Component Design Report for an International Electricity Market Model

Prepared for
US Department of Energy
Energy Information Administration
Office of Electricity Analysis
November 12, 2015

Sustainable Energy Economics
Evelyn Wright and Amit Kanudia
101 West Chester St
Kingston, NY 12401
Evelyn.L.Wright@gmail.com

Executive Summary

The U.S. Energy Information Administration (EIA) Office of Energy Analysis (OEA) requires a new global electricity model, the International Electricity Market Model (IEMM), to be integrated into its World Energy Projection System Plus (WEPS+) tool. The IEMM will take regional electricity demands and fuel prices from WEPS+ and provide to WEPS+ projections of regional electricity capacity and generation by fuel and technology, power sector fuel consumption, and wholesale and retail prices for each of four WEPS+ end-use sectors.

A key design requirement is that the IEMM be very flexible in terms of regional, sectoral and time granularity. Although the IEMM will provide results to WEPS+ at the native WEPS+ level of aggregation, within the IEMM aggregation levels should be fully responsive to data availability and analysis needs and easily controlled by the user on a per-analysis basis. The IEMM should also be fully flexible with regard to fuels and technologies modeled, and must be able to model a wide range of policies impacting the power sector.

To accomplish these goals, we recommend a design for the IEMM based upon the following principles:

- Model data should be maintained at the most detailed level of granularity possible. It should be readily updated, examined, and improved in an incremental fashion.
- A user-friendly model instance generator should prepare model instances from the data at different levels of regional, temporal, and technology aggregation.
- Model structures should be data-driven and should maintain model data separately from the model equations.
- The model framework should allow for the construction of user-driven scenario definitions and be able to manage combinatorial clusters of runs using such scenarios.
- Data exploration and visualization tools should support collaborative analysis of both input data and scenario results data.
- All bulk data creation on both input and output sides should be done using rule-based specifications.

We propose a five-tier integrated knowledge and model management system based on these principles. This component design report (CDR) describes the rationale for this design, describes how it meets the IEMM performance objectives, and provides specifications for each of the components.

Table of Contents

List of Figures

List of Tables

1. Background and Introduction

The U.S. Energy Information Administration (EIA) Office of Energy Analysis (OEA) is seeking to improve its capabilities to analyze global electricity markets and technologies. Specifically, OEA wishes to develop a new International Electricity Market Model (IEMM) to replace the current World Electricity Model (WEM) within the World Energy Projection System Plus (WEPS+), the primary analytical tool used to develop projections for the International Energy Outlook (IEO).

Like the WEM, the IEMM must be capable of receiving regional electricity demands and fuel prices from other WEPS+ modules through the WEPS+ restart file, and providing to other modules projections of regional electricity capacity and generation by fuel and technology, power sector fuel consumption, and wholesale and retail prices for each of the four WEPS+ end-use sectors. These projections must be provided to WEPS+ annually throughout the IEO horizon for each of the WEPS+ regions.

OEA is seeking several enhancements to the capabilities of WEM, including:

- Improve the economic projection of capacity and dispatch, including enabling more realistic representation of operational and system considerations;
- Permit endogenous modeling of all fuels and technologies, including nuclear and renewables, which are exogenously driven in WEM;
- Enable modeling of future technology advancements, including the ability to update model technologies and fuels easily to respond to technology developments and analysis needs;
- Provide both wholesale and retail prices to WEPS+; and
- Expand scenario analysis capabilities, particularly improving the ability to analyze international carbon policy scenarios.

A priority for EIA in the design of the IEMM is flexibility of aggregation. Although integration with WEPS+ imposes a particular temporal and spatial aggregation on the outputs that must be provided to the WEPS+ system (annual, 16-region), EIA wishes the internal level of aggregation within the IEMM to be flexible and responsive to user needs. This flexibility serves several purposes:

- It permits data to be collected, maintained, and updated at its source level of granularity, facilitating the update process and allowing the model to evolve in time as data improves;
- It allows the model's representation of time periods, regions, and technology groupings to respond to analysis needs, using more detail when needed (e.g., for detailed carbon policy or renewable potential assessment within a country or region) and abstracting from detail to improve run time when not needed;
- In doing so, it substantially increases the range of analyses OEA can perform with the IEMM, without any model reprogramming.

The approach described in this component design report (CDR) is driven by these needs and our experience developing and using highly detailed electricity system models in many regions of the world. The next section describes the design principles upon which our approach is based, and Section 3 provides an overview of the approach, which is based upon a five-tier integrated knowledge and model management system. Section 4 describes how the proposed modeling approach meets the IEMM performance objectives, and Section 5 steps through the specifications for each of the system's five tiers. Section 6 discusses hardware, software, and data requirements, and a concluding section summarizes our recommendations.

The body of this CDR provides a platform-independent specification of the proposed approach. In a brief appendix we describe one open source modeling platform, the TIMES framework, which can meet the requirements of the proposed system and whose use may have several advantages over developing a model from scratch.

2. Key Design Principles

Our design is based upon the following six principles.

1. The IEMM should make best use of all available relevant data.

Data to support the IEMM should be obtained, stored, and managed at the most granular level possible. Data describing the existing generating capacity stock should be maintained at the unit level wherever it is available. The global existing stock consists of over 5000 GW of operational units, with another 800 GW under construction. Unit-level description of this capacity is particularly important given EIA's interest in carbon policy analysis, as the costs of a low carbon transition are to a great extent driven by the costs of retrofitting or replacing existing stock, particularly where there are interactions with local air quality policy (Wright and Kanudia, 2015[1]). Similarly, load shapes should be described using hourly historical load curves, where available, and renewable energy potential should be described at the finest level of resolution possible.

Data at different levels of detail may be available for different countries and regions. Model data should not be forced to a lowest common denominator approach, and the process of updating the model when better data becomes available should be friction-free.

Data should be stored in a relational database that supports the model, but is independent from the model structure. This database should also be accessible for browsing and visualization by a much wider group of users than the in-house modeling team. For example, it should be possible for EIA to invite country/regional experts to view, analyze, critique, and suggest updates to the data relevant to their expertise. (If proprietary data is used, data visualization tools should be used to aggregate data to an acceptable level for such use.)

In order to make best use of this data:

2. The capability to prepare different model "instances" from the source database is required.

Between the granular database and a ready-to-run instance of the model, an intelligent layer is needed to query the database, generate the necessary depictions of the data, and prepare it for use by the model. In this sense, the "model" is no longer a static set of data and equations, but becomes the model *plus* knowledge and model management system, within which different instances of the model may be prepared, according to analysis needs. This instance generator needs to aggregate existing capacity by region and technology type, describe new capacity options available in the region, depict the regional grid and interregional transmission capacity, aggregate renewable potential to the appropriate geographical level, and generate sectoral load

---

[1] Wright E.L. and Kanudia A. "Highly Detailed TIMES Modeling to Analyze Interactions between Air Quality and Climate Regulations in the United States." In Giannakidis, G., Labriet, M., Ó Gallachóir, B., Tosato, G. (Eds.), *Informing Energy and Climate Policies using Energy Systems Models: Insights from Scenario Analysis Increasing the Evidence Base*, Springer, 2015. http://www.springer.com/us/book/9783319165394

shapes at the desired time-slice level from 8760 data.

As a consequence of this structure, these levels of aggregation – spatial, temporal, technology, and fuel – will be subject to user control, and may be varied according to the needs of the analysis at hand. For example, an analysis of renewable penetration potential in a particular region may require a finer level of temporal and spatial resolution (even sub-country, for large countries) within the region of interest than a standard IEO run. The instance generator should be controlled by a user-friendly interface (for example, an Excel workbook) enabling the user to readily change each of these levels of granularity and even to experiment with them within a particular study to make his/her own model detail versus performance tradeoffs on the fly.

In order to permit this structure to work:

3. The structure of model instances must be data-driven.

That is, the model equations must be able to accept characterizations of regions, time periods and timeslices, technologies, and fuels at arbitrary and user-adjustable levels of detail from source data. The model framework must contain building blocks (for example, existing and new power plants) that can be readily filled by appropriately structured data, so that the instance generator can prepare and send any number of items to these building blocks, in any number of potentially interacting regions, using data drawn directly from the database. Model equations must be able to operate upon these building blocks in a structured way, again regardless of their number or detail.

Another way to express this principle is that the model and its solution logic must maintain data and equations independently. Many models have model data and equations quite entangled, which makes it difficult to make structural changes. Model equations may be structured to expect a fixed slate of fuels or technologies, or to iterate over a fixed number of regions. For the IEMM, it is essential that the model code contain no data within it. The data management system must be able to deposit data files somewhere, which the model code then retrieves and uses to construct the equations for a particular model instance. Therefore the model equations must be written in a completely generic way, designed to accept any number of items (such as power plants or fuels) of each type, to be filled in by an independent data set. Such a structure eliminates any concern about needing to reprogram or recompile the model in order to change the slate of available technologies, for example, or to reconfigure the model's regional structure.

Data so prepared by the instance generator should be browsable by an input data management system in a way that enables the user to visualize model topology (unit inputs and outputs) and review and manage all data to be sent to model equations.

4. The framework should provide powerful, flexible scenario analysis capabilities.

A key requirement for the IEMM is that it greatly expand EIA's capability to analyze policy, market, and technology scenarios. In particular, EIA is interested in analyzing international greenhouse gas controls or taxes, along with other policies, such as feed-in-tariffs and portfolio standards, that may impact power sector greenhouse gas emissions. The model framework should provide flexible "handles" for the development of user-driven scenario definitions, permitting not only simple changes in input parameters (such as new technology costs) but also allowing the user to construct constraints and/or tax/subsidies involving any arbitrary combination of model variables and using rule-based specifications to pick out precise combinations of affected power plants, fuels, and other options. As recent experience in the US illustrates, as carbon policies move from theoretical proposal towards real implementation, they may grow vastly more complex than the stylized cap-and-trade and carbon tax policies that have been routinely modeled. While EIA may not plan to assess particular country policies in the level of detail required for contemporary US policy analysis, the framework should not impose unnecessary limits on what can be modeled.

Because the costs and emissions impacts of greenhouse gas policies will depend on technology costs and fuel prices, analyzing such a policy benefits from the capability to systematically assess a range of potentially relevant uncertainties. The model interface should facilitate the development and management of run clusters that vary key policy design, market, and technology dimensions of interest systematically. That is, once the desired scenarios are defined, the interface should permit dozens of runs built by their combination to be set up, run, and processed with minimal user intervention.

To analyze the large volume of results data generated from such run clusters:

5. **The IEMM should be supported by a powerful data visualization system.**

Large volumes of results data require sophisticated analysis support tools in order to draw insights from and communicate about the data. The system should provide graphical and geographical tools for visualizing high level model outputs in flexible user-defined formats. It should also permit the user to drill down into the most detailed model results. Like the input data management system, the results data processing system should permit browsing through model network topology so that results may be traced through to their causes. The visualization system should facilitate collaboration between model users and other designated experts, for example, country or region-level analysts, by permitting views of input data and results to be saved, shared, and commented upon.

Finally, to facilitate implementation of each of these other principles:

6. **All bulk data creation necessary for each component of this system should be done in a rule-based manner.**

Each stage in the knowledge and model management process requires development and transformation of data, including:

- Data input for missing pieces of information and user-assigned inputs in the relational database that stores granular data;
- Aggregation of granular data to form model instances;
- Creating sets of processes to manage model input and output data;
- Interpolation/extrapolation of input data from data years to model periods;
- Specification of model bounds and constraints;
- Development of policy, market, and technology scenarios; and
- Creation of high-level variables for the results visualization system.

Use of rule-based specifications greatly streamlines data entry and management and increases model transparency. Simple two-line scenario files, such as those described in Section 5.3.2, create hundreds or even thousands of data values. In the case of missing data, inputs for a particular country can be developed from similar country data using a few lines, as illustrated in Section 5.1, allowing the assumptions used to be easily documented and updated when new data becomes available. In all cases, using rules rather than enumeration of data permits the modeler to manage more data, create more complex scenarios, and avoid the errors that enumeration invariably leads to. This CDR contains many examples of such rule-based specifications.

3. Methodology Overview

To realize an IEMM based upon these principles, we suggest a five-tier, integrated approach, depicted in Figure 1.

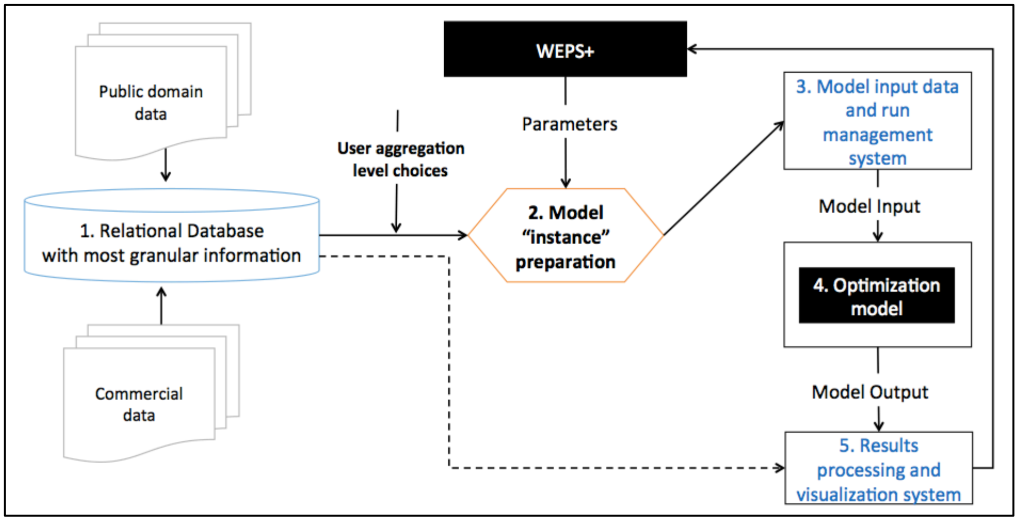


Figure 1 - Proposed IEMM Structure

The first tier is an input data management system in the form of a relational database storing all data that may serve as model input. This data should be gathered and stored at the most granular level available (e.g., unit-level data for existing capacity, hourly historical load shapes by country), so that it can be available for alternative aggregations as needed. This level may vary by region. Section 5.1 describes the types of data to be stored and the structure of the database. It also makes some initial suggestions for data sources and describes how missing data may be assigned and updated.

The second tier is a model instance generator that queries the database and prepares the requested model instance based on user-specified aggregation levels. It prepares system load curves by region and sector by allocating historical load shape data to user specifications of load segments (e.g., season, time-of-day, and peak/off-peak), and it builds model regions from the country- and unit-level data stored in the relational database, aggregating existing capacity, performing base year calibration, and setting up necessary trading relationships between regions. The instance generator also receives data from WEPS+ and prepares it for input to the IEMM.

Model source data may be aggregated into different model instances along the following dimensions:

- **Region**: Countries may be aggregated into any number and combination of regions specified by the user.
- **Process**: Existing capacity may be maintained at the unit or plant level or may be aggregated by fuel and technology type, by user specification.
- **Timeslices**: Annual load curves may be divided into seasons, portions of the week (e.g., weekday and weekend), and portions of the day.

Timeslice and process aggregations may differ by model region, in a particular instance. For example, to conduct a detailed analysis of renewable electricity potential in a particular region, it may be desirable to increase the number of load segments within that region to obtain a finer representation of the load, and to use country-level model regions, in order to capture geographical relationships between load, resources, and transmission. Other regions could then be aggregated at coarser temporal and spatial levels, in order to reduce model size while maintaining responsive global fuel markets. Such runs could be used to inform renewable supply curves and/or constraints to be utilized in later runs with a higher aggregation level. The process of model instance preparation and user control of aggregation is discussed in Section 5.2.

There is a fourth dimension along which model instances may vary:

- **Model years**: The model time horizon and model run years should be freely adjusted by the user. Model period lengths should also be variable. In general, we recommend a one-year period for the base year, in order to facilitate calibration to historical data. Model periods may then get longer as the model time horizon goes on, in order to reduce model size. (Results data may then be broken into annual reporting in post-processing.)

Our recommendation is that flexibility in this dimension be handled within the model front end (tier 3), rather than within the instance generator. This is discussed in Section 5.3, which describes the third tier, the model front end. The front end is an input data management and visualization shell able to browse each model instance numerically and graphically. This shell lets the user examine model inputs and structure, trace inputs back to source data, and define policy and other scenarios. It also serves as the model run manager, allowing the user to define model runs as combinations of scenarios and send complete run data to the model solver in tier four. When running integrated with WEPS+, the shell receives and responds to model run calls from the WEPS+ integrating module. The shell also retains run definitions for re-use and permits the user to set up batches of runs.

The fourth tier is the model itself, which receives model instance data from the front end and computes capacity (including additions, retirements, and retrofits), dispatch, fuel consumption, and emissions. The modeling framework recommended herein is a data-driven optimization model. As described in the previous section, the model framework must also provide a flexible, user-friendly method for specifying user-designed constraints. Section 4 describes how the model will meet the IEMM performance objectives, and Section 5.4 provides a schematic of the core model equations.

The fifth tier is a results data processing and visualization system. It provides powerful, flexible graphical and geographical tools for visualizing both high level and unit-level model outputs, along with a capability to drill down into the most detailed model results by browsing through model network topology (e.g., tracing back from generation to individual units' detailed operating behavior and their fuel consumption). The visualization system should also facilitate collaboration between model users and/or with country-level experts, by permitting views of input data and results to be saved, shared, and commented upon. Section 5.5 describes this system.

4. Modeling Approach

The performance objectives for the IEMM are:

- Convert annual sector electricity demands received from WEPS+ into a system load curve, with sufficient detail to model growing penetration of intermittent renewable generation. The number and timing of segments should be responsive to user control.
- Provide economic projections of dispatch and capacity (including retirements and retrofits), subject to the following considerations:
    - The model should allow endogenous modeling of all fuels and technologies, including nuclear and renewables.
    - It should facilitate modeling of future technology advancements, including the ability to update model technologies and fuels easily to respond to technology developments and analysis needs.
    - It should permit representation of operational, system, and policy constraints.
    - It should provide powerful scenario analysis capabilities, particularly the ability to analyze international carbon policy scenarios.
- Produce wholesale and end-use retail sector prices to be sent back to WEPS+. Retail prices should be able to be calculated using both average cost and marginal cost approaches, and the mix of approaches in each region should be subject to user control.

This section describes the approach to meeting these objectives, providing references to the portions of Section 5 in which the specifications for the necessary components are laid out. We also suggest, in Section 4.4, an additional feature that could decrease iteration time within WEPS+ and allow standalone IEMM operation: price responsive fuel supply and electricity demand.

4.1. Electricity Load

The model should permit the year to be divided by the user into an arbitrary number of *timeslices*, which permit aggregation of a full 8760 load shape into a manageable number of meaningful subdivisions. Typical subdivisions include:

- **Seasons**: the year may be divided into winter, summer, and intermediate; winter, summer, spring, and fall; or even individual months. The first subdivision captures differences in load, particularly space heating, cooling, and lighting, from changes in ambient temperature and day length with a minimum number of segments, whereas further subdividing the year may be valuable where hydropower plays a large role in the generation capacity and/or resource potential.
- **Week**: the week may be divided into weekdays and weekends, or even into individual days, to capture relevant variations in the load.
- **Time of day**: each 24-hour day may be divided into any number of segments, depending on the representation of the load and variable resources required. A typical set of segments would be night, day, peak, and near-peak.

At a minimum, we recommend three seasons and three times of day, but the user will want to experiment with tradeoffs between model size and load disaggregation suitable to each analysis.

The model instance generator will aggregate historical or assigned 8760 hourly load curves to the timeslices requested by the user, using the data and mapping tables described in Sections 5.1 and 5.2.3, to create a base year load shape for each region. As this approach uses the actual peak within the computation process, it also computes the model peak reserve factor that must be charged over and above the capacity reserve margin used by regional planners.

As discussed in Section 5.1, data on hourly historical overall load curves is available for a substantial portion of global load. Assigning load to individual sectors is more problematic. As EIA has documented[2], bottom-up load shape data cannot reliably reproduce system load shapes even in the United States. Nevertheless, the evolution of system load shapes as the share of building loads increases over time is a potentially important feature of the electricity system in countries with rapid economic growth. In this first phase of IEMM development, we recommend inserting scenario handles to analyze the impact of increasing peakiness in the load shape. The degree and shape of this shift should be under user control, and can be driven by econometric projections or parametric scenario analysis.

Transmission and distribution losses, based on historical values and under user control to change over time, will be then be added to the resulting projected load shape for each model year, creating the final load that will drive the capacity and dispatch portion of the model.

4.2. Capacity Planning and Dispatch

In Section 5.4, we describe an integrated dispatch and capacity expansion model based on economic optimization, subject to operational, system, and policy constraints. For flexibility and tractability with varying model size, we recommend a data-driven optimization approach, using linear programming wherever possible to reduce model size. (An exception to linearity is discussed in Section 4.2.4.)

The optimization model is supported by the knowledge and model management system, consisting of the remaining four tiers of the proposed IEMM approach, described in the remainder of Section 5. This system is the key to achieving the flexibility that EIA seeks, permitting different model instances to be prepared from granular input data, facilitating analysis of a wide range of scenarios, and allowing the model to grow over time with updated data and unforeseen analysis requirements. Accordingly, it is possible that EIA could choose to use a different model approach and solution logic, provided that the model accept and provide "hooks" to the front and back-end data management systems. EIA could even choose to use multiple modeling approaches within the overall IEMM management system. Thus the approach described in the current section and Section 5.4 can be thought of as one way to achieve the capacity planning and dispatch performance objectives.

