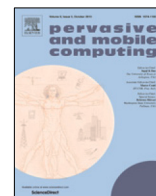




Contents lists available at ScienceDirect

# Pervasive and Mobile Computing

journal homepage: [www.elsevier.com/locate/pmc](http://www.elsevier.com/locate/pmc)

## Dense Indoor Sensor Networks: Towards passively sensing human presence with LoRAWAN<sup>☆</sup>

Jascha Grübel<sup>a,b,c,\*</sup>, Tyler Thrash<sup>d</sup>, Leonel Aguilar<sup>a,e</sup>, Michal Gath-Morad<sup>a,f,g</sup>,  
Didier Hélat<sup>h</sup>, Robert W. Sumner<sup>b,i</sup>, Christoph Hölscher<sup>a</sup>, Victor R. Schinazi<sup>j</sup>

<sup>a</sup> Chair of Cognitive Science, ETH Zürich, Clausiusstrasse 59, 8092 Zürich, Zürich, Switzerland

<sup>b</sup> Game Technology Center, ETH Zürich, Stampfenbachstrasse 114, 8092 Zürich, Zürich, Switzerland

<sup>c</sup> Visual Computing Group, Harvard University, 150 Western Avenue, Allston, MA 02134, United States of America

<sup>d</sup> Department of Biology, Saint Louis University, 1008 Spring Avenue, St. Louis, MO 63110, United States of America

<sup>e</sup> Data Science, Systems and Services Group, ETH Zürich, Stampfenbachstrasse 114, 8092 Zürich, Zürich, Switzerland

<sup>f</sup> Department of Architecture, University of Cambridge, 1 Scroope Terrace, Cambridge CB2 1PX, United Kingdom

<sup>g</sup> Bartlett School of Architecture, University College London, 22 Gordon St, London WC1H 0AY, United Kingdom

<sup>h</sup> OrbiWise SA, Route de la Galaise 34, 1228 Plan-Les-Ouates, Geneva, Switzerland

<sup>i</sup> Disney Research|Studios, Stampfenbachstrasse 48, 8006 Zürich, Switzerland

<sup>j</sup> Department of Psychology, Bond University, 14 University Dr, Robina, Queensland, 4226, Australia

### ARTICLE INFO

#### Article history:

Received 30 September 2021

Received in revised form 25 March 2022

Accepted 14 June 2022

Available online 18 June 2022

#### Keywords:

Dense Indoor Sensor Network

LoRAWAN

Fused Twins

Human presence

Effective Signal Power

### ABSTRACT

Sensors have become ubiquitous in buildings but are rarely connected to a network, and their potential to analyse the performance, use, and interaction with a building is not yet fully realised. In the coming years, we expect sensors in buildings to become part of the Internet of Things (IoT) and grow in numbers to form a Dense Indoor Sensor Network (DISN) that allows for unprecedented analysis of the performance, use, and interaction with buildings. Multiple technologies vie for leading this transformation. We explore Long Range Wide Area Network (LoRAWAN) as an alternative for creating indoor sensor networks that extends beyond its original long-distance communication purpose. For the present paper, we developed a DISN with 390 sensor nodes and four gateways and empirically evaluated its performance for two years. Our analysis of more than 86 million transmissions revealed that DISNs achieve a much lower distance coverage compared to estimations from previous research indicating that more gateways are required. In addition, the deployment of multiple gateways decreased the loss of transmissions due to environmental and network factors. Given the complexity of our system, we received few colliding concurrent messages, which demonstrates a gap between the projected requirements of LoRAWAN systems and the actual requirements of real-world applications given sufficient gateways. We also contribute to the modelling of transmissions with our comparison of attenuation models derived from multiple methodologies. Across all models, we find that robust coverage in an indoor environment can be maintained by placing a gateway every 30 m and every 5 floors. Finally, we also investigate the application of DISNs for the passive sensing and visualisation of human

<sup>☆</sup> This research is funded by ETH Zürich under grant ETH-15 16-2 and supported by OrbiWise with hard- and software. J.G. was supported by an ETH Zurich Doc.Mobility Fellowship.

