Proceedings of ASME 2010 4th International Conference on Energy Sustainability
ES2010
May 17-22, 2010 Phoenix, Arizona, USA

ES2010-90509

ENERGYPLUS SIMULATION SPEEDUP USING DATA PARALLELIZATION CONCEPT

Vishal Garg
IIIT Hyderabad
Hyderabad, Andhra Pradesh, India
vishal@iiit.ac.in

Kshitij Chandrasen
IIIT Hyderabad
Hyderabad, Andhra Pradesh, India

Surekha Tetali
IIIT Hyderabad
Hyderabad, Andhra Pradesh, India

Jyotirmay Mathur
MNIT Jaipur
Jaipur, Rajasthan, India

ABSTRACT

EnergyPlus is an energy simulation tool which models heating, cooling, lighting, ventilation, and other energy flows in buildings. It is based on the popular features of BLAST and DOE-2. Though EnergyPlus is more advanced than DOE-2, it comes with a drawback of larger simulation runtime. A decrease in the simulation runtime of EnergyPlus would increase the usability of the software. In this paper, a concept has been proposed which uses data parallelization to speed up EnergyPlus simulation. Data parallelization is a form of parallelization for computing across multiple processors or multiple computers in a cluster, run under a suitable environment. Data parallelism focuses on distributing the data across different parallel computing nodes by breaking it into smaller chunks, each of which is processed on by the same function, running in parallel on different cores/machines. In the proposed concept, this is achieved by breaking a simulation with annual *RunPeriod* into several simulations of smaller *RunPeriod*, each handled by a separate computer. Each computer instead of running an annual simulation, handles a chunk of smaller *RunPeriod*, say one month, thus taking lesser time. This paper discusses the concept and presents the results of the preliminary investigations. It has been observed that a speed gain of approximately 6.8 times can be achieved by this method.

INTRODUCTION

Energyplus is an extremely powerful building energy simulation tool that was developed with a vision of creating a tool that would model heating, cooling, ventilating and other energy flows in buildings. Since then, it has been popular amongst architects, designers and researchers for performing simulations. Parametric analysis is a common method used to compare various available design options for improving energy efficiency. In a scenario like this, the run time of the software is of major concern. According to a study conducted by Tianzhen Hong, et al [1] , at a 60-minute time step, EnergyPlus runs slower than DOE-2.1E by a factor of 25 for a large office building to 54 for a hospital building. Yin Zhang [2] in his work, proposed to run EnergyPlus on parallel machines specifically for the parametric analysis where multiple design alternatives have to be analyzed simultaneously. The modern Operating Systems are not able to exploit the multi-core computer architectures while running EnergyPlus because of its complex code structure. However, these Operating Systems are able to distribute the load to multiple processors in case of multiple simulations running on the same computer. This vulnerability of the operating systems towards simulations in EnergyPlus can be eradicated if one single simulation is broken into several, each one of which runs as an independent simulation. This division can be achieved by breaking the demanded run period into several smaller run periods. The results of these individual independent simulations can then be compiled together into a single set of output such that the whole process appears as a single simulation. If each chunk except the first one is given an extra 7 days of RunPeriod for warming up, in addition to the warm-up that it does for convergence; it is observed that the chunking of the RunPeriod doesn't deviate the results significantly from the actual ones. Simulating the system on one example file available with the EnergyPlus version 4, the maximum deviations observed in the annual consumption of Interior Lighting is 0 %, for Interior Equipments is 0 %, for Cooling Electricity is 1.36 % and for Heating Gas is -0.42%, while the average speedup achieved was 6.8x. We present in this paper the methodology of parallelization and results are presented together with comparison of single run and parallelization results.

MODEL AND WEATHER FILES USED

To ascertain the correctness of the method, we have performed simulations on one of the Benchmark Model [3] that comes with the EnergyPlus V4.0. We have used the Large Office model. The floor area of the model is 42,757 m^2 with a total of 12 floors. Along with the general configuration of a commercial building, the building has a central air handling equipment, system equipment (auto size), plants with coils, pumps, chillers, boilers and towers.