---

[2] EIA, NEMS Electricity Market Module Documentation, http://www.eia.gov/forecasts/aeo/nems/documentation/electricity/pdf/m068(2014).pdf, p 13.

4.2.1. Model Overview

The proposed optimization model minimizes the net present value cost of meeting the electricity demands received from WEPS+ and shaped into timeslices as described in the previous subsection, taking into account capital, operating, and fuel costs, along with user-imposed taxes and subsidies and optionally welfare loss/gain resulting from changes in quantity of electricity demanded from a reference scenario (see Section 4.4). The optimization is subject to constraints that

- track existing capacity by vintage, allowing for endogenous retirement and retrofit
- enforce fuel-specific unit efficiencies and fuel shares for fuel-flexible units
- permit capacity to be used up to specified availability factors
- require reserve capacity above peak demand to be maintained, by region, and
- impose user-defined policy constraints, including portfolio requirements, emissions limits, minimum and maximum build and operation of unit types.

The model should perform this optimization dynamically, with perfect foresight over user-defined windows of the projection horizon. These windows may range from single model periods to the entire model horizon. Because the model is solved for an entire foresight window simultaneously, this foresight approach does not require iteration over multiple runs.

As a consequence of the linear programming structure, the solution process will produce marginal costs (or shadow prices) on each energy carrier by region and timeslice, as well as on all constraints, including reserve constraints and user-imposed policy constraints. These marginals are valuable for solution evaluation and model diagnosis and debugging and will also be used to develop end-use electricity prices by region, as described in Section 4.3.

It should be noted that, while linear optimization models have some well-known pitfalls, most notably penny-switching behavior – the tendency to switch technology mixes dramatically in response to small changes in technology prices – the model proposed herein will be far less subject to such issues than perhaps anticipated. Penny switching is not so much the result of linear optimization per se, as of a very sparse supply curve. When there are very few steps on the supply curve, small changes in costs may indeed lead to large shifts in model solutions. The model described herein, based on country and unit-level data, will have a very rich supply curve. In particular, the existing capacity base will be described with many steps of different costs and efficiencies corresponding to different vintages within each technology type. As a result, although the linear program will still shift from one "step" of the supply curve to another when the solution changes, the shifts in overall model solution will be more realistic.

As discussed in Section 2, the model framework must support the following capabilities:

- The model structure must be data driven. That is, it must accept any number of items (regions, power generating units, fuels), subject only to data specification.
- The model must permit data specified in any year and produce data for model years by interpolation/extrapolation following rules specified with each parameter.
- The model must allow user-specified constraints to be built out of any linear combination of model variables, including across time periods.

Sections 5.3 and 5.4 describe how these capabilities are to be implemented. The remainder of this section describes how they are used to fulfill the performance requirements.

4.2.2. Technologies and Fuels

A data-driven model represents the power system as a network of fuels and technologies. For each unit, the applicable input fuels must be specified by data input, and the unit may take any combination of its applicable fuels, subject to any user-specified limitations on blending of fuels. Thus changing the fuel suite simply requires updating the specification of which units take which fuels in the input data. At a minimum, the fuel suite will include coal, natural gas, petroleum products, biomass, and nuclear fuels, but these may be augmented or further broken down at any time.

Similarly, the technology suite is data driven. New unit technology types must be described with costs, fuel input(s), efficiency (heat rate), availability, peak contribution, and any limitations on operation. Air pollutant emissions factors may also be specified. Data to describe new technologies may be drawn from EIA, IEA, ETSAP, and other sources. Economic life (for purposes of cost recovery) and technical life may be specified separately by unit. Section 5.1 describes the database structure used to store and manage this data. A minimal suite of technologies might include: subcritical, supercritical, and IGCC coal, with or without carbon capture and storage (CCS); natural gas combined cycle; natural gas/petroleum turbine; new nuclear, potentially of different designs; run-of-river and reservoir hydro; on-shore and off-shore wind; photovoltaic and thermal solar; geothermal, and biomass (with or without CCS). The model should also be able to represent electricity storage processes, which “consume” electricity in one timeslice and release it in another, subject to some storage efficiency and maximum storage capacity.

As with fuels, this suite of technologies may be updated at any time simply by specifying the required data. For example, onshore wind could be broken down further into multiple turbine heights and/or types. The available suite of technologies may also vary by region, again simply by data specification. New technology types that are not considered currently commercially available may be represented with delayed start years, switched on by scenario, or specified with limited exogenous builds. Risk premiums, in the form of technology-specific discount rates, may also be specified by unit type and region.

Existing capacity should be stored in the IEMM database at the unit level and aggregated by fuel, type, and region by the model instance generator, as described in Section 5.2. The aggregation may vary by region, under user control. Because existing units are represented in this way, the choice of new technology types is in no way dependent upon or tied to existing unit types.

Crucially, the model should track all technologies, existing and new, by vintage throughout the model equations. For example, the model should be aware that the existing stock of coal steam units in Region01 consists of x% built in 1974, y% in 1986, and z% in 1999. Key parameters including costs and efficiencies will be assigned by vintage, and the model equations will use these costs and efficiencies for the corresponding portions of the stock. If fixed technology lifetimes are used, the model will also retire the appropriate portion of the stock in the appropriate year. In this way, vintaging allows for a range of costs and heat rates within a plant type, and for existing units, these parameters will endogenously change as vintages of the existing stock retire.

The model should include an equation requiring each region to maintain a specified reserve capacity above the regional system peak. The contribution of different unit types to meeting this reserve requirement may be specified and may be modified over time as system configuration changes.

The model must provide a facility to specify upper and lower bounds on the investment, capacity, inputs and outputs (including emissions), and/or operation of each model plant, and, like all data inputs, these bounds must be specifiable by rule. The model should also permit the user to write constraints consisting of any linear combination of model variables, again specifiable by rule. These bounds and constraints can be used to achieve the following:

- Assign some units must run status;
- Require specified units to operate at a minimum level or face mothballing and recommissioning costs;
- Require an increase in operating reserves with increasing intermittent renewable penetration, by region;
- Provide exogenous specification of builds, retirement, and dispatch by unit type (e.g., for scenarios in which the user chooses to specify exogenous nuclear builds in place of using native endogenous capacity expansion);
- Impose growth rate limitations and/or growth rate capital cost adders by unit type and region; and
- Develop policy scenario specifications, including emissions limits and portfolio standards.

Examples of how such constraints are built and used in scenario files are provided in Section 5.3. All such constraints and limitations should be readily modifiable in scenario analysis, and combined with other scenario specifications in a combinatorial way.

In addition, all model unit outputs should be subject to user-imposed costs (e.g., carbon taxes) and/or further utilization in additional technologies. For example, technologies conducting downstream processing and utilization of carbon captured from new and retrofitted units with CCS in underground storage, for enhanced oil recovery, or for other uses should be able to be represented in the same data-driven way as power generating units, including costs for infrastructure development and transportation of captured carbon from site of capture to site of use.

4.2.3. Renewable Potential and Grid Integration

We recommend that resource potential data for wind, solar, hydro, geothermal, and biomass be developed at the most granular level available and aggregated to the country level, subject to further aggregation into regions, as described in Section 5.2. For non-combustion technologies, this potential will define a bound on capacity, potentially by resource and cost class, that can be utilized by technologies in that class. Using the same rule-based constraints described above, multiple technologies using each resource (e.g., different wind turbines, alternate solar thermal designs) will be able to utilize each resource, dependent only on technology description data.

(For example, a bound can be imposed limiting the sum of all technologies that include "WinOn3" in their names to the total potential for class 3 onshore wind within a given region.) For biomass resources, stepped supply curves should be developed to describe the resource base, which may then be designated as a fuel input to appropriate model plants.

Resources/technologies with seasonal and/or diurnal generation patterns must be represented with time of day capacity factors (for wind and solar) and seasonal availability (for hydro). Like all other model data, such capacity factors should be maintained in the model database at the most granular level possible (e.g., hourly for wind and solar, monthly for hydro) and aggregated to model timeslices in the model instance generator.

4.2.4. Technology Learning

Technology learning may be represented exogenously, through changes in projected technology cost and performance over time, and/or endogenously. Endogenous technology learning employs a learning-by-doing (LBD) approach, in which technology parameters, most often capital costs, change in response to cumulative experience with the technology, as demonstrated by cumulative installed capacity.

Empirical studies have supported a "learning curve" formulation of LBD effects on capital costs, in which the investment cost is an inverse exponential function of cumulative capacity investment C:

$$INVCOST = a \cdot C^{-b}$$

where $\boldsymbol{a}$ is the initial unit investment cost (when $\boldsymbol{C}$ is equal to 1) and $\boldsymbol{b}$ is the learning index, representing the speed of learning. Under this formulation, initial increases in cumulative capacity rapidly bring down investment costs. The rate of learning then slows down considerably after a few capacity doublings, leaving a mature technology whose cost changes very slowly.

This formulation may be piecewise linearized, by specifying the initial cost and capacity, the learning index, a maximum cumulative investment for LBD purposes, and the number of segments into which to break the learning curve for linear approximation. However, binary variables are required to identify which segment of the learning curve the cumulative capacity lies upon in each model period, turning the model into a mixed integer programming (MIP) problem.

Component learning may be modeled by breaking the capital cost of each technology using the component into two parts: one corresponding to the component, which is subject to the LBD formulation, and one corresponding to the remainder of the technology, which may be fixed in time, exogenously changing, or subject to another LBD process. An additional constraint tracks cumulative investment in the component resulting from investment in all technologies that use it, in order to drive the LBD equation. Learning-driven by investment in multiple regions (as for internationally traded components) may be implemented by created a dummy model region with a dummy "learning technology" for each LBD technology. Investment in the real technology in all affected regions is tied to investment in the dummy learning technology by constraint, and it is the investment cost of the dummy technology that is subject to LBD.

We would recommend that analysis goals be evaluated against the potential model structures and model behavior in order to arrive at a learning design. The conversion of the model to a MIP problem can increase runtimes. One possibility is to use the LBD approach to assess technology learning behavior under scenarios of interest and use the results to inform exogenous learning trajectories to be used in later LP runs.

4.3. Producing End-use Prices

As an integral part of the model solution process, the model produces marginal prices (shadow prices) of generation by region and time step, average cost of generation by region, and the shadow price on meeting reserve capacity requirements by region. These building blocks may be combined in post-processing to produce generation prices under cost-of-service and competitive regulatory regimes in a similar manner to the NEMS Electricity Market Module.

Under cost of service regimes, the generation price is calculated as the average cost of generation, which is simply the sum of all system costs (including capital, operating, fuel, and policy compliance costs) divided by total generation. Because this calculation is performed in post-processing, it may be modified to exclude certain costs, or may be subject to other user-designated adjustments. Capital recovery for the existing asset base will be included in the average cost, because the model with track and annualize investment costs for past investments over their economic lifetimes in the same manner as new investments. These investment costs will be stored in the unit database and may be assigned based on costs for new technologies, historical values where available, or some other heuristic. Under competitive pricing, the retail price is the sum of an energy component based on the marginal price on generation, and a capacity component based on the shadow price on the reserve requirement constraint.

The blend of these approaches may be specified by region. To produce end-use prices, average transmission and distribution adders, taxes, and subsidies may be added by sector. These may be specified by country and year, with projections under user control, and aggregated by region. Using these approaches, the breakdown of generation, transmission, and distribution components of prices can also be reported.

4.4. Price Responsive Fuel Supply and Electricity Demand

The IEMM specifications are to receive fixed annual sectoral electricity demand quantities and fuel prices from WEPS+ as model inputs, and to return sectoral electricity prices and fuel consumption to WEPS+ for iteration. The system described meets these requirements. However, we suggest an optional augmentation of the IEMM capabilities that would permit the IEMM to run in standalone mode for faster scenario exploration and may reduce iteration times within WEPS+: to develop the capability for price responsive behavior on both the fuel input and electricity demand sides of the capacity and dispatch model.

On the fuel side, we suggest developing regional stepped supply curves for the fossil fuels represented in WEPS+. These could be developed from runs of WEPS+ and/or its component modules. Each supply "curve" would consist of a series of steps with a resource cost and upper bound, and would represent, in an approximate fashion, the response of WEPS+ to changes in the quantity of fuel demanded by the IEMM, permitting the IEMM to respond to changing prices and a first pass at economic equilibration to occur within the IEMM.

On the electricity demand side, we suggest supplying, along with a quantity demanded, a price elasticity of demand, again derived from WEPS+ runs, perhaps informed by analyst judgment. An IEMM run would include a stepwise linear approximation of the demand curve thus described, with quantity demanded then able to adjust to changes in electricity price. The gains or losses in consumer surplus associated with change in quantity demanded would be included in the model objective function under this formulation.

If included in the IEMM design, either or both of these price responsive adjustment mechanisms could be turned off at runtime by user specification.

5. Component Specifications

This section describes the specifications for each of the five component tiers necessary to realize the proposed design.

5.1. Input Data Management System

The foundation of the recommended IEMM design is a knowledge management (KM) system that stores and manages all data to be used in preparing IEMM model instances. This system should be designed according to the following criteria:

- It should be able to store data at the most granular level available. This level may vary by region.
- It should be easy to update.
- It should be accessible to browsing and visualization using the IEMM data processing system.
- All data that is based on assumptions or inheritance should be clearly distinguished. Further, such data should not be replicated manually across any dimension (region, technology, etc.) This should be done via automated rules processing.

This section describes the types of data to be stored and the structure of the database. It makes some initial suggestions for data sources and describes how missing data may be assigned and updated. First, however, we offer a perspective on data availability and data assignment.

Data available to describe the electricity system may vary greatly by country. Below we make suggestions about data we know to be available at the global level as well as piecemeal by country. For some key data, such as historical 8760 load curves, data may only be available for a handful of countries, and rules for assigning data to other countries will need to be employed. Some data, such as resource potential for biomass resources, may have to be sought out in regional or country-specific studies. At first it may seem that this patchiness of data availability creates an excessive burden to model development.

However, at this point in time, only 20-40 countries host 80-90% of the global generation capacity, as shown in Figure 2. The top 20 countries are listed in Table 1. Many of the countries in Table 1 have excellent data readily available.

This clumpiness of capacity substantially reduces the data availability problem. Efforts to develop country level data should focus on a designated set of the top capacity countries along with a few more countries that EIA judges will likely become important players in the next few decades. Missing information (such as load curve) for the remaining countries can be *inherited* from similar countries in the first list. As described below, this inheritance should be done via automated processing of rules so that country-specific data can be easily injected in a progressive manner, and so that these assumptions can be documented and changed easily.

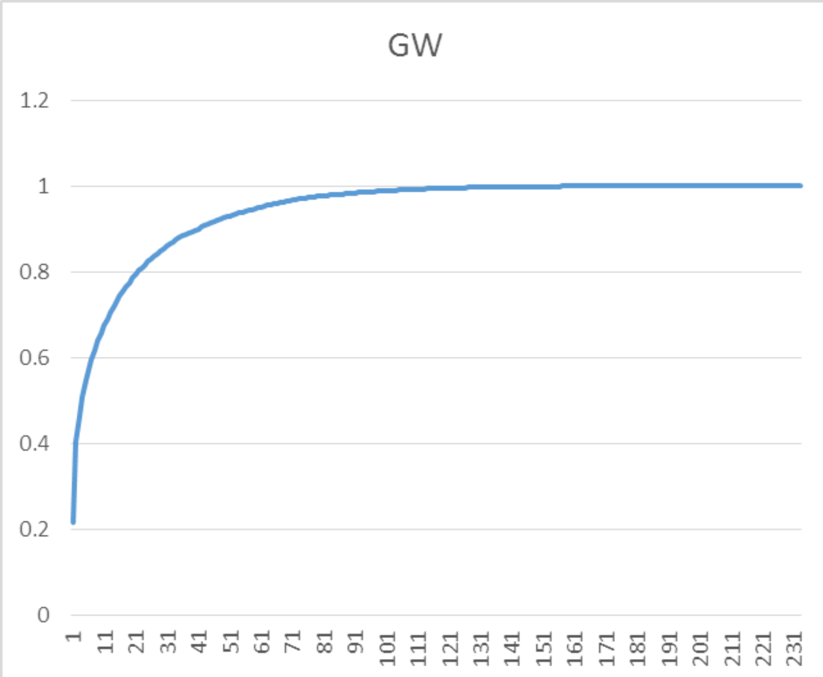


Figure 2 - Cumulative share of generation capacity (operational and planned) by country (Source: Platts 2014)

Table 1 - Share of total operational and planned capacity by country (Top 20) (Source: Platts 2014)

Country	% of Global Capacity
China	21.6%
United States	16.1%
India	9.6%
Russia	3.9%
Japan	3.2%
Brazil	2.2%
Germany	2.1%
Canada	2.0%
France	1.6%
South Korea	1.6%
England & Wales	1.6%
Italy	1.5%
Spain	1.5%
Turkey	1.4%
Saudi Arabia	1.3%
Vietnam	1.2%
Indonesia	1.2%
Iran	1.2%
Australia	1.0%

5.1.1. Database Content and Structure

The system should be able to store the following types of information:

1. Existing electricity generation units (including those under construction)
2. Planned electricity capacity
3. New electricity generation technologies
4. Existing electricity trade networks
5. Potential electricity trade links
6. Electricity demand load curves
7. Wind, solar, biomass, geothermal and hydro potential
8. Data related to electricity price calculations

The following tables provide the database schemas and attribute listings for the information listed above. This database structure could be executed on any standard database platform like MySQL, MS SQL Server or ORACLE.

Table 2 - Database schema for information on existing and new power generating units (data types 1 to 3 above)

FieldName	Remarks
Country	
State	
City	
TechName	A unique short name for existing units; Type/Fuel/Country code for new units
TechDesc	This field will be constructed from other fields (e.g., type, fuel, country) to create handles for rule-based data entry in the model front end
TechType	Steam, CC, GT etc. (This could be used to make assumptions on several dispatch-related attributes)
Input Fuel	Multiple fuels should be possible for each unit
Output	Electricity, waste heat, CO2 and other pollutants
Status	Operational, construction, planned
Attribute	Capacity, efficiency, cost etc.
Year	Online year for existing units and time series for new ones
Currency	
TimeSlice	For qualifying seasonal/diurnal attributes
VALUE	Numeric field for attribute values

In order to ease the process of database updates, all data should be stored in the native units of the data source, as far as possible. Any unit harmonization should be done at the time of visualization or extracting the data for modeling purposes.

The model will need all parameters to be specified for each region/technology/commodity for each period that the model runs. However, it should not be necessary for the database to unnecessarily duplicate data to match the model's requirements. Rather, data should be stored according to its natural years.

For example, assume that the model runs with 5-year periods between 2015 and 2050. For existing hydro units, we would have availability factors only for the historical year when they came online, and we will likely assume them to stay constant in all future periods. Further, assume that we wish to run a larger model (with more regions and more detailed breakout of model plants), where we want the post 2030 periods to be 10 years long. It is easy to imagine the additional data burden if all data were to be generated for each model period at this stage of the process. Further, changing the period definitions would require a recreation of the data to match the model periods every time.

Both issues could be resolved if the task of interpolation/extrapolation of model parameters is performed by the model code instead of the data management system, according to simple interpolation rules specified with each attribute.

For each existing and new (future) power unit, the following attributes should be specified.

Table 3 - Attribute listing for power generating units

Attribute	Remarks	Interpolation Rule
Installed capacity	For existing/under construction unit	None
Maximum availability	By time slice	LI-CE[3]
Minimum operation level	For thermal units	LI-CE
Thermal efficiency at max load	By fuel	LI-CE
Thermal efficiency at min load		LI-CE
Output factors	Flow of output commodities tied to tech activity or to input flows	LI-CE
Ramp rate	Hours to dispatch at full capacity	LI-CE
Variable operation cost	Along with age profile	LI-CE
Fixed operation cost	Along with age profile	LI-CE
Investment cost	Also assigned to existing units for cost recovery	LI-CE
Start-up cost	Cold start and warm start	LI-CE
Technical life	Endogenous retirement should also be an option	LI-CE
Economic life	Period over which investment cost will be amortized	LI-CE
First year of tech availability	For new technologies	None
Tech level discount rate	Risk premium	LI-CE
Peak Contribution factor	Also called capacity value. Corresponds to the probability of being available to dispatch during peak demand, and enters the reserve requirement constraint	LI-CE

[3] Linear interpolation and extrapolation with constant values (backward as well as forward)

Unit-level information on capacity, vintage, fuels, technology type is readily available from Platts[4] for all existing and under construction power plants in the World. For other key parameters, such as thermal efficiency and operating costs, values can be assigned based on the technology type, fuel and vintage, and calibrated where possible to national statistics. For example, IEA national energy balances and, where available, country statistics can be used to check and calibrate efficiency assignments.

It is extremely important that all such assumptions be applied via automated processing of rules and not via manual data entry. Table 4 shows an example of a table used to inject systematic assumptions into the database. The table assigns thermal efficiencies to units by specifying the vintage (Year), Fuel, and type (TehcType) of unit to be affected. The AllRegions column allows the same data to be set for units in all countries, while the subsequent country columns allow alternate assumptions to be set on a country basis. So the example shown sets the same efficiency in all countries for gas and oil-fueled combined cycle (CC) units by year, whereas the coal steam example applies the same values to all countries except India and China.

Factors for output commodities like CO_2 should be linked to the input fuel(s), without even specifying a technology, rather than to activity of each technology. This results in a very terse and accurate specification, as it deals with the flexibility that may exist in input fuels (dual-fuel units) and with efficiency improvements over time. We just need one data value per fuel and region.

