Horizon 2020 Societal challenge 5
Climate action, environment, resource
Efficiency and raw materials

D3.3: DOWNSCALING REPORT

LEAD AUTHOR: TOBIAS CONRADT

OTHER AUTHORS: ROBERTO ROSON, GEORGIOS AVGERINOPOULOS, PILAR MARTINEZ, MARIA BLANCO, EVA ALEXANDRI

DATE: 4 September 2018

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement NO 689150 SIM4NEXUS

PROJECT	SUSTAINABLE INTEGRATED MANAGEMENT FOR THE NEXUS OF WATER-LAND-FOOD-ENERGY-CLIMATE FOR A RESOURCE-EFFICIENT EUROPE (SIM4NEXUS)
PROJECT NUMBER	689150
TYPE OF FUNDING	RIA
DELIVERABLE	D.3.3 DOWNSCALING REPORT
WP NAME/WP NUMBER	THEMATIC MODELS AND INTEGRATION / WP 3
TASK	TASK 3.2
VERSION	3.0
DISSEMINATION LEVEL	PUBLIC
DATE	04/09/2018 (THIS UPDATE) – 31/05/2018 (DUE DATE FOR FINAL VERS.)
LEAD BENEFICIARY	UNEXE
RESPONSIBLE AUTHOR	TOBIAS CONRADT (PIK)
ESTIMATED WORK EFFORT	NOT YET QUANTIFIED
AUTHOR(S)	TOBIAS CONRADT, ROBERTO ROSON, GEORGIOS AVGERINOPOULOS, PILAR MARTINEZ, MARIA BLANCO, EVA ALEXANDRI
ESTIMATED WORK EFFORT FOR EACH CONTRIBUTOR	NOT YET QUANTIFIED
INTERNAL REVIEWER	LYDIA VAMVAKERIDOU-LYROUDIA
DOCUMENT HISTORY	

VERSION	INITIALS/NAME	DATE	COMMENTS-DESCRIPTION OF ACTIONS
1.0	TC	30-NOV-2017	DRAFT COMPLETED
2.0	TC	24-MAY-2018	SUBMITTED FOR INTERNAL REVIEW
2.2	TC	04-JUN-2018	VERSION CONSIDERED AS FINAL, REPLACING PRELIMINARY UPLOAD
3.0	TC	04-SEP-2018	FINAL UPDATE, DATES ADDED IN SECTION 2.3

Table of Contents

Executive Summary

This report addresses modelling data made available by project partners to the entire SIM4NEXUS project. It should act as an overview and directions for use for the data produced within Task 3.2 “Downscaling of climate/climate change and socio-economic scenarios.” For adopting the RCP-SSP scenario framework described in van Vuuren et al. (2014) with its components detailed in Moss et al. (2010), van Vuuren et al. (2011), O'Neill et al. (2017), and Riahi et al. (2017) there is need for downscaling global climate and socio-economic scenarios to the spatial level of the case studies and subsequent modelling with these data within the case studies.

The case study modellers and data users will therefore need to know what data have been produced or downscaled and are available from within the project. They will also need to know the data formats and eventually their reliability. The can often be concluded from the data source (model) and the assumptions made for data generation.

This report aims at giving the necessary overview and providing the metadata reqired. It covers the climate scenario data produced in the ISIMIP framework and their distribution to the case studies as well as the baseline generation for the socioeconomic data. It details the outputs of socioeconomic data from the models CAPRI, E3ME, and OSeMOSYS. Finally, the problem of differing spatial disaggregations between different data sources is addressed and a possible solution communicated.

Changes with respect to the DoA

None.

Dissemination and uptake

The explanations given here are primarily directed to the data users within the project. All project participants should be aware of the data available within SIM4NEXUS and know about the background, generation, and validity of the data products distributed among them.

Short Summary of results (<250 words)

Climate data of the past and climate scenarios are a crucial input to the case studies of the project. Suitable products could be identified and obtained from the ISIMIP project (PIK), these are bias-adjusted gridded data with a 0.5-degree spatial solution originating from five different GCMs. All regional, national, and transboundary case studies have been provided with cut-outs from these data.

For economic data, it was agreed to use the Shared Socio-economic Pathway 2 ("Middle of the Road") in all case studies. As this scenario only provides information on GDP and population, a Computable General Equilibrium model has been used to "downscale" economic variables at a finer level in terms of industries and spatial detail.

A general problem of spatio-temporal data is the diversity of spatial discretisations and, to a lesser extent, different time steps. This is highlighted in an extra chapter.

Evidence of accomplishment

Report uploaded on project repository and data sets distributed through Dropbox (data hosting coordinated by EPSILON)

Glossary / Acronyms

CGE	Computable General Equilibrium (model) – See Section 3.1.
CORDEX	Coordinated Regional Climate Downscaling Experiment – An international research framework for regional climate modelling in defined modelling domains around the globe. See http://www.cordex.org
ETRS89-LAEA	Map projection, originally developed for small-scale equal-area mapping of Europe. Grids based on its coordinate system have been applied for land use mapping (CORINE) and visualisations of spatial statistics.
GCM	Global circulation model or global climate model
GDP	Gross domestic product – the monetary value of all goods and services produced by an economy per period, usually per year.
IEA	International Energy Agency – https://www.iea.org
ILO	International Labour Organization – http://www.ilo.org
ISIMIP	Inter-sectoral impact model intercomparison project – Within this project, a bias-adjusting downscaling generated uniform 0.5-degree fields of meteorological variables. These were made available for the SIM4NEXUS case studies by PIK.
NUTS	*Nomenclature des unités territoriales statistiques* – a hierarchy of spatial sub-divisions of the EU member states and Turkey, developed and used by the EU statistics office EUROSTAT.
OECD	The Organisation for Economic Co-operation and Development – Founded in 1960 by 18 European countries, the United States, and Canada, it grew to 35 member countries around the globe. Its main purpose is to collect and report economic data, define standards, and provide recommendations.
RCM	Regional climate model – Usually covering a single continent or a focus region. Needs GCM or larger RCM (nesting) as driving boundary condition. The advantage is the higher spatial and temporal resolution compared to GCMs.
RCP	Representative Concentration Pathways – Scenarios of greenhouse gas concentrations introduced for the Fifth Assessment Report of the IPCC.
SSP	Shared Socio-Economic Pathways – Socio-economic counterpart to the RCP scenarios. Five SSP were originally defined by narratives only which challenges any downscaling to economic indicators for certain regions and selected time periods.
WATCH	Hydroclimatic data sets based on observations and reanalyses, originally introduced by the Sixth Framework Programme project WATCH (Water and Climate Change). Some formats are still maintained and updated.

