
# Development and analysis of a tool for speed up of EnergyPlus through parallelization


| | |
|---|---|
| Journal: | *Journal of Building Performance Simulation* |
| Manuscript ID: | TBPS-2012-0086.R2 |
| Manuscript Type: | Original Article |
| Keywords: | EnergyPlus, energy simulation, simulation run time, parallel simulation |

# Development and analysis of a tool for speed up of EnergyPlus through parallelization

This paper describes how the EPprarallel tool splits a single annual simulation into 12 simulations of one month each and runs them in parallel. The paper also describes the methodology to prepare input files, enable file sharing between nodes, collate results generated by the nodes, and ensure quality check on the simulations.

The EPparallel tool uses Message Passing Interface (MPI) library and runs on Linux. The tool has been tested on 16 commercial reference buildings over 16 U.S. weather files. The results of these 256 runs which include the runtimes, computing time overheads, speed gains, and accuracy of results are presented in this paper. The speed gain ranged from 2.9x to 7.8x and deviation (percentage of output values obtained in parallel simulation which were off by more than 1% as compared to output values obtained in annual simulation) ranged from 0% to 4%.

KEYWORDS: EnergyPlus parallelization, energy simulation, simulation runtime, simulation speed gain

Nomenclature:

| | |
|---|---|
| BothSync | PreSync & PostSync are collectively referred to as BothSync |
| DOE | Department of Energy |
| DPA_1 | Percentage of data points having deviation greater than 1% between output of annual and parallel simulation |
| DPA_10 | Percentage of data points having deviation greater than 10% between output of annual and parallel simulation |
| EPparallel | Name of the tool developed for EnergyPlus parallelization |
| ESO | EnergyPlus Standard Output |
| IDF | Input Data File of EnergyPlus |
| MPI | Message Passing Interface |
| NFS | Network File System. |
| Overhead | Time taken for data transfer and collection between client and nodes over the network |
| QCDPA_1 | Percentage of data points having deviation greater than 1% between overlapping days a segment with previous segment |
| QCDPA_10 | Percentage of data points having deviation greater than 10% between overlapping days a segment with previous segment |
| S0W0 | No shadow synchronization and zero warmup days |
| S0W7 | No shadow synchronization and 14 warmup days |
| S0W14 | No shadow synchronization and 14 warmup days |

| | |
|---|---|
| S1W0 | With BothSync and 0 days warmup |
| S1W7 | With BothSync and 7 days warmup |
| S1W14 | With BothSync and 14 days warmup |
| Segment Simulation | Annual simulation is segmented into smaller multiple simulations called Segmented Simulations. Each segmented simulation can run on a dedicated CPU in a computer cluster or on different cores of the same computer |
| Segments | Input for each segmented simulation is referred to a segment |
| Speed Gain | Ratio of the run time of annual simulation and the time taken by the slowest segmented simulation including the overhead |



## 1. Introduction

Building energy simulation is playing an increasingly significant role in the operation, diagnostic, and evaluation of building systems leading to an increased use of energy simulation software. EnergyPlus (Building Technologies Program: EnergyPlus Energy Simulation Software, 2012) is a whole building energy simulation program that engineers, architects, and researchers use to model energy and water use in buildings. Modelling the performance of a building with EnergyPlus enables building professionals to optimize the building design to use less energy and water. As per study by Crawley B. D. *et al.* (2008), EnergyPlus not only combines the best features of BLAST and DOE–2 programs, but also represents a significant step forward in terms of computational techniques and program structures. EnergyPlus includes a number of innovative simulation features, such as variable time steps, user-configurable modular systems that are integrated with a heat and mass balance based zone simulation—and input and output data structures tailored to facilitate third party module and interface development.

Although EnergyPlus has various innovative capabilities, its major limitation is that it runs slower than DOE-2. Tianzhen *et al.* (2008) compared computer run time of building simulation programs. As per the study at 15-minute time step, EnergyPlus v2.1.0 runs much slower than DOE-2.1E by a factor of 105 for a large office building to 196 for a hospital building. At 60-minute time step, EnergyPlus v2.1.0 runs even slower than DOE-2.1E by a factor of 25 for a large office building to 54 for a hospital building. For EnergyPlus 6.0, there is up to 25-40%

reduction in execution time in most simulations depending on specific features used (Building Technologies Program: EnergyPlus Energy Simulation Software: Archives, 2013). However this speed gain would not be sufficient to fix the gap.

According to EnergyPlus run time analysis report (Tianzhen *et al.*, 2009) prepared by the Environmental Energy Technology Division of Lawrence Berkeley National Laboratory (LBNL), the value of some simulation variables can have significant impact on EnergyPlus simulation time. These variables include the length of run period, Number_of_Timesteps_per_Hour for load calculations, heat balance solution algorithm, solar distribution and reflection calculation algorithm, system convergence limits, shadow calculation interval and the length of the warm up period.

Researchers have used parallel computing to reduce simulation time for a group of simulations. A tool developed by Zhang Yi (2009) runs multiple instances of EnergyPlus in parallel on multiple machines specifically for parametric analysis where multiple design alternatives have to be analysed simultaneously. Another GenOpt (LBNL, 2008) is an optimization program that is used to carry out the parametric analysis using multiple computers. These tools help in parametric analysis that involves running multiple simulations, one for each design alternative but they do not help in speeding up individual simulations. Garg Vishal *et al.* (2011) presented an approach that uses data parallelization paradigm to increase the speed of single simulation run. In this approach, annual simulation is segmented into smaller multiple simulations, each of which can run on a dedicated CPU in a computer cluster or on different cores on the same computer. This segmentation reduces the simulation run time and increases the speed of simulation. The speed gain in the simulation reportedly ranged from 3x to 6x.

In this paper, enhancement is made to the methodology described by Garg Vishal *et al.* (2011) with the introduction of post synchronization technique for shadow calculation. To simplify the use of EnergyPlus parallelization, a tool EPparallel is developed on Linux platform.

This tool takes an annual Input Data File (IDF) and splits it into 12 monthly IDFs. This tool also gathers & collates the output of each simulation and converts it into a single output file. To access quality of the simulation results from the tool, inbuilt quality assessment module is also developed. Parallelization tool has been tested over 16 different models and 16 different weather files and the output has been used to analyse the accuracy and speed gain due to parallelization. The speed gain ranged from 2.9x to 7.8x and Percentage of data points having deviation greater than 1% between output of annual and parallel simulation (DPA_1) ranged from 0% to 4%.

While designing the tool following were the main considerations:

a) The tool should work on recent EnergyPlus versions and flexible enough to work on future versions with minimum efforts.

b) There should not be any change in EnergyPlus source code.

c) The user need not do any modification in the IDF file created by the user when taking advantage of this tool for simulations.

d) Utilize spare capacity in small and medium offices where several computers are available with multiple cores and most of the time several of these cores are idle.

e) The tool should transparently sit between the user and EnergyPlus. The user need not know anything about the inner working of the tool, he just needs to provide the IDF and the tool returns the output.

## 2. Parallelization methodology and enhancement

This section provides insight into the existing methodology and the improvements that have been made to speed up the simulation process.

### *2.1. Existing methodology*

Existing methodology proposed by Garg Vishal *et al.* (2011) splits a single simulation with run period of one year into multiple simulations with smaller run periods. Since the runtime of a simulation is directly proportional to the run period of the simulation, the segmented simulations with shorter run period take much lesser time to run in parallel over a cluster of computers. Segmentation of simulation has resulted in minor deviations between the results obtained through segmented simulations and annual simulations.

To achieve higher accuracy, warm up days and extra days for shadow calculation synchronisation were added to each segment as shown in Figure 1.

EnergyPlus performs shadowing calculations (sun position, etc) over a period of days, in order to speed up the calculations. This value is very important for determining the amount of solar radiation entering a building and impact the amount of cooling or heating load needed for maintaining the set point temperature within the building. Though termed as "shadowing" calculations, this routine determines the position of the sun for a particular day in a weather file period simulation.

In IDF, the shadowing calculation frequency is mentioned in a field named *'Shadowing Frequency'*. During shadowing calculation, EnergyPlus takes an average of various parameters over a number of days specified in 'Shadowing Frequency', such as sun angle. The results of these calculations are used for finding hourly, daily, monthly and annual results. In Garg Vishal *et al.* (2011), the first day of the segmented simulation was matched with first day of one of the shadowing calculation periods (of the annual simulation) by adding extra days at the beginning of the segmented simulations. By doing this, the first shadowing calculation period comes in sync with one of the shadow calculation periods of the annual simulation, thus reducing the deviations.