\* Corresponding author at: Chair of Cognitive Science, ETH Zürich, Clausiusstrasse 59, 8092 Zürich, Zürich, Switzerland.

E-mail addresses: [jgruebel@ethz.ch](mailto:jgruebel@ethz.ch) (J. Grübel), [tyler.thrash@slu.edu](mailto:tyler.thrash@slu.edu) (T. Thrash), [leonel.aguilar@gess.ethz.ch](mailto:leonel.aguilar@gess.ethz.ch) (L. Aguilar), [michal.gath@gess.ethz.ch](mailto:michal.gath@gess.ethz.ch) (M. Gath-Morad), [didier.helat@orbiwise.com](mailto:didier.helat@orbiwise.com) (D. Hélat), [sumner@disneyresearch.com](mailto:sumner@disneyresearch.com) (R.W. Sumner), [choelsch@ethz.ch](mailto:choelsch@ethz.ch) (C. Hölscher), [victor\\_schinazi@bond.edu.au](mailto:victor_schinazi@bond.edu.au) (V.R. Schinazi).

URLs: <https://gtc.inf.ethz.ch>, <https://cog.ethz.ch> (J. Grübel), <https://https://www.perennialplantdiversity.org/people.html> (T. Thrash), <https://cog.ethz.ch> (L. Aguilar), <https://cog.ethz.ch> (M. Gath-Morad), <https://www.orbiwise.com> (D. Hélat), <https://gtc.inf.ethz.ch> (R.W. Sumner), <https://cog.ethz.ch> (C. Hölscher), <https://research.bond.edu.au/en/persons/victor-schinazi> (V.R. Schinazi).

<https://doi.org/10.1016/j.pmcj.2022.101640>

1574-1192/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

presence using a Digital Twin (DT) and a Fused Twins (FT) representation in Augmented Reality (AR). A passive sensing approach allows us to gather relevant data on human use of a building while still preserving privacy via the aggregation process. Immersive *in situ* visualisations in FT allow for new interactions and new forms of participation. We conclude that DISNs are already technologically feasible today and basing them on Low Power Wide Area Network (LPWAN) offers intriguing possibilities to reduce energy consumption, maintenance cost, and bandwidth use while also enabling new forms of human-building interaction.

© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Internet of Things (IoT) is increasingly permeating spaces, and the increasing density of devices allows researchers to move from specific applications to the analysis of the overall system in a Digital Twin (DT). In this context, a Dense Indoor Sensor Network (DISN) based on Low Power Wide Area Network (LPWAN), as often found in smart buildings and smart cities, provides opportunities to estimate human presence in the wild and their interaction with the building. Originally, LPWANs were developed to cover large outdoor distances with a robust signal at low cost [1–8]. In a LPWAN setup, devices are not easily accessible and must operate with limited maintenance, without external energy supply, and with limited capacity for data transmission [9]. Notably, the features of LPWAN that enable robust long-distance outdoor communication are also attractive indoors because of the low additional overhead for existing infrastructure and easier management of densely crowded Bandwidth (BW). In public spaces, many other communication devices based on BLE and WiFi crowd some Carrier Frequencies (CFs), and shifting IoT transmission to another BW can be beneficial. However, one drawback of LPWAN implementations is that a low Data Rate (DR) may be further diminished in order to increase signal quality.

