Proceedings of BS2013:
13th Conference of International Building Performance Simulation Association, Chambéry, France, August 26-28

GENERATION OF WEATHER FILES USING RESAMPLING TECHNIQUES: AN EXPLORATORY STUDY

Parag Rastogi[1], Marilyne Andersen[1]
[1]Interdisciplinary Laboratoy of Performance-Integrated Design (LIPID),
École Polytechnique Fédérale de Lausanne,
Lausanne, Switzerland

ABSTRACT

Simulating a building to predict its performance over the course of a full year requires an accurate representation of the stable and representative weather patterns of a location, i.e. a weather file. While weather file providers give due consideration to the stochastic nature of weather data, simulation is currently deterministic in the sense that using one weather file always generates one performance outcome (for a given set of building parameters). Using a single time series or aggregated number makes further analysis and decision-making simpler, but this overstates the certainty of the result of a simulation. In this paper, we investigate the advantages and disadvantages of incorporating resampling in the overall simulation workflow by comparing commonly used weather files with synthetic files created by resampling the temperature time series from the same weather files. While previous studies have quantified uncertainty in building simulation by looking at the calculation itself, this paper proposes a way of generating multiple synthetic weather files to obtain better estimates of expected performance. As case studies, we examined the performance of the 'original' and synthetic files for each of a sample of world climates.

INTRODUCTION

Building simulation is a mixture of deterministic and stochastic inputs. The materials from which a building is constructed usually have known properties or behaviours and can, therefore, be counted as deterministic components. Variation of properties with external conditions, linear or non-linear, can mostly be modelled or measured to a reasonable level of accuracy. The energy/comfort performance of a given building is, however, also strongly influenced by weather conditions and occupant behaviour. Both of these are stochastic inputs, and it could be advantageous to treat them as such when predicting the indoor environmental performance of buildings, since they add a large amount of uncertainty to any predictions. In fact, Brohus et al. (2012) argue that it is impossible to have one truly representative number for the expected energy use of a building, and that trying to reduce the complex performance of a building to one number is too simplistic an approach. They suggest generating probability distributions or, at the very least, a mean value and standard deviation. This could, according to the authors, improve the value of simulation for design and provide a "scientific" basis for a "safety factor". de Wit and Augenbroe (2002) argued that, while experts do sometimes factor in their understanding of uncertainty into the design process, contemporary practice gives only passing attention to this issue in predictive simulation, if at all. They found that the potential of quantitative uncertainty analysis is of "virtually no concern" to the simulation community. This is still a problem, ten years after the publication of this article, as features to incorporate stochasticity are not found in popular simulation programs.

An understanding of uncertainty is especially important for the design of high performance buildings, which tend to be finely optimised for particular climatic (and other) inputs. Donn et al. (2012) argue that the performance of a finely tuned building could deviate drastically from predicted values if it is not used by "automatons who behave exactly as the simulation assumed they would". The same fears apply to a building that is optimised to one expected weather pattern. While it may not fail completely in the face of unexpected (especially extreme) weather, its actual performance could fall far short of expectations. An analysis of the influence of user behaviour on indoor performance is outside the scope of this paper, but literature on uncertainty in simulation due to weather files or climate alone is relatively sparse. Several articles have proposed analyses and quantifications of uncertainty in building simulation and design in general. For example, Lomas and Eppel (1992), Fürbringer and Roulet (1995), Lam and Hui (1996), de Wit and Augenbroe (2002), Macdonald (2002), Tian (2013), among others, analyse uncertainties due to inputs such as material properties, user behaviour, and modelling assumptions and simplifications.

de Wilde et al. (2008) examine the same parameters with potential climate change, adding an additional dimension to their analysis of uncertainty in performance projections. Jenkins et al. (2011) investigated the potential of overheating in dwellings in the UK by using probabilistic climate projections. They randomly extracted a hundred equally probable years from a hundred separate time series of thirty years duration each, which were generated from a future weather data-generating algorithm based on the UKCP'09 projections. They then ran a hundred yearly simulations to determine a regression equation that could predict overheating in dwellings without dynamic simulation. While this is not the aim of this study, the idea of incorporating multiple randomly selected files (i.e. time series) is similar. Brohus et al. (2012) proposed the use of stochastic differential equations for simulating the thermal performance of buildings. They modelled input loads as "stochastic processes, each comprising a time-varying mean value function and a stochastic part described as a white noise process scaled by a time-varying standard deviation function". Their methodology is able to incorporate stochasticity into the basic heat balance equation at each node, for both inputs as well as model coefficients (such as specific heat loss coefficients). They concluded that the "impact of a stochastic description compared with a deterministic description may be modest for the dynamic thermal behaviour of buildings", but they feel that it has more value in a model where the effect of air flow is more significant. A major drawback of their methodology is the very long computation times – almost a week for a year-long simulation of a simple room – which would increase significantly with each input modelled stochastically (e.g. wind loads, window opening regime, etc.).