### *2.2. Methodology enhancement*

It has been observed that to reduce deviations, the alignment of shadow is not only important at the beginning of each segment but also at the end of each segment. EnergyPlus shadow calculation frequency was set to 7 days in all simulations.

For the first segment, shadow calculation will be correct for the first 28 days as it is calculated in multiple of 7 days, but for the last 2 or 3 days shadow calculation will not match the calculation based on annual simulation. Similarly for the next segment, shadow calculation starts a few days before the start of the segment and ends before the last day of the segment. To synchronize shadow calculation days with annual simulation, these steps are followed:

- Some days from the end of the previous month are added before the current segment
- Some days from the start of the next month are added after the current segment

In enhanced methodology, adding extra days at the beginning of the segment for the purpose of shadow calculation period synchronization is referred to as PreSync. Adding extra days at the end of the segment for the same period is referred to as PostSync. PostSync & PreSync are collectively referred to as BothSync in this paper. This approach is shown in Figure 2.

## 3. Performance of segmentation alternatives

This section describes the analysis of different parallelization alternatives and how the best alternative is identified.

### *3.1. Segmentation alternatives*

EnergyPlus performs a warm up on the first day of simulation period by repeating it 15 times(user defined) to set the values of certain variables (However, a better approach would be if EnergyPlus treats the warmup period with actual climate data during that period - e.g. a 15 day warmup uses the weather data from the prior 15 days. This would have been better for parallelization because it would have reduced the effect of weather on the first day. We have overcome this limitation by adding warmup days from the actual weather file). Convergence of simultaneous heat balance/Heating Ventilation and Air-Conditioning (HVAC) solution is reached when the criteria for either the loads or temperature is met. In EnergyPlus, in case of an annual simulation, the simulation starts on 1st January, and the warm up calculations are carried out for the first day i.e. 1st January. However, for individual segments, the warm up takes place on the first day of the segment i.e., the first day of each month. If the first day of the month happens to be a holiday, then the warming up of the model for that segment is done according to non-operational conditions. This results in deviated output for the next few working days of that segment and affects the total monthly consumption. To achieve the best alternative, simulation results are compared for three sets of warm up days with shadow synchronization and three set of warm up days without shadow i.e. S0W0, S0W7, S0W14, S1W1, S1W7 and S1W14.

Here are the different alternatives used for segmentation:

(1) **S0W0** (No Sync Zero warmup)

In this case, each segment is made up of exactly one month as shown in Figure 3. Neither shadow calculation synchronized nor warm up days are added for the simulation. Example: Second segment is of 28 days of February and Eleventh segment is of 30 days of November and so on.

(2) **S0W7**(No Sync 7 days warmup):

In this case, each segment (except the first segment) is made up of all the days of a particular month in addition to the last seven days of the previous month as shown in Figure 4. Example: Second segment starts on 25$^{th}$ of January and finishes on 28$^{th}$ of Feb. Eleventh Segment starts on 25$^{th}$ of Oct and finishes on 30$^{th}$ of November.

(3) **S0W14**(No Sync 14 days warmup):

In this case, each segment (except the first segment) is made up of all the days of a particular month in addition to the last fourteen days of the previous month as shown in Figure 5. Example: Second segment starts on 18$^{th}$ of January and finishes on 28$^{th}$of February. Eleventh Segment Start on 18$^{th}$ of October and finishes on 30$^{th}$ of November.

(4) **S1W0**(With *BothSync* 0 days warmup):

In this case, each segment has some extra days from the end of previous month for pre-sync and a few days from the start of next month for post-sync as shown in Figure 6. Example: Second segment starts on 29th January and ends on 4th March. Tenth segment starts on 24th September and ends on 4th November.

(5) **S1W7**(With *BothSync* 7 days warmup):

In this case, each segment has extra 7 days from the end of previous month. Each segment also has a few days from the end of previous month for pre-sync and a few days (if any) from the start of next month for post-sync as shown in Figure 7. Example: Second segment starts on 22nd January and ends on 4th March. Tenth segment starts on 24th September and ends on 4th November.

(6) **S1W14**(With *BothSync* 14 days warmup):

In this case, each segment has extra 14 days from the end of previous month. Each segment also has a few days from the end of previous month for pre-sync and a few days from the start of next month for post-sync as shown in Figure 8. Example: Second segment starts on 15th January and ends on 4th March. Tenth segment starts on 17th September and ends on 4th November.

### *3.2. Simulation models and climate*

This section describes the prototype models used for simulation study and the different climate zone weather data used for the analysis.

### 3.2.1 *Building prototypes*

Department of Energy's (DOE) Building Technologies Program, working DOE's Pacific Northwest National Laboratory, Lawrence Berkeley National Laboratory, and National Renewable Energy Laboratory have developed models for 16 commercial building types in 16 locations representing all U.S. climate zones. These 16 building types represent approximately 70% of commercial buildings in the United States (Deru Michael *et al.*, 2011).

The reference models used in this paper were converted to version 6.0 of EnergyPlus with the available tool on EnergyPlus website (Building Technologies Program: EnergyPlus Energy Simulation Software). These models can be used as standard for the analysis of the results. To analyse the performance of the proposed alternative over different models and climate, reference buildings packaged with the EnergyPlus installation were used for the simulations. The commercial building reference model code and their characteristics are listed in Table 1.

### 3.2.2 *Climatic Zone*

All weather data files have been used in EnergyPlus weather (EPW) format. These weather data files cover all climate zones as defined by ASHRAE 90.1-2007. The weather file and its code are listed in the Table 2.

For each model, one city is selected and simulated for all 6 segmentation alternatives, as shown in Table 3.

## 3.3 ***Deviation and speed gain***

When the annual simulation was segmented into 12 monthly simulations, it was observed that the results from parallel simulations do not match with the results from annual simulation. Therefore, both the result sets were compared to find the percentage of data points having deviation greater than 1% between output of annual and parallel simulation (DPA_1) between the two sets as shown in Figure 9.

The maximum deviation was observed for case S0W0 in which each segment has exactly one month without any sync and warm-up days. This deviation occurred due to inadequate *warm-up* and *no shadow* synchronization.

The minimum deviation was observed with case S1W14 as there was adequate *warm-up* and *shadow synchronization*. On the basis of minimum deviation, case S1W14 seems the best alternative. Note that the speed gain of the simulation is also an important factor for choosing the best alternative. The time taken by annual simulation and slowest segment for all alternatives is shown in Table 4.

Figure 10 shows the speed gain which is the ratio of the run time of annual simulation and the time taken by the slowest segmented simulation including the overhead achieved for all simulation alternatives. It is observed that the minimum speed gain is achieved with case S1W14 and the maximum with case S0W0. With increase in the warm up or the number of sync days, the number of days for simulation in each segment increases, thus reducing the speedup achieved.

### 3.4 *Selection of best alternative*

For selection of the best alternative, a trade-off is required between accuracy of results and speed gain. Figure 11 is plotted between segmentation alternative and number of data points with deviation greater than 1%. In this figure, numbers of data points are large as all data points from all simulations for a particular segmentation alternative are combined. It is obvious from Figure 11 that maximum deviation is in case S0W0 and minimum in case S1W14. Also, the difference between S1W7 case and S1W14 case is not significant. Figure 12 shows the average of speed gain for all simulation runs for all segmentation alternatives.

The conclusion drawn from figure 11 and 12 is that although the accuracy of the results is almost the same as case S1W7 and S1W14, there is a noticeable difference in the speedup of the two cases. After the results of the analysis, case S1W7 is the best choice among all the 6 alternatives.

## 4. Development of parallelization tool: EPparallel

This section explains the architecture and internals of EPparallel. As shown in Figure 13, the tool requires a Linux cluster of 13 computers out of which 12 computers are used to run the simulations in parallel and one system is used for load balancing and sequential tasks, referred to as the master computer in this paper. It also requires the Message Passing Interface (MPI) library used for communication between the computers.

File sharing happens over a shared file system, - Network File System (NFS). MPI (MPI Documents, 2011) is a standard for message-passing for distributed-memory systems (like computer clusters) and several implementations of this standard are available. The MPICH, High-performance and Portable MPI (2011), a popular open source implementation of MPI is used for the development of this tool. MPI is used to communicate between the computers through the network. Tasks, such as triggering simulations and collecting the output have been done using MPI.

