

Random variation and correlation of the weather data series – evaluation and simulation using bounded histograms

Petr Konecny^{1,*} and *David Pustka*²

¹VSB – Technical University of Ostrava, Faculty of Civil Engineering, Department of Structural Mechanics, Ludvika Podeste 1875/17, 708 33 Ostrava-Poruba, Czech Republic

²VSB – Technical University of Ostrava, Faculty of Civil Engineering, Department of Structures, Ludvika Podeste 1875/17, 708 33 Ostrava-Poruba, Czech Republic

Abstract. The contribution is focused on the random variation and correlation of input parameters for the climate data description. The climate data such as ambient temperature, solar intensity, wind speed and direction are of the random nature. The description of ambient temperature can be based on the climate data time series in the form of climatic “load duration curve”. Particular input parameters such as ambient temperature and solar radiation have a significant correlation. The attention is paid especially to the evaluation and simulation application of the histograms using Monte Carlo type process considering correlation of particular parameters.

1 Introduction

Numerical modelling of the wind load, thermal expansion related problems or building thermal balance is affected by the variability of input parameters such as ambient temperature, wind speed, direction and solar radiation.

The parameters for the analysis of the effect of thermal changes-induced deformation on structures such as bridges may benefit from data about the distribution of temperature since the design to most unlikely conditions is frequently too conservative. There are also applications when the wind speed and wind direction plays their role in the quasi static analysis. Then description of the variation of these parameters as well as correlation of wind speed and direction may be of interest in such cases, especially in cases of unsymmetrical objects.

Also the field of the energy performance of the building may benefit from description of variation of input parameters and their correlations. Thermal balance is generally based on the comparison of thermal losses with the capability of heating/cooling system. The system is frequently designed to meet the most demanding conditions. Those most demanding conditions can be rare, thus and design of a single system such as heat pump without supplementary heat source can be expensive. Typical analysis is based either on the deterministic evaluation of the performance in case of extreme summer and winter

* Corresponding author: petr.konecny@vsb.cz

temperatures or on the time-dependent analysis based on the typical year data in case of non-stationary processes (e. g. ground source heat pumps application [1]).

Since the input parameters are of random nature application of probabilistic reliability analysis can give an idea what is the probability that designed structural or thermal system does not meet the design condition with acceptable likelihood. The probabilistic assessment may be conducted using the probabilistic assessment [2-4]. The correlation between random inputs characteristics can be statistically described as well [5-8]. The suitable probabilistic approach in case of time-dependent parameters such as temperature and snowfall can be well described by time pulses (see e. g. [9]) or load-duration curves [2]. It may be worth explanation that load duration curve is a sorted history of loading formed as a frequency histogram.

Lot of such analyses often use the direct Monte Carlo method ([11], [2-4], [12-14]) which is robust, although it may be time consuming. Even though, enhanced probabilistic procedures that use computer power more effectively such as Latin Hypercube Sampling [15], Response Surface Method [16] or Direct Optimized Probabilistic Calculation [17], are not discussed herein. The paper is focused on the analysis of climate random input parameters variability and their mutual correlation necessary for the probabilistic building thermal losses analysis or thermal expansion related problems. This pilot study is based on the weather data time-series available from [7] that cover "so called" typical year.

2 Analysis of data

Time series of weather data hourly records that represents typical year in Ostrava are statistically analysed for mean value, standard deviation, minimum and maximum value. The frequency histograms are created as well. The relationship between the parameters is studied using the correlation index. Strong linear correlation is expressed if the correlation index approaches to value one either positive or negative. Variables are uncorrelated if the correlation index approaches to zero value. Since the weather data are seasonal with variation throughout a day the two time flow-related parameters are added. Parameter *time* represents hours from the beginning of the year while parameter *hours* represents hours from the beginning of the day.

3 Variation of climate data

The resulting statistical parameters are given in Table 1 and frequency histograms are shown in the Fig. 1-3. Statistics for the time parameters are not included in the Table 1.

Relationship between studied parameters expressed through correlation coefficient is given in Table 2. Graphical description of selected correlated parameters is shown in 2D plots Fig. 4 and 5.

