiprobable scenarios; two branches after the first-stage decisions at period time 1, node 1; and two more periods for second-stage decisions.

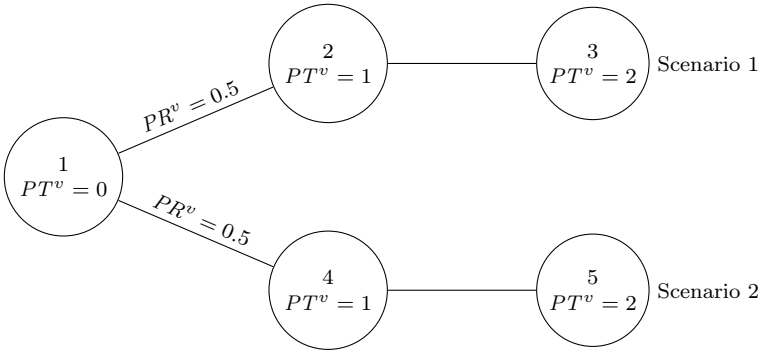


Figure 6.2.: Scenario tree for the multi-stage model numerical example.

6.5.2. Sets

The following sets and subsets elements are used in the instance:

- $\mathcal{A} = \{0, 1, 2\}$;
- $\mathcal{I} = \{PV\}$;
- $\mathcal{K} = \{\text{electricity, radiation}\}$;
- $\mathcal{N} = \{\text{normalRTEp, touRTEp, fictitious}\}$;

6.5. Numerical Example

- $\mathcal{V} = \{1, 2, 3, 4, 5\}$;
- $\mathcal{M} = \{\text{profile1, profile2, profile3, profile4}\}$;
- $\mathcal{T} = \{\text{time1, time2, time3, time4, time5, time6}\}$;
- $\mathcal{I}_{Ds} = \{\text{PV}\}$;
- $\mathcal{A}_{New} = \{0\}$;
- $\mathcal{A}_{Old} = \{1, 2\}$;
- $\mathcal{I}_{Gen} = \{\text{PV}\}$;
- $\mathcal{V}_{Root} = \{1\}$;
- $\mathcal{N}_{Tpur} = \{\text{normalRTEp, touRTEp, fictitious}\}$;
- $\mathcal{K}_{Epur} = \{\text{electricity, radiation}\}$;
- $\mathcal{V}_{Fut} = \{2, 3, 4, 5\}$.

The multidimensional set mapping energy types and contracts is in Table 6.1. Table 6.2 shows the input/output multidimensional sets. The tree structure is managed by the mapping of parent nodes in Table 6.3. The possible technology age at each node is in Table 6.4. Finally, each representative profile has the same short-term periods as shown in Table 6.5.

6.5.3. Parameter Values

In addition to the mapping between parent nodes, the tree structure also needs the period times in Table 6.6 and the nodes' probabilities in Table 6.7. Investment costs and subsidies have been simulated for the two scenarios in the example and are shown in Tables 6.8 and 6.9 respectively. For the sake of space, large tables for operational parameters are omitted and graphical representations are provided instead. Figure 6.3 shows the availability of solar irradiation, Figure 6.4 represents the energy demand, and Figure 6.5 contains the electricity prices. The following parameter values complete the set of data:

6. Stochastic Multi-stage Model

- $DT^m = 4$;
- $DM^m = 90$;
- $G_{PV} = 0.245$;
- $AG_{PV}^a = 1$;
- $EC_{i,k,k'}^v = 1$;
- $LP_{PV}^V = 1500$;
- $ME_{electricity,n} = 276.8$;

6.5.4. Solution

The total expected cost is 71,699 EUR. Table 6.10 shows the optimal strategic decisions on investment. In this extremely simple example, the decision is not to invest now (period 0). Figure 6.6 shows the available units of PV technologies. It can be seen how the foreseen decisions vary from one scenario to another: in one of them investments are recommended during period 1, whilst in the other one investments are delayed one more year. Notice also how the model takes into account technologies age. Figure 6.7 shows the available capacity of PV technologies in the building. Note that this capacity gathers the technologies aging features, as the model takes into account the age of the units as can be seen in Figure 6.6. In any case, for the next period the optimization must be run again and the new information and knowledge about uncertainties might result in different decisions. Moreover, operational decisions are useful to provide the decision maker with insights about the future performance of the building energy systems. For example, Figure 6.8 shows the energy purchases throughout the tree nodes. It can be seen how investment decisions combined with uncertainties also affect tariff selection decisions, as in scenario 2 (nodes 4 and 5) a tariff switch is foreseen the second year (node 4). Of course those strategic decisions also affect operational decisions, as far as when energy is produced, see Figure 6.9, less purchases are needed.

Tables and Charts

Table 6.1.: \mathcal{N}_{Pur}^k multidimensional set

k	n
electricity	normalRTEp
electricity	tourRTEp
radiation	fictitious

Table 6.2.: Energy input/output multidimensional sets

i	\mathcal{K}_{In}^i	\mathcal{K}_{Out}^i	\mathcal{K}_{Po}^i
PV	radiation	electricity	electricity

Table 6.3.: \mathcal{V}_{Pa}^v multidimensional set

v	v'
2	1
3	2
4	1
5	4

6. Stochastic Multi-stage Model

Table 6.4.: $\mathcal{A}_{Ages}^{i,v}$ multidimensional set

i	v	a
PV	1	0
PV	2	0
PV	2	1
PV	3	0
PV	3	1
PV	3	2
PV	4	0
PV	4	1
PV	5	0
PV	5	1
PV	5	2

Table 6.5.: \mathcal{T}_{Tm}^m multidimensional set

m	t					
profile1	time1	time2	time3	time4	time5	time6
profile2	time1	time2	time3	time4	time5	time6
profile3	time1	time2	time3	time4	time5	time6
profile4	time1	time2	time3	time4	time5	time6

Table 6.6.: PT^v parameter

v	value
1	0
2	1
3	2
4	1
5	2

6.5. Numerical Example

Table 6.7.: PR^v parameter

v	value
1	1.00
2	0.50
3	0.50
4	0.50
5	0.50

Table 6.8.: CI_i^v parameter

i	v	value
PV	1	1326.65
PV	2	1061.32
PV	3	849.06
PV	4	1193.99
PV	5	1074.59

Table 6.9.: SU_i^v parameter

i	v	value
PV	1	663.33
PV	2	530.66
PV	3	721.70
PV	4	1014.89
PV	5	913.40

6. Stochastic Multi-stage Model

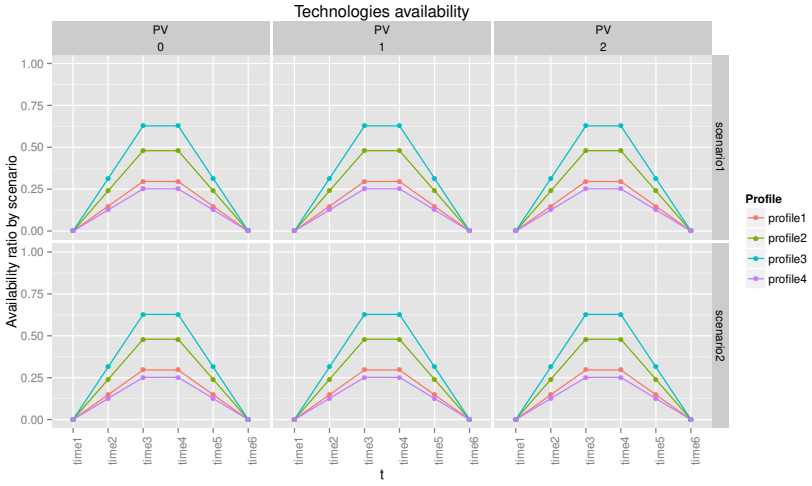


Figure 6.3.: $AF_{PV}^{v,m,t}$ parameter by scenario

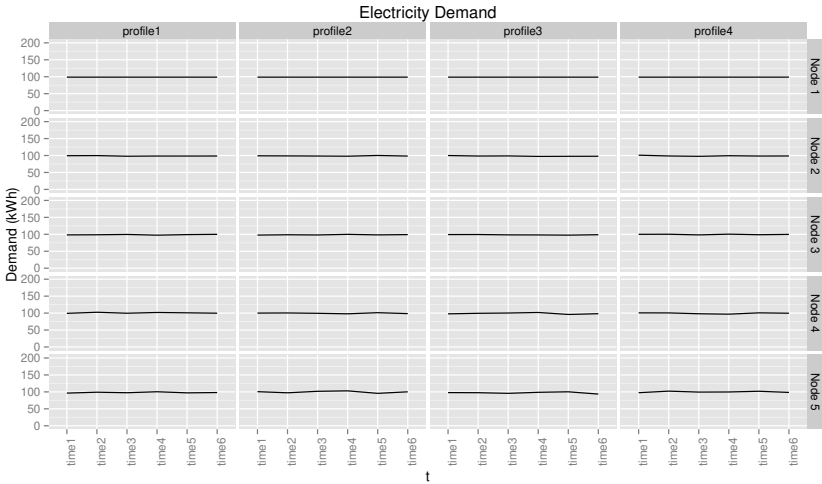


Figure 6.4.: $D_k^{v,m,t}$ parameter by node

6.5. Numerical Example



Figure 6.5.: $PP_{k,n}^{p,m,t}$ parameter by node

6. Stochastic Multi-stage Model

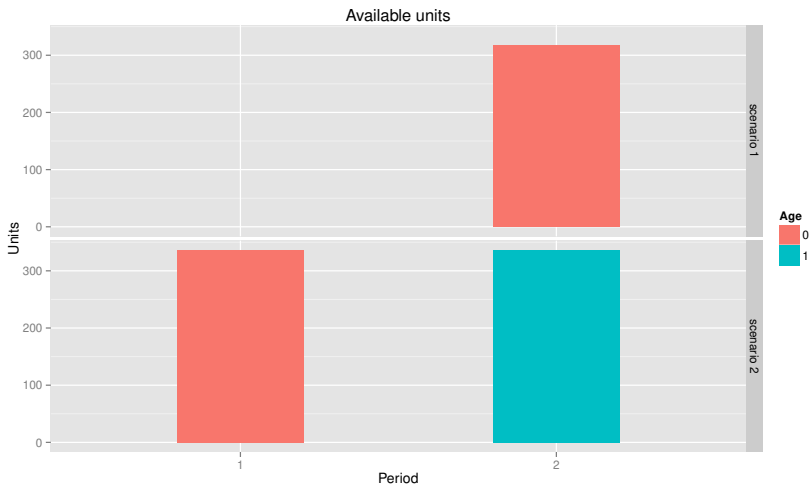


Figure 6.6.: x_i^v variable

Table 6.10.: x_i^v variable

i	v	value
PV	3	317
PV	4	336

6.5. Numerical Example

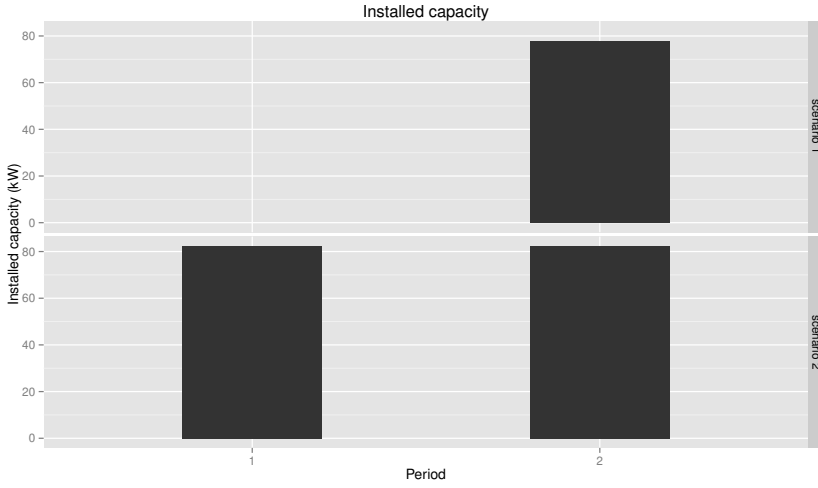


Figure 6.7.: xc_i^v variable

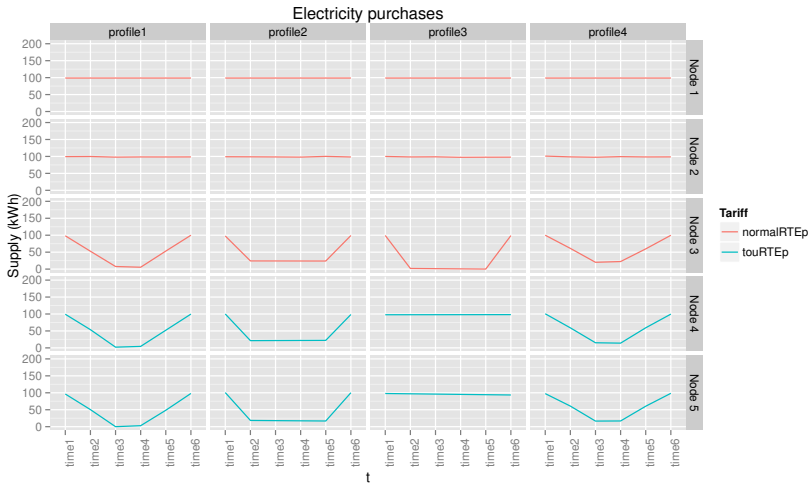


Figure 6.8.: $u_{k,n}^{v,m,t}$ variable

6. Stochastic Multi-stage Model

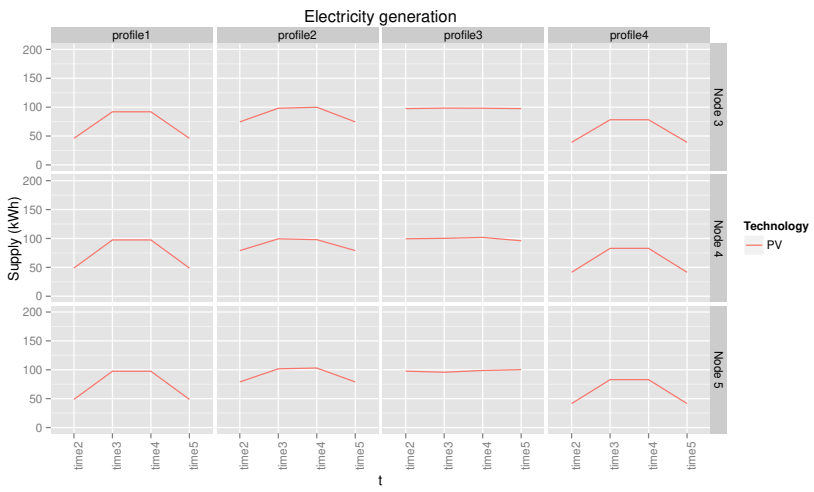


Figure 6.9.: $z_{i,k}^{v,m,t}$ variable

7. Risk Management

7.1. Introduction

In Chapters 4 and 6, uncertainty was added to the optimization problems, formulating the models as SP problems. Indeed, taking into account uncertainty is needed in order to reduce risks. However, the models formulated so far are *risk neutral*, as this risk is not modeled. Ignoring risk management might result on an optimal average value for the objective function, but very bad outcomes for some extreme scenarios. In the case at hand, the *optimal* investment plan leading to the minimum expected cost could be very bad for the actual scenario that eventually occurs. To overcome such drawback, a risk measure should be included in the formulation. In this chapter an overview of the most used risk measures is provided, and the implementation adopted in the models in Chapters 4 and 6 is presented.