Table 4 - Example of rule-based application of assumptions

Attribute	Year	AllRegions	India	China	USA	Fuel	TechType
Thermal Efficiency	1960	0.35				Oil, Gas	CC
Thermal Efficiency	1965	0.36				Oil, Gas	CC
Thermal Efficiency	1970	0.37				Oil, Gas	CC
Thermal Efficiency	1975	0.39				Oil, Gas	CC
Thermal Efficiency	1980	0.40				Oil, Gas	CC
Thermal Efficiency	1985	0.41				Oil, Gas	CC
Thermal Efficiency	1990	0.42				Oil, Gas	CC
Thermal Efficiency	1995	0.43				Oil, Gas	CC
Thermal Efficiency	2000	0.44				Oil, Gas	CC
Thermal Efficiency	2005	0.45				Oil, Gas	CC
Thermal Efficiency	2010	0.46				Oil, Gas	CC
Thermal Efficiency	2015	0.47				Oil, Gas	CC
Thermal Efficiency	1960	0.28	0.25	0.26		Coal	ST
Thermal Efficiency	1965	0.29	0.26	0.28		Coal	ST
Thermal Efficiency	1970	0.30	0.27	0.29		Coal	ST
Thermal Efficiency	1975	0.32	0.28	0.30		Coal	ST
Thermal Efficiency	1980	0.33	0.30	0.31		Coal	ST

[4] http://www.platts.com/products/world-electric-power-plants-database

Attribute	Year	AllRegions	India	China	USA	Fuel	TechType
Thermal Efficiency	1985	0.34	0.31	0.32		Coal	ST
Thermal Efficiency	1990	0.35	0.32	0.34		Coal	ST
Thermal Efficiency	1995	0.37	0.33	0.35		Coal	ST
Thermal Efficiency	2000	0.38	0.34	0.36		Coal	ST
Thermal Efficiency	2005	0.39	0.35	0.37		Coal	ST
Thermal Efficiency	2010	0.40	0.36	0.38		Coal	ST
Thermal Efficiency	2015	0.42	0.38	0.40		Coal	ST

This approach has several significant advantages over the manual option including efficiency, transparency, and the ability to refine assumptions easily and test and maintain alternate assumptions for scenario analysis and model calibration in a parsimonious manner. Use of rule-based entry is particularly important for dealing with missing data. As discussed above, we anticipate being able to access data describing countries with a large portion of the global existing capacity, but data for other countries will be partial or missing, at least at first. Tables like Table 4 should be used to assign values to missing data by rule based on data for similar countries. This approach is self-documenting. It allows anyone to see, and easily edit, the rule used to assign the values.

Information on new technologies should be derived from sources like IEA and EIA. Different sets of new technology data can be maintained as a scenario dimension in model runs. Again, it is important to note that data for new technologies should not be replicated for each country in the core database. As in Table 4, common data should be assigned to all regions, and only parameters that are country specific should be specified at the country level.

Information on existing and new inter-country transmission links should be stored in the structure shown in Table 5.

Table 5 - Database schema for transmission links (data types 4 and 5)

FieldName	Remarks
Exporting country	
Importing country	
TechName	
TechDesc	
Status	Operational, construction, planned
Attribute	
Year	
Currency	
TimeSlice	For qualifying seasonal/diurnal Attributes
VALUE	Numeric field for attribute values

Very good data is available for European countries from ENTSO-E[5]. IRENA could be a possible source for other regions. Although there is limited interregional transmission capacity today at the level of the WEPS+ regions, this may be an important area for scenario analysis in high renewable penetration scenarios.

The following attributes should be captured for transmission links.

Table 6 - Attribute listing for transmission links

Attribute	Remarks	Interpolation Rule
Installed capacity	For existing/under construction links	None
Maximum availability	By time slice	LI-CE
Fixed operation cost	Along with age profile	LI-CE
Investment cost		LI-CE
Transmission efficiency		LI-CE
Technical life	Endogenous retirement should be an option	LI-CE
Economic life	Period over which investment cost will be amortized	LI-CE
Maximum capacity	For new links	LI-CE

Overall electricity demand load curve should be stored at the hourly level, by country, as per the structure shown in Table 7.

Table 7 - Database schema for commodity-based attributes like load curves

FieldName	Remarks
Country	
Commodity	Electricity; may want to add heat later on
Attribute	
Year	For projected transmission efficiency, reserve requirements
Date	
Hour	
VALUE	Numeric field for attribute values

ENTSO-E provides this data for all European countries. Similar data is available from national sources for other major countries including the US, China, India and Japan. It should suffice to cover majority of the global electricity system well.

Table 8 - Commodity-based attributes list

Attribute	Remarks	Interpolation Rule
Hourly power demand	For each hour of a year	None
Transmission efficiency	1 MINUS Transmission and distribution losses	LI-CE
Operational reserve requirement	System reserve requirement set by system operators. May be changed over time.	LI-CE

[5] https://www.entsoe.eu

The database schema for renewable potential is given in Table 9. The potential for each resource should be disaggregated into several resource and cost classes for each country.

Table 9 - Database schema for renewable potential

FieldName	Remarks
Country	
State	
City	
Resource	Wind, solar, hydro, geothermal, biomass
ResourceClass	Drives availability factor
CostClass	Driven by distance from grid and load centers; supply step for biomass
Attribute	
Year	
Currency	
TimeSlice	For qualifying seasonal/diurnal attributes
VALUE	Numeric field for attribute values

The following attributes should be used to characterize renewable technologies:

Table 10 - Attribute listing for renewable technologies

Attribute	Remarks	Interpolation Rule
Maximum Capacity	Potential; size of supply step for biomass	LI-CE
Maximum availability	By timeslice	LI-CE
Variability	Expected variation in availability (by timeslice)	LI-CE
Peak Contribution factor	Based on variability	LI-CE
Variable operation cost	Along with age profile; supply step cost for biomass	LI-CE
Fixed operation cost	Along with age profile	LI-CE
Investment cost		LI-CE
Technical life		LI-CE
Economic life	Period over which investment cost will be amortized	LI-CE
First year of tech availability	For new technologies	None
Tech level discount rate	Risk premium	LI-CE
Electricity Consumption	Historical values (to bound total RE potential under regional aggregation)	LI-CE
GDP	Historical and projected values	

NREL has released country-wise data for onshore and offshore wind potentials in 9 wind and 3-6 cost classes (onshore-offshore). Similar data should be available for solar and wind from IRENA and other sources. IRENA also maintains a database of country-level renewable potential studies and a global atlas of renewable datasets[6]. Global models such as MAgPIE[7] may be a source of

[6] http://www.irena.org/potential_studies/ and http://irena.org/globalatlas

regional biomass supply curves for different levels of technological progress (optimism) and carbon prices. It should be possible to use RE potential data from alternate sources as a scenario dimension in model runs.

Historical electricity consumption and GDP at the country level are available from the World Bank[8]. IMF[9] publishes GDP projections up to 2020. GDP projections for several countries are available from sources like PWC[10].

To calculate end-use electricity prices, as discussed in Section 4.3, we need country-level information on regulatory regime, sector distribution markups, and taxes and subsidies. As discussed above, this data may need to be developed first for a select number of countries of interest, and then inherited by analyst judgment-informed rule to other, similar countries. Tables 11 and 12 provide the schema and attribute listings for this data.

Table 11 - Database schema for price-related data

FieldName	Remarks
Country	
Sector	For T&D adders, taxes, and subsidies
Attribute	
Year	
Currency	
VALUE	Numeric field for attribute values

Table 12 - Attribute listing for price-related data

Attribute	Remarks	Interpolation Rule
Regulatory regime	Competitive or regulated. Can be expressed as fraction of country load competitive for countries with mixed regimes.	LI-CE
T&D adder		LI-CE
Tax/subsidy		LI-CE

5.1.2. Database Analysis and Visualization

The IEMM data visualization system described in Section 5.5 should be able to browse and visualize the database. This capability will support:

- Assessment of data and data development priorities (for example, prioritization of country-specific data, as discussed above);
- Choice of model instance aggregation, for example, by assessing preponderance of unit types in various countries, or comparability of countries for region aggregation;

---

[7] https://www.pik-potsdam.de/research/projects/activities/land-use-modelling/magpie
[8] http://data.worldbank.org/data-catalog/world-development-indicators
[9] https://www.imf.org/external/pubs/ft/weo/data/changes.htm
[10] http://www.pwc.com/gx/en/issues/economy/the-world-in-2050.html

- Quality control and assessment for data updates; and
- Development of scenarios for analysis.

Figures 3-6 show examples of visualizations that should be possible, drawing upon some of the sample data sources mentioned above. Figures 3 and 4 show different views of the Platts database of existing unit, which can be used for various types of explorations of the global electricity generation stock. For example, we can see the operational and planned capacity broken out by fuel at the country level in Figure 3, while Figure 4 shows the mix of different types of generation techs at a global level, separately for operational and under construction units.

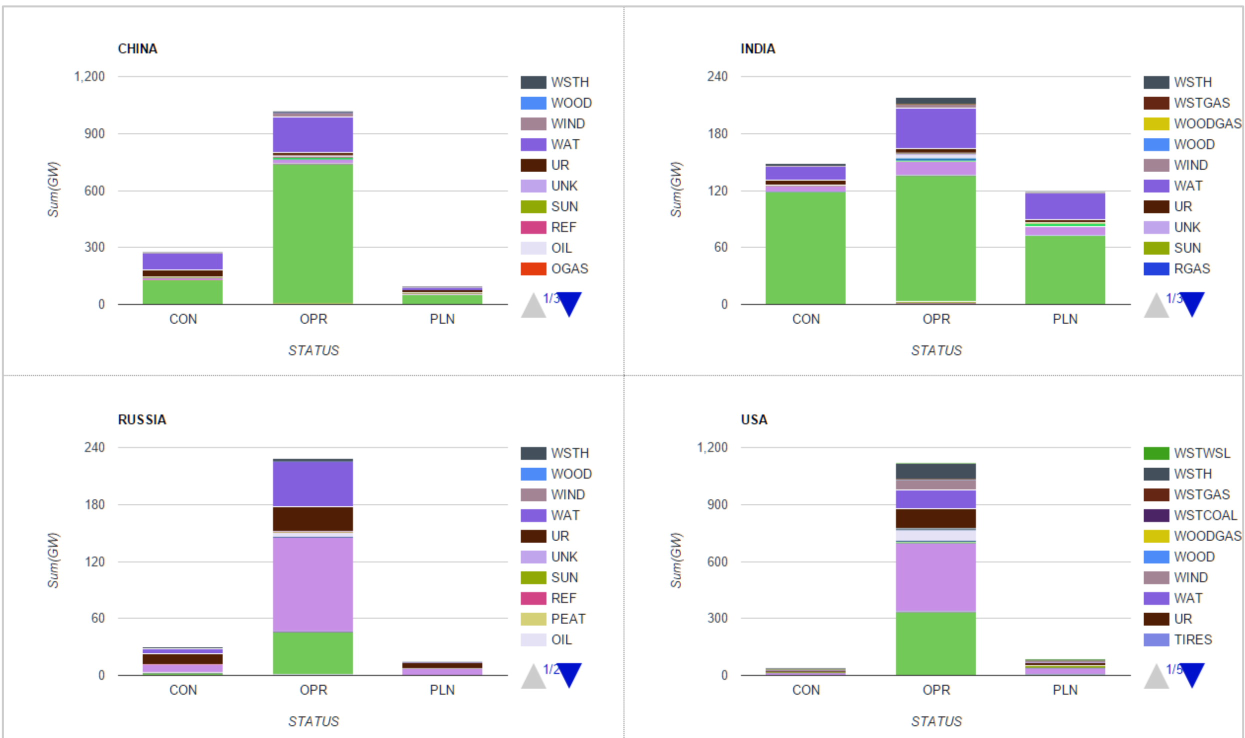


Figure 3 - Total operational (OPR), under construction (CON) and planned (PLN) electricity capacity by primary fuel (Source: Platts WORLD ELECTRIC POWER PLANTS DATABASE, 2014)

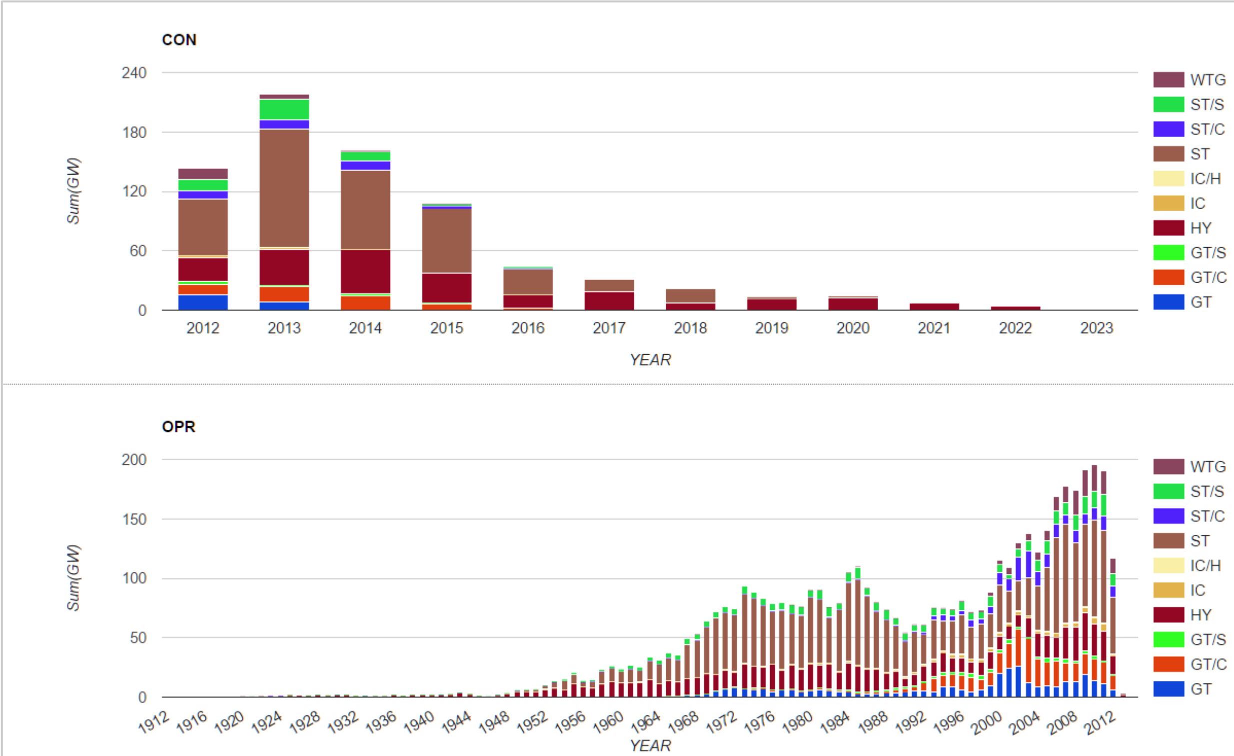


Figure 4 - Global view of operational (OPR) and under construction (CON) capacity, by type of generator (Source: Platts WORLD ELECTRIC POWER PLANTS DATABASE, 2014)

Figures 5 and 6 show two views of NREL global onshore and offshore wind resources, grouped by resource type and distance from load/shore, for the global resource in Figure 5 and by country in Figure 6. Note that we can see that caution must be used when adding across countries to form regions as some countries, like Canada and USA, have resources far above their total consumption. This observation will guide the setting of bounds for country potential to be used when aggregating countries into regions, as described in Section 5.1.2.

Figure 5 - Global onshore and offshore wind resource (Twh) by Depth and distance from shore for offshore and by distance from load for onshore (Source: NREL Global CFDDA-based Onshore and Offshore Wind Potential Supply Curves[11])

[11] http://en.openei.org/datasets/dataset/global-cfdda-based-onshore-and-offshore-wind-potential-supply-curves-by-country-class-and-depth-q

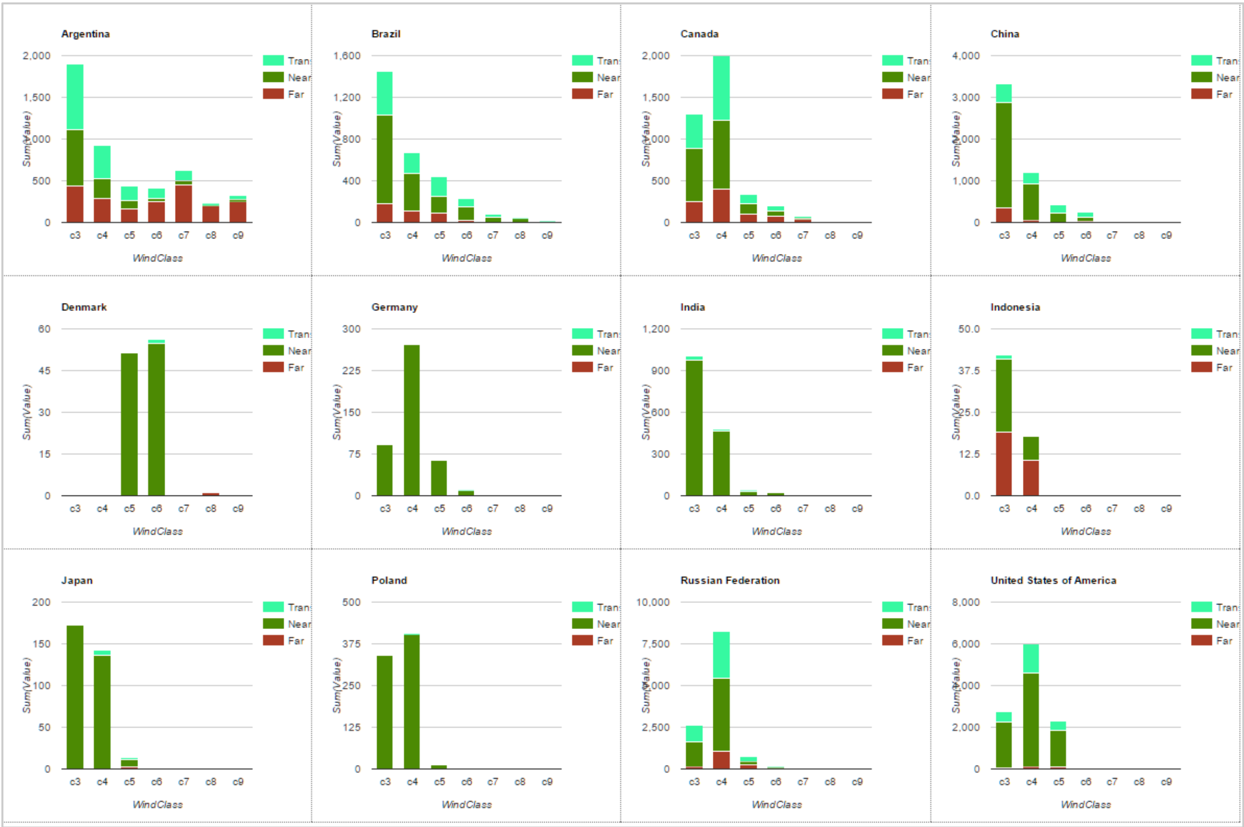


Figure 6 - Onshore wind resource (GW) by cost class and wind class, by country (Source: NREL Global CFDDA-based Onshore and Offshore Wind Potential Supply Curves)

5.2. The Model Instance Generator

The model instance generator prepares data from WEPS+ and the relational database into a model instance, subject to user control. This section describes how data from WEPS+ should be handled and for each data type in the database to be aggregated, the dimensions affected, the tables to be used by the user to control the aggregation level, and the equations to be used to perform the aggregation.

5.2.1. WEPS+ Data Handling

Beginning with the input from WEPS+, the schema and attributes shown in 13 and 14 should be used to capture the electricity demands and fuel prices, by region/sector and year. We make an important recommendation here: instead of using single points for electricity demand and fuel prices, we should develop demand/supply curves to help the process of convergence. We need two additional pieces of information for this: electricity price and fuel quantities, and the elasticities. The first can come from previous iterations or other model runs. And the second could even be based on expert judgment. Elasticities could also be based on parametric runs of WEPS+ with the current electricity module. Developing this data will also allow the IEMM to be run in standalone mode for scenario exploration, if desired.

Table 13 - Database schema to store input from WEPS+

FieldName	Remarks
Region	
Sector	
Commodity	
Attribute	
Year	

Table 14 - Attributes to capture the data (and supporting assumptions) from WEPS+

Attribute	Remarks
Demand	Electricity demand by sector and model region
FuelPrice	
Elasticity	To make a step-wise linear supply curve around the P/Q from WEPS+
FuelQty	To create the supply curve

5.2.2. Aggregation of Process and Region Data

The relational database holds granular data describing existing units at country, unit and year levels, which needs to be transformed to model region and model plant level. Table 15 lists the dimensions that will be affected by the aggregation process along with the notation that will be used in the expressions that follow.

Table 15 - Dimensions for describing power units

Dimension	Description
c	Country
u	Unit
v	Vintage
R	Model region
T	Model plant (by fuel and technology type)

The aggregation process is controlled by means of tables that provide a mapping of dimensions from the granular database to dimensions used in model instances. These tables can be housed in a Microsoft Excel spreadsheet, under user control, and will drive the extraction and transformation of granular data from the database into a particular model instance.

Table 16 shows a sample mapping from database countries to model instance regions. The user enters in column 2 the region that each country belongs to. The example shown here uses numbered regions, but the user may choose any names for the regions.

