A Visual Analytics Approach for Effective Radon Risk Perception in the IoT Era

Sérgio I. Lopes[1,2](✉), Sanne Bogers[3], Pedro M. Moreira[1], and António Curado[4]

[1] ARC4digiT - Instituto Politécnico de Viana do Castelo, Viana do Castelo, Portugal
sil@estg.ipvc.pt
[2] Instituto de Telecomunicações, Campus Universitário de Santiago, Aveiro, Portugal
[3] HAS - University of Applied Sciences, 's-Hertogenbosch, The Netherlands
[4] Prometheus - Instituto Politécnico de Viana do Castelo, Viana do Castelo, Portugal

Abstract. Radon gas is one of the most relevant indoor pollutants in areas of slaty and granitic soils, and is considered by the World Health Organization (WHO) as the second-largest risk factor associated with lung cancer. In the IoT era, active radon detectors are becoming affordable and ubiquitous, and in the near future, data gathered by these IoT devices will be streamed and analyzed by cloud-based systems in order to perform the so-called mitigation actions. However, a poor radon risk communication, independently of the technologies and the data analytics adopted, can lead to a misperception of radon risk, and therefore, fail to produce the wanted risk reduction among the population. In this work we propose a visual analytics approach that can be used for effective radon risk perception in the IoT era. The proposed approach takes advantage of specific space-time clustering of time-series data and uses a simple color-based scale for radon risk assessment, specifically designed to aggregate, not only the legislation in force but also the WHO reference level, by means of a visual analytics approach. The proposed methodology is evaluated using real time-series radon data obtained during a long-term period of 7 months.

Keywords: IoT · Visual analytics · Radon risk

1 Introduction

Along with other indoor air pollutants (smoke produced from solid fuel combustion, volatile organic compounds, etc.), radon gas is responsible for the degradation of air quality in enclosed rooms. However, the World Health Organization (WHO) considers indoor radon exposure as one of the most important causes responsible for lung cancer, right after tobacco smoking [1].

Regarding radon exposure in enclosed environments, the 2013/59/Euratom Directive imposes the so-called reference level of 300 $Bq.m^{-3}$ for the occupational exposure limit value [2]. All European legislation concerning ionizing radiation

H. Santos et al. (Eds.): SmartCity 360 2019, LNICST 323, pp. 90–101, 2020.
https://doi.org/10.1007/978-3-030-51005-3_10

exposure protection goes in the same direction as a result of the transposition of the referred Directive. Furthermore, in [1], the WHO recommends that countries adopt reference levels of $100\,\mathrm{Bq.m^{-3}}$ and if this level cannot be implemented under the prevailing country-specific conditions, WHO recommends that the annual average limit for indoor radon concentration in dwellings, offices, and workplaces must stay below the reference level of $300\,\mathrm{Bq.m^{-3}}$ otherwise, mitigation actions are required to remediate the non-regulatory rooms [2].

Nevertheless, the reference level of $300\,\mathrm{Bq.m^{-3}}$ is the base value to set off some remediation actions in order to reduce indoor radon concentration in a given room, the period of occupancy is a key variable. By way of example, an office where occupants stay on a daily basis for 8 working hours, exposed to an indoor radon level of $300\,\mathrm{Bq.m^{-3}}$, results in a higher risk than a technical room, with the same average level, where workers go there by one hour per day for maintenance purposes. In summary, it can be said that the indoor radon concentration taken in isolation cannot assess radon risk exposure since variables like buildings occupancy, the period of occupation and type of building are of vital importance on radon risk assessment.

Recently, several IoT-based radon detectors have been proposed, cf. [3–5], and in the near future, data gathered by these devices will be streamed and analyzed by cloud-based systems in order to perform the so-called mitigation actions. Having in mind that a poor radon risk communication can lead to a misperception of radon risk, and therefore, fail to produce the wanted risk reduction among the population. Given this, in this work we propose a visual analytics approach that can be used for effective radon risk perception when data is streamed continuously by these IoT radon detectors.

