Stochastic Modeling of Indoor Air Temperature

Joanna Janczura · Monika Maciejewska ·
Andrzej Szczurek · Agnieszka Wyłomańska

Abstract Temperature is one of the main parameters describing thermal comfort and indoor air quality. In this paper we propose an approach, based on a modification of the continuous time random walk, to model the indoor air temperature. We perform a statistical analysis of the recorded time series, that allows us to point out the main statistical properties of the recorded variable. The obtained conclusions about the nature of the process lead to a continuous time random walk, that in contrast to the classical approach, models time dependence of the jumps distribution. Moreover, we show that the waiting times can be modeled by a tempered stable distribution, which yields a subdiffusive behavior in short times and diffusive behavior in longer times. Finally, by conducting a simulation study we illustrate possible applications of the presented approach in the thermal comfort monitoring and forecasting.

1 Introduction

Urban residents in industrialized countries typically spend 70–90 % of their time indoors. The occupant appreciates the indoor environment mainly by its air quality, thermal, acous-

J. Janczura (✉) · A. Wyłomańska
Hugo Steinhaus Center, Institute of Mathematics and Computer Science, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 53-007 Wrocław, Poland
e-mail: joanna.janczura@pwr.wroc.pl

A. Wyłomańska
e-mail: agnieszka.wylomanska@pwr.wroc.pl

M. Maciejewska · A. Szczurek
Institute of Air Conditioning and District Heating, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 53-007 Wrocław, Poland

M. Maciejewska
e-mail: monika.maciejewska@pwr.wroc.pl

A. Szczurek
e-mail: andrzej.szczurek@pwr.wroc.pl

tic and visual comfort. A widely accepted definition of thermal comfort is that *it is a state of mind that expresses satisfaction with the thermal environment* [ASHRAE 55-2004, EN ISO 7730]. Thermal comfort is a psychological phenomenon, subjective feeling affected by the physical laws of heat transfer.

Thermal comfort is influenced by numerous factors. They are usually divided in two groups: environmental and personal. The first group includes air temperature, radiant temperature, air velocity and humidity. The most important personal factors contributing to thermal comfort are: age, gender and health status of occupant, type of clothing (particularly its insulation), activity level (i.e. amount of physical work done) and human physiology resulting in metabolic heat. These factors may be independent of each other, but together they contribute to a thermal comfort.

Thermal environment has been the subject of much international research over many years because the occupants of the buildings frequently suffer from thermal discomfort and this is not a minor problem. Maintaining comfortable thermal conditions inside building is important for health and quality of life of its occupants, appropriate energy consumption profile, as well as ensuring optimum work performance (extremes in air temperature may have adverse effects on productivity). Although a number of modeling approaches exist which help in studying the problem [1], the individual differences in preferences for thermal conditions cause that achieving acceptable comfort level for all occupants is the difficult, if not an impossible, task.

The assessment of thermal environment is often derived for steady state conditions or for minor fluctuations of one or more of the variables [2]. However thermal conditions inside rooms are seldom steady, due to the interaction between building structure, meteorology, occupancy, and heat, ventilation and air conditioning (HVAC) system [3]. The thermal conditions vary throughout a space and also with time (during the day, night and seasons). Hence the real time indoor environment monitoring system based on continuous measurements of selected physical parameters has vast prospects for human thermal comfort assessment.

Temporal variability of temperature causes that the assessment of this parameter based on a single or several periodic measurements may be unrepresentative. Therefore continuous monitoring and statistical analysis of measurement results is preferred [4]. The statistical description of transient states is based on an analysis of time series [5]. The effectiveness of this method depends upon sample sizes, averaging times, and sampling strategies. For example, the statistical description of dynamic conditions can be based on short or long-time average values. Unfortunately, the long-time average value gives little or no information on the thermal episodes that occur infrequently. Measurements that are averaged over shorter periods allow for recognizing these episodes. However, they reflect temporary conditions that are not representative of longer time periods. Therefore, short-term measurements may cause large uncertainties, potentially leading to incorrect conclusions, unsatisfactory performance of HVAC systems, or unnecessary costs [6].

Considering the shortcomings of traditional short-term measurement strategy, we proposed another approach using longer measurement periods and statistical data analysis. In this work, we want to show that a stochastic model of time series of temperature is a source of information about the state of indoor air. This methodology is a continuation of the analysis of indoor air quality time series presented in [7], where a subdiffusive process was analyzed in the context of experimentally recorded data. In this paper we extend the analysis presented in [7] and, based on statistical analysis of the observed time series, propose to use a model that apart from the observed constant time periods accounts also for the seasonal behavior of the underlying time series. The model is based on a modification of the continuous time random walk scheme.

The standard random walk describes a motion of a particle that in each time step makes a jump randomly in one of the possible directions. However, both the random walk and it's continuous time limit—Brownian motion, being a milestone of the statistical thermodynamics, are not capable of explaining a transport phenomena in complex systems such as glasses, liquid crystals, polymers, living cells or ecosystems. In contrast, in the continuous random walk (CTRW) scheme, originally introduced in [8], the waiting times between the jumps are not constant, as in the standard random walk, but are random variables governed by some probability law. As a consequence, the CTRW model seems to be a natural description of transport in crowded environments and complex systems. Limit distribution of the CTRW scheme can be also formulated within the framework of the celebrated fractional Fokker-Planck equation [9]. On the other hand, CTRW—if properly scaled—in the limit leads to a continuous time subordinated processes described by a Langevin type equations, see e.g. [10, 11]. Such processes are in the origin of rapidly developing in the recent years field of anomalous diffusion, see e.g. [12]. In this paper we propose to use the tempered stable distribution [13] as an appropriate for the description of the waiting times and, consequently, the subordinator. This distribution has many interesting properties, like e.g. finite moments of all orders. Moreover it stays close to the purely α-stable family. It is worth mentioning, that this class of distributions allows for modeling data that demonstrate subdiffusive behavior at short times and normal at longer times.