Table 16 - Mapping countries to model regions

Country [c]	ModelRegion [R]
Afghanistan	ModReg12
Albania	ModReg13
Algeria	ModReg01
American Samoa	ModReg16

Country [c]	ModelRegion [R]
Andorra	ModReg07
Angola	ModReg01
Anguilla	ModReg06
Antarctica	ModReg00
Antigua And Barbuda	ModReg06
Argentina	ModReg06
Armenia	ModReg03
Aruba	ModReg06
Australia	ModReg02
Austria	ModReg07
Azerbaijan	ModReg03
Bahamas	ModReg06
Bahrain	ModReg10
Bangladesh	ModReg12
…	…
United Kingdom	ModReg07
United States	ModReg16
United States Minor Outlying Islands	ModReg16
Uruguay	ModReg06
Uzbekistan	ModReg03
Vanuatu	ModReg12
Venezuela	ModReg06
Viet Nam	ModReg12
Virgin Islands, British	ModReg07
Virgin Islands, U.S.	ModReg16
Wallis And Futuna	ModReg12
Western Sahara	ModReg01
Yemen	ModReg10
Zambia	ModReg01
Zimbabwe	ModReg01

Table 17 provides an example mapping of database units by fuel and techtype to model plants. Here the user has specified in column 2 which fuels and techtypes are to be combined into model plants. For example, gas and oil steam units have been combined into a GasOil Steam model plant.

Table 17 - Mapping units to Model Plants

[u]		[T]
Fuel	TechType	ModelPlant
COAL	ST	CoalSteam
WAT	HY	Hydro
UR	ST	Nuc
GAS	GT/C	GasOil-CT
GAS	CC	Gas-CC
COAL	ST/S	CoalSteam
GAS	GT	GasOil-CT
WIND	WTG	Wind
OIL	ST	GasOil-ST
GAS	ST	GasOil-ST
OIL	GT	GasOil-CT
OIL	IC	Oil-IC
…	…	

While in principle database units could also be aggregated by vintage to construct, for example, model plants that differ by decade of age, a better approach would be to require the model formulation to support vintaging of technology attributes. That is, the model should identify different vintages of capacity for the same technology and track technology attributes, such as efficiency and costs, by vintage. This facilitates a more compact depiction of model capacity while allowing accurate representation of the different performance characteristics of units even when they are aggregated into model plants. With vintaging, the performance characteristics of the same process will keep improving as the older vintages retire, endogenously or exogenously.

Once mappings are defined, the aggregation is conducted using the following equations. The capacity of model plants is summed by region and model plant type:

$$Capacity_{R,T,v} = \sum_{\substack{c \in R \\ u \in T}} (Capacity_{c,u,v})$$

All other attributes of model plants are computed as an average over sets of countries (= model region) and units (= model plant), weighted by unit capacity:

$$Attribute_{R,T,v} = \frac{\sum_{\substack{c \in R \\ u \in T}} (Attribute_{c,u,v} \times Capacity_{c,u,v})}{\sum_{\substack{c \in R \\ u \in T}} Capacity_{c,u,v}}$$

Composing model regions from countries also requires setting up transmission links between model regions. For each pair of model regions (R1, R2), the transmission capacity should be computed by summing the transmission capacities between all countries in each region:

$$Capacity_{R1,R2,v} = \sum_{\substack{c1 \in R1 \\ c2 \in R2}} (Capacity_{c1,c2,v})$$

All other attributes for transmission technologies should be computed using weighted averages:

$$Attribute_{R1,R2,v} = \frac{\sum_{\substack{c1 \in R1 \\ c2 \in R2}} (Attribute_{c1,c2,v} \times Capacity_{c1,c2,v})}{\sum_{\substack{c1 \in R1 \\ c2 \in R2}} Capacity_{c1,c2,v}}$$

These equations imply that the transmission capacity between countries that lie in the same model region will be ignored. As a result, the entire generation capacity of that region will be available to meet the entire demand, irrespective of the transmission bottlenecks that will exist in reality. This is the main drawback of having large model regions.

Regional renewable potential should be aggregated in the same way as the existing and new plants, with one important difference. A "haircut" should be taken on the total potential at the country level corresponding to some fraction of its load, so that aggregation does not result in unrealistic total potential for a model region. For example, the huge hydro and wind potential of Canada should be bounded by its own total (projected) consumption so that it is not unrealistically available to the USA, if Canada and USA were aggregated into a single model region. This bounding can be guided by detailed model runs, if desired. To guide this haircut, historical GDP and electricity consumption trends can be used to project electricity intensity of GDP at the country level and projected values of GDP applied to these intensities to project total electricity consumption.

Attributes related to end-use price calculations should be aggregated as averages of country-level data weighted by country load within the region:

$$Attribute_{R,v} = \frac{\sum_{c \in R} (Attribute_{c,v} \times Load_{c,v})}{\sum_{c \in R} Load_{c,v}}$$

5.2.3. Timeslice Aggregation

Tables 18 and 19 show two alternative mappings of 8760 annual hours to model timeslices. Each month, day of the week, and hour is mapped to a season, day type, and time of day. Table 18 shows a finer mapping, in which each month is its own season, weekdays and weekends are separated, and four times of day are used, while Table 19 shows a coarser mapping, with four seasons and three times of day.

Table 18 - Mapping hours of a year to model timeslices – Fine version

Season		Weekly		DayNite	
01	Jan	1	E	01	N
02	Feb	2	W	02	N
03	Mar	3	W	03	N
04	Apr	4	W	04	N
05	May	5	W	05	N
06	Jun	6	W	06	D
07	Jul	7	E	07	D
08	Aug			08	D
09	Sep			09	D
10	Oct			10	D
11	Nov			11	D
12	Dec			12	D
				13	D
				14	D
				15	D
				16	D
				17	Q
				18	Q
				19	P
				20	P
				21	Q
				22	N
				23	N
				24	N

Table 19 - Mapping hours of a year to model timeslices – Coarse version

Season		Weekly		DayNite	
01	Winter	1	W	01	N
02	Winter	2	W	02	N
03	Spring	3	W	03	N
04	Spring	4	W	04	N
05	Spring	5	W	05	N
06	Summer	6	W	06	D
07	Summer	7	W	07	D
08	Summer			08	D
09	Fall			09	D
10	Fall			10	D
11	Fall			11	D
12	Winter			12	D
				13	D
				14	D
				15	D
				16	D
				17	D

Season			Weekly			DayNite	
						18	D
						19	P
						20	P
						21	N
						22	N
						23	N
						24	N

Figures 7 through 9 illustrate the use of these mappings to aggregate the historical load curve of Romania (Figure 7) into the finer (Figure 8) and coarser (Figure 9) aggregations.

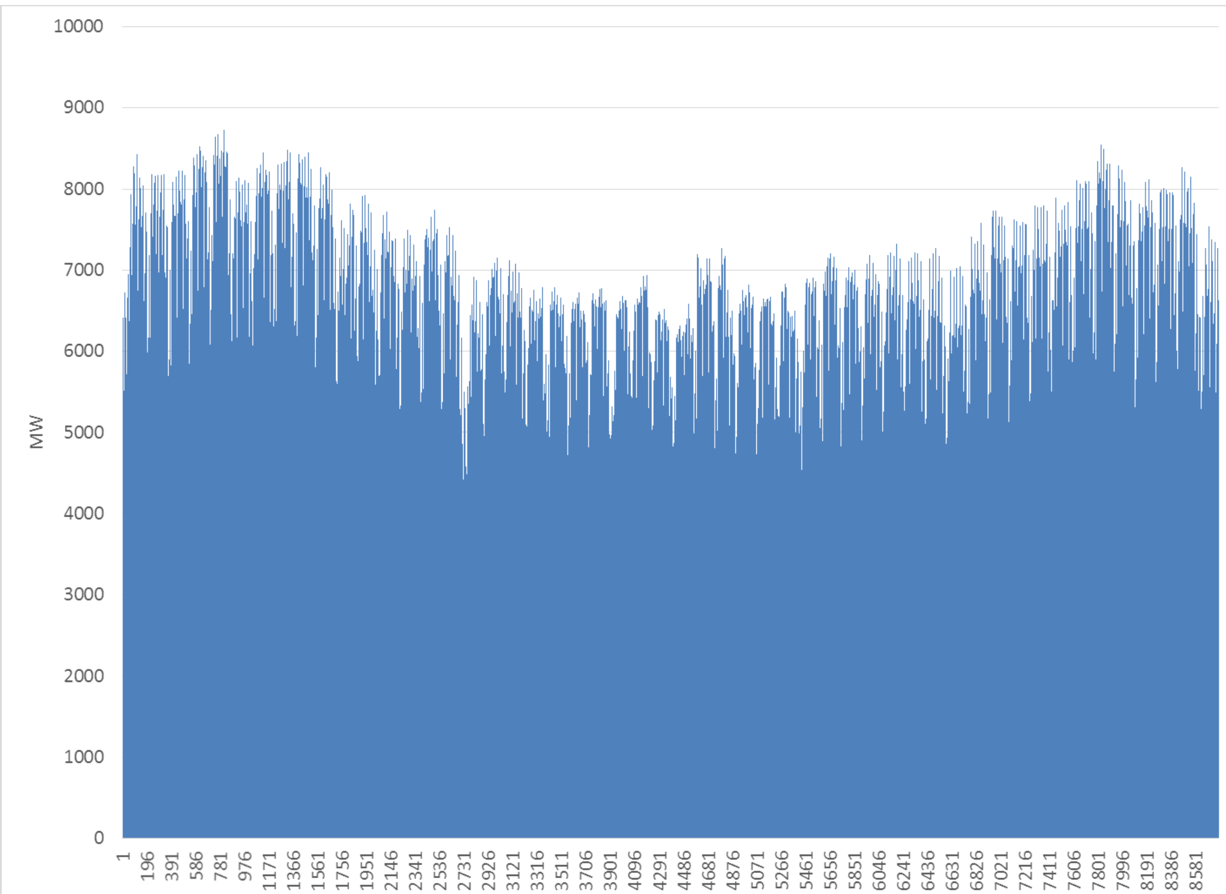


Figure 7 - Hourly load curve for Romania for year 2011 (Source: ENTSO-E consumption data)

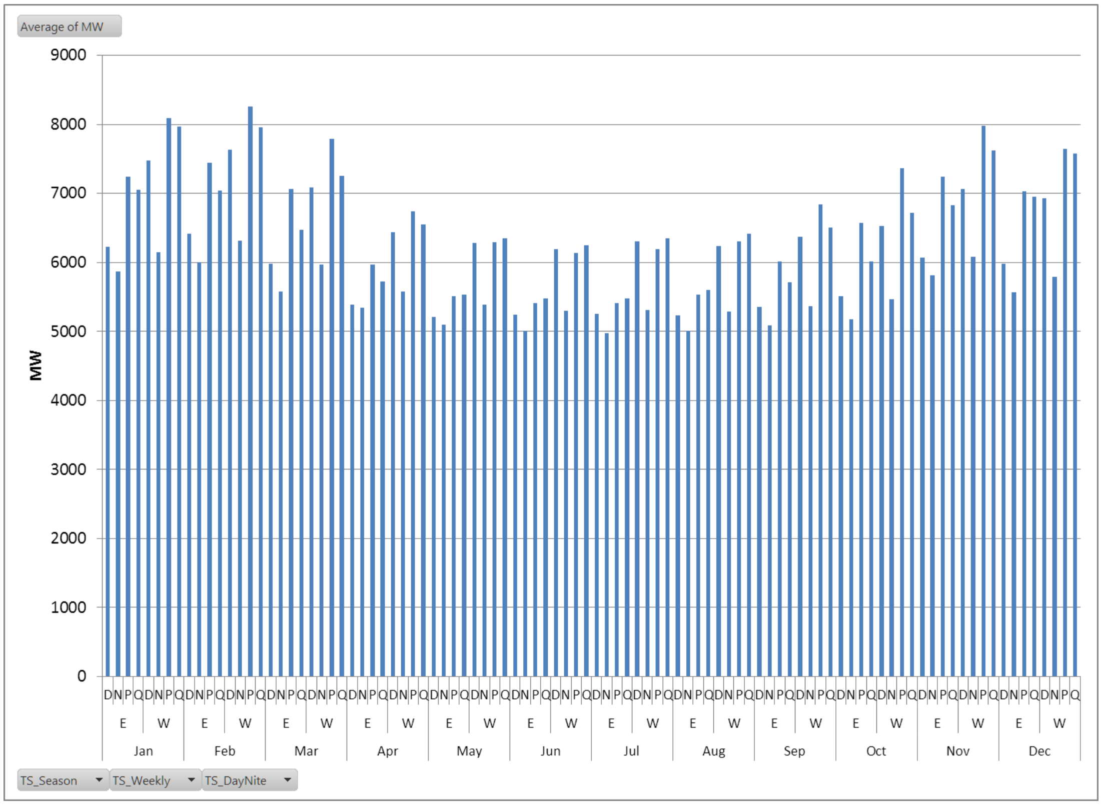


Figure 8 - Detailed timslice definition - 12 seasons, 2 weekly and 4 day-night divisions (96 timeslices)

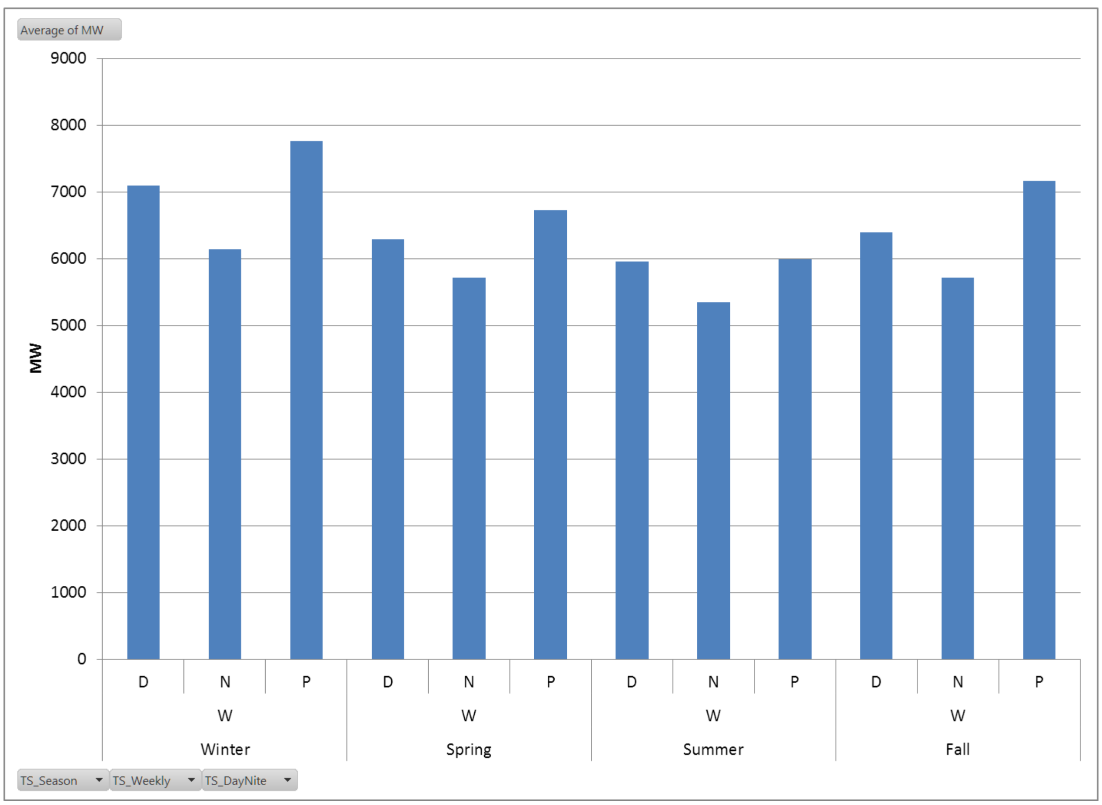


Figure 9 - Coarse timeslice definition - 4 seasons and 3 divisions at day-night level (12 timeslices)

Power demand for each country by model timeslice should be computed by averaging the demand in the included timeslices:

$$MW_{c,TS} = Average_{h \in TS}\left(MW_{c,h}\right)$$

Where ***h*** = hours in a year.

The demand fraction for each model region (R) and model timeslice (TS) is calculated as:

$$DemFraction_{R,TS} = \frac{\sum_{c \in R} MW_{c,TS} \times Duration_{c,TS}}{\sum_{\substack{\forall TS \\ c \in R}} MW_{c,TS} \times Duration_{c,TS}}$$

The denominator in the equation above is the total annual electricity demand for the model region in reference year.

The same approach will be used to migrate the granular capacity factors of Renewable Energy sources that are stored in the input database.

The peak reserve factor, which will create demand for capacity to meet the peak electricity (or heat, if needed in a future version) demand over and above what is needed to meet the energy flow in peak timeslice should be computed from the aggregation as the ratio of the load in the peak hour to the load in the peak timeslice, across the region:

$$PeakReserveFactor_{R} = \frac{\max_{h} \sum_{c \in R} MW_{c,h}}{\max_{TS} \sum_{c \in R} MW_{c,TS}}$$

This value is added to the reserve requirement used by the transmission system operator.

The transmission efficiency, which covers both transmission and distribution losses, show be computed at the model region level by taking a weighted average of the projected values of country-wise transmission efficiency values for each year (y):

$$TransmissionEFF_{R,y} = \frac{\sum_{c \in R} TransmissionEFF_{c,y} \times TotalElecConsumption_{c,y}}{\sum_{c \in R} TotalElecConsumption_{c,y}}$$

Where $\text{TotalElecConsumption}_{c,y}$ are the projected values that were used to compute the "haircut" on total RE potential by model region.

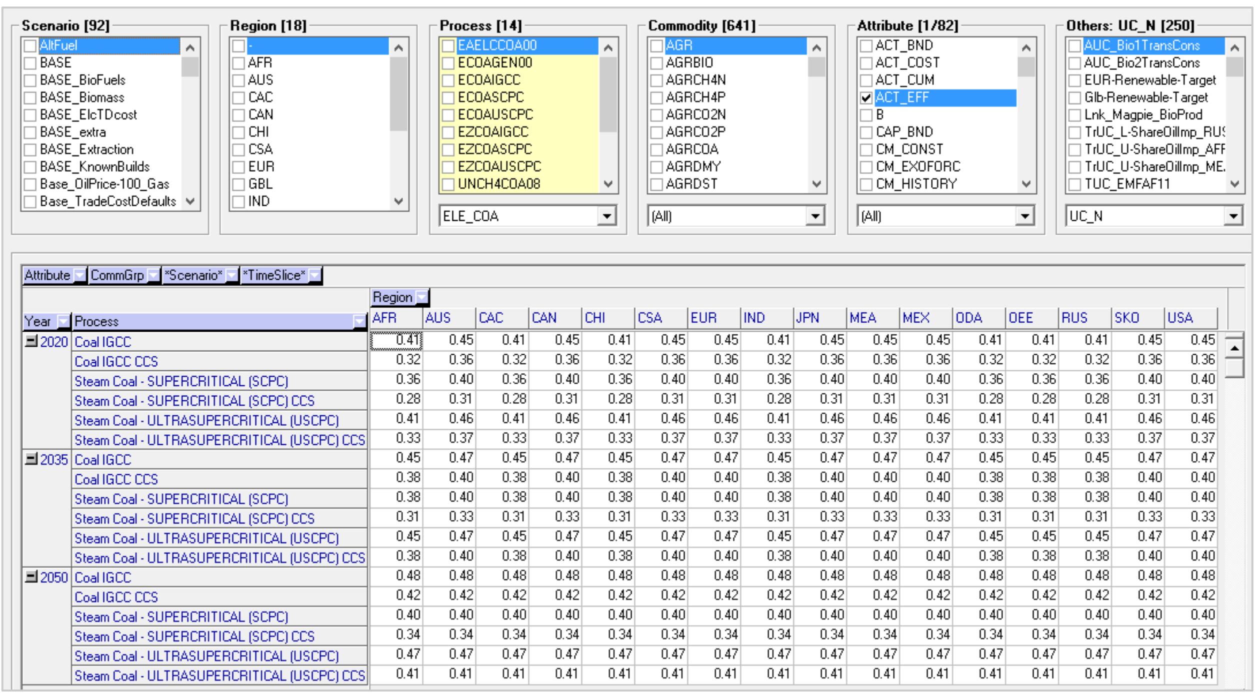
5.3. Model Front End

So far, we have described the storage of data at granular level and the process of transforming it to create a particular model instance. We have proposed several data inputs, which are based on systematic assumptions and inheritance, to be executed via automated processing of rules. To verify and understand the model instance thus created, we need a facility to browse it. Then, we need a way to generate the specifications needed to model various policy targets and measures, along with different state-of-the-world scenarios (such as availability and costs of resources and technologies). Finally, we need a way to combine these pieces of model instance data and scenario assumptions into cases and launch them as model runs. The model input data and run management system, or front end, accomplishes these tasks.

5.3.1. Browsing Model Instance Data

The front end should enable the user to browse all model input data, both from model instances and from any scenario files created as described later in this section. Data should be searchable by scenario, region, process, commodity, attribute, and other dimensions (user-constructed constraints shown in Figures 10 and 11).

Data views should be pivotable, so that the user can easily show, hide, and reorganize dimensions of interest, as shown in Figures 10 and 11. Figure 10 illustrates a data browse configured to show the efficiency of all coal using model plants, across regions and years, while Figure 11 pivots the data another way and shows all attributes of two coal plants.