The remainder of this paper is organized as follows, in Sect. 2 a discussion about related works is undertaken, in Sect. 3 the visual analytics approach used for effective radon risk perception is introduced, in Sect. 4, the case study is presented in detail, and finally in Sect. 5, conclusions are pointed out and discussed.

2 Related Works

Recently, due to the rapid growth of IoT and Big Data technologies, large amounts of data from distinct varieties (timestamps, geolocations, sensor data, images, audio, video, etc.) have been produced. However, such data are not useful without analytic power [6]. Other analytics approaches, notably visual analytics methods, have been explored with success in the IoT and Big Data domains. Visual analytics methods aim to assist users in gaining insights, and therefore to extract knowledge from the data, by means of visual interpretations and interactions in the data analysis process [7]. Note that, an insight, can be seen as the ability of a user to understand a specific cause and effect within a specific context.

In [7], Keim et al. define visual analytics as the combination of automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets, that enable people to i) synthesize information and derive insights; ii) detect the expected and discover the unexpected; iii) provide timely, defensible, and understandable assessments and iv) communicate assessment effectively for action.

In this context, several recent works have been addressing the topic of IoT visual analytics, cf. [8–10], in order to assist users in the knowledge extraction process. In [8], the authors present the Virtual Open Operating System (vf-OS) approach to IoT Analytics, and describe its main components. The proposed approach can be used to capture data from IoT devices to generate and run machine learning models to perform data analytics, not only in the cloud but also on the edge. In [9], Lee et al. present a study that introduces a holistic perspective of storing, processing, and visualizing IoT-generated contents to support context-aware spatio-temporal insight. The study focus on the combination of deep learning techniques with a geographical mapping interface. Visualization is provided under an interactive web-based user interface to enhance the visual data exploration process, by means of a spatio-temporal query-based interface. In [10], the authors propose a framework for visual analytics of geospatial, spatio-temporal time-series data to handle multivariate, multiscale, and time-series data visualization. In the adopted design model they concluded that the most useful patterns are those that show relationships and aggregations of the data in both space and time domains.

3 Visual Analytics for Effective Radon Risk Perception

In order to extract knowledge from radon concentration data, first we need to understand the data under analysis in a space-time context, cf. Fig. 1. Understanding the data will help in the process of selection of appropriate data analysis models, and therefore assist in gaining insights, knowledge generation, and knowledge communication about the data [10].

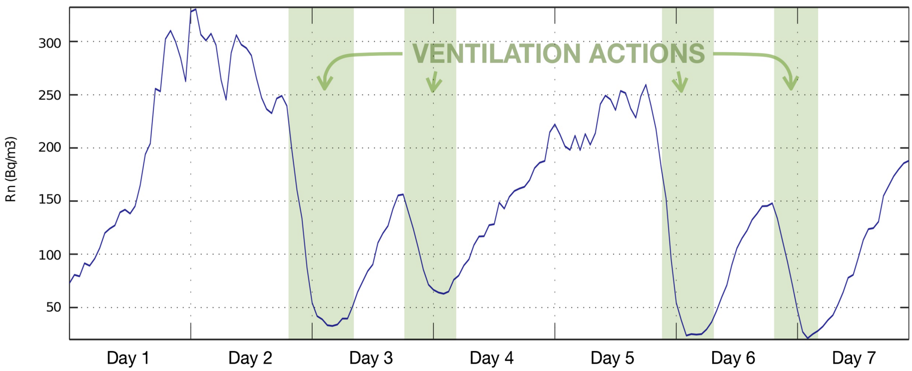

Fig. 1. Indoor radon fluctuation over a week with ventilation actions identified.

Indoor radon levels fluctuate over time depending on the building occupancy and the number of ventilation actions undertaken, cf. [11–13]. Commercial active radon detectors can perform continuous measurements in periods from 10 to 60 min. These detectors normally use an internal averaging mechanism to reduce data dispersion and therefore improve data quality. Figure 1 illustrates the variation over one week of the indoor radon concentration in a room with regular ventilation actions performed.

