# Introduction

This document describes the PHORUM model, a security-constrained unit commitment model that simulates the 2010 PJM Day-ahead energy market[[1]](#footnote-1). The tool has been developed by Carnegie Mellon University’s Carnegie Electric Industry Center (CEIC). The purpose of PHORUM is to be a light-weight, transparent simulation tool for academia, industry, and government. PHORUM allows researchers to investigate how changes to PJM will affect generators, electricity prices, emissions, and human health.

PHORUM has many advantages over similar tools:

* **Free**
* **Open-source:** PHORUM is transparent and peer-reviewed, avoiding the ‘black-box’ problem. We hope users will continue to improve and expand this open source tool. The model only uses publically available data.
* **Easy to use:** PHORUM has been designed with the user in mind. Users can easily customize and run simulations. The tool can be easily adapted to your needs or expanded upon.
* **Verified accuracy:** PHORUM’s results have been verified against actual PJM market results, showing that the model accurately captures market drivers. Detailed verification results are below.
* **Light-weight:** PHORUM has been designed to run on a personal computer. Simulating one day takes less than 5 minutes; simulating a year takes less than 24 hours.

PHORUM is available as an open source software package, protected under the X license. All code and supporting documentation are available free of charge. Users are encouraged to contribute to the project by adding features or improving model accuracy.

# System Requirements

## Required software

* Microsoft Windows
* Microsoft Excel (2007+)
* Matlab
* GAMS 23.6 with CPLEX 12.4

## Minimum recommended hardware

* Intel Core 2 Duo CPU (2.5 GHz) processor
* 6 GB of memory

# Installation Instructions

1. **Download** ‘PHORUM.zip’ from <http://wpweb2.tepper.cmu.edu/phorum/>
2. **Unzip** all file contents into the folder where the model will reside.
3. **Open PHORUMGAMS.gpr:** This is PHORUM’s GAMS IDE project file. Opening the file will set it as the default GAMS project. After opening, close the file.
4. **Update Settings:** Open Settings.xlsx. Change ‘GAMS directory path’ to the folder where gamside.exe resides (likely the folder that GAMS is installed to). The rest of the settings are default to conditions that will allow you test that PHORUM is setup correctly. Do not change them before the first run. Save and close Settings.xlsx.
5. **Open Matlab**. Change the working directory to your PHORUM folder.
6. **Load settings.mat:** Call the LoadSettings function by typing ‘LoadSettings’ into the command line. This will parse the settings.xlsx file and save results as the matlab structure settings.mat, which is used by the simulation.
7. **Load PHORUMdata.mat:** Call the LoadData function by typing ‘LoadData(‘PHORUMdata’,’PHORUMdata’) into the command line. This will take data from the PHORUMdata.xlsx file and save it to the matlab structure PHORUMdata.mat, which is used by the simulation.
8. **Begin simulation:** Call the Main function by typing ‘Main’ in the command line. This should begin the test simulation.
9. Matlab will notify you once the test simulation is complete.

# Model Structure

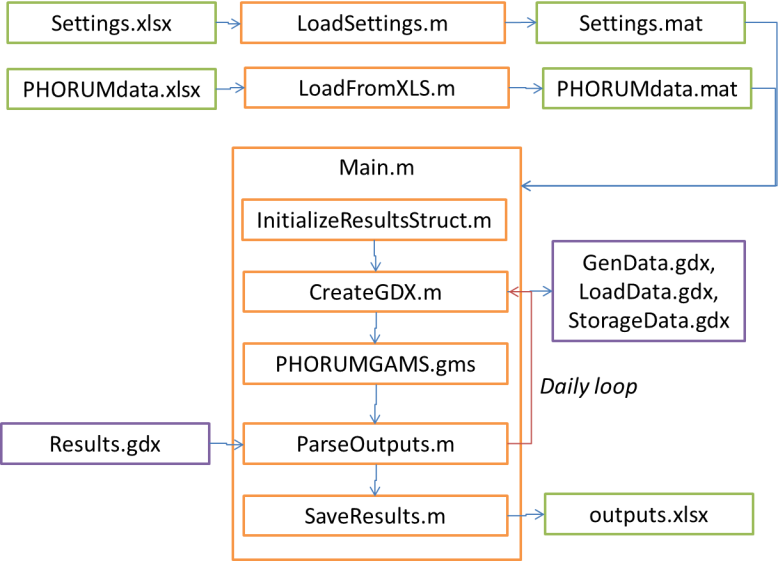


Figure 1: Structure of the PHORUM model.

**Settings.xlsx:** This file controls all model settings. In addition to mandatory settings, the file allows you to set dates over which the optimization will be run, emission prices, and the outputs from the simulation. Outputs are provided either on an hourly, daily, or total basis.

**LoadSettings.m:** This function parses the settings.xlsx file and stores it to settings.mat, which is required for the Main function.

**PHORUMdata.xlsx:** This write-protected file stores all data used by PHORUM, including generator, load, and storage unit data. You can create custom simulations by changing the data and savings to a new file. For example, users can add or remove generators, change fuel prices, and increase or decrease load.

**LoadFromXLS.m:** This function parses the input data file (default to PHORUMdata.xlsx) and stores it to a .mat output file (default to PHORUMdata.mat), which is used by and required for the Main function. Syntax: ‘LoadData(A, B)’, where A is the input file name and B is the output file name. Both A and B should exclude suffixes .xlsx and .mat.

**Main.m:** This is the primary PHORUM Matlab function. The function makes several sub-function calls, and is responsible for calling GAMS.

**InitializeResultsStruct:** Called by Main.m. This function initializes the totalResults.mat and prevDayResults.mat structures, which are used to store results from multiple days and the previous day, respectively.

**CreateGDX.m:** Called by Main.m for each optimization day. This function loads the appropriate data from PHORUMdata.mat and parses it into .gdx files, which are then used as inputs by GAMS. This function uses the prevDayResults.mat structure to set initial conditions for each optimization.

**PHORUMGAMS.gms:** Called by Main.m for each optimization day. The file takes inputs from genData.gdx, loadData.gdx, and storageData.gdx created in the CreateGDX.m subroutine. The optimization outputs results into the results.gdx file.

**ParseOutputs.m:** Called by Main.m for each optimization day. This function parses the outputs returned by GAMS in results.gdx and adds them to the totalResults.m structure, which stores results from all days. The function also populates the prevDayResults.mat file, which is used by CreateGDX.m to set initial conditions for the next optimization.

**SaveResults.m:** Called by Main.m. This function uses the optimization results stored in totalResults.m to calculate all outputs you request. These outputs are then stored into the file you request.

**Cplex.opt:** This is the settings file for the CPLEX solver within GAMS. Default settings are to use 2 threads, 6000 MB of memory, and an optimization time limit of 7200 seconds (2 hours). The epopt and epgap settings control convergence tolerances and should not be changed. You can modify the file improve performance on your computer. More details are at <http://www.gams.com/dd/docs/solvers/cplex.pdf>.

# PHORUM Outputs

PHORUM allows you to specify which outputs are returned from each simulation. You can choose from the following outputs, and you can also specify the granularity of results: hourly, daily, or total.

**System Outputs**

* System costs
* Locational Marginal Prices (LMPs)
* Transmission levels

**Generator Outputs**

* Power generated
* Number of startups and runtime
* Costs (variable, startup, and no-load)
* Net revenues

**Storage Outputs**

* State of charge
* Amount charged and discharged
* Net revenues

**Emissions**

* CO2, NOx, SO2, N2O, CO2-equivalent, CO, NH3, PM10, PM2.5, and VOC emissions, by generator
* Marginal damages to human health from each generator

# Use Instructions

## Standard Use Case

1. Open the **settings.xlsx** file. Make sure that all mandatory settings are correct, and specify any other settings you want. Set the data filename to *PHORUMdata*. Save and close the file.
2. Open Matlab. Change the working directory to the PHORUM folder.
3. Run **LoadSettings** to load settings from settings.xlsx to settings.mat.
4. Run **LoadData(‘PHORUMdata’,’PHORUMdata’)** to load data from PHORUMdata.xlsx and save it to PHORUMdata.mat.
5. Run **main.m** to begin the simulation.
6. When the simulation finishes, outputs will be stored to the output file you specified in settings.