MPI is the backbone of EPparallel. It performs the following tasks:

- Triggers the preparation of 12 IDFs, one for each segment
- Uses NFS to transfer files to the computers in the cluster
- Triggers parallel simulations on each computer
- Collects output of the simulation and copies to the root computer through NFS
- Triggers the collation of output
- Triggers the quality check
- Triggers the report generation

The Perl Programming Language (2011) is needed on the master computer to run the collation and quality-check scripts. Perl is a general purpose programming language which has powerful text processing capabilities using regular expressions. Regular expression is a specific pattern used to match a specific string of text. Perl makes it easier to search specific strings of text even in large text data sets.

The Perl programming language is best suited for processing large volume of text based data. Common text processing tasks, such as collating the output text files and quality analysis is handled efficient in Perl.

To use the tool, at first the user needs to generate an IDF file with run period as the full year. The IDF file is then fed to the tool. The tool makes changes in the file on the master computer and splits it into 12 different IDF files with smaller run periods. Each file is then sent to a computer in a cluster and all simulations are run in parallel over these computers. Once all the simulations are complete, the output of the simulations are brought in to master computer and collated to form a single output file. Quality analysis is done over the output to find out how well the parallel run performed.

### 4.1 *User interface*

The tool has a simple command line interface which takes the IDF file name and the weather file name as command line arguments.

epparallel <*IDF filename*> <*weather filename*>

For example, to specify the IDF file as RefBldgMidriseApartment and the weather data file as SA_FL_Tampa.Intl.AP.722110_TMY3 for the simulation, type:

*epparallelRefBldgMidriseApartment SA_FL_Tampa.Intl.AP.722110_TMY3*

This command runs on the master computer and initiates the IDF pre-processing and splitting.

### 4.2 *IDF processing and splitting*

The tool first creates an MPI, Communication World. Next, it creates 12 IDF files in a sequential order.

The tool makes the following changes in each of the IDF replicas:

1) The Day_of_Week_for_Start_Day element in the *RunPeriod* class is changed to UseWeatherFile.
2) The *RunPeriod* section in the IDF is modified to start the simulation few days before the start of the month. For example, 2nd segment is started a few days before 1st of February. The day on which the simulation has to start is decided as per the number of warmup and shadow synchronization days added before the segment. The simulation end date is the end date of the last shadowing period of the segment.
3) The output variables with '*RunPeriod*' frequency are converted to 'Monthly' frequency.
4) In a separate file, names of the variables requested for both '*RunPeriod*' and '*Monthly*' frequency are stored. In another file, names of the variables requested at '*RunPeriod*' frequency but not at '*Monthly*' frequency are stored. Next, the 12 IDFs are distributed to all the twelve computers over the NFS.

### 4.3 *Simulation runs*

Now each computer runs an instance of EnergyPlus with its segmented IDF file and the weather file as input.

*runenergyplus <IDF filename> <EPW weather file>*

### 4.4 *Collation*

After the simulation is over and output files are generated, each system copies the ESO files from the output directories to the master computer. Then the collation script collates the ESO files which create a single file. This file is viewable in ESO viewer and matches with the ESO that would have been generated if the simulation had run for the full year. The output of the segmented simulations consists of *RunPeriod* values in terms of hours, days, and months.

Since each segmented simulation runs for a few extra days (for warm up and sync), the data for these extra days needs to be discarded. The remaining data is used when collating ESO. Collation generates a single ESO file from all the twelve ESOs of the segmented simulation.

Each of the ESOs of segmented simulations consists of an identical data dictionary which is copied to the collated ESO with some minor changes. Records for hourly, daily and monthly variables are directly copied from the segmented simulation outputs.

*(a) Dealing with RunPeriod variables*

The records for the *RunPeriod* canot be directly copied from the ESOs of segmented simulations. Suppose there is a run period variable X, if the reporting frequency of this variable is not changed to Monthly, that variable in segmented simulations will report output for the whole segment (because that is the segment's run period) and these values can not be taken to compute output value for variable X for the whole year. So, reporting frequency of variable X is changed to monthly before the simulation. After the simulation monthly values of the calendar month from each of the segmented simulations is taken and value for X for whole year is calculated. Calendar month is the month which is wholly present in the segment. For example: calendar month in 2nd segment is February even though this segment may contain some days from January and March.

When the *RunPeriod* variables are converted to monthly variables, one of the two cases may arise. For a particular variable the same variable may already be requested for monthly reporting or the same variable may not be requested for monthly reporting. For both the cases the ESO has a record for that particular variable only once. The difference is that in the second case, the tool converts the monthly records to *RunPeriod* records. But in the first case, the monthly records are kept as it is (since the variable was also requested at monthly frequency) and *RunPeriod* record is prepared from the Monthly values of that variable and added to the collated ESO.

*(b) Preparation of a RunPeriod record for variables requested only at RunPeriod frequency*

Each record in the ESO consists of the report code and the variable value. If the variable is requested at only run period frequency and changed to monthly frequency before the simulation, then the results contain that variable's records only at monthly frequency. To convert that variable to *RunPeriod* variable, all its records are summed or averaged depending on the type of variable to create a *RunPeriod* record. Also, the data dictionary record that declares the *RunPeriod* variable is changed to monthly record.

Here are the steps involved in converting a monthly record back to *RunPeriod* record:

1. Convert the monthly record in data dictionary with the *RunPeriod* record keeping the same report code as that of the monthly variable.
2. Extract all the monthly records of the variable from the ESO files and perform computation on the monthly reported values to prepare the value for whole *RunPeriod*. The maximum and minimum *RunPeriod* values are also generated by evaluating the maximum and minimum monthly values.
3. Add a *RunPeriod* record in the collated file using the report code from Step 1 and the value generated in Step 2.

*(c) Preparation of a RunPeriod record for variables requested at both RunPeriod and Monthly frequencies*

A report code needs to be prepared which is not already used by any of the other variables. This report code is used in declaration of the *RunPeriod* variable in the data dictionary as well as in the *RunPeriod* records in the data section.

## 5. Quality analysis for parallelization tool

The purpose of segmented simulation is to reduce the simulation run time without affecting the accuracy of the simulation results. To know the quality of the results of the simulation output generated by a segment, each segment can be compared with the output values of that part of the segmented simulations which will go in collated ESO with the corresponding values in the annual (non-parallel) simulation. However, as the annual simulation does not exist when using EPparallel tool, it cannot be done. A method has been developed to check the quality of the results which doesn't require annual simulation. In this method overlap days of segment is compared to the end of the previous segment. Figure 14 shows the overlapping days between 2nd and 3rd segment. This overlap occurs because each segment starts few days before the start of the corresponding month and ends few days after the end of the same month.

If the segmented simulation is warmed up adequately prior to the simulation of the calendar month, the deviation for the calendar month should be very less. To check if a segment X is warmed up before the simulation of its calendar month starts, the overlapping output of segment X and its previous segment is compared with each other. For example, if user wants to check if the third segmented simulation has warmed up properly before it starts the simulations for that segment, user can compare the values of the output variables for the first few days of March with the values of same output variables for the last few days given by the February segment simulation.

The tool compares the values of hourly variables of adjacent segments and calculates the number of values that deviate more than 1% (QCDPA_1) and more than 10% (QCDPA_10). Comparison of QCDPA_1 with DPA_1 and QCDPA_10 with DPA_10 is shown in Figure 15 and 16 respectively. As QCDPA_1 and QCDPA_10 are less than DPA_1 and DPA_10 respectively, this indicates the high quality of simulation results. User can access the quality of the output through this data. An output is considered reliable if it has minimal deviation.

Normally, the user may like to run many simulations with small modifications in the input. Currently, a new interface is under development, which will allow the user to run an annual simulation as a benchmark that will give the user an idea about the accuracy of parallelization. While running this initial benchmark, the user might not experience any benefit for parallelization. However, the user will get an idea about the accuracy of the parallelization. After that the user may run multiple simulations with small modifications in the input without running an annual simulation.

When the user is satisfied with the IDF output, the user can prepare and run the parallel and annual simulations again. Running both the simulations will let the user know if the quality indicator is correct. If yes, the user can take the results from the annual simulation since they are without any deviation.

## 6. **Detailed simulation analysis with parallelization tool**

The Parallelization tool was tested on all 16 models across all 16 US climatic zones listed in section 3.2.

Here are the results:

- Summary of the simulation results for the Outpatient Health Care model (M7) for all climatic zones is listed in Table 5.
- Annual simulation run time, slowest segment run time for segmented simulation, overhead, speed gain and deviations are listed.
- Hourly and monthly DPA_1 and DPA_10 are listed.

From all 256 simulations, it has been noted that the run time of the simulation has an effect on the speed gain. Figure 17 depicts the relation between run time and speed gain. It has been observed that speed gain increases with increase in runtime. This effect is due to simulation overhead, which is the extra time taken in transferring files over the network and collating the output.