Histogram of time shows some irregularities due to the number of days within even and odd months throughout the year (Fig. 1 on the left). This can be further refined. Histogram of the temperature has two peaks caused by the winter and summer (Fig. 1 on the right). Average value of temperature is 8.5°C. Histograms of extraterrestrial and global radiation (Fig. 2) show high occurrence of zero value that is caused by the night measurements with average value of 287 W/m² for extraterrestrial and 116 W/m² for global radiation. Maximum values are 1181 W/m² and 890 W/m² respectively. Wind direction is dominating from south west and from north (see Fig. 3 on the left). Wind speed average is 3.8 m/s with extreme value 19 m/s.

The correlation index between the time and temperature (*temp*) is almost zero even though there is a strong seasonal dependence as depicted at Fig. 4. The reason for not

capturing the relationship between time and temperature is its periodic behaviour. Correlation index between radiation and temperature (*temp* is about 0.5 which is significant - see Fig. 4). Correlation between radiation extraterrestrial horizontal (*radia_ah*) and radiation global horizontal (*radia_gh*) around 0.9 illustrates strong influence of extraterrestrial radiation on the global radiation (see Fig. 5 on the left). Small correlation was found between wind speed and direction and between wind speed and extraterrestrial radiation (*radia_ah*).

Table 1. Statistical parameters of typical year weather data from Ostrava [7].

Description	Notation	Units	Mean	Std	Variation	Min	Max
Dry-bulb temperature	<i>temp</i>	[°C]	8.5	8.5	1.0	-14.1	32.9
Radiation extraterrestrial horizontal	<i>radia_ah</i>	[W/m ²]	286.8	369.0	1.3	0.0	1181.0
Radiation global horizontal	<i>radia_gh</i>	[W/m ²]	116.3	190.7	1.6	0.0	890.0
Wind direction	<i>wind_d</i>	[°]	165.8	104.9	0.6	0.0	360.0
Wind speed	<i>wind_s</i>	[m/s]	3.8	2.5	0.7	0.0	19.0

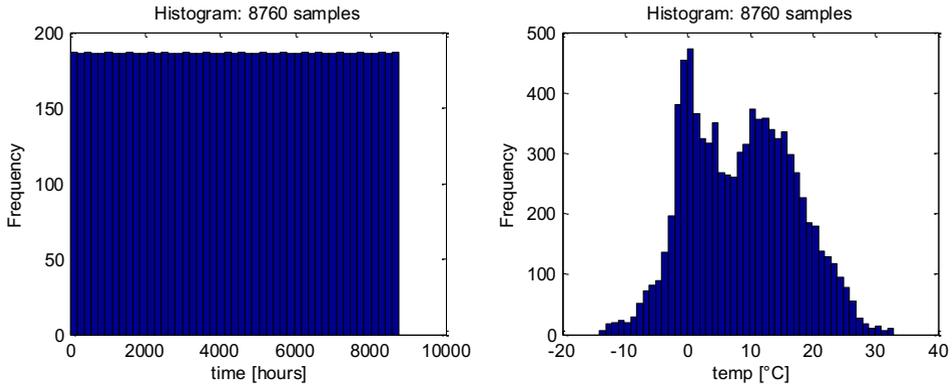


Fig. 1. Histograms of time /on the left/ and dry-bulb temperature [°C] /on the right/.

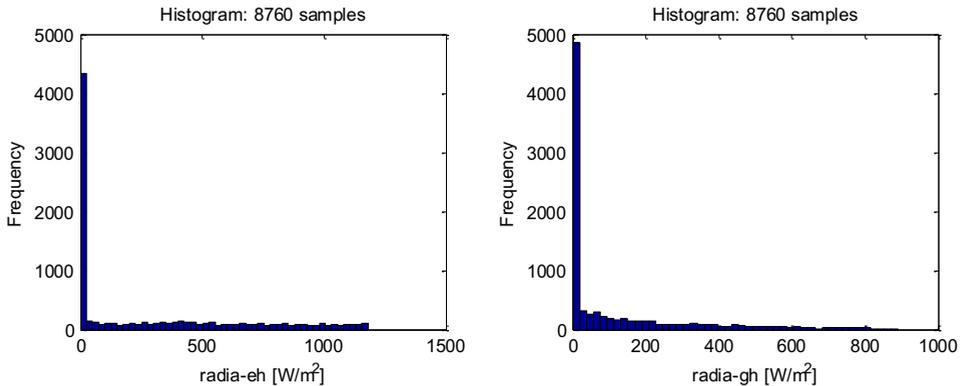


Fig. 2. Histograms of radiation extraterrestrial horizontal [W/m²] /on the left/ and radiation global horizontal [W/m²] /on the right/.