7.2. Risk Measures

A risk measure, or risk function, characterizes the risk associated with a given decision. Throughout the time different risk measures have been proposed for SP. It can be found in the literature detailed descriptions and comparisons between them, see for example Kall and Mayer (2005), Krokmal et al. (2011), or Alonso Ayuso et al. (2009) for general problems; Conejo et al. (2010) focuses on energy markets. However, since Rockafellar and Uryasev (2000) published the minimization of Conditional Value at Risk (CVaR) method, it has been proved as a breakthrough risk measure. It has the desirable properties stated by Artzner et al. (1999), namely: translation

7. Risk Management

invariance, subadditivity, positive homogeneity, and monotonicity, hence becoming a *coherent* risk measure. The more basic risk measure that can be used is the variance. Markowitz (1959) first proposed the mean-variance model that minimized the risk (variance) requiring a given average objective. Variance can be also included in the objective as a weighted combination of risk and reward. Note that this type of models were formerly used for portfolio optimization, where the objective is to maximize reward. Nevertheless, it is straightforward to apply the ideas to minimization of cost, as in the models developed hereby.

A thorough survey of risk measures properties is out of the scope of this work. Some of the more prominent are shortfall probability, expected shortage, or the more recent advances on stochastic dominance, see for example Ogryczak and Ruszczyński (1999), Ogryczak and Ruszczyński (2002), Dentcheva and Ruszczyński (2003), or Dentcheva and Ruszczyński (2006). VaR and CVaR are explained in the following section, as they are used in the developed models.

7.3. Conditional Value at Risk (CVaR)

The approach followed in this work to model risk is the CVaR model by Rockafellar and Uryasev (2000). First, it is needed to define Value at Risk (VaR). VaR is one of the most known measures in financial risk management (see, for example, Duffie and Pan 1997) though it can be applied to other sectors as will be shown later. Even though it is usually defined for profits, i.e., maximization, here it is used the minimization of costs meaning. Given a confidence level α , $0 < \alpha < 1$, typically a large value, e.g., 0.95, the VaR is the lowest cost λ that ensures a probability lower than $1 - \alpha$ of getting a cost higher than such value. Note that this risk measure includes both the loss size and the probability of loss through the use of quantiles, and can be mathematically expressed as:

$$\text{VaR}(\alpha, \mathbf{x}) = \min \{ \lambda : \mathbb{P}[\omega | f(\omega, \mathbf{x}) > \lambda] \leq 1 - \alpha \}, \quad (7.1)$$

where ω is the sample space and $f(\omega, \mathbf{x})$ is the objective function, see Chapter 2. Therefore, the VaR is the $(1 - \alpha)$ -quantile of the cost distribution. For example, a VaR of 100 monetary units for a confidence level of 95% means that the probability of having a cost higher than 100 monetary units is lower than 5%.

In spite of the good properties and widely spread use by banking institutions, VaR has two main drawbacks. First, it is not a coherent measure, as it does not satisfy the subadditivity property. Second, the measure does not provide information about the cost distribution beyond the VaR, and therefore fat tails are not detected. To overcome these shortcomings, CVaR is defined as the conditional expectation of losses Λ that exceed the VaR level λ , i.e.:

$$\text{CVaR} = \min \{ \Lambda : \mathbb{E} [f(\omega, \mathbf{x}) > \lambda] \}. \quad (7.2)$$

According to this definition, a CVaR of 100 monetary units for a confidence level of 95% means that the average of worst case losses

7. Risk Management

that may occur with 5% probability is 100 monetary units. The key to success of the CVaR risk measure is that it is easy to implement in SP problems keeping the good properties of both the risk measure and the SP problem. In general, the CVaR can be added to the optimization problem by adding a weighted risk term \mathbf{R} to the objective function:

$$\mathbf{R} = \lambda - \frac{1}{1 - \alpha} \sum_{\omega \in \Omega} \mathbb{P}[\omega] s(\omega), \quad (7.3)$$

where λ is a decision variable representing the VaR and $s(\omega)$ is an auxiliary variable, which is the solution of:

$$\max \{0, \lambda - f(\omega, \mathbf{x})\}.$$

The following constraints are also needed for all $\omega \in \Omega$:

$$\lambda - f(\omega, \mathbf{x}) \leq s(\omega), \quad (7.4)$$

$$s(\omega) \geq 0. \quad (7.5)$$

The implementation is slightly different in by-scenario formulations and by-node formulations. In the following sections, both approaches are detailed for the models in Chapters 4 and 6 respectively.

7.4. CVaR in the Two-stage Model

In model (4.1)–(4.6) the variability of outcomes is characterized by the stochastic second stage operational decisions $y_{i,j}^t(\omega)$ with unit costs $CO_{i,j}^t(\omega)$. Although the strategic first-stage decisions x_i^t have random costs $CI_i^t(\omega)$, the total cost function (4.6) ignores their variability by using mean values $CI_i^t = \mathbb{E}[CI_i^t(\omega)]$. If the total stochastic costs $\sum_{i \in \mathcal{I}} CI_i^t(\omega) \cdot x_i^t$ cannot well enough approximate a normal distribution, then instead of the mean value of these costs it is advantageous to use in equation (4.6) the median or other

7.4. CVaR in the Two-stage Model

quantiles, especially when these costs may be affected by extreme events. This variability can be easily modeled by additional decision variables $v^t \geq 0$ representing the VaR which jointly with variables $x_i^t, y_{i,j}^t$ minimize the function

$$\sum_{t \in \mathcal{T}} \left(v^t + (\rho^t)^{-1} \cdot \mathbb{E} \left[\max \left\{ 0, \sum_{i \in \mathcal{I}} CI_i^t(\omega) \cdot x_i^t - v^t \right\} \right] + \mathbb{E} \left[\sum_{j \in \mathcal{J}} CO_{i,j}^t(\omega) \cdot y_{i,j}^t(\omega) \right] \right), \quad (7.6)$$

subject to constraints (4.1)–(4.5). The minimization of function (7.6) includes now additional subproblems on minimization with respect to $v^t, t \in \mathcal{T}$,

$$\mathbf{R} = v^t + (\rho^t)^{-1} \cdot \mathbb{E} \left[\max \left\{ 0, \sum_{i \in \mathcal{I}} CI_i^t(\omega) \cdot x_i^t - v^t \right\} \right], \quad (7.7)$$

subject to $v^t \geq 0$, where ρ^t is a risk factor controlling that stochastic costs $\sum_{i \in \mathcal{I}} CI_i^t(\omega) \cdot x_i^t$ do not exceed a desirable robust safety level v^t (VaR). If the safety vector $\rho^t < 1$, then the minimization of the risk function (7.6) yields optimal values v^{t*} satisfying the following safety constraints, known in financial applications as the Value at Risk indicator (Rockafellar and Uryasev 2000):

$$\mathbf{P} \left[\sum_{i \in \mathcal{I}} CI_i^t(\omega) \cdot x_i^t \geq v^t \right] = \rho^t \quad \forall t \in \mathcal{T}. \quad (7.8)$$

Subproblem (7.7) corresponds to the Conditional Value at Risk (CVaR) minimization. This problem has the following interpretation. Foreseen investments costs $\sum_{i \in \mathcal{I}} CI_i^t(\omega) \cdot x_i^t$ at time t are planned to be covered by ex-ante credit at price $\rho^t < 1$ and ex-post borrowing

$$\max \left\{ 0, \sum_{i \in \mathcal{I}} CI_i^t(\omega) \cdot x_i^t - v^t \right\}$$

7. Risk Management

for relatively greater price. This provides more flexibility (robustness) compared only with ex-ante planning of investments using mean values of perceived unit costs $\mathbb{E}[CI_i^t(\omega)]$. It is also important to remark that the ex-ante credit is evaluated by (7.8) as the quantile with desirable safety levels ρ^t , $t \in \mathcal{T}$, rather than mean values which are especially misleading in the case of non-normal distributions.

7.5. Risk Measures in the Multi-stage Model

Several updates are needed in the formulation of the stochastic model (6.1)–(6.44) in order to manage risk. First of all, as CVaR refers to the worst scenarios, it is needed a mapping between nodes and scenarios. In order to deal with this feature, a new set of scenarios $s \in \mathcal{S}$ is defined. Considering that a scenario is actually a path of nodes, whose leaf node is unique, multidimensional sets \mathcal{V}_{Leaf}^s and \mathcal{V}_{Path}^s are defined for the leaf node and the nodes that are in the path of the scenario, respectively. On the other hand, two new variables are needed to model risk: vr for the VaR, that refers to the whole decision horizon; and the auxiliary variable sr^s , per scenario, to calculate the CVaR. Thus, the implementation of the risk term (7.3) is made by a new variable rt :

$$rt = vr + (1 - AL)^{-1} \cdot \sum_{s \in \mathcal{S}} PR^{Leaf(s)} \cdot sr^s, \quad (7.9)$$

where AL is the confidence level α . The risk constraints (7.4) for cost minimization would be the following:

$$\sum_{v \in \mathcal{V}_{Path}^s} (1 + DR)^{-PT^v} \cdot cn^v - vr \leq sr^s \quad \forall s \in \mathcal{S}. \quad (7.10)$$

Finally, the objective is now to minimize the weighted cost and risk through the following objective function:

7.5. Risk Measures in the Multi-stage Model

$$oc = (1 - BE) \cdot c + BE \cdot rt, \quad (7.11)$$

where BE is the risk weight β that models the risk aversion of the decision maker. Note that the problem whose $BE = 0$ is the risk-averse problem, i.e., without risk management.

In Subsection 6.4.5 of Chapter 6, three different objectives were propounded: minimization of cost, emissions, or energy consumption. Even though CVaR is mostly used for monetary risks, this dissertation includes as a novelty taking into account the risk of high emissions, or the risk of high energy consumption. Under the context outlined in Chapter 1, it is necessary to consider emissions and energy consumption as a threat, especially by certain stakeholders such as policy makers. Therefore, keeping the same formulation for the risk term (7.9), one of the following equations replaces the risk constraints (7.10) for either emissions or energy consumption minimization:

$$\sum_{l \in \mathcal{L}, v \in \mathcal{V}_{Path}^s} pn_l^v - vr \leq sr^s \quad \forall s \in \mathcal{S}, \quad (7.12)$$

$$\sum_{v \in \mathcal{V}_{Path}^s} en^v - vr \leq sr^s \quad \forall s \in \mathcal{S}, \quad (7.13)$$

and the counterpart objective function replaces (7.11), namely:

$$op = (1 - BE) \cdot p + BE \cdot rt, \quad (7.14)$$

$$oe = (1 - BE) \cdot et + BE \cdot et. \quad (7.15)$$

The complete model with risk has been tested with real data from the EnRiMa test sites. Some numerical results are shown in Chapter 9.

Part III.

Implementation and
Outcomes

8. An Integrated Framework

8.1. Introduction

The term Optimization is often used to denote the entire process of Operations Research that lead to scientific based decisions. However, the optimization step is only one of the tasks within the whole decision making process. This optimization step may be carried out by a solver, i.e., a piece of software implementing the algorithm, that receives the problem data and returns the problem solution. Indeed, the algorithm implemented on the solver is the actual workhorse for the decision making problem solution. However, the tasks out of the “black box” are not trivial, and they require a large amount of research time.

Many of these tasks are related to statistical analysis, including descriptive statistics, graphics, or inference. More often than not, analysts use several applications during their research, e.g. spreadsheets, statistical software, text editors, etc. Finally, a consolidated report containing the results of the optimization as well as the data analysis has to be generated. The integrated framework proposed in this work allows to perform a whole decision making process regarding energy systems optimization at the building level. This process is summarized in Fig. 8.1¹. The solver is the core of the process. It receives information about:

- The model: symbols and descriptions of the mathematical

¹Note that this is a simplified version of the DSS general scheme in Figure 2.5.

8. An Integrated Framework

formulation;

- The instance: actual elements in the sets, parameters, estimations, etc. It is sometimes also termed as concrete model or case.

The solver returns the solution of the problem, i.e., the optimal values for the decision variables and the objective based on the algorithm selected. This result should be analyzed and end up in a comprehensive report for the decision maker or any other stakeholder. Such a process needs the following elements:

- Statistical software;
- Data visualization;
- Mathematical representation;
- Solver input generation;
- Solver invocation;
- Documentation output.

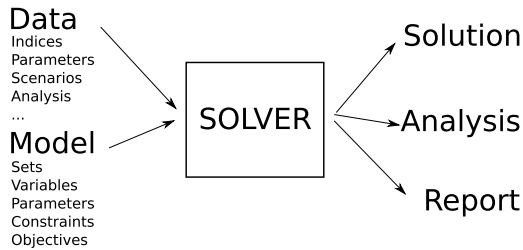


Figure 8.1.: Elements in an optimization decision making process.

As remarked in Chapter 2, the proposed framework relies on the use of Algebraic Modeling Languages (AMLs). Despite AMLs have been selected to build the framework, it is important to remark

that other structured formats, e.g., markup languages, can be used as far as they are useful to accomplish the DSS main mission, i.e., the stakeholders dialog, see Section 2.3 in Chapter 2. For example, Optimization Services (OS)² is a Computational Infrastructure for Operations Research (COIN-OR)³ project that uses the eXtended Markup Language (XML) format to represent optimization problems and that is suitable to effectively communicate within an eventual DSS. OS aims at providing “a set of standards for representing optimization instances, results, solver options, and communication between clients and solvers in a distributed environment using Web Services” (Fourer et al., 2010). Another example is Formal Mathematical Language (FMathL) by Neumaier (2014). There are several AMLs available both commercial and open source. The one selected for the first implementation of the framework has been GAMS. Nonetheless, the classes explained below can be easily extended to other languages. This is possible due to the fact that the SMS is generically represented within the DSS using specific data structures developed in the programming language and statistical software R (R Core Team 2013a). R provides functionality for all the required tasks within the DSS, including data analysis, visualization and representation tasks, allowing communication to different optimization software packages through inner interfaces. For example, the `gdrrw`⁴ package by Dirkse and Jain (2013) can be used for reading and writing data in GAMS `.gdx` binary files. Data cleaning and management can also be done with R. A description on how R can be used for data preparation for optimization can be found in Pitz (2012). Eventually, user interfaces can be easily implemented, both through other technologies such as `php`⁵, or `.NET`⁶,

²<https://projects.coin-or.org/OS>

³<http://www.coin-or.org>

⁴http://support.gams-software.com/doku.php?id=gdrrw:interfacing_gams_and_r. Further examples can be found at <http://www3.lei.wur.nl/gamstools/>.

⁵<http://php.net>

⁶<http://asp.net>

8. An Integrated Framework

or through R libraries devoted to user interfacing, such as `shiny`⁷. Note that the spirit of the framework can be also applied using other programming and analysis tools.

8.2. 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 using the new or updated 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).

Reproducible Research is a relatively recent crossing topic and suitable for any discipline. See for example Gentleman and Temple Lang (2007), Vandewalle et al. (2009), Stodden (2009), Xie

⁷<http://www.rstudio.com/shiny/>

8.2. A Reproducible Research Approach

(2013b), Xie (2013c), Baggerly and Coombes (2009), or Koenker and Zeileis (2009).

In order to fulfill the requirements for a DSS detailed in Section 2.3 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, see for example Cano et al. (2012a). “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” (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 DSSs are:

- It is Free and 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-processing), 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;

8. An Integrated Framework

- There are thousands⁸ 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 like The Comprehensive R Archive Network (CRAN⁹);
- 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 databases, etc;
- 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 accordingly modify it to meet their needs. Moreover, it is easier to detect bugs and errors as one can dig into the code;
- The active and high-qualified R-Core development team and the huge community of R users provide an incredible support level (without warranty, skeptics would say), difficult to surpass by other support schemes. Furthermore, it is also possible to get professional support by commercial companies if needed¹⁰;
- New methods, tools or algorithms can be deployed very quickly. A company or organization can develop and deploy an innovative method from its R&D department or from the result of other published research.

⁸+5000 at the Comprehensive R Archive Network (CRAN) on December, 2013. Other repositories are bioconductor, omegahat, r-forge, and github.

