

Pitfalls in Weather Data Management Strategies of Building Performance Simulation Tools

Livio Mazzarella – Politecnico di Milano – livio.mazzarella@polimi.it

Martina Pasini – Politecnico di Milano – martina.pasini@polimi.it

Abstract

The use of simulation for building design and performance assessment is becoming mandatory if NZEB requirements have to be met. As a matter of fact, only dynamic simulations are able to correctly account for renewable energy exploitation on the building site. Dealing with solar energy source conversion, the correct use of available standard weather data files is still more important than solar gains through the transparent envelope. When using such statistically derived weather data, however, different pitfalls might arise. In fact, when performing sub-hourly simulation, the information provided by hourly-based climatic data is insufficient. Interpolation algorithms are implemented in Building Performance Simulation tools (BPSTs) to provide intermediate weather data, which can affect the quality of the results. Specifically, solar radiation data are insufficiently represented by hourly-based values when dealing with short time-step simulations of complex building systems. In this article, a review of radiation algorithms and weather data management algorithms at sub-hourly simulation time steps will be introduced, as implemented by two well-known software, such as TRNSYS 17 and EnergyPlus 8.6.0. Further considerations will be made upon information exchange among simulation components during the simulation, and upon the possibilities offered or denied by different data management-implementations when multiple actors are involved in the simulation.

1. Introduction

In complex systems like buildings, control strategies become vital when looking for energy use reduction and renewable sources exploitation. However, when dealing with complex systems that exhibit non-linear behaviours which may also depend on thresholds, integral or averaged information over time might not be any more sufficient to assess the

performance of that particular system. In that case, the “instantaneous” time dependence of some of the input data might become important to get a realistic comprehension of the behaviour of the enquired system. Here instantaneous means short, compared to the smallest system characteristic time. Short time step simulations might therefore become the only possibility to compare different systems at design time.

Starting from these considerations, we have tried to evaluate the impact on simulation in “averaging” those inputs that show an extremely variable behaviour, as solar radiation in particular.

When using weather data in a simulation, four main aspects should be taken into consideration:

1. Are the requirements of correct timing and synchronization of weather data with the different nature (instantaneous or averaged) fulfilled?
2. Is the accuracy, by which the solar position is calculated with, consistent with small time steps?
3. What is the loss of information when averaging weather data?
4. What is the impact of different weather data interpolation strategies?

One of the basic requirements, when dealing with information exchange, for both internal and external objects involved in a simulation, is the knowledge of the nature and occurrence timing of the information. In some cases, an “accurate” synchronization of different kinds of information might not be so relevant, such as at pre-design time when the average “variation pattern” is enough to get design values to size the system. In other cases, when studying real situations, such as during the empirical validation of BPSTs against measured data or at operational time, by using monitoring weather data

to optimize the performances of the building system, the correct synchronization between data with different natures, both influencing the behaviour of the system (such as temperature and solar radiation in photovoltaics systems efficiency), becomes important.

Following these considerations, a review of the algorithms implemented in two well-known BPSTs, such as TRNSYS 17 and EnergyPlus 8.6.0, was performed.

2. Aspects to be Considered When Addressing Short or Hourly Time Steps Simulations

2.1 Information Exchange: Timing and Consistency

One of the most important aspects to be addressed when simulating systems comprised of different objects is the exchange of information among the different actors involved. The two main properties that this information should contain are: its nature and its timing.

When speaking of timing, it should be clear that:

- the *time step* is the time interval between two subsequent calculations,
- while the *time stamp* is the exact time to which the results of each calculation are referred (the difference between two subsequent time stamps is thus the time step).

The most commonly used numerical methods in solving partial differential equations, such as finite volumes, finite differences, and conduction transfer functions, provide as their results the instantaneous values of the dependent variables (temperatures and/or fluxes) at a well-defined time stamp, and not an average value on the previous time step. To get to this result it is always necessary to know the instantaneous values of their initial state and – depending on the chosen numerical scheme – of their boundary conditions at the previous and/or current time stamp. For example, to solve the problem of conductive heat transfer in a wall, the fluxes and the temperatures on their boundaries at precise time stamps need to be known.

