NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Time Domain Partitioning of Electricity Production Cost Simulations

Clayton Barrows, Marissa Hummon,
Wesley Jones, and Elaine Hale

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Technical Report
NREL/TP-6A20-60969
January 2014

Contract No. DE-AC36-08GO28308

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Time Domain Partitioning of Electricity Production Cost Simulations

Clayton Barrows, Marissa Hummon,
Wesley Jones, and Elaine Hale

Prepared under Task No. 0664.1301

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

Technical Report
NREL/TP-6A20-60969
January 2014

Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

Acknowledgments

This project was funded internally by the NREL Laboratory Directed Research and Development (LDRD) program. Model datasets in this analysis were derived from publicly available sources; individual utilities modeled in this work did not provide proprietary data or review our interpretation of the data, mode, methods, or results. The authors would like to thank Dr. Glenn Drayton and Energy Exemplar for the helpful feedback and technical assistance with the PLEXOS software package. The following individuals provided valuable input and comments during the analysis and publication process: Nate Blair, Paul Denholm, Jennie Jorgensen, Trieu Mai, David Palchak, Bryan Palmintier, Aaron Townsend and Maggie Mann. Any errors or omissions are solely the responsibility of the authors.

List of Acronyms

CC	combined cycle
CT	combustion turbine
DC	direct current
I/O	input/output
ISO	independent system operator
MIP	mixed integer programming
NRMSD	normalized root mean squared difference
PCM	production cost model
PSCO	Public Service Company of Colorado
PTDF	power transfer distribution factor
RMSD	root mean squared difference
RMPP	Rocky Mountain Power Pool
RTO	regional transmission operator
SAM	System Advisor Model
TEPPC	Transmission Expansion Planning Policy Committee
UC	unit commitment
WACM	Western Area Power Administration
WECC	Western Electricity Coordinating Council

Abstract

Production cost models are often used for planning by simulating power system operations over long time horizons. The simulation of a day-ahead energy market for a one-year horizon can take several weeks to compute. Tractability improvements are often made through model simplifications, such as reductions in transmission modeling detail, relaxation of commitment variable integrality, and reductions in cost modeling detail. One common simplification is to partition the simulation horizon so that weekly or monthly horizons can be simulated in parallel. However, horizon partitions are often executed with overlap periods of arbitrary and sometimes zero length. We calculate the time domain persistence of historical unit commitment decisions to inform time domain partitioning of production cost models. The results are implemented using PLEXOS production cost modeling software in a high performance computing environment to improve the computation time of simulations while maintaining solution integrity. Our results show that including two days of overlap on partitioned problems give adequate solution quality. Additionally, we achieve roughly 25-times speedup in the solver when optimizing 52 weeks in parallel, with overlap, versus a single annual optimization.

Table of Contents

List of Figures

List of Tables

1 Introduction

A production cost model (PCM) simulates the operation of generation and transmission systems by finding, during each time interval, the least-cost solution to generating sufficient energy to meet demand. The solution is restricted by transmission constraints (e.g., linearized de-coupled optimal power flow [DCOPF]), generator operation properties (e.g., startup costs, ramp rates, and maintenance outages), and reserve requirements to meet intra-interval generation-demand imbalances. Utilities, independent system operators (ISOs), and regional transmission operators (RTOs) run PCMs for planning purposes (e.g., to test feasibility of capacity or transmission expansion) and during daily/hourly operation to schedule generating units. Integration studies require the detailed simulation of the system during every hour of multiple years to adequately quantify operations and reliability of proposed projects. This process requires running many annual PCM simulations to assess each integration option under a variety of scenarios. PCMs present a set of mixed integer programming (MIP) problems with dynamic constraints coupling time steps. PCM computation times can often extend into multiple days for large systems. Due to the dynamic nature of generator operating constraints, the integrity of PCM optimization problems is dependent upon the simulation horizon over which the problem is defined. Typically, PCMs are used to simulate operations on an annual horizon to capture the seasonal variability of load and renewable resource availability. However, in practice, it is common to reduce the computational complexity of annual PCM simulations by partitioning the simulation horizon into monthly problems (General Electric Company 2013; Energy Exemplar 2013). Often, these discrete partitions are made with little or no consideration of the system boundary conditions at the endpoints of the partitioned simulation horizons. This paper explores the tradeoffs associated with time domain partitioning of PCM simulations.

While PCMs are models of power system operations, they are commonly used for planning studies and policy analysis. For instance, a PCM can simulate energy and reserve prices for a variety of different generation portfolios or market structures. Traditionally, a single deterministic solution is given based on a PCM simulation horizon. Some studies simulate a suite of scenarios highlighting the system sensitivity to certain input variables (e.g., wind or solar generation profiles) where each simulation is computed simultaneously (Meyers et al. 2011). Other studies require that scenarios be computed serially so that results can inform future simulations. Both cases would benefit from a reduction in computation time via parallelization of simulations. The increasing availability of inexpensive computer hardware and parallel processing platforms provide resources that remain relatively untapped by PCM software. This research focuses on the computational enhancement of a single PCM simulation by partitioning simulation horizons. We present results on real systems using a commercial PCM software package (PLEXOS). PLEXOS was created by Energy Exemplar in 2001and has been developed into a highly extensible PCM package that enables high fidelity modeling of emerging technologies (e.g., storage and renewables). PLEXOS extends the standard features of many production cost modelling software packages by providing the capability to simulate operations at sub-hourly time resolution (Ventyx 2013) (General Electric Company 2013).

Equations 1–7 describe the core PCM optimization problem. The objective function in Equation 1 considers both operating (c) and startup costs (sc) associated with each generation unit (k) for each time period (t). Equation 2 represents the system load balance constraint where the sum of generation output (g) for committed (u) generators must equal the sum of all system loads (l)

plus losses. Equations 3–5 model generator behavior: constraining committed generator outputs to within minimum ($g_k^{t,MIN}$) and maximum ($g_k^{t,MAX}$) output levels, temporal shifts in generation to within ramping capabilities ($ramp_k^{t,MAX}$), and maximum energy production to within daily, weekly, monthly, and yearly limits ($g_h^{day|week|month|year\_max}$). Equations 6 and 7 model electricity transmission using the linearized DC power flow method (Grainger and Stevenson 1994), where *f* represents the megawatt power flow along the system transmission lines. The power transfer distribution factor matrix ($PTDF^s_{k,j}$) describes the distribution of power injected at bus$_k$ and withdrawn at the slack bus (s) that flows along line$_k$. Additional constraint equations are added to this core PCM formulation to model behavior, such as forced outages, generator up/down time constraints, fuel constraints, emissions constraints, and reserve/security constraints.

$$\text{Min} \sum_t \sum_k c_k^t g_k^t u_k^t + sc_k^t (u_k^t - u_k^{t-1}) \quad (1)$$

s.t.

$$\sum_k g_k^t u_k^t = \sum_k l_k^t + \sum_j loss_j^u \quad \forall \text{ t} \quad (2)$$

$$\sum_k g_k^{t,MIN} u_k^t \le g_k^t \le g_k^{t,MAX} u_k^t \quad \forall \text{ t, k} \quad (3)$$

$$g_k^t - g_k^{t-1} \le ramp_k^{t,MAX} u_k^t \quad \forall \text{ t, k} \quad (4)$$

$$\sum_k g_h^t \le g_h^{day|week|month|year\_max} \quad \forall \text{ h} \quad (5)$$

$$f_j^t = \sum_k PTDF_{k,j}^s (g_k^t u_k^t - l_k^t) \quad \forall \text{ t, k, s} \quad (6)$$

$$f_j^{t,MIN} \le f_j^t \le f_j^{t,MAX} \quad \forall \text{ t} \quad (7)$$

PCMs are typically utilized to simulate the annual operation of a power system, and are thus run in annual simulation horizons that are traditionally computed by running a single simulation that finds the optimal solution of each time step given the solution of previous time steps. Without explicit consideration of the simulation start conditions (i.e., unit commitment and generator up times), such annual simulations are themselves approximations of how power systems actually operate. For example, if an annual simulation starts on midnight of January 1 without information about system operation in the previous year, no information about how long generators have/have not been operating is passed to beginning time steps[1], and thus the simulation cannot represent the inter-temporal constraints that define generator operation. Inter-temporal constraints are required to represent the time that it takes to turn on/off a generator and to change the output of a generator. While the accuracy of PCM simulations is dependent, in part, on the enforcement of these inter-temporal constraints, the effects of enforcement in earlier time steps likely diminish. For example, while a certain generator may have a binding minimum on time sometime during January, that constraint is unlikely to affect that generators operation in August. Additionally, the inter-temporal constraints require the sequential simulation of each time step in the horizon, which can lead to extremely long computation times. This theory guides our approach to determine an appropriate partitioning method to speedup PCM simulations.

This paper explores the effects of two key simulation start parameters on the results of partitioned PCM simulations. With reference to an un-partitioned annual simulation, we explore how the time points of partition boundaries and the partition overlap duration (or spin up) affects the partitioned simulation results. Figure 1 shows an illustration of the annual simulation horizon partitioned into four, roughly quarterly, horizons. The partitioned simulations commence at the time corresponding to the beginning of the overlap period and extend the duration of the

[1] By default, PLEXOS considers all inter-temporal constraints non-binding at the beginning of simulation horizons.

partitioned horizon. The simulation results corresponding to the overlap period before the partition boundary are discarded to yield a single complete set of annual simulation data resulting from the completion of all partitioned simulations. We vary the number and location of partition boundaries and the duration of the overlap periods and analyze the effects of these two parameters on the solution quality and run time of time domain partitioned PCM simulations.