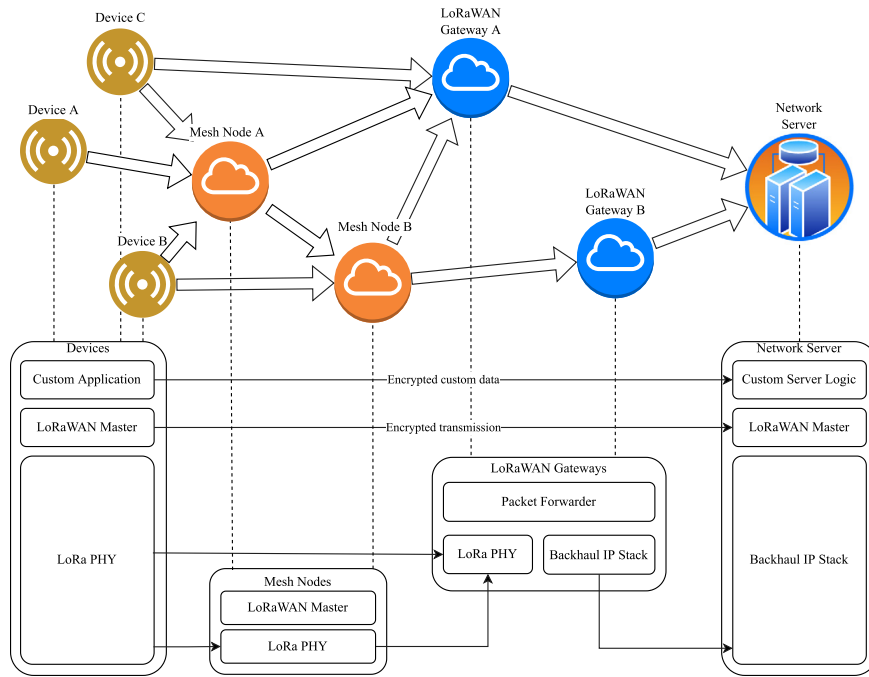
Long Range Wide Area Network (LoRAWAN) [10] has become the most common LPWAN standard [11] and is gaining attention [9] in both research and industry for three reasons. First, LoRAWAN is an open standard, and the only proprietary component from Semtech is a low-cost LoRa Transceiver (LoRa PHY) [10]. Second, LoRAWAN uses the unlicensed Industrial, Scientific, and Medical Frequency Bands (ISM) in contrast to expensive cellular networks that use 5G. Third, LoRAWAN is reliable and easy to use because of its design features such as simple circuitry, robust transmission signal, encryption, and low costs (see Fig. 1).

By allowing for DISNs, LoRAWAN may be suitable for studying human presence at a high spatiotemporal resolution [14]. Indoor human presence can be measured actively or passively [15] (see Fig. 2). Whereas active tracking requires participants to wear sensors [16], passive tracking infers human presence from changes in environmental factors measured by stationary sensors [17]. Usually, a single sensor per room is used to measure human presence as an indication of occupancy [17]. Inferring human presence can be facilitated with a DT [18,19] underpinned by LoRAWAN as a Service-oriented Architecture (SoA) [20] middleware. A DT is a platform for digitally mirroring processes from the physical world [21–23] that is suitable for continuously evaluating human presence indoors [24] and to study human building interaction [25]. DTs can be visualised in Augmented Reality (AR) with the Fused Twins (FT) paradigm [26,27] by which the data is brought back into the physical twin to be placed *in situ* where they were recorded.

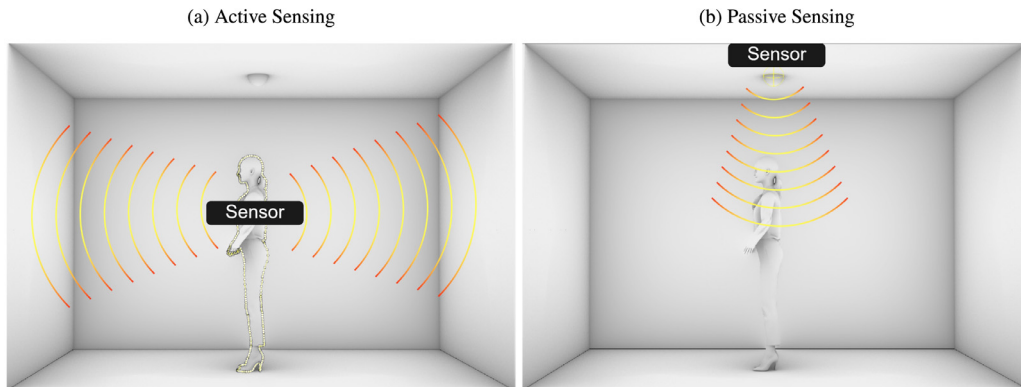
In general, the discussion of dense LoRAWAN applications has been theoretical [28], and empirical testbeds for LoRAWAN with a large number of sensors are scarce [29], although an outdoor scenario with 50 devices [30] and an indoor scenario with 331 devices [31] have been recently tested. Nonetheless, many questions on the scalability of LoRAWAN remain open, especially for DISNs, including the number of sensor nodes that are feasible indoors, the attenuation of the signal by the environment (e.g., walls, floors), the best model for evaluation, and the overall feasibility of LoRAWAN as part of a continuously running system.