## Advanced Use Cases

* **Change database:** PHORUMdata.xlsx is a write-protected file contains default PJM data. If you want to change the data, you can do so in a saved copy. Changing the database allows you do many types of simulations, such as adding or removing generators, changing fuel prices, and increasing or decreasing load. Run LoadData.m to save the updated database to .mat.
* **Scenario analysis:** PHORUM allows you to automate running many simulations. An example of how this can be done with an external function is below:
  + Create a loop
    - Load the PHORUMdata.mat file called by main.m. Update the relevant parameters to your scenario, and save the file to disk.
    - Load the settings.mat file. Update parameters for your simulation and the output filename. Save the file to disk.
    - Call Main.m
  + Repeat the process, changing the input database and output filename as needed for each iteration.

# Function Details

## LoadFromXLS

This function takes data from the specified excel database and converts it into a .mat structure which is used by the PHORUM model.

Inputs:

* inputFileName: File name of excel database (excluding suffix)
* outputFileName: File name of output .mat file (excluding suffix)

Outputs: None

Variables:

* tempGenData: Raw data pulled from the excel database
* genData (structure): All relevant generator data from tempGenData. The structure contains different data elements for each relevant data element.
  + genData.gPlantCode – generator’s unique identifier
  + genData.gHeatRate - Heat rate (BTU/kWh)
  + genData.gCapacitySummer – Max capacity summer months (MW)
  + genData.gCapacityWinter – Max capacity winter months (MW)
  + genData.gVarOM – Variable O&M ($/MWh)
  + genData.gFuelPrice – Monthly fuel price ($/MMBtu)
  + genData.gRampRate – Ramp rate (MW/hr)
  + genData.gMinUp - Minimum time generator must stay up (hrs)
  + genData.gMinDown – Minimum time generator must stay down (hrs)
  + genData.gStartupCost – Generator startup cost ($)
  + genData.gMin – Min capacity (% of maximum)
  + genData.gEAF – Monthly Equivalent Availability Factor (%)
  + genData.gZone – Zone in which generator is located.
  + genData.gNLC – Generator no-load cost ($) (not currently used)
  + genData.gTCR – TCR in which generator is located
* loadData(structure): All relevant load data from the input file. The structure contains different data elements for each relevant data element.
  + loadData.PSEG, .PECO .PPL .BGE .PEPCO .RECO .APS .COMED .AEP .DAY .DUQ .DOM .JCPL .METED .PENELEC .AECO .DPL – Hourly load for each PJM zone
  + loadData.ALTE .ALTW .AMIL .CIN,.CPLE .CPLW .CWLP .DUK .EKPC .FE .IPL .LEE .LIND .MEC .MECS .NEPT .NIPS .NYIS .OVEC .TVA .WEC – Hourly imports/exports across each PJM interconnection to other RTOs
  + loadData.windTCR1 .windTCR3 – Hourly wind generation in TCRs 1 & 3 (only TCRs with installed wind capacity)
  + loadData.EImax .CImax .WImax .DOMImax – Hourly transmission capacity across Eastern, Central, Western, and Dominion interconnections.
  + loadData.LMPTCR1actual .LMPTCR2actual .LMPTCR3actual .LMPTCR4actual .LMPTCR5actual – Actual 2010 hourly LMPs for each TCR
* tempStorageData: Raw data pulled from the excel database
* storageData(structure): All relevant storage data from tempStorageData. The structure contains different data elements for each relevant data element.
  + storageData.sTCR – TCR in which each storage device is located
  + storageData.sCapacity – Capacity of storage device (MW)
  + storageData.sChargeEff – Charge efficiency (%)
  + storageData.sDischargeEff – Discharge efficiency (%)
  + storageData.sDuration – Duration of device (hrs)
  + storageData.sVC – Variable costs of storage ($/MWh) (not currently used)

## LoadSettings

This function parses the settings.xlsx file and stores it to settings.mat, which is required for the Main function.

Inputs: None

Outputs: None

Variables:

* settingsRaw: Raw data pulled from the excel database
* settings(structure) : All relevant settings data from Settings.xlsx. The structure contains different data elements for each relevant data element.
  + settings.GAMSpath – Path to GAMS folder
  + settings.callGAMS – String used to call GAMS from the command line
  + settings.dataFileName – Name of excel data file (excluding suffix)
  + settings.outputFileName – Name of output file
  + settings.numDateRanges – Number of date ranges used in simulation (up to 4)
  + settings.dStart – Start date of each simulated range
  + settings.dEnd – End date of each simulated range
  + settings.priceNOX .priceSO2 .priceN2O .priceCO2 .priceCO2eqv .priceCO .priceNH3 .pricePM10 .pricePM25 .priceVOC – emissions price of each pollutant ($/lb)
  + .isMDNH3 .isMDSO2 .isMDVOC .isMDNOX .isMDPM25 .isMDPM10 – Are marginal damages associated with each pollutant priced (Y/N)
  + settings.systemCosts – Does output file return system costs (daily or total)
  + settings.modeledLMPs – Does output file return modeled TCR LMPs (hourly)
  + settings.actualLMPs – Does output file return actual TCR LMPs (hourly)
  + settings.load – Does output file return load (hourly
  + settings.tLimit – Does output file return transmission limits (hourly)
  + settings.tLevel - Does output file return transmission levels (hourly)
  + settings.gLevel – Does output file return each generator’s production (daily, total)
  + settings.gRuntime – Does output file return each generator’s hours run (daily, total)
  + settings.gStartups – Does output file return each generator’s number of startups (daily, total)
  + settings.gNetRevenue – Does output file return each generator’s net revenue (daily, total)
  + settings.gVC – Does output file return each generator’s variable costs (daily, total)
  + settings.gStartupC – Does output file return each generator’s startup costs (daily, total)
  + settings.gNLC – Does output file return each generator’s no-load costs (daily, total)
  + settings.sSOC – Does output file return each storage device’s state of charge (hourly)
  + settings.sCharge – Does output file return each storage device’s amount charged (daily, total)
  + settings.sDischarge – Does output file return each storage device’s amount discharged (daily, total)
  + settings.sNetRevenue – Does output file return each storage device’s net revenue (daily, total)
  + settings.sVC – Does output file return each storage device’s variable costs (daily, total)
  + settings.gNOX .gSO2 .gN2O .gCO2 .gCO2eqv .gCO .gNH3 .gPM10 .gPM25 .gVOC – Does output file return emissions from each generator (daily, total)
  + settings.MDtotal .MDNH3 .MDSO2 .MDVOC .MDNOX .MDPM25 .MDPM10 – Does output file return marginal damages associated with each generator (daily, total)

## Main

This is the primary PHORUM Matlab function. The function makes several sub-function calls, and is responsible for calling GAMS. The function loads data from the database file and settings file.

Inputs: None

Outputs: None

Variables:

* optWindow – number of hours in each optimization.
* totalResults –
* prevDayResults -

