EMPIRICAL CUMULATIVE DISTRIBUTION FUNCTION
OF CO₂ CONCENTRATION INCREMENTS
AS A DESCRIPTOR OF INDOOR AIR VARIABILITY

Indoor air quality broadly refers to the environmental characteristics inside buildings and characterizes the physical, chemical and biological state of indoor air at some place and time. Usually it is characterized by physicochemical properties of indoor environment such as temperature, relative humidity, airflows and concentration of characteristic pollutants, e.g. carbon dioxide. They are highly variable in time. In the paper, changes of empirical cumulative distribution function (ECDF) of CO₂ concentration were considered as a descriptor of indoor air variability. Using this measure, there may be performed the analysis of CO₂ variation in a short as well as long term. The approach utilizes the idea of classification. A class of variation of CO₂ concentration has its individual ECDF. It indicates a particular case of balance/imbalance between CO₂ delivery and removal processes, including gas dispersion. The analysis is applicable for evaluating the stability of indoor environment and it may provide support for the diagnosis of the performance of the air exchange system. Exemplary results were provided for an open space office and a lecture theater.

1. INTRODUCTION

Worldwide energy crisis in 1970s has revealed the importance of energy saving. However, the intense movement towards energy conservation via energy-efficient designs frequently subjected employees and residents to poor indoor air quality (IAQ). It causes discomfort, leads to health problems, complaints, and in the workplaces contributes to excessive absenteeism as well as lower productivity [1–3]. Indoor air quality broadly refers to the environmental characteristics inside buildings and characterizes the physical, chemical and biological state of indoor air at some place and time [4]. Usually
it is characterized by physicochemical properties of indoor environment such as temperature \((T)\), relative humidity \((RH)\), airflows and concentration of characteristic pollutants, e.g. carbon dioxide.

There is no doubt that IAQ is a serious problem in many buildings, especially if they were built to be airtight and energy efficient [5]. For this reason, there is a strong need for IAQ evaluation [6]. The results of assessment may be an important source of information for building commissioning, proactive building management, diagnosis of complaints about indoor air quality, and investigation of building energy consumption. Various approaches are proposed to evaluate IAQ. Currently, CO\(_2\) monitoring is often used for an indirect indication of indoor air quality problems due to insufficient fresh air delivery rates [7–9]. Various aspects may be taken into account in this approach. One of them is temporal variability of CO\(_2\) concentration [10].

A characteristic feature of indoor air is relative stability or instability of its state. It is reflected in the variability of various parameters describing this component of environment. In our studies, this term refers to variations in the mean state and other statistics (such as standard deviation, occurrence of extremes, etc.) of CO\(_2\) concentration. Changes of this parameter are crucial for evaluation of indoor air quality because they contain information on the emission and removal processes, operation of ventilation system and the influence of occupancy on indoor air quality. The variability of CO\(_2\) concentration also allows one to characterize the ability of this gas to disperse inside the building. It can additionally inform that other volatile contaminants are accumulating inside a building. Variations of CO\(_2\) concentration affect the perception of conditions inside a room. Moreover, the understanding of CO\(_2\) dynamics is important to be capable of designing proper sampling strategy in a monitoring system as well as to evaluate the obtained results [11].

The variation with time of the concentration of pollutants in an indoor air is well known. However, current methods are not potent enough for qualitative and quantitative characterization of this phenomenon. Particularly, they are not credible for the assessment of contaminant variations in long period of time. Indoor air which appears to be stable, may rapidly become unstable following a relatively small perturbation. The sensitivity to even small perturbations can lead to a wide range and complex changes of gas concentrations.

The additional problem results from the accuracy of the measurement. In the case of commonly used CO\(_2\) monitors, this parameter is at the level of ±50 ppm + 3% of the measured value. Hence, measurement errors can lead to significant uncertainties in the CO\(_2\) determination. It means a difficulty in recording small fluctuations of its concentration.

The aim of the presented study is a method to characterize a long-term, temporal variability of indoor CO\(_2\) concentration. In our opinion, the obtained results can be of great importance in many applications. The proposed approach is especially useful for the diagnostic works in building.
2. METHODS