The present paper is the first to present long-term evaluations of the technical feasibility of a LoRAWAN DISN with 390 nodes and 4 gateways in an office building over a prolonged period of nearly 2 years. We expand significantly on a first report of the technical setup [32] by applying additional methods to compare signal quality such as Effective Signal Power (ESP) and the modelling of the transmissions with explanatory (linear models, mixed models, Principal Component Regression (PCR), and Partial Least Squares Regression (PLSR)) and predictive models (Multi-Layer Perceptron (MLP), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer). Using Adaptive Data Rate (ADR), our DISN is consistently capable of maintaining a strong signal on all DRs at distances up to 30 m and 5 floors for a single gateway. At the lowest DR, theoretical distances up to 80 m indoors are possible, but each additional floor halves this distance, limiting sensible setups to 5 floors across all indicators of signal quality. With 390 sensors on 8 floors and up to 64 m of distance from sensor to gateway, the system received an average of 192000 transmissions daily. The percent of concurrent messages (i.e., on different CFs and Spreading Factors (SFs)) was 3.99%, but the percent of colliding concurrent messages (i.e., on the same CF and SF) was negligible at 0.39%.

In this paper, we also explore the effects of increasing gateway density in the building by adding a well-placed fourth gateway to the previous setup which allowed us to reduce the Frame Error Rate (FER) by 73.69% compared to the setup



**Fig. 1.** Overview of LoRaWAN. The Standard [10] includes devices, gateways, and a network server. New research includes forwarding of LoRa PHY transmissions through mesh nodes [12,13] when a gateway is out of reach. The data is encrypted both on the network level and the application level, allowing external parties to forward transmissions.



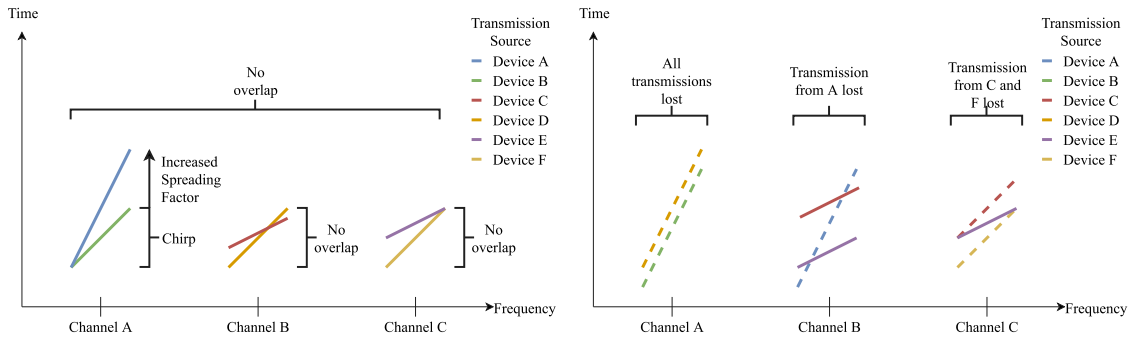
**Fig. 2.** Comparison of active and passive sensing. In active sensing (see Fig. 2(a)), the devices taking measurements are attached to a person and directly record information on human activity. In passive sensing (see Fig. 2(b)), the devices are built into the environment and take measurements of the environment that also allow to infer information regarding human activity.

with only three gateways [32]. The overall system obtained an FER of 2.12%, reduced by a factor of 4 compared to the three gateway setup, which is also far below theoretical boundaries for single-gateway setups. We also show that ESP is a more robust indicator of signal quality and should be used in addition to or instead of Signal to Noise Ratio (SNR) and Received Signal Strength Indicator (RSSI). We report much lower maximal distances and show that previous work on indoor LoRaWAN networks was optimistic, but we confirm that multiple gateways compensate for this shortcoming and enable the deployment of DISN.