This article is an exploration of the possibility of introducing stochasticity into the simulation workflow by the creation of synthetic weather files, rather than a fresh attempt at quantifying uncertainty in modelling. These synthetic weather files, generated with a technique known as bootstrapping, can be used to generate a range of possible values of a performance metric (e.g. energy use for HVAC systems) instead of a single value. Since the final performance of a building has many stochastic inputs (weather being one of them), a range of possible values (due to a range of different weather scenarios) offer a clearer view of likely performance than a single number. The authors have not, in this study, investigated the applicability of resampling-generated files to other sorts of studies like hygrothermal modelling or solar system sizing. Assuming that a study, linear or nonlinear, is carried out using a typical weather file in the first place, we see no reason why it cannot also be carried out using files created by resampling.

Resampling

Resampling is a technique used to create artificial data from existing samples in order to improve estimates of statistical quantities such as the mean or confidence intervals. The procedure begins with a small set of *iid* (independent identically distributed) points, i.e. one's original data. The original dataset is then sampled with replacement (i.e. the probability of selecting any given data point in the set resets after each draw) to create several new sets of 'data' (the resamples). The statistical quantities of interest are calculated for each new dataset, giving a range of values. If the mean value is of interest, for example, one is able to get a range of means (i.e. one mean of each new synthetic set). The shape of the distribution of this set of means is representative of the shape of the original dataset.

Weather Files

To approximate the prevailing weather patterns of a location, building simulation programs use 'typical' or 'design reference years' (DRY). Lund (1991, 1995) enumerated three basic requirements for a DRY to be useful, these being accurate representation of the true *frequencies*, *sequences*, and *correlations* of a climate. The first requirement ensures that a DRY represents true magnitudes of extremes and means. The second, that sequences of distinct values (episodes) recorded most often appear in the DRY. The third, that correlations between different meteorological time series (e.g. temperature, sun, humidity, etc.) are representative. In the report, Lund states that the second and third requirements will be impossible to fulfil satisfactorily. This, the report declares, is because "many of the relationships between individual parameters cannot yet be described mathematically". As part of further work in this direction, the authors intend to survey more literature to see if a satisfactory test for assessing the sequence and correlation requirements can be found. There are several different algorithms currently in use for generating DRY. The principal methods include frequentist methods, i.e. ones that select typical months based on frequency distributions; stochastic generation, which uses long-term means and distributions calculated from observed data to generate synthetic files; and principal component analysis, which separates out parameters based on their contribution to overall variation and selects months that explain best the variation seen in a climate.

In this study we have used two types of DRY: Typical Meteorological Years (TMY), distributed free by the US Department of Energy (Wilcox and Marion, 2008), and weather files generated by a proprietary software called METEONORM (Remund et al., 2012). TMY files are generated from mostly measured data, while METEONORM (Meteo) files are generated from observed distributions of different parameters scaled by long term 'normals' (averages). The implication in testing the resampled files against these two files is that if the resampled time series approximate patterns in measured data as accurately as these established files, then the case for using the resampled files is strong. If the differences in probability distributions between Meteo and TMY are as significant as the differences with the resampled files, there is no reason to suspect that the resampled files represent an unlikely climatic scenario (e.g. unlikely durations or intensity of high temperature episodes).

SIMULATION/EXPERIMENT

Case Study Selection

In this paper, we began with a Typical Meteorological Year file (TMY) and a Meteo-generated file (Meteo) for each of ten locations worldwide. These locations are broadly representative of world climates and population concentration. They were originally selected by Kleindienst et al. (2008) to represent a broad range of solar paths and annual daylighting conditions. They also happen to represent a range of climate types (see Table 1), though temperate/subtropical climate types are over-represented (ASHRAE climate type 4). This apparently arbitrary selection originally came about due to a preference for densely populated zones with sufficiently different solar paths and, for now, we have kept this preference. We did, however, change the original 'local' location (Boston) to Geneva, to represent continental Europe. We may tinker with this set in future work to represent as many significant divisions of the (updated) Koeppen-Geiger climate classification system as possible, while keeping the urban population density preference.