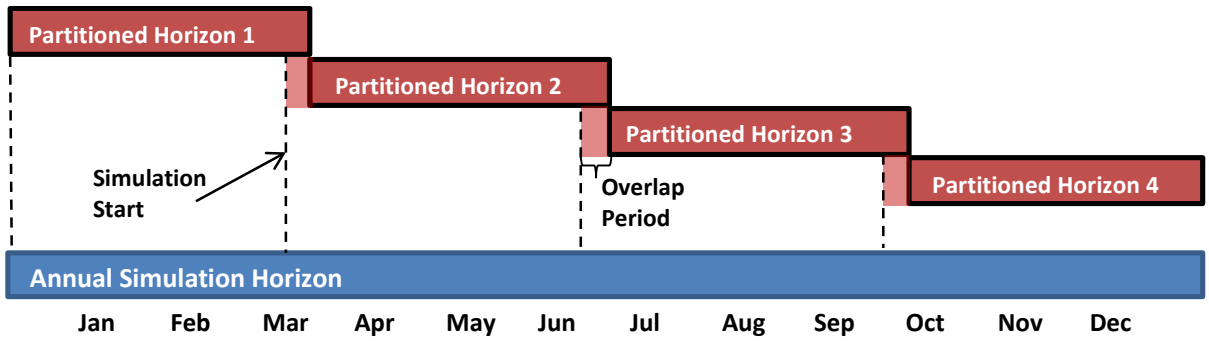

Figure 1. Illustration of annual simulation horizon versus quarterly partitioned simulation horizons. Partitioned simulations start at the beginning of the overlap period and data simulated before the partition boundary is discarded.

2 Solution Time Reduction

Time domain decomposition of PCM simulations is a method for reducing the time to solution for problems that contain many optimization intervals that are tied together by inter-temporal constraints (e.g., generator ramping constraints). By optimizing distinct partitions simultaneously (i.e., in parallel) on separate pieces of computer hardware, solutions can be achieved with a speedup relative to the number of partitions. Amdahl's law argues that the maximum speedup for a given number of partitions occurs when the computation time for each partition is equivalent (Amdahl 1967). Assuming the PCM solution time is dominated by the optimization of the set of MIP steps and that each MIP step can be solved in roughly the same time, a near linear speedup relative to the number of partitions is achievable without overlap. In a first attempt to create partitions with roughly equal solution times, we decomposed the horizon into partitions with equal simulation time horizons (e.g., weekly or monthly). Achieving a linear speedup of an annual simulation decomposed into week-long partitions would allow the simulation of problems that once took a month to be completed in less than a day.

Maximum speedup of time domain decomposed simulations would occur when the simulation is partitioned into simulations the same length as each optimization step (e.g., an annual day-ahead market simulation would achieve maximum speedup when partitioned into 365 single day simulations). Additional problem constraints and characteristics limit the achievable speedup and resultant reduction in runtime. To achieve valid solutions, it is necessary to not create partitions so small as to not capture the longer inter-temporal generation constraints, such as minimum/maximum generator run times. The disruptions caused by the initial conditions of each partitioned horizon may introduce significant inaccuracies because the inter-temporal constraints are not informed by the solutions to previous time steps. To minimize some of the inaccuracies caused by partitioning, redundant computation (e.g., an initialization period at the beginning of each partition) may be required.

Speedup of solution times will be limited by redundant and serial computation requirements, such as model initialization, solution input/output (I/O), reconstruction of the full annual solution from the individual partitions, and writing solutions to disk. For the remainder of this document, we assume that PCM solution time is dominated by the MIP optimization steps, thus the serial components of the simulation time (compilation, validation, reconstruction, and saving) are ignored. Additionally, the computational complexities and time to solution of individual interval or partition solutions are not necessarily equal and may depend on a number of factors. Thus, computational load balancing may also be important.[2]

[2] A point that is made more relevant by our observations that solution times become increasingly uncertain as the MIP gap tolerance is made more precise.

3 Research Approach

3.1 Partitioned Solution Quality

Solutions to the PCM optimization problem are sensitive to input data and many simulation parameters. Figure 2 and Figure 3 show how the generation dispatch results of an un-partitioned annual simulation are changed by solving independent months. Figure 2 shows a generation dispatch stack snapshot from an un-partitioned annual simulation[3], and Figure 3 shows how the partitioned simulation generation deviates from the un-partitioned results. To quantify these differences and benchmark the effects of varying simulation input parameters, we develop as set of comparison criterion metrics.

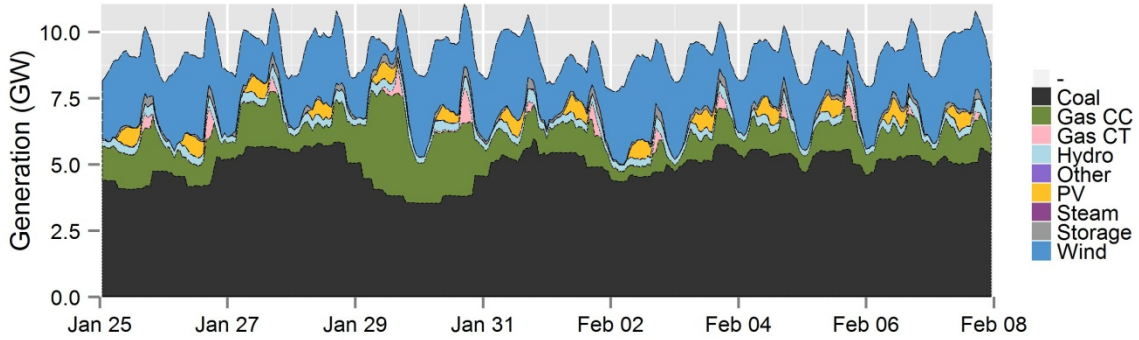

Figure 2. Generation dispatch stack by generator type for the time window starting January 30 and ending February 8

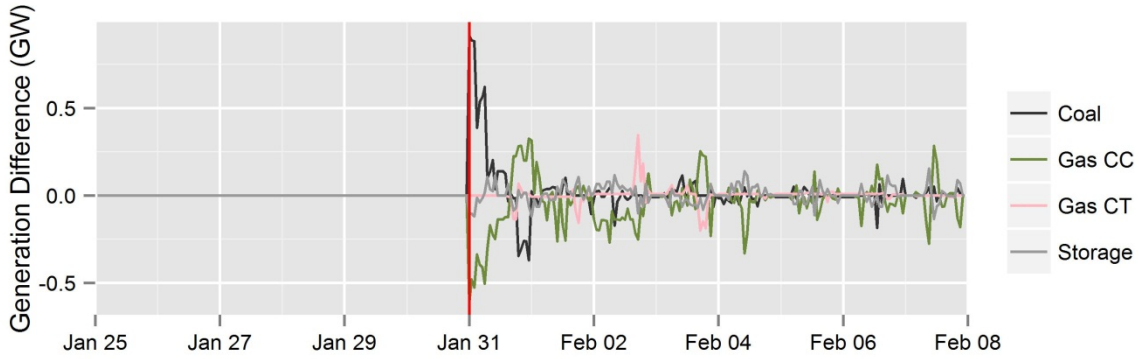

Figure 3. Generation dispatch differences between an annual simulation and a monthly partitioned simulation

Solution quality can be defined using a number of different metrics. The objective of PCMs is to minimize the total system cost. Therefore, a logical solution quality metric exists in comparing the total system cost over the simulation period, with and without parallelization. However, because PCM simulations result in a series of MIPs where each MIP solution contributes to the overall production cost, PCM solutions may have equivalent objective function values while

[3] See Section 3.2 for details on the modeled system.

their solutions differ. Thus, other solution quality metrics need to be considered: total variation of generation dispatch or commitment variables, locational marginal prices (LMPs), and capacity factor by generator type. Here, we aim to measure and quantify the uncertainty introduced by time domain partitioning. Equivalent simulations would yield equivalent dispatch values for every generator during every time step. However, due to common modeling simplifications (e.g., simplified transmission constraints and constant/linear generator heat rate curves) and the similarity of many generator costs, there is no expectation to achieve perfectly equivalent solutions with even minor perturbations in simulation technique. For example, Figure 4 shows differences in generation from two simulations that differ only in the generator random outage seed definition. Rather than a binary test of equivalence, the difference between two solutions is described in (8) using the root mean square deviation (RMSD) of generation.

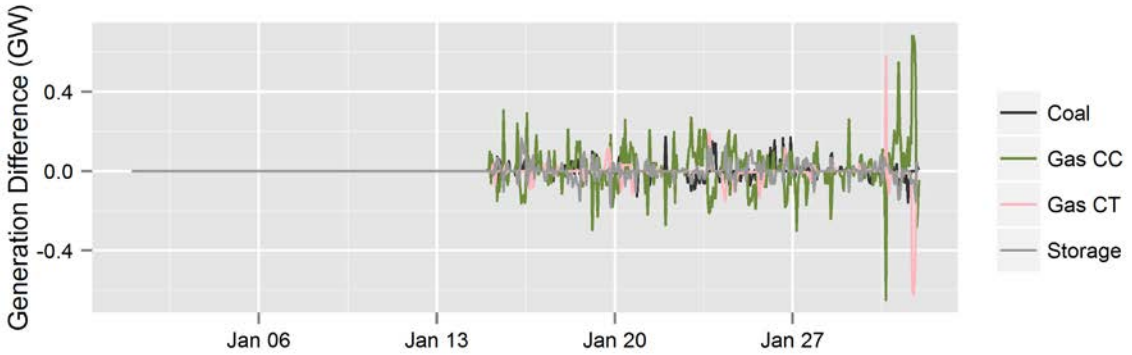

Figure 4. Generation dispatch differences between two simulations with different random generator outage seeds

$$RMSD(g) = \sqrt{\frac{\sum_{t=1}^{n}\left((g_t - g_t')^2\right)}{n}} \tag{8}$$

Where, g and g' represent the outputs of a specific generator resulting from an un-partitioned reference simulation and a partitioned simulation, respectively. The RMSD metric quantifies the difference in the dispatch of specific generators. The RMSD metric can also be calculated for differences in committed generation, generation cost, or differences in reserve provision (the megawatt capacity of a specific generator that is procured for a particular reserve product).

