ISSN 1870-4069

A Physical Variable Data Fusion Approach as Basis for the Reasoning Process in Ambient Intelligence

Julio Muñoz Benítez, Guillermo Molero Castillo, Everardo Bárcenas, Rocío Aldeco Pérez, Alejandro Velázquez Mena

Universidad Nacional Autónoma de México,
Facultad de Ingeniería,
Mexico

juliomunoz@uv.mx, gmoleroca@fib.unam.mx, ebarcenas@unam.mx, raldeco@fib.unam.mx, mena@fib.unam.mx

Abstract. Recently, interest has increased in evaluating environmental quality, for extended periods, in closed spaces of human occupation. This is due to the search for well-being in the health of the users. Therefore, it is important to analyze the behavior of environmental conditions, acceptable to people, in certain periods and workspaces, in accordance with the standards established by the Health sector. This work describes the implementation of a network of sensors to acquire data on certain variables, such as temperature, humidity, and air pollution. In this sense, data fusion provides a way to unify data as support for the analysis of various sources. In addition, Bayesian estimation is presented as a data fusion method of the captured values. The results showed that the data was fused with high precision, which can be used to analyze healthy atmospheric levels for users in closed environments, as a basis for the reasoning process in Ambient Intelligence.

Keywords: Ambient Intelligence, Bayesian Estimation, Data Fusion, Physical Variables, Sensor Network.

1 Introduction

Ambient Intelligence (AmI) is an emerging area in computer science that provides intelligence to real-life environments, where daily activities are performed, making them sensitive to the user and their actions by the use of electronic devices, sensor networks, context awareness, and artificial intelligence [4, 7] At present, the interest in this area has increased due to advances in sensor networks and their integration into daily activities, as well as the growing need to perceive and analyze the environment in order to ensure healthy levels in areas of human occupation environments [11, 10].

Because AmI is designed for real-world environments, the effective use of wireless sensor networks has been important in a variety of applications, such as environmental monitoring. The objective is to obtain useful information for making decisions about a certain event, object, or action [3], where the information is used to identify patterns of

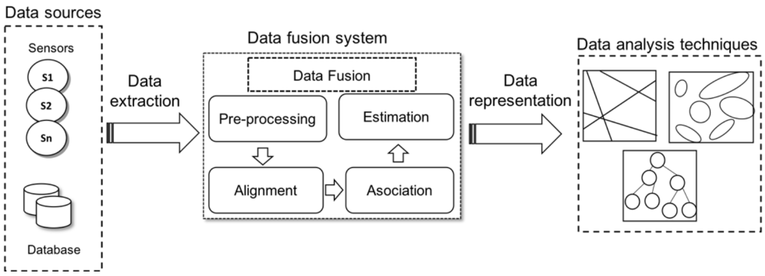

Fig. 1. Conceptual representation of a fusion system.

behavior over time, for example, to analyze concentrations of air pollution, temperature or humidity [14]. For this, it is necessary to analyze different data sources [19], such as those obtained through sensors.

Nevertheless, analyzing data collected through sensor networks is a complex challenge [4, 14, 19], as large volumes of data can be generated, making it difficult to analyze the captured information. Also, if the sensors are faulty or inaccurate, caused by ambient noise, missing or incorrect values can occur, leading to erroneous analysis or incorrect predictions [13]. Therefore, when using data from multiple sources, there must be a way to combine them, prior to analysis, since integrated data provide a comprehensive, unified, consistent, and accurate picture of a situation of interest.

In this sense, data fusion facilitates the detection, association, and integration of data [26], which come from different sources, such as signals from a sensor network [9]. This field of knowledge has been used in various areas, such as [9, 26]: signal processing, estimation, statistics, inference, and artificial intelligence. Today, it has significant advancements in applications for automatic target recognition, autonomous vehicle navigation, remote sensing, and threat identification.

As an example, Fig. 1 shows the diversity of sources, in this case, sensors, and the fusion process, where the data is integrated with the purpose of carrying out an analysis on a certain event.

One of the methods used in data fusion is Bayesian estimation [8, 15, 16, 17]. This method performs the fusion of the probability distribution of subsequent measurement, based on a set of probability distributions from previous measurements [15, 24]. That is, this is a way to represent the dynamism of the environment and the behavior of data sources [13, 15]. Thus, after data fusion, as part of AmI, the analysis of these fused data sources can be useful, for example for the analysis of environmental variables, according to established Health standards, such as the ANSI/ASHRAE 55-2017 (ASHRAE, American Society of Heating, Refrigeration and Air-conditioning Engineers), which defines the ambient thermal conditions for human occupation in a given workspace [1].