Table 1: Cities and their ASHRAE climate types.

CODE	CITY NAME	COUNTRY	TYPE
ADD	Addis Ababa	Ethiopia	3C
BGK	Bangkok	Thailand	4B
GEN	Geneva	Switzerland	4C
HAR	Harare	Zimbabwe	4A
HGK	Hong Kong	PR China	4A
LON	London	UK	4C
PHO	Phoenix	USA	4B
SIN	Singapore	Singapore	1A
STP	St Petersburg	Russia	6A
SYD	Sydney	Australia	4A

Pre-processing

Meteorological time series like dry bulb temperature (TDB) are not stationary *iid* series, and cannot therefore be resampled without some pre-processing. So, we detrended the time series using a localised polynomial fit. For example, Figures 1 and 2 show the original (TMY) and detrended temperature series for Geneva and Singapore respectively. The trend was calculated using the *smooth* function in MATLAB, with the *loess* method[1] and a span of one month. After removing the trend, we found that the remaining time series are stationary (checked using the KPSS test for stationarity developed by Kwiatkowski et al. (1992)), which makes them eligible for resampling. Upon creating the resampled time series (which are also stationary *iid* values), we re-added the original trend to return to a physically valid temperature time series with the proper annual trend.

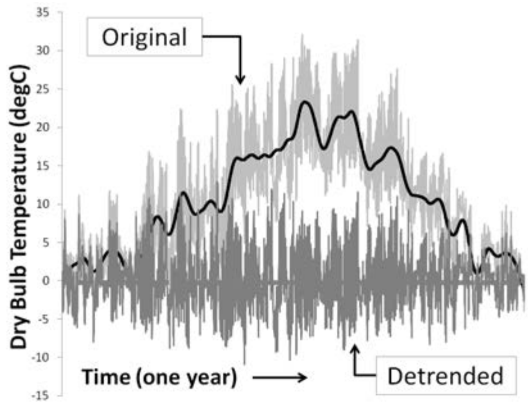

Figure 1: Temperature time series from the Geneva TMY file.

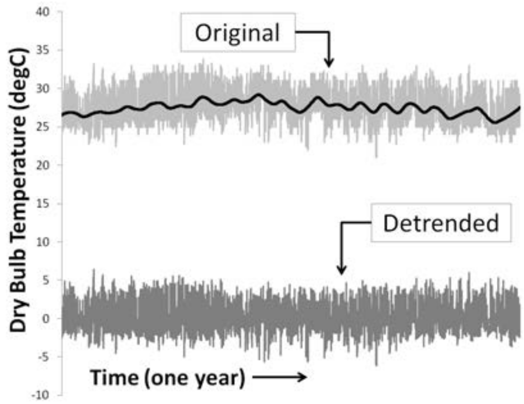

Figure 2: Temperature time series from the Singapore TMY file.

The trend lines in our examples above seem unusually tortuous, since we expect Geneva to have a *cosine* trend and Singapore a constant one. The question of picking an appropriate trend for this procedure is still not satisfactorily resolved for us. In our work, we have so far assumed a simple model consisting of a deterministic and stochastic term, as shown in Equation 1.

$$Y_t = \mu_t + X_t \tag{1}$$

Y_t represents the time series, μ_t is the deterministic trend, and X_t is the stochastic component. In our case, μ_t is a 2nd order polynomial with a period of twelve months, i.e. we expect the monthly means to be the same in subsequent years. The tortuousness of our trend-lines can be explained by our choice of span (also known as bandwidth). Textbooks on time series analysis do not seem to recommend a standard bandwidth, since it is highly dependent on the data set in question (Mudelsee, 2010; Cryer and Chan, 2008; Gluhovsky, 2011). Starting from four-month spans, we found that a span of one month was the largest bandwidth that would leave us a stationary time series after detrending. Whether this span-of-convenience is appropriate will be explored more thoroughly in future work, with more rigorous testing and case studies.

[1] "Local regression using weighted linear least squares and a 2nd degree polynomial model" (The Mathworks, Inc., MATLAB 2013a).