Table 2. Correlation indexes for parameters of typical year weather data from Ostrava [7].

	<i>time</i>	<i>hours</i>	<i>temp</i>	<i>radia_gh</i>	<i>radia_eh</i>	<i>wind_d</i>	<i>wind_s</i>
<i>time</i>	1.00	0.00	0.17	-0.05	-0.05	-0.07	-0.14
<i>hours</i>	0.00	1.00	0.10	-0.02	-0.02	0.03	0.03
<i>temp</i>	0.17	0.10	1.00	0.51	0.53	0.11	0.03
<i>radia_gh</i>	-0.05	-0.02	0.51	1.00	0.88	0.08	0.24
<i>radia_eh</i>	-0.05	-0.02	0.53	0.88	1.00	0.06	0.18
<i>wind_d</i>	-0.07	0.03	0.11	0.08	0.06	1.00	0.26
<i>wind_s</i>	-0.14	0.03	0.03	0.24	0.18	0.26	1.00

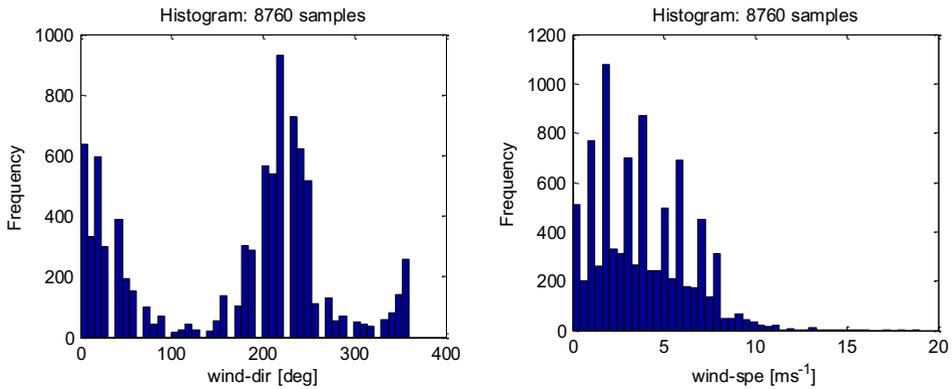


Fig. 3. Histograms of wind direction [°] /on the left/ and wind speed [m/s] /on the right/.

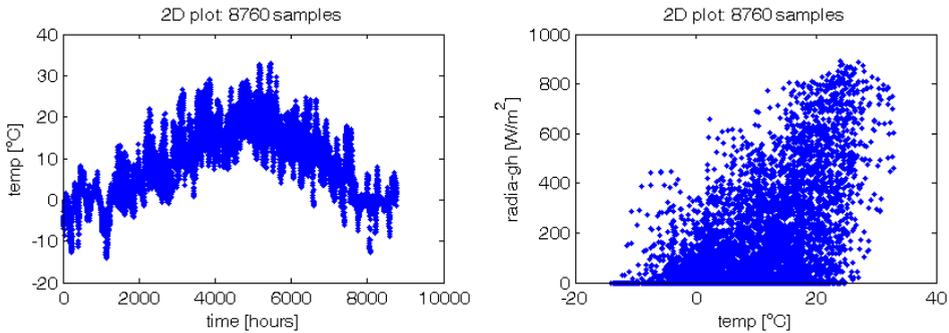


Fig. 4. 2D plot of time [hours] and dry-bulb temperature [°C] /on the left/; 2D plot of dry-bulb temperature [°C] and radiation global horizontal [W/m²] /on the right/.