# Methodology

## Reduced-form Topology

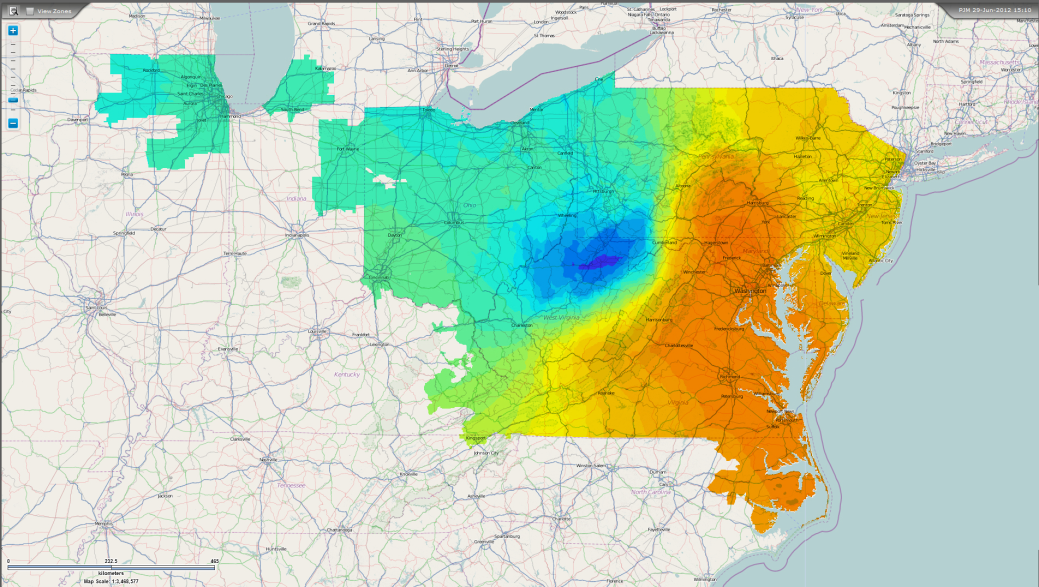
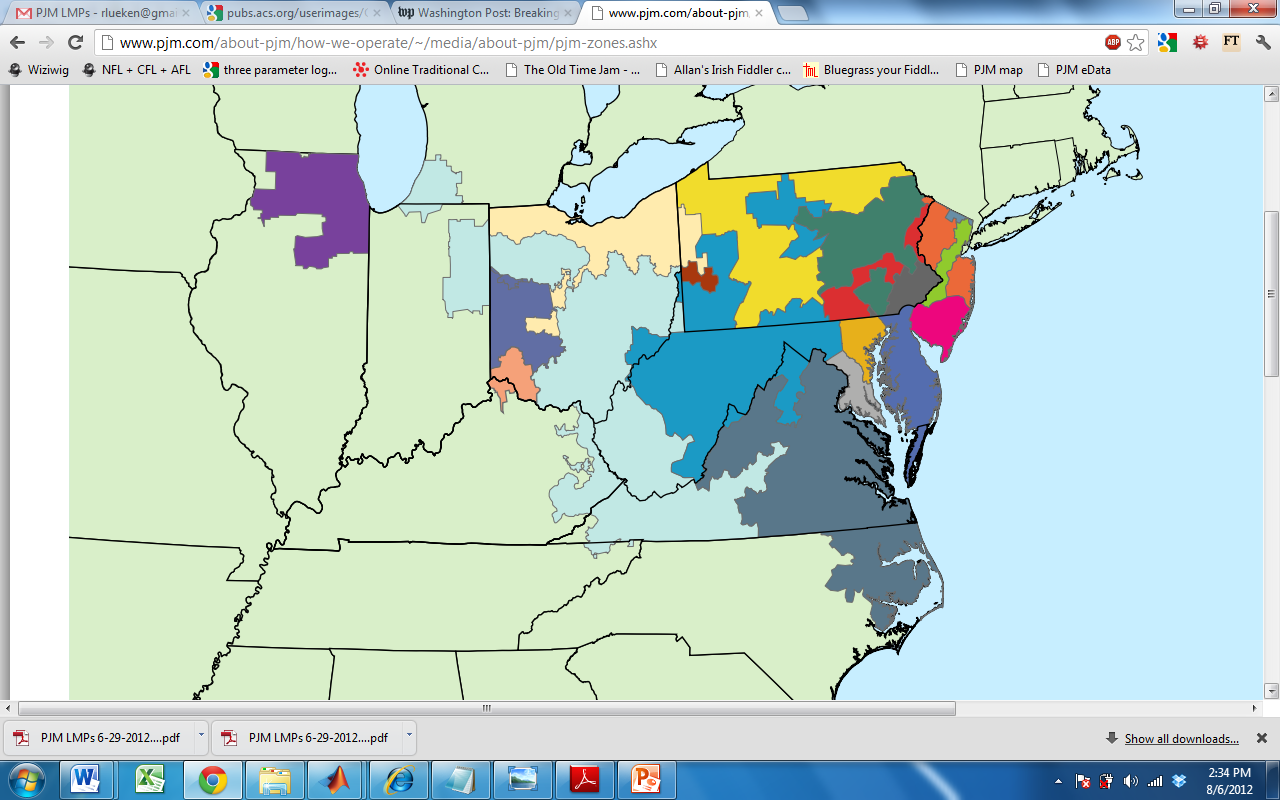
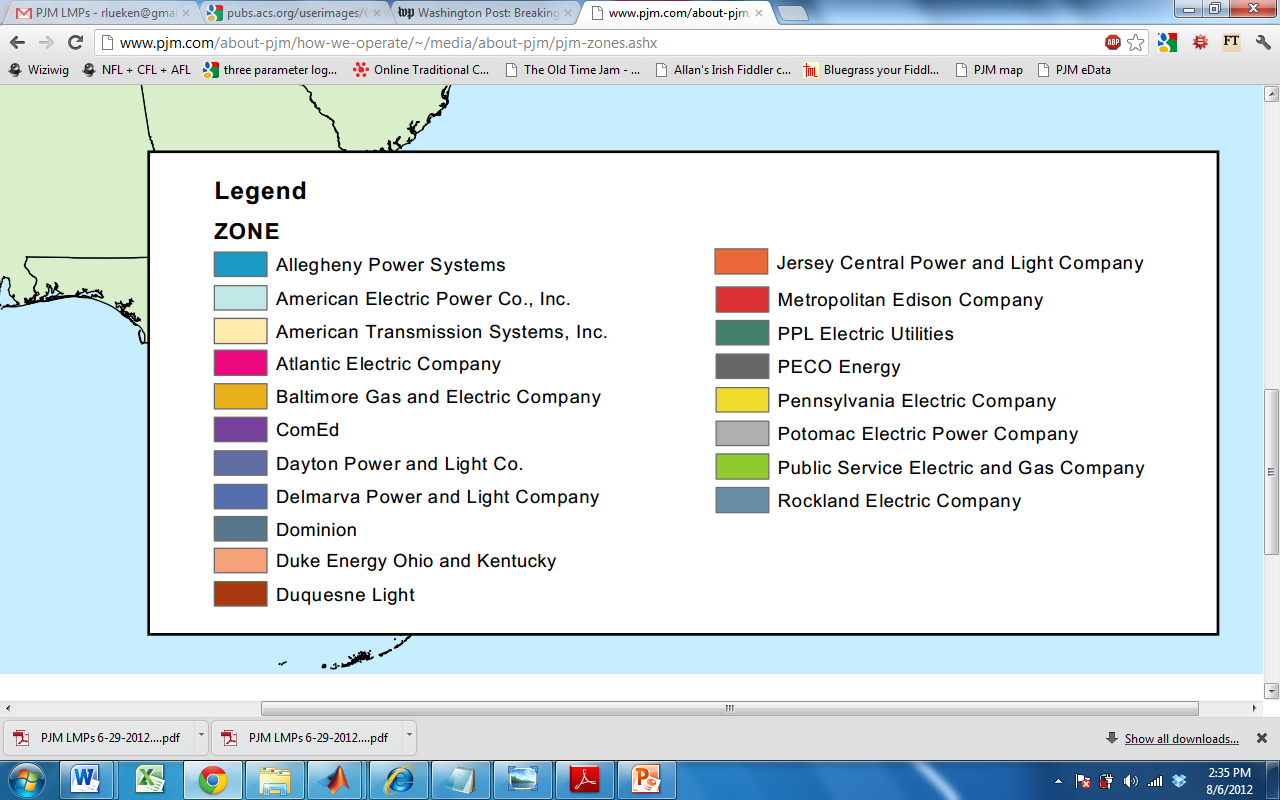
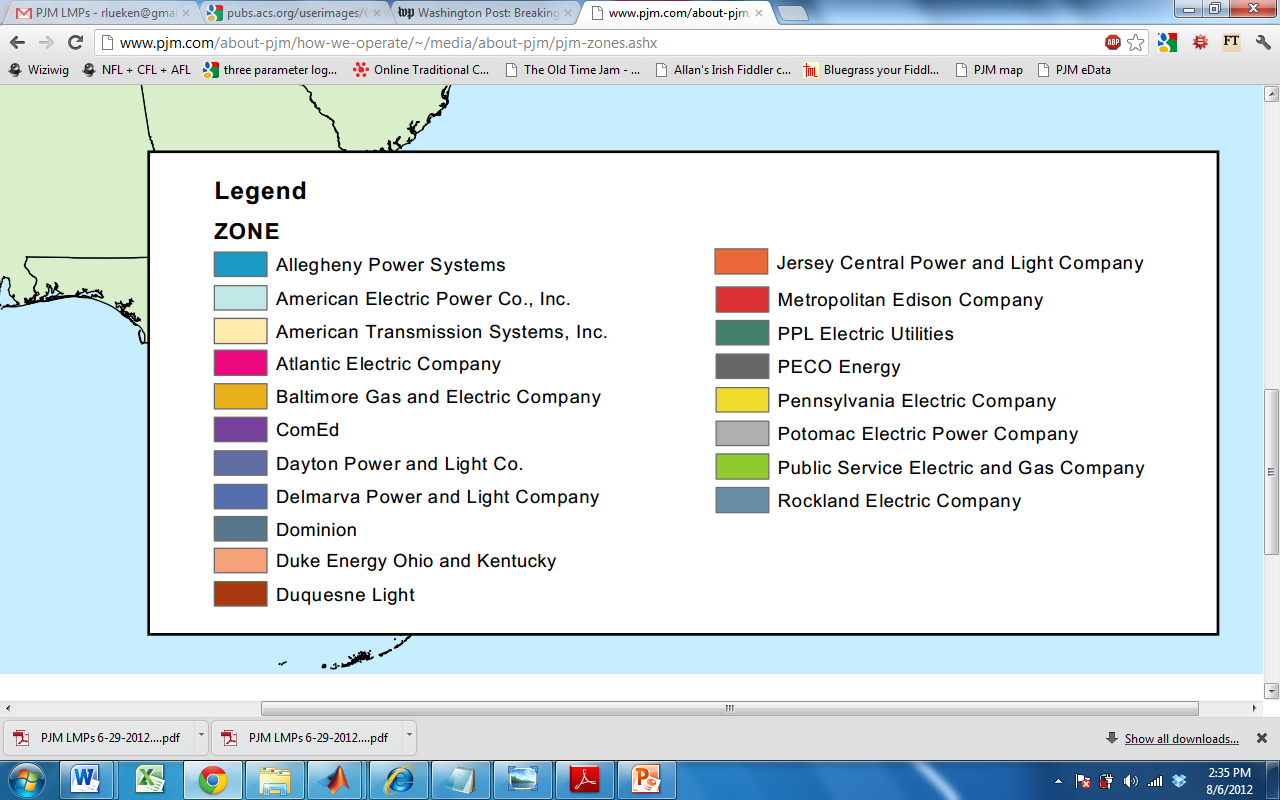