This paper presents the results of Bayesian estimation for the fusion of data from physical variables measured in a work area of a university environment. For the collection of data, humidity, temperature, and air pollution sensors were installed. This data source represents an important field of action as a basis for the generation of knowledge in AmI, specifically for the analysis of the thermal comfort of users in closed environments.

2 Background

At present, the areas that have increased the use of data fusion are data science, pattern recognition, and artificial and environmental intelligence [2, 5. 7], where fusion methods are used in the data extraction, integration, and loading in order to serve as a basis for their analytics.

Precisely, in AmI there is a growing need to perceive and analyze environmental conditions in order to improve the environment of users in the performance of their activities. One of the main characteristics of AmI is the use of electronic devices and communication techniques to perceive the environment, that is, to use sensor networks to monitor environmental conditions. One of these conditions of interest is the monitoring of humidity, temperature, and air pollution [20], which are atmospheric conditions important for the health and well-being of users.

For example, in the case of air pollution, it increases people's risk of contracting a respiratory disease [23, 27], given the presence of harmful particles, which not only depend on the gases generated by vehicles or gases emitted by industrial systems, also depends on environmental conditions such as humidity, temperature, rain, wind, to name a few. Therefore, these environmental conditions must be taken into account for their analysis and subsequent execution of an action plan for the benefit of people and the environment.

In this sense, data fusion is a key and critical aspect of systems with diverse data sources, such as sensors. The goal is to integrate data efficiently to overcome the limitations encountered when using a single source [10]. The fusion of data from various sources represents a significant advantage in the unification of data, to obtain a global, robust, and complete vision of the event or situation that is being observed and analyzed [25].

Data fusion has been used in various disciplines, such as pattern recognition, artificial neural networks, the decision-making process, to name a few. Currently, one of the areas that have driven the use of data fusion is AmI, where fusion methods are used in the extraction, integration, and loading of data to serve as support in the reasoning process from data collected from different sources [19].

On the other side, the analysis of thermal conditions is of importance since the World Health Organization (WHO) defines a set of diseases caused by environmental factors in closed spaces, called sick building syndrome [18, 21]. This syndrome is a set of annoyances caused by poor ventilation that prevents air renewal, temperatures not suitable for user activities, humidity levels that benefit the reproduction of micro-organisms, and the concentration of particles in the air. The symptoms of sick building syndrome affect users, generating discomforts such as [18]: headaches, nausea, dizziness, colds, irritation of the respiratory tract, eyes, skin, as well as persistent allergies.

Thus, thermal analysis is one of the fundamental aspects that determine the indoor environmental quality and is directly related to user satisfaction and energy use in buildings [22]. The purpose is to maintain healthy and adequate environmental levels that guarantee the comfort, health, and productivity of the users in the performance of their activities.

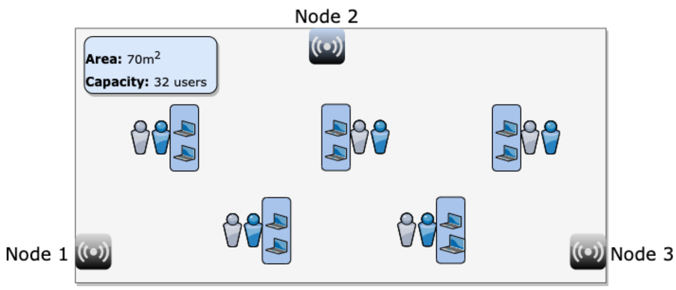

Fig. 2. Sensor network deployment in the workspace.

3 Method

As method, three work stages were defined, the first was the design and implementation of the sensor network for the measurement of environmental variables. In a second stage, the environmental variables were measured with a frequency of one minute, during a period of four weeks, that is, during the last weeks of spring. This period was defined as a period prior to summer, with the objective of obtaining data during the transition days of both seasons (spring and summer), that could be used to compare with a future analysis during winter. The third stage consisted of the use of Bayesian estimation as a data fusion method. This method was based on obtaining estimates based on previously collected data for temperature, humidity, and air pollution.