The building has been simulated for five different cities of USA, as shown in Table 1

TABLE 1: CITIES FOR SIMULATION

City	Climate zone	Climate type
Chicago	5A	Cool Humid
San Francisco	3C	Warm-Marine
Tampa	1A	Very hot- humid
New York	5A	Cool Humid
Houston	2A	Hot- Humid

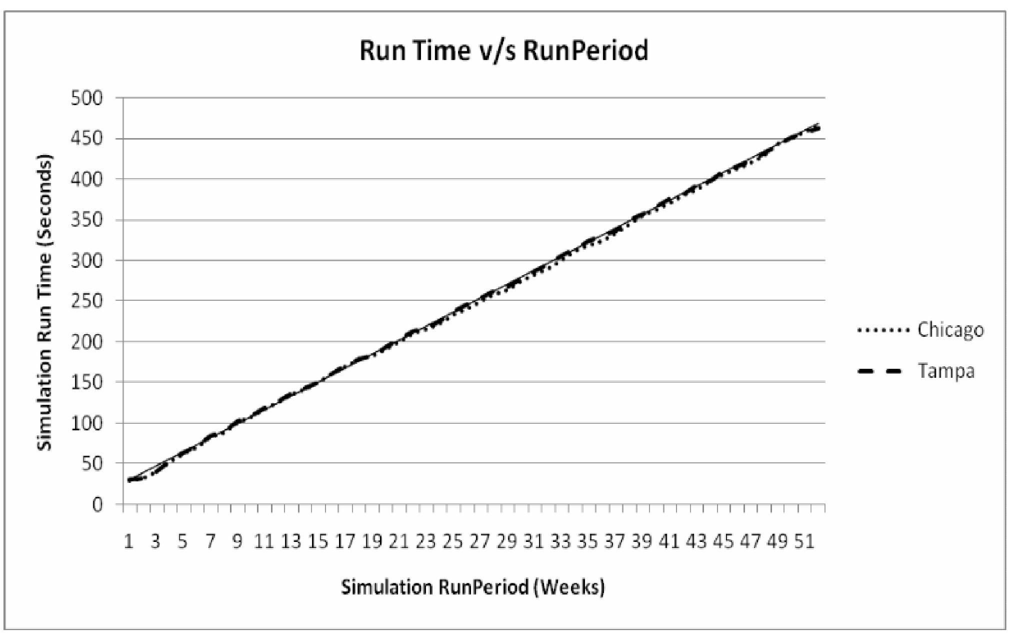
EFFECT OF RUNPERIOD ON SIMULATION RUN TIME

The EnergyPlus runtime analysis report prepared by Simulation Research Group, EETD, LBNL [4], presented the impacts of simulation settings and model features on EnergyPlus run time. To confirm the effect of *RunPeriod* on the simulation run time we simulated the model for varied *RunPeriod* (one week to 53 weeks) and observed the simulation run time. The results clearly show that the simulation runtime is directly proportional to the *RunPeriod* of the simulation. Since, the trend is representational and does not change with the weather file, results for only Tampa and Chicago are shown.

The graph in Figure 1 depicts the runtime behavior of the model with the increase in *RunPeriod* of the simulation. The runtime is found to be linearly dependent on the *RunPeriod*.

For the purpose of our simulations, we have used 2 machines with different computing capabilities. One of the machines has two Intel(R) Pentium(R) D CPU 2.80GHz, with a cache size of 1024 KB each. The other is an Intel(R) Pentium(R) 4 CPU 3.20GHz, with a cache size of 1024 K B. The former uses Fedora Core 5 operating system while the other works on Ubuntu 8.10. Simulations on processors with different processing capabilities and operating systems with different thread management systems strengthen the fact that the methods proposed in the paper are neither architecture nor platform dependent.

