NEMS EMM REGIONAL REASSESSMENT
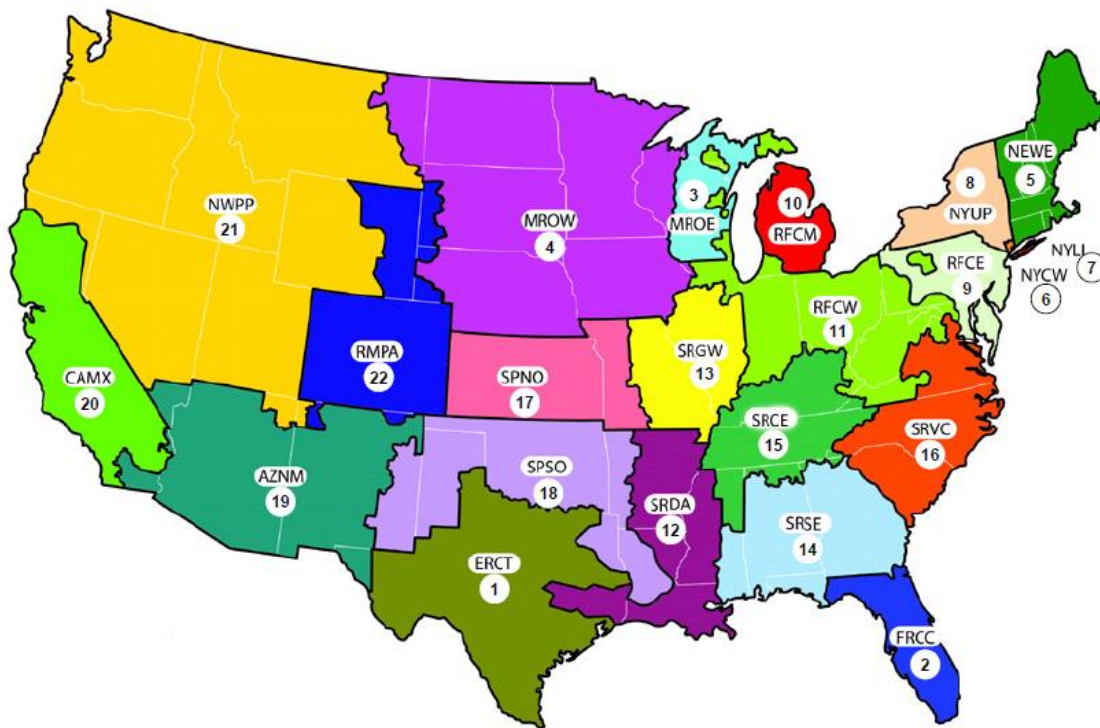## COMPONENT DESIGN REPORT

FINAL REPORT

Prepared for:
U.S. Energy Information Administration

Prepared by:
OnLocation, Inc.
LEIDOS Contract # DE-BP0005365

Revised April 27, 2017

OnLocation

Contents

Tables

Figures

Executive Summary

The EIA NEMS modeling framework is built around a number of regional definitions to improve its forecasting capabilities. The Electricity Market Module (EMM) is responsible for forecasting the capacity, generation, fuel use, emissions and pricing associated with the electric power sector (including renewable and intermittent technologies). The EMM regions are specified in a way that is intended to improve the EMM forecasting by capturing the key regional distinctions that impact the forecasted variables (prices, capacity, etc.). EMM regional specifications have a significant influence on the electric power forecasts produced by NEMS.

Four key overarching dynamics that are addressed in this CDR and must be captured in the regional specification of EMM include:

Regulatory Oversight

1. Pricing - see Sections 4.3 (wholesale electricity markets) and 4.6 (retail markets)
2. Allowed emissions (national, regional, state, and plant/unit) - see Section 4.7 (policy requirements)
3. Grid reliability - see Sections 4.2 (balancing authority) and Section 4.4 (reliability requirements)

Access to fuel

4. Access to fossil fuels and renewable resources - see Sections 4.7 (fuel access and policy requirements) and Section 4.8 (non-fossil energy resources)

Each of these dynamics plays an important role in how EMM regions should be specified. The EMM regions were last revisited in the *Annual Energy Outlook 2011*. At that time there was a need to better capture the disintegration of the historical NERC regions into ISOs and RTOs that were evolving as a result of the move to competitive wholesale markets as stimulated by the FERC. Since then, the power sector markets have evolved further and new regulatory forces regarding emissions, wholesale markets and reliability standards have changed the power industry landscape. This invites a recalibration of the EMM regional definitions.

In the EMM there is a tradeoff between the number of EMM regions and the model performance (memory requirements and solution time). This CDR recommends a new regional specification that is a compromise that balances the need to capture the impact of the dynamics mentioned above with the model performance.

This CDR also addresses in detail the data requirements of the EMM and the implications of changing the EMM regions on the data updating and reconfiguring processes and associated levels of effort. It describes two approaches to updating the EMM for new regional specifications. One short term approach leaves most processing methodologies in place and parallels the past approaches to updating and reconfiguring the regions data inputs in the EMM. The second approach describes a course of action that allows more flexibility in the processes to allow alternative regional specifications to be constructed and experimented with at a lower level of effort. However, to get to this more flexible

approach, an investment in additional processes and routines to support the flexible reconfiguration is required.

1. Overview

The Electricity Market Module (EMM) and the Renewable Fuels Module (RFM) of the National Energy Modeling System represent the electricity sector using 22 geographic regions that have not been redefined since the number of regions was expanded from 13 to 22 for the *Annual Energy Outlook 2011*. Electricity markets have continued to evolve, and EIA has requested a review of the regions along with a recommendation for an update to the processes required for creating the regional inputs to the EMM and RFM.

This Component Design Report starts with section 2.0, which describes the purpose of this task. The current EMM and RFM structure, including input files, preprocessor programs, and data sources, are presented in section 3.0. Factors to consider and motivation for alternative EMM/RFM regionality are described in section 4.0. Section 5.0 provides recommendations for implementation of a new regionality, outlining three phases that can be used to move the EMM from a static regionality to a flexible dynamic design. Issues and limitations in our recommended approach are presented in section 6.0. Finally, section 7.0 provides our conclusions.

2. Statement of Purpose

The purpose of this Component Design Report (CDR) is to:

"Evaluate the current structure of regional electricity supply organizations and recommend the number and geographic representation to account for evolving trends in the electric power industry."

Evaluating the current regional structure of the Electricity Market Module (EMM) requires a review of the following model structures and processes:

- Data inputs to the EMM, their data sources, and the processing methodology
- EMM and RFM source code and their includes
- The relationship between the 22 EMM regions and other regions used by the EMM
- Interfaces between the EMM (and RFM) and other NEMS models

A recommendation for changing the number and geographic representation of EMM regions must consider:

- The geographic borders of existing Independent System Operators (ISOs) and Regional Transmission Organizations (RTOs) and grid connectivity
- The geographic scope of regulatory policies
- Regional variation between resources and loads
- Estimation of impacts on model size, accuracy, and performance of altering number of regions

Two approaches to updating the number of EMM regions will be presented. The first will be a static approach that leaves most processing methodologies in place and could be implemented in a relatively short amount of time. Documentation of this methodology will be provided along with a recommendation for a new regional definition. The second approach will include a more flexible, data driven way of processing the input data and establishing the number of regions. This approach will require a larger effort and longer time period to implement because it will involve a complete redesign of the data processing methods. The approaches are not mutually exclusive, because most of the steps included in the static approach are needed to streamline the data for the flexible approach.

For implementation purposes, the task could be split into three phases:

- The first phase would be similar to the approach taken in the prior re-regionalization effort where the model structure would be modified to accommodate a fixed, alternative regionality with minimal changes to how data are processed. In addition to updating of the data files, this phase could include *some* streamlining in preparation for the next phase. Performing this phase would produce detailed documentation of data sources, aggregations, and transformations necessary to prepare inputs to the model. Also, this phase includes testing the model with the new regionality.

- The second phase would maintain a static approach to the new regionality but would involve altering the model processing to enable a future flexible approach. For example, this phase would involve modifying the input files to represent data at an elemental level used in the original data source. The elemental data could then be aggregated to the new regionality using a static methodology.
- The third phase, if implemented, would build upon the structure put into place in the first and second phases to replace the static regional mapping with a dynamic flexible mapping structure.

All three phases will be discussed in detail in this CDR.

3. Current Structure

In NEMS, the Electricity Market Module (EMM) regions have always been defined as collections of load serving entities. The EMM regional load requirements are a summation of load requirements of its members, that is, the load serving entities assigned to each EMM region. The geographical footprint for the EMM regions does not conform to state boundaries. In fact, the size, shape, and identity of these entities can change and have changed dramatically over time. Further, these entities join together in power pools to dispatch their collective generators across existing transmission capacity to match generation to load. Like the individual entities, these collections can change and have changed over time. Allowing the collection of entities that define the EMM regions to be flexible would improve the model's representation of these entities.

The current 22 EMM region boundaries were created based on the 2006 EPA Emissions & Generation Resource Integrated Database (eGRID)[1] regionality. The EMM uses this regionality to determine electricity load requirements, generation (dispatch), electricity transmission and trade, and electricity prices. The EMM has a second regional definition, 23 fuel regions, which are used to determine delivered fuel prices and, together with the EMM regions, locate new fossil generating plants. The fuel regions are defined by a combination of coal demand regions, natural gas demand regions, Census regions, and four CO_2 market regions, all of which use state boundaries. Investment decisions for new non-fossil (nuclear and renewable) capacity are currently made at the EMM region level (i.e., their fuel regions are identical to the EMM regions). The one exception is for biomass plants, where the biomass regions correspond to the coal regions.

3.1 Preprocessor input files, preprocessor code, and EMM and RFM input files

The inputs to the EMM and RFM include 25 files that have regional data, many of which are produced from a preprocessor program and some that are assembled manually. Many are also processed from numerous pre-preprocessor programs. Table 3.1 provides a list of these files, identifies the source of the data, and indicates the lowest level of aggregation of the data that are available. Appendix A provides a description of the processes that have been used to produce these data files at the EMM regional level.

---

[1] See https://www.epa.gov/energy/emissions-generation-resource-integrated-database-egrid.