Equation 9 defines a normalized RMSD that quantifies the deviation of total generation from a group of generators (S) as a function of system demand (L). The NRMSD metric enables the analysis of all generators or generators categorized by generator type. The definition of NRMSD in Eq. 9 aggregates individual generator errors to allow generation substitutions within generator types. For example, in a system with two coal generators, the NRMSD metric would remain zero despite a shift in generation between the two coal generators, as long as the total coal generation dispatch remained equivalent between simulations for a given time period. A method that preserves individual generator errors is analyzed in the Appendix.

$$NRMSD\big(G(\mathrm{S})\big) = \sqrt{\frac{\sum_{t=1}^{n}\left(\sum_{g\in \mathrm{S}}\frac{g_t - g_t'}{L_t}\right)^2}{n}} \tag{9}$$

Quantifying and analyzing the differences, or similarities, of multiple PCM simulations depends upon the specific goals of the study. Some studies may be more interested in the optimal reserve provision, while others may only be interested in total system cost. To develop a generic metric for differences between PCM simulations, we analyze the differences in three sets of outputs resulting from simulating the operations of the identical systems. To capture the overall difference of the set of objective function values of two PCM simulations, the NRMSD of the reserve provisions (10) and production costs (11). In Eq. 10, *R* represents the sum of reserve provisions for the set of generators (*S*) over the time periods (t∈{1,…,n}) and *r* represents the reserves provided by an individual generator during an individual period. Equation 11 can be calculated for total production cost or for the components that make up production cost (fuel, startup, and variable operation and maintenance [VOM]).

$$NRMSD\big(R(S)\big) = \sqrt{\frac{\sum_{t=1}^{n}\left(\sum_{r\in S} r_t - r'_t\right)^2}{n\left(\sum_{t=1}^{n}\sum_{r\in S} r_t\right)^2}} \tag{10}$$

$$NRMSD\big(C(S)\big) = \sqrt{\frac{\sum_{t=1}^{n}\left(\sum_{c\in S} c_t - c'_t\right)^2}{n\left(\sum_{t=1}^{n}\sum_{c\in S} c_t\right)^2}} \tag{11}$$

3.2 Test System

We evaluate the impact of partitioning on an electric power system on a test system composed of two balancing areas largely in the State of Colorado. The test system is described extensively in (Denholm and Hummon 2012; Denholm et al. 2013; Hummon et al. 2013). The Colorado test system consists of two balancing areas (Public Service of Colorado [PSCO] and Western Area Colorado Missouri [WACM]) using data derived from the database established by the Western Electricity Coordinating Council (WECC) Transmission Expansion Policy Planning Committee (TEPPC) model and other publicly available datasets. Transmission is modeled zonally, without transmission limits within each balancing authority area. Projected generation and loads were derived from the TEPPC 2020 scenario (TEPPC 2011). Hourly load profiles were based on 2006 data and scaled to match the projected TEPPC 2020 annual load. Hourly solar and wind power generation profiles are time synchronized to the load profiles for the year 2006. The system peaks in the summer with a 2020 coincident peak demand of 13.7 GW and annual demand of 79.0 TWh. A total of 201 thermal and hydro generators are included in the test system, with total capacities listed in Table 1. We adjusted the conventional generator mix to ensure the available capacity (after outages) was always at least 9% greater than demand by adding a total of 1,450 MW of natural gas fueled generation (690 MW of combustion turbines and 760 MW of combined cycle units). This adjustment was necessary, in part, because the simulated system does not include contracted capacity from surrounding regions or any capacity contribution from solar and wind resources. The base case of the test system assumes a wind and solar penetration of 16% on an energy basis. For comparison, Colorado received about 11% of its electricity from wind in 2012 (EIA 2013).[4] PV profiles were generated using the System Advisor Model (SAM) (Gilman and Dobos 2012) with 2006 meteorology. Wind data was derived from the Western Wind and Solar Integration Study (WWSIS) dataset (GE 2010).[5] Discrete wind and solar plants

[4] Colorado generated 6,045 GWh from wind in 2012 compared to total generation of 53,594 GWh. See (EIA 2012).
[5] All generation profiles were adjusted to be time synchronized with 2020, which is a leap year.

were added from the WWSIS datasets until the installed capacity produced the targeted energy penetration.[6]

Table 1. Test System Generator Capacity in 2020

	System Capacity (MW)
Coal	6,178
Combined Cycle (CC)	3,724
Gas Turbine/Gas Steam	4,045
Hydro	773
Pumped Storage	560
Wind	3,347 (10.7 TWh)
Solar PV	878 (1.8 TWh)
Demand Response	*293*
Other[a]	513
Total	15,793

[a] Includes oil- and gas-fired internal combustion generators.

Fuel prices were derived from the TEPPC 2020 database. Coal prices were $1.42/MMBtu for all plants. Natural gas prices varied by month and ranged from $3.90/MMBtu to $4.20/MMBtu, with an average of $4.10/MMBtu (EIA 2012). No constraints or costs were applied to carbon or other emissions.

We generated hourly requirements for contingency, regulation, and flexibility reserves.[7] Contingency reserves are based on the single largest unit (a 810 MW coal plant) and allocated with 451 MW to PSCO and 359 MW to WACM, with 50% met by spinning units.[8] Regulation and flexibility reserve[9] requirements vary over time based on the statistical variability of load, wind, and PV, with the methodology described in detail by Ibanez et al. (2012). Hummon et al. (2013) describes the application of the methodology to the test system.

The sum of the total operating reserves (met by spinning units) averages 582 MW, which corresponds to about 6.4% of average load. Table 2 summarizes the general characteristics of the three modeled reserve services. Reserves were modeled as "soft constraints," meaning the system was allowed to not meet requirements if the cost exceeded the high threshold value shown in Table 2. These penalties were chosen to be high enough so that least cost decision is to

[6] The sites were chosen based on capacity factor and do not necessarily reflect existing or planned locations for wind and solar plants.
[7] For additional discussion of these reserves (especially flexibility reserves, which is not yet a well-defined market product), see (Ela et al. 2011).
[8] The PSCO and WACM balancing areas are part of the Rocky Mountain Reserve group, which shares contingency reserves based on these values.
[9] For these services, only the "upward" reserve requirements were evaluated. The need for downward reserves becomes of greater importance at high renewable penetration when conventional thermal generators are operated at or near their minimum generation points for more hours of the year. Future work will evaluate the cost and price of separate up and down reserve products in these scenarios.

start a new unit to provide reserves. Unserved load was also modeled as a soft constraint, with a penalty price of $10,000/MWh.

Table 2. Summary of Operating Reserves in the Base Case of Test System

Operating Reserve Service	System Drivers	Time to Respond [min]	Requirement (% of load) Mean (Min/Max)	Penalty [$/MW-h]
Regulation	PV, wind, load	5	1.33 (1.00/1.71)	9,500
Contingency	largest generator	10	4.54 (2.97/5.95)	9,000
Flexibility	PV, wind	20	0.64 (0.13/1.07)	8,500

The PLEXOS simulations performed in this analysis used day-ahead scheduling with a 48-hour optimization window, rolling forward in 24-hour increments. The extra 24 hours in the unit commitment horizon (for a full 48-hour window) were necessary to properly commit the generators with high start-up costs and the dispatch of energy storage. All scenarios were run for one chronological year using PLEXOS version 6.207 R08, with the model performance relative gap set to 0.5%.

4 Unit Commitment Decision Persistence

The partitioning method described in Figure 1 provides an effective method to partition PCM simulations for parallel execution that achieves roughly linear speedup proportional to the number of partitions. Without overlap, the solution quality is not necessarily preserved. Generator operation is governed by physical constraints, such as minimum and maximum run times, down times, and ramp rates. These dynamic constraints provide a mechanism by which optimal unit commitment decisions made in previous time steps carry forward and affect future unit commitment decisions. The characteristics that drive this commitment persistence vary by generator operating parameters, which are largely governed by fuel source, thus providing a logical grouping for analyzing unit commitment persistence. For instance, the modelling parameters of a coal-fired generator reflect the time it takes to bring a unit up to temperature produce enough steam to begin generating electricity. Similarly, the time required to change the amount of electricity produced (dispatch level) reflects the required temperature and pressure changes needed to achieve the new generation dispatch point. While these constraints are unique to a particular generator, coal-fired steam turbines typically have similar operating characteristics. By comparison, a natural gas-fired combustion turbine does not require steam production and thus requires less time to startup, shutdown, and change its operating point than a coal-fired generator. Figure 5 shows the autocorrelation of the number of committed generators by fuel type for a variety of time snapshots (seasons) throughout the year. The horizontal axis shows autocorrelation lags in hours up to 1 week. The autocorrelation analysis shows that the time it takes for unit commitment decisions to decorrelate varies by generator and by season.