Simulation overhead is the time taken to copy the IDF from the master computer to other computers of the cluster and importing the output (ESO) from other computes to the master computer.

Since collation is heavily based on text processing, collation time is dependent on the number of variables requested for output. This is because large files require large volume of text data to be processed.

Simulation model which takes longer simulation time, requires less time for the overhead thus increasing speed gain of the simulation. The maximum and minimum speed gain across all models is shown in Table 6. Maximum speed gain of 7.8x was achieved in Outpatient health care model for Miami, Florida location and minimum speed gain of 2.9x was achieved in Supermarket for Helena, Minnesota location.

## 7. **Conclusions**

Existing methodology of EnergyPlus parallelization has been improved to reduce the deviation in simulation by adding post shadow synchronization technique which has been tested on different available benchmark simulation models to select the best alternative.

To make EnergyPlus parallelization more user friendly, a tool has been developed which takes an IDF file as input and splits it into 12 IDFs, one to run on each computer of a cluster of 12. This tool also collates the output of each simulation and combines it into a single output file.

To assess the quality of the simulation results, an in-built quality assessment module is provided within the tool. This tool has been tested over 16 different models across 16 different weather files and output data was used to analyse simulation runtime and speed gain. It has been found that for higher runtime, speed gain is also higher. The speed gain achieved for simulation ranged from 2.9x to 7.8x compared to an annual single run. The maximum gain of 7.8x was achieved in the Outpatient health care model for Miami, Florida and minimum speed gain of 2.9x was achieved in the Supermarket for Helena, Minnesota.

## **References**

ASHRAE 90.1-2007. *Energy Standard for Buildings Except Low-rise Residential Buildings*. Atlanta,GA: American Society of Heating, Refrigeration and Air-Conditioning Engineers.

Barroso, Andre Luiz, Dean Jeffrey, and Holzle Urs. 2003. "Web Search for a Planet: The Google Cluster Architecture." *IEEE Computer Society*: 22–28.

"Building Technologies Program: EnergyPlus Energy Simulation Software.", Accessed 10 August 2012, http://apps1.eere.energy.gov/buildings/energyplus/.

"Building Technologies Program: EnergyPlus Energy Simulation Software: Archives.", Accessed December 2012 http://apps1.eere.energy.gov/buildings/energyplus/energyplus_archives.cfm#v6

"Commercial Reference Building Models for New Construction.", Accessed October 11, 2011, http://www1.eere.energy.gov/buildings/commercial/ref_buildings.html

Crawley, Drury B, Jon W Hand, Michaël Kummert, and Brent T Griffith. 2008. "Contrasting the Capabilities of Building Energy Performance Simulation Programs." *Building and Environment* 43 (4): 661–673. doi:10.1016/j.buildenv.2006.10.027.

Deru, Michael, Kristin Field, Daniel Studer, Kyle Benne, Brent Griffith, Paul Torcellini, Bing Liu, Mark Halverson, Dave Winiarski, Michael Rosenberg, MehryYazdanian, Joe Huang, and Drury Crawley. 2011. *U. S. Department of Energy Commercial Reference Building Models of the National Building Stock*. Technical Report NREL/TP-5500-46861, Golden, Colorado: NREL. www.nrel.gov/docs/fy11osti/46861.pdf.

"EnergyPlus Energy Simulation Software: Archives.", Accessed January 06, 2013, http://apps1.eere.energy.gov/buildings/energyplus/energyplus_archives.cfm#v6.

Garg, Vishal, Kshitij Chandrasen, Jyotirmay Mathur, Surekha Tetali, and Akshey Jawa. 2011. "Development and Performance Evaluation of a Methodology, Based on Distributed Computing, for Speeding EnergyPlus Simulation." *Journal of Building Performance Simulation* 4 (3): 257–270. doi:10.1080/19401493.2010.531142.

Hong, Tianzhen, Fred Buhl, and Philip Haves. 2009. *EnergyPlus Run Time Analysis*. http://escholarship.org/uc/item/36h4m5z0.

Hong, Tianzhen, Buhl Fred, Haves Philip, Stephen Selkowitz, and Wetter Michael. 2008. *Comparing Computer Run Time of Building Simulation Programs*. http://escholarship.org/uc/item/6504q6d0.

LBNL, 2008. "GenOpt-generic Optimization Program.", Accessed July 19, 2008, http://simulationresearch.lbl.gov/GO/index.html.

"MPI Documents.", Accessed October 11, 2011, http://www.mpi-forum.org/docs/docs.html.

"MPICH | High-performance and Portable MPI.", Accessed October 19, 2011, http://www.mpich.org/.

"The Perl Programming Language.", Accessed July 19, 2008, http://www.perl.org/.

Torcellini, Paul, Michael Deru , Brent Griffith, Kyle Benne, Mark Halverson, David Winiarski, and Drury Crawley. 2008. "DOE Commercial Building Benchmark Models." In *ACEEE Summer Study on Energy Efficiency in Buildings*, 4–305—4–316. Pacific Grove, California.

Zhang, Yi. 2009. "' PARALLEL ' EnergyPlus and the Development of a Parametric Analysis Tool." In *Eleventh International IBPSA Conference*, 1382–1388. Glasgow, Scotland.

**Changes are as follows:**

Added text highlighted in grey

Table 1 Characteristics of the 16 building models used for the simulation

| S.No. | Building Type Name | Floor Area ($ft^2$) | Number of Floors |
|---|---|---|---|
| M1 | Full Service Restaurant | 5,500 | 1 |
| M2 | Hospital | 2,41,351 | 5 |
| M3 | Large Hotel | 1,22,120 | 6 |
| M4 | Large Office | 4,98,588 | 12 |
| M5 | Medium Office | 53,628 | 3 |
| M6 | Midrise Apartment | 33,740 | 4 |
| M7 | Outpatient Health Care | 40,946 | 3 |
| M8 | Primary School | 73,960 | 1 |
| M9 | Quick Service Restaurant | 2,500 | 1 |
| M10 | Secondary School | 2,10,887 | 2 |
| M11 | Small Hotel | 43,200 | 4 |
| M12 | Small Office | 5,500 | 1 |
| M13 | Stand-alone Retail | 24,962 | 1 |
| M14 | Strip Mall | 22,500 | 1 |
| M15 | Supermarket | 45,000 | 1 |
| M16 | Warehouse | 52,045 | 1 |

Table 2 Climate data used for running the simulation models.

| S.No. | City | Climate Zone | Climate type |
|---|---|---|---|
| W1 | Miami, Florida | 1A | Very Hot -Humid |
| W2 | Houston, Texas | 2A | Hot - Humid |
| W3 | Phoenix, Arizona | 2B | Hot - Dry |
| W4 | Atlanta, Georgia | 3A | Warm -Humid |
| W5 | Los Angeles, California | 3B | Warm - Dry |
| W6 | Las Vegas, Nevada | 3B | Warm - Dry |
| W7 | San Francisco, California | 3C | Warm-Marine |
| W8 | Baltimore, Maryland | 4A | Mixed - Humid |
| W9 | Albuquerque, New Mexico | 4B | Mixed - Dry |
| W10 | Seattle, Washington | 4C | Mixed- Marine |
| W11 | Chicago, Illinois | 5A | Cold - Humid |
| W12 | Boulder, Colorado | 5B | Cold - Dry |
| W13 | Minneapolis, Minnesota | 6A | Cold - Humid |
| W14 | Helena, Montana | 6B | Cold - Dry |
| W15 | Duluth, Minnesota | 7 | Very Cold |
| W16 | Fairbanks, Alaska | 8 | Subarctic |

Table 3 Simulation model and corresponding climatic zone

| S.No. | Building Type Name | Weather data |
|---|---|---|
| M1 | Full Service Restaurant | Miami, Florida |
| M2 | Hospital | Houston, Texas |
| M3 | Large Hotel | Phoenix, Arizona |
| M4 | Large Office | Atlanta, Georgia |
| M5 | Medium Office | Los Angeles, California |
| M6 | Midrise Apartment | Las Vegas, Nevada |
| M7 | Outpatient Health Care | San Francisco, California |
| M8 | Primary School | Baltimore, Maryland |
| M9 | Quick Service Restaurant | Albuquerque, New Mexico |
| M10 | Secondary School | Seattle, Washington |
| M11 | Small Hotel | Chicago, Illinois |
| M12 | Small Office | Boulder, Colorado |
| M13 | Stand-alone Retail | Minneapolis, Minnesota |
| M14 | Strip Mall | Helena, Montana |
| M15 | Supermarket | Duluth, Minnesota |
| M16 | Warehouse | Fairbanks, Alaska |