Besides, in a simulation, a manager must always be “elected” that “conducts the orchestra”, here called *simulation manager*.

The simulation manager should ask all the actors involved in a simulation to perform their calculation and expose to all the other components their properties at each precise time stamp. These time stamps will be referred to as *global time stamps* and the time in between the two subsequent as *global time steps*.

The global time step should be as short as required by the actor that is more sensitive to a rapid variability of its boundary conditions.

That doesn’t mean that all the actors will be “forced” to repeat their calculation at each global time stamp. Each of them will be allowed to expose its properties as unvaried and store the information coming from the other objects at its convenience.

Another possibility, might be the one that sees an object which needs to perform its calculation with shorter time steps to reach more accurate results, but not that much sensitive to the “exact form” of variation of its boundary conditions (linear, random, etc.). In this case it should not “ask” the simulation manager to reduce its global time step. It will collect the needed information at the global time stamps and perform multiple calculations to give out its results at each global time stamp.

A last remark about these definitions relates to the nature of these time stamps, i.e. what kind of time (solar, universal, local, etc.) is used. The choice of the nature of the time stamp has to be made, among the other reasons, to reduce the interpolation needed on the available input data. Therefore, since in the majority of cases, other time-dependent input data, as the schedules defined for describing user habits (such as working hours, etc.), are based on local time, time stamps should be *local* instead of *solar*. Local time can be affected by legal corrections or not, and this aspect should be managed by the simulation manager and well documented to avoid confusion in the input and output reading/writing.

Therefore, we know the input data at specific local time stamps, such as users-schedules, instantaneous weather climatic data, such as dry temperature, relative humidity, wind velocity, and integral or averaged weather data, such as global and diffuse horizontal solar radiation, at a file format dependent time stamp (solar, local, etc.).

2.2 Weather Data Nature and Timing

We have said that the input data are not always available as “instantaneous”. Among the data to be exchanged, the most complex to manage are the weather-related data, due to:

- the different nature of its quantities (some of which are intensive/scalar and other extensive/vectorial)
- the relative time stamp at which these data are available.

A review performed on both the weather data file format manuals (Wilcox and Marion, 2008) and the Guide to Meteorological Instruments and Methods of Observation (Jarraud, 2008) showed that some weather data are recorded as instantaneous values associated to a time stamp, other are integral values of a variable evaluated on the previous time step.

Weather data that are relative to solar radiation are given in almost all weather data file formats as integrated over time, i.e. as the total amount of solar radiation (energy) received during the period ending at the time stamp associated with the datum and starting at the timestamp associated with the previous one.

The weather processor has instead to provide each object with the total irradiance (power) striking a specific oriented surface at each global time stamp. Thus, this information provided in terms of *energy*, should be translated in terms of *power* and used with a correctly aligned sun position, to calculate the correct amount of irradiance reaching each oriented surface.

That means that when a rule is set to assure the time alignment of provided weather climatic data (i.e. the instantaneous data and those transformed from integral to instantaneous), a consistent rule should also be set for the alignment of the sun position. In this regard, before going more into the details of how to manage instantaneous and integral weather data, a review upon sun position algorithms is introduced here to clarify the answer to the following questions:

- what is the timing of the solar position used by the different tools,
- what is the accuracy of the algorithms used for the calculation of the sun position, when the

simulation time step is of the order of magnitude of minutes.

2.2.1 Sun position timing

In some cases (TRNSYS) we have the possibility to define the weather data as instantaneous or averaged and on which time interval it has been averaged, if needed. In other cases we might need to modify the input data to time align them as required. However, there is always something that we cannot change, i.e. the way the tool manages the sun position timing. Both EnergyPlus and TRNSYS expose to all other procedures the sun position that is evaluated at the middle of the global time step regardless any other choice. Accordingly, for time consistency, we should expose to all the objects involved in the simulation all the other input data at that particular *mid time stamp* (i.e. at the mid of the time step) performing interpolations, even if we have instantaneous data available at the global time stamps. In our opinion, this is not the best choice because:

- unneeded interpolations are performed;
- integral weather data need a transformation, which could be better performed when aligning the transformed data to the time stamps of the instantaneous data;
- schedules data are usually defined at the time stamp as well, and not at the mid of the previous time step.