## 2. Related work

### 2.1. LoRaWAN

Long Range Wide Area Network (LoRaWAN) [10] is a standard that uses a LoRa Transceiver (LoRa PHY) to transmit messages on the Industrial, Scientific, and Medical Frequency Bands (ISM) based on a simple ALOHA-like protocol [33]



(a) Channel A is used concurrently by device A and device B. Both messages can be received due to the different SF. Similarly, on Channel B and C there is no overlap between transmissions, but the start time and receive times differ. (b) On Channel A, both devices use the same SF. Therefore, both transmissions are lost. Channels B and C demonstrate different kinds of overlaps that cause a transmission to be lost.

**Fig. 3.** Chirps represent a sweep through a frequency band that is used to encode symbols (shown as coloured lines). Chirps can be different lengths determined by the Spreading Factor (SF). Chirps can also be sent on multiple channels (i.e., different frequency bands). Dotted lines represent chirps that cannot be received. A complete transmission consists of a sequence of chirps to encode multiple symbols forming a frame.

and allows for simple, robust, and cheap devices to create the Internet of Things (IoT) and lay the foundation for Digital Twins (DTs). In this section, we will highlight the design features that enable its use in indoor scenarios and provide the necessary background to understand concepts applied in this paper.

2.1.1. LoRa PHY

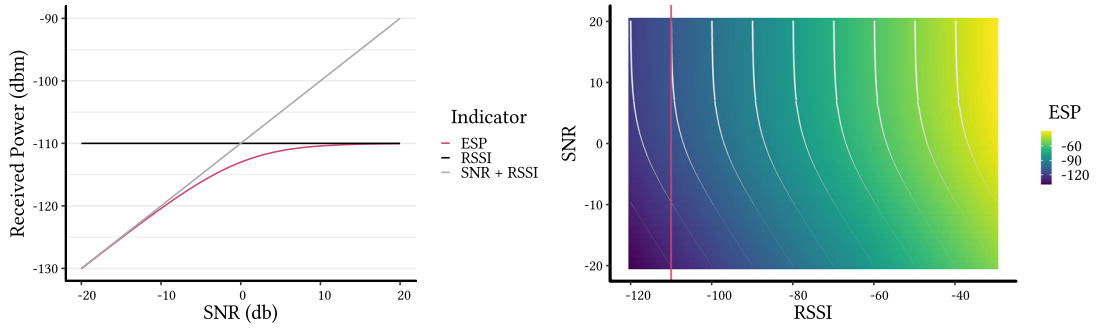
LoRa PHY encodes information with a chirp spread spectrum [34] and an integrated Forward Error Correction (FEC) [35] as specified under IEEE802.15.4a O-QPSK DSSS PHY mode [36]. LoRaWAN can achieve high concurrency because transmissions pseudo-randomly change the CF (often called channel in the context of LoRaWAN) and the SF that determine the chirp shape (see Fig. 3(a)). A chirp traverses a frequency band from low to high or vice versa in a set time window determined by the SF. Therefore, the SF defines the reliability of the transmission where longer transmissions can be received more easily [10]. The SF varies from 7 to 12, with higher values representing a more robust signal, reduced BW [35], a higher chance for overlap in the same CF and SF (see Fig. 3(b)), and thus a higher chance of a concurrent destructive transmission that negates one of the signals via collision. For an in-depth discussion of collisions in LoRaWAN and recoverability, see [37,38].

2.1.2. Spectrum

The unlicensed ISM spectrum is free to use but regulated, and in Europe, the spectrum has a duty cycle limitation of 1% under section 4.3.10.2 of the ETSI EN300.220 V3.2.1 standard [39]. In other regions of the world, different limitations apply, but they all aim to enable everybody to use the ISM in regular intervals. These limitations reduce the number of sensors that can reasonably join a network [40] because the payload size heavily influences transmission time and access. Overall, this kind of setup favours devices that communicate short messages at a low frequency such as environmental sensors. Another limitation is that ultra-reliable services may not be reasonably implemented with LoRaWAN because, in the simplest form, acknowledgements clog the network [40], and the simple MAC protocols of LoRaWAN cannot accommodate the complex channel control [1] that would be required to dynamically allocate resources [41]. While these limitations are inherent to LoRaWAN, their impact can be ameliorated by increasing the number of gateways [42] which may allow researchers to increase the scalability of a LoRaWAN system, especially in the context of DISNs.