Figure 2: PJM LMPs on June 29, 2012, 3:10 pm[[2]](#endnote-1). Prices range from $20/MWh in blue areas to $150/MWh in orange areas, illustrating how transmission constraints limit the supply of power from west to east.

PHORUM uses a reduced-form, 5-bus transmission network to account for transmission constraints within PJM. A reduced-form model is necessary because details on transmission assets are Critical Energy Infrastructure Information and not publically available. We divide PJM into five buses, or Transmission Constrained Regions (TCRs), and assign each PJM zone to a TCR, as discussed below. This process of dividing PJM into regions has been used by other researchers[[3]](#endnote-2),[[4]](#footnote-2). We assume transmission within each TCR is unconstrained and all LMPs are equal. 2010 LMP data shows that within our defined TCRs, zonal LMPs are highly correlated, supporting this assumption (Figure 3).





**APS**

**AEP**

**ATSI**

**AECO**

**BGE**

**COMED**

**DAY**

**DPL**

**DOM**

**DUK**

**DUQ**

**JCPL**

**METED**

**PPL**

**PECO**

**PENELEC**

**PEPCO**

**PSEG**

**RECO**

Figure 3: The PJM Interconnection and its Constituent Zones[[5]](#endnote-3),[[6]](#footnote-3)

Table 1: Assignment of PJM zones to PHORUM Transmission Constrained Regions (TCRs)

|  |  |
| --- | --- |
| **TCR** | **PJM Zones** |
| TCR 1 | AEP, APS, COMED, DAY, DUQ, PENELEC, ATSI |
| TCR 2 | BGE, PEPCO |
| TCR 3 | METED, PPL |
| TCR 4 | JCPL, PECO, PSEG, AECO, DPL, RECO |
| TCR 5 | DOM |



Figure 4: Correlation coefficients between zones for 2010 hourly day-ahead LMPs. High correlation within a TCR supports the assumption that transmission is unconstrained within the TCR.

PJM has identified seven transmission interfaces, each made up of multiple 500 kV lines, which form critical congestion paths. Importantly, PJM provides hourly data on the capacity of these interfaces[[7]](#footnote-4). We model these interfaces as transmission constraints between TCRs. These seven interfaces made up 49% of all congestion costs in 2010[[8]](#endnote-4). We ignore all other transmission constraints, and assume lossless, DC power flow between TCRs. We aggregate the seven PJM interfaces into five Transmission Interconnections (TIs)[[9]](#footnote-5), as shown in Table 2.

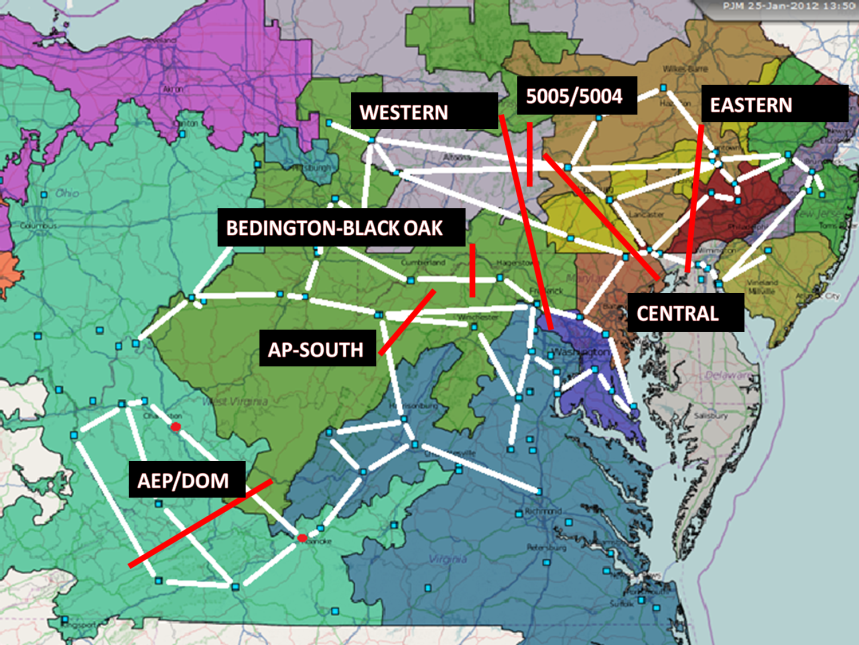
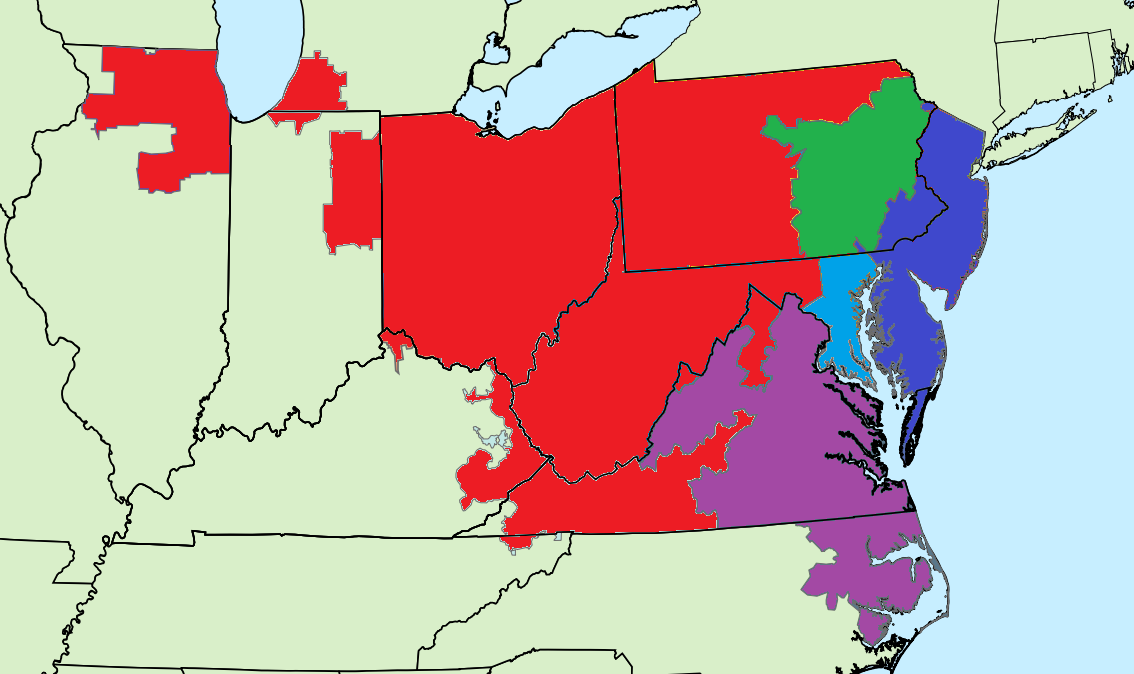


Figure 5: Map of PJM 500kV transmission lines (white lines) and transmission interfaces (red lines)i,[[10]](#endnote-5). Most interfaces contain multiple 500kV lines.