Moreover, indoor radon concentration is also affected by the space dimension, i.e. the soil composition and the building construction materials. Granitic soils and granitic construction materials both contribute, 80% and 20% respectively [1], to high indoor radon levels.

In our case it is expected that users can easily, and based on visual analytics, gain insights about radon risk exposure and the relations of practical situations such as building occupation and ventilation actions in the overall radon risk perception. Figure 2 depicts the proposed visual IoT analytics model that will be used in the visual analytics process.

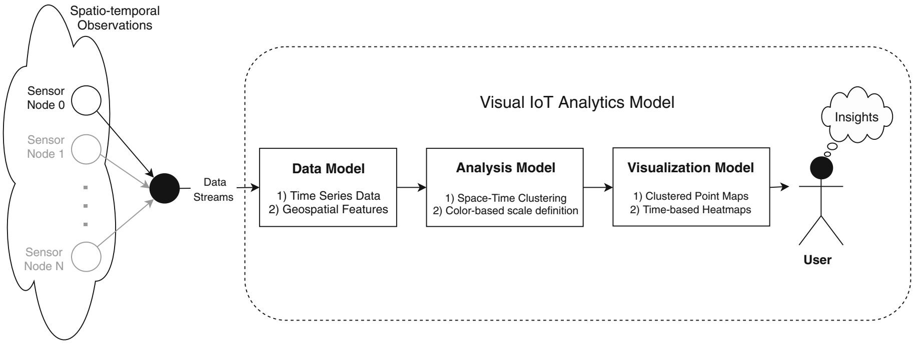

Fig. 2. Visual IoT Analytics Model: from spatio-temporal observations to user insights.

3.1 Data Model

Given the fact that multiple IoT devices can be used in distinct rooms and/or buildings, spatial context is key for radon concentration data exploration, and therefore, spatial-based clustering must be performed having in mind the relation of geographical entities, i.e. District > County > Building > Room > Device, as defined in [3], each having a specific set of geospatial features, such as soil composition, architectonic style, construction materials, etc. Moreover, time-series modeling is appropriate for radon concentration data when multiple devices geographically distributed are considered, not only for temporal clustering, but also for short-term prediction of indoor radon concentration.

3.2 Analysis Model

In [1], short-term measurements are defined as radon concentrations measurements that takes place over a period of not more than 3 months, and long-term measurements as radon concentrations measurements that take place over periods of 3 months up to 1 year. This definition was used as our baseline for clustering IoT time-series radon concentration data when multiple devices geographically distributed are considered. Multiple devices will generate time-series radon concentration data that will be difficult to analyse if no time clustering is performed. Given this, and having in mind the definitions previously introduced for short-term and long-term data clustering, we opted to use a more refined granularity containing five distinct time-based clustering approaches:

1) **RT** - Real Time (Hour);
2) **VST** - Very Short Term (Day);
3) **ST** - Short Term (Week);
4) **LT** - Long Term (Year), for periods always greater than 3 months.

In this analysis, the Euratom reference level of 300 $Bq.m^{-3}$ [2] and the WHO reference level of 100 $Bq.m^{-3}$ [1] were considered in the analysis model for scale definition and color selection, cf. Fig. 3, and therefore to enhance the visual analytics process.

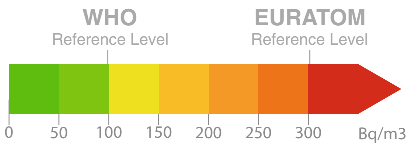

Fig. 3. Scale definition based on seven distinct colors mapped to the WHO [1] and Euratom [2] reference levels.

3.3 Visualization Model

Based on the geographical hierarchy introduced in Sect. 3.1, clustered point maps can be used and controlled by simple user-interface actions, such as zooming in to break a cluster (one point) in a subset of clusters (a group of new points), or zooming out to aggregate a set of clusters (a group of points) in a new cluster (one point).