The paper is structured as follows: in Sect. 2 we describe in details the analyzed time series. Next, in Sect. 3 we introduce the proposed model, i.e., the continuous time random walk that takes into account the properties observed in the considered data such as the apparent constant time periods and the seasonality. Moreover, we present the main properties of the proposed system. In Sect. 4 we analyze the temperature time series and apply the introduced discrete model. Finally, Sect. 5 consists of the simulation study and the last section concludes the paper.

2 Description of the Analyzed Datasets

As it was mentioned above, temperature is an important, measurable parameter associated with thermal comfort. Therefore our attention was focused on an analysis of time series of this variable. The present study was restricted to conditions characteristic for a classroom. The analysis of measurement results covers three days in June 2012, although the more extensive data set is already available. The classroom is involved in a long term indoor air quality (IAQ) monitoring program. The three days were selected to represent indoor conditions, which could be considered as the baseline for the studied space. The classroom was not equipped with mechanical ventilation. Its windows were sealed. In course of measurement series, the windows were not occluded and the room was closed. Therefore, except for cleaning operations in the morning the indoor air was influenced exclusively by the outdoor conditions.

The measurement device was located in the central part of the room, at the height of about 1 m. The temperature data was recorded with 15 s time resolution. The measurement accuracy was ± 0.1 °C.

We plotted the recorded time series in Fig. 1. Looking at the chart one easily notices time dependence of the temperature values. It may be observed that the presented time series exhibits a very characteristic behavior. The first aspect of it is the well pronounced seasonality. The temperature rises during the day and it falls during the night, in cycles. The second characteristic feature of the time series are the noticeable time periods when the temperature has
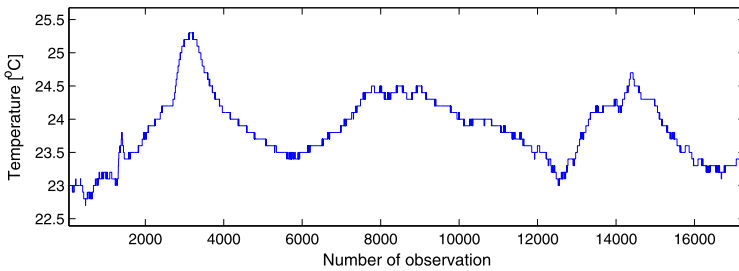
Fig. 1 Temperature [°C] recorded between 12.06.2012 and 14.06.2012

a constant value. These two elements are crucial for the stochastic model of the investigated temperature time series proposed in this paper.

3 The Model

3.1 Continuous Time Random Walk

The observed data is recorded in discrete setting for both time and scale. Hence, in order to describe the dynamics of the analyzed dataset, we use the continuous time random walk (CTRW) methodology [8]. Recall, that in the classical approach the CTRW process is defined as:

$$Y(t) = \sum_{i=1}^{L_t} X_i, \tag{1}$$

where the counting process $\{L_t\}$ is given by

$$L_t = \max\left\{n \in N : \sum_{i=1}^{n} T_i \leq t\right\}$$

with the sequence $\{T_n\}_{n=1}^{\infty}$ of nonnegative independent identically distributed (i.i.d.) random variables representing the waiting times. The sequence $\{X_n\}_{n=1}^{\infty}$ represents the jumps. The process $\{L_t\}$ is often referred to as the renewal process or, alternatively, as the counting process [14]. Moreover the sequences $\{T_n\}$ and $\{X_n\}$ are assumed to be independent.

In this paper we propose to use a modification of the classical CTRW scenario, in which the distribution of the jump sizes depends on the actual time of the jump, while the waiting times are governed by the tempered stable law. In the following we provide a motivation and a detailed description of the proposed model.

The analyzed dataset exhibits seasonality (see Fig. 1) and the jumps can take only two values a and $-a$ (due to the sensitivity of the sensor). Therefore, we consider a binomial model with a particle jumping between two sites, where a jump length is equal to a. The probability of the jump is governed by a periodic function that depends on the actual time t and therefore the jumps X_t for each t have the following form:

$$X_t = \begin{cases} a & \text{with probability } p_t, \\ -a & \text{with probability } 1 - p_t. \end{cases} \tag{2}$$

Such specification is an analogy to the field induced CTRW, analyzed in [16], with $p_t = \frac{1}{2}(1 + f(t))$ for f being a periodic function.

Further, the statistical analysis of the considered dataset (see Sect. 4) indicates that the waiting times can be described by the tempered stable distribution. Therefore we assume that the sequence $\{T_n\}$ constitutes a sample of independent random variables from the tempered stable distribution, i.e., for each n the random variable T_n has the following Laplace transform [17]:

$$E\left(e^{-zT_n}\right) = e^{c(\lambda^\alpha - (\lambda+z)^\alpha)} \qquad (3)$$

for some parameters $c, \lambda > 0$ and $0 < \alpha < 1$. The tempered stable distribution appears to be an important alternative to the power law distributions in confined systems, due to the finite second moment. For the simplicity in the further analysis we assume the scale parameter $c = 1$. The tempered stable distribution is related to the α-stable one through the operation of 'tempering'. Namely, the probability density function (pdf) of tempered stable random variable can be expressed in the following form:

$$g_T(x) = e^{-\lambda x + \lambda^\alpha} g_U(x), \qquad (4)$$

where U is a totally skewed α-stable random variable and g_U is it's pdf. As we observe, for $\lambda = 0$ the tempered stable distribution defined by the Laplace transform in (3) becomes the α-stable random variable U. The tempered stable distribution of waiting times was considered in [17–19] from both theoretical and practical point of view. Moreover the rich class of tempered stable distributions was examined in [13, 20–23]. Some interesting applications of the class of tempered stable distributions one can find also in [24, 25].

To sum up, we assume that the analyzed dataset can be modeled by the process defined as:

$$Y(t) = \sum_{i=1}^{L_t} X_{\Theta_i}, \qquad (5)$$

where $\Theta_j = \sum_{i=1}^{j} T_i$ with the sequence $\{T_n\}_{n=1}^{\infty}$ of i.i.d. tempered stable random variables, $L_t = \max\{n \in N : \Theta_n \leq t\}$ and the distribution of jumps X_j is defined in (2). Observe that, in contrast to the classical CTRW assuming that the waiting times and the jumps are independent, here the distribution of the jump depends on the actual time.