Table 3.1. EMM and RFM input files

#	NEMS Input File	Prep Input File	Source of Data	Level of Aggregation	Pre-preprocessor	Preprocessor	Tables That Are Organized By MNUMNR in File
1	PLNTDAF.DAF	PLTF860.TXT	EIA-860, EIA-923	Boiler, unit, plant, utility owner and operator, balancing authority, NERC region, state	Pre-preprocessor (SAS) combines EIA-860 and 923 and creates PLTF860.TXT file	Preppelt preprocessor creates DAF	Whole file
2	ETTIN.TXT	PRETTIN.TXT	ISOs, NERC winter/summer assessments, EPA Base Case	NERC subregions	Involves considerable manual work to process/compare across multiple sources	Prepetin preprocessor creates ETTIN.TXT	Whole file
3	ETTDEM.DAF	ETCNCT.TXT	ISOs, NERC winter/summer assessments, FERC Form 1	NERC subregions	Involves considerable manual work to process/compare across multiple sources	Prepett preprocessor creates ETCNCT.TXT	Whole file
4		FRMCHRG.TXT	Analyst	EMM region			Whole file
5		CANSPLY.TXT	Analyst	EMM region			Whole file
6	LDSMSTR.TXT		Analyst, EIA-861 for sales matrices and losses	EMM region; for EIA-861: utility, balancing authority, NERC region, state			LSR Names, plus 100+ end use tables, marginal costs of capacity, intermittent demand, elasticities by end use, Census to EMM mapping matrices
7	LDSMDAF.DAF		FERC Form 714 load files	Balancing authority		Preprocessor creates DAF	Uses a table that maps PCA to EMM region
8	CTSINPUT.GDX, CTSSOLN.GDX	CTUS.MDB	NETL capture, transport, storage costs; also uses PLTF860.TXT file			Run CTUS.MDB to create CTSINPUT.GDX and CTSSOLN.GDX	The MDB has a few EMM tables
9	FINREG.TXT		Energy Velocity (FERC Form 1 and 412, RUS 7 and 12), EIA-861	Utility (mapped to balancing authority then to EMM region)	Involves considerable manual work to process multiple input files data		"Form1 section", PINS, overhead multiples, O&M slopes, capital slopes
10	ELDATYR.TXT		Historical data (EIA-860, 923, 861)	Balancing authority, NERC region, state, EMM region			Generation by fuel type, generation by plant type, fuel consumption, electricity sales
11	EMMCNTL.TXT		Analyst	EMM region			Regparm, Co2 STD TN, CO2 STD ET, CO2 GEN NR, CO2 ALL GN, CO2ALL LD, NERC RGN CO2_IM, NERC RGN CO2_DF, NERC RGN CO2_DE, NERC RGN CO2_OS, DM LOSSCHANGE
12	ECPDATX.XLS		Analyst	EMM region			BLD LIM, P2LIM, COF Min Gen, RPRSW, BAS GEN, EXS GEN, GEN BND, SUNSET, PRSBNK, FUEL TRN CST, EMM TRAN CST, CTL REG MAPPING, Regional Tabs
13	NUGPIPE.TXT		Analyst	EMM region			ISHARE, CSHARE
14	EFPCNTL.TXT		Analyst	EMM region			EMM reporting, OMADJ
15	PHASEIN.TXT		Analyst	EMM region			Whole file
16	SALELB.TXT		Analyst	EMM region			Whole file
17	EMMDB.MDB	EMMDBDOC.XLS	Analyst	EMM region			NERC_RG_DEF
18	LAYOUT.TXT	LAYIN.XLS	Analyst	EMM region			Tables 59, 61, 62, 67, 109
19	WFLOORS.TXT			EMM region			Plant capacity
20	WGEPARM.TXT, WGESITE.TXT		NREL report, EIA-860	EMM regions and states			Existing and new geothermal sites
21	SOLARIN.TXT		NREL	EMM region			Solar generation profiles
22	RENDAT.TXT		NREL	EMM region			Wind (and other renewables) cost multipliers
23	WESAREA.TXT		NREL	EMM region			Wind resources, generation profiles
24	HYDSITE.TXT		Idaho Hydropower Resource Economics Database (IHRED)	Site location			New hydro sites
25	MSWDAT.TXT		EPA, Government Advisory Associates, EPRI TAG	EMM region			Population, MSW capacity, costs

3.2 Methodology used for prior re-regionalization of the EMM

The last EMM re-regionalization was implemented in the *Annual Energy Outlook 2011* (AEO2011), where the 13 NERC-based EMM regions were changed to 22 regions by expanding the number of NERC subregions and splitting New York into 3 regions to better capture transmission bottlenecks and limitations of generating technologies that can be built. This was done in two phases. The first phase focused on making structural changes and involved updating the model's source and include files in order to remove as many hard-wired settings as possible. This phase also included simple updates to data files to switch to a 22 region structure, using 13 region data. The second phase involved updating the data in the input files to properly represent the new 22 regions. Appendix B summarizes these activities.

4. Choosing an Alternative EMM Regionality

The process of choosing an alternative regional structure for a model must start by identifying what questions will be addressed by the model. The EMM is tasked with providing the demand modules with prices by the nine census divisions by demand sector and often by end use designation. The EMM is also asked to provide fuel consumption requirements at a variety of regional constructs depending on fuel type: natural gas, coal, oil products, and biomass.

These tasks occur within a variety of regulations and policies, including Renewable Portfolio Standards, technology tax credits, and Clean Air Act regulations, which are implemented at the federal or state level. There are also two major types of electricity pricing regulations—cost of service and competitive—which generally follow state boundaries. The boundaries and relative size of each region can impact the model's choices for placement of new generators and new transmission.

Finally, the most compelling reason for choosing new EMM regions is matching load requirements to the associated generating assets. These generating assets are in general owned by the entities serving the load, and any local transmission bottlenecks have been eliminated. The balancing authority areas or power control areas are the fundamental building blocks of the power grids. The Regional Transmission Organizations (RTOs) and Independent System Operators (ISOs) are a collection of balancing authorities. The three North American power grids are in turn collections of RTOs and ISOs. Allowing the EMM to dispatch in regions that are defined by collections of load serving entities and/or balancing authority areas should result in solutions that more closely mimic actual market behavior.

The following is a discussion of each of these decision criteria and other considerations for determining the optimal regionality for the EMM, including:

1. Regionality of available data sources
2. Electric utility and balancing authority (BA) boundaries
3. Independent System Operators and Regional Transmission Organizations
4. NERC reliability regions and subregions
5. Electric transmission requirements and constraints
6. Retail electricity pricing methodologies
7. Policy requirements for electric generators
8. Location of non-fossil energy sources
9. Relative region sizes
10. Model performance and other considerations

A summary matrix of these 10 criteria is shown in Table 4.1. A comparison of the regionality of other well-known electricity models will also be provided for comparison purposes. Finally, a recommendation for the new set of regions will be described and compared to the current set of regions.

Table 4.1. Summary matrix of EMM regionality criteria

	Criteria	Opportunities/Challenges	Potential Regionality
1	Regionality of available data sources	Need a logical way to map the data into the model regions	Use most elemental data available, including utilities, states, balancing authorities where available
2	Electric utility and balancing authority (BA) boundaries	Balancing authority areas are the fundamental building blocks of the power grids	Use the BAs to aggregate to a set of regions, although some BAs are too large (e.g., MISO)
3	Independent System Operators and Regional Transmission Organizations	Wholesale markets are a collection of BAs; membership changes periodically	Use ISO/RTO pricing zones to help split regions where needed
4	NERC Reliability Regions & Sub-regions	NERC assesses resource and transmission adequacy for each assessment region	Region boundaries should reflect current market structure
5	Electric transmission requirements and constraints	Existing transmission capacity can limit the ability of generators in one area to meet load requirements in another area	Split BAs where needed to reflect significant transmission bottlenecks, using pricing zones as a guide
6	Retail electricity pricing methodologies	Some regions are a mix of competitive and cost-of-service prices	Join Ohio with RFCE; change SRGW to include only Illinois if needed
7	Policy requirements for electric generators	RPS, environmental, other policies are often determined at the state level	Shift non-fossil plants to fuel regions which have state boundaries
8	Location of non-fossil energy sources	Some are concentrated in specific areas of the country	Shift non-fossil plants to fuel regions which have state boundaries
9	Relative region size (retail sales)	Want a more even distribution of sizes than existing regions which range from <1% to 13% of total retail sales	Reduce range of sizes by aggregating smaller regions and/or splitting larger regions
10	Model performance and other considerations	Must determine appropriate balance between number of regions and solution time	Streamlining fuel region definitions and removing MATS decisions can improve model performance

4.1 Regionality of available data sources

The EMM uses many different data sources to calibrate the model to reflect the current electricity system, including but not limited to generating unit costs and operating characteristics, transmission capacity, load and reserve margin requirements, and electricity rate structures. These data sources are often specified at different regional levels, including unit level, utility, balancing authority, state, and North American Electric Reliability Corporation (NERC) region or subregion level. Table 3.1 shows the regionality of each of the key data sources needed for the EMM and RFM input files.

The regionality of the data sources does not necessarily need to match the regionality used by the model to solve, as long as there is a logical way to map the data into the model regions. For example, the data sources for the financial data used for electricity pricing are provided at the utility level. Some

manual work was needed to map these data to the current EMM regions, so the new EMM regionality would ideally make this process easier. This is particularly true if we switch to a flexible regional format.

4.2 Electric utility and balancing authority boundaries

Most balancing authorities are utilities or, in some cases, collections of utilities (such as the RTOs) that are responsible for balancing supply and demand in a designated area. Many of the data sources used for calibrating the EMM provide data at the balancing authority level or can be mapped to that level of aggregation. Therefore, one option would be to use this regionality for the model's input data where possible and use the balancing authorities to aggregate to a set of regions that could be useful for the model's solution.

There are currently 66 balancing authorities (a.k.a., power control areas) in the United States, with additional ones in Canada and Mexico (see Figure 4.1). The balancing authority areas, in most cases, could be easily aggregated to any set of NERC-based regions or subregions for the model solution; however, analyst judgment will be needed in cases where entities claim to belong to more than one balancing authority. Likewise where a balancing authority (such as MISO or PJM) is too large and needs to be split up, utility-level data could be used where possible to aggregate to the new set of regions. While utility pricing and load responsibility are regulated at the state level, utility, balancing authority and NERC region boundaries that govern power flows span state boundaries, making state-based regions less attractive.

Figure 4.1. U.S. balancing authorities
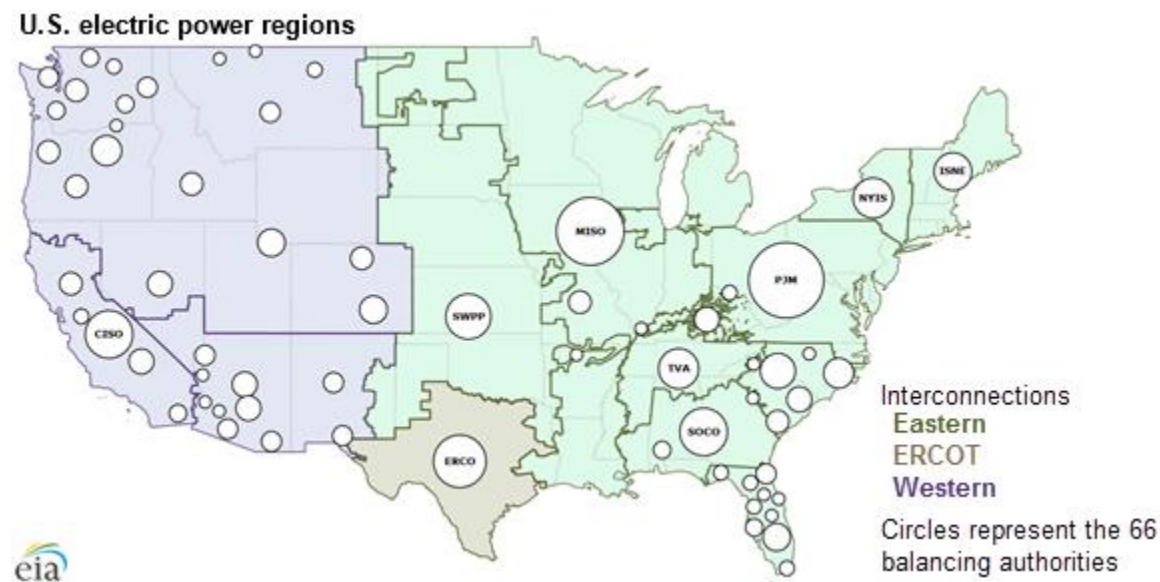
Source: U.S. Energy Information Administration, *Today In Energy*, published July 20, 2016.
Note: U.S. electric system is made up of interconnections and balancing authorities.

4.3 Independent System Operators and Regional Transmission Organizations

Independent System Operators (ISOs) and Regional Transmission Organizations (RTOs) are wholesale electricity markets where members (including transmission operators, generators, and integrated utilities) buy and sell power using shared access transmission systems that are independently operated. The RTOs also function as balancing authorities. According to the Federal Energy Regulatory Commission (FERC), "two-thirds of the nation's electricity load is served in RTO regions."[2]

A map of the existing RTOs is shown in Figure 4.2. In some cases, such as New England, New York, Electric Reliability Council of Texas (ERCT), and California ISO, the organizations generally follow state boundaries. In other cases, particularly the Midcontinent ISO (MISO) and Southwest Power Pool (SPP), the organization can stretch across many states, with some neighboring utilities choosing not to join. The membership of the RTOs can change over time as utilities, cooperatives, and public power authorities choose to join an organization or switch their membership to another RTO.

Since the last EMM re-regionalization, these boundaries have shifted:
- SPP has expanded to include Nebraska, most of South Dakota and parts of North Dakota which are currently included in MROW, including the addition of the Western Area Power Administration (or WAPA).
- The MISO region now includes the SRGW and SRDA regions.
- PJM now includes Dominion Virginia Power which was previously in SRVC.

The optimal set of new regions should reflect the new boundaries. Larger regions could be split into smaller subregions by resource zones used to determine location-specific prices within each region. Each zone includes one or more utilities.