1 Introduction

1.1 WP and Task

This report (D3.3) refers to WP3 “Thematic Models and Integration” and, therein, to Task 3.2 “Downscaling of climate/climate change and socio-economic scenarios.”

1.2 Structure of the document

This report is structured in chapters as follows:

Chapter 2 is about the climate data and presents many details about the ISIMIP data chosen for distribution in Section 2.2.

Chapter 3 is about economic data, the process of baseline generation through computable general equilibrium modelling, and the data provided by the economy-related thematic models.

Chapter 4 showcases and discusses different spatial discretisations used in the SIM4NEXUS which are however following established standards each.

Chapter 5 contains conclusions and recommendations.

Finally, the references cited in this document are listed in Chapter 6.

2 Climate Data

2.1 Available options and SIM4NEXUS selection

The first and foremost data need for the modelling work in the case studies, basis of complexity modelling and the serious game, were spatial fields of climate variables – both for the recent decades and as climate scenarios for the future.

In contrast to the situation one decade ago there is now a large portfolio of climate scenarios available, generated from numerous models and following many different scenario assumptions. However, most products from global circulation models (GCM) are too coarse for our European case studies, and there is no unified grid. Another serious problem of these dynamic models are biases, especially with precipitation.

These problems were addressed at PIK in the framework of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP, Warszawski et al. 2014) by compiling bias-corrected GCM scenarios interpolated to a common grid (Hempel et al. 2013). We will refer to these as ISIMIP data, details are given in the following section. Obviously being the best choice among GCM products ISIMIP data have been selected for and distributed to the SIM4NEXUS case studies.

Another branch of suitable climate data are dynamically downscaled products from regional circulation models (RCM). The international modelling community defined a set of common grids in the framework of the CORDEX initiative (CORDEX 2015, http://www.cordex.org). For analyses requiring a higher spatial resolution the EURO-CORDEX products are a data source to consider, this has however not happened until now (May 2018). With CORDEX, data systematic biases would also be an issue again, and the strange geographical projection of CORDEX data on a rotated pole grid exhibits additional problems for practical utilisation.

2.2 The ISIMIP data

All details of the bias correction and compilation of the ISIMIP climate data are given by Hempel et al. (2013). Important points of their approach are:

- Basic bias corrections like adding a constant offset or applying a linear correction factor may erase the average deviation between simulated and observed time series, but often fail to correct the variability of the data. Non-parametric (empirical) quantile mapping offers more flexibility, but assumptions have to be made for values falling outside the calibration range, usually seen in scenario climate data for the future. Therefore, a parametric approach has been chosen.
- The applied bias correction method preserves the long-term trends of the variables. In a first step, only the monthly variability and mean are corrected using a constant offset or multiplicative correction factor that corrects for long-term differences between the simulated and observed monthly mean data in the historical period. Then the daily variability of the simulated data is modified about their monthly means to match the observed daily variability. This correction is described by calendar-month and grid-cell specific transfer functions. Depending on the variable, these steps may consist of additive correction and linear regression (temperature), or multiplicative correction and non-linear regression with additional touch-ups (precipitation).

- Although no dynamical downscaling was involved, an implicit downscaling to the common, finer resolution of a 0.5-degree grid was achieved using the WATCH Forcing Data (WFD, Weedon et al. 2011) as observation-based reference data.

As WATCH data depend on ground observations, only raster cells over land can be considered for the downscaling. Fig. 1 shows the spatial extent over Europe.

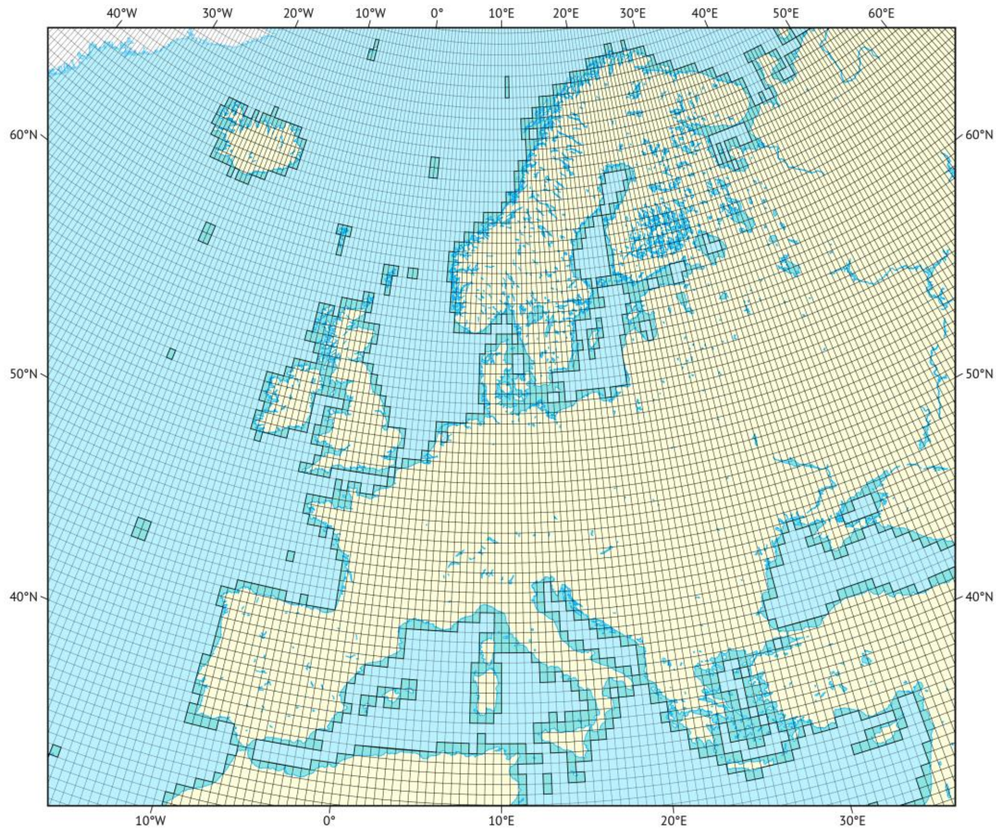

Figure 1. Half-degree grid of ISIMIP climate data over Europe

The bias corrected data are based on simulations of five GCMs from the CMIP5 archive (see Table 1). They consist of historical model runs for the years 1950–2005 and scenarios each for the years 2006–2099, only for the HadGEM2-ES model the historical period ends already in 2004 while the scenario period accordingly starts in 2005.