Resampling and Post-processing

For each climate, we resampled the dry bulb temperature time series from a source file (TMY in this case) to create ten synthetic datasets each, keeping all other meteorological parameters unchanged. We recalculated dew point temperature (TDP) based on the new temperature series and original humidity. The resampled dry bulb temperature time series were smoothed with a moving average filter using a span size of 4-24 hours, until their maximum slope was less than that of the TMY time series. This last step is necessary since random number draws are sometimes so large that successive values (in the synthetic temperature series obtained after re-adding the trend) are physically infeasible.

Building Simulation

To compare the performance of the resampled files with TMY and Meteo files, we simulated a residential building in DesignBuilder software (DesignBuilder Software Ltd, 2011). The building is a single family home in northern Germany, and was initially modelled with its current HVAC systems to verify the accuracy of our model against actual energy bills. Then, we removed auxiliary heating and cooling systems. This allowed us to compare the results from different files quickly and without any bias that may have been introduced due to the design and (non-linear) performance of an HVAC system. The same model was used for all simulations. Some details, assumptions, and characteristics of the model are given below.

1. All templates (construction, occupancy, etc.) were modified from the default templates available in DesignBuilder for the United Kingdom.
2. The site has no significant shading, and is located in a suburban context.
3. Construction
   - The house has a basement, a ground floor, and an attic below a sloped roof. Total occupied floor area is 236.8 m^2 and unoccupied area is 80.8 m^2.
   - Air changes per hour due to infiltration were set at 0.7.
   - Window to wall ratio is 30%. Windows are single-glazed, with clear 3mm panes.
   - U-value of external walls is 2.5 $W/m^2 - K$.
4. Activity and Indoor Environment
   - The house is occupied by a family of four (density 0.0196 $people/m^2$). Equipment density is 2.16 W/m^2 and lighting power density is 5 $W/m^2/100lx$.
   - The building is assumed to run at a common household schedule: not occupied during office/school hours, peak usage in the morning and evening, and occupied all day on weekends and holidays.
   - The heating and cooling set-points were initially set at 21°C and 28°C respectively, although the active systems were included only for initial verification. The ventilation system was left in place to maintain a flow rate of 0.3 air changes per hour.

Figure 3: A rendering of our building model.

RESULTS

Weather Files

We carried out two tests on each temperature time series obtained from resampling to confirm that they are physically valid and reasonably representative substitutions for the source DRY files (in this case, TMY files). The first was to check the correlation of the dry bulb and dew point temperature series of each resampled file and the Meteo file with the TMY file. Correlation is a measure of the 'closeness' in the patterns of two time series. We used the Spearman correlation coefficient (ρ), which takes values between +1 and -1. A value of +1 indicates perfect correlation, though not necessarily linear, while 0 indicates no correlation and -1 perfect anti-correlation.

For our second test, we compared Cumulative Distribution Functions (CDF). A CDF of a random variable is a measure of the frequency of occurrence of each value that the variable can take. The CDF of a variable x is given by $P(x \leq X)$, where X is the value at which the CDF is calculated. So a CDF of 0.85 for X = 13°C means that 85% of the values in a given dataset are less than or equal to 13°C. The TMY-generation algorithm uses the Finkelstein-Shafer (FS) statistic to compare two CDF (Wilcox and Marion, 2008). The FS is a measure of the absolute distance between two distributions (Finkelstein and Schafer, 1971). We used a similar normalised measure called the Normalised Mean Square Error (NMSE, Equation 2) for easier comparison between climates, since the magnitudes of the FS statistic varied significantly. A value of +1 for the NMSE indicates a perfect fit with reference data, 0 indicates that the fit is no better than a straight line, while large negative numbers indicate a bad fit. The NMSE in our test is calculated monthly, like in the TMY algorithm. This is important since the shapes of

CDF for a given climate could vary significantly by season.

$$NMSE(i) = 1 - \frac{|x(i) - xref(i)|}{|x(i) - mean(xref(i))|^2} \quad (2)$$

Figures 4 and 5 show the Spearman correlation coefficient (ρ) values from comparing the Meteo and resampled time series with TMY (the bar representing resampled values is the average from all 10 files). The artificial (i.e. resampled) temperature time series generally show a strong positive correlation with the original time series (i.e. TMY), in some cases even better than the Meteo file. The correlation of the resampled files in Addis Ababa, Bangkok, and Singapore is unacceptably weak (≤ 0.5). However, it should be noted that even Meteo does not compare well to the TMY series for these climates. This could imply that correlation is perhaps not the best comparison for a climate that has a relatively small seasonal trend (compare Figures 1 and 2). If the resampled files perform as badly on the next test for the same climates (i.e. those with small seasonal trend), it would call into question the robustness of this methodology.