Using the relation between the counting process $\{L_t\}$ and the waiting times vector $\{T_n\}$ we can find the distribution of the process $\{L_t\}$:

$$P(L_t = k) = P\left(t - T_{k+1} < \sum_{i=1}^{k} T_i \leq t\right) = \int_0^\infty P\left(t - x < \sum_{i=1}^{k} T_i \leq t\right) g_T(x)dx$$

$$= \int_0^\infty \left(F_T^{*k}(t) - F_T^{*k}(t - x)\right) g_T(x)dx,$$

where $F_T(\cdot)$ is a cumulative distribution function (cdf) for the tempered stable random variable T and $F_T^{*k}(\cdot)$ denotes the cdf of the convolution of k tempered stable distributions. Let us mention, the convolution of k tempered stable random variables with scale parameter $c = 1$ has tempered stable distribution with scale parameter k.

In Fig. 2 we plot the first and the second moment of the process $\{Y(t)\}$ calculated from the simulated trajectories of the considered model with the parameters estimated from the analyzed dataset (for details see Sect. 4).
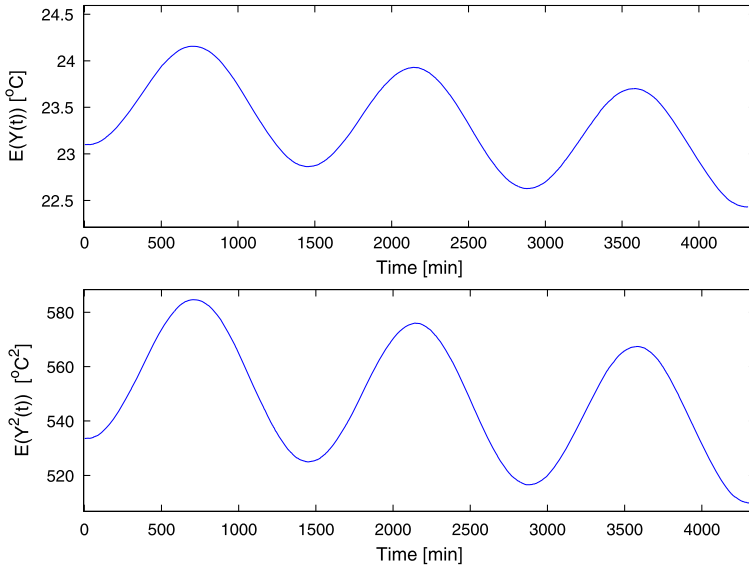
Fig. 2 The first and the second moment calculated from 10000 simulated trajectories of the CTRW model (see (5)) with parameters estimated from the analyzed dataset (for details see Sect. 4)

3.2 The Subordinated Process as a Limit of the CTRW Model

The CTRW model defined in (5) has the following limiting process [10]:

$$Z(t) = \int_0^t f(u)dS(u) + B\big(S(t)\big), \tag{6}$$

where $\{B(t)\}$ is the classical Brownian motion and $\{S(t)\}$ is the inverse subordinator. In the case of the tempered stable distribution of waiting times in the corresponding CTRW model, the process $\{S(t)\}$ is the inverse tempered stable subordinator, being a limit of the counting process $\{L_t\}$, see (5). The inverse tempered stable subordinator is defined as follows, [26]:

$$S(t) = \inf\{\tau : T(\tau) > t\}, \tag{7}$$

where $\{T(\tau)\}$ is a tempered stable Lévy process, i.e., a process with independent stationary increments having the Laplace transform given by:

$$E\big(e^{-zT(\tau)}\big) = e^{\tau(\lambda^\alpha - (\lambda+z)^\alpha)}. \tag{8}$$

Process $\{Z(t)\}$ defined in (6) is a special case of a system defined in [10]. It arises as a combination of the external process $\{X(\tau)\}$ which satisfies appropriate Langevin equation and inverse tempered stable subordinator. More precisely $\{Z(t)\}$ can be expressed as follows:

$$Z(t) = X\big(S(t)\big),$$

where $\{X(\tau)\}$ is a solution of the Langevin equation:

$$dX(\tau) = F\big(T(\tau)\big)dt + dB(\tau).$$

Some properties of the inverse tempered stable process one can find in [27] but also in [17], where the inverse tempered stable subordinator was analyzed in comparison with the

tempered stable Lévy process. Below, we recall a formula for the pdf of the inverse tempered stable subordinator [17], namely

$$g_{S(t)}(x) = \frac{1}{\alpha x} t g_{T(x)}(t) + \lambda \frac{1}{\alpha x} \int_0^t u g_{T(x)}(u) du - \lambda^\alpha \int_0^t g_{T(x)}(u) du. \tag{9}$$

For large and small t the density $g_{S(t)}$, (9), can be approximated using the following relation:

$$g_{S(t)}(x) \sim \frac{t}{\alpha x} g_{T(x)}(t), \tag{10}$$

where $g_{T(x)}(\cdot)$ is a pdf of a tempered stable random variable with the Laplace transform given in (8). The above relation is the extension of the similar behavior of stable subordinator and its inverse. This property is fully examined in [15]. The process $\{Z(t)\}$ defined in (6) is called the subdiffusion process with time-dependent force and was considered in [10] not only for the case of the tempered stable distribution but for all infinitely divisible distributions of the subordinator. The discrete analogue of the process $\{Z(t)\}$ was also analyzed in [16], where the authors presented relation between the pdf of the process and the fractional Fokker-Planck formula. Namely, the pdf $g_{Z(t)}(\cdot)$ of the subdiffusion process with time-dependent force satisfies the following fractional Fokker-Planck equation [16]:

$$\frac{\partial g_{Z(t)}(x)}{\partial t} = \left[-f(t) \frac{\partial}{\partial x} + \frac{1}{2} \frac{\partial^2}{\partial x^2} \right] \Phi g_{Z(t)}(x), \tag{11}$$

where the integro-differential operator Φ is defined as follows:

$$\Phi g(t) = \frac{d}{dt} \int_0^t M(t-y) g(y) dy$$

for sufficiently smooth function $g(\cdot)$. In the above definition the function $M(\cdot)$ is called the memory kernel and in the case of the tempered stable distribution it has the following representation [28]:

$$M(t) = e^{-\lambda t} t^{\alpha-1} E_{\alpha,\alpha}\big((\lambda t)^\alpha\big),$$

where $E_{\alpha,\beta}(\cdot)$ is the generalized Mittag-Leffler function defined as:

$$E_{\alpha,\beta}(z) = \sum_{k=0}^\infty \frac{z^k}{\Gamma(\alpha k + \beta)}.$$