---

[2] https://www.ferc.gov/market-oversight/mkt-electric/overview.asp.

Figure 4.2. Regional Transmission Organizations

Source: Federal Energy Regulatory Commission, *https://www.ferc.gov/industries/electric/indus-act/rto/elec-ovr-rto-map.pdf*.
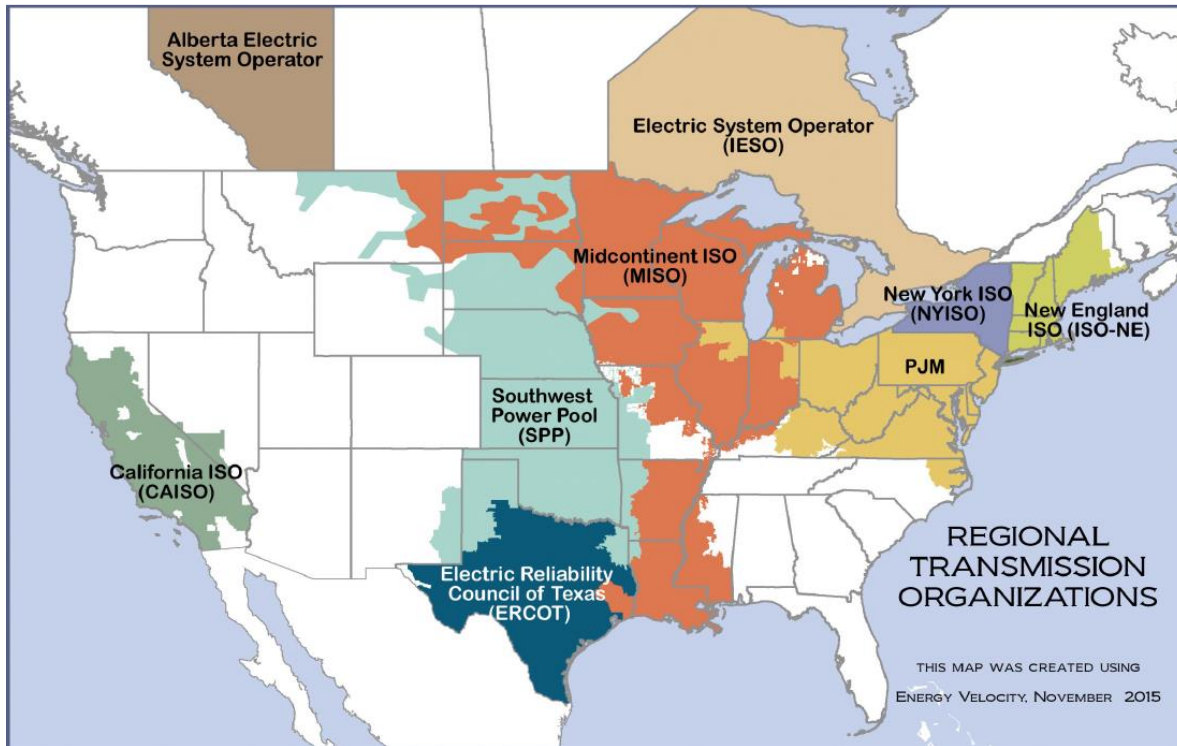
4.4 NERC Regions and Subregions

The North American Electric Reliability Corporation (NERC) is responsible for maintaining the reliability of the North American bulk power system. The organization divides the power system into two sets of regions: 8 Regional Entities (shown in Figure 4.3) and 14 U.S. reliability assessment areas (excluding Canadian provinces) which are generally subregions of the 8 Regional Entities (shown in Figure 4.4). A few assessment areas (such as MISO and SPP) span more than one Regional Entity but are assigned to a single entity for reporting purposes (for example, MRO is responsible for coordinating information for MISO). The assessment areas reflect ISO/RTO boundaries where applicable.

The NERC assessment areas were the primary drivers for the last EMM re-regionalization into 22 regions, and the updated boundaries can also be used in helping determine the new regionality. Changes in boundaries since then include:

- MISO and SPP expansions as discussed in Section 4.3 above
- PJM boundaries now include Dominion Virginia Power, so the new SERC East region now includes only North and South Carolina (vs. VACAR which included Virginia)
- There are now 3 SERC subregions instead of 5, since 2 of the old regions (SRGW and SRDA) are now part of MISO

4.5 Electric transmission requirements and constraints

Transmission requirements are a function of the location of electricity load relative to the generators that can satisfy load requirements. Many load centers are large cities, where the siting of power plants can be problematic. In addition, power plants need access to resources such as water and fuel, which are often located at a distance from load centers. At times, lower cost generation can be considered even if at a greater distance. In some locations, existing transmission capacity constrains the ability of generators in one area to meet load requirements in another area. A classic example is New York City, which has high electricity demand but many of the generators serving this load are located in other parts of the state or in neighboring states, which puts a strain on the transmission capacity entering the city. For this reason, the New York Independent System Operator (NYISO) implemented Locational Capacity Requirements (LCRs) for New York City and Long Island to ensure that sufficient capacity is available locally to maintain system reliability. The locations of transmission bottlenecks and LCRs should be considered when choosing the optimal EMM regionality, given that they can impact where new power plants are located and may signal the need for new or expanded transmission capacity.

4.6 Retail electricity pricing

While wholesale prices are regulated by FERC, retail markets and prices are regulated at the state level by state regulatory commissions. Some states provide partial or full retail competition for electricity service where consumers can choose an alternative power supplier or continue service with the traditional utility provider. Figure 4.5 illustrates a range of different levels of regulation across the United States, with some states that have retail competition (a.k.a., retail choice), some that have traditional rate-based pricing, and a few states where retail competition existed in the 1990s but was suspended after the California Electricity Crisis in 2000–2001.[3] Regulations vary even among competitive states, with some states offering retail choice to commercial and industrial customers only, and others limiting the number of customers or alternative suppliers that can participate. For example, Michigan has fully regulated rates, but up to 10% of load can be served by retail choice.

In the current EMM, each of the 22 regions reflects different levels of regulation in the component states and service territories, with some regions that are fully competitive, some that are fully regulated, and the majority that are a mix of competitive and regulated states and utilities. The new regionality could be used to reduce the number of mixed competitive/regulated regions where appropriate. For example, Ohio could be combined with RFCE (Reliability First Corporation/ East), which is also fully competitive.

---

[3] Electricity Markets Research Foundation, "Retail Choice in Electricity: What Have We Learned in 20 Years?" prepared by Christensen Associates Energy Consulting LLC, February 11, 2016.

Figure 4.3. NERC Regional Entities

Figure 4.4. NERC Assessment Areas

Source: North American Electric Reliability Corporation (NERC), "2016 Summer Reliability Assessment."

Figure 4.5. Status of retail choice

Green – Retail Choice State, Red – Suspended State, White – Traditional State

Source: Electricity Markets Research Foundation, "Retail Choice in Electricity: What Have We Learned in 20 Years?" February 11, 2016.

While utility pricing is regulated at the state level, programs such as "green pricing" are offered at the utility level. Green pricing is an optional program offered by many utilities where customers can choose to support greater investment in renewable capacity by paying a premium on their utility bill. The Federal Trade Commission provides guidance on utility green programs, which exist in many parts of the country. Net metering programs, where utilities purchase excess power from customers with solar energy systems, can also impact retail electricity prices. These programs have been implemented by states in some cases and by individual utilities in other areas.

4.7 Fuel access and policy requirements for electric generators

Many of the policies reflected in the EMM are implemented at federal, state, local, and state-based regional levels. These include environmental policies such as Clean Air Act requirements, the Regional Greenhouse Gas Initiative (RGGI), California AB32, Renewable Portfolio Standards (RPS), Clean Energy Standards (CES), Energy Efficiency Resource Standards (EERS), net metering policies, and many others. These policies and programs are generally reflected in the model's current EMM regions using weighted averages such as load weighting or generation weighting to create region-specific requirements. The drawback to this modeling approach is that some states within a region may not have the same policy requirements as many of their neighboring states but will be subject to the same requirement as all other states within its region. Similarly, some states with aggressive policies will be reflected in the

model with a less-aggressive average target imposed on the region's states. Boundary issues where EMM regions cross state boundaries also complicate this process. Consideration should be given to the potential impact of key state policies when drawing the boundaries of the new set of regions even if the regionality is not strictly state-based. For example, we may want to treat California separately from its neighboring states in any re-regionalization scheme due to its generally more aggressive policies.

In the process of changing EMM regionality, it is important to consider the model's interactions between EMM regions and fuel regions (see Figure 4.6). As mentioned above, the 23 fuel regions are defined by a combination of coal demand regions, natural gas demand regions, Census regions and four CO_2 market regions, all of which use state boundaries. The CO_2 regions were added to account for state and regional carbon abatement initiatives including the Regional Greenhouse Gas Initiative (RGGI) in the Northeast and AB32 in California. Most of the currently identified fuel regions combine one to three states, and the largest is the New England region which combines the six New England states. All existing units in the plant file are specified by fuel region and EMM region. New fossil units are located within both fuel and EMM regions. Fossil units purchase fuel and emit pollutants in geographic regions (i.e., fuel regions) and serve specific EMM regions. Currently, non-fossil resources (renewable and nuclear) are identified by EMM region.

The renewable resources should be geographically defined at the finest level of detail that the data allows. Because the resource data is tied to the geographical location and not to an association with the electrical grid, the data should be aggregated to a geographical region. The existing and new renewable generators would then be connected to both the EMM regions they serve and the geographical region where the resource is located. If the geographical regions share the definition of the fuel regions, the generation of both fossil units and renewable units can be accounted for at both the fuel region and EMM regions. Modeling the state and regional policies would then be much easier to accomplish because both the policies and the resources would be defined by state boundaries. For example, the RPS requirement is specified as a percent of sales which will connect the policy to the EMM regions served, but all the renewable resources needed to comply with the RPS will be built in the fuel region (either a state or collection of states) covered by the requirement instead of potentially across the border due to EMM regional boundaries. This method would also ensure that both fossil and non-fossil generators have a consistent view of the policy requirements.

Specifying renewable resources and RPS policies by fuel region would work well with the existing fuel regions with the exception of the treatment of wind resources in the "WS" region that includes Texas, Oklahoma, Louisiana and Arkansas. Texas and Oklahoma have abundant wind resources whereas the other two states do not have as many opportunities. This suggests that it might be advantageous to split this large fuel region into two regions to accommodate the disparity in available wind resources. While this one exception might argue against the switch to fuel regions, we believe that the more straightforward approach is to specify all resources (not just fossil resources) geographically at the fuel region level.

Figure 4.6. NEMS Fuel Regions
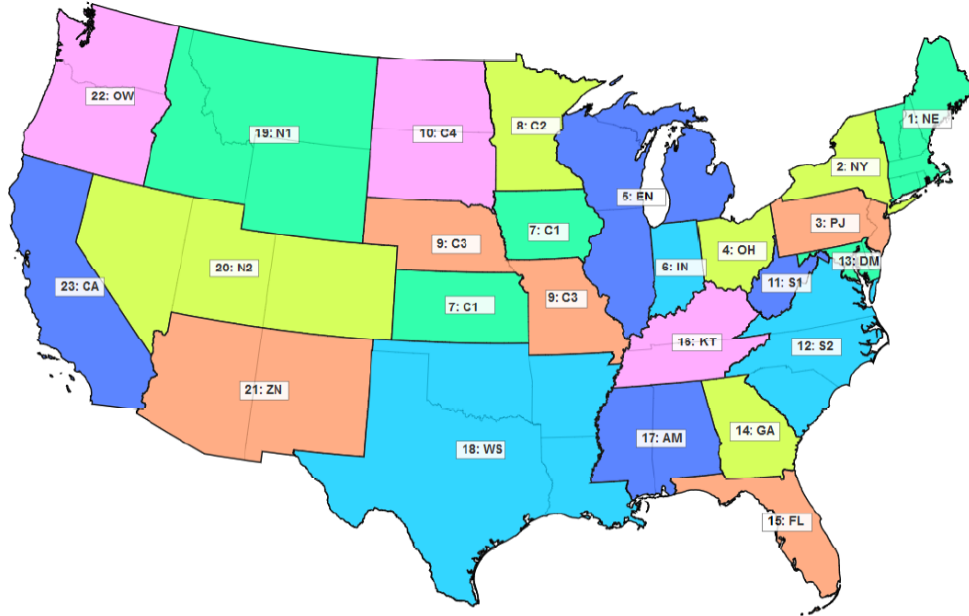
Source: Map created by OnLocation using data from AEO 2015.

