IOŚ-PIB
Environmental Protection and Natural Resources
Ochrona Środowiska i Zasobów Naturalnych
sciendo
Vol. 29 No 3(77): 7-12
DOI 10.2478/oszn-2018-0013

Grit Behrens*, Klaus Schlender*, Florian Fehring*

Data mining methods of healthy indoor climate coefficients for comfortable well-being

* FSP IFE - University of Applied Sciences Bielefeld;
e-mails: Grit.behrens@fh-bielefeld.de, Klaus.schlender@fh-bielefeld.de, Florian.fehring@fh-bielefeld.de

Keywords:
Big data, Web-based application, indoor air quality, machine learning, artificial intelligence, neural network, ventilation behaviour

Abstract
This article provides information about a currently developed measurement and analysis system 'Smart Monitoring', which is used on scientific project in terms of healthy indoor air coefficients, as well as the processing of the collected data for machine learning algorithms. The target is to reduce CO2 emissions caused by wrong ventilation habits in building sector after renovation process in older buildings.

1. INTRODUCTION

The German government has set itself ambitious climate protection targets: By 2020, greenhouse gas emissions have to fall 40% below the level at the state of 1990. Therefore, the building sector is aimed by this project as it is one of the large CO_2-producing sectors. The described work is performed with our partners from the renovation management of Bielefeld Sennestadt. The inhabitants live in pre-renovation buildings and they produce more carbon dioxide as normal, whilst losing heating energy caused by poor building state. After renovation and isolation, buildings lose their energy by wrong heating and ventilation habits.

Machine learning algorithms will help to change the heating and ventilation habits of residents by telling them recommendations to get better room climate and to aim lower CO_2 emission. We created machine learning algorithms that are basically learned by healthy air quality information according to common DIN standards [DIN EN ISO 7730; ASR… 2012; Detlef et al. 2013; DIN EN 13779] and further extended learned by personal feelings about air quality inside the apartments. Therefore, an algorithm **AI** has been developed, which is not only able to decide air quality according to German DIN standards $\mathbf{AI}_{DIN}$, which tells that the indoor climate is healthy or not, but also further with $\mathbf{AI}_{extended}$ if the occupant feels comfortable or not.

These algorithms could prevent unnecessary CO_2 emissions by telling recommendations to residents and influence their heating and ventilation habits in general. Also, they learn a correct and more energy-efficient ventilation and heating behaviour to reduce the CO_2 emission after a renovation process.

The Smart Monitoring system provides an interactive user interface inside the apartments. Residents can view information about actual measured data of the air quality in their apartments and also some statistical evaluations. The developed intelligent algorithms provide a prediction of how an occupant is normally feeling by interpreting the actual sensor data. It includes a preliminary processing of measured datasets, their physical relations, statistical analysis, data reduction, targeted learning of German DIN standards and the extended learning of individual indoor air climate feeling. The project is supported by ministry of North Rhine Westphalia (2016–2020) under Nr. 322-FH Struktur.

2. MATERIAL AND METHODS OF STUDY

The measuring system was developed for the purpose of promoting energy awareness amongst residents and to examine the ageing residential buildings before and after their renovation. It also shows the collected data to occupant's and provide collecting, managing and transferring Big Data by intelligent algorithms to support scientific work on collected data. Within an interdisciplinary team, a cost-effective measuring system was developed. In addition to expandability, scalability and maintainability, Smart Monitoring also provides secure data communication of measured sensors with strongly cryptographic algorithms.

As seen in Figure 1, the measuring system in apartments consists of a set of sensor computers (Raspberry Pi's). Each of them collects measurement data of sensors and actuators and sends them by POST request to a RESTful Webservice interface, which is controlled by main server (Figure 1, left side). The main server handles any type of data and saves it into a local database. The amount of sensor data and standalone computers is scalable. It is possible to have an unlimited amount of measuring systems that send encrypted data from each main server of an apartment to the main 'FH-Server' (Figure 1, right side). We have a 1:n relationship between a data receiving server and a scalable amount of data serving sensor computers in an apartment. Also, we could have a scalable amount of data collected in apartments, which provide sending encrypted data to the FH-Server by calling a special web service.

The FH-Server hosts a web application for data collection, analysis and visualisation. The focus of backend application was an efficient communication and storage of all data types. The handling of large datasets and data analysis algorithms allows us to support big data amounts in our system. All developed backend applications use standard open source technologies of Java EE, such as EJB, JNDI, JPA, JAXB and JAX-RS. JAX-RS are used with RESTful web service interfaces to manage the communication. In the JSON-based client server communication via Web service for high security level, a cryptography support was implemented [Behrens et al. 2017].