To visualise radon concentration data in the time domain we opted to perform Very Short Term (VST), Short Term (ST) and Long Term (LT) time-based clustering through a heatmap visualization approach, were data is visualised through color variations in cells, enabling the easy assessment of its variance using distinct time-based clusters and the identification of relevant patterns.

4 Case Study

In this section, we present the evaluation of the proposed methodology using real data obtained with a certified Airthings Plus Radon detector, between 12/11/2018 and 30/06/2019 in the Lab.1.11 of the School of Technology and Management of the Polytechnic Institute of Viana do Castelo, cf. Fig. 4. The Lab is occupied regularly between 9h00 am and 5h00 pm. Radon concentration data is clustered in time, horizontally, using the periods defined in Sect. 3.2, VST, ST and LT. Vertically, and aligned from top to bottom is presented the evolution in time, week-by-week, with seasons identified.

Fig. 4. The placement of the radon detectors in the Lab.

When occupancy is considered, radon concentration time-series data should be considered for particular time schedules, i.e. when users are effectively exposed to the pollutant. Since that, in public buildings, offices, schools, kindergartens, etc, occupancy is normally restricted to regular schedules during working days, many of the time-series data values must not be considered in the computation of related radon risk metrics and indicators. Common time-series models (e.g. averaging/smoothing models) are inadequate in the case of intermittent time-series because many of the series values must not be considered. Since these models are based on weighted-summations of all past time-series data, they negatively bias the calculus of, not only radon concentration exposure metrics, but also, effective radon risk exposure indicators.

The evaluation is presented based on two time aggregation criteria, i) Very-Short Term (Day) and ii) Short-Term (Week). Additionally, two distinct data visualization approaches were produced, one considering the effective room occupancy and the other considering all the data gathered by the sensors.

4.1 Time-Based Heatmap Data Visualization

Figure 5 illustrates the variation of the radon concentration during 33 consecutive weeks. The data presented was acquired in the Lab and is used here for methodology validation based on two distinct scenarios: a) no occupation considered and b) with occupation considered. The Lab under analysis is a ground-floor office regularly occupied by three people, between 9h00 am and 5h00 pm, during working days. Figure 1 b) illustrates the average radon concentration is obtained directly from the occupancy profile previously defined.