Time-series analysis is one of the most widely applied approaches for evaluating variability, identifying measures of the system variation over time. Generally speaking, it comprises methods for analyzing time series of data in order to extract meaningful statistics and other temporal characteristics of the data. Traditionally, the variability may be expressed by applying standard sample statistics, e.g. range, interquartile range, variance and standard deviation of the observed time series. These parameters are easy to compute. However, a major shortcoming of traditional measures of variability lies in the fact that though the principles of statistical approach are reasonable, the time-averaged parameters which are calculated from time series may not reflect a real character of variation. They represent a coarse quantification of the overall variability. As measures of global variation, they are altered by the duration of measurement. For example, longer series will have greater standard deviation of data. More exhaustive information is included in parameters of models describing time series. Time series models can have many forms and represent different stochastic processes. Unfortunately, in many cases they display inability to identify the complex characteristics of data. This limitation results from the requirement of the stationarity of time series. This term basically means that statistical properties of time series should remain the same throughout the period of observation. In practice, these are features of a flat looking signal, without trend, with constant variance and constant autocorrelation structure over time, as well as no periodic fluctuations (seasonality). Stationarity does not preclude variability. However, it provides the limitation that variability does not change with time or with duration of measurement. This requirement is difficult to fulfill, because indoor air is a time-dependent, non-equilibrium system characterized by intrinsic dynamic, interdependent, nonlinear relationships of its elements and short-term fluctuations [12, 13]. Therefore the time series describing CO₂ variations are not necessarily linear or stationary. In order to avoid this problem, the analysis of time series was based on a segmentation procedure.

The analysis of time series data may be accomplished using single data, patterns or segments. In practice, the analysis of individual results of measurement involves considerable errors, whereas patterns of the data are very often unclear, difficult for identification and interpretation. Our approach is based on a segmentation. This term means a division of time series into a sequence of segments, each having its own, characteristic properties. In time-series segmentation, the key issue was to identify the segment boundary points in the time-series and to describe the variability of CO₂ concentration increments associated with each segment. In this work, the time series were divided into non-overlapping windows, 1 h (for the open space) or 15 min (for the lecture room) long. In other words, the original time series was decomposed into a set of meaningful components, featured by the same time scale depending on the object. The width of segments was selected arbitrarily, but the characteristic temporal dependence of factors
influencing IAQ parameters was taken into account. We assumed that short-term variations of carbon dioxide are first of all a function of an occupant activity pattern and operation of ventilation system. The influence of these factors on indoor air is mainly reflected in short-time periods. We considered 1 h and 15 min long segments. The shorter temporal window was chosen for lecture theater in order to account for breaks between lectures.

Time series reflect the stochastic nature of measurements over time. This fact has a principal importance in our study, because the time series analysis based on statistical methods allows one to obtain information from the measurement results which are characterized by a relatively high uncertainty.

In order to characterize the variability of increments of CO₂ concentration, we proposed to apply the descriptive statistics. More precisely, empirical cumulative distribution function (ECDF) was used for this task. It describes the magnitude and direction of momentary concentration changes, as well as the frequency of their occurrence. In other words, empirical cumulative distribution function allows one to characterize qualitatively and quantitatively the variability of CO₂ concentration in defined, short-time periods. The characterization is based on the shape and location of ECDF in the domain of concentration changes. Its most informative features are the slope and symmetry with respect to zero concentration increment. The ECDF slope indicates the range of changes of CO₂ concentration which were actually observed in the data segment and relative contribution of particular ones (Fig. 1a).

Steep functions are observed when a narrow range of increments is overrepresented in the characterized data segment. Functions with a smaller slope are symptomatic for broader range of increments. The symmetry of cumulative distribution function with
Increments of CO₂ concentration as a descriptor of indoor air variability

respect to the concentration increment equal zero is indicative for the case, when proportions of positive and negative increments in the data segment (Fig. 1b) are comparable. A symmetric ECDF is observed if the time intervals when CO₂ concentration increases and are well balanced by the time intervals when CO₂ concentration decreases in the considered time segment. It basically means that there is an equivalence between the delivery and removal processes. An asymmetric ECDF may indicate either the prevalence of positive or negative concentration increments. In the first case, the data segment is associated with the time when CO₂ delivery processes dominate. In the second case the removal processes prevail in the characterized time segment.

The evaluation of variability of CO₂ concentration in long-term perspective was performed on the basis of segment classification and statistical analysis of the identified classes. We assumed that empirical cumulative distribution functions were the descriptors of variability of CO₂ concentration. Additionally, they could represent CO₂ variation classes. The similarity of these functions in distinct data segments was used as the basis for differentiating classes. The descriptions of variability of CO₂ concentration in distinct data segments were compared using Cramér–von Mises test for two samples.