2.2.2 Sun position accuracy at short time steps

As far as the accuracy of the sun position calculation is concerned, we have done a review of the most broadly diffused algorithms (Duffie and Beckman, 2013) to understand if the simplified hypotheses at their basis are still applicable when dealing with short time step simulations. We detected some ambiguity among different sources concerning the definition of the fractional year; therefore we recovered the original source for the definition of the most accurate equations for the calculation of sun declination and equation of time (Spencer, 1971). We found a witness (Oglesby, 1998) that reported one error in the first coefficient of Spencer’s equation of time. We will report it here only for completeness:

$$E = 0.0000075 + 0.001868 \cos(T) - 0.032077 \sin(T) - 0.014615 \cos(2T) - 0.040849 \sin(2T) \quad (1)$$

In Spencer’s equation of time there is a fractional-year time “T” depending on the day-of-the-year number “d” which ranges from 0 on 1 January to 364 on 31 December, that has the equation: $T = 2 \pi d/365$. Since the sun declination is kept constant over one day and the author (Spencer, 1971) suggested using those equations in years in the middle of a 4-year leap cycle, we wondered:

- what might be the rate of change of declination in one day and
- what would happen if we need to compute the sun position when using monitored weather data near or for a leap year.

We noticed, by using the nautical almanac algorithms for the calculation of the sun position, that the rate of change of its declination in one day is maximum 0.4°.

To answer the second question, we calculated the error of Spencer’s simplified algorithm on the daily-average declination with respect to the more accurate algorithm of the nautical almanac. For a year before the leap one, this error depends on the longitude of the location and varies along the year, reaching maximum absolute values that are of the same order of the maximum daily declination variation. Using, in Spencer’s equations, a fractional day-of-the-year number, calculated at the beginning, mid or end of the day, a bigger or smaller error results depending on the longitude of the location, since universal time is not taken into consideration. However, the accuracy reached with Spencer’s simplified equations is still acceptable when calculating the zenith of the sun even at very short time steps.

2.3 Using Weather Data Recorded on a Particular Time Basis

Referring to weather data availability and “desired” simulation time step two scenarios might occur:

- weather data recorded on a short time step basis (minutes or seconds) are available and hourly simulations are intended,
- hourly weather data are available and short time step simulations are needed.
- In these cases, the following questions have arisen:
 - what do we lose if we use hourly weather data generated from short time step monitored data that preserve integral solar radiation?
 - which interpolation strategy is the most effective when dealing with data estimation between its recorded values?

2.3.1 From data recorded on a short time step basis to hourly data

In this first case we have weather data recorded on a minute or second time basis, but we would like to set the global time step of our simulation to 1 h. This might be the correct approach to reduce calculation time if we imagine that the components involved in the simulation do not need to know the “exact variability” of the data, while only integral values of radiation are of interest. If we want to perform hourly simulation, we cannot use the weather data recorded at those short time steps as they are. Neither of the two tools taken in consideration allows it, and in general it is difficult others might do it. Therefore, those data need to be transformed and we want to understand which error this process introduces. We used the data collected for the year 2016 (in particular the month of April) by the weather station of the Energy department of the Politecnico di Milano. The data were recorded on a 10 s time step basis. We want to know how the integral over the month of April of the solar radiation reaching the most common expositions is influenced by the integration of those data (needed to be able to run a simulation with a global time step of 1 h).

First of all, by doing this operation, we have a smoothing of the original data, as can be seen in Fig. 1.