Table 4 Simulation run time (in seconds) for annual simulation and slowest with all segmentation alternatives

| Simulation model | Location | Simulation alternatives | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | S0W0 | | S0W7 | | S0W14 | | S1W14 | | S1W7 | | S1W14 | |
| | | Annual simulation run time (s) | Slowest segment run time (s) | Annual simulation run time (s) | Slowest segment run time (s) | Annual simulation run time (s) | Slowest segment run time (s) | Annual simulation run time (s) | Slowest segment run time (s) | Annual simulation run time (s) | Slowest segment run time (s) | Annual simulation run time (s) | Slowest segment run time (s) |
| M1 | W1 | 72 | 9 | 72 | 12 | 72 | 13 | 72 | 12 | 72 | 13 | 72 | 14 |
| M2 | W2 | 457 | 77 | 457 | 88 | 457 | 99 | 457 | 83 | 457 | 93 | 457 | 104 |
| M3 | W3 | 668 | 104 | 669 | 117 | 669 | 140 | 669 | 114 | 669 | 131 | 669 | 148 |
| M4 | W4 | 254 | 42 | 255 | 43 | 255 | 49 | 255 | 44 | 255 | 49 | 255 | 54 |
| M5 | W5 | 163 | 27 | 163 | 31 | 163 | 35 | 163 | 31 | 163 | 34 | 163 | 38 |
| M6 | W6 | 899 | 158 | 899 | 187 | 899 | 203 | 899 | 179 | 899 | 212 | 899 | 241 |
| M7 | W7 | 1,697 | 262 | 1,698 | 262 | 1,698 | 242 | 1,698 | 259 | 1,698 | 259 | 1,698 | 258 |
| M8 | W8 | 328 | 58 | 329 | 65 | 329 | 70 | 329 | 59 | 329 | 68 | 329 | 75 |
| M9 | W9 | 103 | 19 | 104 | 20 | 104 | 24 | 104 | 20 | 104 | 22 | 104 | 26 |
| M10 | W10 | 856 | 149 | 856 | 169 | 856 | 189 | 856 | 166 | 856 | 191 | 856 | 210 |
| M11 | W11 | 1,032 | 151 | 1,032 | 177 | 1,032 | 200 | 1,032 | 183 | 1,032 | 210 | 1,032 | 229 |
| M12 | W12 | 150 | 23 | 151 | 25 | 151 | 30 | 151 | 26 | 151 | 29 | 151 | 32 |
| M13 | W13 | 61 | 10 | 62 | 11 | 62 | 13 | 62 | 11 | 62 | 13 | 62 | 13 |
| M14 | W14 | 158 | 24 | 159 | 27 | 159 | 32 | 159 | 28 | 159 | 32 | 159 | 34 |
| M15 | W15 | 74 | 12 | 75 | 13 | 75 | 15 | 75 | 12 | 75 | 15 | 75 | 16 |
| M16 | W16 | 58 | 10 | 59 | 11 | 59 | 15 | 59 | 11 | 59 | 13 | 59 | 16 |

Table 5 Runtime, overhead, speedup and deviation for M7 model with all climate zones

| Model | Location | Annual simulation run time (s) | Slowest segment run time (s) | Overhead(s) | Speed gain | Deviation | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Hourly 1% | Monthly 1% | Hourly 10% | Monthly 10% |
| M7 | W1 | 1729 | 209 | 10.4 | 7.8 | 2.2 | 0 | 0.8 | 0 |
| M7 | W2 | 1691 | 251 | 10.4 | 6.4 | 2.4 | 0 | 0.9 | 0 |
| M7 | W3 | 1748 | 258 | 10.6 | 6.5 | 2.6 | 0 | 1.1 | 0 |
| M7 | W4 | 1781 | 257 | 10.6 | 6.6 | 2.2 | 0 | 0.9 | 0 |
| M7 | W5 | 1754 | 260 | 10.4 | 6.4 | 3.0 | 0 | 1.2 | 0 |
| M7 | W6 | 1911 | 253 | 10.5 | 7.2 | 1.8 | 0 | 0.8 | 0 |
| M7 | W7 | 1683 | 257 | 10.4 | 6.2 | 2.1 | 0 | 0.7 | 0 |
| M7 | W8 | 1716 | 229 | 10.8 | 7.1 | 2.2 | 0 | 0.9 | 0 |
| M7 | W9 | 1877 | 246 | 10.4 | 7.2 | 3.2 | 0 | 1.3 | 0 |
| M7 | W10 | 1710 | 225 | 10.4 | 7.2 | 2.5 | 0 | 0.9 | 0 |
| M7 | W11 | 1876 | 259 | 10.8 | 6.9 | 2.6 | 0 | 0.9 | 0 |
| M7 | W12 | 1759 | 237 | 10.8 | 7.0 | 3.4 | 0 | 1.2 | 0 |
| M7 | W13 | 1775 | 244 | 10.5 | 6.9 | 2.4 | 0 | 0.8 | 0 |
| M7 | W14 | 1758 | 242 | 10.4 | 6.9 | 2.6 | 0 | 1.0 | 0 |
| M7 | W15 | 1802 | 243 | 10.4 | 7.1 | 2.3 | 0 | 0.9 | 0 |
| M7 | W16 | 1748 | 237 | 10.5 | 7.0 | 3.0 | 0 | 1.0 | 0 |

Table 6 Model and climatic zone with maximum and minimum speed gain

| Model | Location | Speed gain |
|---|---|---|
| Outpatient Health Care (M7) | Miami, Florida (W1) | 7.8 |
| Supermarket (M15) | Helena, Minnesota (W14) | 2.9 |

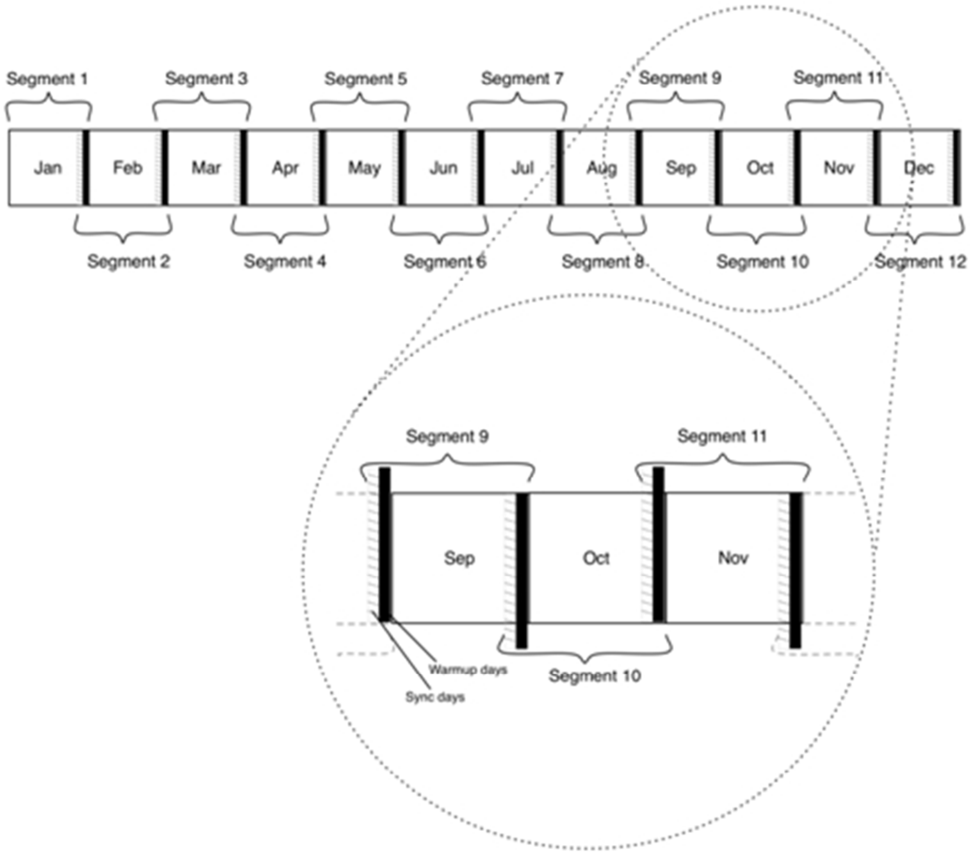


Figure 1 Existing Methodology for Parallelization
176x155mm (72 x 72 DPI)