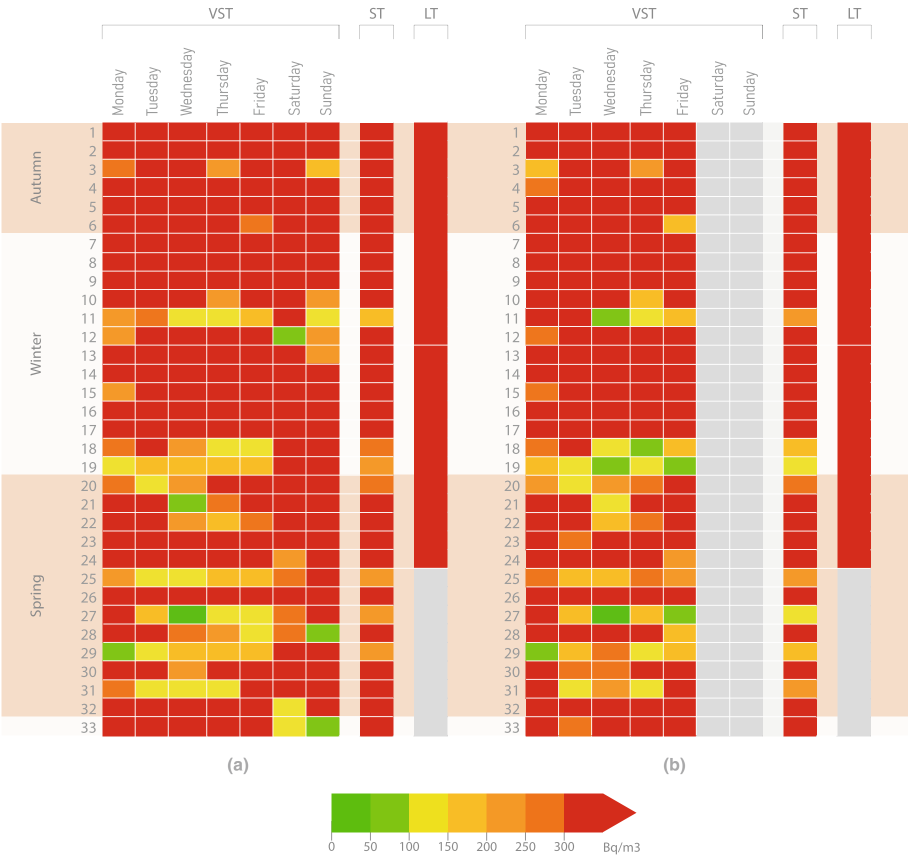

Fig. 5. Time-based Heatmap Visual Analytics: a) no occupancy considered; b) with occupancy considered.

Table 1 presents four evaluation metrics for easy comparison of the two scenarios introduced in Fig. 5. One observation is that scenario B regarding radon concentration data with occupancy considered between 9h00–17h00, results in a variation increase of metrics A and D which reveals that, the number of days below the WHO reference level [1] increased, and the number of days bellow the Euratom reference level [2] also increased. This observation also reveals that the risk perception tends to be overestimated if we look to the heatmap that considers all data, cf. Fig. 5a.

Table 1. Visual analytics performance metrics.

Metric	All data	Occupancy data	Variation	Reference level
	24 h	9h00–17h00		
A = $\sum$ Green cells/$\sum$ All cells	2.6%	4.2%	↑ 1.6%	≤100 Bq.m^{-3} (WHO [1])
B = $\sum$ not(Green cells)/$\sum$ All cells	97.4%	95.8%	↓ 1.6%	>100 Bq.m^{-3} (WHO [1])
C = $\sum$ Red cells/$\sum$ All cells	72.7%	70.9%	↓ 1.8%	>300 Bq.m^{-3} (Euratom [2])
D = $\sum$ not(Red cells)/$\sum$ All cells	27.3%	29.1%	↑ 1.8%	≤300 Bq.m^{-3} (Euratom [2])

4.2 User Evaluation

In order to validate the proposed approach, a set of user evaluation tests were conducted. The main goal of this evaluation test was to understand how users would read and perceive the proposed data visualization approach, and therefore, their ability to effectively perceive risk.

The evaluation protocol was based on the methodology presented in [14]. The evaluation protocol was based in two distinct documents: A) document used to introduce users to the Radon exposure problem, our case study and the main guidelines of WHO and the Portuguese legislation; and B) document with 11 questions, cf. Table 2 in which users have to answer about the visual analytics approach followed in this work, cf. Fig. 5. The questions were split in three main topics regarding Fig. 5 by considering heatmaps a), and b) alone, and considering both heatmaps at the same time.

Before the tests were conducted, the document A) was given to the users to explain the concept of the project to users. After this, the users had 3 min to look at the data visualizations and try to extract knowledge from them. Then, the questionnaire B), cf. Table 2, was handed to the participants. Subjects were then informed that, for each question, while reading the question until an answer was given, an independent observer would collect metrics on time duration and number of errors made. Finally, users were ensured that the evaluation process was about testing the visualizations and not themselves, giving them more confidence and comfort to freely answer the questions.

Table 2. Set of user evaluation questions.

ID	Heatmap	Question
1	a)	Give an example of a day where the Radon concentration was considered good by the WHO
2	a)	Indicate (one of) the best weeks regarding Radon concentration level
3	a)	Indicate the interval of the Radon concentration level observed on Thursday, week 18
4	a)	Indicate the interval of the Radon concentration level observed in week 20
5	a)	Indicate a week where the Radon concentration was better than the three months average
6	b)	Indicate the interval of the Radon concentration level observed on Wednesday week 11
7	b)	Indicate the interval of the Radon concentration level observed in week 19
8	a)+b)	Indicate one day in which the Radon concentration level was above the WHO recommendation level but below the same level, when occupation is considered
9	a)+b)	Indicate the day or days of the week for which the overall Radon Risk is considered higher
10	a)+b)	From all available data, what is the week with less Radon Risk exposure for the workers and what are the less risky days of that week?
11	a)+b)	What is the season with less Radon Risk exposure associated?