Using Proposition 1 in [10] we can calculate the moments of the process $\{Z(t)\}$:

$$E\big(Z(t)\big) = \int_0^t f(u) \frac{d}{du} \int_0^u M(u-z) dz du,$$

$$E\big(Z^2(t)\big) = 2 \int_0^t f(u) \Phi E\big(Z(u)\big) du + \int_0^t M(t-u) du.$$

In the special case when the function $f(x) = 0$ for all $x \in R$ the above formulas reduce to [28]:

$$E\big(Z(t)\big) = 0, \qquad E\big(Z^2(t)\big) = \int_0^t e^{-\lambda u} u^{\alpha-1} E_{\alpha,\alpha}\big((\lambda u)^\alpha\big) du.$$
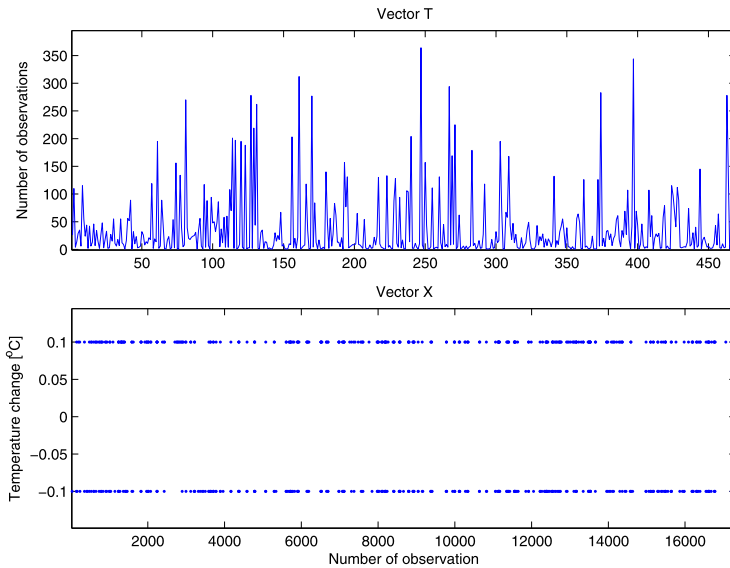
Fig. 3 The constant time periods (vector T) and the jumps (vector X)—two time series related to the temperature data

4 Temperature Data Analysis

According to our assumption of the CTRW scenario, we apply here a procedure adequate for such kind of processes. Namely, in the first step of our analysis we divide the data into two vectors. The first one, vector T, consists of the lengths of the constant time periods and corresponds to the waiting times T_i in the CTRW definition (see Eq. (5)). On the other hand, removing the constant time periods from the dataset, i.e., considering only these time points for which a change in the values of the analyzed process occurs, leads to a vector Y_{t_i} corresponding to the jumps of the CTRW process. Precisely, differentiating the vector Y_{t_i} we obtain the sizes of the jumps X_i, namely $X_i = Y_{t_{i+1}} - Y_{t_i}$. The vectors obtained from the analyzed dataset are plotted in Fig. 3.

4.1 Waiting Times

In the next step, we analyze the waiting times properties. According to the CTRW definition, the vector $\{T_i\}$ should form an independent, identically distributed sample. First, we check the independence assumption. We apply a simple visual test that is based on the autocorrelation function of both, the series and the squared series, for details see [29] or [30]. Recall, that the autocorrelation function of a time series is defined as the correlation between equally distant observations as a function of time lag. If there is no dependence in the analyzed data the autocorrelation values should be close to 0. Moreover, the plot should resemble the values expected for a white noise sequence. In Fig. 4 we plot the obtained autocorrelation functions together with the confidence intervals for a white noise. As can be observed, the calculated values are close to 0 and most of them lie within the white noise confidence intervals. Hence, we may conclude that the waiting times are independent.
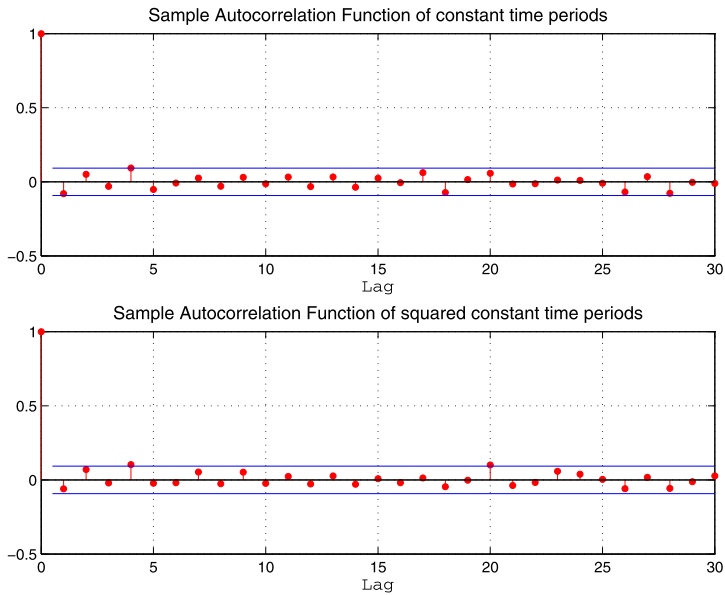
Fig. 4 The sample autocorrelation function of the constant time periods (*top panel*) and the corresponding squared time series (*bottom panel*)—a simple visual test for independence
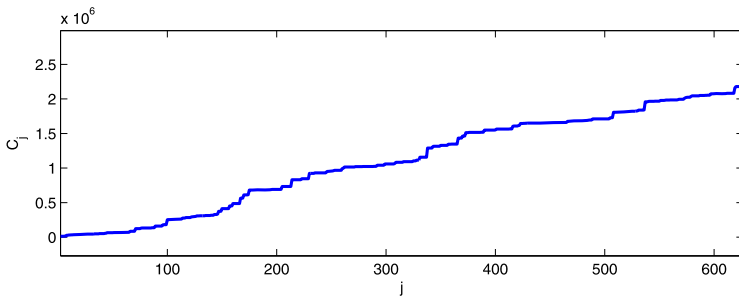