The Cramér–von Mises two-sample test is one of the best-known nonparametric two-sample tests. It is applied to determine if two independent samples were drawn from the same distribution. Let there be two independent samples \( x_1, x_2, \ldots, x_m \) and \( y_1, y_2, \ldots, y_n \) which are drawn independently from two distributions and have continuous cumulative distribution functions \( F(\cdot) \) and \( G(\cdot) \), respectively. Based on those samples, the null hypothesis is to be tested \( H_0: F = G \) against the two-sided alternative \( H_1: F \neq G \). The Cramér–von Mises test statistic \( T_2 \) is given by [14]:

\[
T_2 = \frac{nm}{(n + m)^2} \left( \sum_{i=1}^{m} \left( F_m(x_i) - G_n(x_i) \right)^2 + \sum_{j=1}^{n} \left( F_m(y_j) - G_n(y_j) \right)^2 \right) \tag{1}
\]

where \( F_m \) and \( G_n \) are the empirical distribution functions associated with the respective samples \((x_i)\) and \((y_j)\). The test based on the test statistic consists in rejecting \( H_0 \) if the value of \( T_2 \) is “too large” regarding the accepted significance level \( \alpha \). This statistic and the corresponding test were first studied by Anderson [15] as a 2-sample variant of the goodness-of-fit test introduced by Cramér (1928) and von Mises (1931) [14]. The test is superior regarding its power compared with an alternative two-sample Kolmogorov–Smirnov test [16]. In our calculations, we applied the freeware code developed for Matlab and realizing the Cramér–von Mises two-sample test [16].

The similarity of ECDFs in distinct data segments was evaluated using the two-sample Cramér–von Mises test and the following procedure:

1. Order data segment in time domain.
2. Determine empirical cumulative distribution for each data segment.
3. The first data segment in the sequence is assumed to represent class 1 of variability of CO₂ concentration.

4. Consider first data segment as sample 1 and second data segment as sample 2. Perform the two-sample Cramér–von Mises test (use \( \alpha = 0.01 \)).

4a. If the null hypothesis is not rejected, data sets are considered as being drawn from the same distribution. It is assumed that data sets fulfilling this condition belong to the same class. Label this class using the label of sample with a known class assignment. Aggregate the two data sets and determine ECDF for the newly obtained one.

4b. If the null hypothesis is rejected, data sets are considered as being drawn from different distributions. It is assumed that data segments fulfilling this condition belong to different classes. Label the new class.

5. Consider next data segment in a sequence. Perform the two-sample tests comparing ECDFs for this data and the data sets belonging to the earlier distinguished classes. Depending on test result proceed as mentioned in 4a or 4b.

6. Realize step 5 for the subsequent data segments, one by one, till the last data segment is assigned to the adequate class of variability of CO₂ concentration.

The procedure allows one to distinguish classes of variability of CO₂ concentration in long-term perspective. This is based on the statistical description of variation of this parameter in data segments associated with short periods of time.

3. EXPERIMENTAL

Measurements of CO₂ concentration were carried out in two distinct objects: an open space office and a lecture room. The open space office had a considerable size of several hundred square meters. It was located in an old building, completely renovated and refurbished. External building walls were glassy, without openable windows. Indoor air exchange was realized by a mechanical ventilation system. Air treatment allowed controlling the temperature and humidity. In the period of study, the space was exploited in a typical manner. The presence of people indoors was limited to working days in time interval from 8:00 a.m. to 5:00 p.m. (working hours). Before and after this period, the office was empty, except regular cleaning operations and occasional maintenance works. During working hours, the number of occupants was rather constant. Human load did not change considerably from day to day. Hence, on distinct days, the influence of human factor on indoor air exhibited similar regularity. In the office, CO₂ concentration was monitored in one point, over the period from 27.10.2012 to 29.02.2013. Valid measurement data were obtained for 88 days out of 125. We considered only those days, when the data set was complete. The time resolution of data collection was 2 min. This value was a compromise between the size of memory of the measuring instrument and the object access limitations for data download.
Lecture room had an amphitheatric layout and the dimensions $19 \times 8 \text{ m} \times (4-2.9) \text{ m}$. It was situated in the building from 70s. The room had one external wall, fitted with huge openable windows. Despite availability of mechanical ventilation, air exchange was predominantly realized by means of natural ventilation. In the period of measurements, teaching hours extended from 7:30 a.m. to 9:00 p.m. Classes were held during all working days and on majority of weekends (part time studies). Teaching blocks were typically 1.5 h or 45 min long with the brakes of 15 min in-between. Although designed for 90 students, the lecture room was hardly ever full. In the period of study, the number of listeners changed considerably within a single day as well as between days. It was due to a weakly lecture schedule as well as the general attendance variation. In the lecture theater CO$_2$ concentration was monitored in one point, in the period from 27.10.2013 to 22.01.2014. All measurement data were valid. The time resolution of data collection was 30 s. The maximum available resolution could be used because of an unlimited access to the room for data download.