3.1 Sensor Network

The sensor network was implemented, providing it with sensing, processing, storage, and data transmission capabilities, in order to generate the data stream. Figure 2 shows the deployment of the sensor network in the physical space, where they were installed. Five nodes were installed, three to measure temperature, humidity, and air pollution (nodes 1, 2, and 3).

Each of these nodes was integrated into a NodeMCU board, which consists of an ESP8266 processor with integrated WiFi technology b/g/n [6]. Sensors were connected to the general-purpose inputs and the digital-analog converter, so data was processed and stored in the memory of the NodeMCU board. The sensors used were a) DHT11 for the measurement of relative humidity and temperature, and b) MQ-135 to determine air pollution and concentration of particles in the environment.

3.2 Data Capture

As mentioned, the nodes of the sensor network were configured to measure environmental variables with a frequency of one minute for four weeks. This period was defined according to the ASHRAE 55-2017 standard, which establishes that the acquisition intervals must be five minutes or less and the sampling period greater than two hours or be directly determined based on the occupation schedules of the area of interest [1].

Table 1. Extract of data collected through the implemented sensor network.

Date	Temperature (ºC)			Humidity (%)			Air Pollution (PPM)		
	S1	S2	S3	S1	S2	S3	S1	S2	S3
2020-05-25 07:01:00	24.0	24.0	24.0	45.0	48.0	47.0	83.9	83.3	81.6
2020-05-30 07:02:00	24.0	23.0	23.0	47.0	48.0	47.0	84.3	83.8	81.4
2020-06-21 18:59:00	24.0	24.0	23.0	45.0	47.0	47.0	77.6	76.3	67.0
2020-06-21 19:00:00	23.0	24.0	23.0	46.0	45.0	47.0	77.7	76.5	67.4

Regarding the data collected, these measurements were made in a university workspace, where according to the ASHRAE 55-2017 standard, the activities carried out are classified as typical office tasks, where there is a low metabolic rate, that is, a low level of transformation of chemical energy into heat [1].

In this sense, for data collection, each of the sensors handled a local log in comma-separated values (CSV) format. Sensor data is generated as raw data including date and time data, physical variables, and sensor identifier. This data is extracted and groped into a single table that serves as a source for the fusion method. Furthermore, with the use of WiFi technology, the data were grouped in a single table. As an example, Table 1 shows an extract of the data collected through the sensors on temperature, humidity, and air pollution.

Based on the measurements obtained, the temperature recorded values in degrees Celsius, the humidity as a percentage of water present in the environment, and the air pollution in parts per million.

3.3 Bayesian Estimation

After obtaining the data, Bayesian estimation was used as a data fusion algorithm, based on the Bayes theorem, which uses the conditional probability to estimate an event *X*, after a previous measurement *Y* [13, 15]. The probabilistic information of the measurement *Y* about the event *X* is also described by a probability density function *P*(*x*,*y*), known as conditional probability function:

$$P(x|y) = \frac{P(x) \cdot P(y|x)}{P(y)}.$$

In this sense, Bayesian estimation reflects the probability of the occurrence of measurement based on previous data. This probability can be updated as recent data measurements are made. In this way, when updating the probabilities of the measurements, Bayesian estimation can model the behavior of the environmental variables that generate data streams [12, 15, 19].

Thus, this paper presents the use of the conditional probability function to represent the estimated result from the fusion of sensor measurements, represented as $S_{sensor\ number}^{time}$. When replacing the probabilities of the measurements made by the sensors for Temperature (T), Humidity (H), and Air Pollution (C), we obtain:

$$P(T_i^1|T_i^0) = \frac{P(T_1^1) \cdot\ P(T_2^1) \cdot P(T_3^1) \cdot P(T_1^1|T_1^o) \cdot P(T_2^1|T_2^o) \cdot P(T_3^1|T_3^o)}{P(T_1^0) \cdot\ P(T_2^0) \cdot P(T_3^0)}, \tag{1}$$

$$P(H_i^1|H_i^0) = \frac{P(H_1^1) \cdot P(H_2^1) \cdot P(H_3^1) \cdot P(H_1^1|H_1^o) \cdot P(H_2^1|H_2^o) \cdot P(H_3^1|T_3^o)}{P(H_1^0) \cdot P(H_2^0) \cdot P(H_3^0)}, \quad (2)$$

$$(C_i^1|C_i^0) = \frac{P(C_1^1) \cdot P(C_2^1) \cdot P(C_3^1) \cdot P(C_1^1|C_1^o) \cdot P(C_2^1|C_2^o) \cdot P(C_3^1|C_3^o)}{P(C_1^0) \cdot P(C_2^0) \cdot P(C_3^0)}. \quad (3)$$

In this way, data fused were obtained based on the previous measurements on each of the environmental variables, that is, integrated data were obtained for temperature, humidity, air pollution, respectively. Subsequently, these integrated values will be used for the thermal comfort analysis, under the conditions of the ASHRAE 55-2017 standard.