Fig. 5 The C_j statistic calculated on the basis of the waiting times. The linear (with respect to j) behavior of this statistic indicates that the vector $\{T_i\}$ constitutes a sample of identically distributed random variables

Next, we check if the waiting times have the same distribution. Here we propose to use a simple visual stationarity test based on the behavior of the empirical second moment of the underlying time series $\{T_i\}$, i.e. the C_j statistic defined as follows:

$$C_j = \sum_{i=1}^{j} T_i^2.$$

If the analyzed vector constitutes a sample with elements from the same distribution, then the C_j statistic is a linear function with respect to j. More details of this visual test one can find in [31]. In Fig. 5 we present the behavior of the C_j statistic calculated for the waiting times, i.e. the vector $\{T_i\}$.
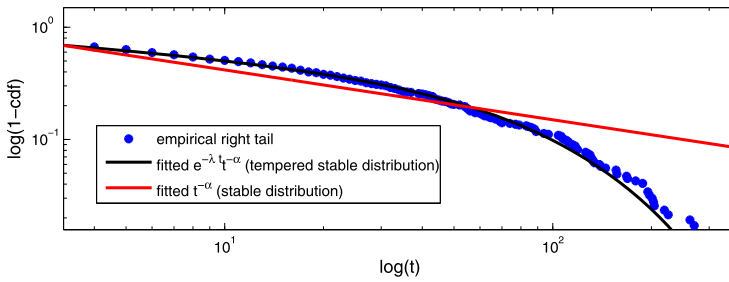
Fig. 6 The empirical right tail of the waiting times sample and the fitted theoretical functions corresponding to the α-stable and tempered stable distributions

On the basis of statistic C_i the rigorous regime variance test can be also constructed. The null hypothesis for this tests is defined as follows: the quantiles of the squared time series do not change in time. The hypothesis is satisfied for example in the case of independent identically distributed random variables. For a detailed description of the test see [31]. In our case the obtained p-value is equal to 0.2, what confirms the hypothesis of the same distribution of the waiting times.

Further, we find the waiting times distribution. We consider two types of distributions, namely α-stable and tempered stable. Since the data consists of only integer values (due to the measurement methodology), in order to estimate the distribution parameters, we propose to use a method based on the tail behavior of the data. Namely, we fit a tail function corresponding to the tested distribution to the right empirical tail of the dataset. A similar method was used in [32, 33] in the context of the α-stable distribution, as well as in [19] for the tempered stable case. If the data comes form the α-stable distribution, the right tail behaves like a power function $t^{-\alpha}$. On the other hand, if the sample follows the tempered stable law, then the right tail can be approximated by $e^{-\lambda t} t^{-\alpha}$, where α and α, λ are the parameters of the α-stable and tempered stable distributions, respectively. In the considered case the above functions are fitted to the right empirical tails of the sample using least squares method. In Fig. 6 we plot the empirical right tail (in double logarithmic scale) and the fitted functions corresponding to both distributions. As we observe, the empirical tail does not behave according to a power law. On the other hand, the tempered stable distribution yields much better fit than the α-stable one and replicates the empirical tail well. The estimated parameters of the tempered stable distribution are $\lambda = 0.01274$ and $\alpha = 0.2082$.

4.2 Jumps

Finally, we analyze the distribution of the jump sizes. The recorded temperature values may jump in each time step upward or downward. Due to the measurement accuracy, the size of the jump is always equal to 0.1 °C. However, as can be observed in Fig. 1, the probability of upward (or equivalently downward) jump varies with time. Based on these observations we assume that the jumps of the process are given by

$$X_t = \begin{cases} +0.1 & \text{with probability } p_t, \\ -0.1 & \text{with probability } 1 - p_t, \end{cases} \tag{12}$$

where p_t is a time varying (periodic) function.

In order to estimate the probability p_t, we use two methods. In the first one we use nonparametric approach and estimate p_t as the proportion of upward jumps within a given

Table 1 Parameters of the function p_t estimated using the maximum likelihood method	a_1	a_2	a_3	a_4
	0.09	5731	-0.03	0.5

hour. Precisely, we divide the dataset into 24 subsets corresponding to each hour of the day. Next, for each hour, we calculate the number of upward jumps and divide it by the overall number of jumps within that hour. An advantage of such method is that it does not require an upfront specification of the form of p_t. However, the results are only reliable, if there are many data points used in the estimation procedure. Here, even though we have recorded in total 17280 time points, the subsets consist of only from 11 observations (for 2 am) up to 47 observations (for 9 am).