From each model, four scenario realisations based on the Representative Concentration Pathway (RCP) emission scenarios (Moss et al. 2010, van Vuuren et al. 2011, Meinshausen et al. 2011) had been released: RCP2.6, RCP4.5, RCP6.0, and RCP8.5. Thus there are 20 different ISIMIP scenario realisations. The climate trajectories of the RCPs do not substantially diverge before the middle of the 21st century, therefore it is principally advisable to use the full set of available realisations for assessments up to the year 2050. This ensures better coverage of model uncertainty and allows for quantile and other statistical analyses.

Table 1. The five GCMs from which ISI-MIP data were produced

Acronym	Origin
HadGEM2-ES	Met Office Hadley Centre (UK) and Instituto Nacional de Pesquisas Espaciais (BR)
IPSL-CM5A-LR	Institut Pierre-Simon Laplace (FR)
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (U Tokyo), and National Institute for Environmental Studies (JP)
GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory (US)
NorESM1-M	Norwegian Climate Centre (NO)

However, on the project meeting in Athens (March 2018) **RCP6.0 was defined as the baseline scenario in SIM4NEXUS.** Thus there are five independent realisations from the different models to consider per climate scenario.

The daily available climate variables are listed in Table 2.

Table 2. Climate variables in the ISIMIP data

Code	Content	Code	Content
tas	Average 2m air temperature	rsds	Shortwave radiation
tasmax	Maximum temperature	rlds	Longwave radiation
tasmin	Minimum temperature	ps	Surface pressure
pr	Total precipitation	wind	Near surface wind speed
prsn	Snowfall	uas	Eastward wind component
rhs	Relative humidity	vas	Northward wind component

The ISIMIP data are produced for geographical 0.5-degree tiles with land surfaces, according to the WATCH calibration data set. The standard data format is NetCDF4 (Unidata 2016). According to the needs of the case studies, monthly averages were compiled and released in ASCII representation.

2.3 Regional climate data cut-outs provided

Regional ISIMIP cut-outs have been prepared and distributed to all regional, national, and transboundary case studies as shown in Fig. 2. Sweden was the last case study to access their data from 18 June 2018 onwards (date of repository upload by EPSILON).

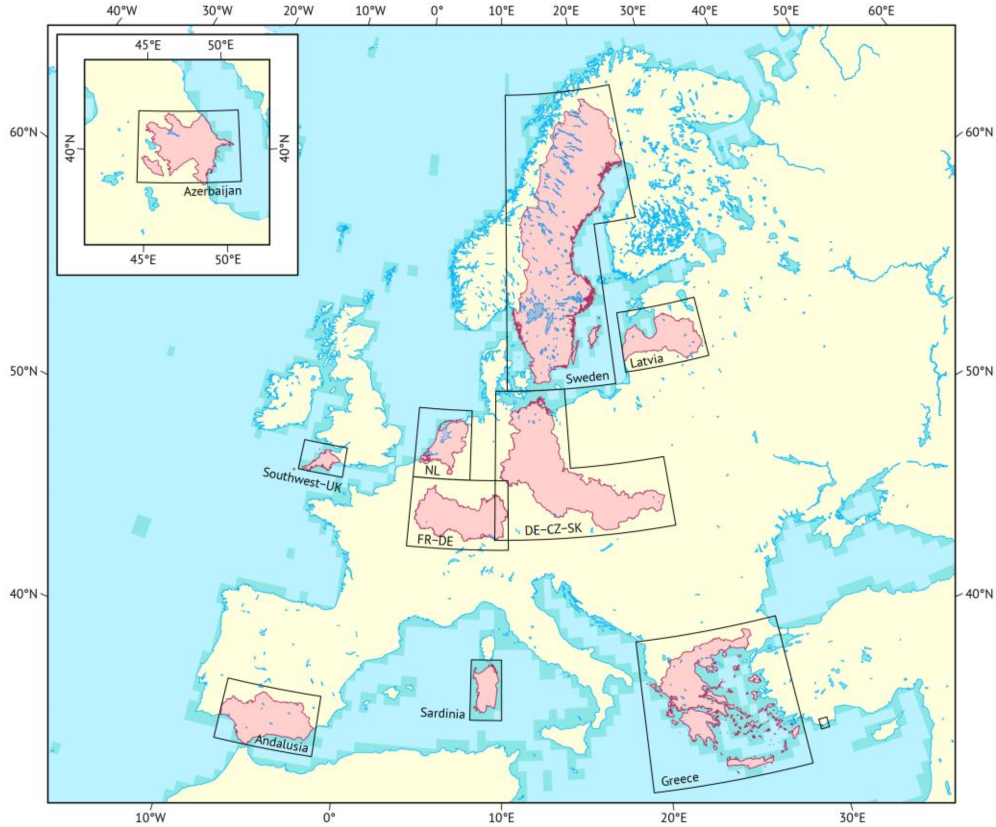

Figure 2. Cut-outs of ISIMIP climate data provided to case studies

In May 2018 it had still been discussed that the European and global case studies would probably need some kind of spatial aggregation of the half-degree grid data because the size of the full-scale global ISIMIP data set is in the Terabyte range and could not have been used directly for modelling. In August it turned out that climate is implicitly considered in these case studies through other climate-dependent input data sets. Therefore the **final decision to not provide any climate data sets to the European and global case studies** was taken in a WP3+4 Skype conference led by Lydia Vamvakeridou-Lyroudia on 3 September 2018.