FIGURE 1: EFFECT OF SIMULATION RUNPERIOD ON SIMULATION RUN TIME



METHODOLOGY

The proposed method of parallelization can be applied on both multi-cores as well as a cluster. The multi-cores would limit the speedup since the general purpose computers are usually not more than quad-cores, while a cluster can easily be set up with as many machines as wanted. Cluster computing would have the drawback of network overheads, but the speedup gained would then compensate for it.

While dividing the *RunPeriod*, we took the split up period to be 1 month because of a simple reason that monthly results are generally used in practice and are also available as one of the default reporting frequencies in the software. However, the proposed method can also be applied with smaller or larger chunks, depending on the computational resources available. The method gives a definite speedup, but there is a need to ensure that the results of individual simulation chunks are within acceptable accuracy limits when compared to an annual simulation result. We will explain some of the factors that need to be kept in mind for practicing this approach. In this paper, the term *'First Chunk Day'* is used for the first day of the month for which the chunk is simulating. For instance, the April 1st will be the *First Chunk Day* for the fourth chunk.

For parallelizing simulations on a multi-core, it must be kept in mind that segmented simulation must be carried in separate folders. EnergyPlus uses SQLite for storing temporary data, which does not support multi-threading. Two simulations running in parallel in the same folder would result in errors indicating failure to acquire locks for these temporary databases. A straight solution to this would be to create separate folders for each simulation. Any operating system that supports the multi-core architecture would distribute the load of subsequent system calls to EnergyPlus on processors available. Thus, if four subsequent calls are made to EnergyPlus on a quad-core machine with an operating system that supports 4 cores, then all the four simulations will be running on different processors. Although it is advisable to use a thread library or an API to distribute the work, since then the tool will not depend on the multi-threading policies of the operating system. Another advantage would be the coordination between the simulations running in parallel which can be better achieved with an API. Once all the simulations are complete, the output of all can then be combined into a single set of files to make it appear like a single simulation.

To start the simulations on each machine, it is important that every machine has the IDF file and the weather file of the region for which the simulation is being performed. Since the data splitting is done only over run period, each machine should know the run period for which it is supposed to simulate the IDF file. The root machine (or the machine with rank 0) scatters the IDF file to all the machines in the ring. Each machine on the ring on receiving the IDF file appends the run period class, which depends on two factors – the number of machines in the ring and its own rank on the cluster. The following formulae calculate the start and the end month for a machine when the rank and the total number of machines are known -

startMonth = Rank*(floor(12/P))+min(Rank,r)
endMonth = (Rank+1)*(floor(12/P))+min(Rank+1,r)

Where,
r = Rank MOD P.
'Rank' is the rank of the machine in the cluster.
P is the total machines in the ring.

The start date is the first day of the 'startMonth' and the end day is the last day of the 'endMonth'. Each machine has a parsing code which parses the idf and appends the appropriate *RunPeriod*.

Once the input files are all set for simulation, energyplus is initiated with the files. The output results are then gathered and manipulated to form the combined output which would be the same as the output we get on running the simulation on 12 months in a single run.

Preliminary investigations showed some major deviations when the results of the segmented simulations were compared to those of the single annual simulation. These errors can be attributed to the fact that the first day of a month can be a weekend or one of the listed holidays, which would result in a warm-up on a holiday. This warm-up will affect the results of some of the days ahead also. The solution to this can be to pad extra 7 days to each simulation. As a result, the First Chunk Day would now be in the previous month. Combining the outputs of EnergyPlus would then be a trivial task as the results are sorted with date. As a proof for the concept, we performed the method on the model mentioned above for the two weather files and observed that the results were in acceptable limits.

RESULTS

Simulations have been performed for the model for all the five weather files as mentioned above. The highest deviations in both cooling and heating energy consumptions were found to be for Chicago, thus the analysis has been shown for the same. There were some interesting observations in the results. Even after padding the extra 7 days, some of the hours still showed major deviations. On further investigation, it was realized that although these deviations were large in fraction, they were very small in value. Since the original values of these results were small, a small deviation showed large percentage differences. The graphs constructed from hourly data make the mentioned facts more apparent. The hourly consumption values of heating and cooling have been analyzed in two different scenarios. One, day on which there is maximum deviation in the results and the day on which there is maximum electricity consumption. Figure 2 and Figure 4 show the hourly data on the day when the maximum deviations are observed for heating and cooling energy consumption respectively. Figure 3 and Figure 5 shows the hourly data of the day on which there is maximum energy consumption for heating and cooling, respectively. All these hourly analysis graphs are plotted for the Chicago climate.