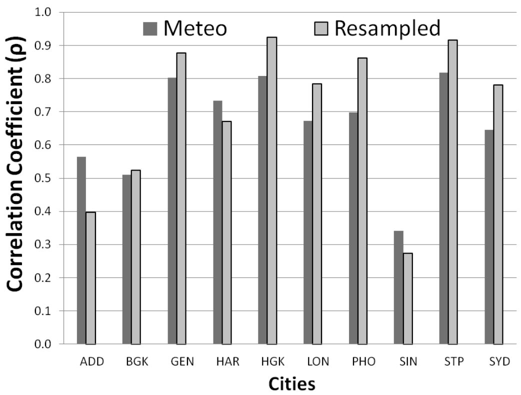

Figure 4: Overall correlation between TMY temperature series and others (Meteo and average resampled). See Table 1 for the names of the cities.

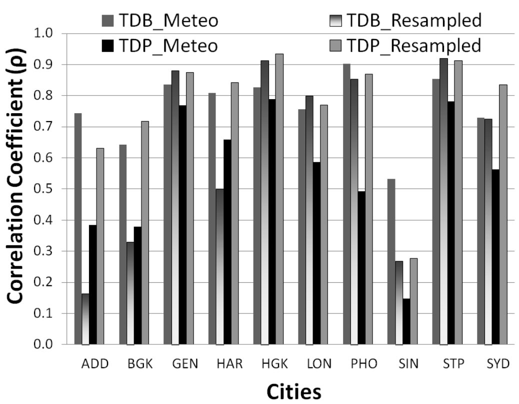

Figure 5: Correlation of dry bulb temperature (TDB) and dew point temperature (TDP) between TMY temperature series and others (Meteo and average resampled). See Table 1 for the names of the cities.

For the second test, we compared the CDF of the resampled time series, TMY, and Meteo with real hourly time series of dry bulb and dew point temperatures obtained from the NCDC (National Climatic Data Center, 2012) for the years 2000-11 (12 in total). The Meteo files we used are based on the latest data available in the software, while the TMY series also use the latest possible data available with the respective meteorological organisations. The reasoning behind comparing CDF is that neither the DRY files nor our resampled files are supposed to ever be able to replicate an entire year exactly. Rather, they should have the same distributions and patterns over a given season.

Figure 6 shows the NMSE-based comparison of the CDF. The NMSE for each month was calculated separately, and the 12 monthly values were averaged. The resampled files show an acceptable performance in this comparison, although the TMY and Meteo files are consistently better. The worst-performing locations are Hong Kong, Harare, and Geneva. Strangely, these three locations are among the best performers when compared using the correlation test above. These results taken together would imply that the resampled time series are 'shifted' from the original, i.e. they maintain the patterns of the original but not the magnitudes. Why this happens for these three climates only is not clear. They are all type 4 climates (ASHRAE, see Table 1), though their secondary label (A, B, or C) is not the same. London and Phoenix are also type 4 climates, and they show acceptable performance. That it is not an artefact of random number sampling is supported by the fact that the result stayed consistent across successive bootstrapping draws. It should be noted that the source files (TMY) for the under-performing locations do themselves compare well to the measured data, except for Hong Kong.

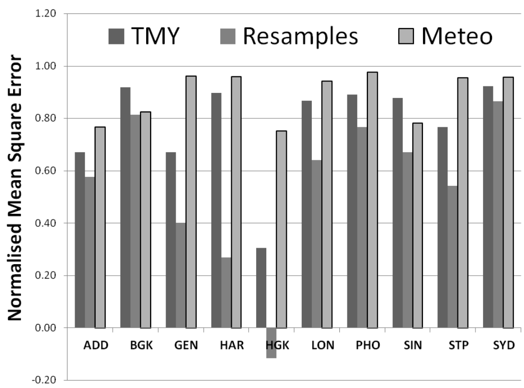

Figure 6: Normalised Mean Square Errors (NMSE) between Cumulative Distribution Functions (CDF) of weather files with reference to measured data.