3 Socio-economic data

3.1 Baseline generation

The definition of the socio-economic reference framework is based on the standard and universally accepted "Shared Socio-economic Pathways", or SSPs. Each SSP identifies a qualitative narrative about the future evolution of the global economy, and a very limited data set consistent with that story-line, which (for our purposes) only include among relevant variables projections of population and real GDP by country.

As GDP and population are too little to serve as a starting point for a case study in SIM4NEXUS, our strategy has been to use a Computable General Equilibrium (CGE) model as a "multiplier of scenario variables". A CGE model is a large non-linear mathematical system, defining equilibrium conditions for interconnected markets, based on microeconomic behavioral assumptions, macroeconomic identities and budget balances. This "multiplication" process has been sometimes termed "downscaling" for its analogy with the one occurring in climate models.

The typical CGE simulation exercise is based on comparative statics: an initial system equilibrium state is assumed, then some exogenous perturbation is considered (change in policy instruments, physical conditions, technology, tastes, etc.) and a counter-factual equilibrium is then obtained and assessed. The idea here is to force a CGE model to be fully consistent with a given national GDP level and population. In practice, this obtained by "swapping" an exogenous productivity parameter (which can be considered as a measure of technological progress) with the naturally endogenous variable of real GDP. In other words, this is like asking the model to compute what degree of progress is necessary to achieve a certain GDP target.

The counterfactual equilibrium produces a wealth of data: industrial output volumes, employment by sector, income sources, trade flows, consumption patterns, trade balances, investment levels, etc. Of course, the results are based on a number of assumptions driving the model (its "closure rules") so they could hardly be considered as "forecasts", also because other shocks and processes may shape the future structure of the economy. Nonetheless, the CGE model can provide a quite detailed picture of the economy, totally consistent with the selected SSP hypotheses.

The following variables were provided for the year 2030 and the SSP2 ("Middle of the Road") scenario, for the Sardinia "fast-track" case study:

- Industry output by sector (57 industries)
- Value added by industry
- Domestic good i demanded in industry j
- Demand for endowment i (primary resources like labor, land, capital) by sector j
- Imported good i demanded in industry j
- Export sales by sector i from home to foreign regions

- Imports by sector
- Household and public sector demand patterns

A recursive dynamic CGE model produces a sequence of temporary equilibria, linked by two inter-temporal relationships: one is the capital accumulation equation, by which investments in one period add to the physical capital stock in the following period; and the other one is the foreign debt accumulation equation, by which any imbalance between regional savings and investments varies the stock of foreign debt (possibly negative), causing an inflow or outflow of interest payments in the next periods.

In SIM4NEXUS we constructed a recursive dynamic model with 5-years steps, from 2010 to 2050. At every step, the GDP growth from the selected SSP scenario is imposed, by endogenously setting a productivity parameter. The model dynamics is, therefore, partly exogenous (forced by the SSP) and partly endogenous (through capital and foreign debt accumulation).

A special feature of the model used in SIM4NEXUS is the generation of macroeconomic data for some sub-national administrative regions (NUTS-2 geographical aggregation scale). For instance, Sardinia is a NUTS-2 region of Italy, Andalusia is a NUTS-2 region of Spain.

These regional data are provided, for each time step, in addition to the national data (which will remain available):

- industrial production volumes;
- demand for primary factors, by industry (labor, capital, land, natural resources);
- regional GDP

At the moment, regional variables are available for the following countries: Italy, Spain, U.K., France and Germany. All NUTS-2 regions for these countries are taken into account, although only a few of them may be of interest for the case studies.

3.2 Data sources

3.2.1 CAPRI

The first authority to ask for economic and agricultural production data are the CAPRI modellers because CAPRI already contains a respective data base for the European NUTS-2 units. In consistently using CAPRI data throughout the project unnecessary work or money for assessing external data collections can be saved and a common data basis is maintained.

CAPRI provides simulated results for the agricultural sector at subnational level in the EU, while considering global agricultural markets, by sequential iteration between supply and market modules (Britz and Witzke 2014). The supply module depicts detailed farming decisions at NUTS-2 in the EU, by representing the interplay between agricultural production, the environment and the political context. The market module simulates bilateral trade flows, as well as bilateral and multilateral border protection instruments for about 60 commodities (primary and secondary agricultural products) and 40 trade blocks.

The CAPRI modelling system is based upon a complete and consistent database, that integrates economic, physical and environmental information from official data sources (EUROSTAT, FAOSTAT, OECD). Data consistency in time and space is guarantee by specific modules of the model.

Simulation results at regional level cover cropped areas, herd sizes, income indicators and environmental indicators (balances for N,P,K, emissions of ammonia, methane and N_2O, greenhouse gas inventories and life-cycle assessment of energy use in agriculture). Prices, supply and demand positions are provided at country level. Producer and consumer prices, supply and demand positions as well as bilateral trade flows with attached prices, transport costs and tariffs globally are derived for each trade block.

CAPRI provides results to SIM4NEXUS case studies at the NUTS-2 level for different sets of agrifood products (cereals, oilseeds, other arable field crops, vegetables and permanent crops as well as animal products) for the years 2010, 2020, 2030, 2040 and 2050. The variables delivered so far are presented in Table 3.