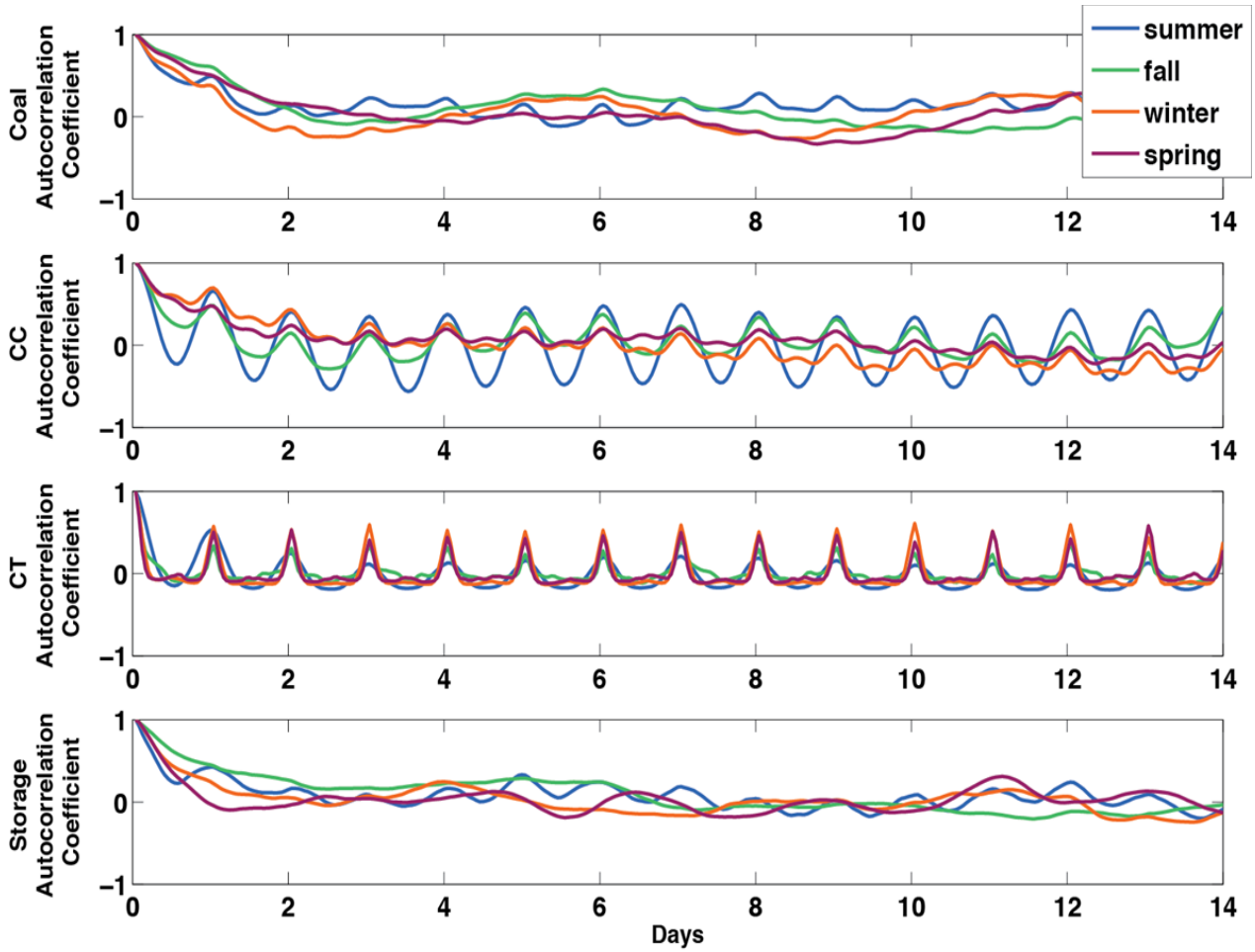

Figure 5. Autocorrelation of generator unit commitment by fuel type and by season

The autocorrelation analysis in Figure 5 suggests that commitment decisions of coal generators and storage units persist for roughly 1–2 days throughout the year while the persistence of other generators (gas combustion turbine [CT] and combined cycle [CC]) is dependent on the time of year. The intervals of high correlation evident in the gas CT and CC autocorrelations suggest that the commitment status may be relatively predictable for gas CTs and CCs during the summer and fall. These results suggest two strategies for achieving high quality solutions through time domain partitioning. The first strategy is to start the simulations far enough ahead of the partition boundaries so that the system has enough time (overlap period) to reduce the effects of the start conditions. The sensitivity of the system to the overlap period and the timing of partition boundaries are explored in the remainder of the document and are informed by the autocorrelation analysis in Figure 5 that suggests the majority of the system commitment decisions become decorrelated after roughly 2 days. The second strategy to enhance partitioned simulation solution quality is to seed the start conditions of each partitioned simulation with an informed system state. This could be done by defining, among other variables, the unit commitment pattern at the beginning of each simulation. This strategy is the subject of ongoing research efforts.

5 Start Overlap and Boundary Location Sensitivity

The analysis presented in Section 4 demonstrates the persistence of various unit commitment decisions in a PCM. Here we investigate the effects of varying both the partition boundary (i.e. the days on which simulations are partitioned) and the overlap period of PCM simulations. To investigate the effects of these two parameters on partitioned PCM simulations, we analyze a suite of test simulations with reference to an un-partitioned annual "base case" simulation where, other than simulation horizon, all other model parameters are equivalent. The suite of test simulations is generated by simulating 30 days of operation starting at midnight on each day of the year. This process results in a suite of 336 thirty-day simulation results. The 30-day horizon is chosen to provide enough data to allow analysis of overlap periods of several days to several weeks. By analyzing the differences between each simulation in our test suite and the base case, we can quantify the effects of varying the overlap period. The effect of varying the number of overlap days is described by the daily RMSDs in normalized generation dispatch, normalized reserve provision, and generation cost of coal, gas CC, and gas CT generators[10] as functions of the number of days since the start of each simulation (overlap days).

Figure 6 shows that coal and gas CC generation differences are significant for up to 2 days of overlap. The figures describe the distribution of 336 sample differences for 1–30 days of overlap where the red dot represents the median and the bottom and top of the whiskers represent the 25th and 75th percentiles. After the partitioned simulations experience 2 days of spin up, differences in generation dispatch are largely minimized. The density plots on the left detail the shape of each distribution described by the box plots on the left for 0 and 3 days overlap. Relative to the 0-day overlap plots, the increased occurrence of smaller differences in the 3-day overlap plots shows that the majority of generator dispatch differences are very small after the simulation has some time to spin up.

[10] While other generator types exist in the Rocky Mountain Power Pool (RMPP) system, the system is dominated by three generator types (coal, gas CC, and gas CT), and storage operation is particularly difficult to model, thus figures for other generation types are omitted.

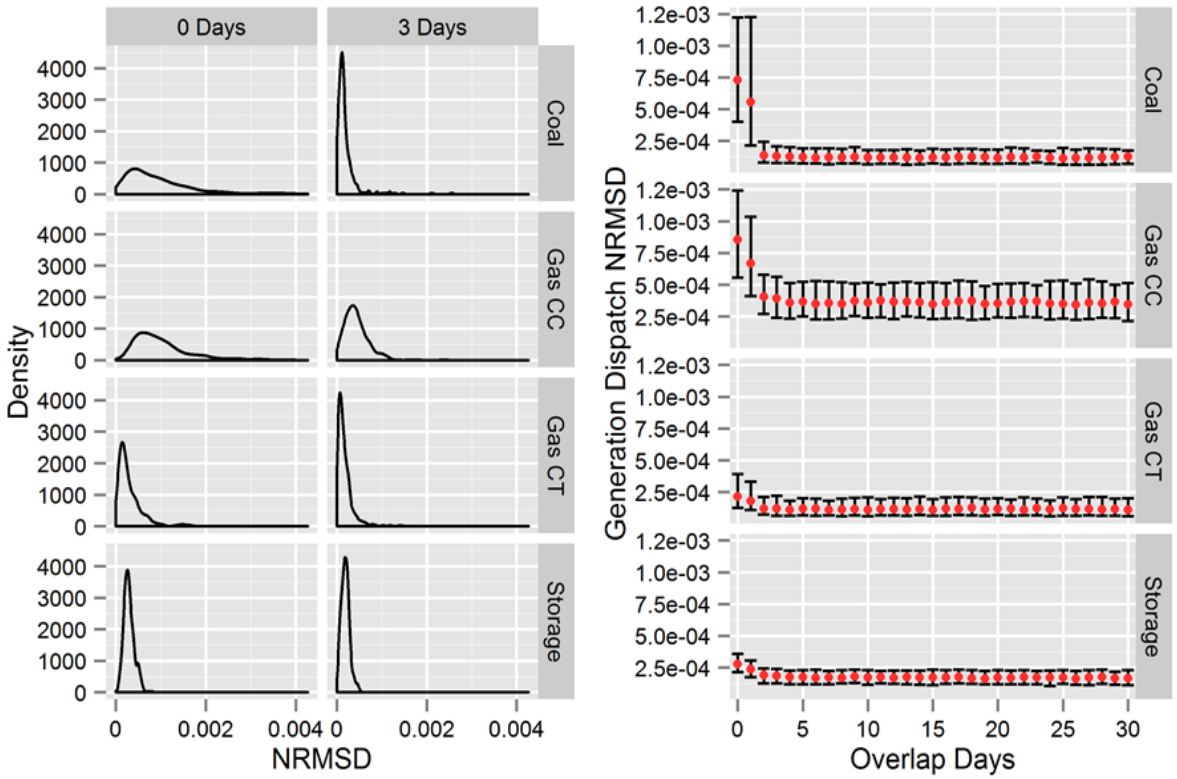

Figure 6. (Left) Density of generation dispatch normalized root mean squared difference (NRMSD) for 0 and 3 days of overlap; (Right) Generation dispatch NRMSD as a function of the number of days since the partitioned simulation has started where the lower and upper whisker edges represent the 25th and 75th percentiles and the red dots represent the medians

The seasonal variation of errors is demonstrated in Figure 7 where the mean NRMSD of generation is plotted for the 2020 calendar versus the number of overlap days. The darkest tiles for each generator technology show that differences in generation typically persist for two overlap days. The figure again shows that differences are largely minimized after two days of overlap. Additionally, the seasonal variation shows smaller differences in late July and August. This result suggests that fewer unique optimal solutions exist during the periods of highest load than when the system has more available generation capacity (i.e., when [virtually] all generators are operating, the number of possible commitment decisions is reduced).