4 Monte Carlo simulation

Since the probabilistic simulation of observed climate phenomena is of interest considering their mutual correlation, the model for generation of correlated distributions is applied with 50 thousands simulation steps as described by Phoon et al. [6, 10] and applied also in [8]. The advantage of this procedure is direct generation of random occurrences without dependence on previous simulations. This approach is suitable for the direct Monte Carlo method. The disadvantage of this approach is the ambiguous description of the resulting joint distribution despite the quality of the marginal distributions being maintained. The correlation coefficient of the generated distributions is close to the specified value.

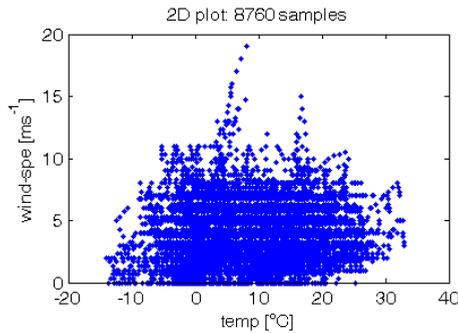


Fig. 5. 2D plot of temperature [°C] (horizontal axis) and wind speed [m/s] (vertical axis).

The simulation process using the Monte Carlo method can be enhanced using knowledge of the marginal distributions and the correlation between the parameters. The correlation matrix is modified to better preserve fractile correlation.

The first step in the simulation is to generate independent variables described by the normalized Gaussian distribution, which is adjusted by the Choleski decomposition into the correlated unit Gaussian distribution. The normal (Gaussian) distribution can be converted to a uniform distribution through its inverse transformation of the distribution function. The correlated uniform distribution is equivalent to the primary generator of random variables described by a frequency histogram.

Simulation steps are following:

- **Transformation of correlation matrix.** Due to differences in the normal distribution and the general distribution characterized, for example, by a histogram, it is appropriate to modify the obtained correlation index. The resulting correlation matrix R must satisfy the condition of positive definiteness to generate vectors of the correlated normal distribution.
- **Generation of correlated normal distributions.** The correlated Gaussian distributions may be, according to [5], [10] generated using either of two methods – Eigen decomposition, or Choleski factorization.
- **Transformation of a normal distribution into a uniform distribution.** Two generated occurrences of a normal distribution with a suitable correlation need to be transformed into a uniform distribution Y_F , using the inverse distribution function for the normal distribution.
- The probability density obtained by orthogonal transformation of a Gaussian random occurrence is a uniform distribution. It is suitable for generation of samples described by the frequency histogram.

Transformation of a uniform distribution into a general distribution (non-Gaussian process). A relevant general distribution is consequently obtained using the inverse transformation of the distribution function of the sought distribution.

- The probability density obtained by orthogonal transformation of a Gaussian random occurrence is a uniform distribution. It is suitable for generation of samples described by the frequency histogram.
- Transformation of a uniform distribution into a general distribution (non-Gaussian process). A relevant general distribution is consequently obtained using the inverse transformation of the distribution function of the sought distribution.

Since the resulting distribution follows the same marginal function described by bounded histogram (cumulative density function, CDF) as shown in [8]. It is not necessary to show the plots. The fractile correlation of respective CDFs that is more informative is shown in Table 3.

All parameters except of time show very good agreement with fractile correlation almost 1 (complete match). In case of the time, the difference is quite high here even though resulting histogram is approaching the original one on Fig 1 (on the left).

Table 3. Fractile correlation between original and simulated histograms.

<i>time</i>	-0.0890
<i>hours</i>	0.9994
<i>temp</i>	0.9991
<i>radia-eh</i>	1.0000
<i>radia-gh</i>	1.0000
<i>wind-dir</i>	0.9999
<i>wind-spe</i>	0.9995