Year	Process	AFR	AUS	CAC	CAN	CHI	CSA	EUR	IND	JPN	MEA	MEX	ODA	OEE	RUS	SKO	USA
2020	Coal IGCC	0.41	0.45	0.41	0.45	0.41	0.45	0.45	0.41	0.45	0.45	0.45	0.41	0.41	0.41	0.45	0.45
	Coal IGCC CCS	0.32	0.36	0.32	0.36	0.32	0.36	0.36	0.32	0.36	0.36	0.36	0.32	0.32	0.32	0.36	0.36
	Steam Coal - SUPERCRITICAL (SCPC)	0.36	0.40	0.36	0.40	0.36	0.40	0.40	0.36	0.40	0.40	0.40	0.36	0.36	0.36	0.40	0.40
	Steam Coal - SUPERCRITICAL (SCPC) CCS	0.28	0.31	0.28	0.31	0.28	0.31	0.31	0.28	0.31	0.31	0.31	0.28	0.28	0.28	0.31	0.31
	Steam Coal - ULTRASUPERCRITICAL (USCPC)	0.41	0.46	0.41	0.46	0.41	0.46	0.46	0.41	0.46	0.46	0.46	0.41	0.41	0.41	0.46	0.46
	Steam Coal - ULTRASUPERCRITICAL (USCPC) CCS	0.33	0.37	0.33	0.37	0.33	0.37	0.37	0.33	0.37	0.37	0.37	0.33	0.33	0.33	0.37	0.37
2035	Coal IGCC	0.45	0.47	0.45	0.47	0.45	0.47	0.47	0.45	0.47	0.47	0.47	0.45	0.45	0.45	0.47	0.47
	Coal IGCC CCS	0.38	0.40	0.38	0.40	0.38	0.40	0.40	0.38	0.40	0.40	0.40	0.38	0.38	0.38	0.40	0.40
	Steam Coal - SUPERCRITICAL (SCPC)	0.38	0.40	0.38	0.40	0.38	0.40	0.40	0.38	0.40	0.40	0.40	0.38	0.38	0.38	0.40	0.40
	Steam Coal - SUPERCRITICAL (SCPC) CCS	0.31	0.33	0.31	0.33	0.31	0.33	0.33	0.31	0.33	0.33	0.33	0.31	0.31	0.31	0.33	0.33
	Steam Coal - ULTRASUPERCRITICAL (USCPC)	0.45	0.47	0.45	0.47	0.45	0.47	0.47	0.45	0.47	0.47	0.47	0.45	0.45	0.45	0.47	0.47
	Steam Coal - ULTRASUPERCRITICAL (USCPC) CCS	0.38	0.40	0.38	0.40	0.38	0.40	0.40	0.38	0.40	0.40	0.40	0.38	0.38	0.38	0.40	0.40
2050	Coal IGCC	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
	Coal IGCC CCS	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
	Steam Coal - SUPERCRITICAL (SCPC)	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
	Steam Coal - SUPERCRITICAL (SCPC) CCS	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
	Steam Coal - ULTRASUPERCRITICAL (USCPC)	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47
	Steam Coal - ULTRASUPERCRITICAL (USCPC) CCS	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41

Figure 10 - Thermal efficiency of all coal plants by model region and year

To enable the user to fully understand the model energy system and to facilitate debugging, it is essential that the data browser be model topology-aware and be able to step through model topology via process inputs and outputs. Figure 12 shows an example of this functionality. The view is centered on Centralized Electricity. Columns to left and right show all processes that process this commodity (left) and consume it (right), shown as the labeled boxes. Further to left and right are the inputs and outputs to these processes, shown as the labeled lines. The user should be able to click on any of these inputs and outputs and jump to a view of that process/commodity's inputs/outputs, and/or switch back to the data browser to inspect the data view for the item of interest.

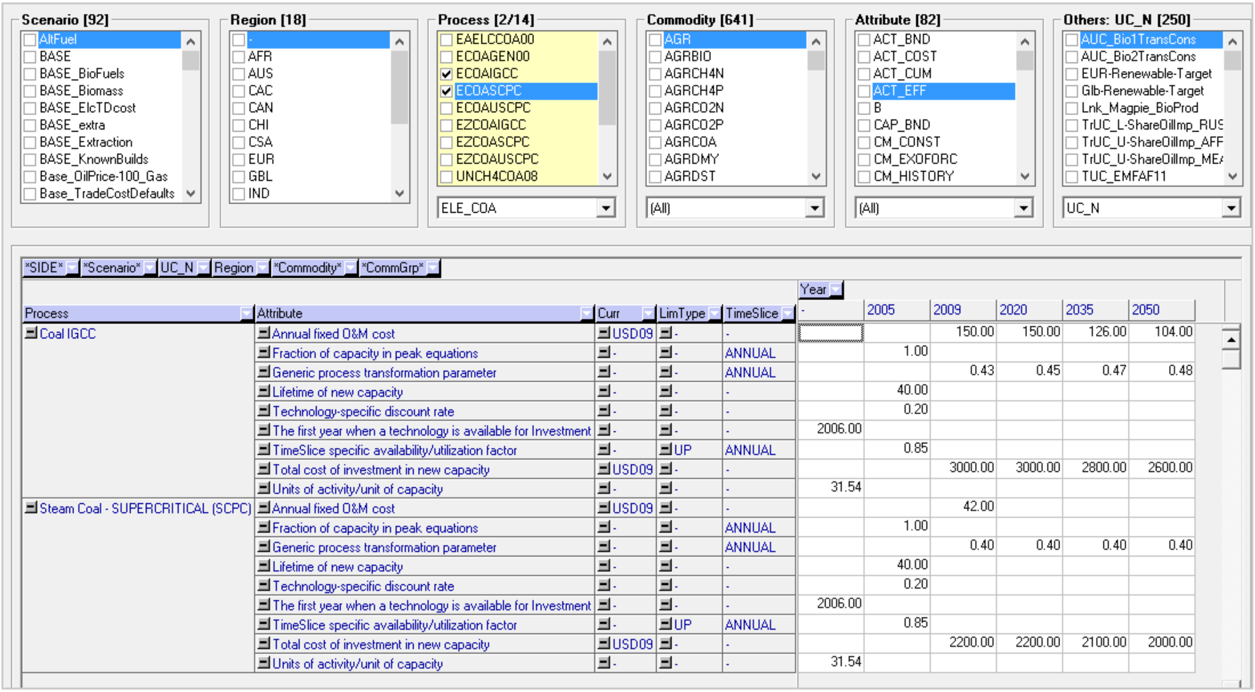


Figure 11 - All attributes of two coal plants

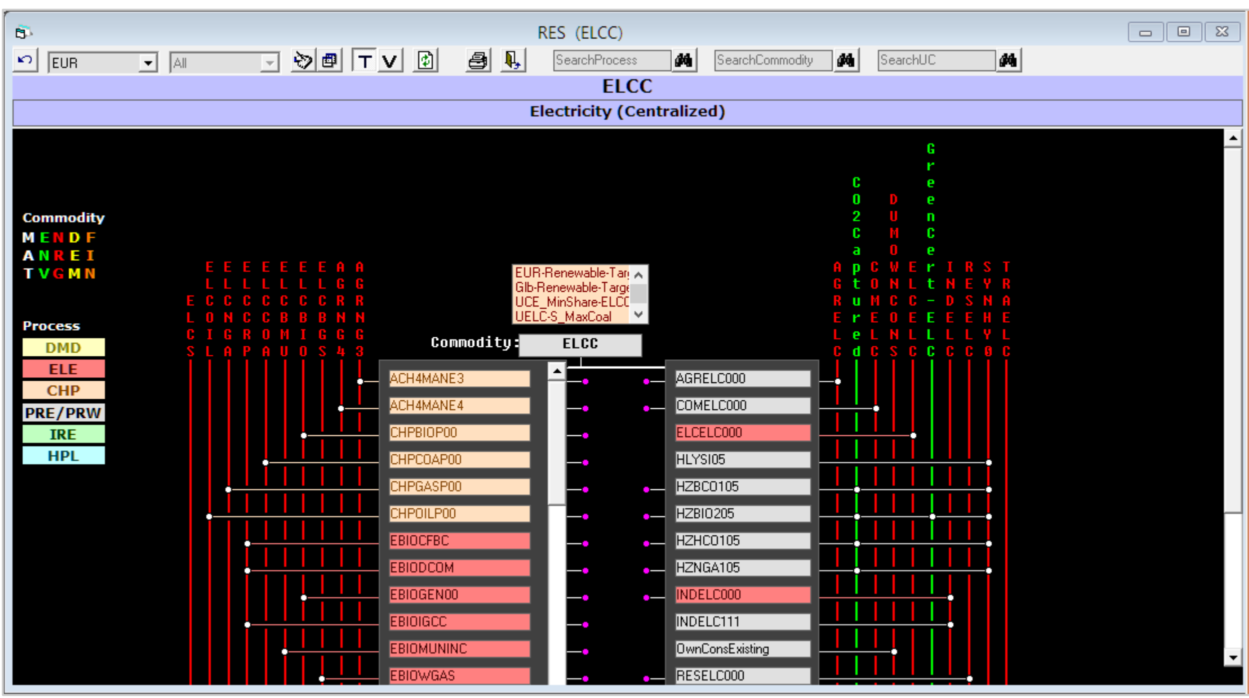


Figure 12 - The Reference Energy System view showing all sources and uses of centralized electricity (along with the inputs of sources and outputs of consumers)

5.3.2. Scenario Specification

Let us consider generating scenario specifications for the following use cases:

1. Prohibit further investment in nuclear power in certain model regions
2. Impose a maximum limit on nuclear capacity in certain model regions
3. Apply different discount rates, by technology and region
4. Create a renewable portfolio standard
5. Impose a CO2 intensity limit
6. Model CCS retrofit of existing coal plants

Above all, it is crucial to be able to work with sets of processes in order to generate scenario specifications efficiently. For example, it should be possible to read the Excel worksheet shown below to create sets of processes.

Table 20 - Rule-based creation of sets of processes

SetName	Tech Name	Tech Desc	Input Commodity	Output Commodity	Comments
NewCoal		*New*	Coal	Electricity	all techs that consume "Coal" and have "New" in their description
NewNuc		*New*	NucFuel	Electricity	all techs that consume Nuclear fuel and have "New" in their description
Elec-RE			Hydro,Geo, Wind,Solar	Electricity	all hydro, geothermal, wind and solar electricity plants
Elec-All				Electricity	all electricity plants
Elec-Fossil			Coal,Gas,Oil	Electricity	all fossil fuel plants
Elec-CCS				Electricity AND Captured CO2	All CCS plants

Different scenario specifications should be created by processing Excel files with high-level rules specification, as described through the use cases below. Transparency and efficiency are the main advantages of this approach, over fully enumerated data specification. Note that depending upon the number of aggregated regions and model plants, each of these small files would create hundreds or even thousands of data values.

Model formulation must support creation of such constraints dynamically, based only on data declarations. In essence, there should be parameters to create links among capacities/flows of different technologies, or with total flows of commodities, by region and period.

Use case 1: Prohibit further investment in nuclear power in certain model regions

This can be achieved by declaring an attribute <Bound on New Capacity> = 0 for each new nuclear technology in the concerned model regions. This should be possible by processing a rule declared in an Excel file, named *Scen_NoNewNuc*, for example, using the simple table shown in Table 21.

Table 21 - Excel table to prohibit new nuclear investment in six regions

Attribute	LimType	Process Set	Years	MR16	MR07	MR01	MR16	MR05	MR03
Bound on New Capacity	UP	NewNuc	2010-2100	0	0	0	0	0	0

Use case 2: Impose a maximum limit on nuclear capacity in certain model regions

Let us assume that we need to set to different limits for different time spans, for some of the model regions. Here we add an additional row to the new Excel file shown in Table 22, *Scen_MaxNewNuc*, specifying the relevant time range. The region columns now show different bound values in each region.

Table 22 - Excel table to impose bound on new nuclear capacity

Attribute	LimType	Process Set	Years	MR16	MR07	MR01	MR06	MR03
Bound on New Capacity	UP	NewNuc	2010-2050	20	25	30	35	40
Bound on New Capacity	UP	NewNuc	2051-2100	30	35	40	45	50

Use case 3: Apply different discount rates, by technology and region

Let us say we have to declare specific high values for technology discount rates for all new CCS technologies in some regions, and 10% in all other regions. This should be possible by processing a rule declared in an Excel file, *Scen_CCSDiscRate*, as shown in Table 23.

Table 23 - Excel table to process-specific discount rates

Attribute	Process Set	Years	AllRegions	MR16	MR07	MR01	MR06	MR03
DiscountRate	Elec-CCS	AllYears	0.1	0.15	0.2	0.15	0.2	0.2

Use case 4: Create a renewable portfolio standard

Assume that we have to model a renewable portfolio standard as a policy target. So, we need to impose the following constraint in each model region and period:

(Electricity generation from RE sources) >= Target% x (Total Electricity generation)

Further, we may have different targets for 2025 and 2050, and different targets for some regions, with a default value of the target for the remaining regions.

All this should be possible to accomplish by processing the table shown in Figure 13, in an Excel file called *Scen_RPS*, for example. The default targets are 20% in 2025 and 30% in 2050. (The target percentage values are shown with a minus sign, because for simplicity of processing we bring that term over to the left hand side of the inequality, making the right hand side 0.)

ConstraintName	Attribute	Process Set	Commodity	LimType	Years	AllRegions	MR16	MR07	MR01	MR06	MR03
RenPortStandard	Production	Elec-RE	Electricity		2025	1					
	Tot Production		Electricity		2025	-0.2	-0.25	-0.3	-0.25	-0.22	-0.35
	Production	Elec-RE	Electricity		2050	1					
	Tot Production		Electricity		2050	-0.3	-0.35	-0.4	-0.35	-0.32	-0.45
	RightHand Side			LO	2025-2050	0					

Figure 13 - Excel table to create a renewable portfolio standard

Use case 5: Impose a CO_2 intensity limit

Assume that we have to model a CO_2 emission intensity limit as a policy target, by model region and periods. So, we need to impose the following constraint in each model region and period:

(CO_2 emission from electricity generation) <= RateTarget% x (Total generation)

Further, we may have different targets for 2025 and 2050, and different targets for some regions, with a default value of the target for the remaining regions.

All this should be possible to accomplish by processing the table shown in Figure 14, in an Excel file called *Scen_CO2IntensityLimit*, for example. The default targets are 850 ktCO2/Twh in 2025 and 680 in 2050 in the example shown below.

ConstraintName	Attribute	Process Set	Commodity	LimType	Years	AllRegions	MR16	MR07	MR01	MR06	MR03
MaxCO2IntOfElec	Production	Elec-All	CO2		2025	1					
	Tot Production		Electricity		2025	-850	-757	-831	-781	-822	-833
	Production	Elec-All	CO2		2050	1					
	Tot Production		Electricity		2050	-680	-605	-665	-624	-658	-666
	RightHand Side			UP	2025-2050	0					

Figure 14 - Excel table to impose a CO_2 intensity limit

Use case 6: CCS retrofit of existing coal plants

To model retrofitting existing coal plants with carbon capture equipment, we can create a new technology, E_CCSRetrofit, with incremental investment cost and all other parameters (efficiency, emission coefficients and operational costs) of the retrofitted plant. If endogenous retirement is enabled on the host plant – E_HostPlant, then the following equation, in each year, will make sure that the model is able to retrofit existing capacity:

(Capacity of host plant) + (Capacity of CCS retrofit) <= <original capacity of host plant>

It should be possible to create this equation by processing the declarations shown below.

Table 24 - Excel declarations to model CCS retrofit of existing coal plants

ConstraintName	Attribute	TechName	LimType	Years	AllRegions
CCSRetrofitCoal	Capacity	E_HostPlant		2025 onward	1
	Capacity	E_CCSRetrofit			1
	RightHandSide		UP		<Cap of host plant>

5.3.3. Case Manager

Each of these scenario files essentially creates a "data layer", which one should be able to overlay on the core model instances. Figures 15 and 16 show two examples of a case manager form in which the user may choose which model instance and scenario files are to be used in a particular model run. Figure 15 shows a case in which a model instance based on NREL renewable potential data is combined with the No New Nuclear, Discount Rate, and RPS scenarios, and Figure 16 shows a case in which an IRENA-based model instance is combined with the Maximum Nuclear Capacity, Discount Rate, and CO_2 Intensity Limit scenarios.

	ModelCase1
X	Model Instance with NREL RE data
	Model Instance with IRENA RE data
X	Scen_NoNewNuc
	Scen_MaxNewNuc
X	Scen_CCSDiscRate
X	Scen_RPS
	Scen_CO2IntensityLimit

Figure 15 - Example model case generation form #1

	ModelCase2
	Model Instance with NREL RE data
X	Model Instance with IRENA RE data
	Scen_NoNewNuc
X	Scen_MaxNewNuc
X	Scen_CCSDiscRate
	Scen_RPS
X	Scen_CO2IntensityLimit

Figure 16 - Example model case generation form #2

The case manager should allow the user to name, store, manage, and modify these cases for later use. It should also provide a facility to allow the user to select scenario files to be automatically combined in a combinatorial fashion to systematically explore the uncertainty and policy design space under study.

The front end writes data from each of the data layers in the case manager to separate text files, a list of which should be submitted to the model code via a simple script. After each model run, a script will be launched that will process the output file as per rules declared in an Excel file, which will compile the input needed by WEPS+ and prepare output data for the results data processing system described in Section 5.5.

5.4. Optimization Model

For Tier 4, the core of the model itself, we recommend an integrated dispatch and capacity expansion model based on economic optimization, using linear programming wherever possible to reduce model size. The optimization is subject to operational, system, and policy constraints, including constraints that are core to the model equations and others written by the user. This section[12] describes the model topology, variables, objective function and constraints (also called "equations", although they are generally inequalities, rather than equalities.)

5.4.1. Model Topology and Data Specification

Our design principle #3 is that the model must be data-driven. That is, the model framework must consist of intelligently-connected building blocks that can be readily filled with data by using simple Excel table and rule-based specifications, enabling many different electricity systems consisting of different numbers of regions, model plants, and time-slices to be constructed within the same model structure, using only different data preparations.

The fundamental construct of a data-driven model is the model topology, constructed via three types of building blocks:

- *Technologies* (also called *processes*) are representations of physical power plants, transmission lines, or other devices that transform some commodities into other commodities, changing their form or their location. For example, a power plant changes its input fuel(s) into electricity, and a transmission line changes electricity in one location to electricity in another. Dummy processes may also be used to change the names of commodities, track commodities for scenario analysis purposes, or directly satisfy the final demand for electricity in a region.
- *Commodities* connect processes in the model topology. A commodity is produced by some process(es) and/or consumed by other process(es). They may be of several different types, including: energy carriers, such as fuels and electricity/heat; energy services, such as lighting or space heating, in a model representing end use service demand detail; materials; monetary flows; and emissions.
- Commodity *flows* are the links between processes and commodities. A flow is of the same nature as a commodity but is attached to a particular process, and represents and tracks one input or one output of that process. For instance, electricity produced by wind turbine type A at period p, time-slice s, in region r, is a commodity flow.

Each technology, commodity, and flow in the model must be specifiable by standard data parameters. Together, these building blocks can then be connected by data specification into an energy system network of arbitrary size and detail, simply through data provision to the model. As discussed in Section 5.1, the model code should also be able to accept data specifications in any year and perform interpolation/extrapolation of model parameters according to rules provided with each parameter. In particular, the model should accept data on investments in technology capacity that occur before the beginning of the model horizon, tracking this capacity identically to capacity invested in within the model horizon, in order to eliminate any difficulties in updating model base year. The model code must also be able accept data dictionaries in the form prepared by the front end.

---

[12] This section draws heavily on Loulou et al. (2005) and uses the TIMES nomenclature and notation. TIMES offers additional features not described herein. We have described the minimum capabilities necessary to meet the IEMM performance objectives and the design principles identified in Section 2. Other frameworks may describe these variables and equations differently, but should provide similar functionality.

Using this data-driven approach permits new model structures to be created simply by declaring the relevant items and providing their input data. For example, although the current specifications call for the IEMM to receive fixed sectoral electricity demands from WEPS+, EIA may wish at some point to conduct analyses of end-use efficiency for reducing these demands. Efficiency supply curves could be built into the model by declaring and providing data for "dummy" power plants that are capable of satisfying (avoiding) a given percent of electricity demand each year at a specified capital and/or variable cost. These efficiency measures would then compete with model power plants to meet the demands provided by WEPS+. No changes to the model code would be required, only changes to the input data.

Similarly, depictions of intra-country fuel supply and electricity transmission infrastructure could be created for scenarios that wish to resolve that level of detail, again simply by providing the necessary data to describe the technologies, commodities, and flows involved.

5.4.2. Variables

In the description of model variables and equations that follows:

- Model variables are prefixed by ***VAR_***
- Model equations (constraints) are prefixed by ***EQ_***
- The following indices are used:
  - ***r:*** region
  - ***t:*** the current time period
  - ***v:*** the vintage year of a technology investment
  - ***p:*** process (or technology, for example, a model power plant)
  - ***s:*** time-slice
  - ***c:*** commodity (for example, fuel, electricity, or emission).

The key decision variables needed are:

VAR_NCAP(r,v,p)**:** New capacity addition (investment) for technology ***p***, in period ***v*** and region ***r***. Tracking the vintage year ***v***, in which a technology was invested in enables, for example, the efficiency of a new wind plant to change over time with exogenous technology learning. The vintage index is optional. If input parameters for a technology do not change over time, the technology does not need to be vintaged.

VAR_RCAP(r,v,t,p): Amount of capacity of technology ***p***, vintage ***v*** in region ***r*** that is newly retired at period ***t*** . The new retirements will reduce the available capacity of vintage ***v*** in period ***t*** and in all successive periods $t_i > t$ by the value of the variable.

VAR_CAP(r,v,t,p): Total installed capacity of process ***p***, in region ***r*** and period ***t,*** optionally with vintage ***v***, considering the residual capacity at the beginning of the modeling horizon, adding new investments made prior to and including period ***t*** that have not reached their technical lifetime, and subtracting retired capacity.

VAR_FLO(r,v,t,p,c,s): The quantity of commodity ***c*** consumed or produced by process ***p***, in region ***r*** and period ***t*** (optionally with vintage ***v*** and time-slice ***s***). Power plant related flows would include fuels consumed and electricity and emissions produced.