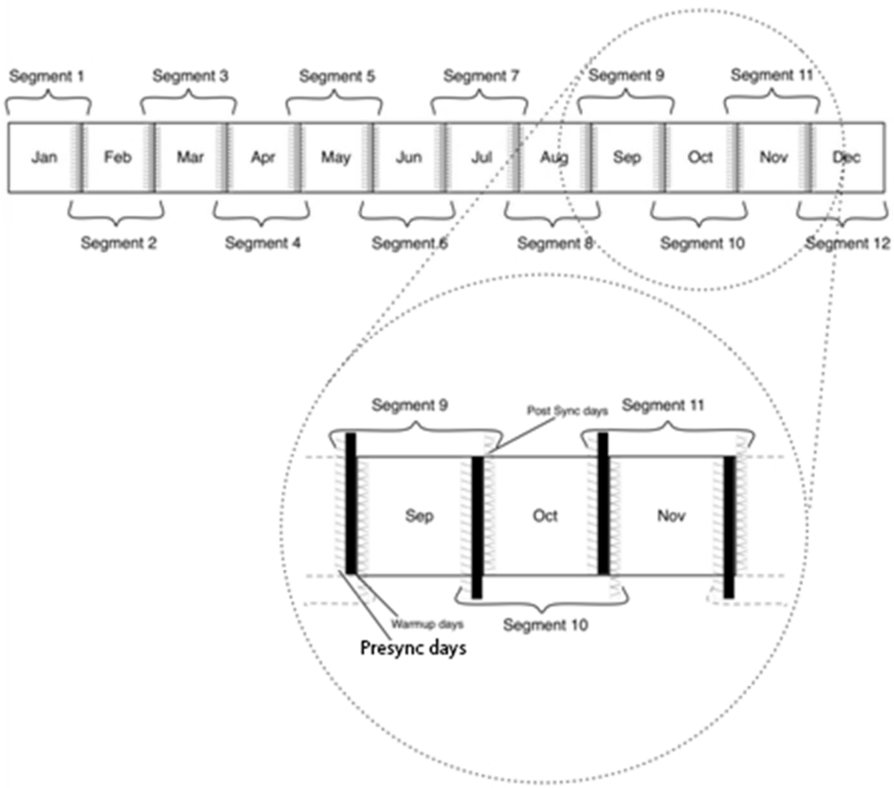


Figure 2 Methodology enhancement with Post Synchronization
162x142mm (72 x 72 DPI)

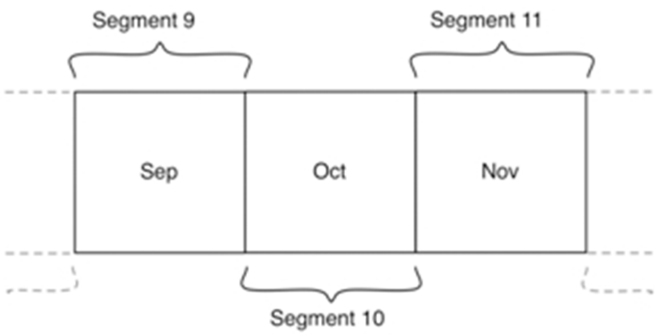


Figure 3 Representative segments for S0W0
162x142mm (72 x 72 DPI)

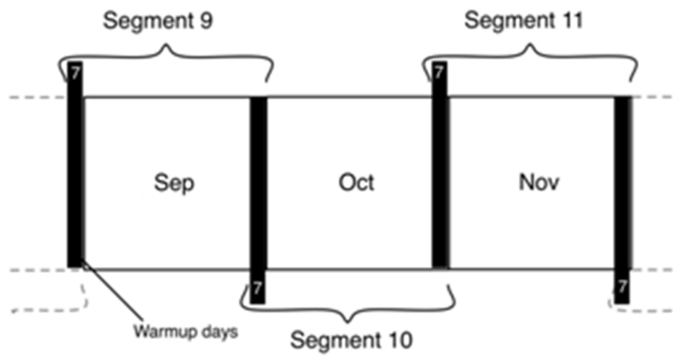


Figure 4 Representative segments for S0W7
123x65mm (72 x 72 DPI)

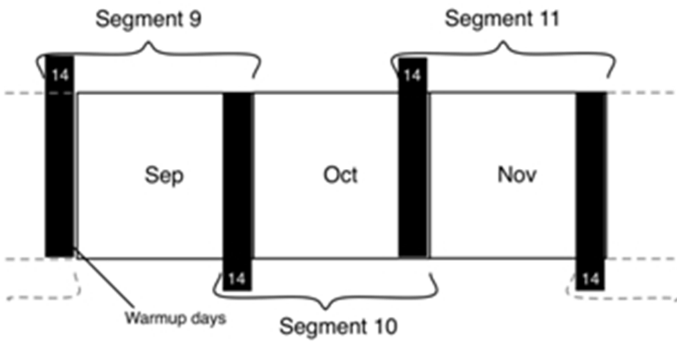


Figure 5 Representative segments for S0W14
131x66mm (72 x 72 DPI)

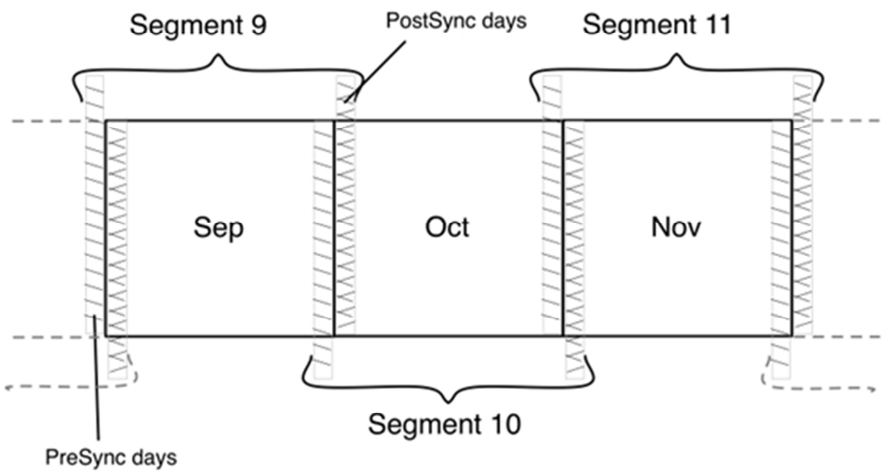


Figure 6 Representative segments for S1W0
168x95mm (96 x 96 DPI)

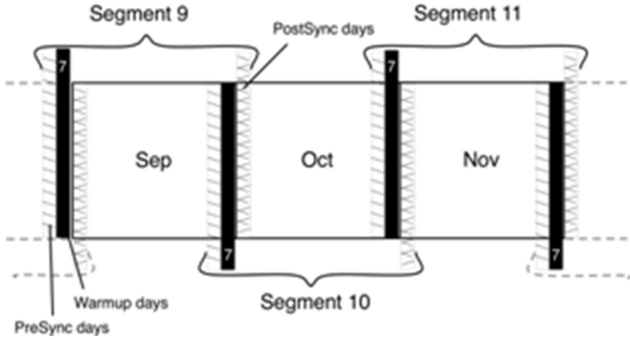


Figure 7 Representative segments for S1W7
116x62mm (72 x 72 DPI)

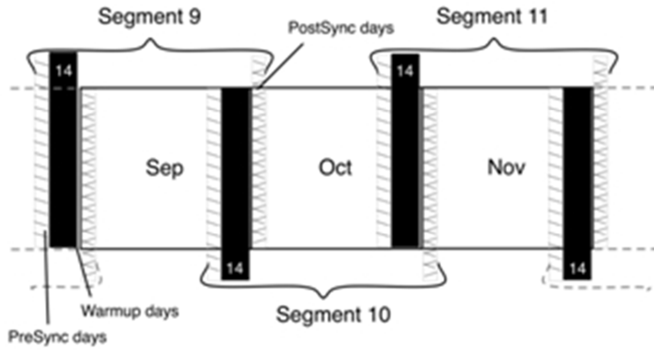


Figure 8 Representative segments for S1W14
118x65mm (72 x 72 DPI)