## 4.8 Location of non-fossil energy sources

The availability of non-fossil energy sources, including wind, solar, geothermal, and hydropower, vary considerably by region. Wind resources are currently specified in the model by EMM region and by wind class based on data provided by the National Renewable Energy Laboratory (NREL), and regional solar capacity factor profiles were also provided by NREL. The NREL Regional Energy Deployment System (ReEDS) model has 356 wind and solar resource regions[4] (see Figure 4.7), so the relevant input files could be specified at this level of aggregation or, with guidance from NREL, these resource regions could be mapped in a variety of different ways to suit the needs of the EMM. Geothermal and hydropower sites are specified primarily at the state level. Geothermal sites are primarily located in the western (WECC) regions. A large percentage of hydropower sites exist in the NWPP region, especially in Washington and Oregon. Biomass units are treated the same as the fossil units and use fuel regions for consumption and serve EMM regions.

As noted in Section 4.7, the renewable resources should be geographically defined at the finest level of detail that the data allows and should be aggregated to a geographical region (i.e., fuel region) rather than to the EMM region they serve. An added advantage is that the definition of the EMM regions would no longer need to be tied to the location of renewable resources, especially wind and solar. This would require modifying the wind and solar input files, the RFM code that reads and processes the data,

---

[4] National Renewable Energy Laboratory, "Regional Energy Deployment System (ReEDS) Model Documentation: Version 2016," NREL/TP-6A20-67067, November 2016.

and the communication between the EMM and RFM. Geothermal and hydro input files are already site specific with a defined location. This is true for both existing and potential new units.

**Figure 4.7. ReEDS model regional structure**
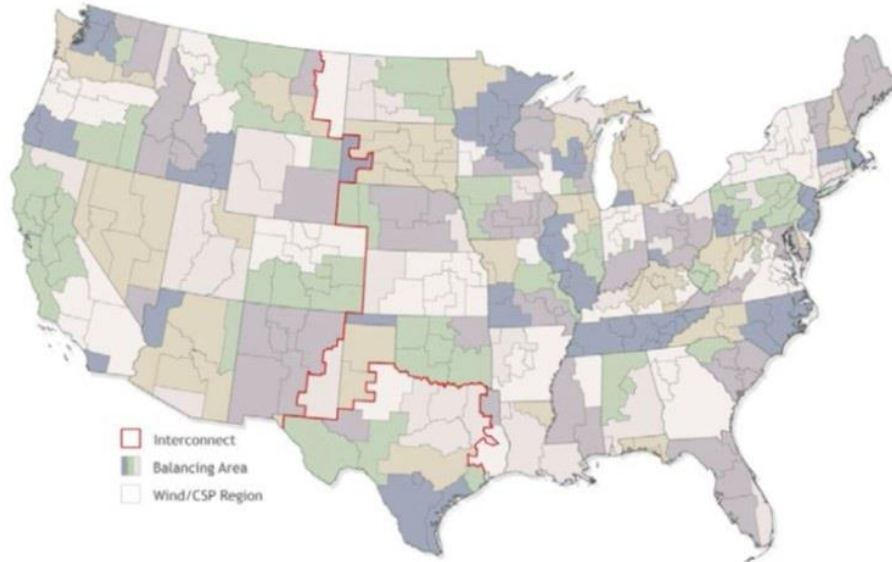
ReEDS includes 3 interconnections, 134 model BAs, and 356 wind and CSP resource regions.

4.9 Relative region sizes

Another consideration is the relative size of the regions. If each EMM region is a collection of load serving entities, then we can measure the size of each region using retail electricity sales. Figure 4.8 illustrates the size of the current 22 regions, which range from less than 1% (NYLI) to 14% (RFCW) of sales. Since the model does not build transmission capacity within a region, transmission investment may be underestimated if the regions are too large. If regions are too small, however, transmission investment could be overstated.

Figure 4.8. Retail sales by current EMM region in 2015
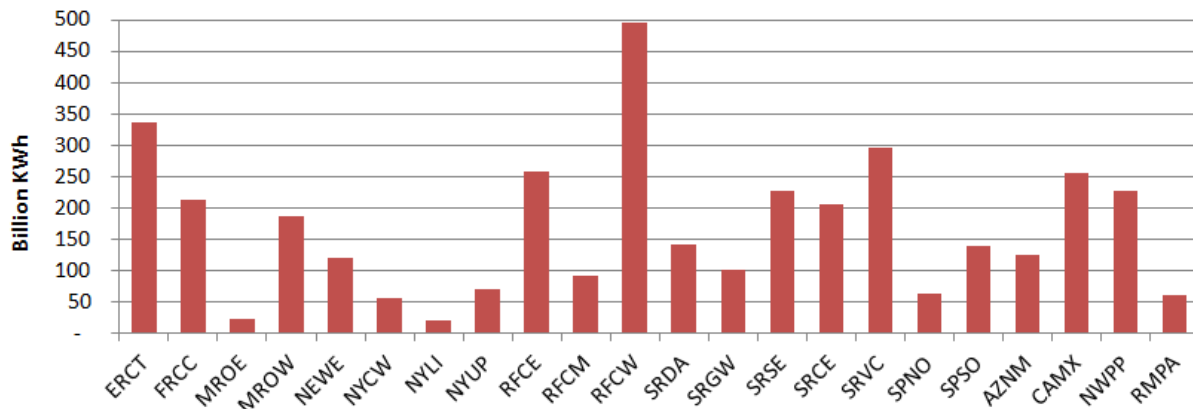
Source: EIA Annual Energy Outlook 2017

4.10 Model performance and other considerations

As noted above, the EMM currently operates with 22 regions. While it can likely handle more than 22 regions without a significant reduction in processing speed, it cannot operate at the full state or balancing authority level without some impact on model performance. The level of aggregation is important for reasons beyond those related to model performance. While too many small regions can impact model performance, regions that are too large will be unable to represent many of the complexities of the electricity system. For example, a model with only one national region would not be able to adequately represent issues such as transmission needs and bottlenecks, the considerable variation in fuel prices and generation mix across the country, policy requirements that impact some parts of the country but not others, and other considerations. The optimal regionality would allow many of these issues to be addressed.

Allowing each balancing authority area to be a separate EMM region and each state a separate fuel region would create a very easily understandable, defensible, and flexible system to examine new and existing policy scenarios. However, the resulting ECP and the Electricity Fuel Dispatch (EFD) LPs would exceed the available memory in NEMS and solution times would likely be excessive even if it were possible to get the models to fit in the available memory. Currently, there are 22 EMM regions and 23 fuel regions and the solution times are in an acceptable range. Increasing the number of either or both of the regions will increase both the EFD and ECP matrices, but the exact impact on size is still hard to predict. In the ECP the key is the number of EMM / fuel region pairs and the number of plant types associated with each pair. In the EFD, increases in the number of regions will have less impact because the dispatch is already more closely tied to the number of individual units. Units are combined into groups where fuel region and EMM region are part of the grouping criteria, but the number of additional groups should be secondary to the total number of units.

Solution times are even more difficult to predict. In general, a linear increase in matrix size can increase solution times geometrically. That is the general rule, but actual increases in solution time still depend on the difficulty in finding a solution. If there are many solutions at or near the optimal, solution times are likely to be large. If there are few solutions near the optimal solution, then times can be dramatically less. The best advice is to experiment. If regional compositions are flexibly defined, the cost of experimentation can be small. Because the model has changed considerably over the years, we cannot know exactly what the impact of adding more regions might be. But as a point of reference, when the number of regions was changed from 13 to 22 (which is a 70% increase in the number of regions), the solution time increased 18%. This estimate applies to the current OML version; it is even more difficult to predict how solution times will change if the EMM is converted to an AIMMS platform.

In order to improve the model's flexibility to handle additional EMM regions without a time penalty and to simplify the new regional structure, some streamlining of the existing structure is recommended.

1. One recommendation is to reevaluate the current fuel regions because the impact on performance is at least partially dependent on the number of combined fuel and EMM regions. The main issue is the relationship with the fuel region definition and the ECP's selection of EMM region and fuel region for endogenously decided new generation units. For example, if an EMM region has a choice of fuel regions for a new natural gas combined cycle unit and these fuel regions share a common natural gas region, then the natural gas prices will be identical. In this situation the ECP LP will be indifferent between the two fuel regions. This allows the ECP to randomly assign the fuel region and makes comparison of the resulting natural gas consumption or related carbon emissions between scenarios unreliable, if viewed by fuel region. Currently, there are three sets of regions in the Midwest that share not only the same natural gas regions but also the same coal regions and census regions: EN/IN, C1/C3, and C2/C4 (see Figure 4.6). The only difference is the "carbon" region; however, both carbon regions are currently non-constraining. At the time the carbon regions were defined, there was a proposal to create a regional carbon limit. As of yet, that proposal has not been implemented, so these three sets of fuel regions receive exactly the same fuel prices. At a minimum, the extra carbon region should be eliminated and the fuel regions combined, which would also free up some space in the LP to expand the number of EMM regions.

2. However, recommendation #1 does not eliminate the issue for individual generator types across the remaining fuel regions, if the technology uses a fuel that is identically identified over multiple fuel regions. The challenge is to eliminate the ECP indifference between fuel regions. This can be accomplished in a number of ways. The supply modules (CMM and NGTDM) could be expanded to provide prices at a more disaggregated regional level. Since the NGTDM module is currently being modified, some coordination is needed when making regionality choices for the NGTDM and EMM. The EMM can specify mark-ups / mark-downs to the fuel prices provided by the supply modules in order to establish a preferred EMM / fuel region pairing for new units. The factors used to regionalize new generator investment costs are currently specified by EMM regions. If these factors are instead provided by fuel regions this would also differentiate between similar investments in alternative fuel regions.

3. The cost to connect new units located in a fuel region to the transmission grid that serves an EMM region is another way to differentiate investment cost in a specified fuel region. In the EMM interregional transmission data is specified between adjacent EMM regions. The data includes an inventory of current transmission capacity, contract flows and the costs to add new transmission capacity. To facilitate flexible region definitions, the transmission data needs to be kept by the entities with transmission capacity. Mapping these entities to EMM regions defines capacity and contract flow between EMM regions. The cost of new transmission between EMM regions is problematic because, at least in part, the costs depend on the distance between regions, the population density of the regions and the terrain. Currently, the costs are manually assigned for adjacent regions. Full automation will require the development of an algorithm to calculate distances between regions and combine density and terrain indexes to estimate new transmission costs.

4. Transmission lines across international borders have similar issues. Capability and contract flows come from the same sources. The economy sales currently flow between Canadian provinces and the EMM regions. To maintain this capability requires attaching a province to each transmission line that crosses between the United States and Canada. Currently, there are no economic flows to/from Mexico.

5. Another option for streamlining is to reduce the size of the matrix by eliminating the Mercury and Air Toxics Standards (MATS) compliance decisions from the LP. The MATS compliance deadline has passed so the retrofit decisions have been completed and should not be changed in the model. Eliminating these retrofit options for coal plants would simplify the matrix and reduce solution times. We would preserve the SO_2 and NOx retrofit options In the event that further changes are made to these air emission requirements, but the differences in ECP types related to particulate equipment (e.g., electrostatic precipitators and fabric filters) may no longer be needed now that the MATS decisions have been completed. Collapsing these ECP types and retrofit options would simplify the existing coal types and reduce the model's retrofit decision matrix.

4.11 Regionality of other electricity models

It may be instructive to compare the regionality of other U.S. electricity market models in use. Table 4.1 lists some common models along with their level of regionality. The models shown here employ a variety of regional definitions. Some models (DIEM, FACETS and US-REGEN) define regions based on states, other models (IPM, Haiku) build their regions around NERC regions, and some models (ReEDS, NewERA) use electricity market regions or balancing authority areas. In some cases a combination of these regional definitions is used (e.g., the IPM and ReEDS models).