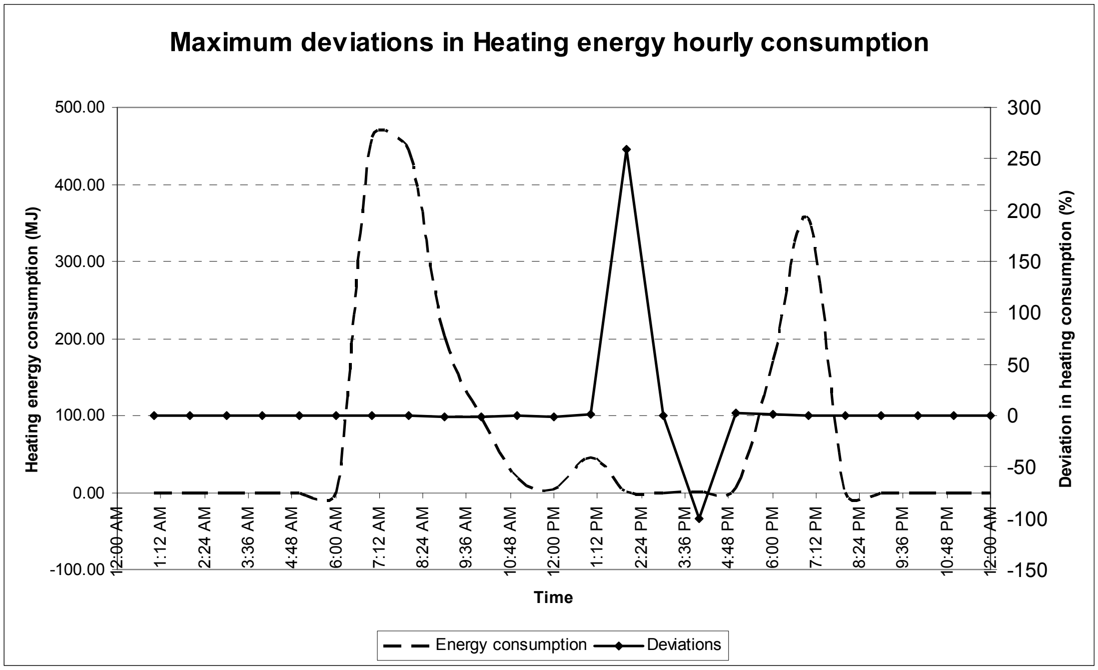


FIGURE 2: PERCENTAGE DEVIATIONS IN HEATING ENERGY HOURLY CONSUMPTION OF THE CHUNKED SIMULATION WHEN COMPARED WITH THE ANNUAL SIMULATIONS, FOR CHICAGO WEATHER. THE MAXIMUM DEVIATIONS IN HOURLY CONSUMPTION NOTED ON THIS DAY IS AROUND 250%.DECEMBER 11TH SHOWS THE MAXIMUM DEVIATIONS IN HEATING ENERGY HOURLY CONSUMPTION.