⁹<http://cran.r-project.org>

¹⁰Visit, for example, <http://www.revolutionanalytics.com/>

8.3. The *optimr* R Package

8.3.1. Overview

An R package¹¹ called *optimr* has been developed as part of the framework described in this chapter to deal with the model, the data, and the solutions. The *optimr* package revolves around two classes of objects: *optimSMS* and *optimInstance*. The former contains the 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 8.2 shows an outline of the package structure. The *optimSMS* and *optimInstance* classes are S4 classes, which allows to use Object Oriented Programming (OOP) techniques through the use of slots and methods, see R Core Team (2013b) for details about the R Language.

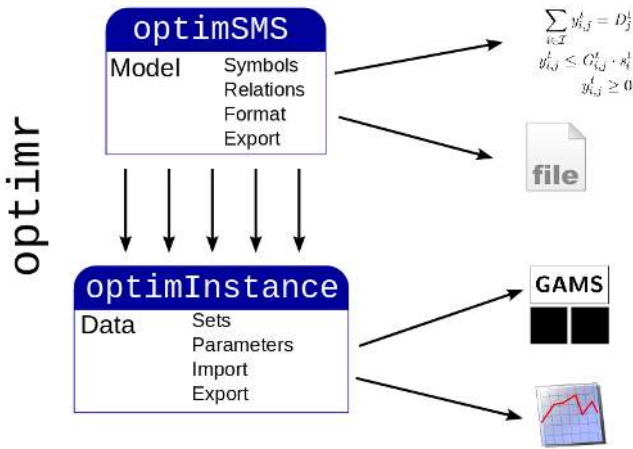


Figure 8.2.: The *optimr* package structure.

The model in an *optimSMS* object can be represented in both human and machine readable formats through standard data struc-

¹¹'package' and 'library' are used interchangeably in the R jargon.

8. An Integrated Framework

tures such as R *data frames*, which are similar to database tables, with rows and columns whose data is of a given type. Thus, from a data-driven perspective, all the elements in the SMS are stored in data objects, and they can be used for any purpose. Likewise, the specific data in an `optimInstance` object, which are accurately linked to a generic SMS through common symbols, can be analyzed, visualized, and exported to be used by eventual stakeholders. Note how this feature links with the *stakeholders dialog* emphasized in Section 2.3 of Chapter 2. Important actions that will help during the optimization research are:

- Fetch the symbols and descriptions to represent the items and equations of the model, formatted in the appropriate syntax (e.g. \LaTeX);
- Form complex structures as equations;
- Generate instances for solvers in the required format, e.g., GAMS and MPS;
- Solve the optimization problem, calling the appropriate interfaces;
- Import the solution, making it available for post-treatment.

A SMS is stored in `optimSMS` class objects. The basic information stored for each item of a SMS is: The symbol that represents the item (`symbol`); A short description (`sDes`); A long description (`lDes`). Variables and parameters need also the indices (`ind`) that apply for them. Other columns such as `units`, `dataType`, `nature`, or `tag` complete the information about each entity. Sets, variables and parameters are stand-alone objects. However, in order to represent the equations, a more sophisticated structure is needed. Each equation has its own symbol and descriptions in addition to equation-specific information, such as the domain or the relation (equality or inequality). Moreover, an equation can be formed by an undetermined number of terms (variables and parameters), adding and

multiplying at several levels, and even nested among them. Therefore, a tree structure defining the terms in the equations and their relationship is implemented in the `optimSMS` class. On the other hand, a specific instance of the optimization problem, composed by the actual elements in the sets and the values for each parameter, are stored in `optimInstance` class objects, where the results returned by the solver will be also stored for representation and analysis purposes.

8.3.2. Details

The `optimSMS` class is composed by the following slots:

- 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 most relevant are (the names are self-meaningful):

- To get expressions: `getAliases`, `getConsts`, `getEq`, `getEqs`, `getExpr`, `getModel`, `getMultiSets`, `getPars`, `getSets`, `getSubsets`, `getSymbol`, `getVars`. The main arguments to be passed are the `optimSMS` object, the entity identifier, and the format, e.g. 'tex' or 'gams';
- To get the R data frames with the information: `SMSconsts`, `SMSeqs`, `SMSpars`, `SMSsets`, `SMsterms`, `SMSvars`.

The creation and addition of elements in a SMS is made through the specific functions `newSMS`, `newSMSconst`, `newSMSseq`, `newSMSpar`, `newSMSset`, and `newSMSvar`.

8. An Integrated Framework

As for the instance, i.e., the concrete model to be solved using specific data, it is stored in `optimInstance` class objects, whose slots are:

- Descriptive information: characters `name`, `sDes`, `lDes`; integer `id`;
- An `sms` slot with the `optimSMS` object to which the instance is referred. Note that this is not a link, but a *copy* created when the `optimInstance` object is created. In this way, changes can be done to the SMS within the instance;
- Input data frames: lists `sets` and `pars` containing the data frames with the data;
- Output data frames: list `vars` containing the data frames with the data imported from the solver;
- Equations definition for the model instance: a list `eqs` with two vectors `objectives` and `constraints` containing the equations identifiers that apply for the instance;
- Result summary: A list with three single-valued vectors: `solve` and `model` for the solver and model output codes from the optimization software respectively, and `obj` for the optimal value of the objective function (if any).

The names of the data frames mentioned above are the symbols of their counterpart entities (set, variable, or parameter). Similarly to the `optimSMS` class, the slots of an `optimInstance` object can be also accessed using self-explained functions: `instanceSets`, `instancePars`, and `instanceVars`. Other methods of the `optimInstance` class are:

- `wProblem`: writes the problem in a given format to be solved;
- `exportPars`: exports parameter values in a given format to be used for other software, for example XML or `.gdx` (GAMS binary format);

- `importGams`: imports the solution from `.gdx` files into an `optimInstance` object;
- `getTree`: returns a `data.frame`, a plot, or both for the scenario tree of a SP problem.

8.4. Illustrative Example

In this section, a complete example of the framework implementation is presented. The objects are created sequentially following the logic of a real case. Finally, an example of a reproducible report containing the model, the data, and the analysis is provided. Models in Chapters 3 and 4 are used for the example, and a new instance will be generated simulating 100 scenarios. Some excerpts of the code are shown within the text. The complete reproducible code can be found in Appendix B.

8.4.1. The Model

To create the deterministic model in Chapter 3 the following code should be used¹²:

```
model1SMS <- newSMS(name = "Deterministic1",
  sDes = "A Basic Case",
  lDes = "The simplest model, only electricity")
newSMSset(sms = model1SMS,
  list(symbol = "i",
    sDes = "Technology",
    setType = "set"))
# ... ..
newSMSpar(sms = model1SMS,
  list(symbol = "D",
    sDes = "Demand Level",
    units = "kW",
```

¹²For the sake of space, only one example of each entity is shown.

8. An Integrated Framework

```
        ind = as.array(list(c(2,3))))
# ... ..
newSMSvar(sms = model1SMS,
          list(symbol = "x",
               sDes = "Capacity to be installed",
               units = "kW",
               positive = TRUE,
               ind = as.array(list(c(1,3))))))
# ... ..
newSMSseq(sms = model1SMS,
          list(id = 5,
               symbol = "eqCapacity",
               sDes = "Technologies capacity",
               relation = "lte",
               nature = "constraint",
               domain = as.array(list(c(1,2,3))))),
          list(id = 1,
               side = "l",
               nature = "vars",
               item = 2),
          list(id = 2,
               side = "r",
               nature = "pars",
               item = 3),
          list(id = 3,
               side = "r",
               nature = "vars",
               item = 3,
               parent = 2))
```

Once the SMS is in an `optimSMS` object, any expression can be easily obtained. For example, the last command of the code above adds Equation (3.4) to `model1SMS` object, whose class is `optimSMS`, using the `newSMSseq` function. The equation expression in GAMS

format¹³ is retrieved as follows:

```
cat(getEq(object = model1SMS, getid = 5,
      format = "gams"))

## eqCapacity(i,j,t) ..
##   y(i,j,t) =l= G(i,j,t)*s(i,t)
## ;
```

Note that combining different expressions and working with text in R, complex representations of the models can be produced. In fact, this is what more elaborated methods, e.g., `wProblem`, do to create comprehensive representations of the model.

8.4.2. The Instance

An instance always refers to a model, and therefore to create an `optimInstance` object it is needed an `optimSMS` object to exist. Once the instance has been created, elements (actual sets, parameter values and equations to include) are added to the instance, related to its SMS¹⁴:

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

newInstanceSet(instance = model1Instance1,
              set = "i",
              c("RTE", "PV", "CHP"))

# ... ..

newInstancePar(instance = model1Instance1,
              par = "DT",
              data.frame(
                j = rep(model1Instance1@sets[["j"]][,2],
```

¹³And similarly in L^AT_EX or, potentially, further formats.

¹⁴Again, only one example of each type is printed.

8. An Integrated Framework

```
        each = 5),
    t = model1Instance1@sets[["t"]][,2],
    value = 91*8))
# ... ..
defInstanceEqs(instance = model1Instance1,
               constEqs = c(3, 4, 5),
               objEqs = 6)
```

8.4.3. The Solution

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 first command calls the `wProblem` method over the `model1Instance1` object, whose class is `optimInstance`. This command creates an input file for GAMS; the second command calls the GAMS R interface included in the `gdrrw` library, creating an output file with the solution; the last command imports the solution from the file to the `model1Instance1` object. Note that at any point data analysis and visualization can be straightforwardly performed over the data, as they are stored in homogeneous and consistent data structures. For example, Figures 8.3 and 8.4 show possible visualizations for the solution of the deterministic model in Chapter 3.

8.4. Illustrative Example

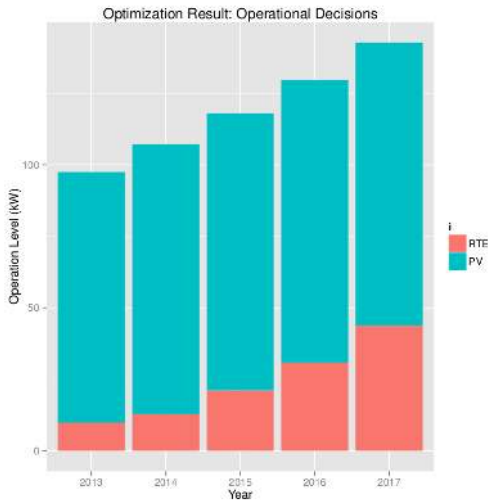


Figure 8.3.: Visualization of operational decisions example.

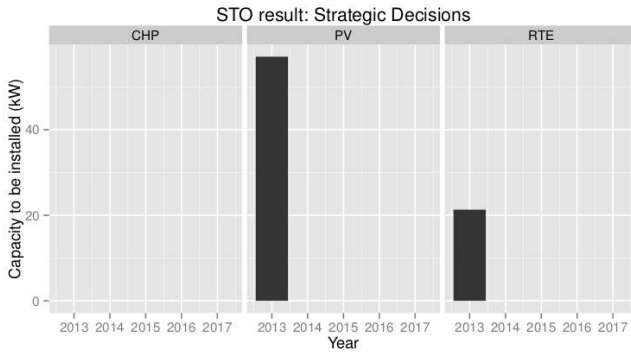


Figure 8.4.: Visualization of strategic decisions example.

8. An Integrated Framework

8.4.4. Extending the Instance

One step beyond, it is not needed to create the model in Chapter 4, i.e., equations (4.8)–(4.13), from scratch. It is enough to create a copy of `model1SMS` object and add the new information in the SMS regarding scenarios:

```
model2SMS <- model1SMS

newSMSset(sms = model2SMS,
          list(symbol = "n",
               sDes = "Scenario",
               setType = "set"))

model2SMS@vars[2, ]["ind"] <- list(c(1,2,4,3))
# ... ..

newSMSpar(sms = model2SMS,
          list(symbol = "P",
               sDes = "Scenario Probability",
               ind = as.array(list(c(4))))))
# ... ..
```

Instead of using the numerical example in Section 4.3, a new instance will be created to illustrate data treatment concerns. In this new numerical example, the forecasted evolution of the stochastic parameters CI_i , $CO_{i,j}^t$, and $D_{i,j}^t$ is that in Table 8.1: from base present values, the future values increase or decrease at a given average ratio, with a given standard deviation. Normal distributions for these annual variations is assumed.

Scenarios can be generated within the framework using the R capabilities, for example to simulate the demand parameter the following code can be used¹⁵:

¹⁵And similarly the rest of stochastic parameters, see Appendix B.

8.4. Illustrative Example

Table 8.1.: Base parameter values an uncertain evolution

Parameter	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

```

setJ <- instanceSets(model2Instance2, "j")
setN <- instanceSets(model2Instance2, "n")
deltaMean <- c(-0.10, -0.05, 0.10, 0.03, 0.10)
deltaSD <- c(0.05, 0.06, 0.04, 0.02, 0.05)
baseD <- ((213.50*(10^3))/(365*24))
corrSeason <- c(0.7, 0.9, 1.4, 1)