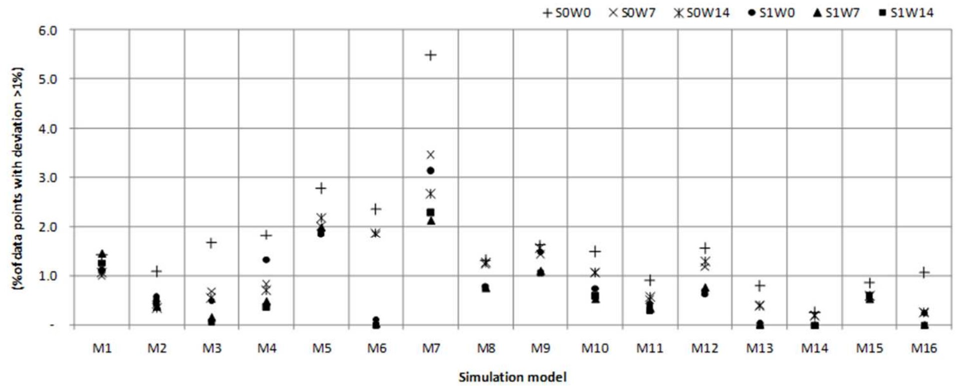


Figure 9 Percentage of data points having deviation greater than 1% and simulation model run
236x102mm (96 x 96 DPI)

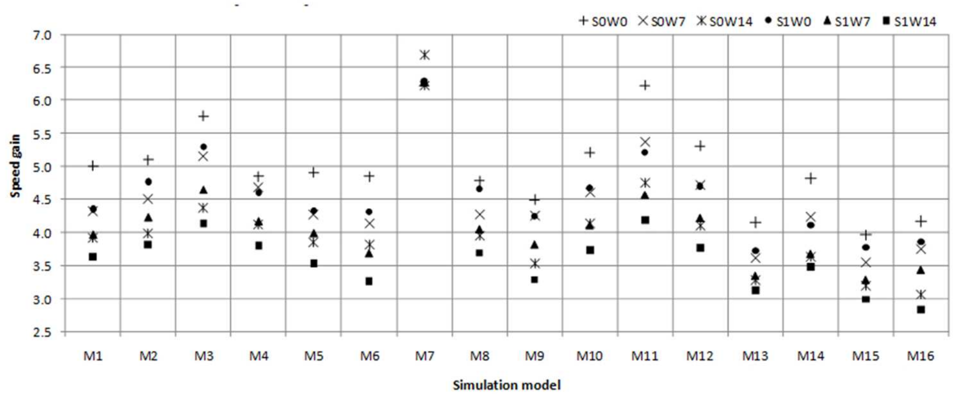


Figure 10 Speed gain for all simulation models
236x99mm (96 x 96 DPI)

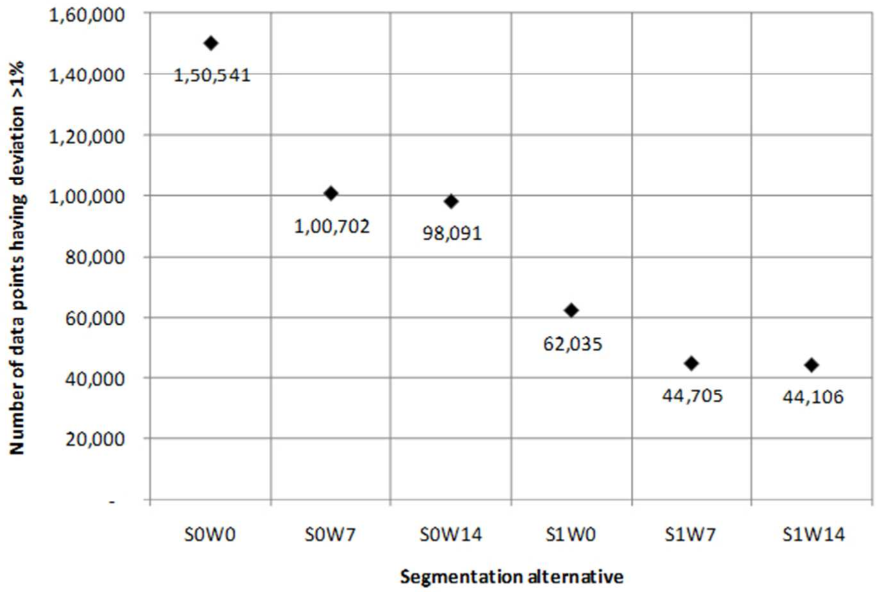


Figure 11 Number of data points having deviation greater than one percent with segmentation alternatives
188x120mm (96 x 96 DPI)

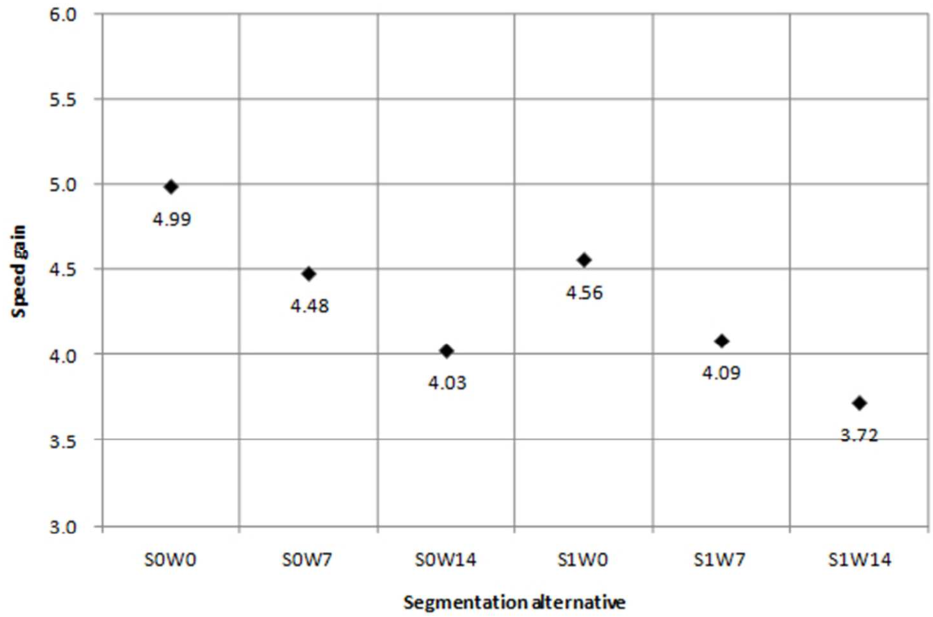


Figure 12 Simulation speed gain with different alternatives
161x106mm (96 x 96 DPI)

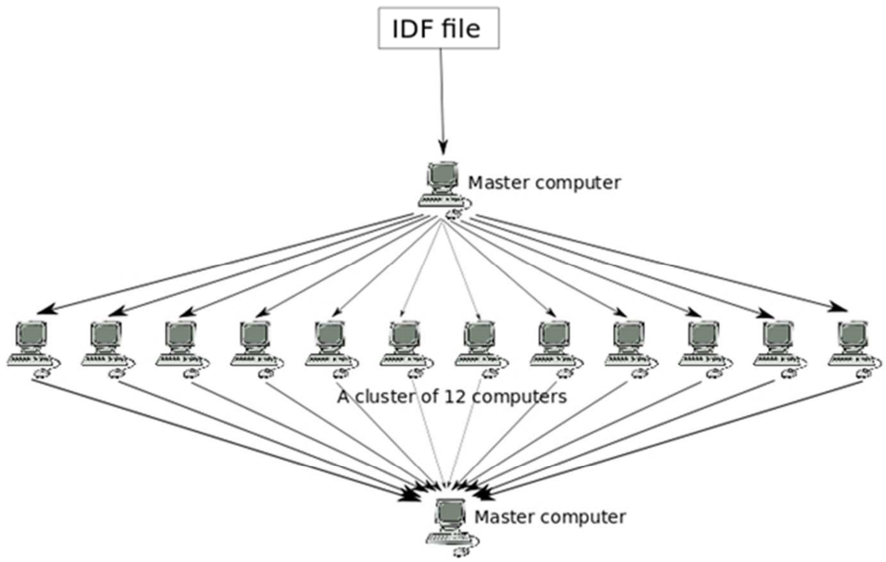


Figure 13 Cluster setup for Parallelization
187x113mm (96 x 96 DPI)