4 Results

As a result of the fusion process, based on the Bayesian estimate, integrated values were obtained for each of the environmental variables. These integrated values were projected in graphs that represent the conditional probability distribution of previous and future measurements, which were obtained by the sensor network.

4.1 Bayesian Estimation

Figure 3, part A, shows the temperature values measured in nodes 1, 2, and 3, through DHT11 sensors. The measurements of sensors 1 (blue) and 2 (orange) have similar behavior, with slight variations (values of 24.5 and 24.9 ºC, respectively), while sensor 3 (green) presented a difference with respect to the other sensors (23.2 ºC). This could be due to the characteristics of the workspace and the activities carried out, that is, a closed work environment with ventilation through a window and the main access door. Thus, based on the measurements, integrated values (red) were obtained. These fused values represent a general behavior of the sensation of heat in the environment. The mean value of this fusion was 23.8 ºC, with a minimum and maximum of 19.6 and 28.7 ºC respectively. It is important to note, that the fusion takes into account the probability of the measured values, in this way, the fusion is not based on the average of the

Table 2. Estimated root-mean-square error for the fused data.

Variable	Sensor	RMSE
Temperature	1	1.429
	2	1.424
	3	1.677
Humidity	1	1.577
	2	3.113
	3	1.760
Air pollution	1	1.447
	2	1.459
	3	2.040

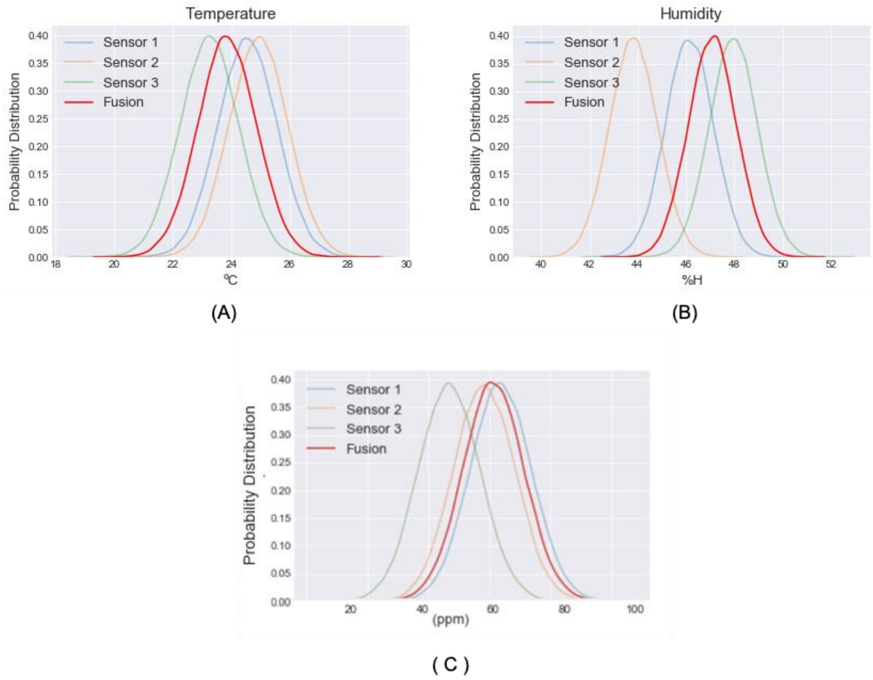

Fig. 3. Data fusion of variables temperature, humidity, and air pollution.

measurements, it represents the estimated values based on the probabilities of previous values.

Similarly, Figure 3 –part B– shows the values of the humidity percentage in the environment, acquired by the DHT11 sensors (nodes 1, 2, and 3). The measurements of sensor 1 (blue) and 3 (green) had high measurements, with the highest values recorded by sensor 3, presenting an average humidity sensation of 46 and 47.9%, respectively; while sensor 2 (orange) showed an average of 43.8%, recording lower values with respect to the measurements of sensors 1 and 3, the latter with the highest values.