In the second method we assume that the upward jump probability is given by a periodic function of the form $p_t = a_1 \sin(2\pi/a_2(t + a_3)) + a_4$. In order to estimate the parameters a_i, we use the maximum likelihood method, i.e. we maximize the function $L(\mathbf{x}, a_1, a_2, a_3, a_4) = \prod_{t=t_0}^{T} P(X_t = x_t | a_1, a_2, a_3, a_4)$, where $\mathbf{x} = (x_{t_0}, x_{t_1}, \ldots, x_T)$ is the observed sample and the possible values for x_t are ± 0.1. Observe that if $x_t = 0.1$, then $P(X_t = x_t | a_1, a_2, a_3, a_4) = p_t = a_1 \sin(2\pi/a_2(t + a_3)) + a_4$. On the other hand, if $x_t = -0.1$, then $P(X_t = x_t | a_1, a_2, a_3, a_4) = 1 - p_t = 1 - a_1 \sin(2\pi/a_2(t + a_3)) - a_4$. Therefore, using the indicator function $\mathbf{1}_{\{\cdot\}}$, we may write that:

$$L(\mathbf{x}, a_1, a_2, a_3, a_4) = \prod_{t=t_0}^{T} P(X_t = x_t | a_1, a_2, a_3, a_4)$$

$$= \left[a_1 \sin\left(\frac{2\pi}{a_2}(t + a_3) \right) + a_4 \right]^{\mathbf{1}_{\{x_t = 0.1\}}}$$

$$\times \left[1 - a_1 \sin\left(\frac{2\pi}{a_2}(t + a_3) \right) - a_4 \right]^{\mathbf{1}_{\{x_t = -0.1\}}}. \quad (13)$$

In practice the above likelihood function has to be maximized numerically. It should be noted, that the numerical maximization may lead to parameter estimates yielding a value of p_t that is negative or higher than one. Hence, in order to avoid such situation, some constraints should be used in the procedure. Here, we assume that $a_4 > 0$, $|\frac{a_4}{a_1}| \geq 1$ and for $a_1 > 0 \ \frac{a_4 - 1}{a_1} \leq -1$, while for $a_1 < 0 \ \frac{a_4 - 1}{a_1} \geq 1$. The estimated parameters are provided in Table 1. Note that the obtained period of 5731 observations approximately corresponds to a daily periodicity, as there are 5760 observations each day. This is in compliance with the intuition and the visual investigation of Fig. 1.

The values of p_t obtained using both proposed methods are plotted in Fig. 7. Observe that, although the curves are not identical, they preserve similar shape with a higher probability of an upward jump at the beginning of the day and lower in the late afternoon and night hours. This result is in compliance with the temperature properties apparent in Fig. 1. Analogously, there is an upward trend of temperature in the morning/early afternoon, which changes into a downward trend in the late afternoon.

5 Simulation Study

In this section we perform a simulation study using the proposed model and the parameters estimated in the previous section. First, in Fig. 8 we plot a sample simulated CTRW trajec-
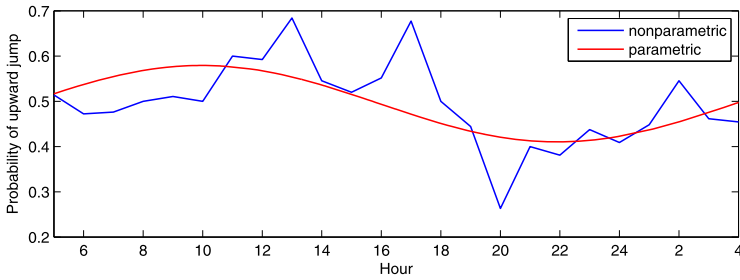
Fig. 7 The estimated probability of an upward jump using nonparametric method (*blue line*) and sine specification (*red line*)
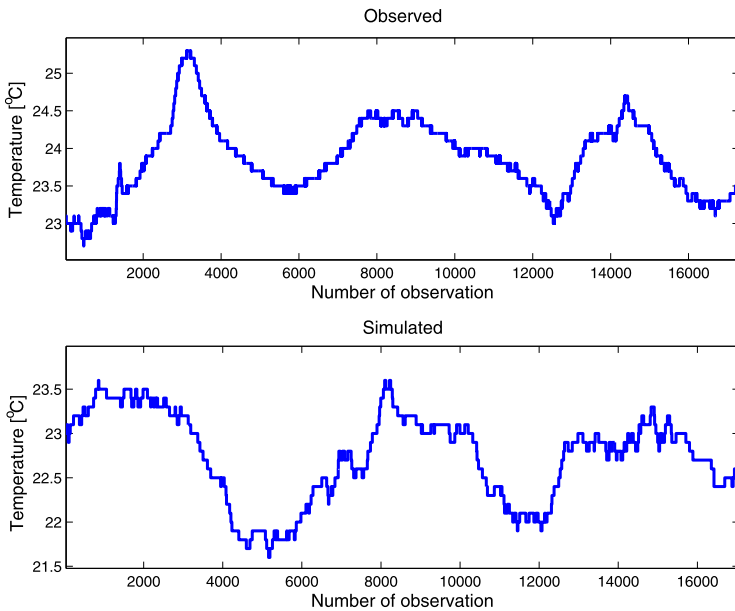

Fig. 8 The observed (*top panel*) and simulated (*bottom panel*) values of temperature

tory and compare it with the observed values of temperature. As can be noted, the simulated values replicate the properties of the analyzed dataset.

Next, we calculate the quantile lines, i.e., the curves describing the values that with a given probability will not be exceeded by the corresponding process, see [34] for details. The quantile lines of levels 10 %, 20 %, ..., 90 % obtained using 10000 simulated trajectories are plotted in Fig. 9. Additionally, in the same figure we plot the recorded temperature values. Observe that the observed temperature does not exceed the bounds given by the 10 % and 90 % quantile lines.

Further, using 10000 simulated trajectories of the CTRW model (5), we plot the time evolution of the model distribution. The obtained probabilities $P(Y(t) = x)$ are plotted in Fig. 10.