To measure datasets in different apartments, a standalone measuring system based on Raspberry Pi with Debian Linux is used **(figure 2)**. We use sensors to measure datasets of the relativity humidity $\mathbf{RH_{in}}$, pressure **p**, air temperature $\mathbf{T_{air}}$, inside temperature of the outer wall $\mathbf{T_{wall}}$, inside temperature of the heating $\mathbf{T_{heat}}$, carbon dioxide level $\mathbf{CO_2}$ and window ventilation states $\mathbf{ST_{window}}$, and A weather station measures the outside relativity humidity $\mathbf{RH_{out}}$, outside temperature $\mathbf{T_{out}}$ and outside wind speed $\mathbf{W_{out}}$, as shown in **Figure 2** [Behrens et al. 2017; Sonntag… 1990].

To measure the personal room climate feelings, a visual display unit (**figure 3**) is located inside monitored apartment room. Residents reported their air climate feelings **AQF** by pressing one of three feedback buttons with smiley icons – green, yellow and red. Feeling states good $\mathbf{AQF_{good}}$, neutral $\mathbf{AQF_{neutral}}$ and bad $\mathbf{AQF_{bad}}$ are generated, which are used for supervised machine learning algorithms.

After pressing a Feedback button as input data, the user gets a view of actual measured air quality data. For training $\mathbf{AI_{DIN}}$, we interpreted indoor air quality based on a given DIN standard as summarised in **Table 1**.

The German DIN standard [DIN EN ISO 7730; ASR… 2012; Detlef et al. 2013; DIN EN 13779] gives information about how the actual measurement data in a room should be **(Table 1)** to have a healthy room climate by interpreted categorisations ranges: $\mathbf{AQF_{good}}$, $\mathbf{AQF_{neutral}}$ and $\mathbf{AQF_{bad}}$.

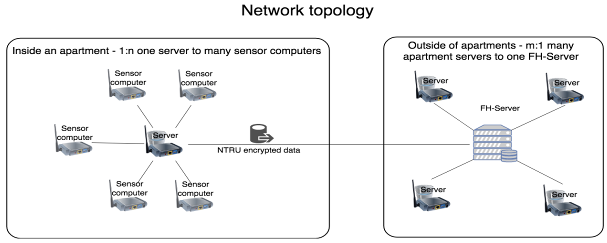

Figure 1. Network topology of measuring sensor systems.

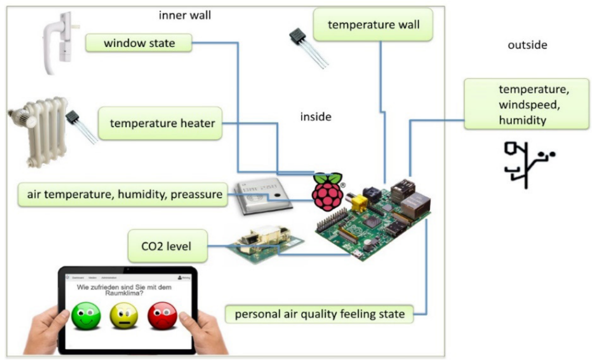

Figure 2. Example of a measuring system in an apartment.

Figure 3. Visual display unit logging personal room climate feelings AQF.

DIN standard deals with the sensors $\mathbf{RH_{in}}$ [ASR… 2012], $\mathbf{T_{air}}$ [Detlef et al. 2013], $\mathbf{CO_2}$ [ASR… 2012; DIN EN 13779] and $\mathbf{T_{wall}}$ [DIN EN ISO 7730]. Interpreting air quality state and room climate feeling as good, neutral or bad was made by sensor data fusion, getting multidimensional input feature vector to supervised learning algorithms $\mathbf{AI_{DIN}}$ and $\mathbf{AI_{extended}}$. Sensor data fusion means that all records made by the measuring system are merged by matching the timestamp. The $\mathbf{AI_{DIN}}$ algorithm is mainly used to train the neural network on almost all given datasets collected by sensors and then it serves as a base to learn the $\mathbf{AI_{extended}}$ algorithm by extending itself. Finally, $\mathbf{AI_{extended}}$ is based on the already learned $\mathbf{AI_{DIN}}$ as a base.

This method allows the algorithm to learn air quality ($\mathbf{AI_{DIN}}$) from all measured datasets (high frequency by sensors) and to further correct itself as a new branch by learning air feelings ($\mathbf{AI_{extended}}$) of inhabitants (low frequency by input).