Building Simulation

Since this is an exercise in examining the feasibility of using resampled weather files rather than testing the performance of the building itself, we report the results of energy simulation using a crude metric similar to heating/cooling degree days. We start with the hourly outdoor temperature and calculate the distance of each point from some arbitrary static comfort zone (15º - 20º C in this case). We then aggregate these hourly differences into one value per climate – the total number of degree-hours when a climate may require heating or cooling. We repeat the procedure for the indoor climate, which gives us an indoor value for aggregate degree-hours. Since the same building is used for simulating each file (in each climate), the difference between the indoor and outdoor degree-hour values (for a given climate) should be caused purely by the building (we call this the building's 'performance').

The mean performance value obtained by simulating the ten resampled files is compared (for each climate) with the values obtained from the Meteo and TMY files in Figure 7. The standard deviations for performance obtained from the set of resampled files tended to be less than 10% of the mean for all climates. The range of values is somewhat smaller than would be expected from a stochastic process (e.g. locations shown in Figure 8). This could be because the resampled files were smoothed before being used for energy simulation, causing them to resemble each other more than raw random samples would. Bangkok presents the only problematic case in this comparison. Given that it performed well in the NMSE test (i.e. the relative difference in hourly temperature values is not large), and mediocre in the correlation test (i.e. its episodic patterns are not well aligned), we see no reason why absolute (aggregated) performance should be so starkly different. We hope to test more climates of this sort in future work to verify whether this is a problem with our methodology or an outlier.

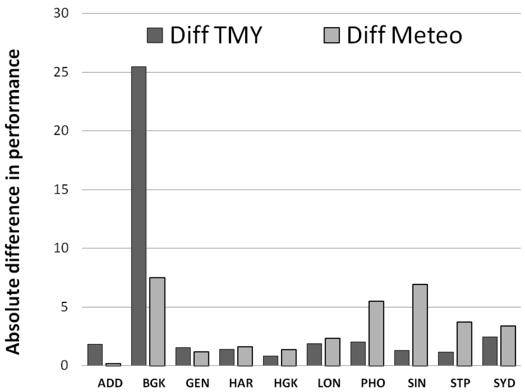

Figure 7: Difference in (averaged) performance values obtained from resampled files and the two reference files. 'Diff TMY' is with reference to the TMY file, and 'Diff Meteo' is with reference to the Meteo file.

The results presented here are for one 'realisation'. In a realisation, a random series of indices is used to pick the values that make up a given draw. This means that we obtain different datasets with successive runs of a resampling script. However, the random nature of this process ensures that the datasets (ten each) generated by each realisation have similar average performance. That is, the results were consistent across different bootstrap realisations.

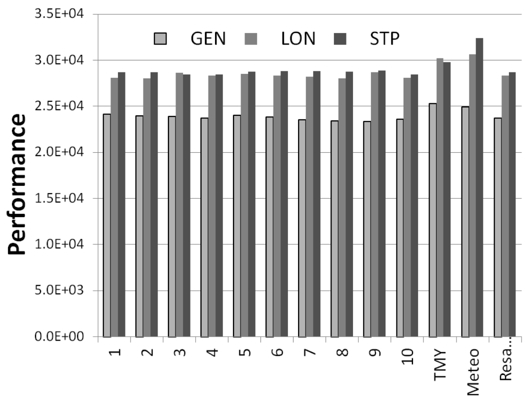

Figure 8: Performance values for Geneva, London, and St Petersburg. Results plotted are from all resampled values (for an arbitrary realisation), TMY, Meteo, and an average of all resampled files.

DISCUSSION

Our exploration of resampling to generate weather files for simulation generally resulted in physically viable files. This procedure does not, naturally, convey the underlying physical model. However, given the random nature of weather data and the difficulty of modelling it with closed-form models, we find that bootstrap draws from a source DRY file generate acceptable time series. Observing the mismatch of the TMY and Meteo files against measured data, we see that the Design Reference Years should not be treated as purely deterministic inputs. While they reflect historic climatic normals and patterns well, they are not meant to be exact representations of any given year a building might experience in its lifetime. While the performance obtained from the various resampled files did not vary significantly in this experiment, that in itself is not reason enough to disregard the potential usefulness of generating files from resampling. We only resampled temperature values, whereas a building's performance is also strongly influenced by solar radiation, humidity, and wind speed (with other meteorological factors such as infra-red radiation from the sky making smaller contributions). The issues raised by this initial study, and discussed below, point toward several avenues for improvement.