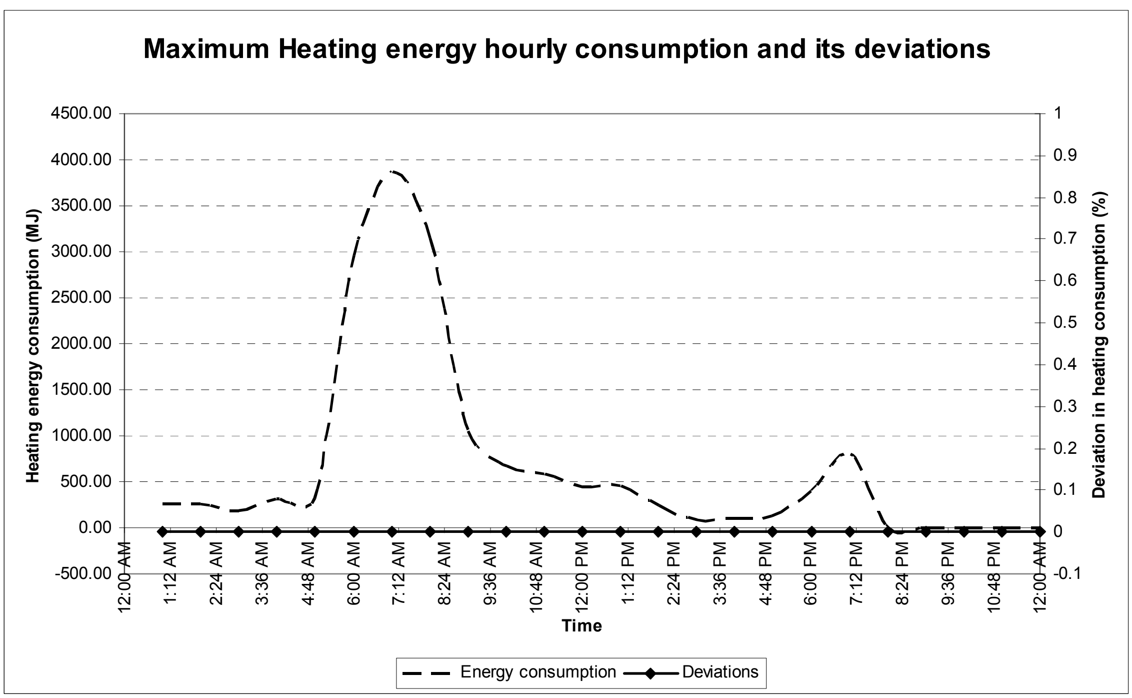


FIGURE 3: PERCENTAGE DEVIATIONS IN HEATING ENERGY HOURLY CONSUMPTION OF THE CHUNKED SIMULATION WHEN COMPARED WITH THE ANNUAL SIMULATIONS, FOR CHICAGO WEATHER. THERE IS NO DEVIATION IN THE RESULTS ON THIS DAY, WHEN THE HEATING CONSUMPTION IS MAXIMUM.JANUARY 9TH IS THE DAY WHEN MAXIMUM HEATING CONSUMPTION VALUES ARE OBSERVED.