Table 2: Assignment of PJM Interfaces to PHORUM Transmission Interfaces (TIs)

|  |  |
| --- | --- |
| **TI** | **PJM Interface** |
| TI 1-2 | ¼ of Western Interface capacity |
| TI 1-3 | ½ of Western Interface capacity |
| TI 5-2 | ¼ of Western Interface capacity |
| TI 1-5 | Bedington-Black Oak, AP South, AEP-DOM |
| TI 2-3 | Central Interface |
| TI 3-4 | Eastern Interface |
| Not modeled | 500X(5004+5005) |



**TCR 4**

**TCR 5**

**TCR 3**

**TI1-2**

**TI2-3**

**TI1-5**

**TI5-2**

**TI3-4**

**TI1-3**

**TCR 2**

**TCR 1**

Figure 6: Map of Transmission Constrained Regions (TCRs) and Transmission Interconnections (TIs)

**Day-Ahead Market**

PJM operates several electricity markets, the largest of which are the day-ahead (DAH) and real-time energy markets.We model the DAH market instead of the real-time market for two reasons. First, the DAH market is larger, with generally lower and less volatile prices[[11]](#footnote-6), serving as a conservative lower bound on storage profitsiv. Secondly, prices in the Real-Time market are highly influenced by factors outside the capability of PHORUM, such as sudden changes in the weather, forced generator outages, and transmission outages[[12]](#footnote-7). We assume all available generators participate in the DAH market[[13]](#footnote-8).

**Reserve Requirements**

To maintain reliability, PJM maintains spinning reserves that are synchronized to the grid at all times[[14]](#footnote-9). PHORUM adds these reserve requirements to hourly load[[15]](#footnote-10). This differs from PJM’s model, which co-optimizes the day-ahead energy market and a separate day-ahead scheduled reserve (DASR) market. The DASR market includes synchronized reserves, quick-start reserves (available within 10 minutes), and supplemental reserves (available within 10 – 30 minutes).

## Unit Commitment Methodology

PHORUM simulates each individual generator and finds the least-cost combination to meet load at every hour[[16]](#footnote-11), thereby minimizing the total social cost of providing electricity. This optimization, described in detail below, is subject to generator and transmission constraints. PJM and all major restructured markets use UCMs to dispatch power. PHORUM uses a mixed integer linear programming (MILP) optimization technique. Linear programming cannot model the non-convexities of grid operations, which include:

* **Startup Costs:** Starting a generator incurs a fixed startup cost.
* **Minimum generation:** Once a generator is called upon, it must provide power between some minimum and maximum level.
* **Minimum runtime/downtime:** Once a generator is turned on or off, it must remain on/off for a minimum amount of hours.

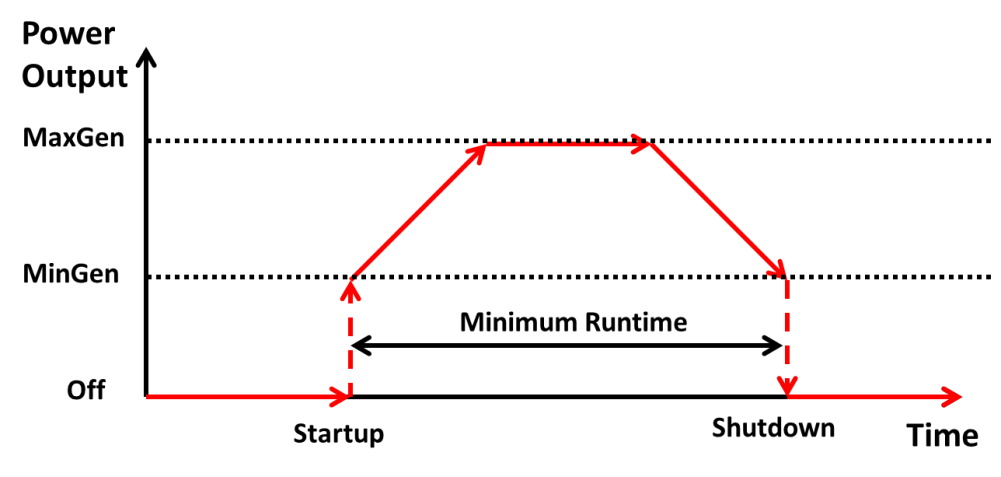


Figure 7: Illustrative dispatch of a generator. Once ordered on, output increases nonlinearly to the minimum generation level. The generator must remain on for the minimum runtime. When ordered off, power output nonlinearly drops from minimum generation to zero.

PHORUM simulates 1,017 generators[[17]](#footnote-12), 4 pumped hydro storage facilities[[18]](#footnote-13), 5 buses and 6 transmission interconnections. We run 365 separate optimizations, each minimizing costs over a 48-hour period. The optimizations are rolled over, as shown below. This rollover ensures that minimum runtime/downtime constraints hold between days. PJM’s dispatch process minimizes costs over one day only; cross-day decisions are made manually by the day-ahead operator[[19]](#endnote-6),[[20]](#footnote-14).

Each optimization is passed four variables from the last hour of the previous day’s optimization:

* The on/off state of each generator
* How long each generator must remain on/off
* The power output of each generator
* The state of charge for each storage unit

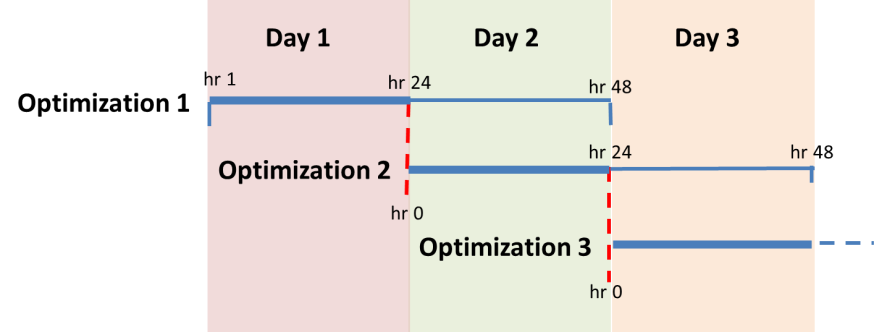


Figure 8: Illustration of how model handles day boundaries. Each optimization runs for a full 48 hours, but only the first 24 hours of results are retained. Variables are passed from the 24th hour of the first optimization to hour 0 of the second.

## Formulation

Table 3: Model Nomenclature

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Constants** | |  | **Variables** | |
|  | System demand at time t in TCR r [MW] |  |  | Power imported (+) or exported (-) from TCR r to TCR m ( |
| (t) | Max power flow on interface l |  |  | Power flow on interface rm at time t |
|  | Variable cost of storage unit k for generating / storing power at time t |  |  | State of charge (SOC) of storage unit k at time t [MWh] |
|  | Fixed cost of storage unit k for time t |  |  | Initial SOC of storage unit k |
| , | Max and min charge/discharge rate from storage unit k at time t |  |  | Power discharged (+) or charged (-) by storage unit k at time t |
|  | Max SOC of storage unit k [MWh] |  |  | Power generated by unit I at time t |
|  | Round-trip efficiency of storage units |  |  | Initial state of thermal unit i (1 if it is online, 0 otherwise)[[21]](#endnote-7) |
|  | Variable cost of unit i for generating power at time t |  |  | Binary variable that is equal to 1 if unit i is online in period t and 0 otherwisevii. |
|  | Fixed cost of unit i for time t |  |  | Startup cost of unit i at period t (positive variable) |
|  | Startup cost of unit i |  | **Sets** | |
| , | Max and min generation levels from unit i at time t |  | *­* | Set of indices of the generating units in TCR r |
|  | Ramp rate of thermal unit i |  |  | Set of indices of the storage units in TCR r |
|  | Minimum uptime of unit Ivii |  |  | Set of TCRs |
|  | Minimum down time of unit Ivii |  |  | Set of transmission interfaces |
|  |  |  |  | Set of indices of the time periods |

Table 4: Model Formulation

|  |  |
| --- | --- |
| **Minimize** | (1) |
| **subject to** |  |
| **System Constraints** |  |
|  | (2) |
|  | (3) |
| **Storage Constraints** | |
|  | (4) |
|  | (5) |
|  | (6) |
|  | (7) |
| **Generator Constraints** | |
|  | (8) |
|  | (9) |
|  | (10) |
|  | (11) |
|  | (12) |
|  | (13) |
|  | (14) |
|  | (15) |
|  | (16) |
|  | (17) |
| ,[[22]](#endnote-8) | (18) |

PHORUM’s formulation is derived from multiple sources[[23]](#endnote-9),vii, and is similar to the model used by PJM to dispatch power on the DAH market[[24]](#endnote-10),[[25]](#footnote-16). The objective function is to minimize total social cost of providing electricity (1). Separate supply/demand constraints are set for each TCR (2). The LMPs at each TCR are the negative Lagrange Multiplier (shadow price) of these constraints.