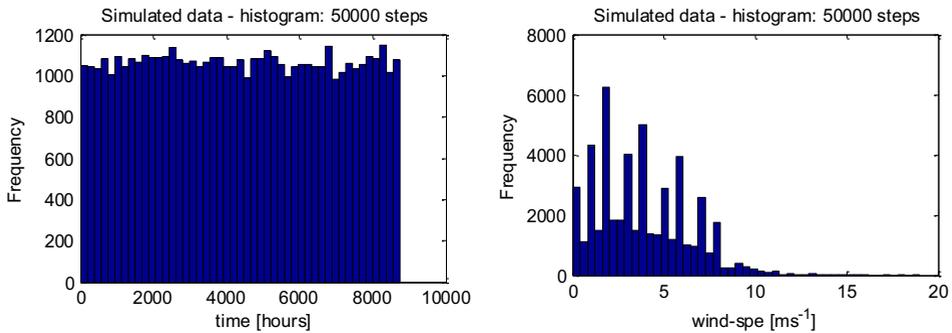


Fig. 6. Histograms of time [hours] /on the left/ and wind speed [m/s] /on the right/; both 50,000 simulations.

To verify the simulation process, the correlation matrix of randomly generated correlated occurrences is subsequently verified and compared with the original matrix. The Pearson's correlation coefficient of simulated correlation matrix is shown in Table 4.

Table 4. Correlation matrix of parameters simulated by Monte Carlo simulation.

	<i>time</i>	<i>hours</i>	<i>temp</i>	<i>radia_eh</i>	<i>radia_gh</i>	<i>wind_d</i>	<i>wind_s</i>
<i>time</i>	1.00	0.00	0.17	-0.04	-0.04	-0.07	-0.14
<i>hours</i>	0.00	1.00	0.10	-0.02	-0.02	0.03	0.03
<i>temp</i>	0.17	0.10	1.00	0.47	0.46	0.11	0.03
<i>radia_eh</i>	-0.04	-0.02	0.47	1.00	0.82	0.07	0.22
<i>radia_gh</i>	-0.04	-0.02	0.46	0.82	1.00	0.04	0.15
<i>wind_d</i>	-0.07	0.03	0.11	0.07	0.04	1.00	0.25
<i>wind_s</i>	-0.14	0.03	0.03	0.22	0.15	0.25	1.00

The resulting correlation indexes after simulation are similar to original ones as shows Table 4. The highest difference is in case of solar radiations and temperature as shows Table 5.

Table 5. Difference between original and simulated correlation matrices.

	<i>time</i>	<i>hours</i>	<i>temp</i>	<i>radia_gh</i>	<i>radia_eh</i>	<i>wind_d</i>	<i>wind_s</i>
<i>time</i>	0.00	0.00	0.00	-0.01	-0.01	0.00	0.00
<i>hours</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>temp</i>	0.00	0.00	0.00	0.04	0.08	0.00	0.00
<i>radia_eh</i>	-0.01	0.00	0.04	0.00	0.07	0.01	0.02
<i>radia_gh</i>	-0.01	0.00	0.08	0.07	0.00	0.01	0.02
<i>wind_d</i>	0.00	0.00	0.00	0.01	0.01	0.00	0.01
<i>wind_s</i>	0.00	0.00	0.00	0.02	0.02	0.01	0.00

5 Conclusion

The paper was aimed at the initial analysis of the typical year weather data from Ostrava [18] with respect to application of probabilistic structural and thermal balance assessment. It shows statistical parameters and frequency histograms of temperature, solar radiation, wind direction and speed. Correlation index between studied parameters and time is examined as well.

Significant correlation between solar radiation and temperature that would have to be considered in the probabilistic assessment is observed. Global radiation and extraterrestrial radiation are correlated very well. Correlation between time and temperature was not found due to the applied linear correlation coefficient. Suitable transformation shall be applied to adjust for the seasonal changes in temperature in order to simulate the relationship between season and the temperature in case of non-stationary processes such as ground source heat pumps application.

The analysis of typical year can illustrate the behaviour of selected parameters. There is a difference between typical year data used for the presented analysis and actual data sets with a decades of data. Storing a data in the form of histograms and correlation indexes can help to cover longer period of time with lower amount of storage space required. The suitable correlation field process shall be used in order to reproduce also the time history data variation related to seasonal changes relationship with sufficient quality and feasibility.

Moreover testing Monte Carlo simulation process is demonstrated. The process of simulation consists of preparation of a set of correlated normally distributed vectors. These vectors are transformed into two correlated uniform distributions that are used to produce a general distribution (a histogram). The histograms are created by a standard procedure, i.e., by the inverse transformation of a uniform distribution.