set.seed(1111)
newInstancePar(instance = model2Instance2,
  par = "D",
  rbind.fill(lapply(seq(along = setN),
    function (y) {
      data.frame(j = setJ,
        n = setN[y],
        t = rep(setT, each = 4),
        value = rep(sapply(0:4,
          function(x){
            (1 + rnorm(1, deltaMean[5],
              deltaSD[5]))^x
          } ),
        each = 4) * rep(baseD * corrSeason, 5))
    })))

```

8. An Integrated Framework

Finally, the new instance is solved like in Subsection 8.4.3. Post analysis and reporting over the data has then endless possibilities, for example charts like those in Figure 8.5 can be produced. Integration in a reproducible research framework is possible using specific file formats such as html, pdf or markdown.

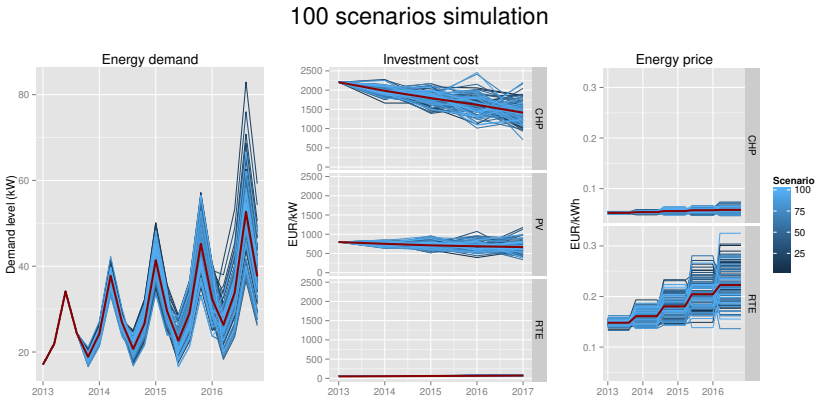


Figure 8.5.: Integrated graphics for input data.

8.4.5. Comprehensive Reproducible Example

Given the decision making problem described so far, in this section the different actions to be made over the data are described, and some of the usual procedures are reviewed. Then, a detailed example gathering all the tasks performed during a research is presented, using the integrated framework proposed in this work. Please note that some of the topics covered have been already treated throughout the dissertation, but they are included again for the example completeness.

Data Preparation and Pre-analysis

In Table 8.1, a series of values are given to solve the problem instance. But, where do they come from? In the case at hand, there

8.4. Illustrative Example

are data regarding costs, energy demand and technologies. Energy costs may come from annual contracts with energy companies. In deregulated environments, it is more and more usual to have daily and intra-day prices that can be estimated analyzing time series. Regarding energy demand, depending on the Information and Communication Technologies (ICT) infrastructure installed in the building, there may be sensors gathering data about the building occupancy and energy consumption. Jointly with the information about the building envelope, the demand can be estimated using statistical techniques such as regression or data mining. Finally, the information about technologies can be stored in some database within the building ICT infrastructure.

One way to analyze data and calculate the parameter values is to acquire raw data from the sources using different methods, and then treat them separately, for example using spreadsheets or statistical software. When the process must be documented, a text editor can be used to present graphics, describe the analysis and, eventually, publish the document.

Problem Solution

Once all the information needed to create the problem instance is ready, it is time to look for the solution. Depending on the problem type, an appropriate solver must be selected. The solver contains the algorithm that will return the solution, i.e., the optimal values and further information, e.g., slackness, using the instance of the problem as input. Furthermore, maybe an optimization or modeling software is going to be used to call the specific solver. The problem instance must be packaged in a compatible format, either for the solver or for the optimization software. For example, Mathematical Programming System (MPS) is a well known format accepted by many solvers and optimization programs, though formats using AMLs such as AMPL or GAMS are replacing “matrix-like” formats in the last decades, see Section 2.3. In any case, the parameter values and the equations must be arranged into an input file for the solver. Then, a call to the solver is made, sending the input infor-

8. An Integrated Framework

mation and returning the output, usually in form of files and direct output to the device from which it was called, e.g., Graphical User Interface (GUI) or console.

The input file can be created using programming languages to manipulate data and write files directly, for example C++. Specialized software packages and programming languages such as **Matlab** or **R** are another option. Optimization and modeling software, e.g., **GAMS**, use their own language to manipulate and prepare data in the appropriate format before calling the solver. The call to the solver can be made through a command on the system console, or from the optimization software using its own capabilities, e.g., executables or GUI).

As far as optimization software is concerned, **GAMS** is one of the most used in industry. Mathematical software such as **Matlab** and its open source counterpart **Octave**, or **Mathematica** are other powerful options. **R** itself can be used as optimization software. There is a “Task View” within the **R** project devoted to Optimization¹⁶, maintained by Theussl (2014), where a set of packages and resources are listed and briefly outlined. **R** can communicate with a number of solvers and optimization software, for example: **CPLEX**¹⁷; `lp_solve`, see Berkelaar and others (2013); or GNU Linear Programming Kit (**GLPK**)¹⁸, see Theussl and Hornik (2013), among others. The **COIN-OR**¹⁹ project also contains optimization solvers, some of them interfaced from **R**.

Post-analysis and Reporting

Similarly to the pre-analysis tasks, results must be analyzed and presented to the decision maker in a meaningful way. Moreover, results should be stored within the building’s ICT infrastructure for future use. For example, data can be saved in databases, and

¹⁶<http://cran.r-project.org/web/views/Optimization.html>

¹⁷<http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/>

¹⁸<http://www.gnu.org/software/glpk/>

¹⁹<http://www.coin-or.org/>

8.4. Illustrative Example

reports in documentation repositories.

The copy-paste approach mentioned in Sec. 8.2 is, unfortunately, the most spread way of finalizing a long and hard research. Researchers gather all the “chunks” of information scattered throughout their computers and start composing a document merging data, charts, results, text and other stuff in a final report. In the short term, this procedure may seem the faster one. But this is an illusion, because some of the following facts may arise, resulting in a painful and discouraging ending of the work:

- Several people have been working in the research. Some problems may arise, for example different location of the resources, or undocumented code only understandable for the one who wrote it;
- Different levels of reporting needed. Summaries and technical reports are very different, but the sources should be the same;
- Changes on initial data that affect subsequent results, see Section 9.2.4 in Chapter 9;
- New intermediate actions or methods to include. For example, a change on the way the parameters are estimated including regression analysis result;
- Validation and/or verification by internal or external reviewers. Many reasons can prompt these actions, for example quality control or regulations.

On the contrary, the example presented in the following subsection makes use of reproducible research techniques, avoiding the drawbacks of an unstructured procedure. On the downside, this way of work requires some training at the beginning, and a commitment to be disciplined when applying it. Nonetheless, this will end up in better results and a more gratifying undertaking. Furthermore, the integrated framework proposed using R provides a coherent environment to speed up the research’s outcomes.

8. An Integrated Framework

All Together

A whole decision making process described above is now put in practice. The Task View at CRAN entitled “Reproducible Research” (Kuhn, 2014) gathers a set of resources which can be used in R for reproducible research. The core of reproducible research capabilities in R is the `Sweave` function (Leisch, 2002). It is based on the “noweb syntax” (Ramsey, 1998), which is usually used for literate programming (Knuth, 1984). Inside Sweave files (with extension `.Rnw`), chunks of R code are inserted into \LaTeX code, resulting in outstanding documents mixing the best of two great worlds: \LaTeX for high-quality typesetting, and R for data analysis. It is important to remark that the R capabilities for reproducible research are not constrained to \LaTeX : Open Document and HTML formats are also supported. Nevertheless, \LaTeX produces stunning printable documents in PDF, ready to be delivered to DSS stakeholders. Moreover, splitting the source code files and using them appropriately, they can be combined into different final documents, such as HTML pages, PDF documents, or presentations. There are several contributed packages that extend, improve or ease the Sweave base capabilities. Some of the most valuable ones are: `Hmisc` (Harrell Jr et al., 2014), `knitr` (Xie, 2013a), or `xtable` (Dahl, 2013).

A basic skeleton of a Sweave file is the following:

```
%
\documentclass[a4paper]{article}
\usepackage{Sweave}
\title{<Here a title>}
\author{<Author/s>}
%
\begin{document}
  \maketitle
  %<here LaTeX code and text>
<< [Sweave options] >>=
  #here R code
@
  %More LaTeX code
```

```
\end{document}
```

Chunks of R code are delimited by the lines `<<>=` and `@`. Between them any R command can be included and control the effect in the document through the options written between the `<<` and `>>` symbols on the preamble line of the chunk. Thus, the classes and packages can be loaded with no effect in the pdf document by including the options `echo=FALSE`, `results=hide`. Similarly, the graphics, output or layout can be managed, see the Sweave manual for in depth details in Leisch and R-core (2013). It is good practice to have the logic of the analysis and the results of the research in different files when the code is lengthy. Thus, a large R script may be saved in a separate file, whose contents can be run with a single command within a chunk. For the example, whose complete code can be found in Appendix B, the following pieces will be used:

1. A chunk of code to load the classes, data, and packages to be used;
2. An introductory text and a chart with some of the parameter values;
3. A chunk providing the document with one of the equations of the model;
4. The code to solve the model with GAMS, including the instance file generation;
5. Presentation of the solution in tables and charts.

It is important to remark that these pieces can be as complex as needed, even though for the sake of simplicity very simple ones are used. Furthermore, other formats and solvers might be used, see Chapter 9. Provided the `.Rnw` file is in the R working directory, scripts are in a `data` subdirectory, and GAMS appropriately installed, the report is automatically generated, and therefore any change is updated, just through the following commands:

8. An Integrated Framework

```
Sweave("comprehensiveExample.Rnw")  
library(tools)  
texi2pdf("comprehensiveExample.tex")
```

The content of the `.Rnw` file as well as all the R scripts used can be consulted in Appendix B. The final document obtained after “sweaving” the `.Rnw` file is reproduced in Figures 8.6 to 8.8. Note that it is needed having installed a distribution of \LaTeX , which is freely available for different platforms, as well as some additional R packages, whose installation is straightforward. Appendix C contains the session information in which the example and the rest of the code in this work was executed.

Comprehensive Example for “An Integrated Framework for the Representation and Solution of Stochastic Energy Optimization Problems”

Emilio L. Cano

January 6, 2014

1 Introduction

This document is an example on how to use R as an integrated environment for optimization. It is assumed that the `optimr` package is installed.

Here we can include any statistical analysis, for example a time series analysis to forecast the future energy prices, saving the values as parameters. We can also show graphical representations of the parameter values, as in Figure 1, or tables with data, e.g. Table 1.

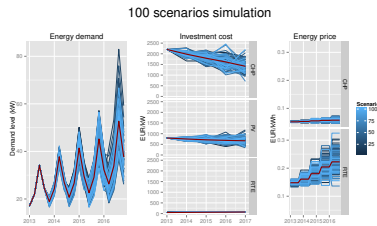


Figure 1: Parameter values for the stochastic parameters.

The equations or any item of the model can be printed automatically from the `model2SMS` object. For example, the following command fetches the objective function:

```
> cat("$$", getEq(model2SMS, 6, "tex", only = "rExpr"), "$$")
```

$$\sum_{n \in \mathcal{N}} P_n \cdot \sum_{i \in \mathcal{I}} \left(\sum_{i \in \mathcal{I}} CI_{i,n}^t \cdot x_i^t + \sum_{i \in \mathcal{I}, j \in \mathcal{J}} CO_{i,j,n}^t \cdot DT_j^t \cdot y_{i,j,n}^t \right)$$

Figure 8.6.: Page 1 of the report generated through the example.

8. An Integrated Framework

	j	n	t	value
1	winter	1	2013	17.06
2	spring	1	2013	21.93
3	summer	1	2013	34.12
4	autumn	1	2013	24.37
5	winter	1	2014	19.89
6	spring	1	2014	25.58

Table 1: Example D parameter values (first 6 values).

2 Solving the problem

Once we have the instance in an `optimInstance` object, it can be solved and the solution imported (see source code). Results checking is also possible as this information is also stored:

We can embed calculations within the text, for example the value of the objective function (68595), or we can print pretty \LaTeX tables with the optimal values, as the ones in Tables 2 and 3, or any other analysis and representation (see Figure 2). See the `.Rnw` source file to see the code.

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

Table 2: Optimal values for x

i	j	n	t	value
RTE	winter	1	2014	2.31
RTE	winter	1	2015	4.27
RTE	winter	1	2016	8.96
RTE	winter	1	2017	7.92
RTE	winter	2	2014	1.76
RTE	winter	2	2015	5.21

Table 3: Optimal values for y (first 6 values)

3 Conclusion

This document can be compiled at any time, by any researcher. Note that if any value is changed, for example in the script that contain the parameters ("`./data/model2Instance2.R`"), the whole report is updated automatically (including tables, equations and charts). If we use simulation during the research, we can simply fix the seed to allow the verification of the results by third parties. Different reports for different stakeholders can be produced using a common structure and tailoring the outputs.

Figure 8.7.: Page 2 of the report generated through the example.

8.4. Illustrative Example

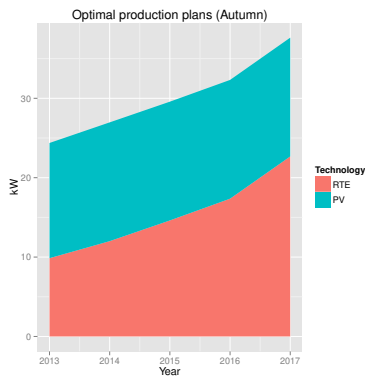


Figure 2: Output data representation.

Figure 8.8.: Page 3 of the report generated through the example.

9. Conclusions

9.1. Results

This thesis tackles Stochastic Optimization (STO) from two quite different scopes. On the one hand, the application to a specific problem such as energy systems planning at the building level. On the other hand, a global perspective of Decision Support System (DSS). Regarding the former, comprehensive models including energy systems modeling, uncertainty, and risk management have been developed and tested on real sites. As for the DSS extent, new ideas have been contributed regarding stakeholders dialog and reproducible research. The following is a summary of the contributions of this work:

- Innovative energy systems modeling at the building level under a holistic approach. The models integrate different types of energy and technologies and include novel features such as technologies' aging, refurbishment, efficiency and emissions in a single model;
- Both the deterministic and the stochastic models have been tested for real sites getting coherent results with the actual performance of the buildings, hence demonstrating their validity;
- The uncertainty modeling demonstrated the usefulness of SP in energy systems optimization and how deterministic models lead to unfeasible scenarios;
- A new application of risk management through CVaR is proposed, not only considering economic risks, but also environ-

9. Conclusions

mental and social risks, i.e., emissions and energy consumption;

- The application of Reproducible Research methods in Operations Research is also something new as a real world implementation;
- This Reproducible Research approach, jointly with the stakeholders dialog viewpoint proposed, must empower DSSs that adopt the developed framework;
- The own library developed for the proposed framework is another outcome, though the framework is suitable to be implemented using further technologies;
- A comprehensive example using the models, the data, and the framework demonstrates a whole reproducible research cycle applying the developments in this thesis.

As an example of these results, Figure 9.1 shows the energy balance for heating in the FASAD EnRiMa test site after running the calibrating optimization with real data and existing technologies. Likewise, Table 9.1 shows the solution results for several values of the risk-aversion weight β in the risk model in Chapter 7. It can be seen how the risk aversion affects the total cost and the operational decisions (Energy generated and energy purchased) induced, in turn, by the strategic ones. The solution time for each run completes the table.

Table 9.1.: Optimization results for the complete risk model.

β	Sol.Time	Cost	En.Gen	En.Pur
0.00	2.84	284,903	96,692	99,867
0.25	3.27	288,037	96,557	100,003
0.50	3.08	291,161	96,215	100,345
0.75	3.18	294,218	95,789	100,770
1.00	4.17	297,277	90,742	105,817

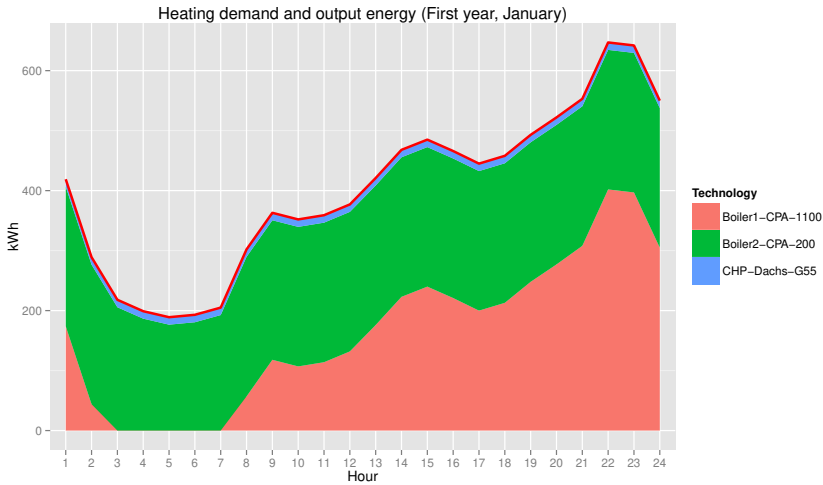


Figure 9.1.: Energy Heating demand and output from technologies in the test site.

9.2. Discussion

9.2.1. Stochastic Energy Systems Planning

Improving energy efficiency in public buildings is a critical component of the EU's policy for reaching its climate target goals for 2020, i.e., reduction by 20% of the total energy consumption, 20% contribution of renewable energies to total energy production, and 20% reduction of greenhouse gases such as CO₂ below 1990 emissions. The EnRiMa project seeks to develop an ICT-based DSS with enhancements to the existing state-of-the-art research in terms of modeling energy flows, generating scenarios for dealing with uncertainties, and handling stochastic optimization at the building level. In this thesis, the focus was on strategic energy planning decisions. To this aim new models have been developed using stochastic programming techniques, which aim to find robust solutions taking into account all potential scenarios. The described models are flexible

9. Conclusions

enough to include strategic decision variables such as equipment installation, financial positions and retrofits involving passive measures, and therefore can be adapted to different configurations of buildings and new arising technologies. The numerical examples presented in this work show how stochastic models can be applied to decision making in strategic energy planning.

9.2.2. Deterministic Modeling

The deterministic version of the strategic model to be used in the DSS of the EnRiMa project was presented in Chapter 5. The approach of the EnRiMa project allows to consider the short-term performance of the building throughout a long-term decision time resolution. The model contains, in a comprehensive way, innovative modeling of several important aspects at the building level, namely: technologies aging, emissions, efficiency, and the energy-balance requirements. It can also deal with several types of energy, as well as different objectives depending on the decision maker interests or needs. An illustrative numerical example was solved combining real data from the project and simulated data to demonstrate the usefulness of the eventual DSS results.

The deterministic model was a preliminary step to develop the final stochastic model presented in Chapter 6. During early stages of the EnRiMa project, a two-stage stochastic programming problem was propounded, see Chapter 4.

9.2.3. Stochastic Multi-stage Modeling

The stochastic strategic model faces uncertainties pertaining mainly to energy prices and building occupancy. This uncertainty is tackled by the generation of scenario trees and a full formulation using a compact structure was presented in Chapter 6. In spite of the model size reduction attained through the use of representative mid-term periods (profiles), the stochastic problem complexity increases very fast with the number of scenarios considered, and can eventually result into computational times which are not acceptable. In

this regard, the multi-level scenario trees generation of the EnRiMa project has been assumed.

The stochastic strategic model was built from the basis of the deterministic problem defined in Chapter 5. Some changes were needed, which are explained in Chapter 6. Risk Management through CVaR was finally incorporated to the model for different objectives in Chapter 7, extending the traditional financial scope to environmental threats considering risk management on emissions and energy consumption.

9.2.4. A Framework for DSS

A research on optimization is a complex process that requires not only getting the optimal values for a given problem, but also a set of complementary tasks aimed at achieving consistent and high-quality results. Energy systems optimization problems can become really huge projects, whose related information may come from several sources, and must result in coherent reports for DSS stakeholders. The traditional “copy-paste” way is clearly inefficient, and innovative techniques as the ones presented in this thesis can improve the achievements of the research, as well as rise the overall satisfaction of stakeholders, e.g., researchers, companies, or customers, through the necessary stakeholders dialog. To this aim, an integrated framework has being developed under the reproducible research approach, which is flexible enough to include different tools in order to cover a wide range of needs.

The integrated framework for DSS proposed in Chapter 8, starts from an atomic representation of the optimization problem components: just symbols and values. Thus, any evolved format can be reached, given that the syntax can be implemented through programming languages. Moreover, by attaining a trade-off between a layer-based research and comprehensive environments, the breakthrough capabilities of reproducible research can be exploited. The environment used for the framework implementation, R, is used all over the world by scientists, statisticians, mathematicians and computer scientists. Besides the statistical and data analysis ca-

9. Conclusions

pabilities, it comprises a programming language. Definitely, R is a good choice to apply the ideas in this thesis. Nonetheless, there is room for improvement. Though recent advances have been made, it is a challenge for R to reach the computational efficiency level of other systems. In this sense, the last versions of R include advanced parallel computing and “Big Data” capabilities, which will lead to overcome the current limitations. Other important issue the R community is taken up is certification. R detractors blame on the “ABSOLUTELY NO WARRANTY” phrase when starting R to use commercial software. But, is there anything more verifiable or certifiable than an open source program? In any case, the goal of this work is to use R as a framework, allowing the researcher using different tools at any stage of the project, including commercial solvers and/or optimization software.

The `optimr` R package is available at the author’s webpage (<http://www.proyectum.es>) and it is already being used in the EnRiMa DSS. However it is not finished work and new features will be added as an aftermath work, see Section 9.3.

As a final consideration, a recall of the examples presented throughout the dissertation and the final report in chapter 8 is proposed. It is a real problem, with thousands of variables, several months of data analysis work, and so on. The real report would probably consist on hundreds of pages. If some time after delivering the report the model must be re-shaped, for example by adding a new set which was not considered at the beginning of the research, a high percentage of the results, both final and intermediate, would change. One can decide how to tackle this issue: starting again to generate partial results and cut-copy-paste throughout the document, or just adding the new information in the correct place, change only the affected parts, and rebuild the report getting automatically the updated layout and content.

9.3. Further Research

The results of this thesis are the starting point for future research. Regarding the optimization models, beyond the EnRiMa project lifetime they will be applied to further public buildings. The undertaking of this task will be defined by the project's exploitation plan, which is being done at the time this thesis is being finished. On the other hand, new fields of application other than public buildings are foreseen by the research group in which the author is involved. At the time this is written, project proposals are being prepared that hopefully will result in new funded projects. Regardless the scenario generation was out of the scope of this thesis, new directions of post-doc research might be new methods of estimation for long-term perspectives for the stochastic parameters considered in the models, specially to determine probability distributions as accurate as possible. Furthermore, the benchmarking between the two-stage approach in Chapter 4 and the multi-stage approach finally adopted in the EnRiMa project, i.e., Chapter 6 model, is likely one of the forthcoming topics to tackle by this researcher.

As for the framework, there is a wide field to develop and extend it. Even though the current implementation includes \LaTeX and \GAMS formats, the extension to other formats for both humans and machines is foreseen, e.g., HTML, GLPK, etc. Even though the package is already available, it will be also contributed to public repositories once the documentation is improved, which is another task to tackle.

Bibliography

- A. Alonso Ayuso, L. Escudero, and C. Pizarro Romero. *Introduction to Stochastic Programming*. Colección Ciencias Experimentales y Tecnología. Editorial Dykinson, S.L., 2009. ISBN 9788498494549.
- P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, (3):203–228, Jul 1999.
- K. A. Baggerly and K. R. Coombes. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *The Annals of Applied Statistics*, 3(4): 1309–1334, Dec. 2009.
- D. Bell, H. Raiffa, and A. Tversky. *Decision Making: Descriptive, Normative, and Prescriptive Interactions*. Cambridge University Press, 1988. ISBN 9780521368513.
- M. Berkelaar and others. *lpSolve: Interface to Lp_solve v. 5.5 to solve linear/integer programs*, 2013. URL <http://CRAN.R-project.org/package=lpSolve>. R package version 5.6.7.
- J. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer series in operations research and financial engineering. Springer New York, 2011.
- J. R. Birge. The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming*, (24): 314–325, 1982.
- J. Bisschop. *Aimms – Optimization Modeling*. Lulu.com, 2006. ISBN 9781411698994.

Bibliography

- J. Bisschop and A. Meeraus. On the development of a general algebraic modeling system in a strategic planning environment. In J.-L. Goffin and J.-M. Rousseau, editors, *Applications*, volume 20 of *Mathematical Programming Studies*, pages 1–29. Springer Berlin Heidelberg, 1982. ISBN 978-3-642-00851-1.
- A. Boza, A. Ortiz, and L. Cuenca. A framework for developing a web-based optimization decision support system for intra/inter-organizational decision-making processes. In A. Ortiz, R. Franco, and P. Gasquet, editors, *Balanced Automation Systems for Future Manufacturing Networks*, volume 322 of *IFIP Advances in Information and Communication Technology*, pages 121–128. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-14340-3.
- W. Buehring, C. Huber, and J. M. de Souza. *Expansion planning for electrical generating systems: a guidebook*. Technical reports series. International Atomic Energy Agency (IAEA), Vienna, 1984.
- Y. P. Cai, G. H. Huang, Z. F. Yang, Q. G. Lin, B. Bass, and Q. Tan. 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.
- E. L. Cano, J. M. Moguerza, and A. Redchuk. *Six Sigma with R. Statistical Engineering for Process Improvement*, volume 36 of *Use R!* Springer, New York, 2012a. ISBN 978-1-4614-3651-5. URL <http://www.springer.com/statistics/book/978-1-4614-3651-5>.
- E. L. Cano, J. M. Moguerza, A. Siddiqui, T. Ermolieva, and Y. Ermoliev. Strategic model for energy systems optimisation: aspects of energy efficiency and risk management. In *eceee 2012 Industrial Summer Study proceedings*, pages 603–610, Arnherm (Netherlands), September 2012b. European Council for an Energy Efficient Economy. ISBN 978-91-980482-0-9 (printed), 978-91-980482-1-6 (digital). URL <http://proceedings.eceee.org/visabstrakt.php?event=2&doc=4-125-12>.

- E. L. Cano, J. M. Moguerza, T. Ermolieva, and Y. Ermoliev. Energy efficiency and risk management in public buildings: strategic model for robust planning. *Computational Management Science*, 11(1-2):25–44, 2014. doi: 10.1007/s10287-013-0177-3. URL <http://dx.doi.org/10.1007/s10287-013-0177-3>.
- CBS. Deception at Duke. TV Report, February 2012. URL <http://www.cbsnews.com/video/watch/?id=7398476n>. [Retrieved 20120626].
- A. J. Conejo, M. Carrión, and R. García-Bertrand. Medium-term electricity trading strategies for producers, consumers and retailers. *International Journal of Electronic Business Management*, 5(3):239–252, 2007.
- A. J. Conejo, M. Carrión, and J. Morales. *Decision Making Under Uncertainty in Electricity Markets*. International Series in Operations Research and Management Science Series. Springer, 2010.
- D. B. Dahl. *xtable: Export tables to LaTeX or HTML*, 2013. URL <http://CRAN.R-project.org/package=xtable>. R package version 1.7-1.
- G. Dantzig. Linear programming under uncertainty. *Management Science*, (1):197–206, 1955.
- G. Dantzig. *Linear Programming and Extensions*. Landmarks in Physics and Mathematics. Princeton University Press, 1965. ISBN 9780691059136.
- G. Dantzig and M. Thapa. *Linear Programming: 1: Introduction*. Springer, 1997. ISBN 9780387948331.
- G. Dantzig and M. Thapa. *Linear Programming: 2: Theory and Extensions*. Springer, 2003. ISBN 9780387986135.

Bibliography

- E. Delage, S. Arroyo, and Y. Ye. The value of stochastic modeling in two-stage stochastic programs with cost uncertainty. Technical Report G-2012-05, GERAD (HEC Montreal), 2012.
- D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2): 548–566, Feb 2003.
- D. Dentcheva and A. Ruszczyński. Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance*, 30(2):433–451, 2006.
- S. Dirkse and R. Jain. *gdxrrw: An interface between GAMS and R*, 2013. URL http://support.gams-software.com/doku.php?id=gdxrrw:interfacing_gams_and_r. R package version 0.3.1.
- D. Duffie and J. Pan. An overview of value-at-risk. *Journal of Derivatives*, 1997(4):7–49, 1997.
- J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming. *Mathematical Programming*, 95(3): 493–511, 2003. doi: 10.1007/s10107-002-0331-0.
- W. El-Khattam, K. Bhattacharya, Y. Hegazy, and M. Salama. Optimal investment planning for distributed generation in a competitive electricity market. *Power Systems, IEEE Transactions on*, 19(3):1674–1684, aug. 2004.
- F. Engdahl and D. Johansson. Optimal supply air temperature with respect to energy use in a variable air volume system. *Energy and Buildings*, 36(3):205–218, 2004.
- EnRiMa project consortium. Energy efficiency and risk management in public buildings (EnRiMa), 2010-2014. URL <http://www.enrima-project.eu>. FP7 Research project number 260041.
- Y. Ermoliev and R. Wets, editors. *Numerical techniques for stochastic optimization*. Springer series in Computational Mathematics. Springer-Verlag, 1988.