Table 1. Ranges of sensor data as good, neutral and bad.

DIN standard ranges for winter							
Features	**Unit**	**Range good**		**Range neutral**		**Range bad**	
RHin	**[%]**	**40**	**68**	**40**	**62**	**<40**	**>70**
Tair	**[°C]**	**21**	**23**	**20**	**24**	**<20**	**>24**
CO_2	**[ppm]**	**0**	**1,000**	**1,000**	**2,000**	**2,000**	**>2,000**
Twall	**[°C]**	**13**	**23**	**23**	**35**	**>13**	**>35**
DIN standard ranges for summer							
Features	**Unit**	**Range good**		**Range neutral**		**Range bad**	
RHin	**[%]**	**40**	**62**	**40**	**55**	**<40**	**>70**
Tair	**[°C]**	**23.50**	**25.50**	**22**	**26**	**<22**	**>26**
CO_2	**[ppm]**	**0**	**1,000**	**1,000**	**2,000**	**2,000**	**>2,000**
Twall	**[°C]**	**13**	**23**	**23**	**35**	**>13**	**>35**

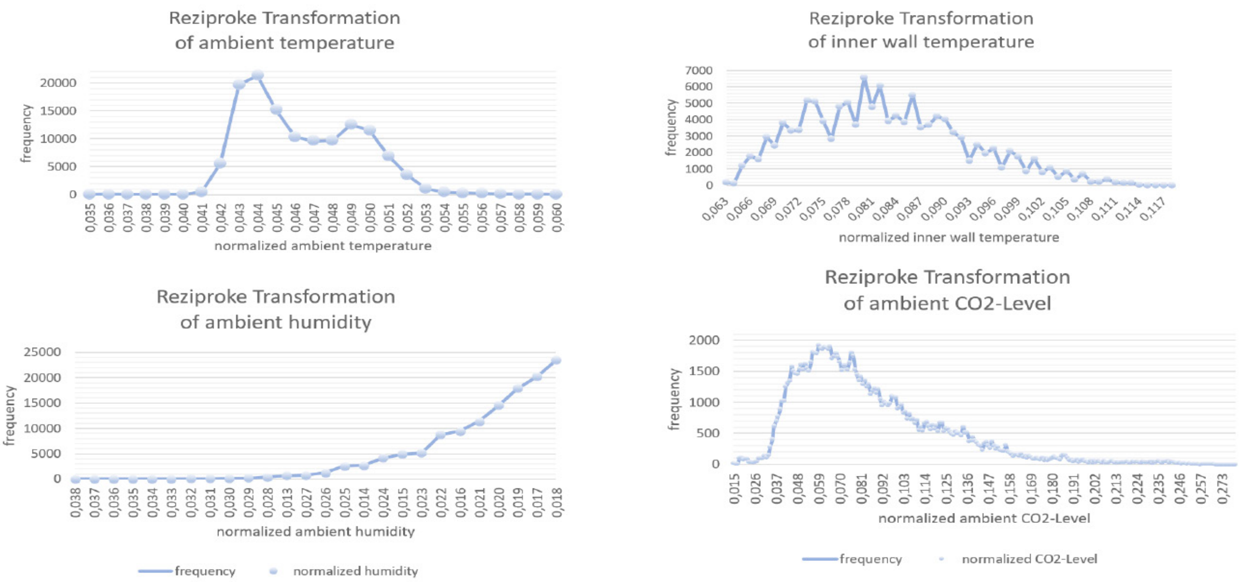

Figure 4. Interpolation and normalisation within reciprocal transformations.

3. RESULTS AND DISCUSSION

Supervised machine learning comes with a data mining process described in this section. We collected data from apartments and laboratory. After collecting datasets, we started pre-processing by filtering them within matching timestamps, cleaning up faulty data and normalising them **(Figure 4)**.

After pre-processing, we made data reduction and data breakdown to get data according to heating season (winter) and the remaining data (summer). The measuring system collected datasets at a rate of every minute in the winter period and at a rate of every half-minute in the summer period. The ranges are given in Table 2. Overall, we have used 129,047 merged records in the winter period (approximately September to March) and 138,914 merged records in the summer period (approximately April to June) to train the algorithms described next.