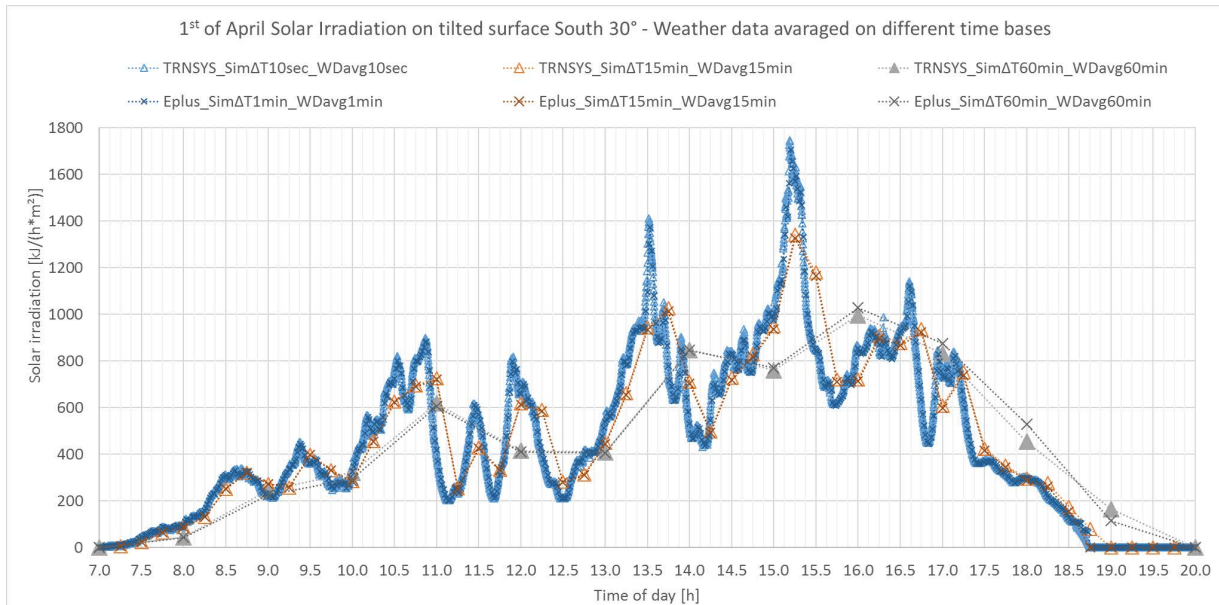


Fig. 1 – Effect of the choice of different time bases for averaging monitored data

A note about all the figures shown here: All the graphs are represented with dotted lines for an easy time series reading, even if they are intended to show discontinuous, averaged over the global simulation time step, pseudo-instantaneous values as provided by the investigated tools. These values have been associated to the time stamp provided by the tool, even if the tool calculates them as average values over the simulation time step.

Table 9 – Weather data recorded at a short time basis used on hourly simulation

Exp	WD ΔT [min]	April Sol.Rad. [MJ/m²]		Diff [%]	
		TRNSYS	Eplus	TRNSYS	Eplus
South 30°	15	565	519	0.46	0.36
	60	567	526	0.89	1.61
Hor	15	498	461	0.32	-0.1
	60	501	465	0.98	1.07
North	15	128	124	0.30	-0.47
	60	132	122	3.43	-2.12
South	15	351	333	0.55	1.16
	60	353	335	1.09	1.86
West	15	409	373	1.33	3.26
	60	430	361	6.41	-0.08
East	15	219	207	0.28	-2.82
	60	222	208	1.47	-2.36

Together with the smoothing, also the integral of the solar radiation that reaches the different expositions

is influenced. Taking as reference values those computed with the smallest global simulation time step allowed by the tool (10 seconds for TRNSYS and 60 seconds for EnergyPlus), we can see in A note about all the figures shown here: All the graphs are represented with dotted lines for an easy time series reading, even if they are intended to show discontinuous, averaged over the global simulation time step, pseudo-instantaneous values as provided by the investigated tools. These values have been associated to the time stamp provided by the tool, even if the tool calculates them as average values over the simulation time step.

Table 1 that we get a difference in the integral value of the total solar radiation incident on each exposition which depends on the particular kind of exposition. EnergyPlus gives lower solar radiation integrals with respect to TRNSYS. We report these considerations to show that preserving the integral on those input data related to solar radiation is not a guarantee of an “accurate” integral on the computed radiation reaching different expositions.