CO$_2$ monitoring was realized with the instruments dedicated for continuous measurements and data logging. They are based on a non-dispersive infrared (NDIR) CO$_2$ sensor. Its measuring characteristics is as follows: measuring range $0-5000 \text{ ppm}$; accuracy $\pm 50 \text{ ppm} + 3\%$ of measured value and resolution: 1 ppm. This level of performance may be currently considered as a standard in continuous measurements of indoor air quality.

During indoor air study, the instrument was placed 1 m above the floor, out of direct influence of occupants, heating and ventilation system as well as outdoor environment (far from windows).

4. RESULTS AND DISCUSSION

Based on our approach, there were identified classes of variability of CO$_2$ concentration in indoor air. Further we demonstrate that their statistical description, the number of distinguished classes, their frequency of occurrence as well as the actual occurrence in time, all provide information on the stability of conditions in indoor air. This information may be applied when evaluating the capability of ventilation system to maintain stable conditions indoors. We used examples of extreme systems: the open space office, featured by the regular occupancy pattern and mechanical ventilation as well as the lecture theater, featured by the irregular occupancy pattern and natural ventilation.

Daily time series of CO$_2$ concentration in the open space office and in the lecture theater are shown in Figs. 2 and 3, respectively. Data referring to working days and days-off exhibited distinctive behavior and they were presented separately. As may be seen in Figs. 2b and 3b, in both objects, during days-off the CO$_2$ concentration was usually low and nearly invariant over entire period of 24 h. In these circumstances, indoor sources of CO$_2$ were absent, therefore the temporal variation of CO$_2$ concentration
indoors reflected the influence of ambient air. Similarly invariant indoor conditions were observed in the two objects during night time on working days, see Figs. 2a and 3a. Utterly different behavior of CO₂ concentration was characteristic for working days daytime. In this period, temporal variability of this parameter was considerable. Moreover, it had distinctive features in each of the examined indoor spaces. Considering the open space office (Fig. 2a), CO₂ concentration displayed quite regular daily rhythm of change. Three stages could be easily distinguished in this cycle. The first stage was a rapid increase of concentration in the morning hours, before the noon. Secondly, during the day (up to about 5:00 p.m.) there typically occurred fluctuations of concentration around the quite stable average level.

Finally, in the evening, there was visible the phase of concentration decay. The observed changes of CO₂ concentration were a resultant of the influence of humans (major source of CO₂ emission indoors) in conditions of constant air exchange rate and the
uncontrolled influence of all other factors, meaningful for CO₂ concentration. The essential phases of the indicated cycle of the CO₂ concentration changes were related to the daily schema of work that is typical of a corporation. The schema is fixed and consequently the associated CO₂ concentration change cycle was similar on distinct working days.

In the lecture theater, there was not observed a cycle of changes of CO₂ concentration during daytime of working days. The concentration showed raises and falls which varied on distinct days regarding their magnitude as well as time of the day (Fig. 3a). This could be associated with highly variant number of humans entering the lecture room, staying there and leaving it over the day. Obviously the attendance depends on the popularity of the individual classes. Moreover, the lecture schedule differs among weekdays, which is typical of the academic environment. Also the classes, and duration of breaks could vary a lot over time.

Based on the collected measurement data, in long-term perspective (4 months) there were distinguished classes of variation of CO₂ concentration. The total number of classes in the open space office was 18. For the lecture theater, there were recognized 34 classes. Considering the number of classes as indicative for the complexity of indoor conditions, the open space office represented a more stable indoor environment.

The frequency of class occurrence contributed with an additional information. As shown in Fig. 4, only one class dominated in each objects. It covered 50% of the study period in the open space office and over 70% of the measurements time in the case of the lecture theater. However, based on the hours of the day when this dominating class was observed (Figs. 5 and 6), it could be associated with conditions of lacking human influence on indoor air. In principle, this period was not interesting regarding the ability to maintain stable air quality indoors. Contrary, the remaining period was most important, when the CO₂ concentration was jointly influenced by the human factor and
ventilation system. Its ability to compensate human impact could be observed. This period constituted 50% of the study period in the open space office and 30% in the lecture theater. From our analysis, multiple types of CO₂ concentration variation were encountered during that time, 17 – in the open space office and 33 – in the lecture theater. The mentioned facts indicate that indoor conditions were more stable upon regular occupancy pattern and mechanical ventilation of an indoor space, compared with an irregular pattern of human presence and air exchange induced by a natural ventilation.