Table 3: List of output variables in CAPRI

Category	Variable	Unit
Agriculture	Agricultural Production	million t DM/yr
Agriculture	Agricultural Production|Crops	1000 t/yr
Agriculture	Agricultural Production|Livestock	1000 t/yr
Agriculture	Yield|Rainfed crops	Kg/ha
Agriculture	Yield|Irrigated crops	Kg/ha
Agriculture	Price|Agriculture|Crops and Livestock	Euro/t
Agriculture	Agricultural Demand|Food	1000 t/yr
Agriculture	Agricultural Demand|Feed	1000 t/yr
Agriculture	Agricultural Demand|Biofuels	1000 t/yr
Agriculture	Agricultural Demand|Processing	1000 t/yr
Land cover	Land Cover|Cropland	1000 ha
Land cover	Land Cover|Irrigated cropland	1000 ha
Land cover	Land Cover|Pasture	1000 ha
Food	Food Demand	kcal/cap/day
Food	Food Demand|Crops	kcal/cap/day
Food	Food Demand|Livestock	kcal/cap/day
Food	Import|Agriculture	1000 t/yr
Food	Export|Agriculture	1000 t/yr
Water	Water Withdrawal|Irrigation	1000 m3/yr
Energy (primary)	Production|Primary Energy |Biofuels	1000 t/yr
Energy (trade)	Import|Primary Energy|Biofuels	1000 t/yr
Energy (trade)	Export|Primary Energy|Biofuels	1000 t/yr
Energy (price)	Price|Primary Energy|Biofuels	Euro/t

3.2.2 E3ME

E3ME is a computer-based model of the world's economic and energy systems and the environment. It was originally developed through the European Commission's research framework programmes and is now widely used in Europe and beyond for policy assessment, for forecasting and for research purposes.
This model description in this section provides only a short summary of the E3ME model. For further details, the reader is referred to the full model manual available online from www.e3me.com.

The structure of E3ME is based on the system of national accounts, with further linkages to energy demand and environmental emissions. The labour market is also covered in detail, including both voluntary and involuntary unemployment. In total, there are 33 sets of econometrically estimated equations, also including the components of GDP (consumption, investment, international trade), prices, energy demand and materials demand. Each equation set is disaggregated by country and by sector.

3.2.2.1 The main dimensions of the model

The main dimensions of E3ME (see also Table 4) are:

- 59 countries – all major world economies, the EU28 and candidate countries plus other countries' economies grouped
- 44 or 70 (Europe) industry sectors, based on standard international classifications. The 70-sector classification for Europe follows the NACE Rev 2 2-digit classification from Eurostat.
- 28 or 43 (Europe) categories of household expenditure
- 22 different users of 12 different fuel types
- 14 types of air-borne emission (where data are available) including the six greenhouse gases monitored under the Kyoto protocol
- Yearly time steps up to 2050

Table 4: Main dimensions of the E3ME model

	Regions	Industries (Europe)	Industries (non-Europe)
1	Belgium	Crops, animals, etc	Agriculture etc
2	Denmark	Forestry & logging	Coal
3	Germany	Fishing	Oil & Gas etc
4	Greece	Coal	Other Mining
5	Spain	Oil and Gas	Food, Drink & Tobacco
6	France	Other mining	Textiles, Clothing & Leather
7	Ireland	Food, drink & tobacco	Wood & Paper
8	Italy	Textiles & leather	Printing & Publishing
9	Luxembourg	Wood & wood prods	Manufactured Fuels
10	Netherlands	Paper & paper prods	Pharmaceuticals
11	Austria	Printing & reproduction	Other chemicals
12	Portugal	Coke & ref petroleum	Rubber & Plastics
13	Finland	Other chemicals	Non-Metallic Minerals
14	Sweden	Pharmaceuticals	Basic Metals
15	UK	Rubber & plastic products	Metal Goods
16	Czech Rep.	Non-metallic mineral prods	Mechanical Engineering
17	Estonia	Basic metals	Electronics
18	Cyprus	Fabricated metal prods	Electrical Engineering
19	Latvia	Computers etc	Motor Vehicles
20	Lithuania	Electrical equipment	Other Transport Equipment
21	Hungary	Other machinery/equipment	Other Manufacturing
22	Malta	Motor vehicles	Electricity
23	Poland	Other transport equip	Gas Supply
24	Slovenia	Furniture; other manufacture	Water Supply
25	Slovakia	Machinery repair/installation	Construction
26	Bulgaria	Electricity	Distribution
27	Romania	Gas, steam & air cond.	Retailing
28	Norway	Water, treatment & supply	Hotels & Catering
29	Switzerland	Sewerage & waste	Land Transport etc
30	Iceland	Construction	Water Transport
31	Croatia	Wholesale & retail MV	Air Transport
32	Turkey	Wholesale excl MV	Communications

33	Macedonia	Retail excl MV	Banking & Finance
34	USA	Land transport, pipelines	Insurance
35	Japan	Water transport	Computing Services
36	Canada	Air transport	Professional Services
37	Australia	Warehousing	Other Business Services
38	New Zealand	Postal & courier activities	Public Administration
39	Russian Fed.	Accommodation & food serv	Education
40	Rest of Annex I	Publishing activities	Health & Social Work
41	China	Motion pic, video, television	Miscellaneous Services
42	India	Telecommunications	Unallocated
43	Mexico	Computer programming etc.	Forestry
44	Brazil	Financial services	Hydrogen supply
45	Argentina	Insurance	
46	Colombia	Aux to financial services	
47	Rest Latin Am.	Real estate	
48	Korea	Imputed rents	
49	Taiwan	Legal, account, consult	
50	Indonesia	Architectural & engineering	
51	Rest of ASEAN	R&D	
52	Rest of OPEC	Advertising	
53	Rest of world	Other professional	
54	Ukraine	Rental & leasing	
55	Saudi Arabia	Employment activities	
56	Nigeria	Travel agency	
57	South Africa	Security & investigation, etc	
58	Rest of Africa	Public admin & defence	
59	Africa OPEC	Education	
60		Human health activities	
61		Residential care	
62		Creative, arts, recreational	
63		Sports activities	
64		Membership orgs	
65		Repair comp. & pers. goods	
66		Other personal serv.	
67		Hholds as employers	
68		Extraterritorial orgs	
69		Unallocated/Dwellings	
70		Hydrogen supply	

Source(s): Cambridge Econometrics.

3.2.2.2 Main data sources for E3ME

This subsection outlines the main data sources for E3ME's historical data on economy, energy and emissions.

Economic data

The E3ME economic data is made consistent across countries and in the same units. For monetary data the euro is used. The main sources for the economic data are:

- The Eurostat national accounts branch is the primary source for European countries and provides a consistent data source across countries. The OECD's STAN data set also provides some sectoral disaggregation.
- For non-EU countries UN and World Bank databases are consulted, as well as national statistics websites
- Data from the AMECO database are used to provide macroeconomic figures and to check totals.
- Comtrade and WITS database is used for bilateral trade data
- ILO statistics database is used to fill in any gaps in the labour market data, once above sources have been exhausted.