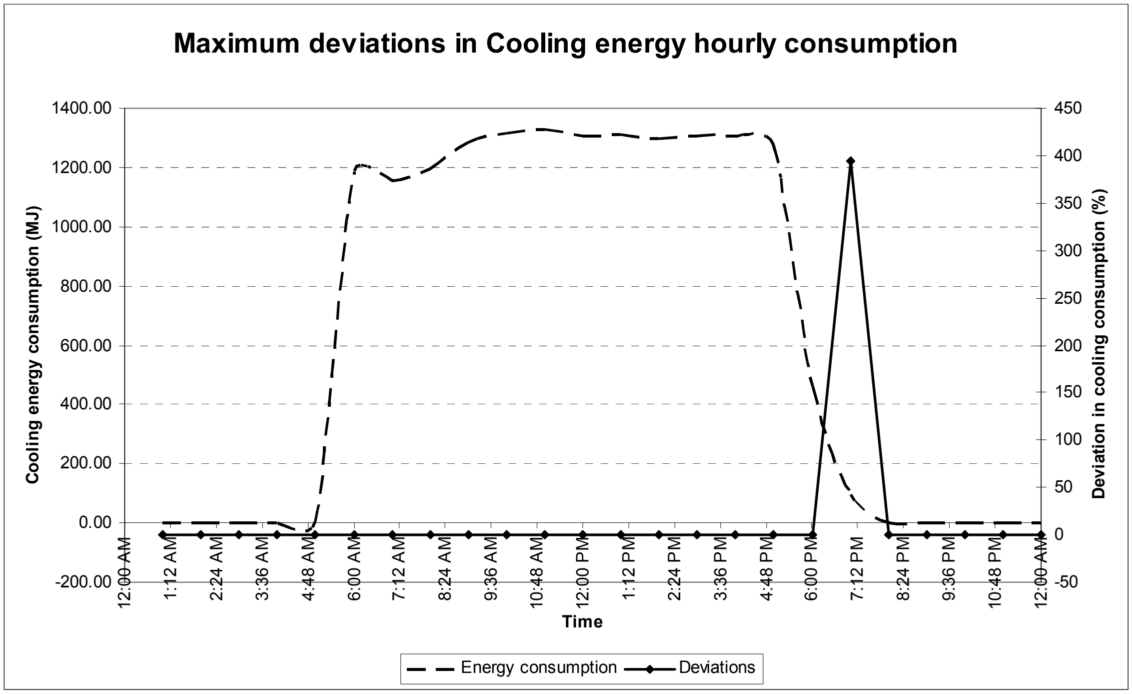


FIGURE 4: PERCENTAGE DEVIATIONS IN COOLING ENERGY HOURLY CONSUMPTION OF THE CHUNKED SIMULATION WHEN COMPARED WITH THE ANNUAL SIMULATIONS, FOR CHICAGO WEATHER. THE MAXIMUM DEVIATIONS IN HOURLY CONSUMPTION NOTED ON THIS DAY IS AROUND 400%.AUGUST 12TH SHOWS THE MAXIMUM DEVIATIONS IN COOLING ENERGY HOURLY CONSUMPTION.

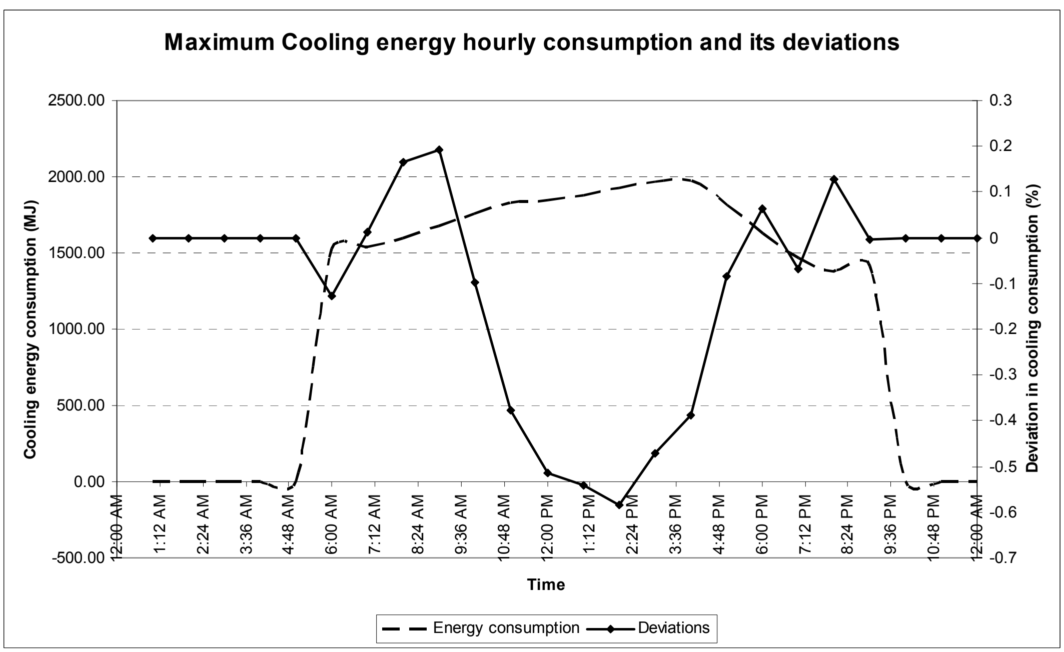


FIGURE 5: PERCENTAGE DEVIATIONS IN COOLING ENERGY HOURLY CONSUMPTION OF THE CHUNKED SIMULATION WHEN COMPARED WITH THE ANNUAL SIMULATIONS, FOR CHICAGO WEATHER. THE DEVIATION VARY IN RANGE OF -0.6% TO 0.2% ON THIS DAY WHEN THE COOLING CONSUMPTION IS MAXIMUM. JULY 31TH IS THE DAY WHEN MAXIMUM COOLING CONSUMPTION IS RESULTED.