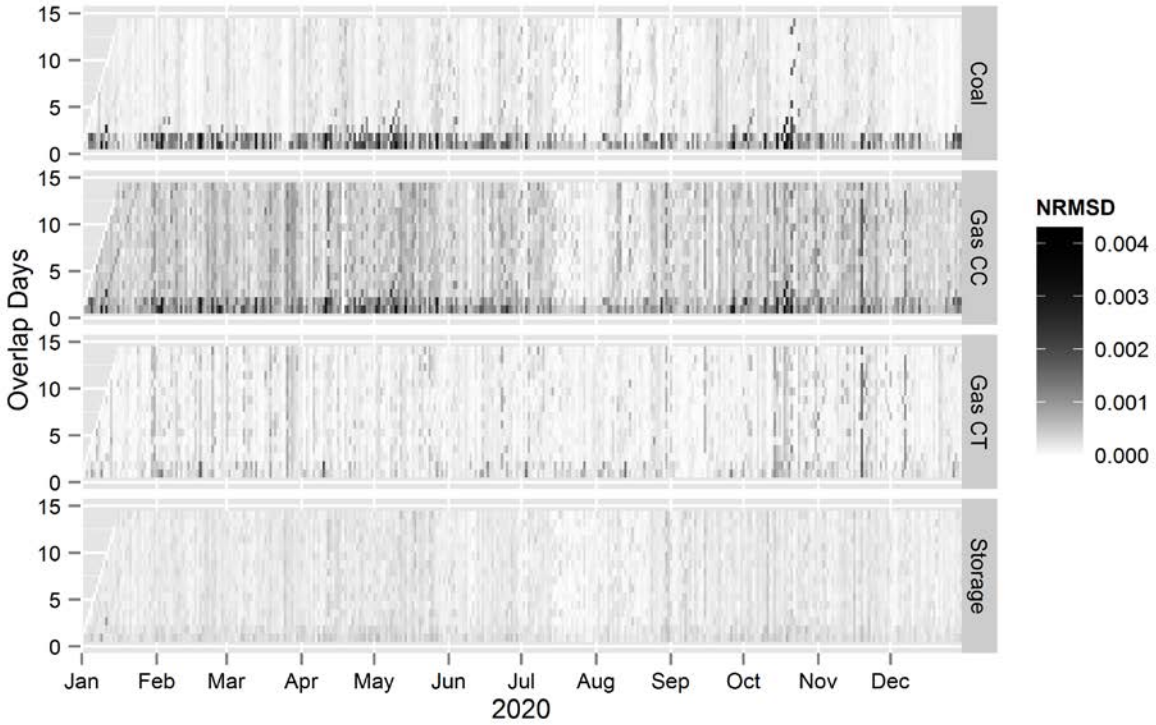

Figure 7. Mean generation normalized root mean squared difference values as a function of the number of days since the partitioned simulation has started and the date

Figure 8 shows similar results for the provision of the three reserve products: contingency reserves, and upward flexibility and upward regulation reserves for the RMPP system. Reserve NRMSD values again converge toward a minimum value after roughly two days. Figure 9 demonstrates a similar, although more pronounced, seasonal dependency of upward regulation reserve provision NRMSD values from coal generators to the pattern demonstrated in Figure 7. The NRMSD tiles again show a slightly darker band during the first 2 days of overlap. The zero values and blank shading for gas CT Regulation in Figures 8 and 9 are because we don't allow CTs to provide regulation reserves in the model.

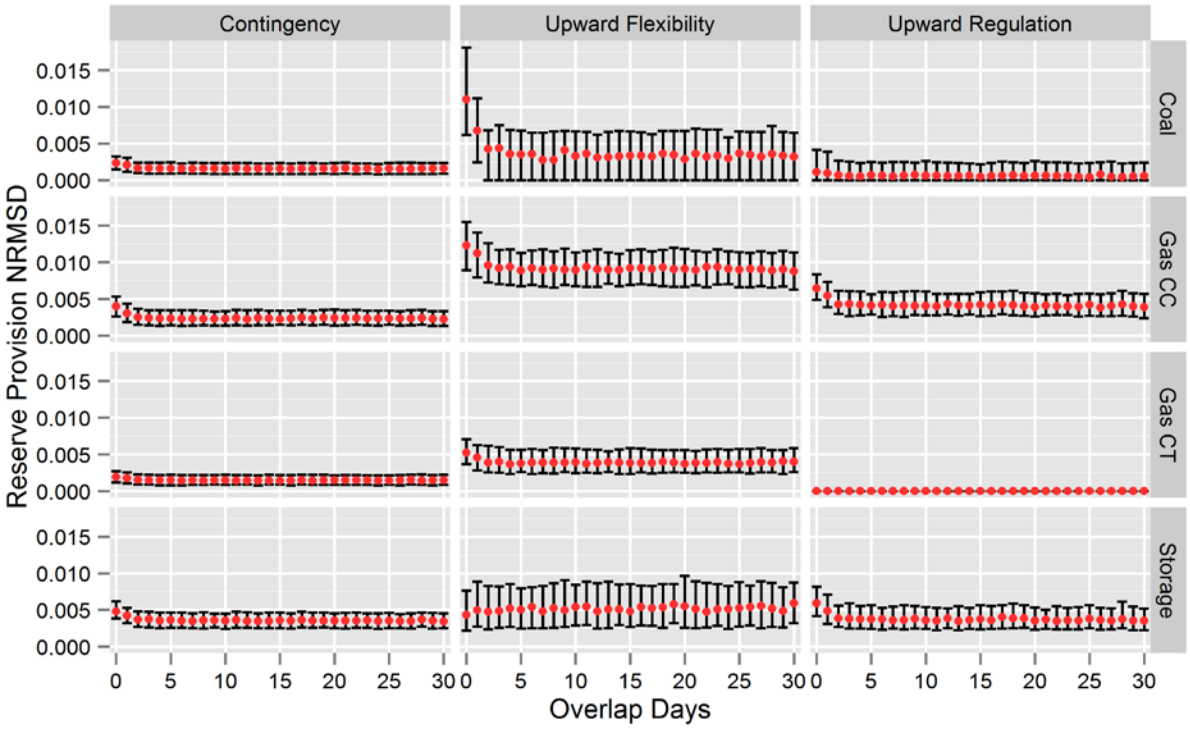

Figure 8. Reserve provision NRMSD as a function of the number of days since the partitioned simulation has started where the lower and upper whiskers represent the 25th and 75th percentiles and the red dots represent the medians

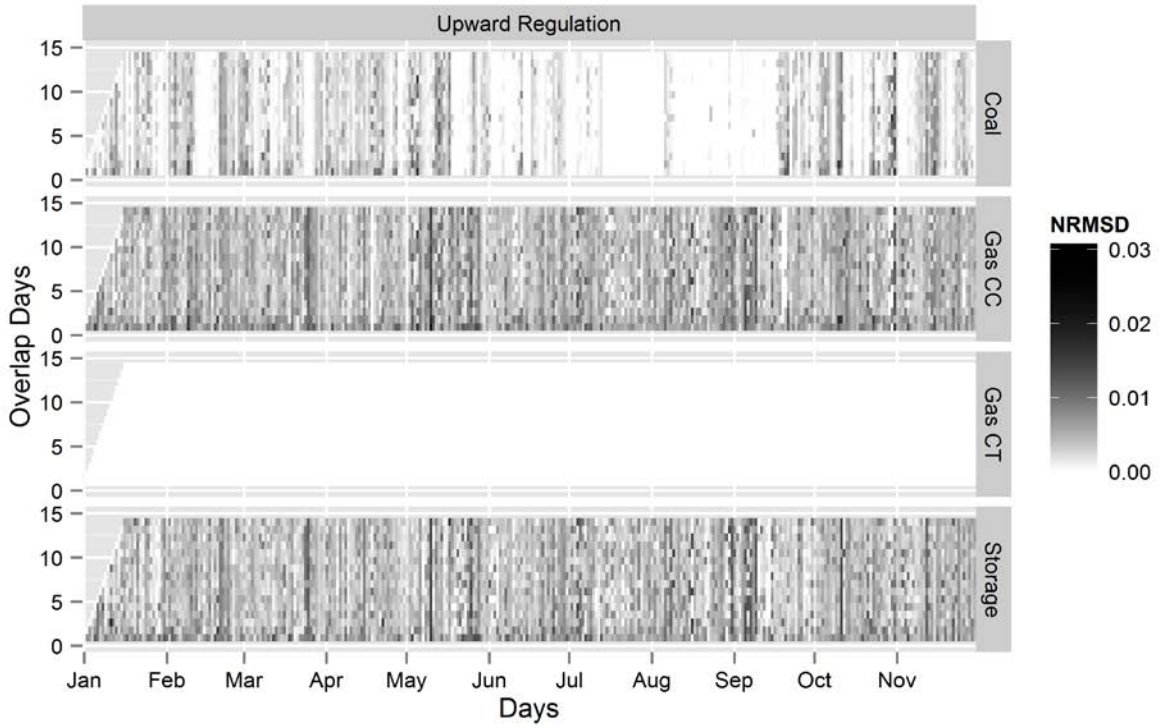

Figure 9. Seasonal reserves NRMSD heat map as a function of the number of days since the partitioned simulation has started and the date