Energy and emissions data

The energy data in physical units have been provided by IEA energy balances. The IEA energy balances are given as time-series data. The 54 energy carriers provided by IEA have been aggregated into the twelve energy carriers in E3ME, and some inconsistencies (comparing items year by year) have been removed. The energy price data is also taken from IEA statistics.

The CO2 emissions data, disaggregated by energy user, are on an annual basis and are obtained from the European Environment Agency or the EDGAR database. These are allocated to fuels using standard coefficients and then scaled to be consistent with the total. Adjustments may be made for international shipping and aviation to maintain consistency between the energy and emissions data. Non-CO2 emissions data are also taken from the EDGAR database.

3.2.2.3 Downscaling E3ME data

The majority of case studies are at a national level, so the E3ME data can be used directly by the case studies. In the instances where the case study is at a regional level, the E3ME data is downscaled to the sub-national level. The data used for downscaling are generally based on the NUTS 2 or 3 classification. The process for downscaling E3ME data to sub-national level is as follows:

1. The NUTS regions that make up the case study region are identified in discussion with the case study lead.
2. Economic, energy and emission data sources (or actual data) are provided by the case study lead or collected by the E3ME team. The data are generally collected for the case study region and the national level from the same source.
3. Last available year ratios are calculated between regional and national data based on the data collected as outlined at in the point above.
4. E3ME classifications are mapped to as closely as possible to the classification in which the case study data are available. In most cases E3ME data needs to be aggregated to further to match the classification of the case study data.

5. The ratios calculated at point 3 are then applied to the aggregated E3ME data (as outlined in point 4).
6. Trends in data are then checked to see if they make sense and whether adjustments are required.

E3ME data are usually provided as CSV or in EXCEL format.

3.2.3 OSeMOSYS

The Open Source energy Modelling System (OSeMOSYS) is a bottom-up optimisation modelling framework used primarily for long-term energy systems analysis and planning. Over the past years, it has been used to create sub-national, national, trans-national/continental and global energy and integrated assessment (Climate, Land-use, Energy and Water) models (Brouwer et al. 2018). Within the large suite of existing modelling frameworks, its mission is to provide a fully open source, freely available, flexible and accessible tool for capacity building and uptake by planners in developed and developing countries.

The first code was made available to the public in 2008 and it has been further developed ever since. For the purpose of accessibility and engagement of different scientific communities, OSeMOSYS is written in three modelling languages: GNU MathProg (original version), GAMS and Python. Its code is structured modules which can be plugged in or out depending on the scope of the analysis. Each module is constituted by a number of equations representing physical relations and constraints between elements of a generic system (e.g. the processes and commodities in the energy supply chain). Such structure provides users with the flexibility to run diverse analyses and to add new equations to the code.

Different interfaces have been used/developed to run OSeMOSYS, facilitating the creation of models, results visualisation and import/export of inputs/outputs to Microsoft Excel. However, regardless of the interface, an OSeMOSYS model can be built and run through simpletext-files.. The results of an optimisation through OSeMOSYS provide the lowest-cost energy and resources mix (in terms of installed capacity and generation) for each year and time step of the time and space domain chosen by the user. The time frame of the studies can be any, but most applications have focused on 30–50 years.

The units in OSeMOSYS are exogenously defined. In other words, the user needs to consider a set of units for the different datasets. Consequently, ensuring the consistency in the units amongst the datasets is of critical importance. For instance, for the case of the energy sector if electricity generation is expressed in PJ, the emission factors for the various fuels used to produce electricity should be expressed in (k)tons of emitting substance per PJ of fuel used.

Regarding the spatial scale, OSeMOSYS is indeed unrestricted. It has already been used on a village scale. In SIM4NEXUS, it is currently (May 2018) used in the global and Azerbaijan national case study.

Table 5: OSeMOSYS output data

Model inputs	Model outputs
Capital and operation costs for all technologies and commodities	Power generation and fuel consumption by energy source and by technology
Techno-economic characteristics of the technologies (e.g. input and output commodities, efficiency, operation limits, availability factors and emission factors)	Installed capacity by energy source
	Emissions per region, year and technology
Demand projections for user-defined commodities	Energy system cost, aggregated per region and disaggregated per technology (both investment and operating)
Renewable energy targets, emissions targets, emission penalties, reserve margin constraints and reserve capacity demand	

4 Spatially incongruous data sets

Socio-economic data are usually collected on administrative levels while geoscientific models are often raster-based. We observe at least three kinds of spatial discretizations of data over Europe:

- NUTS units (Fig. 3),
- River basins and their subcatchments (Fig. 4), and
- Gridded data (Fig. 5 shows an example of the equal-area raster whose higher resolution versions are used for pan-european land use mapping (CORINE) and eurostat's population density maps (GEOSTAT)).

On the Athens meeting of the SIM4NEXUS project in March 2018 Eugene Westerhof (WUR-LEI) announced a computing data repository capable of transferring any area-related spatial data into each desired spatial discretisation which will be followed.