VAR_ACT(r,v,t,p,s): Activity level of technology ***p*** with vintage ***v***, in region ***r***, period ***t*** and time-slice ***s*)**. The activity is a user-designated sum of one or more input or output flows from a technology. (For example, for a power plant, the activity is the generation of electricity.) The activity is limited by the available capacity via the Use of Capacity equation. The activity variable also determines variable operation and maintenance costs.

VAR_SIN(r,v,t,p,c,s)/VAR_SOUT(r,v,t,p,c,s): if electricity storage processes are used, the quantity of commodity ***c*** stored or discharged by storage process ***p,*** in time-slice ***s***, period ***t*** (optionally with vintage ***v)***, and region ***r***.

VAR_TRADE(r,v,t,p,c,s,exp) and VAR_TRADE(r,v,t,p,c,s,imp): quantity of commodity ***c*** sold (***exp***) or purchased (***imp***) by region ***r*** through trade process ***p*** in period ***t*** and time-slice ***s***). Note that the topology defined for the exchange process ***p*** specifies the traded commodity ***c***, the region ***r***, as well as the region ***r'*** with which region ***r*** is trading commodity ***c***. These variables are used to describe flows across interregional transmission lines, for example.

VAR_DEM(r,t,d): demand for commodity ***d*** in region ***r*** and period ***t.*** For the IEMM, these commodities correspond to the sector electricity loads.

Additional commodity related variables should be produced for reporting purposes and to create handles for applying bounds, including: ***VAR_COMPRD(r,t,c,s)***, ***VAR_COMCON(r,t,c,s)***, and ***VAR_COMNET(r,t,c,s)***, the total amount of commodity ***c*** produced, consumed, and net, respectively, in region ***r***, at time period ***t***, in timeslice ***s***.

5.4.3. Objective Function

The model's objective function is the minimization of the total discounted system cost of meeting the electricity load defined in Section 4.1. The components of this cost include:

- *Capital Costs* for investment in new capacity. These costs should be annualized over the unit's economic life, at either the system discount rate or a technology-specific discount rate imposed by the user;
- Fixed and variable annual *Operation and Maintenance (O&M) Costs*;
- Costs for *fuel* consumed, including any *delivery costs* the user may wish to impose on top of the regional fuel price;
- Any *taxes* and *subsidies* the user wishes to impose; minus
- *The salvage value* of capacity with remaining economic life.

These costs reflect the minimum set of costs that should be included in the model objective function. It is also possible to include other costs, such as:

- *Decommissioning costs*, which may be capital expenditures and/or annual operation and maintenance costs, for capacity that must incur costs for dismantling at the end of its useful life;
- *Damage costs* from pollutant emissions, for example from acid rain or health impacts; and
- *Welfare loss/gain* resulting from changes in quantity of electricity demanded from a reference scenario.

The final option is used when the IEMM has been configured to allow the quantity of electricity demanded in each region to be elastic to price, rather than fixed at the value received from WEPS+. As discussed in Section 4.4, under this approach, a piecewise linearized demand curve is used to drive the model, in place of a fixed demand. In conjunction with price elastic fuel supply curves, this option would enable the IEMM to be run in a standalone mode for scenario exploration and may reduce iteration time within WEPS+. In this case, the minimization of system costs, including welfare gains/losses, is equivalent to maximizing total consumer plus producer surplus.[13]

5.4.4. Model Constraints

This optimization is then subject to the following constraints, which impose physical, operational, and policy restrictions:

Capacity Transfer

When computing the available capacity in some time period, the model takes into account the capacity resulting from all investments up to that period, some of which may have been made prior to the initial period but are still in operating condition (embodied by the residual capacity of the technology), and others that have been decided by the model at, or after, the initial period, up to and including the period in question. Investing in a particular technology increases its installed capacity for the duration of the technical life of the technology. At the end of that life, or when the technology is retired, the total capacity for this technology is decreased by the same amount.

The total available capacity for each technology *p,* in region *r*, in period *t* (all vintages), is thus given by:

EQ_CPT(r,t,p) - Capacity transfer

> ***VAR_CAP(r,t,p)*** *=* ***Sum*** *{over all periods t' preceding or equal to t such that t-t'<LIFE(r,t',p) of* ***VAR_NCAP(r,t',p)****} + PASTI(r,t,p) -* ***VAR_RCAP(r,v,t,p)****}*
>
> where:
> *LIFE* is the technical lifetime of technology *p,* and
> *PASTI(r,t,p)* is the (exogenously provided) investments in technology *p* made prior to the initial model period and still exist in region *r* at time *t*.

Use of capacity

In each time period the model may use some or all of the installed capacity according to the Availability Factor of that technology, which specifies a maximum usage. (The option to provide a BOUND to force specific technologies to use their capacity to any minimum level of operation, including full potential is discussed in Section 5.4.5.)

---

[13] See for example, Loulou, R., Remne, U., Kanudia, A., Lehtila, A., Goldstein, G., 2005. *Documentation for the TIMES Model Part I*, Energy Technology Systems Analysis Programme, http://www.iea-etsap.org/web/Docs/TIMESDoc-Intro.pdf, Section 3.2.3.2, for an explanation of the equivalence.

For each technology ***p*** and vintage ***v***, in region ***r***, in period ***t*** and time-slice ***s***, the activity of the technology may not exceed its available capacity, as specified by a user defined availability factor:

EQ_CAPACT (r,v,t,p,s) - Use of capacity

VAR_ACT (r,v,t,p,s) ≤

AF(r,v,t,p,s) CAP2ACT(r,p))*FR(r,s)* **** VAR_CAP(r,v,t,p)***

where:

CAP2ACT(r,p) is the conversion factor between units of capacity and activity (often equal to 1. For power plants with capacities in GW and activities in TBTU, CAP2ACT = 29.889.)

FR(r,s) is equal to the (fractional) duration of timeslice s

AF is the availability factor.

Defining flow relationships in a process

A process may have one or more input and output flow variables. We therefore require one or more constraints relating the sum of some of its output flows to the sum of some of its input flows is equal to a constant. In the case of a single commodity in, and a single commodity out of a process, this equation defines the traditional efficiency of the process. With several commodities, this constraint may leave some freedom to individual input (or output) flows, as long as their sum is in fixed proportion to the sum of output (or input) flows. For example, we may have a fuel flexible power plant that can burn different combinations of its applicable fuels.

Because units may have additional inputs and outputs unrelated to the relationship to be described by a given constraint, the modeler must identify which commodities the constraint relates. These are known as *commodity groups.* The equation below relates an input commodity group *cg*1 to an output commodity group *cg*2, with an overall efficiency ratio named *FLO_FUNC(p,cg1,cg2)* and commodity-specific efficiency factors *CEFF(r,v,t,p,c,s)*:

EQ_PTRANS(r,v,t,p,cg1,cg2,s) **– Efficiency definition**

$$\textbf{\textit{SUM}}\ \{c \text{ in } cg2 \text{ of} : \textbf{\textit{VAR\_FLO(r,v,t,p,c,s)}} * \textbf{\textit{CEFF(r,v,t,p,c,s)}}\} =$$

$$FLO\_FUNC(r,v,cg1,cg2,s) *$$

$$\textbf{\textit{SUM}}\ \{c \text{ within } cg1 \text{ of}: \textbf{\textit{VAR\_FLO(r,v,t,p,c,s)}} / \textbf{\textit{CEFF(r,v,t,p,c,s)}}\}$$

where ***CEFF*** may default to 1 if not otherwise specified.

When either of the commodity groups *cg*1 or *cg*2 contains more than one element, the previous constraint allows freedom on the values of flows. A flow share constraint can be used to limit the flexibility, by constraining each flow within its own group to an upper, lower, or fixed share, as follows:

EQ_INSHR(c,cg,p,r,t,s) **and** ***EQ_OUTSHR(c,cg,p,r,t,s)*** **– Flow share constraints**

VAR_FLO (c) $\leq, \geq, =$

FLO_SHAR(c) * *Sum {over all c' in cg of:* ***VAR_FLO (c')*** *}*

The commodity group *cg* may be on the input or output side of the process.

Commodity Balance Equation

In each time period and time-slice, the production of a commodity within a region plus any imports from other regions must balance the amount consumed in the region and exported to other regions. That is, for each commodity ***c,*** in time period ***t*** and time-slice ***s***, in region ***r***, we have:

EQ_COMBAL(r,t,c,s) - Commodity Balance

[Sum *{over all* ***p,c*** *∈ TOP(r,p,c,"out") of:* **[*VAR_FLO (r,v,t,p,c,s)*** + ***VAR_SOUT(r,v,t,p,c,s)***STG_EFF(r,v,p)*] *}* +

Sum *{over all* ***p,c*** *∈ RPC_IRE(r,p,c,"imp") of:* ***VAR_TRADE(r,t,p,c,s,imp)*** *}* ***]***
* *COM_IE(r,t,c,s)*

$\geq$

Sum *{over all* ***p,c*** ***∈ TOP(r,p,c,"in")*** *of:* ***VAR_FLO(r,v,t,p,c,s)*** +
VAR_SIN(r,v,t,p,c,s) *}* +

Sum *{over all* ***p,c*** *∈ RPC_IRE**(r,p,c,exp)**} of:*

VAR_TRADE(r,t,p,c,s,'exp") +

FR(c,s) * ***VAR_DEM(c,t)***

where:
TOP(r,p,c,"in/out") identifies that there is an input/output flow of commodity c into/from process *p* in region *r;*
RPC_IRE(r,p,c,"imp/exp") identifies that there is an import/export flow into/from region *r* of commodity *c* via process *p;*
STG_EFF(r,v,p) is the efficiency of storage process p;
COM_IE(r,t,c) is the infrastructure efficiency of commodity c;
FR(s) is the fraction of the year covered by time-slice ***s***

Note that the last term of the right-hand side includes sectoral electricity demands in the equation.

Peaking Reserve Constraint

For time-sliced commodities such as electricity, this constraint requires that the total capacity of all processes producing a commodity in each time period and in each region must exceed the average demand in the time-slice where peaking occurs by a certain percentage, *COM_PKRSV(r,t,c,s)*, which is chosen to insure against several contingencies, such as: possible commodity shortfall due to uncertainty regarding its supply (e.g. water availability in a reservoir); unplanned equipment down time; and random peak demand that exceeds the average demand during the time-slice when the peak occurs. Technology capacity may be adjusted a factor to represent that can be relied upon to contribute to the peak.

For each time period ***t*** and for region ***r***, there must be enough installed capacity to exceed the required capacity in the season with largest demand for each commodity ***c*** by a safety factor *E* called the *peak reserve factor*.

EQ_PEAK(r,t,c,s) - Commodity peak requirement

> ***Sum*** *{over all* **p** *producing* **c** *as its activity-defining output of CAP2ACT(r,p) * PEAK(r,v,p,c,s) * FR(s)* ****VAR_CAP(r,v,t,p) * VAR_ACTFLO(r,v,p,c)*** *} +*
>
> ***Sum*** *{over all* **p** *producing* **c** *along with other energy outputs of PEAK(r,v,p,c,s)* ****VAR_FLO (r,v,t,p,c,s)****} +*
>
> ***VAR_TRADE(r,t,p,c,s,i)***
>
> ≥
>
> *[1+ COM_PKRSV(r,t,c,s)] * [* ***Sum*** *{over all* **p consuming c** *of* ***VAR_FLO(r,v,t,p,c,s)*** *+* ***VAR_TRADE(r,t,p,c,s,e)*** *}]*
>
> where:
>
> *COM_PKRSV(r,t,c,s)* is the region-specific reserve coefficient for commodity *c* in time-slice *s*
>
> *PEAK(r,v,p,c,s)* specifies the fraction of technology *p*'s capacity in a region *r* for a period *t* and commodity *c* that is allowed to contribute to the peak load in slice *s***.** Many types of supply processes are predictably available during the peak and thus have a *PEAK* coefficient equal to 1, whereas others (such as wind turbines or solar plants) are attributed a *PEAK* coefficient less than 1, since they are on average only fractionally available at peak.

Note that the first and second terms distinguish between two cases:

- For technologies where the peaking commodity is the output defining the technology activity (such as a typical power plant), the capacity of the process may be assumed to contribute directly to the peak.
- For processes where the peaking commodity does not define the activity, for example in the case of a combined heat and power plant with flexible electricity + heat outputs defining the activity, the capacity as such cannot be used in the equation. In this case, the

actual production of the peaking commodity is taken into account in the contribution to the peak, instead of the capacity.

The third term assumes that imports of the commodity are contributing to the peak of the importing region, that is that interregional electricity trades are of the *firm power* type.

5.4.5. Bounds and Constraints

In addition to these core model constraints, the user should be able to easily impose bounds and constraints on any desired model variable or combination of variables. A bound is an upper, lower, or fixed limit on a single variable, and should be able to be imposed on:

- Capacity, investment, generation, fuel consumption, or emissions from any technology. These bounds can be used to set operational or policy limit technology investment and behavior.
- Total production, consumption, or net level of any commodity (for example, to set emissions caps). Such bounds should be able to set for a single model period or cumulatively across specified periods.

More complex relationships between variables should be permitted using a simple, user friendly structure for building linear relationships between any combination of model variables across user-specified sets of processes and/or commodities. Common applications of such constraints include:

- Impose bounds on individual technology/commodity for a group of regions. For example, put a common emission cap on a group of regions that are allowed to trade.
- Impose renewable portfolio, clean energy standard, carbon emissions intensity and other policies based on relationships between fractions of generation by different plant types or between emissions and generation. These constraints can permit quite detailed specifications, including different rules for different plant types.
- Create banking schemes for cap and trade programs.
- Model capacity retrofits, by tying capacity of the retrofitted device to the capacity of the original plant, requiring the old plant to retire when the "new" plant is invested in.
- Impose build-rate limits: total creation of wind capacity not to exceed ***x*** GW per year, across different cost and wind classes.
- Model build-rate penalties: any capacity created beyond ***x*** GW per year will cost an additional $***y*** / GW.

Constraints that relate variables across periods or timeslices can be used to:

- Impose a limit on the rate of growth (or decay) in capacity/activity of a new technology type, from one period to the next; and
- Limit the ramping rate – the percentage increase or decrease in the utilization level – of a thermal power plant from one timeslice to the next.

Section 5.3.2 provides some examples of these bounds and constraints.

5.5. Results Data Processing and Visualization System

We recommend two subsystems to process and view results data: a model back end that permits viewing all model outputs at the most detailed level, similar to the model front end data browser, and a visualization system that aggregates outputs into high-level variables intelligible to users beyond the core modeling team, coupled with sophisticated, collaborative visualization tools to utilize it.

5.5.1. Detailed Results Data Browser

The detailed results data browser is a topology-aware database that stores model output data. This interface should also be able to pull information from the model instance data. For example, it should be possible to see (input) technology characteristics like costs and efficiency alongside (output) capacity and generation.

The primal and dual solution values from model runs should be stored in the table shown below.

Table 25 - Schema to store all the reporting variable values produced by each model run

FieldName	Remarks
Attribute	Activity, Capacity, Investment, Flow, Marginal values of commodity balance and other constraints etc
Commodity	
Technology	
Period	
Region	
Vintage	
TimeSlice	
UserConstraint	
Scenario	
VALUE	Numeric field for attribute values

Descriptions of technologies and commodities should be stored as shown below. It should be possible to choose between <shortname>, <description> and <shortname – description> in the results analysis interface.

Table 26 - Elements dictionary

FieldName	Remarks
DimName	All non-numeric fields from results table
EleName	Name in the results table
EleDesc	Description

The set definitions should be created in the same rule-based way that was described in the scenarios section and stored in the schema shown below.

Table 27 - Schema for sets definitions

FieldName	Remarks
DimName	All non-numeric fields from results table

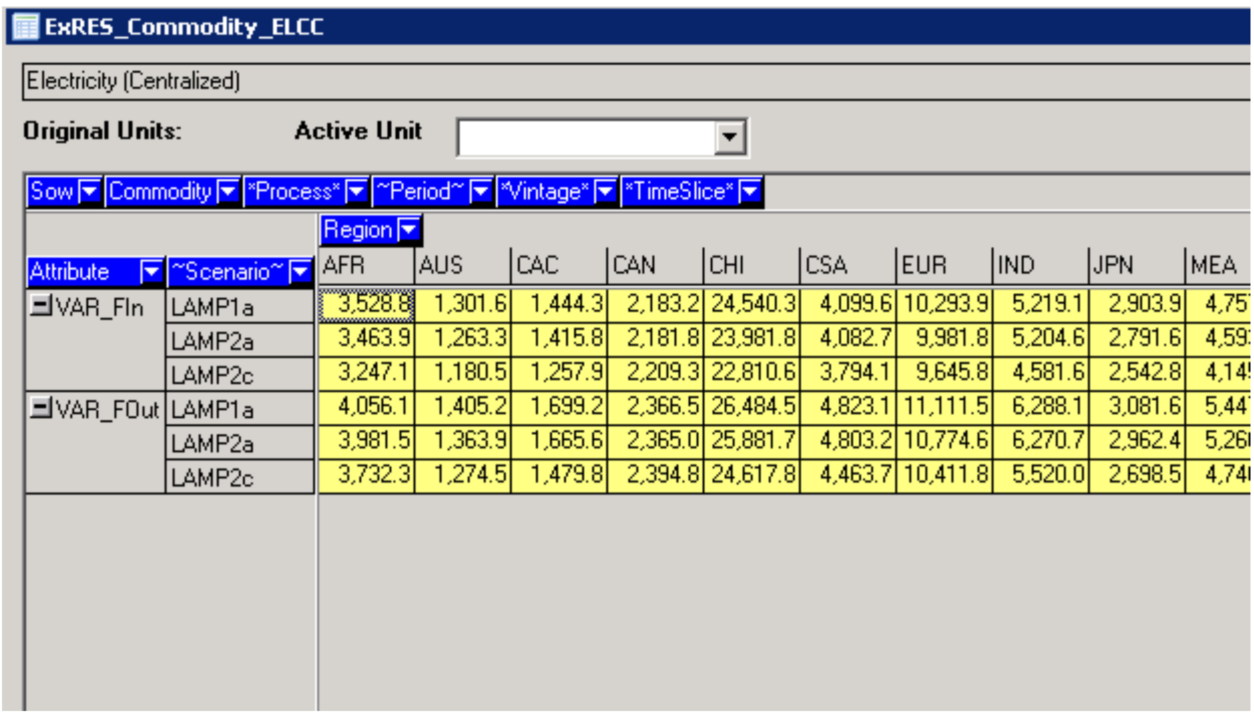
SetName	Name of the set
EleName	Members of the set

As in the front end data browser, it should be possible to navigate the solution along the Reference Energy System of the model using the topology information stored in the table shown below. This feature greatly facilitates model diagnosis and the understanding of results.

Table 28 - Storing topology

FieldName	Remarks
Technology	
Commodity	
IO	IN/OUT

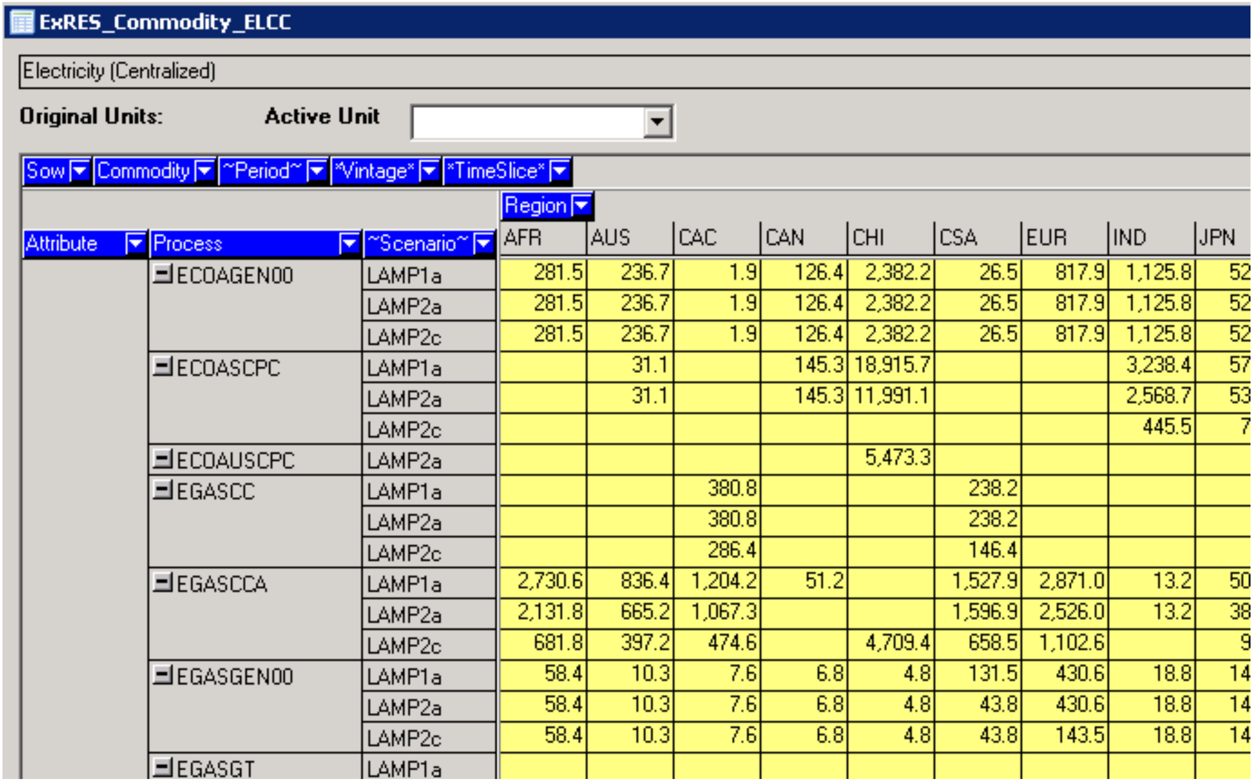
Figures 17-20 show examples of this process at work. Figure 17 shows total consumption (VAR_Fin) and production (VAR_Fout) of centralized electricity, by region, in a selected model period for three different scenarios.