Climate Change

A CIBSE report on the use of climate change scenarios for building simulation says that "at present there is no accepted methodology for carrying out climate change risk assessments for the environmental design of buildings..." (Hacker et al., 2009). Shamash et al. (2012) reviewed available guidance on, and approaches to, using probabilistic climate projections in simulation in the building industry in the UK. The methods they discuss inevitably involve multiple simulations with extensive pre- and post-processing. The approach proposed in this paper is not an acceptable substitute for simulation with probabilistic climate projections in its current form, since it cannot recreate weather 'episodes' (e.g. heat waves), extreme events (e.g. overheating), and changes in climatic normals. Resampling, with proper adjustment and calibration (e.g. the smoothing carried out in this paper), is a good candidate for modelling a stochastic natural process. Successive realisations of this procedure could help generate patterns that better reflect the true random nature of weather. What resampling cannot do by itself, though, is generate new trends or create episodes. This means that its applicability to modelling probable climate change is limited in the raw form we explored here. Future work in this regard is expected to involve other modelling methods (e.g. ARMA, ARCH, etc.) to overcome these shortcomings. Detailed climatic models can predict future weather with some accuracy when given realistic starting conditions and inputs like emissions scenarios. They could also form a more physics-based basis for generating future weather files in combination with a random draw technique like resampling.

Post-processing

The resampling methodology used in this paper does not account for correlation between parameters. This correlation is difficult to model in principle, and is not always apparent even in complex general climate models. In any case, running climate models to generate files for building simulation, even if these models were well characterised and localised, would be virtually impossible with the computation power usually available for energy modelling. Correlation between parameters and auto-correlation of each parameter with itself are usually modelled based on measured data, akin to a CDF. Assuming that this resampling method would be used with only one short time record (e.g. a TMY file), developing correlation or auto-correlation functions on the fly could be difficult and unreliable. Future work involving the resampling of multiple parameters from one file (e.g. solar radiation, humidity, etc.) might need more post-processing to maintain physically valid correlations between the parameters. For example, checking of the correlation functions of the resampled files against those found in the source files. In addition to correlation, resampled files must be post-processed to respect the physical constraints of each individual meteorological parameter in question. For example, in this paper, we smoothed the resampled temperature datasets to have the same maximum slope as measured data (represented by TMY for now). Relative humidity, to take another example, cannot be less than 0% since that would be physically meaningless. We feel that the post-draw smoothing carried out in this procedure is justified, since that is a reflection of the tendency of temperature time series to have long memory (i.e. change slowly in time). A brute force approach to addressing both these problems could be to keep running realisations until a sufficiently large pool of acceptable time series is obtained. Diagnostic tests could include matching correlation and auto-correlation functions as well as physical validity with the original files and underlying physics.

Computational Issues

If resampling is deemed a satisfactory method of creating new weather files, the added computational cost naturally depends on the size and complexity of one's building model. Resampling itself contributes minimal extra computational load, although rerunning the energy simulations in *DesignBuilder* software multiple times was very time-consuming in our implementation. The extra effort involved in preparing multiple weather files manually and simulating them far exceeded the time for resampling and even for the actual simulation itself. However, this can be overcome by running energy simulation programs in 'batch mode' (multiple inputs and outputs with one command), which we were unable to do for this study. We feel that the added time needed for multiple simulations is not significant considering the potential advantages of getting a distribution of performance data (assuming weather file generation and energy modelling can be done in batches).

CONCLUSION

In this paper we presented an initial exploratory study for using resampling to generate multiple weather files for simulation. Based on our tests of correlation with, and deviation from, TMY and Meteo weather files and measured data, we found that resampling does not produce unrealistic time series for dry bulb and dew point temperature. This study was limited to temperature, though solar radiation, humidity, and wind speed are also potential candidates for resampling-based generation of synthetic time series. We found that resampling is not computationally-intensive in itself.

Given the nascency of this approach, and of probabilistic simulation in general, it is difficult to predict how this approach could be best incorporated in simulation. There are possibly several other sources of variation (like occupant behaviour) that may not be adequately described by equations. It seems unlikely that it would be feasible to run all likely variations in different inputs (e.g. climate, programmatic usage, etc.) in all possible combinations. A possible solution could be to demonstrate the robustness of one's design with several input files generated, and preliminary results described here indicate that resampled files are strong candidates for generating probabilistic inputs, especially if expected trends(e.g. UKCP'09 projections), episodes, and events are properly integrated. This would result in a 'building performance range', reflecting the range of outcomes that could be caused by any number of sources of variation. This would demarcate confidence intervals for building performance instead of relying on single output values. The range of values generated by this method could be used to improve predictions for individual buildings. They could also be used to calculate scientific "safety factors" or probabilistic performance criteria for common building types that designers may be required to demonstrate as having been fulfilled.