This may be due to the location of the sensor, which is close to the window, where there is a greater amount of heat due to solar intensity and, at the same time, in contact with the exterior, concentrates a greater amount of water present in the environment.

Thus, based on humidity measurements, fused values (red) were obtained. These values were 47%, with a minimum of 42.8% and a maximum of 51.4%.

Air pollution, determined by levels of concentration of particles in the environment, is another of the physical variables that were analyzed due to its importance in determining healthy levels for closed environments. In Figure 3 –part C– it was observed that the values of sensors 1 (blue) and 2 (orange) present a similar probability distribution, with averages of particles per million (PPM) of 62 and 58, respectively. While sensor 3 (green) showed a behavior different from the other two sensors, whose PPM average was 47. The variation of this sensor is attributed to its location, which is close to a window that encourages airflow and ventilation. On the other hand, the result of the fused values (red) was influenced by measurements from sensors 1 and 2, with an average of 60 PPM.

4.2 Data Fusion Validation

As part of the data fusion validation process, the root-mean-square error (RMSE) was used. In this sense, the estimated data were compared with the measurements acquired by the sensor network. The summary of the results obtained for each variable is shown in Table 2.

Based on the results obtained, it was observed that the RMSE of data fusion reached low error levels. For example, for the case of measured temperatures, there are error percentages that range between 1.4 and 1.7%, which represents that the data fusion reached a high level of accuracy, that is, above 98.4%. These ensure that the precision of the data fusion method is trustworthy. A similar case occurs with the other measurements, whose levels of accuracy exceed 97.8% for humidity, 98.4% for air pollution. Consequently, the error results obtained allow us to validate, with a high confidence level, the Bayesian estimation method used for data fusion.

5 Conclusions

Data fusion is an area of research that has received great interest, due to the incorporation of sensors in computer systems and the need for these systems to interact with the environment that surrounds them, adapting to situations to provide appropriate services whenever necessary.

A sensor network was designed and implemented to collect data on physical variables in a certain university workspace. For this, DHT11 sensors were used to measure the percentage of water in the environment and the thermal sensation, and MQ-135 to measure the concentration of particles in the environment. These sensors were connected to a NodeMCU-type controller board, with storage and connection capabilities through WiFi technology, with the aim of storing information about the environment, exploiting the versatility that this board offers when programmed in the Python language.

Bayesian estimation was used as a fusion method, in order to obtain integrated data from previous measurements (real values). The purpose of obtaining fused values is to make subsequent analyzes about the thermal comfort of users in closed workspaces.

To validate the fusion method, the RMSE was used, whose quantification of the difference to the table of the fused values for each physical variable, with respect to the values acquired by the sensors, allowed to determine the level of accuracy of the data fusion. In this way, data fusion results were obtained, with a confidence level above 98% on average in each of the variables analyzed.

Environmental analysis is of importance to avoid unhealthy levels that propitiate the symptoms caused by the syndrome of the sick building, having as priority users' health in closed environments.

As future work, we contemplate expanding the extension of the sensor network to monitor other areas of interest, in order to analyze the thermal comfort of users in different work areas. Likewise, it is contemplated to extend the monitoring period to define work cycles or seasons in order to adapt parameters associated with each climatic season of the year.

Acknowledgment. This work was supported by UNAM-PAPIIT IA105320.