Attribute	~Scenario~	AFR	AUS	CAC	CAN	CHI	CSA	EUR	IND	JPN	MEA
VAR_FIn	LAMP1a	3,528.8	1,301.6	1,444.3	2,183.2	24,540.3	4,099.6	10,293.9	5,219.1	2,903.9	4,75
	LAMP2a	3,463.9	1,263.3	1,415.8	2,181.8	23,981.8	4,082.7	9,981.8	5,204.6	2,791.6	4,59
	LAMP2c	3,247.1	1,180.5	1,257.9	2,209.3	22,810.6	3,794.1	9,645.8	4,581.6	2,542.8	4,14
VAR_FOut	LAMP1a	4,056.1	1,405.2	1,699.2	2,366.5	26,484.5	4,823.1	11,111.5	6,288.1	3,081.6	5,44
	LAMP2a	3,981.5	1,363.9	1,665.6	2,365.0	25,881.7	4,803.2	10,774.6	6,270.7	2,962.4	5,26
	LAMP2c	3,732.3	1,274.5	1,479.8	2,394.8	24,617.8	4,463.7	10,411.8	5,520.0	2,698.5	4,74

Figure 17 - Total consumption and production of electricity in 3 scenarios, by region (2030)

Figure 18 shows the same view, but now the Process block has been pulled into the table, and we see a portion of the production data by power plant type.



ExRES_Commodity_ELCC

Electricity (Centralized)

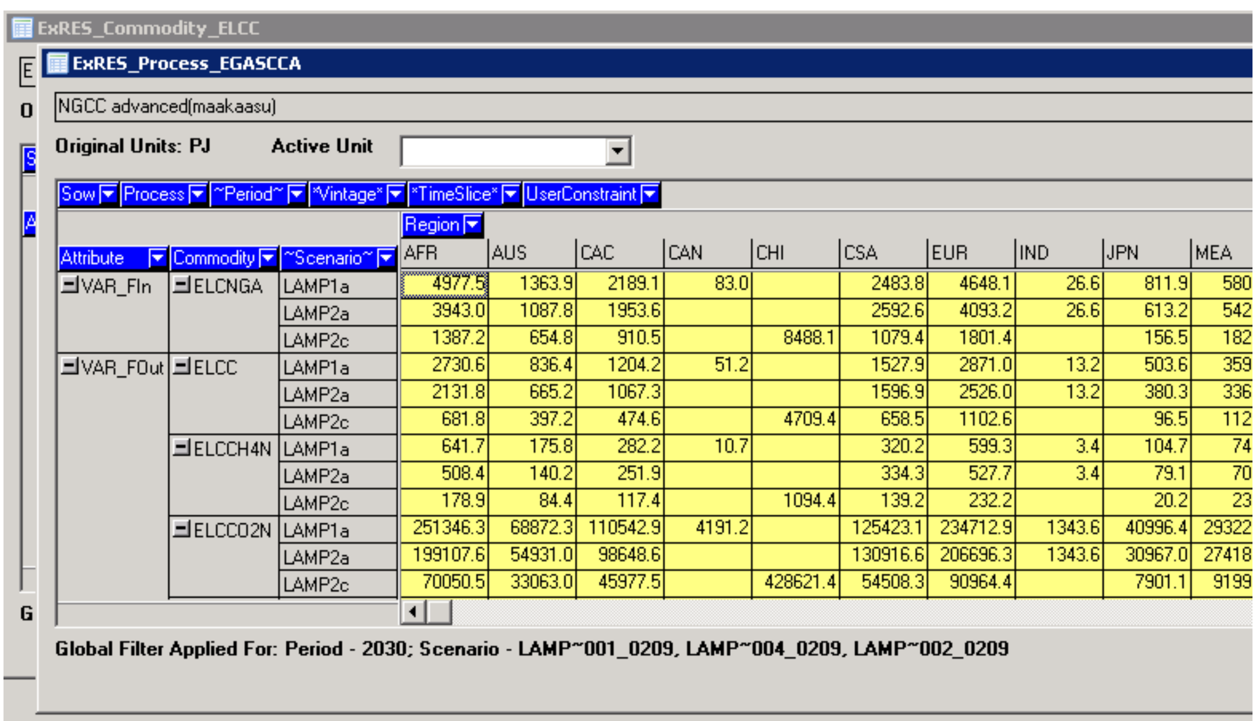
Original Units: Active Unit

Sow | Commodity | ~Period~ | *Vintage* | *TimeSlice*

Attribute	Process	~Scenario~	AFR	AUS	CAC	CAN	CHI	CSA	EUR	IND	JPN
	ECOAGEN00	LAMP1a	281.5	236.7	1.9	126.4	2,382.2	26.5	817.9	1,125.8	52
		LAMP2a	281.5	236.7	1.9	126.4	2,382.2	26.5	817.9	1,125.8	52
		LAMP2c	281.5	236.7	1.9	126.4	2,382.2	26.5	817.9	1,125.8	52
	ECOASCPC	LAMP1a		31.1		145.3	18,915.7			3,238.4	57
		LAMP2a		31.1		145.3	11,991.1			2,568.7	53
		LAMP2c								445.5	7
	ECOAUSCPC	LAMP2a					5,473.3				
	EGASCC	LAMP1a			380.8			238.2			
		LAMP2a			380.8			238.2			
		LAMP2c			286.4			146.4			
	EGASCCA	LAMP1a	2,730.6	836.4	1,204.2	51.2		1,527.9	2,871.0	13.2	50
		LAMP2a	2,131.8	665.2	1,067.3			1,596.9	2,526.0	13.2	38
		LAMP2c	681.8	397.2	474.6		4,709.4	658.5	1,102.6		9
	EGASGEN00	LAMP1a	58.4	10.3	7.6	6.8	4.8	131.5	430.6	18.8	14
		LAMP2a	58.4	10.3	7.6	6.8	4.8	43.8	430.6	18.8	14
		LAMP2c	58.4	10.3	7.6	6.8	4.8	43.8	143.5	18.8	14
	EGASGT	LAMP1a									

Figure 18 - Electricity production by technology

Figure 19 steps backward into the model energy system network by selecting one of the power plants. The view now shows all the inputs and output of this process.



ExRES_Process_EGASCCA

NGCC advanced(maakaasu)

Original Units: PJ Active Unit

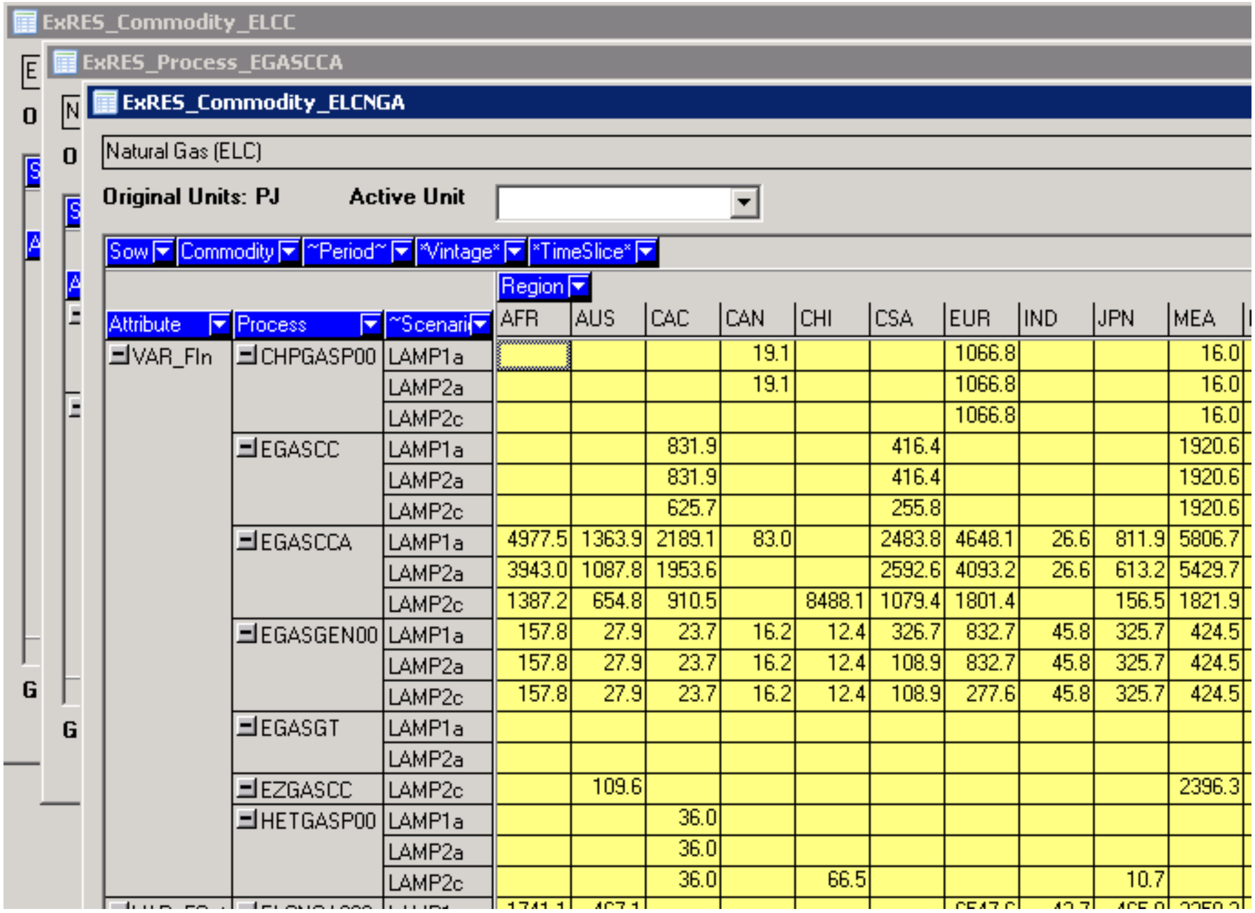
Sow | Process | ~Period~ | *Vintage* | *TimeSlice* | UserConstraint

Attribute	Commodity	~Scenario~	AFR	AUS	CAC	CAN	CHI	CSA	EUR	IND	JPN	MEA
VAR_FIn	ELCNGA	LAMP1a	4977.5	1363.9	2189.1	83.0		2483.8	4648.1	26.6	811.9	580
		LAMP2a	3943.0	1087.8	1953.6			2592.6	4093.2	26.6	613.2	542
		LAMP2c	1387.2	654.8	910.5		8488.1	1079.4	1801.4		156.5	182
VAR_FOut	ELCC	LAMP1a	2730.6	836.4	1204.2	51.2		1527.9	2871.0	13.2	503.6	359
		LAMP2a	2131.8	665.2	1067.3			1596.9	2526.0	13.2	380.3	336
		LAMP2c	681.8	397.2	474.6		4709.4	658.5	1102.6		96.5	112
	ELCCH4N	LAMP1a	641.7	175.8	282.2	10.7		320.2	599.3	3.4	104.7	74
		LAMP2a	508.4	140.2	251.9			334.3	527.7	3.4	79.1	70
		LAMP2c	178.9	84.4	117.4		1094.4	139.2	232.2		20.2	23
	ELCCO2N	LAMP1a	251346.3	68872.3	110542.9	4191.2		125423.1	234712.9	1343.6	40996.4	29322
		LAMP2a	199107.6	54931.0	98648.6			130916.6	206696.3	1343.6	30967.0	27418
		LAMP2c	70050.5	33063.0	45977.5		428621.4	54508.3	90964.4		7901.1	9199

Global Filter Applied For: Period - 2030; Scenario - LAMP~001_0209, LAMP~004_0209, LAMP~002_0209

Figure 19 - All inputs and outputs of one technology - advanced NGCC

Figure 20 steps back one more step in the network. It shows all the processes that consume ELCNGA (natural gas to power plants).



ExRES_Commodity_ELCC
ExRES_Process_EGASCCA
ExRES_Commodity_ELCNGA

Natural Gas (ELC)

Original Units: PJ   Active Unit

Sow | Commodity | ~Period~ | *Vintage* | *TimeSlice*

Attribute	Process	~Scenario	AFR	AUS	CAC	CAN	CHI	CSA	EUR	IND	JPN	MEA
VAR_FIn	CHPGASP00	LAMP1a				19.1			1066.8			16.0
		LAMP2a				19.1			1066.8			16.0
		LAMP2c							1066.8			16.0
	EGASCC	LAMP1a			831.9			416.4				1920.6
		LAMP2a			831.9			416.4				1920.6
		LAMP2c			625.7			255.8				1920.6
	EGASCCA	LAMP1a	4977.5	1363.9	2189.1	83.0		2483.8	4648.1	26.6	811.9	5806.7
		LAMP2a	3943.0	1087.8	1953.6			2592.6	4093.2	26.6	613.2	5429.7
		LAMP2c	1387.2	654.8	910.5		8488.1	1079.4	1801.4		156.5	1821.9
	EGASGEN00	LAMP1a	157.8	27.9	23.7	16.2	12.4	326.7	832.7	45.8	325.7	424.5
		LAMP2a	157.8	27.9	23.7	16.2	12.4	108.9	832.7	45.8	325.7	424.5
		LAMP2c	157.8	27.9	23.7	16.2	12.4	108.9	277.6	45.8	325.7	424.5
	EGASGT	LAMP1a										
		LAMP2a										
	EZGASCC	LAMP2c		109.6								2396.3
	HETGASP00	LAMP1a			36.0							
		LAMP2a			36.0							
		LAMP2c			36.0		66.5				10.7	

Figure 20 - All technologies that consume gas

5.5.2. Results Visualization System

The data browse system describe in Section 5.5.1 is an essential tool for the direct model users to debug scenarios and fully understand model behavior. However, it is a cumbersome way for model users to get an overview of model results, and it does not facilitate communication and exploration of the results by a wider circle of domain experts. Exploring the raw results stored above would require deep knowledge of the model topology. For example, the analyst would need to know the model naming conventions to be able to construct a view of electricity generation fuel mix. Well-documented sets of processes would address the issue to some extent, but the attributes names might remain esoteric to all outside the core modeling team.

Therefore we recommend use of a results visualization system that uses rule-based aggregation/transformation of raw results to create intuitive high-level variables that most analysts will look for while analyzing model runs. These variables can then be used in a data visualization platform that can rapidly produce graphs of the aggregated variables and provide a platform for communication and collaboration with other designated experts, both inside and outside of EIA.

The aggregated variables should be created from the raw results by specifying rules, along with variable names, descriptions, and units, as shown in Table 29. For example, the first row creates a variable named *Generation_NewCoal* that sums the activities of all processes in the set *NewCoal* within each region and model period.

Table 29 - Rules for creating standard variables out of raw results

Attribute	SetName	Units	VarblName	VarblDesc	Comment
Activity Variable	NewCoal	Twh	Generation_NewCoal	NewCoal.Gen	
Activity Variable	NewNuc	Twh	Generation_NewNuc	NewNuc.Gen	
Activity Variable	Elec-RE	Twh	Generation_Elec-RE	Elec-RE.Gen	
Activity Variable	Elec-Fossil	Twh	Generation_Elec-Fossil	Elec-Fossil.Gen	
Capacity Variable	NewCoal	GW	Capacity_NewCoal	NewCoal.Cap	
Capacity Variable	NewNuc	GW	Capacity_NewNuc	NewNuc.Cap	
Capacity Variable	Elec-RE	GW	Capacity_Elec-RE	Elec-RE.Cap	
Capacity Variable	Elec-Fossil	GW	Capacity_Elec-Fossil	Elec-Fossil.Cap	

Table 30 provides a schema for the database storing these variables.

Table 30 - Schema to store high-level variables for web visualization

FieldName	Remarks
Scenario	
Model	
Varbl	
Region	
Year	
Unit	
Val	

The variables defined in this manner would be fully flexible to EIA specification, but should include such high-level results as:

- Generation, capacity, investment, fuel consumption, and emissions by plant type;
- Electricity and fuel prices;
- Electricity consumption by sector; and
- Interregional electricity trade.

The system should be designed to support three types of uses: individual analysis, collaborative analysis, and dissemination of scenarios results. Visualization for analysis can be seen as a three-step process: filter, configure and render. The system should allow filtering of multiple elements from each of the following dimensions: Scenario, Variable, Region and Year. Lists of unique elements for each dimension should be made available for selection by the user with search and dictionary facilities to readily locate elements of interest (for example, all variables related to price, or emissions). Users should also be able to define sets and save of elements often selected together, for example, variables that together make up a complete picture of electricity generation by fuel.

To construct a chart, the user then determines how elements are configured for a view, selecting from the following options for each dimension: assign to the x-axis, make a separate chart for each selected element (small multiples), sum over selected elements, make a separate tab for each element, or display on y-axis. Figures 21 and 22 show how this facility can be used to pivot information on the view to highlight and compare different dimensions. Figure 21 shows a simple stacked column graph of generation by unit type evolving across time, with four different scenarios shown in small multiples.

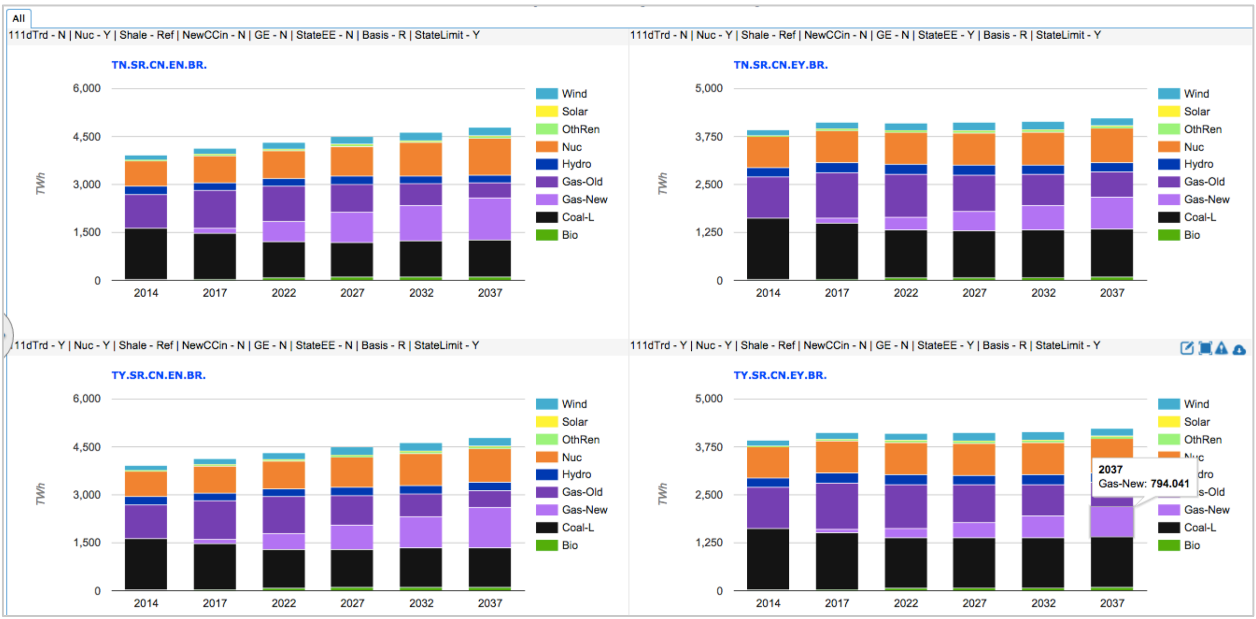


Figure 21 - Sample visualization with years on the x-axis and scenarios in small multiples

The call-out in the lower right corner of the figure illustrates that these charts should be interactive, allowing the user to see the data behind the chart by hovering over any portion of it. (In this case, we can see that generation from new gas units in 2037 is 794 TWh in the depicted scenario.) The system should also permit the user to convert an entire view to tabular form in a pivot table and download in to Excel.

Figure 22 pivots the information in Figure 21 to show scenarios on side-by-side on the x-axis, for different years in small multiples, now allowing a more direct comparison of the generation mix between scenarios.