Figure 10 shows the generation cost differences between the base case and the partitioned simulations. Total generation costs (Total) are broken down into startup and shutdown costs (Start), VOM, and fuel costs (Fuel). While, the total cost values are driven primarily by the fuel cost values, the total cost NRMSD values are driven by differences in startup costs. The largest deviations in startup cost can be attributed to the initial conditions of the simulation horizon in the PCM optimization. At the beginning of each partitioned simulation, all generators are committed by default, without an associated startup cost. Many generators must be de-committed during the first day of the simulation. Partitioned simulations commitment patterns converge to the annual simulation as the partitioned simulation inter-time generator commitment constraints become binding; this behavior is amplified by the startup costs of different generation technologies and is evident in Figure 10 and Figure 11. The relatively (and consistently) high NRMSD values for Gas CC and Storage generators is likely due to the high level of degeneracy (cost similarity) in Gas CCs and extreme sensitivity of Storage operation to even minor simulation perturbations. Figure 11 again shows a seasonal dependence of NRMSD values and again highlights the errors that occur before two days of overlap with the dark band at the bottom of the tile representing coal generation. However, unlike reserve provisions where we see a strong seasonal dependence in coal generators, the cost NRMSD seasonal dependence is not nearly as dramatic as the overlap period dependence. By contrast, the gas CC seasonal dependence is more evident than the overlap dependence. These results are consistent with the autocorrelation results of Figure 5.

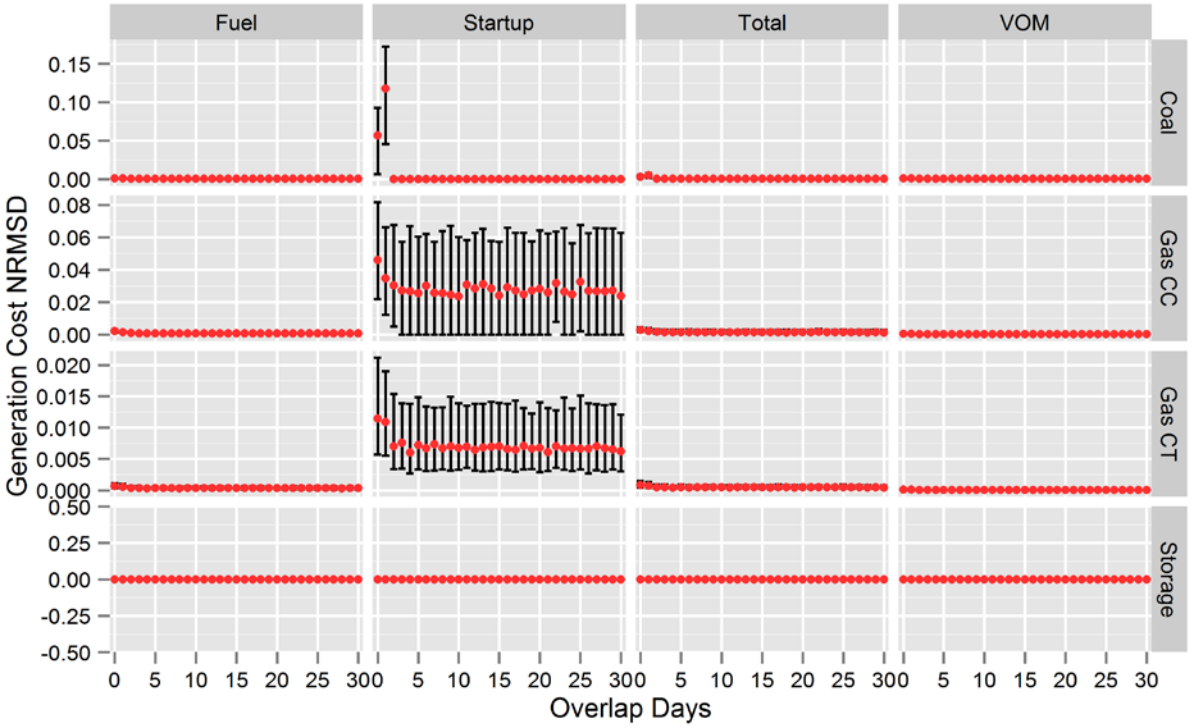

Figure 10. Generation cost NRMSD as a function of the number of days since the partitioned simulation has started where the lower and upper whiskers represent the 25th and 75th percentiles and the red dots represent the medians

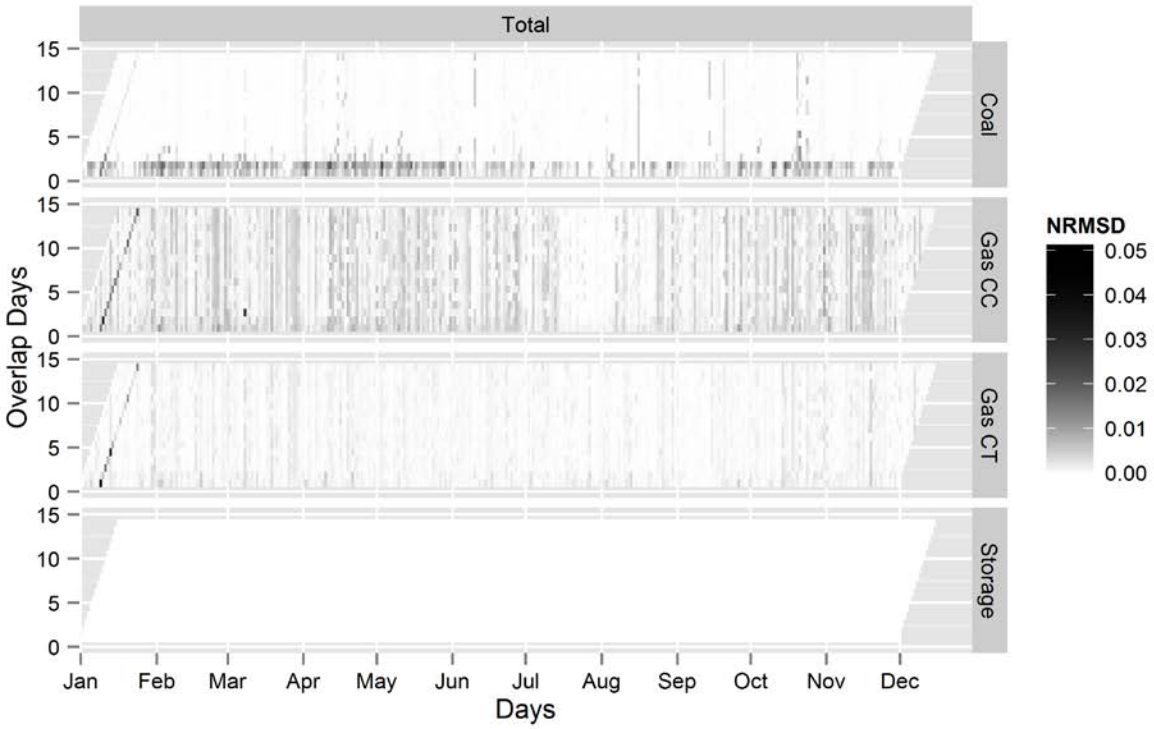

Figure 11. Mean total generation cost NRMSD heat map as a function of the number of days since the partitioned simulation has started and the date

6 Results of Partitioned Annual Simulations

The operating constraints of individual generators dictate that there exists some minimum duration where unit commitment decisions continue to affect system operation. Without knowing the time decay of unit commitment (UC) decision persistence, we partitioned a known annual simulation horizon into 52-week-long simulation horizons. All partition boundaries occur at midnight.

Simulations were executed using PLEXOS version 6.208 on NREL's Peregrine high performance computer running Linux. Peregrine has 1440 nodes with a total of 65,664 cores. Each PLEXOS simulation was executed on one of Peregrine's 144 Intel SandyBridge processor nodes that have 16 cores and 32GB memory. We executed these partitioned simulations simultaneously. In annual PCM simulations, the vast majority of the computation time is typically dedicated to solving the set of optimization problems. Therefore, the following analysis presents computation times as the time required to compute optimal solutions neglecting serial computation steps such as read/write activities.

Default initial conditions at time t_0 for any simulation horizon consider every unit committed and all time-dependent constraints (see Eq. 4) to be non-binding at time t_0. These initial condition assumptions have the potential to introduce operating point deviations for some unknown time period extending forward from the partition boundary. This effect is demonstrated in Figure 3. To mitigate the effects of the initial condition assumptions, a time overlap can be added to each partitioned horizon (see Figure 1). As a result, operations are simulated for longer durations and overlapping results are discarded. Each of the partitioned simulations was performed with overlap ranging from 0 (no overlap) to 5 days of overlap at the beginning of the horizon. The solution times for each partitioned simulation are shown in Figure 12. The marginal solution time penalty for an additional day of overlap is roughly 45 seconds on this test system. Assuming the availability of computer hardware is not a constraint on parallel execution of partitioned simulations, the time to solution is driven by the longest single partition solution (i.e., for the no-overlap partitioned simulation, the total optimization time is roughly 4 minutes).