In Figure 2, the energy consumption at the point of maximum deviation is as low as 0.33 MJ. Small deviations in the consumption at such low values bring about a large percentage deviation in the results. Similar is the case with Figure 4, the energy consumption at the point of maximum deviations is 93.62 MJ. In Figure 3, which shows the maximum heating energy consumption day, where the consumptions are as high as 3877.92 MJ, the percentage deviation is 0. Similarly in Figure 5, maximum consumption was found to be at 458.36 MJ, whereas the deviations were low at -0.3%.

With the investigations done in order to find an efficient way of parallelizing EnergyPlus, it was realized that breaking the *RunPeriod*, padding 7 days to each segment, simulating them in parallel would be an appropriate method. To emphasize on the correctness of the method, some of results and deviations have been shown on both, monthly and annual basis in the Table 2 and Table 3, for cooling and heating respectively. The maximum deviations for cooling resulted in March for Chicago and the minimum deviations resulted in month of February for New York, as shown in Table 2. For heating, maximum deviations resulted in the month April for Tampa, and the minimum deviation resulted in the month of July for Chicago as shown in Table 3. The effective annual speed gain achieved is the minimum speed gain achieved for the segmented simulation. The time taken for the simulation and the speed up achieved for each segment simulation and the effective speed gain are as shown in Table 4

TABLE 2: PERCENTAGE DEVIATIONS IN COOLING ELECTRICITY CONSUMPTION

Month	Chicago	Tampa	New York	Houston	San Francisco
January	0.0001	-0.0196	-0.0331	-0.0043	0.0000
February	0.0173	-0.0934	-0.4277	0.0524	-0.0606
March	**1.3665**	-0.0957	0.1967	-0.1915	-0.0208
April	-0.2974	-0.1789	0.1208	-0.0428	-0.2290
May	-0.2842	-0.0504	-0.0173	-0.0867	-0.0829
June	-0.0309	-0.0048	-0.0020	0.0018	-0.0413
July	0.0790	0.0139	-0.0131	-0.0677	0.1469
August	-0.0076	0.0240	0.1381	-0.0021	0.1776
September	-0.1937	0.1008	0.0353	0.0472	0.1069
October	0.0856	0.1125	-0.0963	0.0470	0.1249
November	0.0064	-0.0088	0.0813	-0.0748	0.0748
December	0.0371	-0.0725	0.0013	-0.0578	-0.1096
Annual	-0.0429	-0.0073	0.0343	-0.0220	0.0482

TABLE 3: PERCENTAGE DEVIATIONS IN HEATING ELECTRICITY CONSUMPTION

Month	Chicago	Tampa	New York	Houston	San Francisco
January	0.0007	0.0019	0.0070	0.0043	0.0020
February	0.0743	-0.2017	0.1440	0.4383	0.0668
March	0.2540	-0.0977	0.1475	-0.1890	0.0955
April	0.1783	1.3131	-0.7155	-0.3357	0.2191
May	0.1620	0.5588	0.2038	0.1973	0.5044
June	-0.1628	0.1509	0.1767	-0.0895	0.0612
July	**-1.8882**	0.1116	0.3124	-0.0281	0.4823
August	0.0389	-0.1754	1.2339	-0.7242	0.4669
September	-0.6531	-0.2822	0.1970	0.6710	-0.4395
October	-0.0111	0.0871	0.1557	0.1074	0.0890
November	-0.0459	-0.1654	0.0852	-0.1304	-0.1425
December	-0.0187	0.2110	-0.0319	0.1945	0.1069
Annual	0.0298	0.0593	0.0193	0.0605	0.0763

TABLE 4: THE TABLE SHOWS THE RUNTIME OF THE SEGMENTED SIMULATIONS AND THE ANNUAL SIMULATION, AND FRACTION OF THE FORMER WITH THE LATTER. THE MINIMUM FRACTION IS THE NET SPEEDUP ACHIEVED FOR THAT CITY.