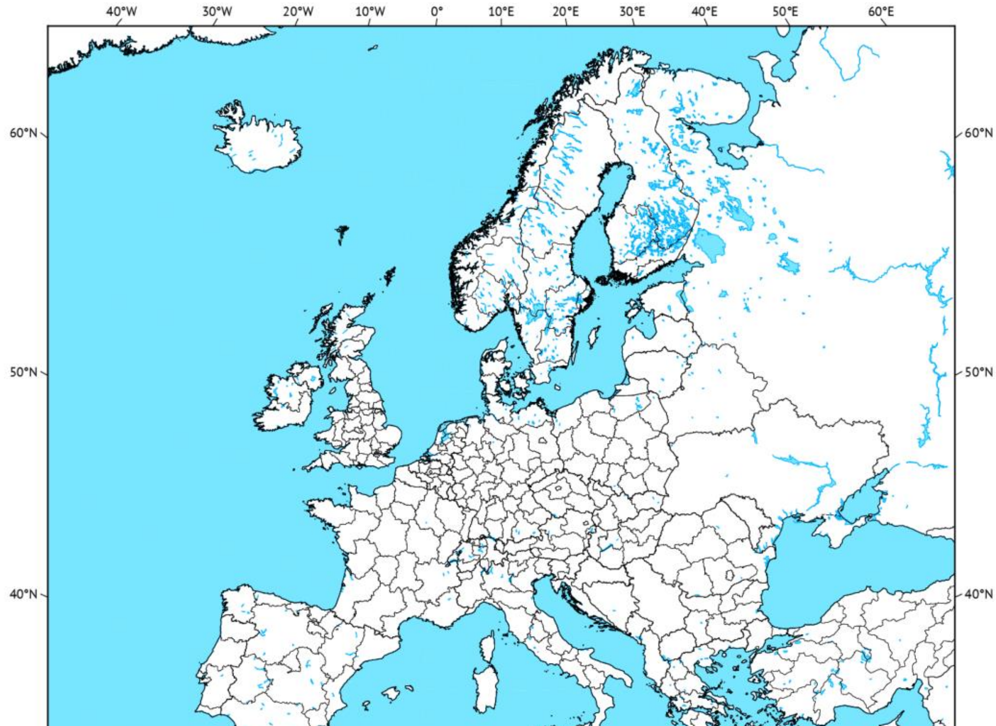

Figure 3. NUTS-2 units in EU countries and Turkey

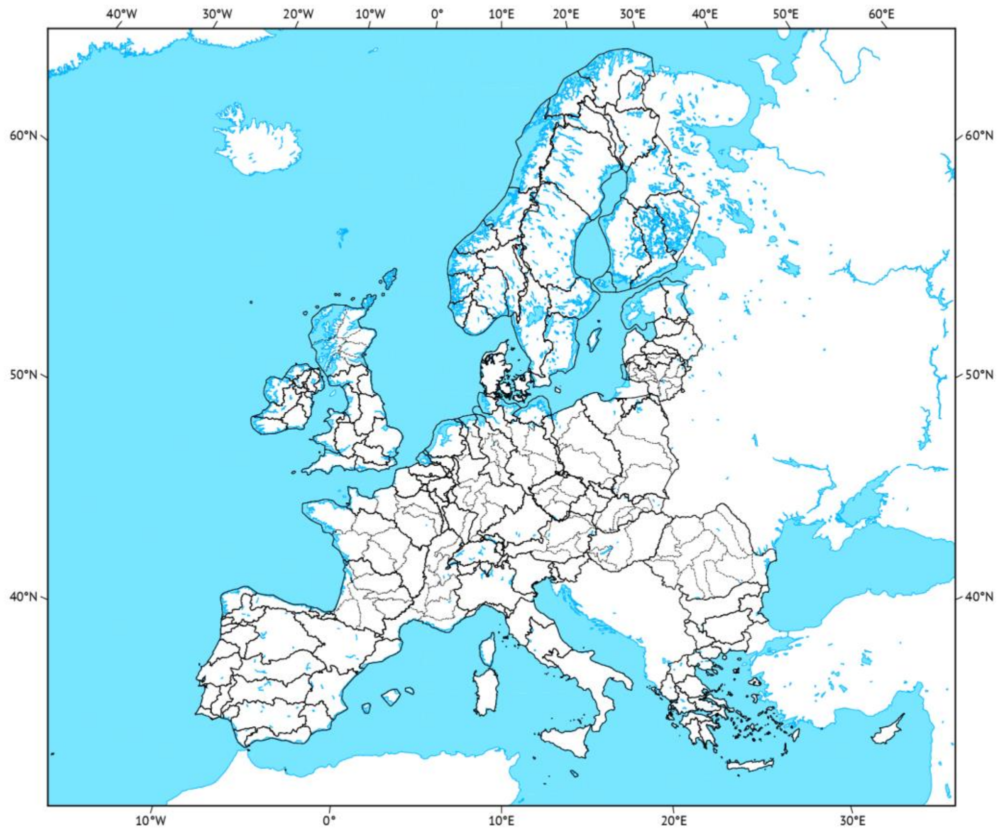

Figure 4. River Basin Districts (RBD) and their Sub-Units (RBDSU) as implemented by the Water Framework Directive