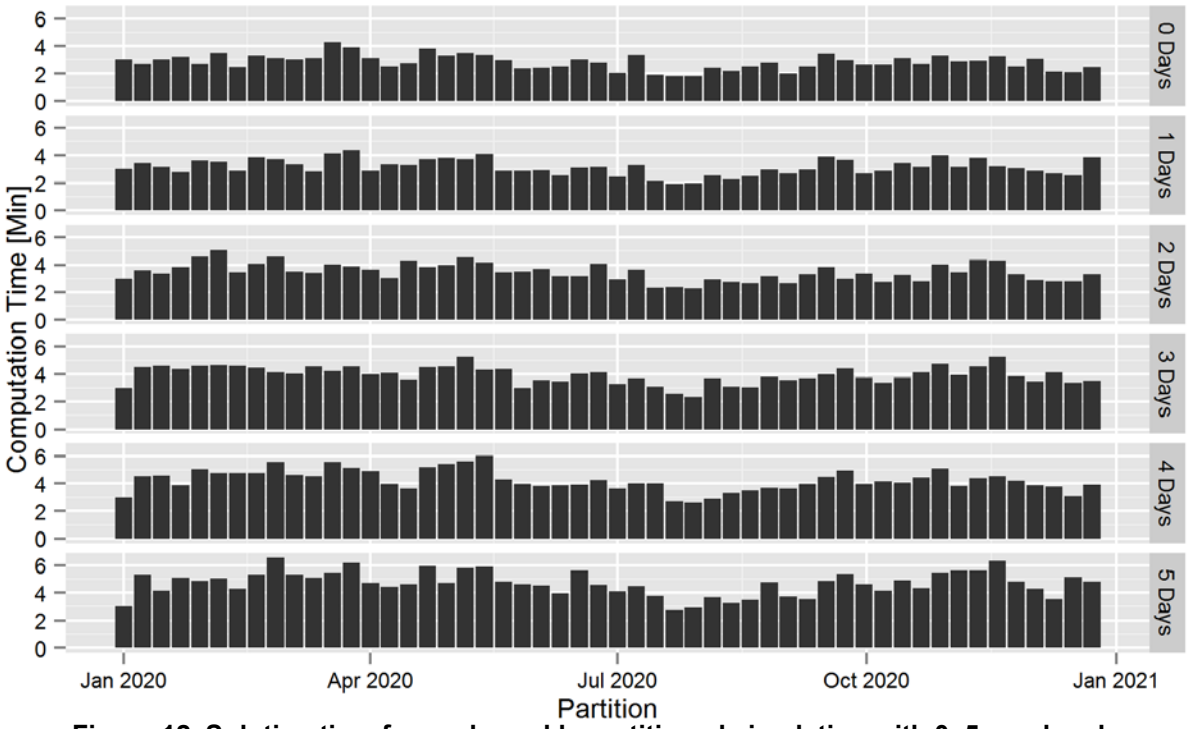

Figure 12. Solution time for each weekly partitioned simulation with 0–5 overlap days

Figure 13 shows the tradeoffs between speedup and NRMSD. The reserve NRMSD values are averages of the NRMSD values for the three reserve products (upward flexibility, upward regulation, and contingency) while the cost NRMSD values reflect total production cost NRMSD. The left panel of Figure 13 shows the speedup (annual simulation computation time divided by the maximum partitioned simulation computation time) achieved with 0–4 days of overlap. The results show that adding 2 days of overlap at the beginning of partitioned solutions largely minimizes differences in generation, costs, and reserve provisions, while achieving completion roughly 25 times faster than the un-partitioned simulation. The diminishing returns of reduced NRMSD relative to increasing NRMSD values suggests that two days of overlap provides a reasonable tradeoff between accuracy and computation time reduction.

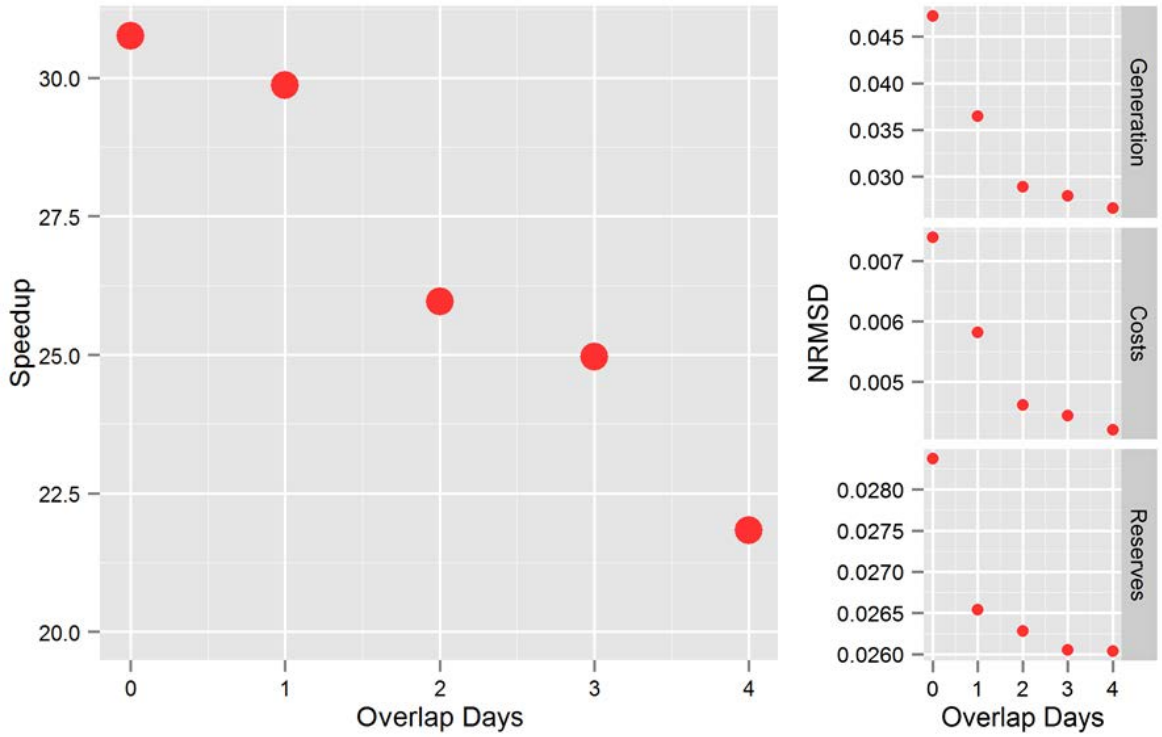

Figure 13. (Left) Speedup factors for annual day-ahead unit commitment simulations; (Right) Annual generation NRMSD values for simulations partitioned into 52 weekly partitions. Average values are reported for the three reserve product NRMSD values (contingency, upward regulation, and upward flexibility).

7 Conclusions

The coupling of unit commitment time steps through inter-temporal generation constraints makes PCM simulations difficult to partition and parallelize. The unit commitment decision persistence analysis presented in Section 4 demonstrates that unit commitment decisions have a finite duration where they continue to affect future system behavior. This suggests that partitioning PCM simulations can yield accurate results if the system is given enough time to forget after partition boundaries. In the process of reassembling the results of partitioned simulations, we denote this time during which unit commitment decisions made at the beginning of a simulation persist as the "overlap period," where system operation is simulated, but the results are not considered as part of the reassembled solution. The amount of overlap required to achieve accurate simulations, as described by the solution quality metrics in Section 3.1, is sensitive to the temporal location of the partition boundary. The partition boundary and overlap sensitivity analysis in Section 5 shows that differences between un-partitioned and partitioned simulation results are largely minimized when partitions are given at least 2 days of overlap. This result may differ depending on the generator mix specific to different systems, a point that will be addressed in ongoing research.

The tradeoffs between overlap duration and solution quality are highlighted in Section 6 where simulations partitioned into 52 weeks are reassembled and compared against and un-partitioned annual simulation. While the marginal cost of adding a day of overlap is relatively small—roughly 45 computation seconds—the marginal benefits, in terms of reduced NRMSD values, diminish beyond 2 days of overlap. The need for 2 overlap days constrains the speedup according to Amdahl's since the amount of redundant computation increases. Simulations partitioned into 52 nine-day (7 days + 2 days overlap) simulations solve roughly 25 times faster than un-partitioned simulations when implemented on a cluster capable of handling all partitioned simulations simultaneously.

While additional speedup is achievable through the creation of shorter partitions, the diversity in partition solution times (see Figure 12) becomes the limiting factor as individual solution times decrease. This diversity in solution time stems from the fact that the computational complexity of the partitioned MIPs are not simply functions of the simulation horizons but instead depend on the difficulty for the optimization algorithm to find a suitably optimal solution. Because the parallel execution of partitioned simulations is limited by the longest solution time for a single partition, an intelligent temporal partitioning algorithm that creates partitioned simulations with varying horizons but (more) equivalent solution times could further enhance simulation speedup. Ongoing research aims to develop more intelligent partitioning algorithms to achieve uniform solution times across partitions and further parallelize the PCM simulation.

The 2 days of overlap required to largely minimize NRMSD values for partitioned simulations is achieved when all simulations start with "hot start" initial conditions. That is, all unit commitment variables are set to one at the beginning of each simulation. A more accurate prediction of the unit commitment patterns would reduce the initial difference between partitioned and un-partitioned simulations, thus reducing the overlap duration required for quality solutions.

This paper focuses on methods for time domain partitioning for parallelizing PCM simulations. While significant speedup is available through these methods, additional opportunities exist in the spatial domain. Similar to how individual operators simulate and control a portion of a larger interconnect, spatial decomposition would make assumptions about the interaction between a focus region and the remainder of an interconnect to enable parallel simulation of operations within distinct zones. The accuracy of spatially decomposed simulations depends upon the ability of zonal boundaries to mitigate the inter-zonal effects of operation decisions. In combination with temporal decomposition, spatial decomposition could further reduce PCM computation time.

References

Amdahl, G. M. (1967). "Validity of the Single Processor Approach to Achieving Large Scale Computing Capabilities." *AFIPS Conference Proceedings,* Atlantic City, NJ.