Table 4.2. Regionality of select U.S. electricity models

Model Name	Primary User/Client	Regionality
Regional Energy Deployment System (ReEDS)	National Renewable Energy Laboratory	134 balancing authority area subregions divided by state boundaries
Integrated Planning Model (IPM)	EPA / ICF International	64 regions based on NERC regions and RTO/ISO boundaries
Dynamic Integrated Economy/Energy/Emissions Model (DIEM)	Nicholas Institute for Environmental Policy Solutions at Duke University	14 regions defined along state lines that approximate distinct electricity markets
Framework for Analysis of Climate-Energy-Technology Systems (FACETS)	Amit Kanudia and Evelyn Wright	MARKAL/TIMES framework with flexible multi-region structure built from the state level up
Haiku	Resources for the Future	13 NERC subregions circa 1999
NewERA Electricity Sector Model	NERA Economic Consulting	32 electricity market regions
U.S. Regional Economy, Greenhouse Gas, and Energy (US-REGEN)	Electric Power Research Institute	15 regions but can be configured at any aggregation of the lower 48 states

4.12 Recommendation for new regionality

As noted above, the recommended regionality involves defining EMM regions as collections of load serving entities and defines fuel regions as collections of states. This construct, which maintains the original concept of the two sets of regions, provides the most flexibility for meeting regulatory policy requirements, ensuring grid connectivity, and matching generators to load requirements. Defining the EMM regions as a collection of states would be problematic because the utilities and balancing authorities, which are the fundamental building blocks of the power grids, cross state boundaries.

The recommended approach involves redesigning the data inputs to be represented at an elemental level in order to facilitate a flexible approach. Location-related fields that should be included in the stored elemental data are state, utility, balancing authority, and RTO/ISO where available. Streamlining the processes needed to achieve maximum flexibility is also an important step, including reducing the size of the solution matrix, which will increase the number of regions that the model can accommodate.

If we choose to implement a flexible approach, a final decision on optimal regionality is not needed. However, if we implement the phased approach, a decision will need to be made on a new static regional definition. That decision must take into account the likely uses of the model going forward, factoring in the policies, energy sources, market changes, and other considerations outlined above that will drive future scenarios.

Our recommended set of 25 new static EMM regions and their descriptions are summarized in Table 4.3. The table lists the name and description of each new region and how it relates to the current regionality. The new regions follow NERC assessment region and ISO region boundaries, with some larger regions split into smaller subregions based on each region's pricing zones and/or balancing authorities. The approach involved using the latest EIA-861 data to evaluate the size (i.e., retail sales) of each of the current NERC regions and subregions. This information was then used to determine which regions were candidates for splitting up further into smaller regions while maintaining the boundaries of the current ISOs and NERC assessment areas. A combination of NERC assessment regions and pricing zones within ISOs along with transmission information was used to split the larger regions.

Following is a description of each set of regions:

- New England (NEWE), Florida (FRCC), and Texas (ERCT) were not changed. Since ERCT is a large region, we considered splitting the state into 2 regions: the western and northern (primarily Dallas/Fort Worth) portions of the state, and the southern and eastern (primarily Houston) portions. This split would result in 2 regions with large load centers and place most of the existing wind generation (located in the west) in one of the two regions. However, because Texas supports full retail competition with most of its retail sales sold by non-utility service providers, there does not appear to be sufficient data to divide retail sales between regions of the state. Therefore it is recommended that ERCT remain as a single region.
- California has 2 major load centers: San Francisco (north) and Los Angeles (south) which were used to split the CAMX region (see Figure 4.10). The northern region would include balancing authorities BANC and TIDC as well as utilities Sacramento Municipal Utility District (SMUD) and Pacific Gas and Electric Company from the California ISO, and the southern region would include balancing authorities LADWP and IID as well as utilities San Diego Gas & Electric and Southern California Edison from the California ISO. The 2 subregions would be treated as a single region for state-wide policies such as SB32 (the extension of AB32).
- The New York ISO was split into 2 regions instead of the current 3 regions. Because the challenges facing NYC-Westchester and Long Island such as transmission limitations and siting limitations for new power plants are similar from the model's perspective, we recommend combining these 2 regions.
- The new boundaries of PJM were used to replace the RFCE and RFCW regions. This expanded ISO was split into 4 regions that are collections of utilities which also act as pricing zones: eastern PJM (eastern PA, NJ, Delmarva), western PJM (west PA, OH, WV, part of KY), Commonwealth Edison, and Dominion Virginia Power. See Figure 4.11 for a map of the utilities/zones included in PJM.
- The expanded MISO region, which includes MROE, SRGW and SRDA and parts of MROW and RFCW (Indiana), was split into 4 regions based on the ISOs pricing zones (see Figure 4.12).
- The expanded SPP region, which includes SPNO, SPSO and the western part of MROW, was split into 3 regions. The boundaries of these new regions were determined by collections of utilities as shown in Figure 4.13.

- The SERC region no longer includes SRGW and SRDA, so it was reduced from 5 regions to 3 regions. In addition, the SRVC region no longer includes Dominion VA Power which joined PJM. The new regions maintain the original boundaries of SRCE and SRSE, and reflect the new boundaries of SRVC. The 3 new regions match the NERC assessment regions (see Figure 4.4 above).
- WECC was split into 6 regions to replace the existing 4 regions. AZNM (now called SRSG), and RMPA (now called RMRG) remain unchanged, California was split into 2 regions (CANO and CASO), and NWPP was split into 2 regions: the Basin area (largely the utilities in NV, UT, ID), and NWPP (the remaining utilities in the former NWPP) as shown in Figure 4.14.

A rough estimate of the retail sales for each of these new regions was calculated using the 2015 EIA-861 data is shown in Figure 4.9. The sizes of the current regions (illustrated in Figure 4.8) range from a minimum of 21 BKWh or 0.6% of total sales (NYLI) to a maximum of 495 BKWh or 13% of sales (RFCW). The new regions have a smaller range of sizes, from a minimum of 46 BKWh or 1.3% of sales (SPPN) to a maximum of 358 BKWh or 9.7% of sales (TRE). Note that the scale of the new figure has been adjusted to be the same as the prior figure illustrating the size of the current regions for comparison purposes.

Figure 4.9. Retail sales by new EMM region
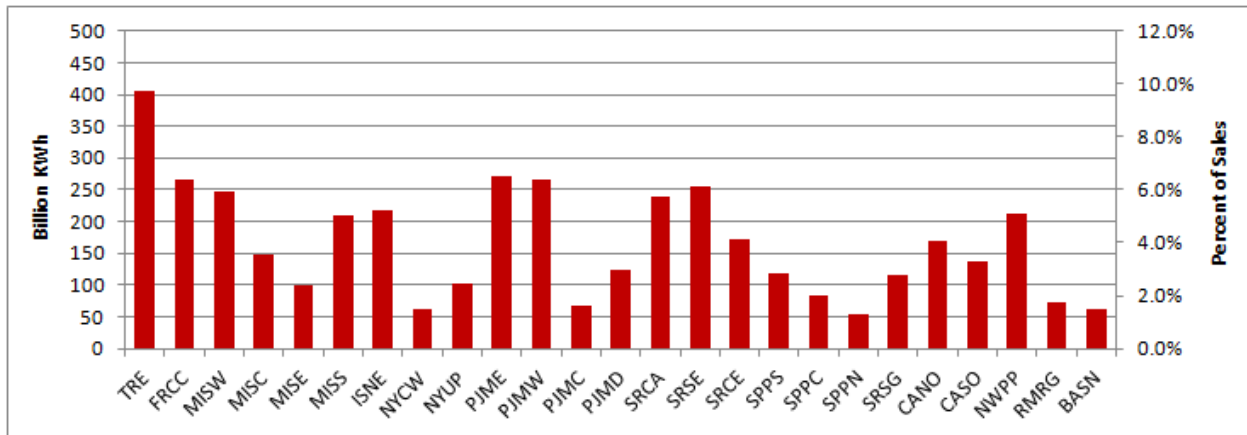
Source: EIA-861 data for year 2015

Table 4.3. New EMM regions

	New EMM Regions	Description	Boundaries	NERC Regional Entity	Old EMM Regions	% of Total Retail Sales
1	TRE	Texas Reliability Entity	Texas Reliability Entity	TRE	ERCT	9.7%
2	FRCC	Florida Reliability Coordinating Council	Florida Reliability Coordinating Council	FRCC	FRCC	6.4%
3	MISW	MISO-West	MRO Zones 1, 2, 3	MRO	MROE + part of MROW	6.0%
4	MISC	MISO-Central	MRO Zones 4, 5, 6	MRO	SRGW + parts of MROW and RFCW	3.5%
5	MISE	MISO-East	MRO Zone 7	MRO	RFCM	2.4%
6	MISS	MISO-South	MRO Zones 8, 9, 10	MRO	SRDA	5.0%
7	ISNE	ISO New England	ISO New England	NPCC	NEWE	5.2%
8	NYCW	NYC-Westchester & Long Island	Consolidated Edison and LILCO	NPCC	NYCW + NYLI	1.5%
9	NYUP	Upstate New York	Rest of NY	NPCC	NYUP	2.4%
10	PJME	PJM East	PJM major utilities - PENELEC, PPL, METED, PECO, JCPL, PSEG, RE, BGE, Delmarva, PEPCO, ACE	PJM	RFCE	6.5%
11	PJMW	PJM West	PJM major utilities - APS, AEP, ATSI, DPL, Duke, DuQ, LGEE	PJM	part of RFCW	6.4%
12	PJMC	PJM ComEd	PJM - Commonwealth Edison	PJM	part of RFCW	1.6%
13	PJMD	PJM Dominion	PJM - Dominion Virginia Power	PJM	part of SRVC	3.0%
14	SRCA	SERC - East	SERC - North and South Carolina	SERC	part of SRVC	5.7%
15	SRSE	SERC - Southeast	SERC - Southeast	SERC	SRSE	6.1%
16	SRCE	SERC - Central	SERC - TVA	SERC	SRCE	4.1%
17	SPPS	SPP - South	SPP major utilities - Oklahoma G&E, SWPS, AEP West	SPP	SPSO	2.8%
18	SPPC	SPP - Central	SPP major utilities - Sunflower, Westar, Kansas City, Empire	SPP	SPNO	2.0%
19	SPPN	SPP - North	SPP major utilities - WAPA, Basin Electric, Heartland, Nebraska, Omaha	SPP	part of MROW	1.3%
20	SRSG	SRSG	WECC - Desert Southwest	WECC	AZNM	2.8%
21	CANO	CA - North	WECC - California North	WECC	part of CAMX	4.0%
22	CASO	CA - South	WECC - California South	WECC	part of CAMX	3.3%
23	NWPP	NWPP	WECC - Northwest	WECC	part of NWPP	5.1%
24	RMRG	RMRG	WECC - Rockies	WECC	RMPA	1.7%
25	BASN	BASIN	WECC - Basin subregion	WECC	part of NWPP	1.5%

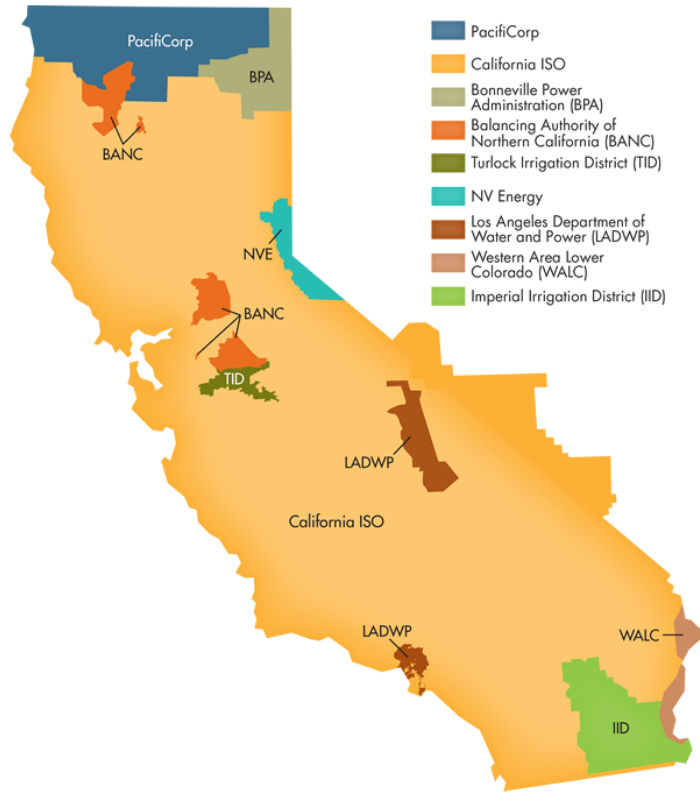**Figure 4.10. California ISO (CAISO) balancing authorities map**