ACKNOWLEDGEMENT

We would like to acknowledge the guidance provided by Prof. A Gluhovsky from Purdue University in exploring statistical techniques. He provided the final push to take this project from idea to reality. We would also like to thank Soenke F. Horn for allowing us to use his energy simulation models as a basis for our analyses. Discussions with Soenke were invaluable in getting out of software- and procedure-related jams.

REFERENCES

Brohus, H., Frier, C., Heiselberg, P., and Haghighat, F. 2012. Quantification of uncertainty in predicting building energy consumption: A stochastic approach. *Energy and Buildings*, 55:127–140.

Cryer, J. and Chan, K. 2008. *Time Series Analysis: With Applications in R*. Springer Texts in Statistics. Springer.

de Wilde, P., Rafiq, Y., and Beck, M. 2008. Uncertainties in predicting the impact of climate change on thermal performance of domestic buildings in the UK. *Building Service Engineering Research and Technology*, 29(1):7–26.

de Wit, S. and Augenbroe, G. 2002. Analysis of uncertainty in building design evaluations and its implications. *Energy and Buildings*, 34(9):951–958.

DesignBuilder Software Ltd 2011. DesignBuilder Printable Documentation v3.0.

Donn, M., Selkowitz, S., and Bordass, B. 2012. The building performance sketch. *Building Research & Information*, 40(2):37–41.

Finkelstein, J. and Schafer, R. 1971. Improved goodness-of-fit tests. *Biometrika*, 58(3):641–645.

Fürbringer, J. M. and Roulet, C. A. 1995. Comparison and combination of factorial and Monte-Carlo design in sensitivity analysis. *Building and environment*, 30(4).

Gluhovsky, A. 2011. Statistical inference from atmospheric time series : detecting trends and coherent structures. *Nonlinear Processes in Geophysics*, 18(4):537–544.

Hacker, J., Capon, R., and Mylona, A. 2009. *Use of climate change scenarios for building simulation: the CIBSE future weather years*. CIBSE Chartered Institute of Building Services Eng.

Jenkins, D. P., Patidar, S., Banfill, P. F. G., and Gibson, G. J. 2011. Probabilistic climate projections with dynamic building simulation: Predicting overheating in dwellings. *Energy and Buildings*, 43(7):1723–1731.

Kleindienst, S., Bodart, M., and Andersen, M. 2008. Graphical representation of climate-based daylight performance to support architectural design. *Leukos*, 5(1):1–28.

Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54:159–178.

Lam, J. C. and Hui, S. 1996. Sensitivity analysis of energy performance of office buildings. *Building and Environment*, 31(I):27–39.

Lomas, K. J. and Eppel, H. 1992. Sensitivity analysis techniques for building thermal simulation programs. *Energy and Buildings*, 19(1):21–44.

Lund, H. 1991. The Design Reference Year. In *Proceedings of Building Simulation 1991*, pages 600–606, Nice, France. IBPSA.

Lund, H. 1995. The Design Reference Year User Manual. Technical report, Technical University of Denmark, Copenhagen, Denmark.

Macdonald, I. A. 2002. *Quantifying the Effects of Uncertainty in Building Simulation*. Phd, University of Strathclyde.

Mudelsee, M. 2010. *Climate Time Series Analysis*, volume 42 of *Atmospheric and Oceanographic Sciences Library*. Springer Netherlands, Dordrecht.

National Climatic Data Center 2012. Climate Data Records.

Remund, J., Mueller, S., Kunz, S., and Schilter, C. 2012. METEONORM Handbook Part II : Theory.

Shamash, M., Mylona, A., and Metcalf, G. 2012. What Guidance Will Building Modellers Require For Integrating. In Wright, J. and Cook, M., editors, *First Building Simulation and Optimization Conference, IBPSA-England*, number September, pages 253–260, Loughborough, UK. IBPSA-England.

The MathWorks Inc. 2013. MATLAB.

Tian, W. 2013. A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews*, 20:411–419.

Wilcox, S. and Marion, W. 2008. Users Manual for TMY3 Data Sets. Technical report May, National Renewable Energy Laboratory.