2.3.2 From hourly weather data to short time step simulation

We have seen in the previous paragraph what happens in the first scenario identified in Paragraph 2.3. Now we will consider the second scenario.

Naturally, when dealing with short time step simulations, the best solution should be to work with

weather data recorded with the same frequency. However, we are still working with hourly weather data, even if Crawley, Hand and Lawrie (1999) pointed out that this kind of data is no longer enough since interpolating between hourly observations does not accurately represent weather conditions that change much more frequently.

Given that, we need to assess the possibilities offered by the different tools to overcome such lack of information.

While for the instantaneous variable, the interpolation algorithm commonly chosen is linear interpolation with solar radiation; different tools have chosen different strategies.

In particular TRNSYS and EnergyPlus chose different interpolation algorithms for solar radiation. EnergyPlus decided to convert the integral value of solar radiation (energy) associated with a time stamp in the weather data file to an average instantaneous value of solar irradiance (power) associated to a time stamp in the middle of the current and previous ones (backward middle time). After that, the instantaneous value at each time stamp (irradiance) is calculated as linear interpolation between the average irradiance attributed to the backward middle time and that attributed to the forward middle time (Fig. 2).

TRNSYS instead, chose to interpolate the values gained from the weather data file for horizontal solar radiation by using the curve for extra-terrestrial radiation. This kind of interpolation is more heavy computationally, than the one implemented in EnergyPlus, therefore its greater effectiveness should be evaluated.

However, care should be taken when importing user defined weather data in TRNSYS, because the following two possibilities are allowed, i.e. using:

- one object, i.e. the Type 99 that combines external data reader and solar processor;
- two objects, i.e. the Type 9 for external data reading and the Type 16 for processing solar information.

These two ways give different results, as can be seen in Fig. 3 where we have the comparison between solar irradiance (power) computed by the two Types for a simulation with 15-minute global time steps, using input data averaged on a 60-minute time basis, and a 15-minute simulation using instead input data averaged on a 15-minute time basis. As written in the manual, Type 99 performs radiation "smoothing", while Type16 does not, in both cases the sky model of Perez 1999 was used, but site altitude could not be set. Table 2 shows the difference in respect to the reference values, of the integrals of the solar radiation striking different expositions as given by the smoothed, non-smoothed and linear interpolations, according to the exposition.

Table 10 – Integrals of the solar radiation with different interpolation algorithms of hourly weather data at time steps of 15 min

Exp	Diff [%] over the month of April		
	Type 99	Types 16+9	EnergyPlus
South 30°	1.13	0.94	1.34
Hor	0.86	0.74	0.75
North	1.02	1.77	-1.14
South	1.27	1.07	1.69
West	2.55	4.26	0.44
East	2.16	1.79	-1.79