Basically, the overall supervised machine learning $\mathbf{AI}_{\mathbf{DIN}}$ and $\mathbf{AI}_{\mathbf{extended}}$ was made by using classified and randomly picked 70% of training data and 30% of testing data to create the **AI**s that are based on neural network approaches. The neural network used to train the algorithms was developed within this project and the chosen topology contained five layers. Between the first layer (an input layer) and the last layer (an output layer), we used three hidden layers,

Table 2. Overview of minima and maxima of collected measurement datasets for winter and summer season.

Features	Unit	Winter ranges		Summer ranges	
STwindow	[state]	0.00	2.00	0.00	2.00
RHin	[%]	14.27	77.10	35.71	67.45
p	[hPa]	960.33	1029.31	1006.13	1015.32
Tair	[°C]	15.89	28.66	21.48	28.93
Wout	[km/h]	0.00	11.40	0.10	2.35
Tout	[°C]	−10.50	17.60	14.70	32.70
RHout	[%]	22.20	99.00	30.70	86.20
CO2	[ppm]	350.00	6865.00	400.00	1230.00
Theat	[°C]	14.13	54.44	22.56	28.94
Twall	[°C]	8.50	27.31	24.13	32.06
AQF	[state]	0.00	2.00	0.00	2.00

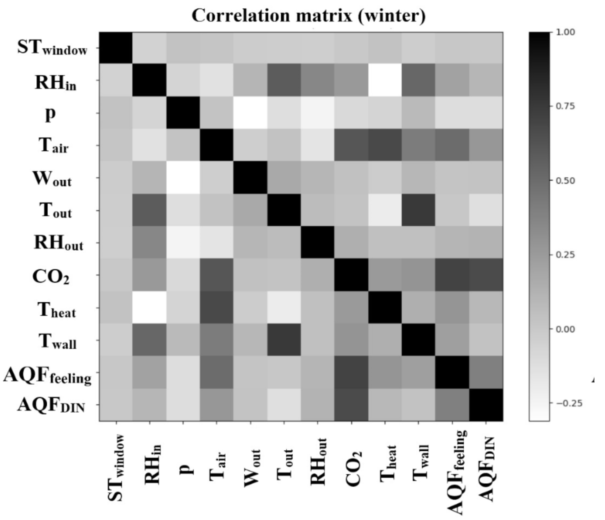

Figure 5. Correlation matrix on winter datasets.

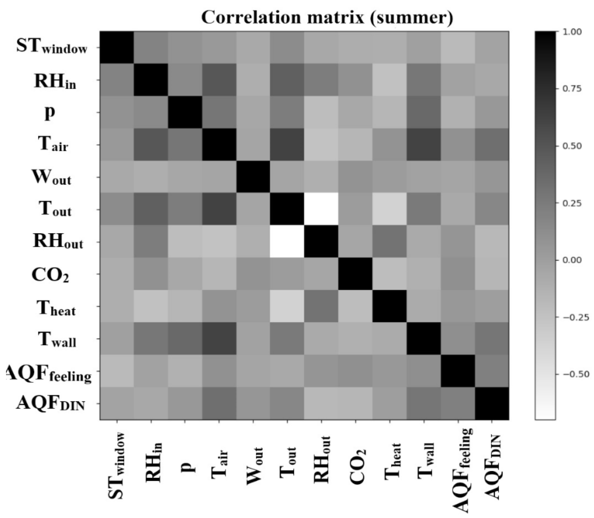

Figure 6. Correlation matrix on summer datasets.

of which the first holds 8 neurons, the second holds 16 neurons and the third holds 8 neurons.

Input feature vector (**ST_{window}, RH_{in}, p, T_{air}, W_{out}, T_{out}, RH_{out}, CO_2, T_{heat}** and **T_{wall}**) and target output vector (**AQF_{good}, $AQF_{neutral}$** and **AQF_{bad}**) were used, and a correlation matrix for each dataset is presented in **Figures 5 and 6**. The correlation matrix provides information about the relationship of sensor data to each other. The darker connection fields are more correlated than the lighter connection fields.

To differentiate the learned **AI**s, we named them by alias. **$AI_{DIN-winter}$** learned DIN standard at heating season, **$AI_{extended-winter}$** learned resident feelings at heating season, **$AI_{DIN-summer}$** learned DIN standard at not heating season and **$AI_{extended-summer}$** learned resident feeling at not heating season.