Eqn (5) sets transmission limits between TCRs. Storage constraints (4-7) limit the capacity, charge/discharge rates, and set initial charge levels. Eqn (8) triggers startup costs when a generator turns on, and (9) constrains generation capacity while the generator is online. Eqns (10-11) constrain generator ramp rates. Eqns (12-14) ensure generators satisfy uptime constraints: (12) sets initial uptimes, (13) constraints uptimes for subsequent hours, and (14) forces generators that turn on near the end of the day to stay on over the final time periods. Eqns (15-17) are analogous to (12-14), but for generator downtimes. Equation (18) calculates variable costs.

## Environmental Damages

PHORUM tracks CO2, NOx, and SO2 emissions from all generators. We assume emission rates are independent of output level, time of year, or ambient temperature[[26]](#footnote-17). The marginal damages of each generator’s NOx and SO2 emissions on human health and the environment are quantified with the Air Pollution Emission Experiments and Policy Analysis Model (APEEP) model[[27]](#endnote-11) (see Appendix III).

## Storage

We also assume storage operators are free to buy and sell on the day-ahead market, and have perfect information[[28]](#footnote-18) of prices over the 48-hour optimization period[[29]](#footnote-19). Charge level is set to half of maximum capacity for the first hour of the year and the last hour of each 48-hour optimization. We do not dynamically model storage degradation, and assume no standby losses. We assume there are no variable costs associated with charging or discharging, other than the price of electricity that is purchased to charge. We do not explicitly model a minimum depth of discharge.

# Data

PHORUM’s input data can be broken down into two categories: generator data and hourly data. Generator data describes each generator in PJM. Hourly data describes factors that change hourly. The table below details each data element used in the model and its source.

The PHORUMdata.xlsx file holds all data used by PHORUM. This file contains data that represents the state of PJM in 2010. The file is read-only; modifications can be made by creating a copy of the file. The database has the following worksheets:

* **Sources:** Identifies the source for all data
* **genData:** Data on the PJM generator fleet. Many of the data are not currently used; unused columns are hidden.
* **Load:** Hourly 2010 PJM load, by zone.
* **Imports:** Hourly transfers across all PJM interfaces.
* **LMPs:** Actual hourly zonal LMPs. Used to test PHORUM’s accuracy.
* **Transmission constraints:** Hourly transfer limits across PJM transmission interfaces.
* **Wind:** Hourly PJM wind generation
* **storageData:** Characteristics of all PJM storage units.

We have made several modifications to the generator database in order to improve accuracy. First, for generators in the PJM EIA-411 generator database but not in the NEEDS database, we have assumed values for NEEDS and eGRID data elements. These assumptions are based on values for similar plants. Similarly, the PJM database occasionally combines two generators that NEEDS calls out separately. In these cases, we combine the generators as in the PJM database and assume values based on the constituent generators. Finally, for hydro plants we combine all generators at one plant in order to improve model performance.

We assume generators have linear heat rates, variable O&M costs, ramp rates, and emission factors over their operating range.

Table 5: PHORUM data sources

|  |  |  |
| --- | --- | --- |
| **Data Element** | **Source** | **Notes** |
| **Generator Data** | | |
| ORIS Plant ID | NEEDS[[30]](#endnote-12) |  |
| Plant Type | NEEDSxii |  |
| State Name | NEEDSxii |  |
| State Code | NEEDSxii | [1] |
| County Code | NEEDSxii | [1] |
| Heat Rate [Btu/kWh] | NEEDSxii |  |
| Modeled Fuels | NEEDSxii |  |
| Capacity [MW] (Summer] | PJM[[31]](#endnote-13) | [2] |
| Capacity [MW] (Winter] | PJMxiii | [2] |
| Variable O&M Cost [2010 $/MWh] | Handbook of Electric Power Calculations[[32]](#endnote-14) | [3] |
| Monthly Fuel Price: Jan 2010 – Dec 2010 [2010 $/MMBtu] | EIA Electric Power Monthly, 2010[[33]](#endnote-15), EIA-423[[34]](#endnote-16) | [4] |
| Ramp Rate [MW/hr] | NERC Generating Availability Data System[[35]](#endnote-17) | [5] |
| Min uptime [hrs] | Monitoring Analytics[[36]](#endnote-18). | [6] |
| Min downtime (hrs] | Monitoring Analyticsxviii | [6] |
| Startup cost adder [$] | InterTek[[37]](#endnote-19), CAISO[[38]](#endnote-20) | [7] |
| Minimum Generation [% of maximum generation] | PJM Training[[39]](#endnote-21) |  |
| Monthly Equivalent Availability Factor: Jan 2010 – Dec 2010 | PJM correspondencexxxii & NRC[[40]](#endnote-22) | [8] |
| Stack Height [ft] | National Emissions Inventory[[41]](#endnote-23) |  |
| CO2 emission rate [lb/MMBtu] | eGRID[[42]](#endnote-24) |  |
| Plant annual NOx total output emission rate [lb/MWh] | eGRIDxxiv |  |
| Plant annual SO2 total output emission rate [lb/MWh] | eGRIDxxiv |  |
| **Hourly Data** | | |
| Load | PJM[[43]](#endnote-25) | [9] |
| Imports/Exports [MW] | PJM[[44]](#endnote-26) | [10] |
| Zonal Locational Marginal Prices (LMPs) [$/MWh] | PJM[[45]](#endnote-27) |  |
| Transmission Capacity [MW] | PJM[[46]](#endnote-28) |  |
| Wind Generation [MW] | PJM[[47]](#endnote-29) | [11] |
| Reserve Requirement [MW] | PJM[[48]](#endnote-30) | [12] |

[1] Plants are assigned to zones by state and county codes.

[2] Generator capacities listed for different databases (PJM EIA-411, eGRID, and NEEDS) vary widely. We use data listed in the PJM EIA 4-11 report. Hydro generator capacities are derated by their annual capacity factor.

[3] Variable O&M costs are 2010 values. LFG and MSW costs are based on EIA data[[49]](#endnote-31)

[4] Fuel prices are aggregated by state and by month for each fuel. This aggregation captures both location and seasonal variation in fuel price. Prices are primarily based on the EIA’s Electric Power Monthly data for coal, petroleum liquids, and natural gas delivered price. These databases intentionally exclude some entries in order to maintain anonymity for data providers. Excluded prices are assumed to be the Census Division average, with the exception of West Virginia coal prices, which are derived from the EIA-423 reporting. Prices are assumed to be the same for all types of coal (BIT, SUB, waste coal, etc) and liquid petroleum (DFO, RFO). Fuel price for LFG, MSW, and NUC are assumed to be zero.

[5] Ramp rates are derived from the GADS database. Ramp rates are assumed to be equal for up-ramping and down-ramping. The GADS data was used to identify how the ramp rate of each plant type was correlated to the plant’s capacity. We used an OLS regression of ramp rate against generator capacity. Results are as follows:

* Combined cycle: 0.22 MW/h ramp / MW capacity
* Steam Turbine: 0.14 MW/h ramp / MW capacity
* Gas Turbine: 0.34 MW/h ramp / MW capacity
* Combustion Turbine: 0.33 MW/h ramp / MW capacity

[6] Minimum runtime for small (<150MW) coal plants have been adjusted to account for the fact that these plants are used within PJM as shoulder plants. Runtimes for LFG and MSW plants are assumed to be equal to combined cycle plants.

[7] Based on InterTek and CAISO data, startup costs are assumed at $25/MW for combustion turbine, $50/MW for combined cycle, $100/MW for coal, and $500/MW for nuclear

[8] PJM provided monthly 2010 EAF data, aggregated by generator type (coal 0-249 MW, coal 250-499 MW, coal 500+MW, gas CC, and gas CT). Nuclear EAF was derived from NRC data, using generators in PJM. EAF for LFG and MSW was assumed to be equal to natural gas combustion turbine plants.