Figure 4.11. PJM zonal map

Figure 4.12. MISO zonal map

Local Resource Zones

LRZ	Local Balancing Authorities
1	DPC, GRE, MDU, MP, NSP, OTP, SMP
2	ALTE, MGE, MIUP, UPPC, WEC, WPS
3	ALTW, MEC, MPW
4	AMIL, CWLP, SIPC
5	AMMO, CWLD
6	BREC, DUK(IN), HE, IPL, NIPSCO, SIGE
7	CONS, DECO
8	EAI
9	CLEC, EES, LAFA, LAGN, LEPA
10	EMI, SME

Source: MISO

Figure 4.13. SPP zonal map
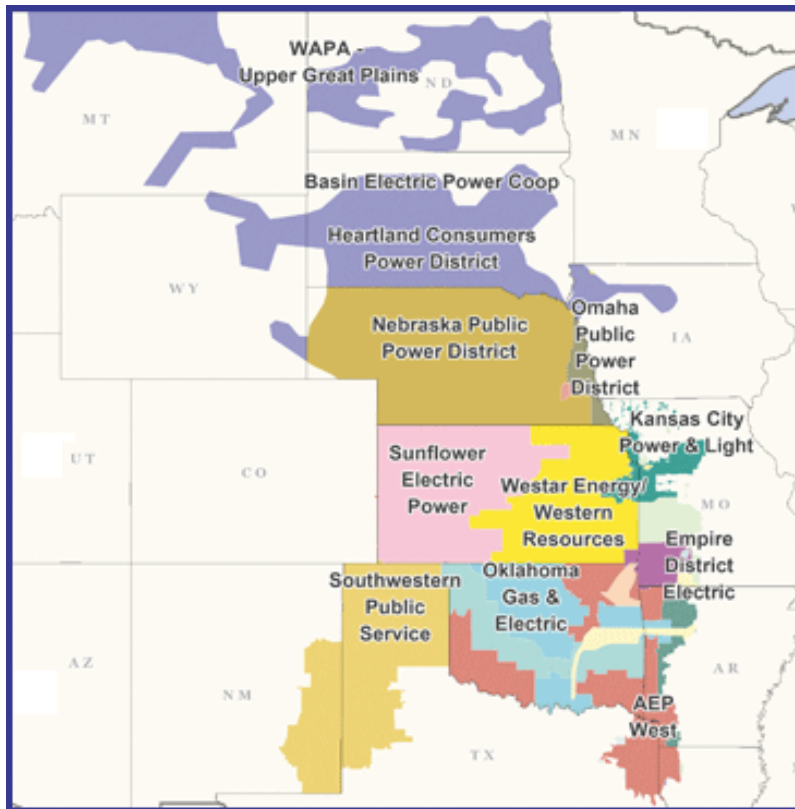
Figure 4.14. NERC Regional Entities and Regional Councils (2010)

Source: https://www.eia.gov/electricity/data/eia411/
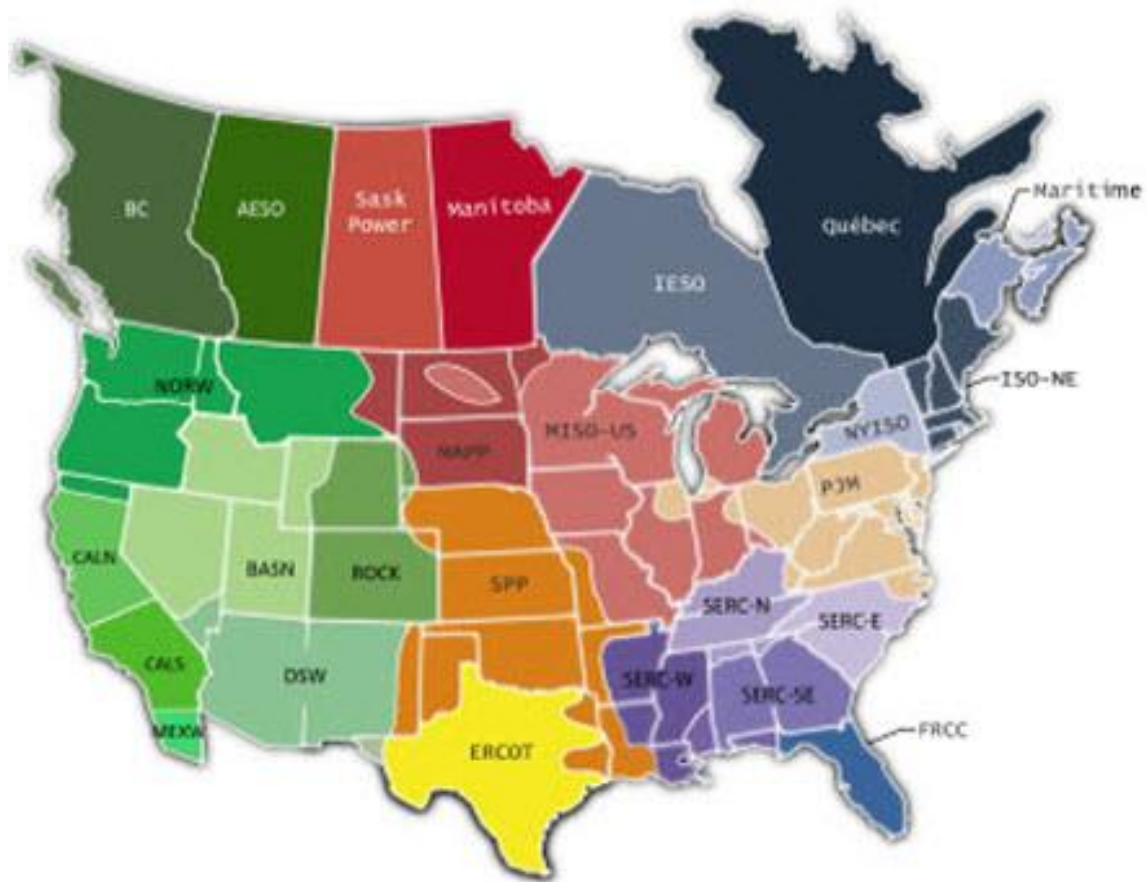
5. Implementing the EMM Re-regionalization

The three phases of implementation will be discussed in detail in this section.

- The first phase would be similar to the approach taken in the prior re-regionalization effort where the model structure would be modified to accommodate a fixed, alternative regionality with minimal changes to how data are processed.
- The second phase would maintain a static approach to the new regionality but would involve altering the model processing to enable a future flexible approach
- The third phase, if implemented, would build upon the structure put into place in the first and second phases to replace the static regional mapping with a dynamic flexible mapping structure.

5.1 Phase I – Maintain current regional design

Appendix C describes the steps needed to implement redefined regions, that is, to update the data input files, source code, and includes for a new regionality. It uses a similar approach to what was done last time.   Note that some updates are more involved than others.

5.2 Phase II – Streamlined static regional design

The second phase is a continuation of the static approach outlined in Phase I—meaning that one new set of regions will be defined and used. However, this phase is a precursor to a flexible approach in that it modifies how the input data files are organized in order to facilitate a set of flexible regions in the future. The concept here is to transfer all EMM regional data from all input files to database tables. For example, all of the data tables at the EMM regional level would be moved from the ECPDAT file, leaving behind those that are at the ECP technology and other levels, and all 20 or so tables in the EMMCNTL file would be moved, and so on.

In addition, all data that are used to produce these EMM regional tables would be stored in tables at their most elemental level. Then a program could be developed that would process the elemental level data to the EMM regional level. A program would also facilitate allowing the user to set assumption data for those tables that require regional assumptions.

Finally, a program would be written that would read these EMM regional tables (or Comma-Separated Value (CSV) files or whatever format is decided) to bring the data into the EMM (and RFM).

5.3 Phase III – Flexible regional design

The proposed Phase III modeling approach to the EMM is a data-driven concept built around flexibility, transparency, and a new GIS browser-based interface.

The core concept of this new design is to store all model data at the finest level of regional aggregation possible, as proposed in Phase II. In the case of the EMM, this finest level of regional aggregation that

could be attached to each data item in the model may be the power control areas (PCAs), also known as balancing authority areas. In some cases, it is the NERC region or subregion, and many sources also include state. Each physical item in the data set will have a set of coordinates attached, along with other defining information so that it can be displayed visually in the GIS system.

By storing the key data in a GIS-focused database, the datasets can be analyzed by modern geographic information systems that allow the user to inspect, verify, and validate the datasets. Updating the inputs to the model is then an exercise of updating and maintaining the GIS data. Furthermore, this approach lends itself to distributed upkeep by allowing subject matter (ECP, EFD, EFP, ELD, RFM) experts to be responsible for the content of each dataset, while technology experts maintain those datasets and manage their input into the EMM. Finally, to the extent applicable, Application Program Interfaces (APIs) can be employed to periodically refresh the datasets automatically from online sources[5]. The resultant regional datasets can be displayed graphically and investigated by the modelers to validate and verify the model specification. By using the spatial anchoring and GIS framework we can auto-generate the abstract logical model, which then can be used to generate and solve the EMM linear program (LP).

As long as each entity in the network is geographically located in the GIS, we can add any aggregation scheme needed, ad hoc, and without re-collecting or estimating any data. The values can be "rolled up" to the higher level of aggregation by summing all the constituent parts. The network flow then can populate an optimization model at the required resolution and can run immediately with no further editing needed by a user.

There are two primary advantages of modeling at the finest level of geospatial detail. The first is that the graph algorithms benefit from the specificity of the more accurate graph, and then can report results at an aggregate level. Calculating at the balancing authority level and then aggregating to larger regional aggregations yields better results than calculating at the larger level from the start. The second advantage is that the more disaggregated model can be "rolled up" or aggregated to whatever reporting level is appropriate. The initial graph optimization model does not need to be altered or re-run to accommodate these various on-the-fly reporting aggregations.

While there is a myriad of approaches to addressing this data modeling problem, we believe the combination of a geographical information system, auto-generation of input datasets, and updated LP matrix generation algorithms will provide the benefits outlined in the following sections.

5.3.1 Flexible LP modeling framework

The EMM LP will be written in a data-driven framework so that changes to key problem dimensions such as regionality would require no changes to the code describing the LP variables, constraints, and objective function. This will allow the users to focus on the preparation of alternate datasets and analysis of model results instead of spending time making technical changes to the core LP model.

---

[5] The EIA Open Data API found at https://www.eia.gov/opendata/ is one example of an online data source around which automated data updates could be created.

Taking this approach, the model will support evolutionary development, early prototype delivery, and continuous improvement and encourage rapid and flexible response to change.

5.3.2 Responsive software design

Our proposed approach is to use the open-source language Python, in conjunction with GIS software and SQLite database software to build out the EMM input files. We will separate the raw data from the model inputs so that alternative sets of model inputs may be created easily. By taking this approach, we will be able to test alternative levels of aggregation quickly and with relative ease.

Python has an extensive set of libraries dealing with the reading and writing of data in various formats, manipulation of data, and connection to various databases. Our recommendation is to use SQLite, as it is already in use at EIA (in the WEPS+ framework, for example). We will use Python to read in and prepare the model input data. The user-specified parameters will indicate the model's regional aggregation level and the datasets will be prepared accordingly.

All of the datasets can be kept in CSV format, which provides the following advantages:

- CSV files are ubiquitous and nearly every software package and database system has built-in capability for import and export.
- CSV files are easy to compare for differences and therefore naturally lend themselves to version control systems such as SVN and Git.