The first step was to create an intelligent **AI_{DIN}** that learns the DIN standard air climate quality classes. The second step was to use the already learned **AI_{DIN}** to create a second **$AI_{extended}$** that was additionally based on the interpretation of personal room climate feeling states (**AQF_{good}, $AQF_{neutral}$** and **AQF_{bad}**). In the conclusion, we got two kinds of **AI**s, **AI_{DIN}** and **$AI_{extended}$**. **AI_{DIN}** represents the actual state of the air quality inside an apartment for healthy purpose and **$AI_{extended}$** the actual state of personal feeling of an apartment resident. The classification experiments show the suitability of the learned algorithms for their determination (**Figures 7 and 8**).

The classification experiment 1 (winter) shows the algorithm **$AI_{DIN-winter}$**, which has learned the air quality conditions, and it made 126614 correct interpretations towards 129047 given datasets.

The classification experiment 2 (winter) shows the extended algorithm **$AI_{extended-winter}$**, which has learned the air feeling of residents, and it made 1,265 correct interpretations towards 1,274 given datasets.

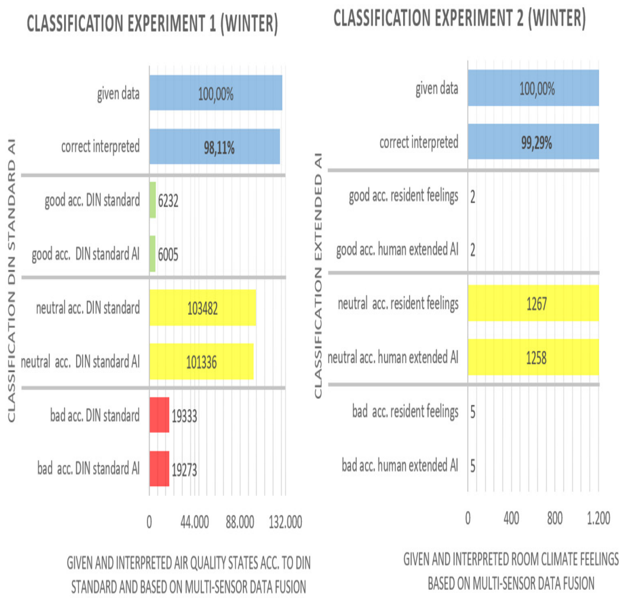

Figure 7. Comparing suitability of **AI**$_{\textbf{DIN-winter}}$ according to the DIN standard.

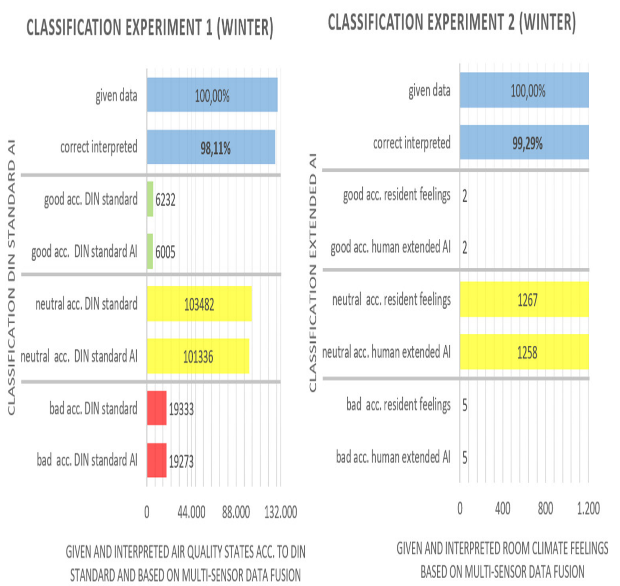

Figure 8. Comparing suitability of **AI**$_{\textbf{extended-winter}}$ according to the air feeling.

Similar classification experiments (**Figures 9 and 10**) were performed on the not heating season (summer) by learning the **AI**$_{\textbf{DIN-summer}}$ and extended learning of the **AI**$_{\textbf{extended-summer}}$. The classification experiment 1 (summer) shows the algorithm **AI**$_{\textbf{DIN-summer}}$, which has learned the air quality conditions, and it made 126,614 correct interpretations towards 129,047 given datasets.
The classification experiment 2 (summer) shows the extended algorithm **AI**$_{\textbf{extended-summer}}$, which has learned