Finally, in order to illustrate how beneficial might be the simulations of the proposed model in the air quality monitoring, we calculate the temperature forecast for the next hour
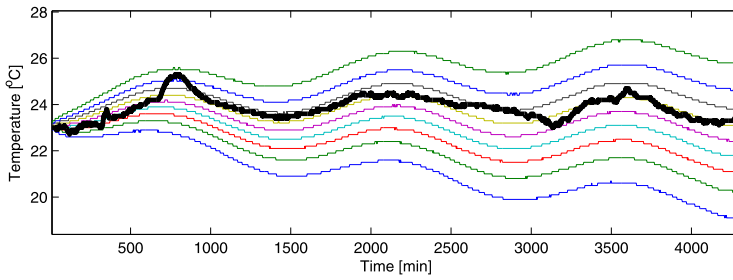
Fig. 9 Quantile lines of level 10 %, 20 %, ..., 90 % and the measured temperature (*black thick line*)
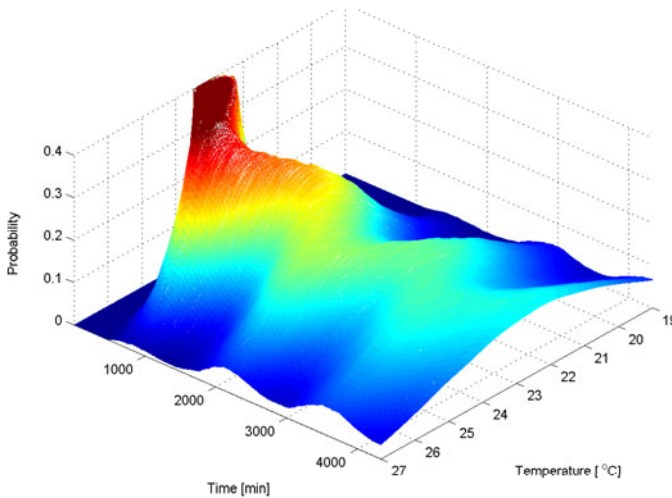

Fig. 10 Time evolution of the temperature probability in the considered CTRW model

and the next day. The obtained forecast is then validated by comparing it with the actual measured values. To this end, we fit the model to the first 11520 observations (i.e. 48 hours) and then based on the obtained estimates we calculate the forecast. The values obtained using 10000 simulated trajectories of the considered model are plotted in Fig. 11. Moreover, we calculate the 10 % and 90 % confidence bounds for the forecast, i.e., the curves within which a future value should lie with 80 % probability. As can be observed in the figure, the actual measured values lie within the calculated confidence interval and the forecast visually resembles the actual values. Next, we perform a more rigorous validation. Namely, we calculate the mean squared error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) of the forecast for the next 24 hours and the next hour. The obtained values are given in Table 2. The mean percentage error does not exceed 0.02 % for the 24 hours forecast and is no more than 0.002 % for the one hour forecast, so the temperature values are predicted quite accurately. In absolute terms the mean 24 hour error is about 0.4 °C, while the mean hourly error about 0.04 °C.
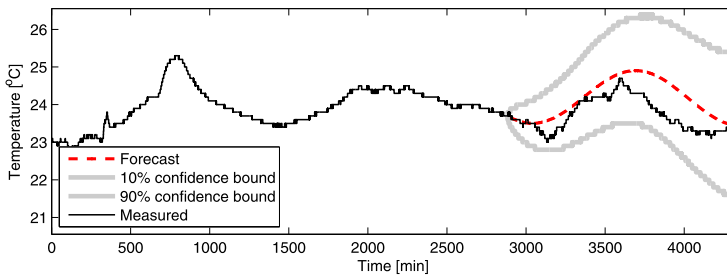
Fig. 11 Measured values of temperature together with the 24 hour forecast. Additionally, the 10 % and 90 % confidence bounds are given

Table 2 Mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) of the forecast for the next 24 hours (first row) and the next hour (second row)

Time horizon	MSE	MAE	MAPE
24 hours	0.2713	0.4234	0.0179 %
1 hour	0.0028	0.0357	0.0015 %

6 Conclusions

Measurements of temperature may be a source of information to characterize the transient state of the thermal environment. The temperature is extremely important to the occupant's perception of indoor air quality and thermal processes occurring inside room [35]. It reflects transient state of thermal environment because of an inherent link with building usage and surroundings [36]. The temperature in a building is dependent on outside temperature, sun loading, plus heating and cooling added by the HVAC system and other sources, e.g. occupants add heat to the room since the normal body temperature is much higher than the room temperature. Therefore, temperature value is the result of a range of interactions affected by seasonal and daily changes in meteorology and by the requirements of occupants varying in time and space [37].

In principle, two major kinds of information are useful for the indoor microclimate specialists. The first kind of information is applicable to the direct control of heat ventilation and air conditioning systems. The data which is useful for this purpose shall be acquired continuously and in real time. The other kind of information serves the assessment of the indoor thermal conditions in relatively long-term perspective. The evaluation shall be based on the time series of the measurement data. Therefore, in this paper we search for mathematical methods which would allow for the complex and quantitative assessment of indoor thermal conditions based on the time series of the measured values of physical/chemical parameters e.g. the temperature. In our opinion this is the way to provide the second kind of information.

In this paper we considered a stochastic system that allows for modeling such time series as the indoor temperature. We took into account the seasonal behavior but also the constant time periods observable in the indoor temperature measurements. More precisely, we proposed to use a generalized CTRW scheme in which the probability of jumps is given by a (periodic) function of time. Moreover we extended the classical model by introducing the tempered stable distribution of waiting times. For such a model we proposed a procedure to

verify the model assumptions and to estimate the model parameters. This procedure can be divided into following steps:

– divide the analyzed data into two vectors—first corresponding to the waiting times and second to the jump sizes,
– check the assumption on the independence of waiting times using autocorrelation function,
– check if the waiting times constitute sample from the same distribution by using tests based on the empirical second moment of the underlying time series,
– estimate the parameters of the tempered stable distribution using the empirical tail behavior and check the fit of the theoretical tail,
– estimate the parameters of the jump probabilities using the maximum likelihood method.

Finally, we have shown in the simulation study that fitting an appropriate model to the analyzed data set might be beneficial in analyzing, monitoring and forecasting the thermal environment indoors. It is worth mentioning, that the presented methods were used for the measurements over short periods that gives the greater reliability of the obtained results.