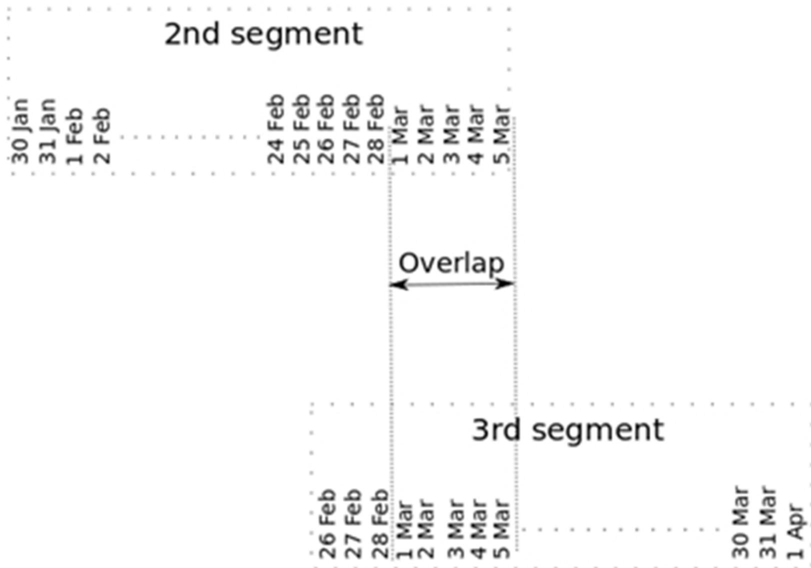


Figure 14 Overlap days for 2nd and 3rd segment to access Quality of results
118x81mm (96 x 96 DPI)

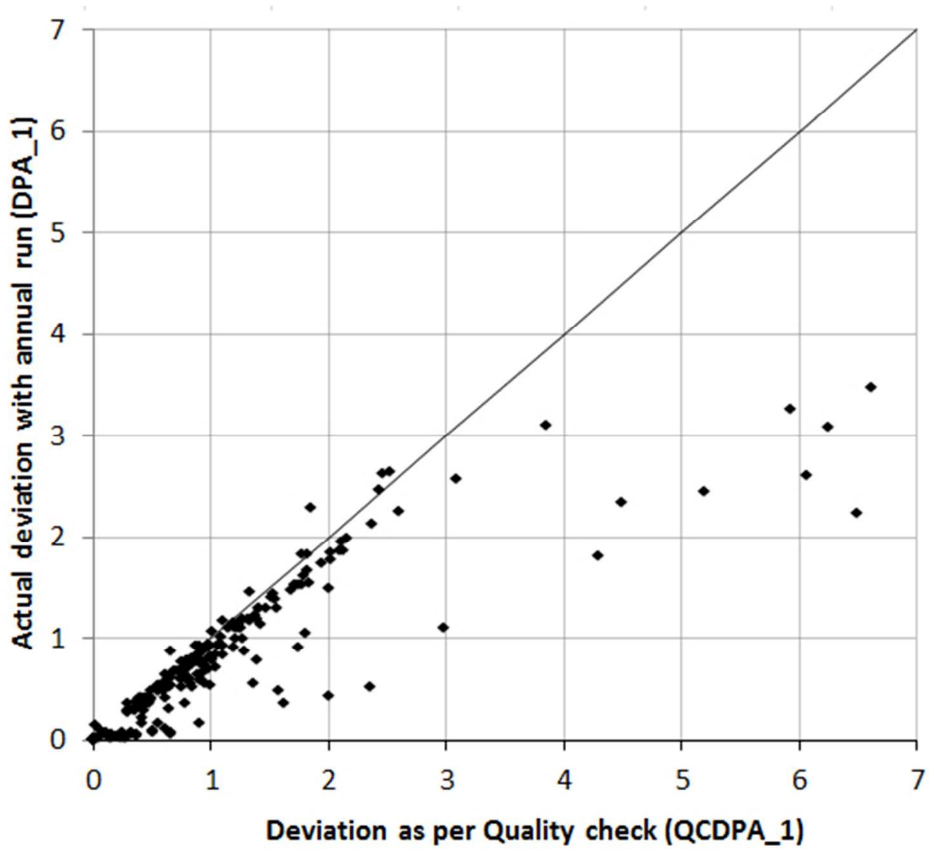


Figure 15 Deviation as per Quality check (QCDPA_1) and actual deviation with annual run simulation (DPA_1)
148x127mm (95 x 95 DPI)

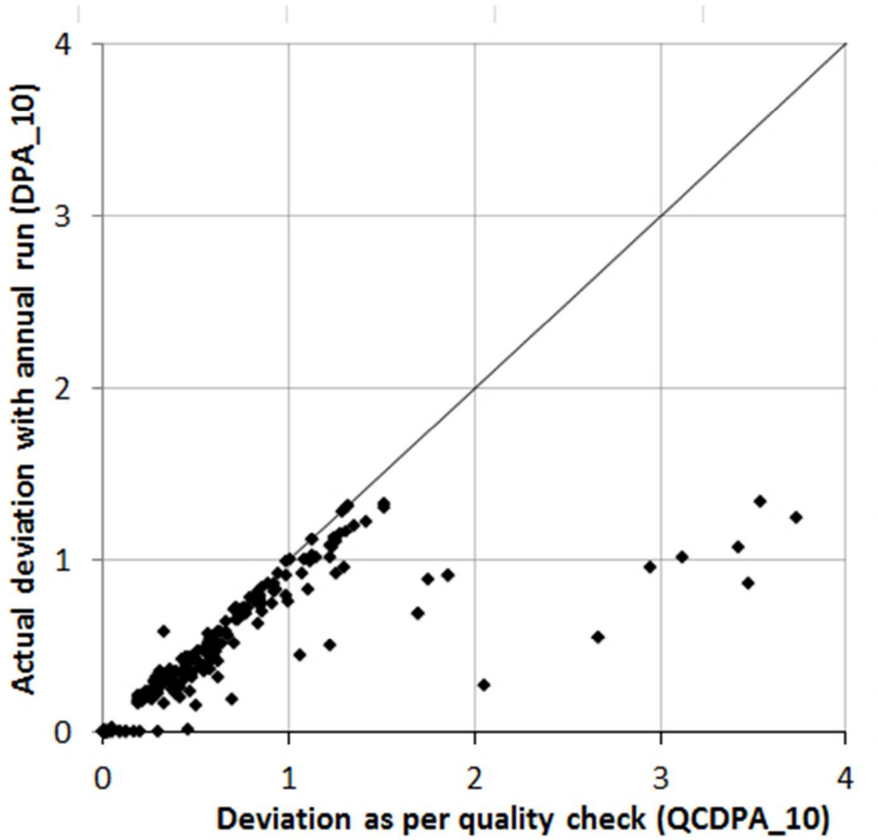


Figure 16 Deviation as per Quality check (QCDPA_10) and actual deviation with annual run simulation (DPA_10)
120x111mm (96 x 96 DPI)

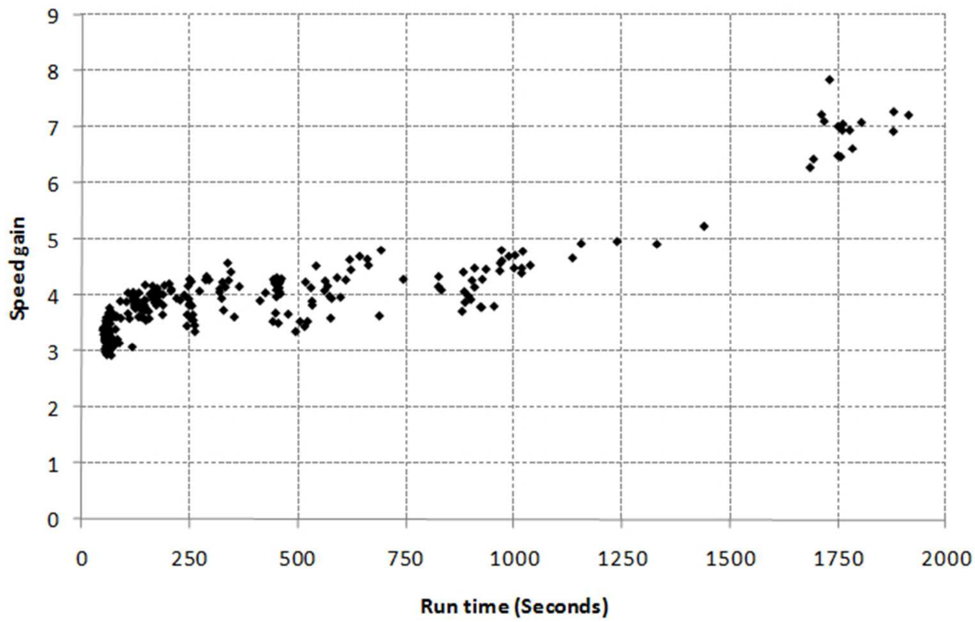


Figure 17 Simulation Run time vs. Speed gain, the time taken by the Annual simulation (in seconds) and the effective speed gain for all 256 runs
183x124mm (96 x 96 DPI)