Good agreement between the input data and the results generated using the Monte Carlo method was observed. The correlation coefficient of the simulated values approached the required degree of correlation, and the prescribed marginal distributions were almost identical to the simulated distributions.

Financial support from VŠB-Technical University of Ostrava by means of the Czech Ministry of Education, Youth and Sports through the Institutional support for conceptual development of science, research and innovations for the year 2017 is gratefully acknowledged.

References

1. J.D. Spitler. Ground-source heat pump system research - Past, present, and future, HVAC&R Research, Vol: **11**, 2, 165-167 (2005)
2. P. Marek, M. Guštar, T. Anagnos, *Simulation-Based Reliability Assessment for Structural Engineers* (CRC Press, Inc., Boca Raton, Florida, 1995)
3. R.E. Melchers, *Structural reliability analysis and prediction* (2nd ed., John Wiley, Chichester, New York, 1999)
4. P. Marek, J. Brozzetti, M. Guštar, P. Tikalsky (ed.), *Probabilistic Assessment of Structures using Monte Carlo simulation. Basics, Exercises, Software*, (2nd edition, CD ROM attached, ITAM Czech Academy of Sciences, Prague, 2003)
5. M. Grigoriu, Simulation of non-Gaussian translation processes. J. Engng. Mech, ASCE, **124**, 2, 121-6 (1998).
6. K.K. Phoon, S.T. Quek, H. Huang, Simulation of non-Gaussian processes using fractile correlation, Probabilist. Eng. Mech., **19**, 287-292 (2004)
7. M. Vořechovský, D. Novák, Correlation control in small-sample Monte Carlo type simulations in a simulated annealing approach, Probabilist., Eng., Mech. **24**, 3, 452-462 (2009)
8. P. Konečný, Correlation of Steel Yield Stress and Ultimate Strength – Monte Carlo Simulation via Non-Gaussian Random Variables, Mater. Sci. Forum. (to be published in 2018)
9. M. Sýkora, *Probabilistic Analysis of Time-Variant Structural Reliability* (Doctoral thesis, CTU Prague, 2005)
10. K.K. Phoon, S.P. Huang, S.T. Quek, Simulation of second-order processes using Karhunen - Loeve expansion, Comput. Struct., **80**, 12, 1049-60 (2002)
11. N. Metropolis, S. Ulam, S. The Monte Carlo Method. Journal of American Statistical Association, **44**, 247, 335-341 (1945)
12. H.J. Pradlwarter, G.I. Schuëller, Assessment of low probability events of dynamical systems by controlled Monte Carlo simulation, Probabilist. Eng. Mech. **14**, 3, 213-227 (1999)
13. P. Konečný, J. Brožovský, P. Ghosh, Evaluation of Chloride Influence on the Cracking in Reinforced Concrete Using Korozeeneck Software, Transactions of the VŠB – Technical University of Ostrava, Civil Engineering Series, **11**, 1, 1-7 (2011)
14. P.S. Marsh, D.M. Frangopol, Reinforced concrete bridge deck reliability model incorporating temporal and spatial variations of probabilistic corrosion rate sensor data, Reliab. Eng. Syst. Saf. **93**, 3, 394-409 (2008)
15. R.L. Iman, W.J. Conover, A Distribution-Free Approach to Inducing Rank Correlation among Input Variables. Communications in Statistics Part B-Simulation and Computation, **11**, 3, 311-334 (1982)
16. Kralik, A RSM Method for Probabilistic Nonlinear Analysis of Reinforced Concrete Bubbler Tower Structure Integrity. In: *Proceedings of European Safety and Reliability Conference (ESREL)*, Reliability, Risk and Safety: Theory and Applications, 2010, 1-3, 1369-1372. Taylor & Francis (2009)
17. M. Krejsa, P. Janas, R. Čajka, Using DOProC Method in Structural Reliability Assessment, Appl. Mech. Mater. **300-301**, 860-869 (2013)
18. EnergyPlus Energy Simulation Software available on-line at: <http://apps1.eere.energy.gov/buildings/energyplus/> (2013)