References

1. García, J.A.O.: A review of general and local thermal comfort models for controlling indoor ambiences. In: Kumar, A. (ed.) Air Quality. InTech, Rijeka (2010)
2. Ozbalta, T.G., Sezer, A., Yildiz, Y.: Models for prediction of daily mean indoor temperature and relative humidity: education building in Izmir, Turkey (2011). doi:10.1177/1420326X11422163
3. Zhao, R.: Investigation of transient thermal environments. Build. Environ. **42**, 3926–3932 (2007)
4. Hossain, A., Fernandez-Gonzalez, A.: Development and validation of a thermal network model to predict indoor operative temperatures in dry roofpond buildings. In: World Renewable Energy Forum, Denver, Colorado, 13–17 May (2012)
5. Scartezzini, J.L., Faist, A., Liebling, Th.: Using Markovian stochastic modelling to predict energy performances and thermal comfort of passive solar systems. Energy Build. **10**(2), 135–150 (1987)
6. Riederer, P., Marchio, D., Visier, J.C., Husaunndee, A., Lahrech, R.: Room thermal modeling adapted to the test of HVAC control systems. Build. Environ. **37**, 777–790 (2002)
7. Maciejewska, M., Szczurek, A., Sikora, G., Wyłomańska, A.: Diffusive and subdiffusive dynamics of indoor microclimate. A time series modeling. Phys. Rev. E **86**, 031128 (2012)
8. Montroll, E.W., Weiss, G.H.: Random walks on lattices. II. J. Math. Phys. **6**, 167–181 (1965)
9. Metzler, R., Klafter, J.: The random walk's guide to anomalous diffusion: a fractional dynamics approach. Phys. Rep. **339**, 1–77 (2000)
10. Magdziarz, M.: Langevin picture of subdiffusion with infinitely divisible waiting times. J. Stat. Phys. **135**, 763–772 (2009)
11. Magdziarz, M.: Stochastic representation of subdiffusion processes with time-dependent drift. Stoch. Process. Appl. **119**, 3416–3434 (2009)
12. Klages, R., Radons, G., Sokolov, I.M.: Anomalous Transport: Foundations and Applications. Wiley, New York (2008)
13. Rosinski, J.: Tempering stable processes. Stoch. Process. Appl. **117**, 677–707 (2007)
14. Magdziarz, M., Weron, K.: Anomalous diffusion schemes underlying the cole-cole relaxation. The role of the inverse-time. Physica A **367**, 1–6 (2006)

15. Magdziarz, M., Weron, A., Weron, K.: Fractional Fokker-Planck dynamics: stochastic representation and computer simulation. Phys. Rev. E **75**, 016708 (2007)
16. Sokolov, I.M., Klafter, J.: Field-induced dispersion in subdiffusion. Phys. Rev. Lett. **97**, 140602 (2006)
17. Wyłomańska, A.: Arithmetic Brownian motion subordinated by tempered stable and inverse tempered stable processes. Physica A **391**(22), 5685–5696 (2012)
18. Orzeł, S., Wyłomańska, A.: Calibration of the subdiffusive arithmetic Brownian motion with tempered stable waiting-time. J. Stat. Phys. **143**, 447–454 (2011)
19. Gajda, J., Wyłomańska, A.: Geometric Brownian motion with tempered stable waiting times. J. Stat. Phys. **148**, 296–305 (2012)
20. Terdik, G., Woyczynski, W.A.: Rosinski measures for tempered stable and related Ornstein-Uhlenbeck processes. Probab. Math. Stat. **26**(2), 213–243 (2006)
21. Chakrabarty, A., Meerschaert, M.M.: Tempered stable laws as random walk limits. Stat. Probab. Lett. **8**, 989–997 (2011)
22. Mantegna, R.N., Stanley, H.E.: Stochastic processes with ultraslow convergence to a Gaussian: the truncated Lévy flight. Phys. Rev. Lett. **732**, 946–2949 (1994)
23. Kim, Y.S., Rachev, S.T., Bianchi, M.L., Fabozzi, F.J.: A new tempered stable distribution and its application to finance. In: Bol, G., Rachev, S.T., Wuerth, R. (eds.) Risk Assessment: Decisions in Banking and Finance. Physika, pp. 77–110. Springer, Berlin (2009)
24. Palmer, K.J., Ridout, M.S., Morgan, J.T.: Modelling cell generation times by using the tempered stable distribution. Appl. Stat. **57**(4), 379–397 (2008)
25. Terdik, G., Gyires, T.: Lévy flights and fractal modeling of Internet traffic. IEEE/ACM Trans. Netw. **17**(1), 120–129 (2009)
26. Stanislavsky, A., Weron, K., Weron, A.: Diffusion and relaxation controlled by tempered-stable processes. Phys. Rev. E **78**, 051106 (2008)
27. Piryatinska, A., Saichev, A.I., Woyczynski, W.A.: Models of anomalous diffusion: the subdiffusive case. Physica A **349**, 375–420 (2005)
28. Janczura, J., Wyłomańska, A.: Anomalous diffusion models: different types of subordinator distribution. Acta Phys. Pol. B **43**(5), 1001–1016 (2012)
29. Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting, 2nd edn. Springer, New York (1996)
30. Wyłomańska, A.: How to identify proper model? Acta Phys. Pol. B **43**(5), 1241–1253 (2012)
31. Gajda, J., Sikora, G., Wyłomańska, A.: Regime variance testing—a quantile approach. Acta Phys. Pol. B **44**(5), 1015–1035 (2012)
32. Janczura, J., Wyłomańska, A.: Subdynamics of financial data from fractional Fokker-Planck equation. Acta Phys. Pol. B **40**(5), 1341–1351 (2009)
33. Janczura, J., Orzeł, S., Wyłomańska, A.: Subordinated α-stable Ornstein–Uhlenbeck process as a tool for financial data description. Physica A **390**, 4379–4387 (2011)
34. Janicki, A., Weron, A.: Simulation and Chaotic Behaviour of α-Stable Stochastic Processes. Marcel Dekker, New York (1994)
35. Fang, L., Clausen, G., Fanger, P.O.: Impact of temperature and humidity on the perception of indoor air quality. Indoor Air **8**, 80–90 (1998)
36. Yun, G.Y., Steemers, K.: Time-dependent occupant behaviour models of window control in summer. Build. Environ. **43**, 1471–1482 (2008)
37. Hazyuk, I., Ghiaus, Ch., Penhouet, D.: Optimal temperature control of intermittently heated buildings using model predictive control: part II—control algorithm. Build. Environ. **51**, 388–394 (2012)