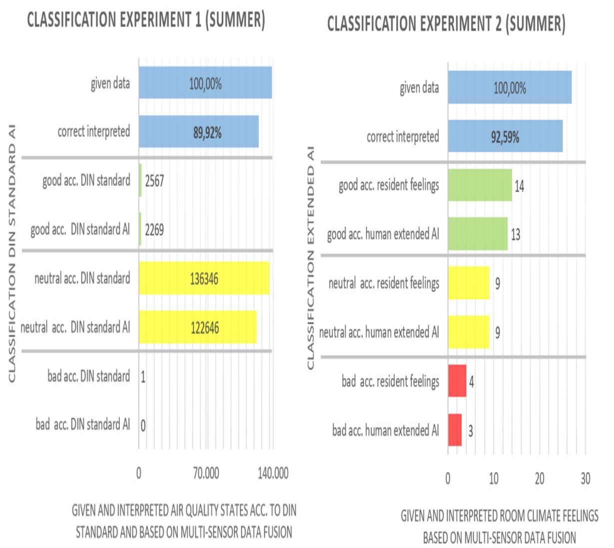

Figure 9. Comparing suitability of **AI**$_{\textbf{DIN-summer}}$ according to the DIN standard.

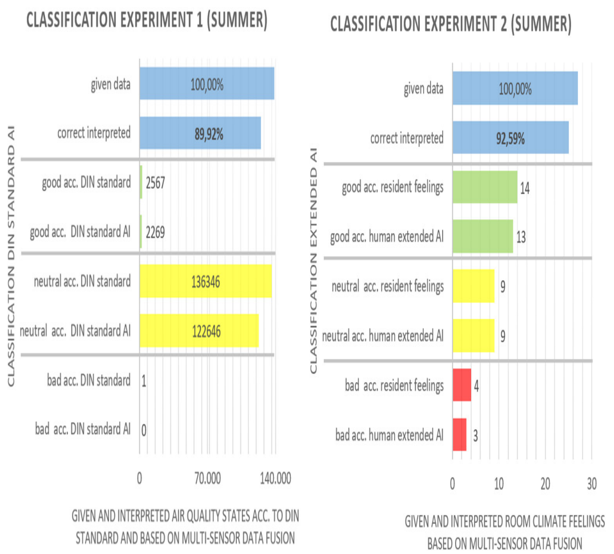

Figure 10. Comparing suitability of **AI**$_{\textbf{extended-summer}}$ according to the air feeling.

the air feeling of residents, and it made 1,265 correct interpretations towards 1,274 given datasets.

4. CONCLUSIONS

The developed smart monitoring system is based on RaspberryPi platform for measuring the air quality and wall humidity in apartment houses and delivers big sensor data. It recognises actual air quality in an apartment room by further telling us the actual personal air climate feeling of the residents.

In future works, a new academic research project based on the smart monitoring system developed in Bielefeld-Sennestadt for active assistance of the users during renovation process, called Environ and Approved, is needed. The benefit of this work is to use deep learned artificial intelligence for assistance purpose inside of apartments. Therefore, an apartment could be equipped with the algorithm AI_{DIN} (as the base algorithm) to serve information about actual air quality in the beginning and the resident could teach the second algorithm $AI_{extended}$ (as a branch) about his/her personal air feelings by sending input data. In addition to information on measured data, residents will also receive suggestions of how to interact with their environment to get better room climate and to reduce CO_2 emissions.

REFERENCES AND LEGAL ACTS

www.umweltbundesamt.de/themen/klima-energie, Klima und Energie.

BEHRENS G. et al. 2017. Smart Monitoring System of Air Quality and Wall Humidity Accompanying an Energy Efficient Renovation Process of Apartment Buildings. In: Otjacques B., Hitzelberger P., Naumann S., Wohlgemuth V. (eds) From Science to Society. Progress in IS. Springer, Cham, ISBN 978-3-319-65687-8: 181 – 189

SONNTAG, D. 1990. Important new values of the physical constants of 1986, vapour pressure formulations based on the IST-90, and psychrometer formulae. Z.Meteorol. 70 (5): 340-344

DIN DEUTSCHES INSTITUT FUER NORMUNG 2003. DIN EN ISO 7730 Ergonomie des Umgebungsklimas.

ASR A3.6 2012. Technische Regeln für Arbeitsstätten, www.baua.de/DE/Angebote/Rechtstexte-und-Technische-Regeln/Regelwerk/ASR/pdf/ASR-A3-6.pdf?__blob=publicationFile, (state 08.05.2018)

DETLEF MAKULLA, DIETMAR ROSSBRUCH 2013. Auslegungskriterien für thermische Behaglichkeit, www.heinze.de/media/65228/pdf/19726555px595x842.pdf, (state 02.05.2018)

DIN DEUTSCHES INSTITUT FUER NORMUNG 2005, E. 13779. Lüftung von Nichtwohngebäuden–Allgemeine Grundlagen und Anforderungen an Lüftungs-und Klimaanlagen.