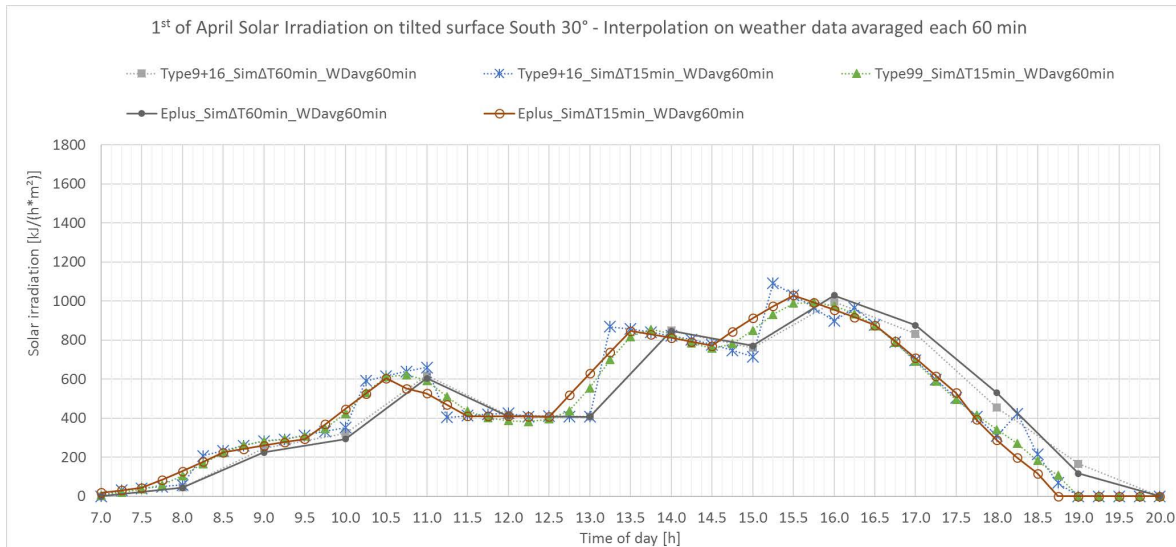


Fig. 2 – Effect of a different interpolation of 60 minutes averaged weather data for simulation with 15 min. global time steps

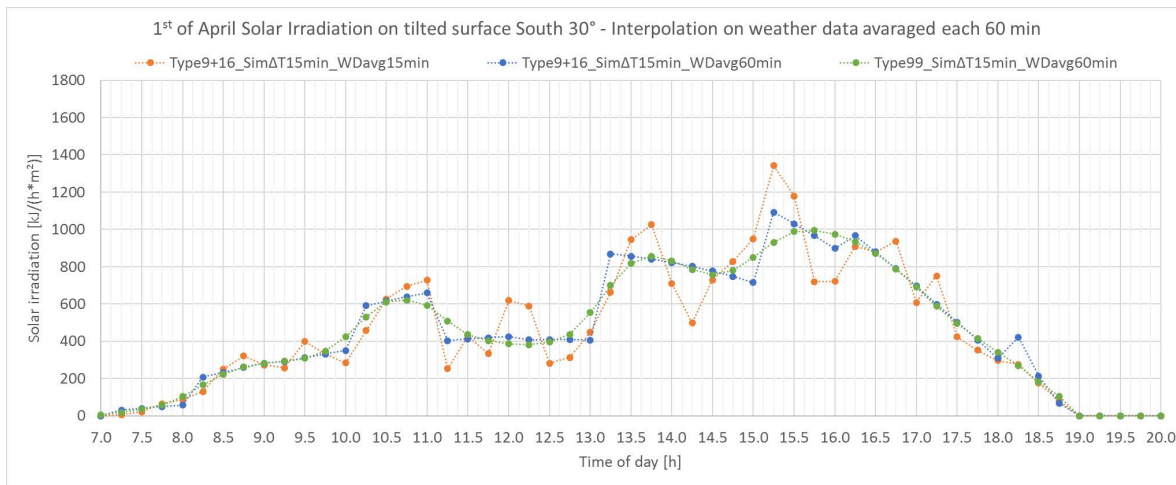


Fig. 3 – Comparing simulations with global time step of 15 min in case of weather data averaged over a time base of 15 min. or 60 min

Actually, with the exception of the west exposition, the two interpolation routines might be considered quite equivalent in terms of integral solar radiation calculation (Table 2).

However, if we want to check the performance of our solution against solar irradiance variability, the second non-smoothed interpolation might be preferable, since it is a bit more “discontinuous”.

Of course this second strategy is still not sufficient to emulate those random phenomena that occur in the atmosphere that are quite relevant when dealing with solar radiation. If we use the weather data file averaged on a 15-minute time basis, we have a different pattern of variability with respect to the interpolated ones starting from an hourly averaged weather data file (Fig. 3). Energy Plus does not allow

any choices and its linear interpolation has quite the same effect as the radiation smoothing of TRNSYS Type99.

While trying to ensure the conservation of energy received on a horizontal surface, a valuable algorithm for the estimation of those unknown values might consider the statistical variability of that specific variable. Variables such as solar radiation or wind velocity might be better estimated by applying a more realistic “pattern” to their “interpolation” than linear or smoothed regression. Statistics might be used to define a reduced number of patterns that might be applied to variables that show similar variability. Of course this apparently random behaviour should be deterministic and repeatable. Otherwise each time we repeat a simulation after having

changed a component to evaluate a different technology, we would get results that are non-comparable.