Denholm, P.; Hummon, M. (2012). *Simulating the Value of Concentrating Solar Power with Thermal Energy Storage in a Production Cost Model.* TP-6A20-56731. Golden, CO: National Renewable Energy Laboratory.

Denholm, P.; Jorgenson, J.; Hummon, M.; Jenken, T.; Palchak, D.; Kirby, B.; Ma, O.; O'Malley, M. (2013). *Value of Energy Storage for Grid Applications.* TP-6A21-58465. Golden, CO: National Renwable Energy Laboratory.

EIA. (2012). *Electric Power Monthly with Data, December.* Washington, DC: Energy Information Administration.

EIA. (2013). *Electric Power Monthly with Data, June.* Washington, DC: Energy Information Administration.

Ela, E.; Milligan, M.; Kirby, B. (2011). *Operating Reserves and Variable Generation.* TP-5500-51978. Golden, CO: National Renewable Energy Laboratory.

Energy Exemplar. (August 2013). *PLEXOS for Power Systems*. Accessed January 17, 2014: http://energyexemplar.com/software/plexos-desktop-edition/.

GE. (2010). *Western Wind and Solar Integration Study.* NREL/SR-550-47434. Golden, CO: National Renewable Energy Laboratory.

General Electric Company. (2013). *MAPS*. Accessed January 17, 2014: http://geenergyconsulting.com/practice-area/software-products/maps.

Gilman, P.; Dobos, A. (2012). *System Advisor model, SAM 1011.12.2: General Description.* Golden, CO: National Renewable Energy Laboratory.

Grainger, J. J.; Stevenson, W. D. (1994). *Power System Analysis.* New York, NY: McGraw-Hill Inc.

Hummon, M.; Denholm, P.; Jorgenson, J.; Kirby, B.; Ma, O.; Palchak, D. (2013). *Fundamental Drivers of Operating Reserve Cost in Electric Power Systems.* TP-6A20-58491. Golden, CO: National Renewable Energy Laboratory.

Ibanez, E.; Brinkman, G.; Hummon, M.; Lew, D. (2012). A Solar Reserve Methodology for Renewable Energy Integration Studies Based on Sub-Hourly Varibility Analasys. *2nd Annual International Workshop on Integration of Solar Power Into Power Systems Conference, 12-13,* Lisban Portugul.

Meyers, C.; Streits, F.; Yao, Y.; Smith, S.; Lamont, A. (2011). *Using Supercomputers to Speed Execution of the CAISO/PLEXOS 33% RPS Study.* Livermore, CA: Lawrence Livermore National Laboratory.

TEPPC. (2011). *TEPPC 2012 Study Program 10-Year Regional Transmission Plan.* Salt Lake City , UT: Western Electricity Coordinating Council Transmission Expansion Planning Policy Committee.

Ventyx. (2013). *PROMOD IV*. Accessed January 17, 2014: http://www.ventyx.com/en/enterprise/business-operations/business-products/promod-iv.

Appendix: Alternative NRMSD Formulation

Section 3.1 defines the NRMSD metrics using Eqs. 9–11 such that differences between generation, reserve provisions, and costs at individual generators are abstracted to generator type. The purpose of this abstraction is to allow for substitutions within a generator type because generators of the same time often have similar if not equivalent cost curves. In the absence of binding transmission constraints, this redundancy can lead to multiple optimal solutions. Here, we perform similar analysis while preserving solution differences at the individual generator level (see Eq. 12). The following analysis provides some insight into the uncertainty introduced by partitioning versus the uncertainty inherent in optimizing a system with many redundant costs.

$$NRMSD\big(G(\mathrm{S})\big) = \sqrt{\frac{\sum_{t=1}^{n}\sum_{g\in \mathrm{S}}\left(\frac{g_t - g'_t}{L_t}\right)^2}{n}} \tag{12}$$

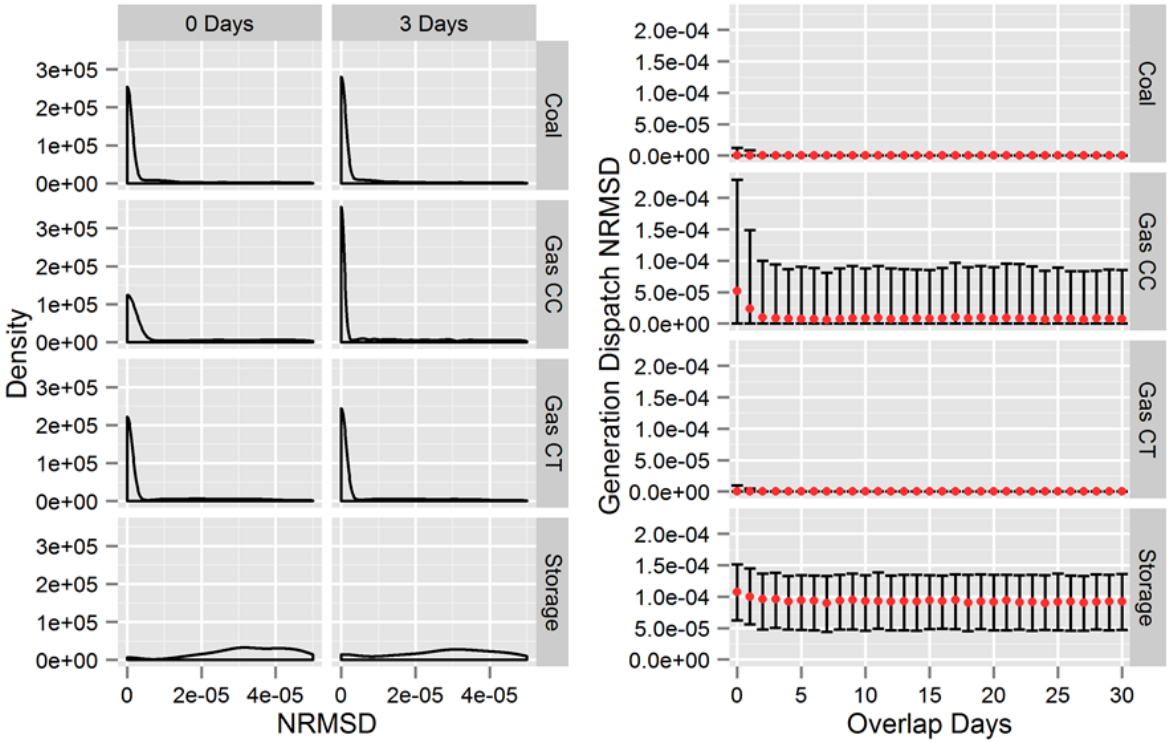

Figure 14. (Left) Density of generation dispatch NRMSD for 0 and 3 days of overlap; (Right) Generation dispatch NRMSD as a function of the number of days since the partitioned simulation has started where the lower and upper whisker edges represent the 25th and 75th percentiles and the red dots represent the medians

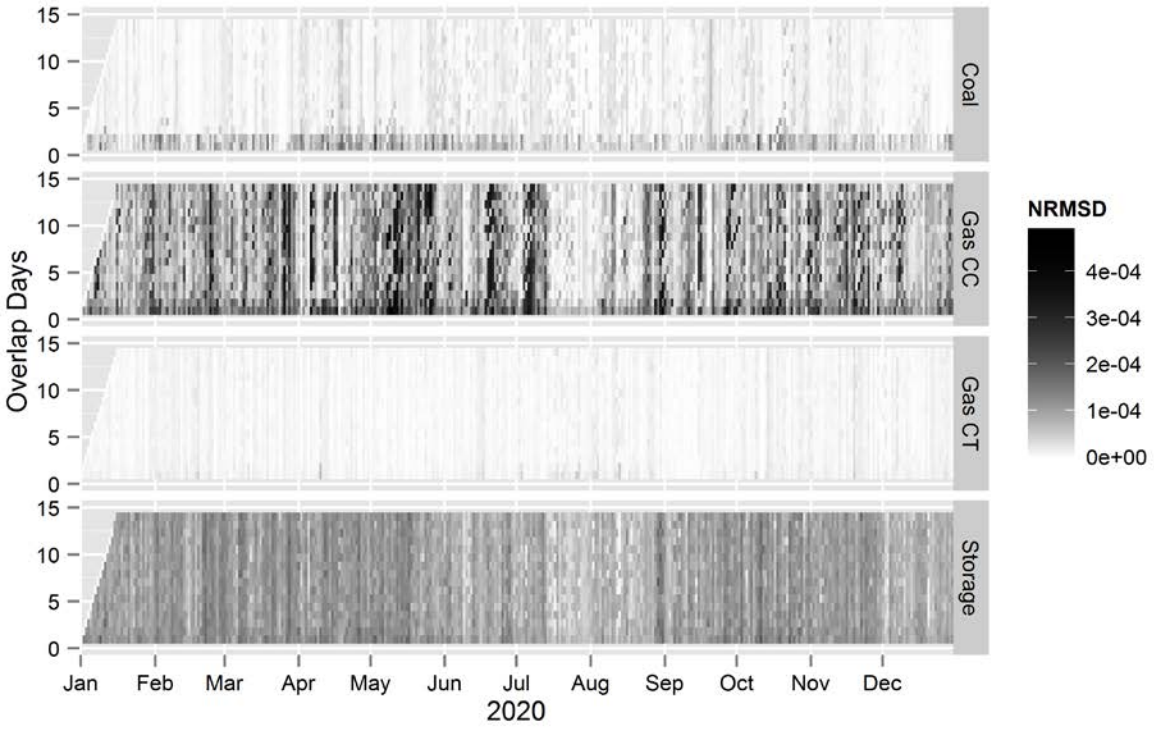

Figure 15. Mean generation NRMSD values as a function of the number of days since the partitioned simulation has started and the date