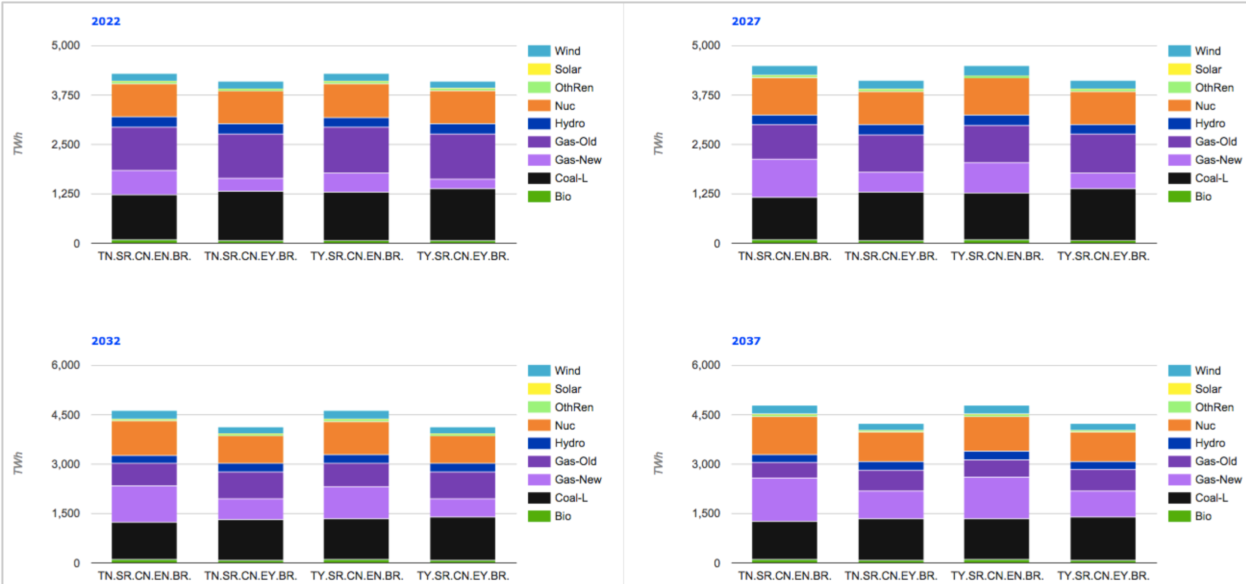


Figure 22 - Sample visualization with scenarios on the x-axis and years in small multiples

The system should also have a facility for creating scenario differences, that is, sets of results that show the differences between values in two scenarios that differ along a single user-specified dimension (such as renewable energy potential scenarios or policy choices), along with differences from a business-as-usual scenario. Figure 23 shows an example of a scenario difference view, which was created to examine the impact of low shale resource in four different cases, on electricity generation mix by state, in four different years for USA. In this view, different model regions (here, states) are shown in small multiples, and the years are put on tabs. Each of the four "scenarios" shown along the x-axis is really the difference between the results for two scenarios whose specifications are identical except for the assumed shale resource. These scenario differences are calculated in the database and stored as "scenarios" to be used in charts just as other scenarios are.

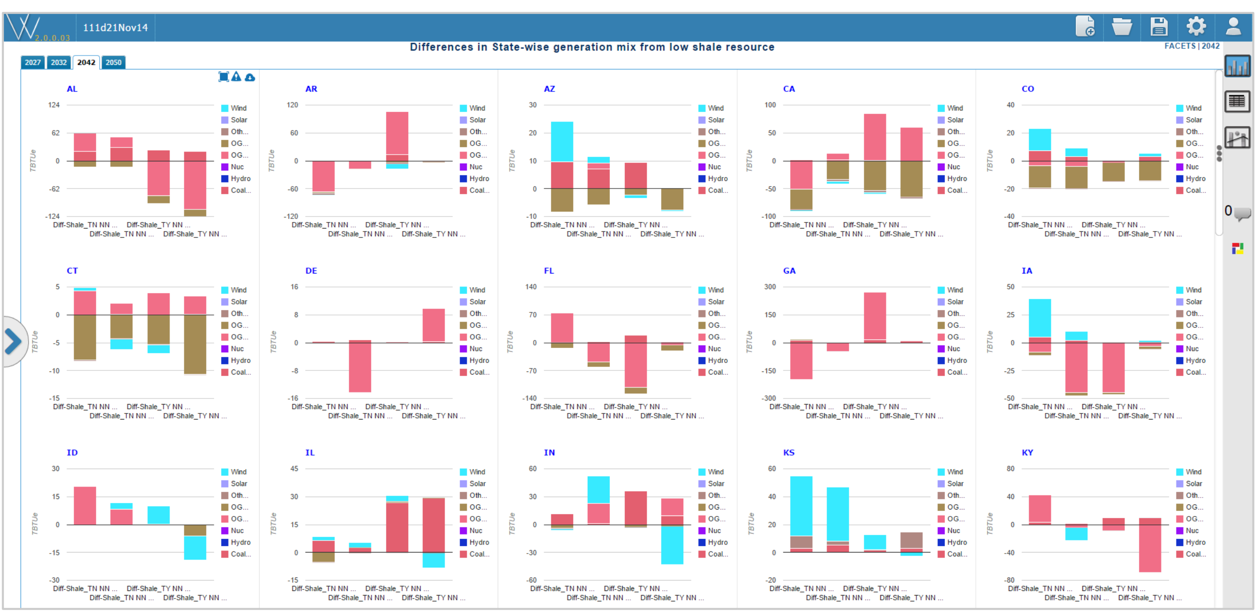


Figure 23 - Sample scenario difference view

In addition to column and line graphs, should be able to use additional formats, including:

- Maps, to visualize regional information, including trade flows between regions;
- Sankey diagrams, to visualize flows within the energy system; and
- Parallel coordinates, scatter matrices and animated bubble charts, to analyze patterns across large numbers of scenarios.

Figure 24 illustrates the use of a map view with CO_2 flows to explore carbon capture and storage scenarios for the Iberian Peninsula. In this view, red nodes show the CO_2 emission sources and blue nodes show the storage sites. The size of the nodes represents the size of the sources and sinks, and the size of the arrows indicates the transportation capacity and flow direction.

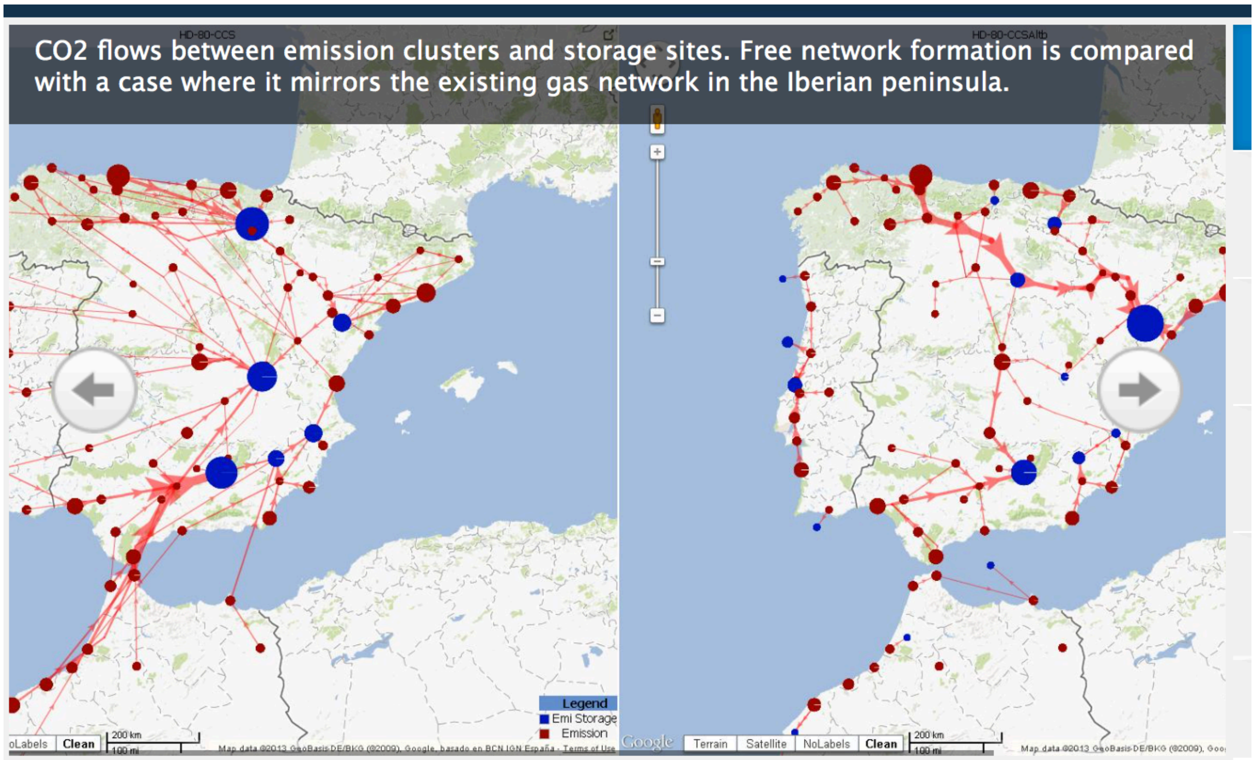


Figure 24 - Sample map view

Figure 25 shows an example of a bubble chart, which is a time-animated scatterplot of results from many scenarios that allows the user to set the variables on the x and y-axes, color scenarios according to scenario definition dimensions, and play the view forwards in time. The example shown here includes 100 scenarios that differ across dimensions including carbon price, technology availability and cost, and transmission grid expansion within a Japan multi-region electricity system model. In this figure, capacity of onshore and offshore wind have been placed on the x and y-axes, and the time control along the bottom of the view has been set to 2050.

Each circle depicts one scenario's results. We see a wide range of results for both on and offshore capacity. By setting the color of the scenario circles to represent scenario dimensions, the user can explore the reasons for differences across scenarios. Here the scenarios have been colored according to whether they permit transmission grid expansion. A clear bifurcation in the results can be seen. Without grid investment (dark blue circles), onshore wind development is limited, and offshore wind is further developed. With grid expansion (light green circles), onshore wind can be further developed. Visualization tools like these are essential to facilitate extracting information and insights from large numbers of runs, so that uncertainty space can be well explored.

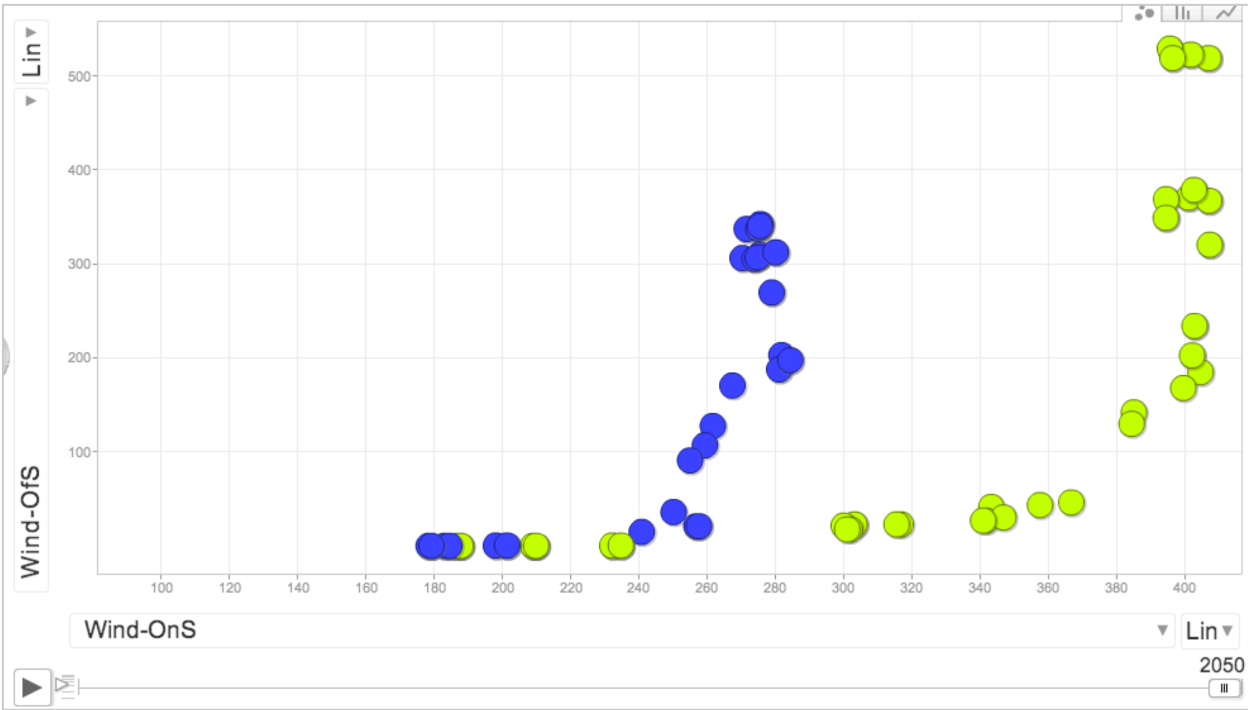


Figure 25 - Sample bubble chart view

To support collaborative analysis, the system should allow different users to save their views, as public or private, along with observations. Each view should support a discussion page, where users can discuss the points of interest.

To support dissemination of scenario results and findings, the system should support at least three options: saving views as web links, conversion of selected views and comments to a presentation, and development of portals to allow users to explore selected results directly. A direct web link should be able to be generated for any saved view, which can be distributed via email or used within reports or on websites. Selected views should be able to be grouped into a presentation, along with itemized observations stored for each view. An example of such a presentation view is shown in Figure 26. Here the observations have been placed into a bulleted list on an overlay layer, which can be collapsed or enlarged, with the buttons in the center of the picture.

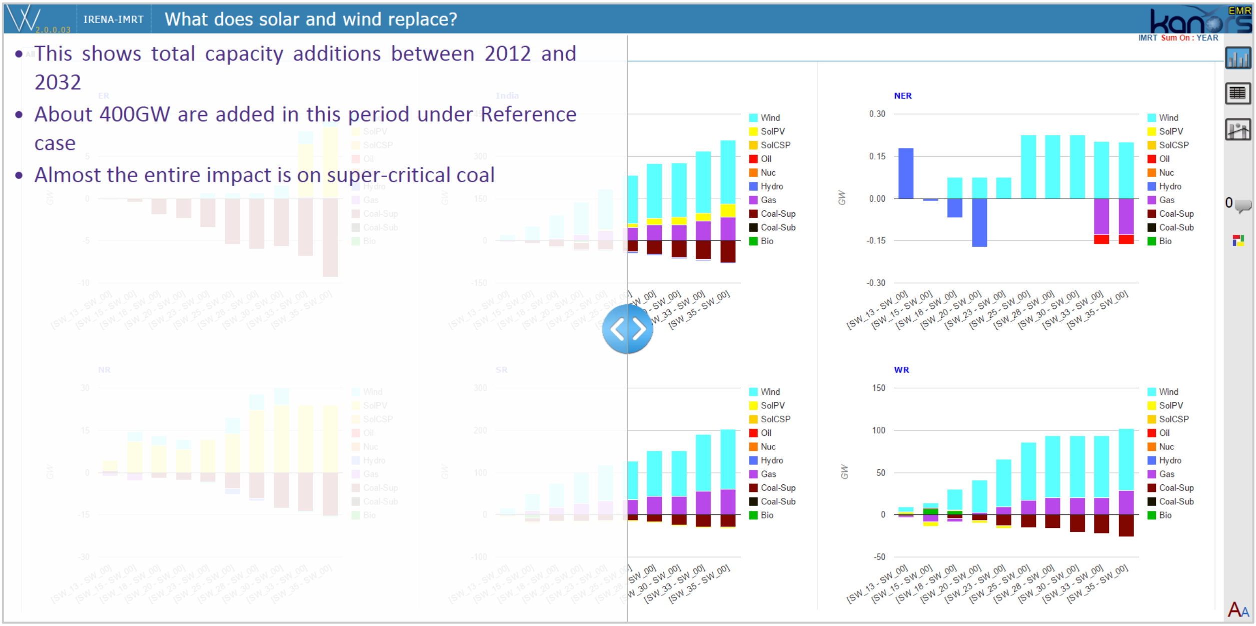


Figure 26 - Sample presentation view

Figure 27 shows a sample of a portal constructed from six selected views of scenario results data. It has been set up to allow the user to explore global energy and emissions scenarios while changing key dimensions including CCS, nuclear, and renewable available and cost. Like the DECC calculator, this view enables exploration of hundreds of scenarios with minimal effort. However, unlike the DECC calculator, the portal is not performing simulations to carry out calculations, but combining scenario elements to draw from model runs that have already been conducted. In this way, a large set of parametric runs, along with such a portal, can convert a model into an expert system that can promote and support discussion among policy makers and stakeholders.

Figure 27 - Sample portal view

The use of this visualization system is not confined to analyzing IEMM results. Section 5.1.2 provided examples of how it could be used to analyze the granular data stored in the IEMM input database, including data on existing units, electricity load curves, and renewables resources, and thereby to help inform assignment of missing data values and aggregation choices for model instances. It can also be used to display all WEPS+ input into IEMM and the full WEPS+ output. For example, it could be used to create a presentation tool for IEO results and used to compare previous years' IEO results to the current IEO. In fact, it should be easy to use such a platform to view any structured dataset, such as EIA survey data, State Energy Data System (SEDS) data, and NEMS results.

6. Implementation Considerations

6.1. Software, Hardware, and Runtime

The granular input data (Tier 1) and the results processing system (Tier 5) will be housed in a relational database format, which can be implemented in executed on any standard database platform like MySQL, MS SQL Server or ORACLE. The model instance generator (Tier 2) and all front end (Tier 3) scenario specifications will be controlled by user-friendly Excel spreadsheets. The user interfaces for Tiers 3 and 5 could be developed in VisualBasic or VisualBasic.NET, with the relational database as the backend and Excel files for all data input. Note that we do not recommend any data input to be done outside Excel.

The proposed model (Tier 4) can be written in AIMMS, GAMS or other similar mathematical modeling language. The online visualization platform (Tier 5) can be built with a mix of ASP.NET, Javascript, and open source visualization widgets from D3JS[14].

Windows CMD files can be used to launch any of the processes described in this document.

Since we are recommending Excel as the basic user interface for data specification, this system would work best on the Windows platform. Model solve times will of course vary depending on the chosen level of aggregation. With around 20 regions and existing stock aggregated to around 40-50 plant types in each region, the model should take between 10 and 20 minutes to solve on a fast windows machine with 16GB RAM. For bigger model instances, the front end should be able to produce platform-independent input for the solvers and run the model on other machine(s), permitting multiple scenario runs in parallel.

6.2. Data Requirements and Sources

All inputs and outputs for the model are communicated through text files produced by the system. But all model input from WEPS+, including electricity demand by region, sector, and year, and fuel prices by region and year, and all model outputs required by WEPS+, including capacity, generation, fuel consumption, and prices, should be read from and written into the WEPS+ restart file under SQLite database.

Some third party data, notably the Platts database of existing capacity, has been recommended for this approach. We have assumed that EIA can utilize its standard procedures to work with this data. However, the model developers must work with EIA to meet EIA's needs for the use of third-party data and/or to find alternatives if the third-party data cannot meet EIA's needs.

[14] http://d3js.org/

7. Conclusions and Recommendations

Major advancements in data processing and visualization techniques in recent years, combined with a well-designed model framework can enable a quantum leap forward in EIA's international electricity market modeling capabilities. In this CDR, we've recommended an approach based on the following design principles:

1. The IEMM should make best use of all available relevant data.
2. The capability to prepare different model "instances" from the source database is required.
3. The structure of model instances must be data-driven.
4. The framework should provide powerful, flexible scenario analysis capabilities.
5. The IEMM should be supported by a powerful data visualization system.
6. All bulk data creation should be done in a rule-based manner.

In Section 5, we've provided specifications for a sophisticated knowledge and model management system supporting a capacity and dispatch model based on (mostly) linear optimization. As noted there, however, EIA could elect to use a different model structure and solution logic within this system, either instead of or alongside the optimization model described herein, provided that it supports these design principles (most notably, that it is data-driven) and connects properly with the model management system.

In the following Appendix, we briefly describe one possible framework for implementing the proposed optimization model, the TIMES framework.

Appendix on TIMES

The TIMES framework is an open-source, engineering-economic energy systems model generator developed by the International Energy Agency's Energy Technology Systems Analysis Program[15] (ETSAP), a consortium of nearly 20 national labs and research institutions. TIMES, and its predecessor MARKAL, have been used by more than 300 institutions in nearly 70 countries over the past 40 years to model energy systems ranging from municipal to global. The framework continues to evolve in response to the analytical needs of its user base, with significant new features regularly developed.

TIMES provides a set of data-driven model building blocks similar to those described in Section 5.4 to model energy resources, energy transformation processes (such as power plants), energy transmission, and devices that use energy to satisfy end use service demands. TIMES was designed to incorporate many of the capabilities described in this CDR.

For example, TIMES was developed specifically to enable data-equation independence and to permit run-time adjustment of model years with no need to adjust data years. The TIMES code handles interpolation/extrapolation by rules, as described herein, and the tracking of investment costs, in particular, is quite nuanced, in order to support such time period flexibility. It has a full suite of handles for developing complex user constraints on linear combinations of model variables, including the ability to support cumulative constraints and constraints that relate variables across model periods and timeslices. It also supports endogenous technology learning, endogenous economic retirement decisions for processes, and flexible-input and -output processes.

TIMES has several other features that may also be useful for the IEMM down the road, even if not envisioned for near term use, including:

- A robust representation of electricity storage, allowing daily and seasonal charge/discharge cycles;
- Parameters and equations to capture issues connected to dispatching of thermal units to support more nuanced modeling of renewables integration, including ramping constraints, minimum up and down time, startup/shutdown costs, and cost and efficiency penalties for partial loading;
- The option to use integrated climate equations to model policy scenarios with maximum global temperature/forcing increase;
- The ability to use environmental damage costs for criteria pollutant emissions in the model objective function;
- Use of mixed integer programming to model to "lumpy" investment decisions, where new capacity can only be built in integer multiples of some minimum viable capacity;
- Multiphase tradeoff analysis, in which alternate objective functions can be used along with a user-specified maximum deviation from the least cost solution; and
- Multi-stage stochastic analysis, which permits derivation of strategies to hedge against probabilistic states of the world before they get resolved.

---

[15] http://www.iea-etsap.org/

The TIMES code is written in GAMS and is solved by a standard GAMS solver. TIMES is also supported by the VEDA front and back end data management shells and the VEDAViz data visualization system, that carry out the functions called for in Sections 5.3 and 5.5 for TIMES-based models. These systems are based on the same rule-based data generation philosophy that plays a central role in the approach we recommend here.

In addition to the reduced development effort and cost associated with using an existing, well-tested and documented system, using TIMES brings the benefit of participating in an active community of users and developers who continue to advance the tools and the modeling paradigm. However, while TIMES can support all of the requirements laid out in this CDR, nothing described herein requires the use of TIMES, provided the model(s) selected can provide similar data interfacing capabilities.