This statistical approach should have the goal to better assess the efficiency of a system and its control strategy as realistically as possible.

3. Discussion and Result Analysis

We have seen that different requirements might occur depending on:

- the sensitivity of the simulation components to input variability;
- the availability of weather data recorded at different time bases;
- the goal of the dynamic simulation (validation or evaluation at design time).

We have seen that in some cases it is necessary to describe in detail the variability of input data by using very short time step simulation (minutes), while other times, it might be sufficient to communicate to a simulation component an average value of the input required. In some cases, it might be better to align the different input data with precision, other times this synchronization might not be vital.

In the majority of the cases we have simulation components with very different time bases (heavy building construction and HVAC systems), but we do not want to use a very short time step simulation because it takes too much time.

How could we fulfil all the different requirements, in the best possible way?

A starting point might be to help the user to describe needs and input data as precisely as possible and let the tool choose what has to be done, in the most consistent way. For example, the user should not be asked to manipulate its weather data if the recorded time base is smaller than the global simulation time step.

After that, it might be better to reduce the required assumption. Since different kinds of uncertainties are already ingrained in numerical simulation, while defining the strategy to handle time-variant information inside the simulation, a relevant pursued goal should be to avoid assumptions not strictly needed, as interpolating values at mid time steps systematically.

The last but most important suggestion would be related to avoid information annihilation. If we have input data described on a short time basis, they should be kept available for whichever component might need them, without compelling all the other components as well to perform not "strictly needed" calculations. This can be managed by the singular component that should decide "by itself" if it needs to perform its calculation each time stamp or not.

4. Conclusion

The review here presented has been focused on the routines that handle weather data, as implemented by TRNSYS 17 and EnergyPlus 8.6.0. Our aim has been to point out how weather data provided on a certain time basis are used in the simulation, when the simulation time step is smaller, equal or larger than the weather data recording time basis.

When the simulation time step is the same as the weather data recording time basis, a clear architectural choice for the alignment of the different types of weather data and sun position is needed. While some possibilities are provided to the user to describe the available weather data, sun position timing is not modifiable. The more common strategy implemented by the analysed tools is to exchange the values of the simulation variable averaged over the time steps. However, this strategy prevents the numerical scheme implemented in each simulation component to directly handle the process of weighting its boundary conditions at different time stamps. A good architectural choice should preserve all the information available in input data, while managing rules should be defined and used only if strictly needed, as we have seen for solar radiation. When the simulation time step is smaller than the weather data recording time basis, interpolation is needed. We have seen that the currently available interpolation routines might not be "significant enough" to test complex components and control strategies. A stochastic "interpolation" algorithm, derived by statistical analysis on weather data fluctuations, might overcome this lack of information. This algorithm will have the purpose of mimicking the variability of variables with similar

capacity/patterns, to better evaluate the effectiveness of a system and of its control strategy when subjected to “realistic” boundary condition fluctuations.

References

- Crawley, D.B., J.W. Hand, L.K. Lawrie. 1999. “Improving the Weather Information Available to Simulation Programs”. In: *Proceedings of Building Simulation '99*. Kyoto, Japan: IBPSA.
- Duffie, J.A., W.A. Beckman. 2013. *Solar engineering of thermal processes*. Hoboken, U.S.A: John Wiley & Sons, Inc.
- Jarraud, M. 2008. *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8). Geneva, Switzerland: World Meteorological Organisation.
- Oglesby, M. 1998. Personal correspondence. Accessed the 30th of January 2016 <http://www.mail-archive.com/sundial@uni-koeln.de/msg01050.html>.
- Spencer, J.W. 1971. “Fourier Series Representation of the Position of the Sun”. *Search* 2(5): 172.
- Wilcox, S., W. Marion. 2008. *Users manual for TMY3 data sets*. Golden, U.S.A.: National Renewable Energy Laboratory.