Month	Chicago		Tampa		New York		Houston		San Francisco	
	Time (sec.)	Speed Up	Time (sec.)	Speed Up	Time (sec.)	Speed Up	Time (sec.)	Speed Up	Time (sec.)	Speed Up
January	81.66	6.3	78.94	5.9	28.768	8.5	25.94	9.2	26.495	9.1
February	73.58	7.0	69.67	6.7	30.328	8.1	26.783	8.9	27.202	8.9
March	71.99	7.1	68.87	6.8	30.197	8.1	28.245	8.4	29.894	8.1
April	70.69	7.3	67.5	6.9	28.517	8.6	27.89	8.5	29.173	8.3
May	71.79	7.2	68.34	6.9	33.609	7.3	32.296	7.4	32.419	7.4
June	71.52	7.2	68.91	6.8	28.138	8.7	28.497	8.4	29.21	8.3
July	71.08	7.2	69.63	6.7	30.272	8.1	28.608	8.3	29.472	8.2
August	78.05	6.6	78.13	6.0	28.574	8.6	28.734	8.3	29.877	8.1
September	71.8	7.1	70.22	6.7	28.554	8.6	28.706	8.3	30.27	8.0
October	76.32	6.7	71.05	6.6	28.532	8.6	29.234	8.1	29.415	8.2
November	75.12	6.8	66.8	7.0	28.304	8.7	29.082	8.2	28.236	8.5
December	70.14	7.3	62.59	7.5	31.183	7.9	32.514	7.3	30.332	8.0
Annual	**513.3**	**6.3**	**468.51**	**5.9**	**245.649**	**7.3**	**238.055**	**7.3**	**241.396**	**7.4**

Although there are deviations to the score of 250% in heating and 400% in cooling in the hourly comparison, the monthly and the annual deviations show the numbers to be not more than 1.37% in cooling and 1.89% in heating. These figures suggest that if the speedup attained is good, there can be a giveaway for such low deviations. When experimented, it was found that the speedups ranged from 5.9 times to 7.4, whereas the average was 6.8 times the actual speed of a sequential run. Although the average speedup of each chunk obtained is as high as 7.7, the effective average speedup obtained is low because of the colder months taking longer simulation time, hence creating a bottleneck. This is probably because of the fact that the colder months require more HVAC iterations for warm up.

CONCLUSION

Comparison of simulations of the single run and parallelization show that the deviations found in the results on monthly and annual basis are less. There are rare cases of major deviations on the hourly basis though. This effect on the hourly data will not affect analysis at the macro level. The model of a typical commercial building generally takes several minutes for simulation. During the initial designing/modeling of a building, a lot of simulations are required to be done. For such cases, a speed up of 6-7 times can save a lot of time. For the final analysis however, a singe run simulation as per conventional approach can be performed to get precise results. Further investigation needs to be done in order to look for other reasons causing the deviations during parallelization.

REFERENCES

[1] Tianzhen Hong, Fred Buhl, Philip Haves, Stephen Selkowitz, Michael Wetter. Comparing Computer Run Time of Building Simulation Programs. Build Simul (2008). July 2008. p.210-213.

[2] Yi Zhang. “Parallel” EnergyPlus and the development of a parametric analysis tool. Eleventh International IBPSA Conference. July 2009.p.1382-1388.

[3] Paul Torcellini, Michael Deru, Brent Griffith, Kyle Benne, Mark Halverson, David Winiarski, Drury Crawley. Building Benchmark Models. National Renewable Energy Laboratory. July 2008.

[4] Tianzhen Hong, Fred Buhl, Philip Haves. EnergyPlus Run Time Analysis. Ernest Orlando Lawrence Berkley National Laboratory. December 2008.