While the raw form of many of the EMM tables will remain as CSV files, the plan is to implement a system that brings all of the data that are used by the EMM (including the restart variables) into an SQLite database, and that will have a front end that allows the user to select parameters to be used to group the data into whatever regional aggregation is desired. Once the user is satisfied that the regional setup is acceptable, the program will produce the data inputs needed by the EMM LP model.

We recommend that the front-end interface include a GIS that provides feedback to the user regarding the aggregate properties of the user's chosen regional aggregation for validation and verification purposes. In addition, the GIS system can be used to facilitate the "benchmarking" and associated imputation of missing or mis-scaled data.

5.3.3 Nodal geographic representation

As introduced above, in the analysis and optimization of routing and examining flows of commodities, we find it useful to represent the critical entities in that flow as nodes in a network, or graph. A node is simply an entity that has attributes and can be connected to other nodes by way of links. In flow networks, these links are directed, meaning flows only travel in one direction at a time across the link. These node and link representations have important mathematical characteristics and have been studied and formalized since the 1700s, and modern computer science and analytics make use of this past work to analyze and optimize systems. The result of all this background work is that today we can

leverage these systems to perform optimization and organize data about large and complex flow systems using these graph constructs.

In the case of the electricity flow network, our nodes correspond to physical parts of the energy infrastructure such as generating units and high-power transmission lines connecting them.

Figure 5.1 illustrates how the geographic information system can catalog the physical infrastructure and process this into a flow network that is suitable for forecasting with iterative optimization models, at the balancing authority level in this example. These groupings of generating units at the balancing authority level are linked by transmission lines to complete the transport infrastructure. A key element of the computational framework provided by modern geographic information systems is that it allows us to link sources, transport links, and destinations in ways that are not explicitly collected in survey data on global trade. In this way, the resulting models are greatly improved as they more closely relate to actual ground truth, as well as needing to make fewer assumptions in the model. Rather than relying on rules of thumb or analyst's tuning of starting parameters for optimization, the needed parameters are calculated directly from the infrastructure geography and characterizations. This geography can be easily represented as a logical graph by connecting the nodes to the relevant links. In practice this requires collecting data through observation or estimation of the routes through the electricity grid.

Figure 5.1. Balancing authorities and major transmission interfaces
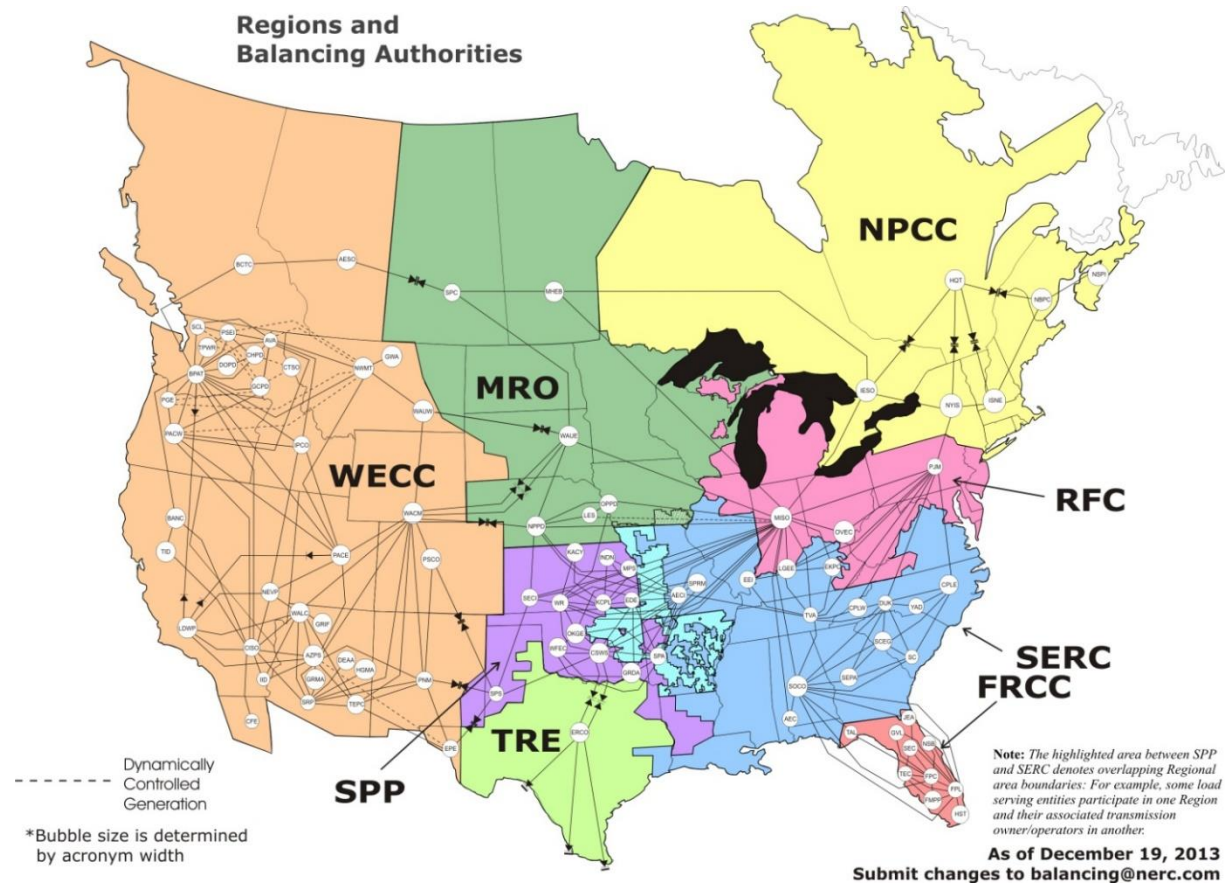
Source: *North American Electric Reliability Corporation*

Updating the data inputs to the model becomes part of the GIS data QA process and is done by a GIS operator, rather than a modeling expert. This allows the task of keeping the inputs current to be done by those working with the data on a regular basis. The GIS approach also lends itself to distributed upkeep, with each region or domain expert responsible for that part of the dataset, rather than one entity tasked to update the entire database. The transmission infrastructure can be maintained separately from the generating units by people who are familiar with each subject matter.

There are currently many sources of data, each representing only a part of the overall picture.

- GIS data
  - Generating units
  - Transmission network nodes
  - ISO, RTO, NERC regions, utility territories
- Historical data
  - Generation
  - Consumption
  - Sales

By leveraging the information held in each of these data sources, they combine to produce the information necessary for a data-driven model. By generating these models automatically and directly from the source data, the inputs, connections, constraints, and outputs of energy supply are established. In this way, we define the system automatically, and eliminate a tedious modeling step.

Using real-world data to generate the network flow model provides the opportunity to quickly go from an analytical question to a computed answer, which can then be used to refine the original questions.

6. Uncertainty and Limitations

There are several issues associated with revised EMM region definitions, and some limitations with using a flexible approach to defining the EMM regions.

The current mapping of PCA (or balancing authorities) to EMM regional definitions was composed from multiple sources across many different years of data, including FERC, EIA, EPA eGRID, and NERC. PCA names and ID numbers were not always consistent across all of these data sources, and some data sources used a name from a prior year before a company merged with another. These mappings can hopefully be improved now that more of the data filings needed (such as the EIA-860 and FERC 714) include balancing authority detail and will likely be more consistently defined, but there is no guarantee that all data sources will align perfectly.

There is still much manual work needed to process data into tables for the EMM input files. Many of the data processes in MDBs and Excel programs are not production worthy and involve many manual steps. In addition, some data are obtained by conducting research on the web or sifting through reports. These things will simply not be able to be automated.

One other limitation is that the FTAB reporting has two tables that are hardwired for the EMM regions (63 and 64). So once a different set of EMM regions is defined, these table definitions need to be modified in the LAYIN/LAYOUT files and the FTAB.f source code needs to be updated to load the information for the reports.

7. Conclusions

Our recommended set of 25 new static EMM regions, described in Table 4.3, follow updated NERC assessment region and ISO region boundaries.   The new boundaries take into consideration the decision criteria outlined in Table 4.1 and offer a reasonable representation of current electricity markets. The new regions result in a modest increase in the number of regions the model currently uses so they should not have a significant impact on model processing time.

The level of effort for each proposed implementation phase will depend on the desired scope of work to be included. We recommend that the phases of the EMM re-regionalization effort include the following:

- **Phase I.a:** Implement an updated set of input files, include files, and source code for the new set of 25 EMM regions using the existing data and existing processes/file configuration. This should be relatively easy to implement as many of the current regions will not change and only a modest number of new regions will be added. Additional time is needed for research on transmission limits and some other parameters needed for the new set of regions. Updating fuel regions and/or carbon regions increases the work effort but is recommended. Removing the MATS structure should also be considered as it should help improve the model's solution time.
- **Phase I.b:** Switch to an updated set of source data for the new set of regions. This step adds a good deal of additional time and effort to the schedule, because it introduces considerably more research, data manipulation and pre-preprocessing that must be done.
- **Phase II:** Streamline the data processing with new programming and a revised file structure and data manipulation. This process is needed to reduce the time needed to change regionality in the future, whether or not we implement a fully flexible design. This phase would be a medium-size effort.
- **Phase III:** Implement a fully flexible design by replacing all of the data files and associated processing routines as well as developing the code for the GIS, and additional updates to the EMM and RFM to use the new data. This would be a very large effort, as it would involve more than the initial streamlined data process from Phase II.

Each phase is designed to result in a fully functioning model. We will work with EIA staff to determine the appropriate implementation schedule, taking into consideration other model development activities and AEO release dates.

APPENDIX A - Processes Used to Create EMM and RFM input files

1. *PLTF860.TXT and PLNTDAF.DAF – Plant data*

The sources of these files are EIA-860 and EIA-923 data, which are organized by boiler, unit, and plant. The owner and operator utility companies, the balancing authority, North American Electric Reliability Corporation (NERC) region, and state associated with the unit are also provided. A mapping table of BA to EMM region is used to attach the EMM region. These data are processed (in SAS) into the PLTF860.TXT file, which is used by the plant preprocessor (prepplt.f), to create the PLNTDAF.DAF file that is input to the EMM.

2. *PRETTIN.TXT – Transmission constraints*

These data used to come from the NERC Summer (and Winter) Assessments, which provided a figure showing Normal Base Electricity Transfers and First Contingency Incremental Transfers (in MWs) between NERC subregions, and was supplemented by other data from the Independent System Operator (ISO) websites and various reports and studies (like the Western Interconnection 2006 Congestion Assessment Study, May 8, 2006). While the NERC Assessments are still published, the figure showing transmission capability stopped being provided after 2004. Data for transmission capability are now updated based on data and reports found on ISO / Regional Transmission Organization (RTO) websites and in other sources, which is a labor-intensive manual process.

3. *ETCNCT.TXT – Firm power contracts*

The NERC Summer (and Winter) Assessments contain net firm power contracts by NERC region. These data are supplemented by researching ISO/RTO websites, FERC Form 1 (Federal Energy Regulatory Commission), the Canadian National Energy Board, and EIA's *Electric Power Monthly*, which, again, is a labor-intensive manual process.

4. *FRMCHRG.TXT – Demand and energy charges*

This file contains manually set demand and energy charges by EMM and Canadian regions.

5. *CANSPLY.TXT – Canadian projects*

This file contains manually set Canadian supply available to the United States organized by Canadian region to EMM region.

A Fortran preprocessor (Prepett.f) reads the PRETTIN.TXT, ETCNCT.TXT, FRMCHRG.TXT, and CANSPLY.TXT files and creates two files. The ETTIN.TXT is a file of transmission constraints that is organized by EMM region and year and is net of firm power contracts. The ETTDEM.DAF file contains firm and economy charges, Canadian projects data, and each EMM region's import and export transmission constraint.

6. LDSMDAF.DAF

The LDSMDAF.DAF contains over 200 representative load shapes (organized by 24 hours by 3 day types for 12 months – or 864 hourly loads) by end use, sector, and NERC region, as well as system load shapes by EMM region.