4.3 Results

The user evaluation was performed with 10 subjects aged between 22 and 48 years old, and was based on the protocol introduced in Sect. 4.2. Figure 6 depicts the statistical results regarding the users' response time through a standardized box plot representation. Additionally, in the same figure at right, the percentage of wrong answers was added. From the results presented in Fig. 6, one can observe that the spreading of the response time regarding questions 4, 6, 7 and 10B are the smallest. The results also show that most of the participants needed more time to answer question 8, with an average response time of 106 s. As shown in Fig. 6, this is the first question that regards both heatmaps. In this question, the user was asked to make connections between both heatmaps which naturally took more time to relate and gain an insight. It seems that this question was quite hard for the majority of the users, due to the fact that only half of the users answered this question correctly. Question 6 and 7 were answered rather quickly, i.e. all users answered this question in less than one minute. This relative quick response can be due to the fact that both questions are similar to question 3 and 4, which also resulted in a better success rate. Question 6 was even better understood by the users than question 3, i.e. every user answered question 6 correctly while 2 users answered question 3 wrongly. This shows that users were generally able to answer more quickly because previously gained insight earlier questions. While the users were conduction the test, it was observed that the difference between the time scales were sometimes unclear (VST-ST-LT) and it took the user some time to figure out which time scale was relevant for the question.

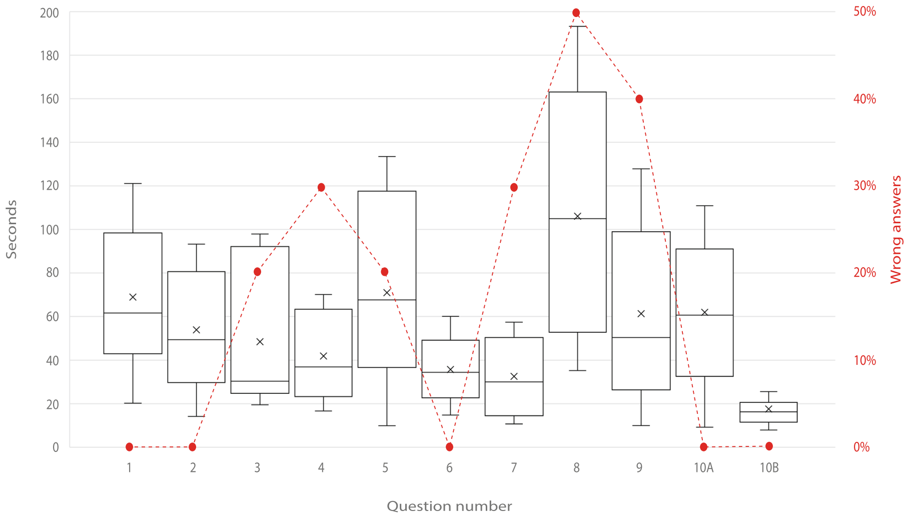

Fig. 6. User evaluation results, response time and percentage of wrong answers.

5 Conclusions

In this work, we proposed a visual analytics approach that can be used for effective radon risk perception in the IoT era. The proposed approach took advantage of specific space-time clustering of time-series data and used a simple color-based scale for radon risk assessment, specifically designed to aggregate, not only the legislation in force but also the WHO reference level.

The field results obtained after evaluation with users show that 83% of the overall questions were answered correctly with an overall average response time of 49 s. Another relevant observation was regarding similar questions made intermittently, resulting in a considerable reduction of the response time and also in a better success rate. Moreover, this study revealed that the performance of the proposed visual analytics method increases when occupancy is considered because when considering the heat map with all data available, we are inducing users to overestimate radon risk, and therefore emphasize their risk perception.

The study allowed to conclude that a proper radon risk communication is key for an effective radon risk perception, which results in a natural increase of the radon risk awareness among the population. As a consequence of this awareness increase, an overall radon risk reduction can be achieved based on these two main factors: i) increase of regular ventilation actions and ii) performing proper building occupancy management.