- Y. Ermoliev, M. Makowski, and K. Marti, editors. *Managing Safety of Heterogeneous Systems: Decisions under Uncertainties and Risks*. Lecture notes in economics and mathematical systems. Springer, 2012a. ISBN 9783642228841.
- Y. Ermoliev, M. Makowski, and K. Marti. *Managing Safety of Heterogeneous Systems*, chapter Robust management of heterogeneous systems under uncertainties. Springer-Verlag, Heidelberg, Germany, February 2012b.
- L. Escudero, A. Garín, M. Merino, and G. Pérez. The value of the stochastic solution in multistage problems. *TOP*, 15(1):48–64, 2007. ISSN 1134-5764.
- A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Research Analysis Corporation, 1968. Republished in 1990 by the Society for Industrial and Applied Mathematics.
- R. Fourer. Modeling languages versus matrix generators for linear programming. *ACM Trans. Math. Softw.*, 9(2):143–183, 1983.
- R. Fourer, D. M. Gay, and B. W. Kernighan. A modeling language for mathematical programming. *Management Science*, 36(5):519–554, 1990.
- R. Fourer, J. Ma, and K. Martin. Optimization services: A framework for distributed optimization. *Operations Research*, 58(6):1624–1636, 2010. URL <http://or.journal.informs.org/content/58/6/1624.abstract>.
- S. French. Decision support systems. In D. Rios Insua and S. French, editors, *e-Democracy*, volume 5 of *Advances in Group Decision and Negotiation*, pages 65–82. Springer Netherlands, 2010. ISBN 978-90-481-9044-7. doi: 10.1007/978-90-481-9045-4_5.
- S. French, J. Maule, and N. Papamichail. *Decision Behaviour, Analysis and Support*. Cambridge University Press, 2009. ISBN 9780521883344.

Bibliography

- L. Garcés, A. Conejo, R. García-Bertrand, and R. Romero. A bilevel approach to transmission expansion planning within a market environment. *Power Systems, IEEE Transactions on*, 24(3):1513–1522, 2009.
- R. Gentleman and D. Temple Lang. Statistical analyses and reproducible research. *Journal of Computational and Graphical Statistics*, 16(1):1–23, 2007.
- A. M. Geoffrion. The SML language for structured modeling: Levels 1 and 2. *Operations Research*, 40(1):38–57, 1992a.
- A. M. Geoffrion. The SML language for structured modeling: Levels 3 and 4. *Operations Research*, 40(1):58–75, 1992b.
- P. Gill, W. Murray, M. Saunders, J. Tomlin, and M. Wright. On projected newton barrier methods for linear programming and an equivalence to karmarkar’s projective method. *Mathematical Programming*, 36(2):183–209, 1986. doi: 10.1007/BF02592025. URL <http://dx.doi.org/10.1007/BF02592025>.
- J. R. González, D. A. Pelta, and A. D. Masegosa. A framework for developing optimization-based decision support systems. *Expert Systems with Applications*, 36(3, Part 1):4581 – 4588, 2009.
- A. Gritsevskii and Y. Ermoliev. An energy model incorporating technological uncertainty, increasing returns and economic and environmental risks. In *Proceedings of the International Association for Energy Economics 1999 European Conference, 30 September - 1 October, Paris, France, 1999*.
- A. Gritsevskii and Y. Ermoliev. *Managing Safety of Heterogeneous Systems*, chapter Modeling technological change under increasing returns and uncertainty, pages 109–136. Springer-Verlag, Heidelberg, Germany, February 2012.
- A. Gritsevskii and N. Nakicenovic. Modeling uncertainty of induced technological change. *Energy Policy*, 28(13):907–921, 2000.

- M. Groissböck, S. Heydari, A. Mera, E. Perea, A. Siddiqui, and M. Stadler. Optimizing building energy operations via dynamic zonal temperature settings. *Journal of Energy Engineering*, 2013a. doi: 10.1061/(ASCE)EY.1943-7897.0000143.
- M. Groissböck, E. López, E. Perea, A. Siddiqui, and A. Werner. Improving energy efficiency and risk management in EU public buildings. *IAEE Energy Forum*, 22:17–20, 2013b. URL <http://www.iaee.org/en/publications/newsletterdl.aspx?id=193>.
- L. Hardesty. US favors Net-Metering while Europe, Japan like feed-in-tariffs. online [accessed 2013-07-25], May 2013. URL <http://goo.gl/xCNiAf>.
- F. E. Harrell Jr, C. Dupont, and other unspecified contributors. *Hmisc: Harrell Miscellaneous*, 2014. URL <http://CRAN.R-project.org/package=Hmisc>. R package version 3.14-0.
- W. Hart, C. Laird, J. Watson, and D. Woodruff. *Pyomo – Optimization Modeling in Python*. Springer optimization and its applications. Springer, 2012. ISBN 9781461432265.
- HCE, CET, TecNALIA, IIASA, SU, UCL, URJC, and SINTEF. Requirement assessment. EnRiMa Project Deliverable D1.1, December 2011.
- D. Henning. MODEST – An energy-system optimisation model applicable to local utilities and countries. *Energy*, 22(12):1135 – 1150, 1997. doi: 10.1016/S0360-5442(97)00052-2.
- P. Hernandez and P. Kenny. From net energy to zero energy buildings: Defining life cycle zero energy buildings (LC-ZEB). *Energy and Buildings*, 42(6):815 – 821, 2010.
- S. Heydari and A. Siddiqui. 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.

Bibliography

- B. Hobbs. Optimization methods for electric utility resource planning. *European Journal of Operational Research*, 83(1):1–20, 1995.
- K. Hornik. The R FAQ, 2013. URL <http://CRAN.R-project.org/doc/FAQ/R-FAQ.html>.
- International Atomic Energy Agency. *Wien Automatic System Planning (WASP) Package. A Computer Code for Power Generating System Expansion Planning*. International Atomic Energy Agency, Vienna, 2001. Version WASP-IV User’s Manual.
- ISO. Statistics – vocabulary and symbols – part 1: General statistical terms and terms used in probability. Standard ISO 3534-1:2006, International Organization for Standardization, Geneva, Switzerland, 2006.
- T. Jamasb and M. Pollitt. 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/journal/2005se-a02.html>.
- R. Jelliffe. Climate wars: Global warming, climate-gate, 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].
- P. Kall and J. Mayer. *Stochastic Linear Programming. Models, Theory, and Computation*. Springer, 2005.
- J. Kallrath, editor. *Algebraic Modeling Systems: Modeling and Solving Real World Optimization Problems*. Applied Optimization. Springer Berlin Heidelberg, 2012a.
- J. Kallrath. A practioner’s wish list towards algebraic modeling systems. In J. Kallrath, editor, *Algebraic Modeling Systems*, volume 104 of *Applied Optimization*, pages 213–222. Springer Berlin Heidelberg, 2012b.

- J. Kallrath. Algebraic modeling languages: Introduction and overview. In J. Kallrath, editor, *Algebraic Modeling Systems*, volume 104 of *Applied Optimization*, pages 3–10. Springer Berlin Heidelberg, 2012c.
- B. G. Karlsson, M. Soderstrom, and D. Henning. Simulation of the nuclear phase-out by direction of the energy commission. Technical Report LiTH-IKP-R-888, Dept. of Mech. Engng., Linköping Institute of Technology, Sweden, 1995.
- N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984. doi: 10.1007/BF02579150. URL <http://dx.doi.org/10.1007/BF02579150>.
- M. Kaut, K. Midthun, A. Werner, A. Tomaszewski, L. Hellemo, and M. Fodstad. Multi-horizon stochastic programming. *Computational Management Science*, 2013. doi: 10.1007/s10287-013-0182-6.
- D. King and M. Morgan. Adaptive-focused assessment of electric power microgrids. *Journal of Energy Engineering*, 133(3):150–164, 2007.
- G. Klein, J. Orasanu, and R. Calderwood. *Decision Making in Action: Models and Methods*. Cognition and Literacy. Ablex Publishing Corporation, 1993. ISBN 9780893919436.
- D. E. Knuth. Literate programming. *The Computer Journal*, 27(2): 97–111, 1984.
- R. Koenker and A. Zeileis. On reproducible econometric research. *Journal of Applied Econometrics*, 24(5):833–847, 2009.
- P. Krokhmal, M. Zabaranin, and S. Uryasev. Modeling and optimization of risk. *Surveys in Operations Research and Management Science*, 16(2011):49–66, 2011.

Bibliography

- M. Kuhn. CRAN Task View: Reproducible Research. <http://cran.r-project.org/web/views/ReproducibleResearch.html>, 2014. [retrieved 2014-01-31].
- C. Kuip. Algebraic languages for mathematical programming. *European Journal of Operational Research*, 67(1):25 – 51, 1993.
- G. Kumbaroğlu and R. Madlener. 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.
- F. Leisch. Sweave: Dynamic generation of statistical reports using literate data analysis. In W. Härdle and B. Rönz, editors, *Compstat 2002 — Proceedings in Computational Statistics*, pages 575–580. Physica Verlag, Heidelberg, 2002. URL <http://www.stat.uni-muenchen.de/~leisch/Sweave>.
- F. Leisch and R-core. *Sweave User Manual*, December 2013. URL <http://stat.ethz.ch/R-manual/R-devel/library/utils/doc/Sweave.pdf>.
- Y. Liang, D. Levine, , and Z. Shen. Thermostats for the smart-grid: Models, benchmarks, and insights. Working Paper, 2011. University of California Energy Institute.
- H. Lund. Large-scale integration of wind power into different energy systems. *Energy*, 30(13):2402 – 2412, 2005.
- A. Madansky. Inequalities for stochastic linear programming problems. *Management Science*, (6):197–204, 1960.
- H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Wiley, New York, 1959.
- C. Marnay, G. Venkatarmanan, M. Stadler, A. Siddiqui, R. Firestone, and B. Chandran. Optimal technology selection and operation of commercial-building microgrids. *IEEE Transactions on Power Systems*, 23(3):975–982, 2008.

- K. Marti. *Stochastic Optimization Methods*. Springer, 2008. ISBN 9783540794578.
- K. Marti, Y. Ermoliev, M. Makowski, and G. Pflug, editors. *Coping with Uncertainty: Modeling and Policy Issues*. Lecture notes in economics and mathematical systems. Springer, 2006. ISBN 9783540352624.
- K. Marti, Y. Ermoliev, and M. Makowski, editors. *Coping with Uncertainty: Robust Solutions*. Lecture notes in economics and mathematical systems. Springer, 2010. ISBN 9783642037351.
- A. Neumaier. FMathL - formal mathematical language, 2014. URL <http://www.mat.univie.ac.at/~neum/FMathL.html>. [retrieved 2014-01-31].
- W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: semideviations as risk measures. *European Journal of Operational Research*, (116):33–50, 1999.
- W. Ogryczak and A. Ruszczyński. Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization*, (13): 60–78, 2002.
- H.-J. Pitz. Data preparation for optimization with R. In J. Kallrath, editor, *Algebraic Modeling Systems*, volume 104 of *Applied Optimization*, pages 171–183. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-23592-4_9.
- G. Platt, J. Li, R. Li, G. Poulton, G. James, and J. Wall. Adaptive hvac zone modeling for sustainable buildings. *Energy and Buildings*, 42:412–421, 2010.
- S. Pless and P. Torcellini. 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.

Bibliography

- D. Power. A brief history of decision support systems. DSS-Resources.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, 2013a. URL <http://www.R-project.org/>. [software version 3.0.2].
- R Core Team. *R Language Definition*. R Foundation for Statistical Computing, Vienna, Austria, 2013b. URL <http://www.R-project.org/>. [software version 3.0.2].
- N. Ramsey. *Noweb man page*. University of Virginia, USA, 1998. URL <http://www.cs.tufts.edu/~nr/noweb/>.
- T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41, 2000.
- K. A. Salewicz and M. Nakayama. Development of a web-based decision support system (DSS) for managing large international rivers. *Global Environmental Change*, 14, Supplement(0):25 – 37, 2004.
- M. Salvador and S. Grieu. 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.
- T. Schleiff and M. Steglich. *CMPL (<Coliop/Coin> Mathematical Programming Language)*. COIN-OR, June 2013. URL <http://cml.th-wildau.de/download/CMPL.pdf>. Manual Version 1.8.0.
- M. Schmidt. The sankey diagram in energy and material flow management. *Journal of Industrial Ecology*, 12(1):82–94, 2008. doi: 10.1111/j.1530-9290.2008.00004.x.

- J. Shapiro. *Mathematical programming: structures and algorithms*. A Wiley-Interscience publication. Wiley, 1979. ISBN 9780471778868.
- J. Shim, M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda, and C. Carlsson. Past, present, and future of decision support technology. *Decision Support Systems*, 33(2):111 – 126, 2002.
- A. Siddiqui, C. Marnay, J. Edwards, R. Firestone, S. Ghosh, and M. Stadler. Effects of carbon tax on combined heat and power adoption. *Journal of Energy Engineering*, 131(1):2–25, 2005.
- A. Siddiqui, C. Marnay, R. Firestone, and N. Zhou. Distributed generation with heat recovery and storage. *Journal of Industrial Engineering*, 133:181–210, 2007.
- M. Stadler, A. S. Siddiqui, C. Marnay, H. Aki, and J. Lai. 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.
- V. Stodden. Enabling Reproducible Research: Open Licensing for Scientific Innovation. *Social Science Research Network Working Paper Series*, Mar. 2009.
- V. Stodden, F. Leisch, and R. D. Peng, editors. *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2013.
- H. Taha. *Operations Research: An Introduction*. Pearson Education, Limited, 2011. ISBN 9780131391994.
- M. Tanaka, H. Watanabe, Y. Furukawa, and T. Tanino. 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.

Bibliography

- 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].
- S. Theussl. CRAN Task View: Optimization and Mathematical Programming. Internet, 2014. URL <http://cran.r-project.org/web/views/Optimization.html>. [retrieved 2014-01-31].
- S. Theussl and K. Hornik. *Rglpk: R/GNU Linear Programming Kit Interface*, 2013. URL <http://CRAN.R-project.org/package=Rglpk>. R package version 0.5-2.
- UCL, URJC, CET, Tecalia, IIASA, SU, SINTEF, and HCE. A mathematical formulation of energy balance and flow constraints. EnRiMa Project Deliverable D2.2, March 2012. URL <http://www.enrima-project.eu/sites/default/files/D2.2.pdf>.
- URJC, HCE, IIASA, SU, UCL, SINTEF, CET, and Tecalia. Symbolic model specification. EnRiMa Project Deliverable D4.2, March 2012. URL <http://www.enrima-project.eu/sites/default/files/D4.2.pdf>.
- E. Van Sambeek. Distributed generation in competitive electricity markets. Working Paper no. 00-S4, 2000. Center for Energy and Environmental Policy, Newark, DE, USA.
- P. Vandewalle, J. Kovacevic, and M. Vetterli. Reproducible research in signal processing - what, why, and how. *IEEE Signal Processing Magazine*, 26(3):37–47, May 2009. URL <http://rr.epfl.ch/17/>.
- J. Villumsen and A. Philpott. Investment in electricity networks with transmission switching. *European Journal of Operational Research*, 222(2):377 – 385, 2012.
- J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 1953.

- C. Weinberg, J. Iannucci, and M. Reading. The distributed utility: Technology, customer, and public policy changes shaping the electrical utility of tomorrow. *Energy Systems Policy*, 15(4):307–322, 1991.
- Y. Xie. *knitr: A general-purpose package for dynamic report generation in R*. R, 2013a. package version 1.5.
- Y. Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, 2013b.
- Y. Xie. knitr: A comprehensive tool for reproducible research in R. In V. Stodden, F. Leisch, and R. D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2013c.
- B. Xu, L. Fu, and H. Di. Dynamic simulation of space heating systems with radiators controlled by trvs in buildings. *Energy and Buildings*, 40:1755–1764, 2008.

Appendices

A. Glossary

Alias A second name for a set, usually used in order to distinguish between different parameter or variable indices within an equation, or even within the same variable or parameter. For example, for energy input and energy output.

Conditional set A special type of subset whose content depends on other sets. Example: Input (or output) energy for each technology. When the conditional set only contains one element, it can be seen as a function of other sets' elements.

Constant A fixed scalar value that does not depend on any set index.

Constraint An equation that restricts the possible values of variables.

DSS Kernel The DSS component that is responsible for providing functionality needed to manage the system data and to run the scenario generator tool and solver.

Equation A mathematical relationship between parameters and variables. For simplicity, we gather both mathematical concepts of equation (equality relationship) and inequation (inequality relationship) under the common notion "equation".

Expression Combination of variable and parameter symbols with their set symbols as super- or sub-indices.

Multidimensional set A way of implementing conditional sets in the optimizer. Multidimensional sets provide mappings between elements of different sets. Thus, one of the dimensions

A. Glossary

is a subset of a given set and the rest of the dimensions are the values of other sets that are related to the former. There may be one-to-one (e.g., parent nodes) and many-to-many (e.g., output types of energy for a technology) mappings.

Objective A function of variables and parameters that the decision maker wants to optimize (minimize or maximize).

Optimizer Software application capable to deal with optimizations problems. Optimizers usually can be configured to use different *solvers*. Besides the optimization itself, these applications can perform other related tasks, such as modeling, data analysis, or data visualization. Examples: GAMS, AMPL, Matlab, Pyomo.

Parameter A known or uncertain characteristic inherent to a set element or a combination of set elements, which is fixed and cannot be changed by the decision maker. Examples: Generation capacity of a given model of PV generator, the energy demand during a specific time span.

Set Collection of elements of a given class that are related and combined with other sets elements to parameters and decision variables. Example: technologies, types of energy.

Solver Computational algorithm that solves an optimization problem, i.e., receives input data and returns a problem solution. Examples: CPLEX, lp_solve.

Solver Manager (SM) The module of the EnRiMa DSS that runs the Optimizer using the input provided by the Interface and prepares the solution returned by the Optimizer for the Interface.

Solver Manager Core Script The main script within the Solver Manager. It is started by the SM Interface, prepares input and output data, manages files, and interacts with the Optimizer.

Solver Manager Interface The component of the Solver Manager that is responsible for the communication between the core of the Solver Manager and the DSS Kernel.

Stochastic Optimization For the scope of this document, *Stochastic Optimization* has the same meaning as *Solver Manager*, for consistency with the deliverable names in the DoW. Stochastic optimization model, stochastic optimization problem, and similar compound forms refer to the inherent mathematical concepts.

Subset Elements in a set that are similar with respect to some characteristic. Example: Generation technologies, storage technologies.

Symbol The representation for a set, variable, parameter, or equation within a SMS.

Symbolic Model Specification The mathematical representation (composed of variables, parameters, and relations between them) of the stochastic model of all relevant energy subsystems and their interactions.

Variable A variable, or decision variable, is a characteristic of a set element or combination of set elements, which is unknown and that the decision maker can change, given that the constraints are satisfied. Example: number of PV panels to be installed at a given year. For convenience, *calculated variables* can be defined through an equation. For example, the available capacity of a given technology during a given year is calculated from other variables and parameters.

B. Code for example in Chapter 8

B.1. File comprehensiveExample.Rnw