There are some preprocessor programs (in .NET) that allow a user to replace the system load shapes with updated ones. There are also Fortran/EViews (Econometric Views) programs that use the multiple years of FERC 714 hourly load data along with weather data to weather-normalize the load shapes.

7. LDSMSTR.TXT

The LDSMSTR.TXT file includes a list of load shape representation (LSR) names that each EMM region should use, as well as a table that allows the user to arrange the 864 hours into load slices.

8. CTSINPUT.GDX and CTSSOLN.GDX – Capture, transport, and storage costs for CTS plants

The CTSINPUT.GDX file is produced by running an MDB preprocessor. This file is input to a CTS preprocessor GAMS program that then produces the CTSSOLN.GDX file. While neither GDX file has any data that are divided by EMM region, the MDB does use some tables that are manually organized by EMM region.

9. FINREG.TXT – Financial data

This is a text file that is manually organized and contains multiple tables of data. The tables that were derived from the FERC Form 1, EIA-412, RUS 7 and RUS 12 data were most recently processed in Excel and MDB programs from an Energy Velocity database. The financial data are at the utility level, which is mapped to PCA (power control area, now called balancing authority area), and then mapped to EMM region. The age and difficulty in updating these data are a real problem that needs to be highlighted (EIA-412 data are not available after 2004, and the RUS 7 and 12 data are no longer released from the RUS).

10. ELDATYR.TXT – Historical generation, consumption, and sales

These data are from the EIA-860, EIA-923, and EIA-861, which are at the plant, utility and balancing authority level. The data are processed and mapped to EMM region.

11. EMMCNTL.TXT – EMM control information

This text file contains over 200 tables of data that are manually inserted into the file, approximately 20 of which are organized by EMM region.

12. ECPDAT.XLS – Electricity capacity planning assumptions

This Excel spreadsheet contains 35 tabs that are manually set. The first 13 tabs of assumption data contain multiple tables, about a dozen of which are organized at the EMM region level. The last 22 tabs are organized by a tab for each EMM region.

13. NUGPIPE.TXT – Nonutility history and planned additions

This text file is manually set and includes multiple tables of information, two of which are organized by EMM region.

14. EFPCNTL.TXT – EFP control file

This text file is manually set and includes multiple tables of information, two of which are organized by EMM region.

15. PHASEIN.TXT – Phase-in plans

This text file is manually set and includes multiple tables of information, all of which are organized by EMM region.

16. SALESLB.TXT – Sales-leaseback transactions

This text file is manually set and includes one table of information that is organized by EMM region.

17. EMMDBDOC.XLS / EMMDB.MDB – EMM Access database

This Microsoft Access database contains numerous tables of output data from the ECP, EFD, and EFP. The EMMDBDOC.XLS defines the tables and includes one that provides the EMM regional definitions and is manually updated.

18. LAYOUT.TXT / LAYIN.XLS – NEMS report-writer layout file

The LAYIN.XLS file defines the layouts of the NEMS FTAB reports which are saved to the LAYOUT.TXT file. The XLS file has numerous tables that are organized by EMM region and need to be manually updated when EMM region definitions change.

19. WFLOORS.TXT – Planned capacity additions

This text file is used as a supplement for the PLTF860.TXT file and includes select planned capacity additions, especially renewable units. The additions are specified by EMM region. The file is read by the RFM.

20. WGEPARM.TXT and WGESITE.TXT – Geothermal sites

These text files are used by the RFM to specify geothermal sites that are currently in use (WGEPARM.TXT) and sites that are available for new development (WGESITE.TXT). In both files, the

geothermal sites are specified by state and EMM region. Sources are the National Renewable Energy Laboratory (NREL) and the EIA-860.

21. SOLARIN.TXT – Solar supply shapes

This text file is used by RFM to specify the solar photovoltaic (PV) and concentrating solar power (CSP) supply shapes (capacity factors) by month and time of day, EMM region, and year. The supply shapes were developed using NREL data.

22. RENDAT.TXT – Wind step multipliers

RENDAT.TXT provides the RFM with long-term capital cost adjustment factors for the wind technologies. These factors are a function of available wind resources by EMM region and how much of each supply step has been used by existing wind capacity. The adjustment factors are used by the ECP module to determine which wind class is the most economic to develop in each EMM region and year. These data were provided by NREL.

23. WESAREA.TXT – Wind resources

WESAREA.TXT contains the available wind resources by EMM region and wind class as well as the wind supply shapes (capacity factors) by month and time of day, EMM region, and year, for both onshore and offshore wind. Wind data are provided by NREL.

24. HYDESITE.TXT – Hydropower sites

HYDESITE.TXT contains hydropower sites available for development, specified by state and EMM region. The site data come from the Idaho Hydropower Resource Economics Database (IHRED) and FERC.

25. MSWDAT.TXT – Landfill gas resources

MSWDAT.TXT contains parameters used to determine the amount of municipal solid waste that is available for use in landfill gas-to-electricity plants. Parameters that are specified by EMM region include regional population share factors, share of total waste streams available for combustion by electric generators, and existing landfill capacity. Also specified by EMM region are the capital and Operations and Maintenance (O&M) cost multipliers by supply step. Sources for these data are the U.S. Environmental Protection Agency (EPA), Government Advisory Associates, and Electric Power Research Institute (EPRI) Technical Assessment Guide (TAG).

APPENDIX B - Activities performed for EMM update from 13 to 22 regions

A. A list of **utilities** to **PCA** to new **EMM regions** was developed. (As noted earlier, the term *power control area* has since been renamed *balancing authority area*). A few other lists for "simple" updates—NERC subregions to new EMM regions, and old EMM regions to new EMM regions—were also developed.

B. There were 23 input file changes:

PLTF860.TXT	ECPDAT.TXT	HYDSITE.TXT
PRETTIN.TXT	EFPCNTL.TXT	MSWDAT.TXT
CANSPLY.TXT	ELDATYR.TXT	RENDAT.TXT
ETCNCT.TXT	EMMCNTL.TXT	SOLARIN.TXT
FRMCHRG.TXT	FINREG.TXT	WESAREA.TXT
LDSMSTR.TXT	NUGPIPE.TXT	WFLOORS.TXT
DICT.TXT		WGEPARM.TXT
		WGESITE.TXT
		WODSUPP.TXT
		WODTECH.TXT

Most input files were updated with data as described in section 3.1, with a few exceptions:
   a. PLTF860.TXT – Plant data from the previous PLTF860.TXT file were merged with eGRID data by PCA.
   b. NOPER and NOWN region fields in the plant file were updated for new 22 region indexes for plants whose PCAs matched.
   c. For plants that did not match, plants were manually looked up in the eGRID file or simply assigned a region that made sense.

C. There were 14 include files changed, mostly for hardwired "13"s changed to "mnumnr-3," and a few other small changes:

bldctl	control	dsmdimen	dsmsectr	efple	efprp2	efptemp
emmbuf	emmparm	ftable	omlbuf	parametr	run	wwind

D. Preprocessors were updated and rerun:

Prepplt.f – Rgn_fac data were manually updated, the code was recompiled, and rerun
Prepett.f – TDLS data were manually updated, the code was recompiled, and rerun

E. There were 11 source code files changed and recompiled:

Uldsm.f	Uread.f	Renew.f	Ucape.f	Uecp.f	Uefp.f
Udat.f	Uefd.f	Util.f	Ftab.f	Ftab2.f	

Changes included:

  a. Removing hardwired code
  b. Updating buffer space (uread.f)
  c. Special code associated with HI,AK regions from 14,15 to 23,24 (renew.f and util.f)
  d. Special code associated with regions 11-13 changed to 19-22 (renew.f)
  e. Special code associated with 6 (NY) updated to 6,7,8, (udat.f)
  f. Special code associated with 13 (CA) updated to region 20 (udat.f)
  g. Data statements expanded from 13 regions to 22 (uefd.f and uefp.f)
  h. Data statements expanded from 13 regions to 22 (uefp.f)
  i. Updated parameters regarding regional tables and NERCNAM (ftab.f, ftab2.f)

APPENDIX C - Activities needed to update EMM for a new set of regions

A. Develop mapping of **utilities** to **BA** to new **EMM regions**. Develop a few other lists for "simple" updates—NERC subregions to new EMM regions, and old EMM regions to new EMM regions.
B. Update input files
    1. PLTF860.TXT – Use BA on each latest EIA-860 plant record to obtain EMM region for NOPER region. Use utility owner ID (mapped to BA) to obtain EMM region for NOWNER region.
    2. PRETTIN.TXT – Change EMM region number based on NERC subregion (this can be very involved if the new EMM regions are very different than before, and/or very different from the NERC subregion definitions).
    3. ETCNCT.TXT – Change EMM and Canadian region numbers old EMM to new EMM region mapping.
    4. FRMCHRG.TXT – Change EMM and Canadian region numbers old EMM to new EMM region mapping.
    5. CANSPLY.TXT – Change EMM and Canadian region numbers based on old EMM to new EMM region mapping.
    6. LDSMSTR.TXT – Update the LDSMSTR.TXT file to change the nNercReg setting, the EMM regions to LSR names tables, and the sales matrices, and other tables.
    7. LDSMDAF.DAF – Rerun the program that processes the FERC 714 data to the new EMM regionality based on BA.
    8. CTS files – As a minimum, update the EMM region tables associated with the CTUS MDB, rerun the MDB program to produce the CTSINPUT.GDX, and rerun the GAMS preprocessor code to produce a new CTSSOLN.GDX. This involves more updates if the fuel regions change.
    9. FINREG.TXT – Rerun the MDB and Excel programs with a revised mapping of utility to BA to new EMM region.
    10. ELDATYR.TXT – Rerun the programs with a revised mapping of utility to BA to new EMM region.
    11. EMMCNTL.TXT – Manually update the 20 tables that are by EMM region. This involves more updates if the fuel regions change.
    12. ECPDATX.XML – Manually update the tabs that have EMM regional data.
    13. NUGPIPE.TXT – Manually update the two tables that are by EMM region.
    14. EFPCNTL.TXT – Manually update the two tables that are by EMM region.
    15. PHASEIN.TXT – Manually update the tables that are by EMM region.
    16. SALELB.TXT – Manually update the tables that are by EMM region.
    17. EMMDBDOC.XLS/EMMMDB.MDB – Manually update the tables that are by EMM region.
    18. LAYOUT.TXT/LAYIN.XLS – Manually update the tables that are by EMM region.
    19. WFLOORS.TXT – Manually update the EMM regions.
    20. WGEPARM.TXT / WGESITE.TXT – Manually update the EMM regions.
    21. SOLARIN.TXT – Manually update the EMM regions.

22. RENDAT.TXT – Manually update the EMM regions.
23. WESAREA.TXT – Manually update the EMM regions.
24. HYDESITE.TXT – Manually update the EMM regions.
25. MSWDAT.TXT – Manually update the EMM regions.

C. Update include files:
1. Efprp2 – Buffer space is set to a multiple of the number of regions.
2. Emmparm – Set EFD$MFRG and WPLT$RGN to number of regions.
3. Parametr – Set MNUMNR to number of regions.
4. Omlbuf – If more than 25 regions, check that there is enough buffer space.
5. Wwind – Set EXNR to one plus number of regions.

D. Run preprocessors:
1. Prepplt.f – Manually update Rgn_fac data which are hardwired in the code first, recompile, rerun.
2. Prepett.f – Manually update the TDLS data which are hardwired in the code first, recompile, rerun.

E. Update source code:
1. Renew.f – Update places where code is hardwired for "22" regions for output reports, for HI and AK (23, 24), for special geothermal settings for regions 19-22.
2. Uread.f – Buffer space has already been updated from 132 to 240 characters for files that have regions across the page. Check to see if buffer space needs to be revised again if number of regions is much larger than 25.
3. Udat.f – Check portions of code where special adjustments are made for NY and CA regions.
4. Uefd.f – Update CHCOD and REGCD for new number of regions.
5. Uefp.f – Update data statements and Growth data for new number of regions.
6. Util.f – Update for AK and HI.
7. Ftab.f and Ftab2.f – Update parameters for regional tables, update NERCNAM, update CTL_CD_NR.

F. Recompile code.
G. Run NEMS test case.
H. Compare national and regional results to baseline case.