[9] PJM sums the DAH load for all zones within the MIDATL region (PENELEC, BGE, PEPCO, METED, PPL, JCPL, PECO, PSEG, AECO, DPL, and RECO) into one entry. Therefore, we divide MIDATL load into its constituent zones. We do this by analyzing the Real Time load data, which is provided separately for all MIDATL zones. For each MIDATL zone, we find the percentage of MIDATL total its load contributes. We then assume that this percentage is the same for DAH and RT loads. Finally, we use that percentage to find the DAH load for each MIDATL zone.

[10] Imports and export data is provided for each interface. We assign these interfaces to the appropriate zones as follows. We assume imports and exports do not change based on PJM prices.

|  |  |
| --- | --- |
| **Zone** | **Interfaces** |
| AEP | ALTE, ALTW, CPLW, CWLP, DUK, EKPC, IPL, LGEE, MEC, MECS, NIPS, OVEC, TVA, WEC |
| PENELEC | FE |
| PSEG | NEPT, NYIS, LIND |
| DOM | CPLE |
| DAY | CIN |

[11] PJM provides hourly wind generation for WEST & MIDATL PJM regions. All WEST wind generation is assigned to TCR1, all MIDATL wind is assigned to TCR3, which is the location of most Mid-Atlantic wind capacity[[50]](#footnote-20). We assume wind generation is must take, and subtract it from load.

[12] For RFC (TCR 1), DOM (TCR 5) and Mid-Atlantic (TCR 2-4), the synchronized reserve requirement is the single largest unit. This is 1300 MW for TCR1, 1170 MW for TCRs 2-4, and 1170 MW for TCR5. The 1170 MW reserve for TCRs 2-4 is apportioned amongst the TCRs based on their loads. Reserve requirements are added to zonal loads.

# Validation

The PJM DAH market is a complex system influenced by countless factors, making it impossible to perfectly simulate prices. However, it is important to validate that PHORUM captures the important factors that determine electricity price and dispatch order. To validate the model, we construct a business as usual (BAU) scenario that simulates the 2010 DAH market. We then compare the LMPs from our BAU simulation to the actual 2010 DAH LMPs.

We measure accuracy with two metrics: mean hourly error and daily arbitrage error. The first gives a general sense of the model’s accuracy:

The second metric tracks how well PHORUM predicts the four highest and lowest priced hours of each day. Because storage profits in energy markets by arbitraging between high and low priced hours, it is important to predict them accurately. Multiple hours are used because storage devices cannot fully discharge in one hour.

*, where*

*X = the Xth highest/lowest priced hour of the day; e.g. X=2 is the 2nd highest and 2nd lowest priced hours.*

Table 6: Validation Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **TCR1** | **TCR2** | **TCR3** | **TCR4** | **TCR5** | **Total** |
| **Hourly Error (%)** | | | | | | |
| Mean | 1% | 0% | 8% | 5% | -15% | 0% |
| Standard Deviation | 25% | 24% | 24% | 24% | 21% | 23% |
| **Arbitrage Error [$/MWh]** | | | | | | |
| Mean | -3.4 | -4.4 | 2.8 | 2.8 | -17.2 | -3.9 |
| Standard Deviation | 13.1 | 29.8 | 29.1 | 29.5 | 17.2 | 23.7 |

PHORUM predicts hourly prices reasonably well, but consistently under-predicts prices in TCR5. The model consistently under-predicts arbitrage potential, and therefore under-predicts the value of storage.

PHORUM’s accuracy is highly sensitive to how much generation is offline for maintenance. PJM provided monthly data on maintenance outages, aggregated by generator type and month[[51]](#endnote-32). Including all of these outages caused large model errors by forcing too much capacity offline. Halving the outages resulted in the lowest average error. Therefore, the base model and validation results in Table 6 only include half of the offline generation identified by PJM.

Better data could further improve accuracy. In particular, better information on when generators are offline for maintenance, more detailed transmission constraints, and more refined TCRs would improve the model. We hope model users will continue to improve the model’s accuracy.

# Performance

The model was run in parallel on ten 8-core, 8 GB memory virtual machines at the Carnegie Mellon University Parallel Data Laboratory. Each 48-hour optimization took an average of 4 minutes; 365-day runs took 24 hours.

# Future improvements

PHORUM is a platform that researchers can adapt to their own needs, or continue to expand upon and improve. We have designed PHORUM to be flexible and easy to expand upon. Below are suggestions of future work for PHORUM:

## Accuracy Improvements

Better data could further improve accuracy. In particular, better information on when generators are offline for maintenance, more detailed transmission constraints, and more refined TCRs would improve the model.

## Runtime improvements

Improving the optimization’s formulation could decrease convergence time and overall runtime.

## More markets

PHORUM could be expanded to simulate more markets, such as the day-ahead scheduled reserve (DASR) market, real-time energy market, and regulation markets. Simulating more markets would both improve the model’s accuracy and expand the types of investigations that it could be used for.

## Capacity planning

PHORUM could be used as the basis for a capacity planning model. Such a model would investigate how the PJM generation fleet and operations might change over time.