```
\documentclass[a4paper]{article}
\usepackage{Sweave}

\title{Comprehensive Example for
``An Integrated Framework for the
Representation and Solution of Stochastic
Energy Optimization Problems''}
\author{Emilio L. Cano}

<<intro , echo=FALSE, results=hide>>=
## System requirements
## - The R software and the packages loaded below
## - A licenced GAMS installation. If GAMS directory
##   is other than "~/app/gams23.9", change the line
##   'igdx("~/app/gams23.9")'
## - A LaTeX distribution for the current system

## Load needed packages
library(knitr)
library(optimr)
library(gdxrrw)
library(xtable)
library(ggplot2)
library(grid)

## Tell gdxrrw where GAMS is installed
igdx("~/app/gams23.9")

## Run scripts with data
## Deterministic model
source("../data/model1SMS.R")
## Stochastic extension
source("../data/model2SMS.R")
## Instance with 100 scenarios
source("../data/model2Instance2.R")
@
```

B. Code for example in Chapter 8

```
\begin{document}
\SweaveOpts{concordance=TRUE}

\maketitle

\section{Introduction}
This document is an example on how to use \textsf{R} as an
integrated
environment for optimization. It is assumed that the \textsf{R}
\textsf{optimr} package is installed.

Here we can include any statistical analysis, for example a
time series analysis
to forecast the future energy prices, saving the values as
parameters. We can
also show graphical representations of the parameter values,
as in Figure~\ref{fig:examplepar}, or tables with data, e.
g. Table~\ref{tab:exampletable}.

\begin{figure}[htp]
\begin{center}
<<examplepar, fig=TRUE, echo=FALSE, width=10>>=
## Demand
dtoPlot <- instancePars(model2Instance2, "D")
dtoPlot$id <- 1:20
pD <- ggplot(data = dtoPlot, aes(x = id, y = value, group=n,
col=n))
pD <- pD + geom_path()
pD <- pD + scale_x_discrete(name = "", breaks = seq(1,21, by
=5), labels = 2013:2017)
pD <- pD + ggtitle("Energy demand")
pD <- pD + scale_y_continuous(name = "Demand level (kW)")
pD <- pD + theme(legend.position = "none")
pD <- pD + stat_summary(fun.y=mean, colour="darkred", geom="
line", aes(group=1), size=1)

## Investment cost
dtoPlot <- instancePars(model2Instance2, "CI")
pCI <- ggplot(data = dtoPlot, aes(x = t, y = value, group=n,
col=n))
pCI <- pCI + geom_path()
pCI <- pCI + facet_grid(i ~ .)
pCI <- pCI + scale_x_continuous(name = "")
pCI <- pCI + ggtitle("Investment cost")
pCI <- pCI + scale_y_continuous(name = "EUR/kW")
pCI <- pCI + theme(legend.position = "none")
pCI <- pCI + stat_summary(fun.y=mean, colour="darkred", geom="
line", aes(group=1), size=1)

## Operation cost
dtoPlot <- instancePars(model2Instance2, "CO")
dtoPlot$id <- 1:20
pCO <- ggplot(data = dtoPlot, aes(x = id, y = value, group=n,
```

B.1. File `comprehensiveExample.Rnw`

```
col=n))
pCO <- pCO + geom_path()
pCO <- pCO + facet_grid(i ~ .)
pCO <- pCO + scale_x_discrete(name = "", breaks = seq(1,21, by
=5), labels = 2013:2017)
pCO <- pCO + ggtitle("Energy price")
pCO <- pCO + scale_y_continuous(name ="EUR/kWh")
pCO <- pCO + guides(color = guide_colourbar(title = "Scenario
"))
pCO <- pCO + stat_summary(fun.y=mean, colour="darkred", geom="
line", aes(group=1), size=1)

grid.newpage()
vpAll <- viewport(layout = grid.layout(2, 3,
widths = c(1/3, 1/3, 1/3),
heights = c(0.1, 0.9)))
vpT <- viewport(layout.pos.col = 1:3, layout.pos.row = 1, name
= "title")
vpD <- viewport(layout.pos.col = 1, layout.pos.row = 2, name =
"D")
vpCI <- viewport(layout.pos.col = 2, layout.pos.row = 2, name
= "CI")
vpCO <- viewport(layout.pos.col = 3, layout.pos.row = 2, name
= "CO")
splot <- vpTree(vpAll, vpList(vpT, vpD, vpCI, vpCO))
pushViewport(splot)
seekViewport("title")
grid.text("100 scenarios simulation", gp = gpar(cex=2))
seekViewport("D")
print(pD, newpage = FALSE)
seekViewport("CI")
print(pCI, newpage = FALSE)
seekViewport("CO")
print(pCO, newpage = FALSE)
@

\caption{Parameter values for the stochastic parameters.}
\label{fig:examplepar}
\end{center}
\end{figure}

<<echo=FALSE, results=tex>>=
xtable(head(instancePars(model2Instance2, "D")),
label = "tab:exampletable",
caption = "Example $D$ parameter values (first 6 values
).")
@
```

The equations or any item of the model can be printed automatically from the `\textsf{model2SMS}` object. For example, the following command fetches the objective function:

B. Code for example in Chapter 8

```
<<results=tex>>=
cat("$$", getEq(model2SMS, 6, "tex", only = "rExpr"), "$$")
@
```

```
\section{Solving the problem}
```

Once we have the instance in an `\texttt{optimInstance}` object, it can be solved and the solution imported (see source code). Results checking is also possible as this information is also stored:

```
<<sol, echo=FALSE, results=hide>>=
wProblem(model2Instance2,
  filename = "./data/model2Instance2.gms",
  format = "gams",
  solver = "LP")
res <- gams("./data/model2Instance2.gms --outfile=./data/
  model2Instance2.gdx")
data(gamsOut)
if (res == 0){
  importGams(model2Instance2) <- "./data/model2Instance2.gdx"
  message("Optimization ok\n",
    "\tModel Status: ",
    as.character(subset(gamsModelStatusCode,
      id = model2Instance2@result$model, desc, drop =
        TRUE))),
    "\n\tSolver Status: ",
    as.character(subset(gamsSolverStatusCode,
      id = model2Instance2@result$solve, desc, drop =
        TRUE)))
} else{
  warning("Check the listing file, something was wrong: ",
    subset(gamsOutCode, id == res, desc, drop = TRUE))
}
@
```

We can embed calculations within the text, for example the value of the objective function (`\Sexpr{round(model2Instance2@result$obj)}`), or we can print pretty `\LaTeX` tables with the optimal values, as the ones in Tables `\ref{tab:x}` and `\ref{tab:y}`, or any other analysis and representation (see Figure `\ref{fig:res}`). See the `\texttt{.Rnw}` source file to see the code.

```
<<results=tex, echo=FALSE>>=
print(xtable(instanceVars(model2Instance2, "x"),
  "Optimal values for  $x$ ",
  "tab:x"), include.rownames = FALSE)
print(xtable(head(instanceVars(model2Instance2, "y")),
  "Optimal values for  $y$  (first 6 values)",
  "tab:y"), include.rownames = FALSE)
@
```

B.2. File model1SMS.R

```
\begin{figure}[htp]
\begin{center}
<<bar, echo=FALSE, fig=TRUE>>=
df2plot <- subset(instanceVars(model2Instance2, "y"), j == "
  autumn")
df2plot <- aggregate(value ~ i + t, data = df2plot, FUN = mean
)
df2plot$t <- as.integer(as.character(df2plot$t))
p <- ggplot(df2plot, aes(x=t))
p <- p + geom_area(aes(y=value, fill=i))
p <- p + labs(title = "Optimal production plans (Autumn)", x =
  "Year", y = "kW")
p <- p + scale_fill_discrete("Technology")
print(p)
@
\end{center}
\caption{Output data representation.\label{fig:res}}
\end{figure}

\section{Conclusion}
This document can be compiled at any time, by any researcher.
  Note that if any
value is changed, for example in the script that contain the
parameters
(\texttt{"../data/model2Instance2.R"}), the whole report is
  updated automatically
(including tables, equations and charts).
If we use simulation during the research, we can simply fix
  the seed to allow
the verification of the results by third parties. Different
  reports for different stakeholders can be produced using a
  common structure and tailoring the outputs.

\end{document}
```

B.2. File model1SMS.R

```
## SMS
model1SMS <- newSMS("Deterministic1",
  "A Basic Case",
  "The simplest model only with electricity
  ")

## Sets

newSMSset(model1SMS, list(symbol = "i", sDes = "Technology",
  setType = "set"))
addItem(model1SMS, "sets") <- list(symbol = "j", sDes = "
  Period", setType = "set")
addItem(model1SMS, "sets") <- list(symbol = "t", sDes = "Year
```

B. Code for example in Chapter 8

```
    ", loc = "sup", setType = "set")

##Variables

newSMSvar(modelSMS, list(
  symbol = "x",
  sDes = "Capacity to be installed",
  units = "kW",
  positive = TRUE,
  ind = as.array(list(c(1,3))))))
addItem(modelSMS, "vars") <- list(
  symbol = "y",
  sDes = "Production plan",
  units = "kW",
  positive = TRUE,
  ind = as.array(list(c(1,2,3))))
addItem(modelSMS, "vars") <- list(
  symbol = "s",
  sDes = "Available capacity",
  units = "kW",
  positive = TRUE,
  ind = as.array(list(c(1,3))))

## Constants

#unity constant to get previous year
addItem(modelSMS, "consts") <- list(
  symbol = "1",
  value = 1)

#Parameter for life time
addItem(modelSMS, "pars") <- list(
  symbol = "LT",
  sDes = "Lifetime",
  units = "years",
  ind = as.array(list(c(1))))

#Equation to calculate precedent year:  $t = t-1$ 
addItem(modelSMS, "eqs") <- list(
  symbol = "aux1",
  sDes = "The precedent year",
  relation = "eq",
  nature = "aux")
addItem(modelSMS, "terms") <- list(
  eq = 1,
  side = "l",
  nature = "sets",
  item = 3)
addItem(modelSMS, "terms") <- list(
  eq = 1,
  side = "r",
  nature = "sets",
  parent = 0,
```


B.2. File model1SMS.R

```
    item = 3)
addItem(model1SMS, "terms") <- list(
  eq = 1,
  side = "r",
  nature = "consts",
  parent = 0,
  sign = "-",
  item = 1)

#Equation to get when to decommission obsolete technology
addItem(model1SMS, "eqs") <- list(
  symbol = "aux2",
  sDes = "Obsolete Technology decommission",
  relation = "eq",
  nature = "aux")
addItem(model1SMS, "terms") <- list(
  eq = 2,
  side = "l",
  nature = "sets",
  item = 3)
addItem(model1SMS, "terms") <- list(
  eq = 2,
  side = "r",
  nature = "sets",
  parent = 0,
  item = 3)
addItem(model1SMS, "terms") <- list(
  eq = 2,
  side = "r",
  nature = "pars",
  sign = "-",
  parent = 0,
  item = 1)

#Equation to calculate the available capacity
addItem(model1SMS, "eqs") <- list(
  symbol = "eqAvail",
  sDes = "Available technologies capacity calculation",
  relation = "eq",
  nature = "constraint",
  domain = as.array(list(c(1,3))))
addItem(model1SMS, "terms") <- list(
  eq = 3,
  side = "l",
  nature = "vars",
  item = 3)
addItem(model1SMS, "terms") <- list(
  eq = 3,
  side = "r",
  nature = "vars",
  parent = 0,
  item = 3,
  setSubEq = as.array(list(c(1))))
addItem(model1SMS, "terms") <- list(
```

B. Code for example in Chapter 8

```
eq = 3,
side = "r",
nature = "vars",
item = 1,
parent = 0,
sign = "+"
addItem(modelSMS, "terms") <- list(
  eq = 3,
  side = "r",
  nature = "vars",
  item = 1,
  parent = 0,
  sign = "-",
  setSubEq = as.array(list(c(2))))

#Parameter for demand
newSMSpar(modelSMS, list(
  symbol = "D",
  sDes = "Demand Level",
  units = "kW",
  ind = as.array(list(c(2,3)))))

#Equation for demand
addItem(modelSMS, "eqs") <- list(
  symbol = "eqDemand",
  sDes = "Production plan for demand",
  relation = "eq",
  nature = "constraint",
  domain = as.array(list(c(2,3))))
addItem(modelSMS, "terms") <- list(
  eq = 4,
  side = "l",
  nature = "vars",
  setSums = as.array(list(c(1))),
  item = 2)
addItem(modelSMS, "terms") <- list(
  eq = 4,
  side = "r",
  nature = "pars",
  item = 2)

#Parameter for availability
addItem(modelSMS, "pars") <- list(
  symbol = "G",
  sDes = "Technology Availability",
  units = "kW/kW",
  ind = as.array(list(c(1,2,3))))

#Equation for capacity
newSMSeq(modelSMS,
  list(id = 5,
    symbol = "eqCapacity",
    sDes = "Technologies capacity",
```

B.2. File model1SMS.R

```
      relation = "lte",
      nature = "constraint",
      domain = as.array(list(c(1,2,3))),
list(id = 1,
     side = "l",
     nature = "vars",
     item = 2),
list(id = 2,
     side = "r",
     nature = "pars",
     item = 3),
list(id = 3,
     side = "r",
     nature = "vars",
     item = 3,
     parent = 2))

#Variable cost
addItem(model1SMS, "vars") <- list(
  symbol = "z",
  sDes = "Total cost",
  units = "EUR",
  positive = FALSE,
  nature = "objective")

#Parameters for investment and operation cost
addItem(model1SMS, "pars") <- list(
  symbol = "CI",
  sDes = "Investment Cost",
  units = "EUR/kW",
  ind = as.array(list(c(1,3)))
addItem(model1SMS, "pars") <- list(
  symbol = "CO",
  sDes = "Operational Cost",
  units = "EUR/kWh",
  ind = as.array(list(c(1,2,3)))

addItem(model1SMS, "pars") <- list(
  symbol = "DT",
  sDes = "Duration time of period",
  units = "hours",
  ind = as.array(list(c(2,3)))

#Equation for total cost
addItem(model1SMS, "eqs") <- list(
  symbol = "Cost",
  sDes = "Total Cost",
  nature = "objective",
  relation = "eq",
  objective = "min")
addItem(model1SMS, "terms") <- list( #1.z
  eq = 6,
```

B. Code for example in Chapter 8

```
    side = "l",
    nature = "vars",
    item = 4)
addItem(model1SMS, "terms") <- list( #2.1
  eq = 6,
  side = "r",
  nature = "consts",
  item = 1,
  setSums = array(list(c(3))))
addItem(model1SMS, "terms") <- list( #3.CI
  eq = 6,
  side = "r",
  nature = "pars",
  parent = 2,
  item = 4,
  setSums = as.array(list(c(1))))
addItem(model1SMS, "terms") <- list( #4.x
  eq = 6,
  side = "r",
  nature = "vars",
  item = 1,
  parent = 3)

addItem(model1SMS, "terms") <- list( #5.CO
  eq = 6,
  side = "r",
  nature = "pars",
  parent = 2,
  item = 5,
  sign = "+",
  setSums = as.array(list(c(1,2))))
addItem(model1SMS, "terms") <- list( #6.DT
  eq = 6,
  side = "r",
  nature = "pars",
  item = 6,
  parent = 5)
addItem(model1SMS, "terms") <- list( #7.y
  eq = 6,
  side = "r",
  nature = "vars",
  item = 2,
  parent = 6)
```

B.3. File model2SMS.R

```
# Creation of the stochastic model from the deterministic one
model2SMS <- model1SMS
model2SMS@name <- "Stochastic1"

#New set for scenarios
addItem(model2SMS, "sets") <- list(
```

B.3. File model2SMS.R

```
symbol = "n",
sDes = "Scenario",
setType = "set ")

# update indices
model2SMS@vars[2, ][["ind "]] <- list(c(1,2,4,3))
model2SMS@pars[2, ][["ind "]] <- list(c(2,4,3))
model2SMS@pars[4, ][["ind "]] <- list(c(1,4,3))
model2SMS@pars[5, ][["ind "]] <- list(c(1,2,4,3))

#Parameter for probability
addItem(model2SMS, "pars") <- list(
  symbol = "P",
  sDes = "Scenario Probability",
  ind = as.array(list(c(4))))

#update cost equation
addItem(model2SMS, "eqs") <- list(
  id = 6,
  symbol = "CostSto",
  sDes = "Total Cost",
  nature = "Objective",
  relation = "eq")
addItem(model2SMS, "terms") <- list( #1.z
  id = 1,
  eq = 6,
  side = "l",
  nature = "vars",
  item = 4)
addItem(model2SMS, "terms") <- list( #2.p
  id = 2,
  eq = 6,
  side = "r",
  nature = "pars",
  item = 7,
  setSums = array(list(c(4))))
addItem(model2SMS, "terms") <- list( #3.1
  id = 3,
  eq = 6,
  side = "r",
  nature = "consts",
  item = 1,
  parent = 2,
  setSums = array(list(c(3))))
addItem(model2SMS, "terms") <- list( #4.CI
  id = 4,
  eq = 6,
  side = "r",
  nature = "pars",
  parent = 3,
  item = 4,
  setSums = as.array(list(c(1))))
addItem(model2SMS, "terms") <- list( #5.x
  id = 5,
```

B. Code for example in Chapter 8

```
    eq = 6,
    side = "r",
    nature = "vars",
    item = 1,
    setSums = NULL,
    sign="",
    parent = 4)
addItem(model2SMS, "terms") <- list( #6.CO
    id = 6,
    eq = 6,
    side = "r",
    nature = "pars",
    parent = 3,
    item = 5,
    sign = "+",
    setSums = as.array(list(c(1,2))))
addItem(model2SMS, "terms") <- list( #7.DT
    id = 7,
    eq = 6,
    side = "r",
    nature = "pars",
    item = 6,
    parent = 6)
addItem(model2SMS, "terms") <- list( #8.y
    id = 8,
    eq = 6,
    side = "r",
    nature = "vars",
    item = 2,
    parent = 7)

# Update constraints domain
model2SMS@eqs[4, "domain"][[1]] <- list(c(2,4,3))
model2SMS@eqs[5, "domain"][[1]] <- list(c(1,2,4,3))
```

B.4. File model2Instance2.R