References

1. American Society of Heating, Refrigeration and Airconditioning Engineers (ASHRAE): Thermal Environmental Conditions for Human Occupancy Standard (2017)
2. Bernardos, A., Tarrio, P., Casar, J.: A data fusion framework for context-aware mobile services. In Multisensor Fusion and Integration for Intelligent Systems, pp. 606–613 (2008)
3. Bogaert, P., Gengler, S.: Bayesian maximum entropy and data fusion for processing qualitative data: theory and application for crowdsourced cropland occurrences in Ethiopia. Stochastic Environmental Research and Risk Assessment, 31(1) (2017)
4. Cook, D., Augusto, J., Jakkula, V.: Ambient Intelligence: Technologies, applications, and opportunities. Pervasive and Mobile Computing, 5(4), pp. 277–298 (2009)
5. De Paola, A., Ferraro, P., Gaglio, S., Re, G., Das, S.: An adaptive bayesian system for context-aware data fusion in smart environments. IEEE Transactions on Mobile Computing, 16(6), pp. 1502–1515 (2017)
6. Espressif Systems IOT Team: ESP8266EX Datasheet (2015)
7. Gellersen, H., Schmidt, A., Beigl, M.: Multi-sensor context-awareness in mobile devices and smart artifacts. Mobile Networks and Applications, 7(5), pp. 341–351 (2002)
8. Guerriero, M., Wheeler, F., Koste, G., Dekate, S., Choudhury, N.: Bayesian data fusion for pipeline leak detection. In: 19th International Conference on Information Fusion, 1(1), pp. 278–285 (2016)
9. Hall, D., Llinas, J.: An introduction to multisensor data fusion. Proceedings of the IEEE, 85(1), pp. 6–23 (1997)
10. Jenkins, M., Gross, G., Bisantz, A., Nagi, R.: Towards context aware data fusion: modeling and integration of situationally qualified human observations to manage uncertainty in a hard+soft fusion process. Information Fusion, 21, pp. 130–144 (2015)
11. Jin, L., Zhang, Y., Zhang, Z.: Human responses to high humidity in elevated temperatures for people in hot-humid climates. Building and Environment, 114, pp. 257–266 (2017)
12. Kabir, G., Demissie, G., Sadiq, R., Tesfamariam, S.: Integrating failure prediction models for water mains: bayesian belief network-based data fusion. Knowledge-Based Systems, 85, pp. 159–169 (2015)
13. Khaleghi, B., Khamis, A., Karray, F., Razavi, S.: Multisensor data fusion: a review of the state-of-the-art. Information Fusion, 14(1), pp. 28–44 (2013)
14. Khattak, A., Akbar, N., Aazam, M., Ali, T., Khan, A., Jeon, S., Hwang, M., Lee, S.: Context representation and fusion: advancements and opportunities. Sensors, 14(6), pp. 9628–9668 (2014)
15. Kumar, M., Garg, D.: Multi-sensor data fusion in presence of uncertainty and inconsistency in data. In: Milisavljevic, N. (ed.) Sensor and Data Fusion, pp. 225–244 (2009)
16. Lahat, D., Adaly, T., Jutten, C.: Challenges in multimodal data fusion. In: Signal Processing, pp. 101–105 (2014)
17. Lee, P., You, P., Huang, Y., Hsieh, Y.: Inconsistency detection and data fusion in USAR task. Engineering Computations, 34(1), pp. 18–32 (2017)
18. Maddalena, R., Mendell, M., Eliseeva, K., Chan, W., Sullivan, D., Russell, M., Satish, U., Fisk, W.: Effects of ventilation rate per person and per floor area on perceived air quality,

sick building syndrome symptoms, and decision-making. Indoor air, 25(4), pp. 362–370 (2015)
19. Muñoz-Benítez, J., Molero-Castillo, G., Benítez-Guerrero, E.: Data fusion architecture of heterogeneous sources obtained from a smart desk. In: Intelligent Data Sensing and Processing for Health and Well-Being Applications, pp. 23–40, Elsevier (2018)
20. Ohtsuki, T.: A smart city based on ambient intelligence. IEICE Transactions on Communications, 100(9), pp. 1547–1553 (2017)
21. Redlich, C., Sparer, J., Cullen, M.: Sick-building syndrome. The Lancet, 349(9057), pp. 1013–1016 (1997)
22. Schiavon, S., Hoyt, T., Piccioli, A.: Web application for thermal comfort visualization and calculation according to ASHRAE Standard 55. In: Building Simulation, 7, pp. 321–334, Springer (2014)
23. SEMARNAT: Programa de gestión para mejorar la calidad del aire en el estado de Veracruz de Ignacio de la llave 2016-2025 (2017)
24. Siciliano, B., Khatib, O.: Handbook of Robotics. Springer (2016)
25. Voida, S., Patterson, D., Patel, S.: Sensor Data Streams. In: Olson, J., Kellogg, W. (eds.) Ways of Knowing, pp. 291–322, Springer Science & Business (2014)
26. White, F.: Data Fusion Lexicon. Tech. Rep., Joint Directors of Laboratories (1991)
27. World Health Organization: Who guidelines for indoor air quality: selected pollutants (2010)