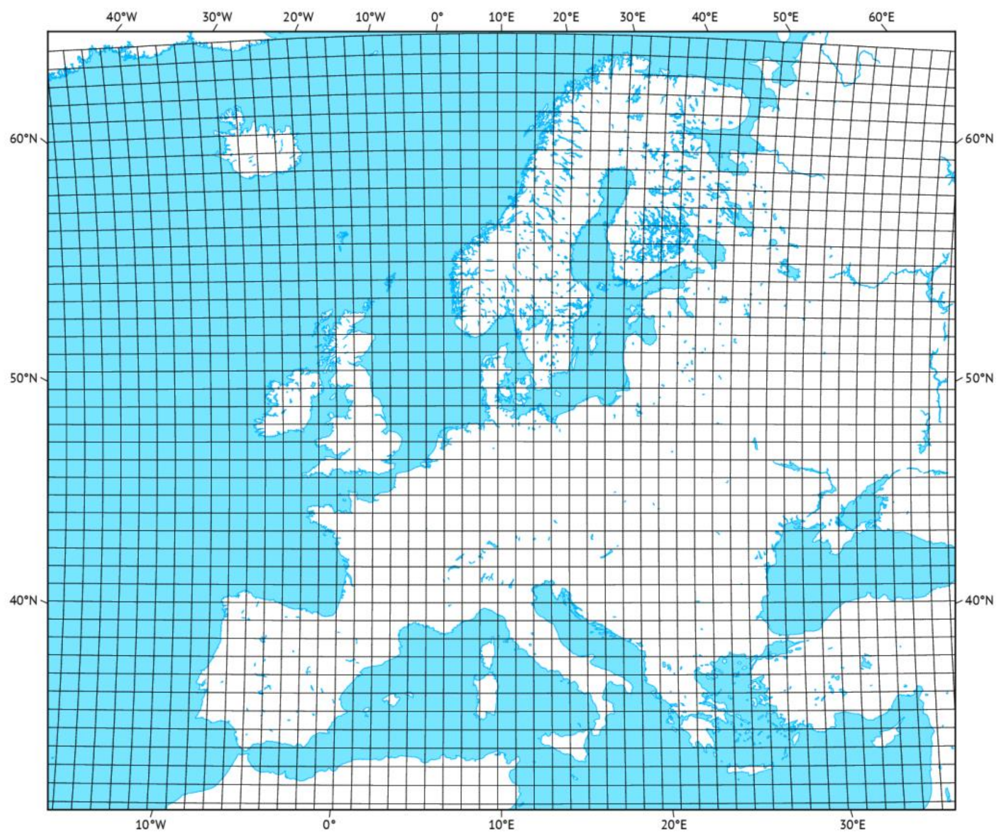

Figure 5. ETRS89-LAEA grid over Europe

5 Conclusions and recommendations

This report gives an impression of the data galore produced within SIM4NEXUS after two years, half of the full project runtime. Albeit it might seem challenging to keep the synopsis in mind when searching for a certain information, the majority of case studies are probably well aware of what they need, what there is available, and whom to ask.

Nevertheless, this compendium should be kept as reference. It can serve as metadata repository for every data user and ease

- deciding if and how to apply certain data,
- citing data sources correctly, and
- indicating the fidelity and credibility of certain data.

The aim of this report is reached when it is not just stored as “D3.3” in a project repository but when it is copied and regularly used by many project partners, thus avoiding unnecessary requests on metadata details, processing options or even general data availabilities.

6 References

- Brouwer, F., G. Avgerinopoulos, D. Fazekas, C. Laspidou, J.-F. Mercure, H. Pollitt, E. P. Ramos & M. Howells (2018): Energy modelling and the Nexus concept. Energy Strategy Reviews 19: 1–6. doi:10.1016/j.esr.2017.10.005
- CORDEX (2015): CORDEX domains for model integrations (updated 23/10/15). The Coordinated Regional Climate Downscaling Experiment of the World Climate Research Programme (WCRP CORDEX), Norrköping, Geneva, and Paris. URL: http://cordex.org.preview.binero.se/wp-content/uploads/2012/11/CORDEX-domain-description_231015.pdf. Last accessed in November 2017.
- Hempel, S., K. Frieler, L. Warszawski, J. Schewe & F. Piontek (2013): A trend-preserving bias correction – the ISI-MIP approach. Earth System Dynamics 4 (2): 219–236. doi:10.5194/esd-4-219-2013
- Meinshausen, M., S. J. Smith, K. Calvin, J. S. Daniel, M. L. T. Kainuma, J.-F. Lamarque, K. Matsumoto, S. A. Montzka, S. C. B. Raper, K. Riahi, A. Thomson, G. J. M. Velders & D. P. P. van Vuuren (2011): The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Climatic Change 109 (1–2): 213–241. doi:10.1007/s10584-011-0156-z
- Moss, R. H., J. A. Edmonds, K. A. Hibbard, M. R. Manning, S. K. Rose, D. P. van Vuuren, T. R. Carter, S. Emori, M. Kainuma, T. Kram, G. A. Meehl, J. F. B. Mitchell, N. Nakicenovic, K. Riahi, S. J. Smith, R. J. Stouffer, A. M. Thomson, J. P. Weyant & T. J. Wilbanks (2010): The next generation of scenarios for climate change research and assessment. Nature 463 (7282): 747–756. doi:10.1038/nature08823
- O'Neill, B. C., E. Kriegler, K. L. Ebi, E. Kemp-Benedict, K. Riahi, D. S. Rothman, B. J. van Ruijven, D. P. van Vuuren, J. Birkmann, K. Kok, M. Levy, W. Solecki (2017): The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. Global Environmental Change 42: 169–180. doi:10.1016/j.gloenvcha.2015.01.004
- Unidata (2016): Network Common Data Form (NetCDF). Online documentation, last accessed in November 2017. URL: http:www.unidata.ucar.edu/software/netcdf/. doi:10.5065/D6H70CW6
- Riahi, K., D. P. van Vuuren, E. Kriegler, J. Edmonds, B. C. O'Neill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink, O. Fricko, W. Lutz, A. Popp, J. C. Cuaresma, S. KC, M. Leimbach, L. Jiang, T. Kram, S. Rao, J. Emmerling, K. Ebi, T. Hasegawa, P. Havlik, F. Humpenöder, L. A. Da Silva, S. Smith, E. Stehfest, V. Bosetti, J. Eom, D. Gernaat, T. Masui, J. Rogelj, J. Strefler, L. Drouet, V. Krey, G. Luderer, M. Harmsen, K. Takahashi, L. Baumstark, J. C. Doelman, M. Kainuma, Z. Klimont, G. Mangaroni, H. Lotze-Campen, M. Obersteiner, A. Tabeau, M. Tavoni (2017): The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change 42: 153–168. doi:10.1016/j.gloenvcha.2016.05.009
- van Vuuren, D. P., J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. C. Hurtt, T. Kram, V. Krey, J.-F. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S. J. Smith & S. K. Rose (2011): The representative concentration pathways: an overview. Climatic Change 109 (1–2): 5–31. doi:10.1007/s10584-011-0148-z
- van Vuuren, D. P., E. Kriegler, B. C. O'Neill, K. L. Ebi, K. Riahi, T. R. Carter, J. Edmonds, S. Hallegatte, T. Kram, R. Mathur, H. Winkler (2014): A new scenario framework for Climate Change Research: scenario matrix architecture. Climatic Change 122 (3): 373–386. doi:/10.1007/s10584-013-0906-1
- Warszawski, L., K. Frieler, V. Huber, F. Piontek, O. Sedeczny & J. Schewe (2014): The inter-sectoral impact model intercomparison project (ISI-MIP): Project framework. Proceedings of the National Academy of Sciences of the United States of America 111 (9): 3228–3232. doi:10.1073/pnas.1312330110

- Weedon, G. P., S. Gomes, P. Viterbo, W. J. Shuttleworth, E. Blyth, H. Österle, J. C. Adam, N. Bellouin, O. Boucher & M. Best (2011): Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. Journal of Hydrometeorology 12 (5): 823–848. doi:10.1175/2011JHM1369.1