```
library(plyr)

## create instance of the problem
model2Instance2 <- newInstance(model2SMS)

## Sets of the instance
newInstanceSet(model2Instance2, "i", c("RTE", "PV", "CHP"))
newInstanceSet(model2Instance2, "j", c("winter", "spring", "
    summer", "autumn"))
newInstanceSet(model2Instance2, "n", 1:100)
newInstanceSet(model2Instance2, "t", c(2013:2017))

## Short variables for convinience
setI <- instanceSets(model2Instance2, "i")
setJ <- instanceSets(model2Instance2, "j")
```

B.4. File `model2Instance2.R`

```
setN <- instanceSets(model2Instance2, "n")
setT <- instanceSets(model2Instance2, "t")

## Parameters
## Lifetime
newInstancePar(model2Instance2, "LT",
  data.frame(i = setI,
    value = 20))

## Duration of periods (8 h. per day)
newInstancePar(model2Instance2, "DT",
  data.frame(j = rep(setJ, each = 5),
    t = setT,
    value = 91*8))

#scenarios 0.4 0.1 0.2 0.1 0.2
#probScn <- c(0.4,0.1,0.2,0.1,0.2)
#probScn <- c(0.24,0.19,0.18,0.20,0.19)
#symSets <- model2Instance2@sets[["n"]][,2]
newInstancePar(model2Instance2, "P",
  data.frame(n = setN,
    value = 1/length(setN)))

## Distribution of the variations
## CI_CHP, CI_PV, CO_RTE, CO_RIG, D

deltaMean <- c(-0.10, -0.05, 0.10, 0.03, 0.10)
deltaSD <- c(0.05, 0.06, 0.04, 0.02, 0.05)

#valScn <- matrix(c(
#   -0.10, -0.05, 0.10, 0.03, 0.10, # example1 scenario
#   -0.15, -0.10, 0.10, 0.05, 0.05, # low inv costs,
#   normal oper costs, normal demand
#   -0.15, -0.08, 0.12, 0.03, 0.10, # normal inv costs,
#   high gap between oper costs, high demand
#   -0.10, -0.05, 0.06, 0.03, 0.05, # high inv costs,
#   low oper costs, low demand
#   -0.15, -0.08, 0.12, 0.05, 0.10), # low inv costs,
#   high oper costs, high demand
#   byrow = TRUE, nrow = 5 )
#valScn <- cbind(valScn, probScn)
#rownames(valScn)<- paste("s", 1:5, sep="")
#colnames(valScn) <- c("invCHP", "invPV", "operRTE", "operNG",
#   "Demand", "Probability")

baseD <- ((213.50*(10^3))/(365*24))
corrSeason <- c(0.7, 0.9, 1.4, 1)

set.seed(1111)
newInstancePar(model2Instance2, "D", rbind.fill(lapply(seq(
  along = setN), function (y) {
    data.frame(j = setJ,
      n = setN[y],
```

B. Code for example in Chapter 8

```

        t = rep(setT, each = 4),
        value = rep(sapply(0:4, function(x){
            (1 + rnorm(1, deltaMean[5],
                deltaSD[5]))^x
        } ),
            each = 4) * rep(baseD * corrSeason, 5))
    })))

availPV <- (c(0.39 + 0.57 + 0.78,
    0.88 + 0.94 + 1,
    1.7 + 1.03 + 0.96,
    0.66 + 0.41 + 0.41) / 24) / 0.245
newInstancePar(model2Instance2, "G", data.frame(i = rep(setI,
    each = 20),
    j = setJ,
    t = setT,
    value = c(rep(1, 20), rep(availPV, 5), rep(1, 20))))

subsidies <- c(0, 0.4, 0.4)
baseCI <- c(50, 325.03 / 0.245, 20205.72/5.5) * (1-subsidies)
#parCI <- rbind.fill(lapply(seq(along = setN), function (y) {
#     data.frame(i = rep(setI, each = 5),
#         n = setN[y],
#         t = rep(2013:2017, 3),
#         value = c(t(baseCI * sapply(0:4, function(x){
#             (1 + valScn[y, c(3, 2, 1)])^x
#         })))
#     })))
#     })))

newInstancePar(model2Instance2, "CI",
    rbind.fill(lapply(seq(along = setN), function (y) {
        data.frame(i = rep(setI, each = 5),
            n = setN[y],
            t = rep(2013:2017, 3),
            value = c(t(baseCI * sapply(0:4, function(x)
                {
                    (1 + rnorm(3, deltaMean[c(3,
                        2, 1)], deltaSD[c(3,2,1)])
                )^x
                })))
        })))
    })))

baseCO <- c(0.134571, 0.05056)
newInstancePar(model2Instance2, "CO",
    rbind.fill(lapply(seq(along = setN), function (y) {
        data.frame(i = rep(setI[-2], each = 20),
            j = rep(setJ, 5),
            n = setN[y],
            t = rep(2013:2017, each = 4),
            value = rep(c(t(baseCO * sapply(1:5,

```


B.5. File comprehensiveExample.Rnw

```
function(x){
  (1 + rnorm(2, deltaMean[c
    (3,4)], deltaSD[c(3,4)
    ]))^x
  })),
  each = 4))
  )))

defInstanceEqs(modelInstance2, constEqs = c(3, 4, 5), objEqs
  = 6)
```

B.5. File comprehensiveExample.Rnw

```
\documentclass[a4paper]{article}
\usepackage{Sweave}

\title{Comprehensive Example for
``An Integrated Framework for the
Representation and Solution of Stochastic
Energy Optimization Problems''}
\author{Emilio L. Cano}

<<intro, echo=FALSE, results=hide>>=
## System requirements
## - The R software and the packages loaded below
## - A licenced GAMS installation. If GAMS directory
## is other than "~/app/gams23.9", change the line
## 'igdx("~/app/gams23.9")'
## - A LaTeX distribution for the current system

## Load needed packages
library(knitr)
library(optimr)
library(gdxrrw)
library(xtable)
library(ggplot2)
library(grid)

## Tell gdxrrw where GAMS is installed
igdx("~/app/gams23.9")

## Run scripts with data
## Deterministic model
source("../data/model1SMS.R")
## Stochastic extension
source("../data/model2SMS.R")
## Instance with 100 scenarios
source("../data/model2Instance2.R")
@

\begin{document}
\SweaveOpts{concordance=TRUE}
```

B. Code for example in Chapter 8

```
\maketitle

\section{Introduction}
This document is an example on how to use \textsf{R} as an
integrated
environment for optimization. It is assumed that the \textsf{R}
\textsf{optimr} package is installed.

Here we can include any statistical analysis, for example a
time series analysis
to forecast the future energy prices, saving the values as
parameters. We can
also show graphical representations of the parameter values,
as in Figure~\ref{fig:examplepar}, or tables with data, e.
g. Table~\ref{tab:exampletable}.

\begin{figure}[htp]
\begin{center}
<<examplepar, fig=TRUE, echo=FALSE, width=10>>=
## Demand
dtoPlot <- instancePars(model2Instance2, "D")
dtoPlot$id <- 1:20
pD <- ggplot(data = dtoPlot, aes(x = id, y = value, group=n,
col=n))
pD <- pD + geom_path()
pD <- pD + scale_x_discrete(name = "", breaks = seq(1,21, by
=5), labels = 2013:2017)
pD <- pD + ggtitle("Energy demand")
pD <- pD + scale_y_continuous(name="Demand level (kW)")
pD <- pD + theme(legend.position = "none")
pD <- pD + stat_summary(fun.y=mean, colour="darkred", geom="
line", aes(group=1), size=1)

## Investment cost
dtoPlot <- instancePars(model2Instance2, "CI")
pCI <- ggplot(data = dtoPlot, aes(x = t, y = value, group=n,
col=n))
pCI <- pCI + geom_path()
pCI <- pCI + facet_grid(i ~ .)
pCI <- pCI + scale_x_continuous(name = "")
pCI <- pCI + ggtitle("Investment cost")
pCI <- pCI + scale_y_continuous(name="EUR/kW")
pCI <- pCI + theme(legend.position = "none")
pCI <- pCI + stat_summary(fun.y=mean, colour="darkred", geom="
line", aes(group=1), size=1)

## Operation cost
dtoPlot <- instancePars(model2Instance2, "CO")
dtoPlot$id <- 1:20
pCO <- ggplot(data = dtoPlot, aes(x = id, y = value, group=n,
col=n))
pCO <- pCO + geom_path()
pCO <- pCO + facet_grid(i ~ .)
```

B.5. File comprehensiveExample.Rnw

```
pCO <- pCO + scale_x_discrete(name = "", breaks = seq(1,21, by
=5), labels = 2013:2017)
pCO <- pCO + ggtitle("Energy price")
pCO <- pCO + scale_y_continuous(name = "EUR/kWh")
pCO <- pCO + guides(color = guide_colourbar(title = "Scenario
"))
pCO <- pCO + stat_summary(fun.y=mean, colour="darkred", geom="
line", aes(group=1), size=1)

grid.newpage()
vpAll <- viewport(layout = grid.layout(2, 3,
widths = c(1/3, 1/3, 1/3),
heights = c(0.1, 0.9)))
vpT <- viewport(layout.pos.col = 1:3, layout.pos.row = 1, name
= "title")
vpD <- viewport(layout.pos.col = 1, layout.pos.row = 2, name =
"D")
vpCI <- viewport(layout.pos.col = 2, layout.pos.row = 2, name
= "CI")
vpCO <- viewport(layout.pos.col = 3, layout.pos.row = 2, name
= "CO")
splot <- vpTree(vpAll, vpList(vpT, vpD, vpCI, vpCO))
pushViewport(splot)
seekViewport("title")
grid.text("100 scenarios simulation", gp = gpar(cex=2))
seekViewport("D")
print(pD, newpage = FALSE)
seekViewport("CI")
print(pCI, newpage = FALSE)
seekViewport("CO")
print(pCO, newpage = FALSE)
@

\caption{Parameter values for the stochastic parameters.}
\label{fig:examplepar}
\end{center}
\end{figure}

<<echo=FALSE, results=tex>>=
xtable(head(instancePars(model2Instance2, "D")),
label = "tab:exampletable",
caption = "Example  $D$  parameter values (first 6 values
).")
@
```

The equations or any item of the model can be printed automatically from the `\textsf{model2SMS}` object. For example, the following command fetches the objective function:

```
<<results=tex>>=
cat("$$", getEq(model2SMS, 6, "tex", only = "rExpr"), "$$")
```

B. Code for example in Chapter 8

@

```
\section{Solving the problem}
```

Once we have the instance in an `\texttt{optimInstance}` object, it can be solved and the solution imported (see source code). Results checking is also possible as this information is also stored:

```
<<sol, echo=FALSE, results=hide>>=
wProblem(model2Instance2,
  filename = "./data/model2Instance2.gms",
  format = "gams",
  solver = "LP")
res <- gams("./data/model2Instance2.gms --outfile=./data/
  model2Instance2.gdx")
data(gamsOut)
if (res == 0){
  importGams(model2Instance2) <- "./data/model2Instance2.gdx"
  message("Optimization ok\n",
    "\tModel Status: ",
    as.character(subset(gamsModelStatusCode,
      id = model2Instance2@result$model, desc, drop =
        TRUE)),
    "\n\tSolver Status: ",
    as.character(subset(gamsSolverStatusCode,
      id = model2Instance2@result$solve, desc, drop =
        TRUE)))
} else{
  warning("Check the listing file, something was wrong: ",
    subset(gamsOutCode, id = res, desc, drop = TRUE))
}
}
```

@

We can embed calculations within the text, for example the value of the

```
objective function (\Sexpr{round(model2Instance2@result$obj)})
, or we can print pretty
\LaTeX tables with the optimal values, as the ones in Tables \ref{tab:x} and \ref{tab:y}, or any
other analysis and representation (see Figure \ref{fig:res}).
See the \texttt{.Rnw} source file to see the code.
```

```
<<results=tex, echo=FALSE>>=
print(xtable(instanceVars(model2Instance2, "x"),
  "Optimal values for  $x$ ",
  "tab:x"), include.rownames = FALSE)
print(xtable(head(instanceVars(model2Instance2, "y")),
  "Optimal values for  $y$  (first 6 values)",
  "tab:y"), include.rownames = FALSE)
```

@

```
\begin{figure}[htp]
```

B.5. File comprehensiveExample.Rnw

```
\begin{center}
<<bar, echo=FALSE, fig=TRUE>>=
df2plot <- subset(instanceVars(model2Instance2, "y"), j == "
  autumn")
df2plot <- aggregate(value ~ i + t, data = df2plot, FUN = mean
)
df2plot$t <- as.integer(as.character(df2plot$t))
p <- ggplot(df2plot, aes(x=t))
p <- p + geom_area(aes(y=value, fill=i))
p <- p + labs(title = "Optimal production plans (Autumn)", x =
  "Year", y = "kW")
p <- p + scale_fill_discrete("Technology")
print(p)
@
\end{center}
\caption{Output data representation.\label{fig:res}}
\end{figure}

\section{Conclusion}
This document can be compiled at any time, by any researcher.
Note that if any
value is changed, for example in the script that contain the
parameters
(\texttt{"../data/model2Instance2.R"}), the whole report is
updated automatically
(including tables, equations and charts).
If we use simulation during the research, we can simply fix
the seed to allow
the verification of the results by third parties. Different
reports for different stakeholders can be produced using a
common structure and tailoring the outputs.

\end{document}
```


C. Session Info

All materials of this thesis have been produced under the *Reproducible Research* approach (see Leisch, 2002). Thus, the source files contain both the text and the data analysis, as well as all the related code. The interface used for the development has been Eclipse 4.2¹ and StatET 3.3². The information of the R³ session where all the code has been run is the following:

```
print(sessionInfo(), locale = FALSE)

## R version 3.0.2 (2013-09-25)
## Platform: x86_64-pc-linux-gnu (64-bit)
##
## attached base packages:
## [1] splines    grid        tools
## [4] stats      graphics   grDevices
## [7] utils      datasets   methods
## [10] base
##
## other attached packages:
## [1] Hmisc_3.14-0    Formula_1.1-1
## [3] survival_2.37-7 lattice_0.20-24
## [5] lpSolve_5.6.7   gdxrrw_0.3.1
## [7] reshape2_1.2.2  plyr_1.8
## [9] psych_1.4.1     lubridate_1.3.3
## [11] knitr_1.5       ggplot2_0.9.3.1
```

¹<http://www.eclipse.org>

²<http://www.walware.de/goto/statet>

³<http://www.R-Project.org>

C. Session Info

```
## [13] xtable_1.7-1      optimr_0.0-10
## [15] rj_1.1.3-1
##
## loaded via a namespace (and not attached):
## [1] cluster_1.14.4
## [2] colorspace_1.2-4
## [3] dichromat_2.0-0
## [4] digest_0.6.4
## [5] evaluate_0.5.1
## [6] formatR_0.10
## [7] gtable_0.1.2
## [8] highr_0.3
## [9] labeling_0.2
## [10] latticeExtra_0.6-26
## [11] MASS_7.3-29
## [12] memoise_0.1
## [13] munsell_0.4.2
## [14] proto_0.3-10
## [15] RColorBrewer_1.0-5
## [16] rj.gd_1.1.3-1
## [17] scales_0.2.3
## [18] stringr_0.6.2
## [19] tcltk_3.0.2

Sys.time()

## [1] "2014-01-31 18:39:09 CET"
##
## Machine:
##   Intel(R) Core(TM) i5-2430M CPU @ 2.40GHz
##   MemTotal:      8090440 kB
```

The GAMS ⁴ version used was:

⁴<http://www.gams.com>


```
## GAMS program:  
## LEX-LEG 23.9.5 x86_64/Linux
```

Febrero 2014

Tesis doctoral de Emilio López
Cano. Dirigida por Javier
Martínez Moguerza y Yurii
Yermoliev



Universidad
Rey Juan Carlos

Dpto. de Estadística e
Investigación Operativa