![Fig. 5. Occurrence of CO₂ concentration variation class 1 during: (a) working days and (b) days-off in the open space office](image)

![Fig. 6. Occurrence of CO₂ concentration variation class 1 during: (a) working days and (b) days-off in the lecture theater](image)

Descriptive characteristics of CO₂ concentration variation classes combined with their occurrence during the day, offer another contribution to the examination of the stability of indoor air condition in a particular space. The empirical cumulative distribution functions for six major classes distinguished in the open space office are presented in Fig. 7a and in the lecture theater – in Fig. 7b. The daily occurrence of classes is demonstrated in Fig. 8 for the open space office and in Fig. 9 for the lecture theater.
The distinguished classes described by the associated ECDFs (Fig. 7) could be easily interpreted in terms of counteracting impacts of the human factor and ventilation on indoor air. Class 1 was observed late at night, when the consequences of daytime human presence waned (Fig. 8). It was after people left the office and the ventilation completely removed the accumulated CO2. When people arrived to work, from 7:00 to 8:00 a.m., i.e. upon emerging emission, the variation of CO2 concentration was described by class 4 (Fig. 8). Interestingly, two initial hours of working time, from 8:00 to 10:00 a.m., were featured by a distinctive type of variation of CO2 concentration labeled as class 6 (Fig. 8). It applied to the situation when CO2 emission was not balanced by the removal through ventilation and this gas accumulated indoors. In a broad time interval, from 10:00 a.m. to 5:00–6:00 p.m., the variation of CO2 concentration belonged to class 2 (Fig. 8). It most likely represented the quasi-equilibrium between human influence (constant number of people and their activity during deep working hours) and ventilation. A separate class 5 (Fig. 8) of variation of CO2 concentration, falling on evening hours (6:00–7:00 p.m.), could be linked with the phenomenon of people leaving from work. Later hours, typically until midnight, could be associated with waning human influence on indoor air. Sometimes, this process extended to the early morning on a subsequent day, as shown by the time span of the corresponding class 3 (Fig. 8). Based on Fig. 8, often-times an individual class was observed longer than 1 h. Hence, in the open space office a particular type of indoor air variation was maintained for a relatively long period of time.

Compared with the open space office, in the lecture theater there was none well pronounced daily sequence of CO2 variation classes (Fig. 9). Based on their ECDFs, the distinguished classes (Fig. 7b) could be viewed as representing conditions of: 1) dominating CO2 delivery processes – class 4, 5 and 6, 2) domination of CO2 removal processes – class 3, and 3) dynamic equilibrium between the delivery and removal process – class 1 and 2. The first group could be associated with the time when students entered the classroom for lectures and the beginning of their presence.
The second group would involve periods of time when the classroom was emptied or more intensely weathered. Finally, the third group would refer to the periods when human impact was in equilibrium with the ventilation rate. But in principle, the method...
offered distinction finer than that, as demonstrated by the presence of many classes in each group. The individual classes most frequently occupied short periods of time, typically 15 min in turn. The layout of their daily occurrence varied among different working days.

Fig. 9. Occurrence of variation of the CO₂ concentration classes from class 2 to class 6, during working days in the lecture theater
5. CONCLUSIONS

Empirical cumulative distribution function (ECDF) of the changes of CO₂ concentration was considered as a descriptor of indoor air variability. We showed that using this tool, there may be performed the analysis of CO₂ variation in a short as well long term. The approach utilizes an idea of classification. A class of variation of CO₂ concentration is statistically described by its individual ECDF. It indicates a particular case of balance/imbalance between CO₂ delivery and removal processes including gas dispersion. The analysis is applicable for evaluating the stability of indoor environment and it may provide support for the diagnosis of the performance of the air exchange system.

The approach was demonstrated using examples of indoor environments which are very different regarding their stability, namely, the open space office and the lecture theater. Based on 4 months of continuous monitoring, we demonstrated that in the case of the open space office, the number of classes of variation of CO₂ concentration was eleven. Individual classes emerged in a regular, daily cycle and their duration was from one to several hours. For the lecture theater, there were distinguished 34 classes. None daily cycle of their occurrence could be noticed. Moreover, classes changed frequently, typically every 15 min.

The introduced approach qualitatively indicated that upon regular occupancy pattern and mechanical ventilation indoor environment is more stable compared with irregular pattern of human presence and air exchange induced by a natural ventilation.

In our work, we have shown that time series analysis based on the segmentation approach allows one to obtain information about variation of CO₂ concentration in spite of relatively poor accuracy of the measurement equipment.

ACKNOWLEDGMENTS

This work was financially supported by the National Science Center, Poland, under the contract No. DEC-2012/07/B/ST8/03031.

REFERENCES