Acknowledgements. This contribution has been developed in the context of the Project "RnMonitor - Online Monitoring Infrastructure and Active Mitigation Strategies for Indoor Radon Gas in Public Buildings on the Northern Region of Portugal (Ref. POCI-01- 0145-FEDER-023997)" funded by FEDER (Fundo Europeu de Desenvolvimento Regional) through Operational Programme for Competitiveness and Internationalization (POCI).

References

1. WHO - World Health Organization: Handbook on Indoor Radon: A Public Health Perspective. World Health Organization (2009). ISBN 978-92-4154-767-3
2. European Union 2013: Council Directive 2013/59/EURATOM. E.U. Official Journal, 5 December 2013
3. Lopes, S.I., et al.: On the design of a Human-in-the-Loop Cyber-Physical System for online monitoring and active mitigation of indoor Radon gas concentration. In: 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA, pp. 1–8 (2018). https://doi.org/10.1109/ISC2.2018.8656777
4. Lopes, S.I., Pereira, F., Vieira, J.M.N., Carvalho, N.B., Curado, A.: Design of compact LoRa devices for smart building applications. In: Afonso, J.L., Monteiro, V., Pinto, J.G. (eds.) GreeNets 2018. LNICST, vol. 269, pp. 142–153. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12950-7_12
5. Blanco-Novoa, O., Fernández-Caramés, T.M., Fraga-Lamas, P., Castedo, L.: A cost-effective IoT system for monitoring indoor radon gas concentration. Sensors **18**, 2198 (2018). https://doi.org/10.3390/S18072198
6. Marjani, M., et al.: Big IoT data analytics: architecture, opportunities, and open research challenges. IEEE Access **5**, 5247–5261 (2017). https://doi.org/10.1109/ACCESS.2017.2689040
7. Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: definition, process, and challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 154–175. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70956-5_7
8. Anaya, V., Fraile, F., Aguayo, A., García, O., Ortiz, Á.: Towards IoT analytics. A vf-OS approach. In: 2018 International Conference on Intelligent Systems (IS), Funchal, Madeira, Portugal, pp. 570–575 (2018). https://doi.org/10.1109/IS.2018.8710476
9. Lee, J., Kim, K.-S., Lee, R., Lee, S.-H.: Visual insight of spatiotemporal IoT-generated contents. In: Proceedings of the 2018 International Conference on Advanced Visual Interfaces (AVI 2018), vol. 70, p. 3. ACM, New York (2018). https://doi.org/10.1145/3206505.3206575

10. Sibolla, B.H., Coetzee, S., Van Zyl, T.L.: A framework for visual analytics of spatio-temporal sensor observations from data streams. ISPRS Int. J. Geo-Inf. **7**(12), 475 (2018). https://doi.org/10.3390/ijgi7120475
11. Curado, A., Silva, J.P., Carvalho, L., Lopes, S.I.: Indoor Radon concentration assessment in a set of single family houses: case study held in Barcelos, North of Portugal. Energy Procedia **136**, 109–114 (2017). https://doi.org/10.1016/j.egypro.2017.10.295
12. Lopes, S.I., Silva, J.P., Antão, A., Curado, A.: Short-term characterization of the indoor air Radon concentration in a XII century monastery converted into a school building. Energy Procedia **153**, 303–308 (2018). https://doi.org/10.1016/j.egypro.2018.10.036
13. Curado, A., Silva, J.P., Lopes, S.I.: Radon Risk Management in public buildings in northwest Portugal: from short-term characterization to the design of specific mitigation actions. Int. J. Recent Technol. Eng. **8**(1), 90–96 (2019)
14. Nascimento, T., Gama, S.: FishEye: marine species' recognition and visualization: 24º Encontro Português de Computação Gráfica e Interação (EPCGI). In: Guimaraes 2017, pp. 1–8 (2017). https://doi.org/10.1109/EPCGI.2017