1. PHORUM stands for PJM Hourly Open-source Reduced-form Unit Commitment Model [↑](#footnote-ref-1)
2. PJM eData. <https://edata.pjm.com/eContour/#app=ecca&e929-selectedIndex=2> (accessed June 29, 2012). [↑](#endnote-ref-1)
3. *A Comparative Analysis of Actual Locational Marginal Prices in the PJM Market and Estimated Short-run Marginal Costs: 2003-2006;* London Economics International LLC, 2007; <http://appanet.cms-plus.com/files/PDFs/LEIexecutivesummary2012007.pdf> (accessed Aug 3, 2012). [↑](#endnote-ref-2)
4. We make some changes because more data are now available. The LEI report includes three interfaces (Western, Central, and Eastern), and five regions. PHORUM includes three more interfaces: Bedington – Black Oak, AEP-DOM and AP South. We include an additional region, TCR5 (Dominion/VA), but do not model Delmarva Power and Light (DPL) as a separate region. Finally, we do not divide the METED zone across multiple regions. [↑](#footnote-ref-2)
5. PJM Zones. <http://www.pjm.com/about-pjm/how-we-operate/~/media/about-pjm/pjm-zones.ashx> (accessed Aug 6, 2012). [↑](#endnote-ref-3)
6. DUK (Duke Energy) zone was integrated into PJM Jan 1, 2012 and is not included in the analysis [↑](#footnote-ref-3)
7. We make the simplifying assumption that the capacity of each line is not dependent on how much current it is carrying. [↑](#footnote-ref-4)
8. *PJM State of the Market Report, 2010, Volume 2;* Monitoring Analytics, LLC, 2011; <http://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2010/2010-som-pjm-volume2.pdf> [↑](#endnote-ref-4)
9. We make three modifications: (1) The Western Interface is made up of four 500kV lines, each connecting different TCRs. Therefore, we divide the Interface’s capacity into quarters and apportion the capacity to TIs as appropriate; (2) The 500X(5004+5005) Interface is made up of two 500kV lines that are contained within the Western Interface. Therefore, we ignore the 500X(5004+5005) interface as it is included in the Western Interface; and (3) We combine the Bedington-Black Oak, AP South, and AEP-DOM Interfaces into a single TI between TCRs 1 and 5. [↑](#footnote-ref-5)
10. *Transmission and Voltage Emergencies;* PJM 2011; <http://pjm.com/training/~/media/training/core-curriculum/ip-ops-101/ops101-transemer.ashx> (accessed Aug 6, 2012). [↑](#endnote-ref-5)
11. On average, 2010 day-ahead load was 91 GW; real-time load was 80GW**Error! Bookmark not defined.**. [↑](#footnote-ref-6)
12. According to PJM, “The price difference between the Real-Time and the Day-Ahead Energy Markets results, in part, from volatility in the Real-Time Energy Market that is difficult, or impossible, to anticipate in the Day-Ahead Energy Market”iv. [↑](#footnote-ref-7)
13. In reality, 2010 DAH load was met by a combination of bilateral contracts (4.9%), self-supply from the load-serving entity’s own generation (75.8%), and spot purchases on the DAH market (19.3%)iv. This assumption is equivalent to assuming that bilateral contracts and self-supply do not cause out-of-merit-order dispatch. [↑](#footnote-ref-8)
14. Synchronized reserve requirements are equal to the largest unit in RFC (TCR1), the largest unit in the Mid-Atlantic control zone (TCRs 2-4), and the largest unit in Dominion (TCR5)xxx [↑](#footnote-ref-9)
15. PHORUM is not strictly “N-1” security constrained, as the modeled synchronized reserve requirements are sufficient for the grid to recover from only the largest generator contingency, but may not be large enough to cover the largest transmission contingency. [↑](#footnote-ref-10)
16. We assume demand is perfectly inelastic (does not change based on price) [↑](#footnote-ref-11)
17. Generators smaller than 5 MW are excluded [↑](#footnote-ref-12)
18. PJM contains four pumped hydro facilities: Bath County (VA, 2.8 GW), Yards Creek (NJ, 400 MW), Muddy Run (PA, 1 GW), and Smith Mountain (VA, 240 MW). The Bath County facility is the world’s largest. [↑](#footnote-ref-13)
19. *Resource Commitment and Dispatch in the PJM Wholesale Electricity Market;* M. Ward, 2011; <http://www.ferc.gov/eventcalendar/Files/20110628072854-Jun28-SesA2-Ward-PJM.pdf> (accessed Aug 3, 2012). [↑](#endnote-ref-6)
20. “DA objective is to minimize production cost over one day only. Such market design doesn’t handle day boundaries well . It is often up to DA operator to stop a unit with long minRun /minDown time or run through the midnight, manual adjustments for units limited by the number of starts per week”vi [↑](#footnote-ref-14)
21. Carrión, M., and J. M Arroyo. “A Computationally Efficient Mixed-integer Linear Formulation for the Thermal Unit Commitment Problem.” *Power Systems, IEEE Transactions On* 21, no. 3 (2006): 1371–1378. [↑](#endnote-ref-7)
22. *A Review of Generation Compensation and Cost Elements in the PJM Markets*; PJM, 2009; <http://www.pjm.com/~/media/committees-groups/committees/mrc/20100120/20100120-item-02-review-of-generation-costs-and-compensation.ashx> (accessed Aug 3, 2012). [↑](#endnote-ref-8)
23. Guan, X., P. B Luh, H. Yen, and P. Rogan. “Optimization-based Scheduling of Hydrothermal Power Systems with Pumped-storage Units.” *Power Systems, IEEE Transactions On* 9, no. 2 (1994): 1023–1031. [↑](#endnote-ref-9)
24. Streiffert, D., R. Philbrick, and A. Ott. “A Mixed Integer Programming Solution for Market Clearing and Reliability Analysis.” In *Power Engineering Society General Meeting, 2005. IEEE*, 2724–2731, 2005. [↑](#endnote-ref-10)
25. The primary differences are PHORUM’s simplified handling of reserve requirements and reduced-form topology [↑](#footnote-ref-16)
26. In reality, NOx and SO2 emission rates vary based on output level; in particular, emission rates are high when generators start up or shut down. PHORUM also does not account for variations in NOx output for ozone season. [↑](#footnote-ref-17)
27. AP2 (APEEP) Model – Nick Muller’s Homepage. <https://sites.google.com/site/nickmullershomepage/home/ap2-apeep-model-2> (accessed Aug 12, 2012). [↑](#endnote-ref-11)
28. On the DAH market, perfect information is the DAH forecast. The system operator has perfect information of the first 24 hours of each period (the DAH forecast), but not the hours 25-48 (the DAH forecast for the second day). [↑](#footnote-ref-18)
29. Optimizing storage operations over a longer period of time would increase the value of storage; however, accurately predicting load more than 48 hours out may be difficult. [↑](#footnote-ref-19)
30. EPA National Electric Energy Data System (NEEDS) v.4.10. <http://www.epa.gov/airmarkt/progsregs/epa-ipm/BaseCasev410.html#needs> (accessed Aug 10, 2012). [↑](#endnote-ref-12)
31. PJM 2010 EIA 411 Report. <http://www.pjm.com/documents/reports/~/media/documents/reports/2006-pjm-411-data.ashx> (accessed Aug 10, 2012). [↑](#endnote-ref-13)
32. Beaty, H. W., and Inc NetLibrary. *Handbook of Electric Power Calculations – Section 8: Generation of Electric Power*. McGraw-Hill, 2001. <http://energysystems.princeton.edu/EnergyResources/GenerElectPower__Shalaan.pdf> (accessed Aug 3, 2012). [↑](#endnote-ref-14)
33. EIA, Electric Power Monthly (2010), Table 4.10.A, 4.11.A, 4.13.A. <http://205.254.135.7/electricity/monthly/index.cfm> (accessed Aug 10, 2012). [↑](#endnote-ref-15)
34. EIA-423 Monthly Nonutility Fuel Receipts and Fuel Quality Data, 2010.: <http://www.eia.gov/cneaf/electricity/page/eia423.html> (accessed Aug 10, 2012). [↑](#endnote-ref-16)
35. North American Electric Reliability Corporation (NERC) Generating Availability Data System (GADS) <http://www.nerc.com/page.php?cid=4%7C43> (accessed Aug 10, 2012). [↑](#endnote-ref-17)
36. *Updated Operating Parameter Matrix;* J Bowring, 2007; <http://www.monitoringanalytics.com/reports/Presentations/2007/011107-rmwg.pdf> (accessed Aug 10, 2012). [↑](#endnote-ref-18)
37. *Estimating Power Plant Cycling Costs;* InterTek, 2011; <http://www.wecc.biz/Lists/Calendar/Attachments/3868/11082011_WECC_DWG_Lower.pdf> (accessed Aug 10, 2012). [↑](#endnote-ref-19)
38. *MRTU Market Power Mitigation: Options for Bid Caps for Start-Up and Minimum Load Costs;* CAISO Department of Market Monitoring, 2007; <http://www.caiso.com/1b87/1b87a5451d380.pdf> (accessed Aug 10, 2012). [↑](#endnote-ref-20)
39. *Scheduling Process and eMKT*; PJM, 2011; <http://www.pjm.com/~/media/training/core-curriculum/ip-gen-201/gen-201-scheduling-process.ashx> (accessed Aug 10, 2012). [↑](#endnote-ref-21)
40. Power Status Reports 2010; Nuclear Regulatory Commission;   
    <http://www.nrc.gov/reading-rm/doc-collections/event-status/reactor-status/2010/2010PowerStatus.txt> (accessed Aug 10, 2012). [↑](#endnote-ref-22)
41. 2005 National Emissions Inventory - ALLNEICAP Annual 113020078; EPA; <http://www.epa.gov/ttnchie1/net/2005inventory.html> (accessed Aug 10, 2013). [↑](#endnote-ref-23)
42. EPA eGRID2012. <http://www.epa.gov/cleanenergy/energy-resources/egrid/index.html> (accessed Aug 10, 2012). [↑](#endnote-ref-24)
43. PJM Historical Load Forecasts, 2010. <http://www.pjm.com/markets-and-operations/~/media/markets-ops/ops-analysis/2010-rto-forecasts-zip.ashx> (accessed Aug 10, 2012). [↑](#endnote-ref-25)
44. PJM Interchange Actual/Schedule Summary Report, 2010. <http://www.pjm.com/markets-and-operations/~/media/markets-ops/ops-analysis/2010-act-sch-summary.ashx> (accessed Aug 10, 2012). [↑](#endnote-ref-26)
45. PJM Monthly Locational Marginal Pricing. <http://www.pjm.com/markets-and-operations/energy/real-time/monthlylmp.aspx> (accessed Aug 10, 2012). [↑](#endnote-ref-27)
46. PJM Historical RTO Transfer Limit and Flows, 2010. <http://www.pjm.com/markets-and-operations/~/media/markets-ops/ops-analysis/2010-flows.ashx> (accessed Aug 10, 2012). [↑](#endnote-ref-28)
47. PJM 2010 Hourly Wind. <http://www.pjm.com/markets-and-operations/~/media/markets-ops/ops-analysis/2010-hourly-wind.ashx> (accessed Aug 10, 2012). [↑](#endnote-ref-29)
48. *Reserves Scheduling, Reporting, and Loading*; PJM, 2011; <http://www.pjm.com/~/media/training/core-curriculum/ip-ops-101/ops-101-reserves.ashx> (accessed Aug 10, 2012). [↑](#endnote-ref-30)
49. EIA **- Levelized Cost of New Generation Resources in the Annual Energy Outlook 2011.** <http://www.eia.gov/oiaf/aeo/electricity_generation.html> (accessed Aug 10, 2012). [↑](#endnote-ref-31)
50. Curtailment of Wind Farms Output Breakout Session; D Schweitzer; <http://renewableenergy.illinoisstate.edu/wind/conferences/speaker%20presentations/072111%20IWWG%20Annual%20Conference/2F%20PM%200130%20B2%20David%20Schweizer.pdf> (accessed Aug 10, 2012). [↑](#footnote-ref-20)
51. Bresler, S. PJM. Personal communication, 2012. [↑](#endnote-ref-32)