2.1.3. LoRaWAN protocol

LoRaWAN is reliable and easy to use because of its simple circuitry, robust transmission signal, encryption, and low costs. A battery life of 10 years is possible with LoRaWAN Class A sensor nodes because of an asymmetrical uplink communication from sensor to server with an ALOHA-like [33] time window for uplink communication followed by a dedicated downlink time window [10]. While communication may be bidirectional, uplink communication from sensor to server is favoured by this specification in order to reduce battery use. LoRaWAN is therefore more practical for sensors than actuators. However, if energy-saving is not the highest priority, then LoRaWAN offers two further device classes (i.e., B and C) with an increasing downlink capacity at the cost of higher energy consumption. Interaction capacity for Class B devices is increased by listening for downlink messages at specific times in addition to the dedicated downlink time window after an uplink communication. Maximal interactivity is achieved with Class C devices that listen at all times except when they are transmitting information [10], but this feature usually requires a permanent energy supply (e.g., a wire). Under the standard, all LoRaWAN devices communicate with servers in a star-of-stars topology according to which



(a) The dynamic range that ESP introduces by combining SNR with RSSI (see Eq. 1). For example, at an RSSI signal strength of -110 dBm, the corresponding ESP can vary between -130 dBm and -110 dBm depending on the SNR signal strength of the transmission. The grey line representing the addition of SNR and RSSI is added as context to explain the ESP curve. (b) The overall range of ESP in relation to the respective measures of SNR and RSSI. The contour lines indicate levels of equal signal strength. The magenta ESP line from Fig. 4a is overlaid. ESP manages to widen the received power range up to -140 dBm by adding the range of the SNR to the RSSI (see Eq. 1). For high signal strengths, ESP defaults to the RSSI.

**Fig. 4.** Visualisations of ESP to explain its relations to SNR and RSSI. Source: These visualisations are inspired by an example from seminal work on ESP [47].

gateways relay messages and translate them into higher order protocols such as IP [10] which can be easily consumed by services on the Internet, see Fig. 1. Recent research also investigates mesh topologies where gateways without Internet access pass messages to other gateways based on LoRA PHY [12,13].

In typical applications, LoRA PHY can send data over distances of 10–40 km in rural areas, 1–5 km in urban areas [43], and 200–800 m in dense urban areas [6,44]. At the best possible DR, only low BW (0.3-50 kb/s) applications [35] are feasible. Theoretically, millions of devices can be supported within a single LoRAWAN network because of the star-of-stars topology [43]. However, there is a trade-off between the number of devices and the number of transmissions which is imposed by the duty cycle limitation and interference from colliding concurrent messages [45].

## 2.2. Signal measurement

Previous research has nearly exclusively focused on measuring signal quality as either the SNR or the RSSI. First, SNR is used to indicate whether a message can be received at all. In other transmission technologies, the noise floor is usually at 0 dB beyond which demodulation cannot occur. However, LoRA PHY allows for a lower noise floor at which messages are lost between -7.5 dB and -20 dB because of the robustness that the duration of chirp signal modulation provides [35] (see Fig. 3). Second, RSSI is used to indicate the level at which the signal weakens, which strongly depends on the transceiver. A general convention is that signals at -60 dBm are considered strong and below -100 dBm are considered weak, and any measurement below the noise floor of -120 dBm disappears [46].

Recent research has proposed a third measure to overcome some limitations of RSSI [47,48]. Effective Signal Power (ESP) is more sensitive at measurements below -120 dBm where RSSI can no longer provide additional information [48]. ESP achieves this by enriching the RSSI measurement with information of the SNR as shown in Eq. (1) [47,48]. This formulation effectively extends the signal floor with the negative range of SNR and allows for more differentiation between different signal strengths (see Fig. 4), especially near the noise floor. There is a need for more empirical studies that compare the different signal quality measures for modelling purposes to clearly understand the benefits of ESP.

$$P_{RX(dBm)} = RSSI_{(dBm)} + SNR_{(dB)} - 10 * \log_{10} (1 + 10^{0.1 * SNR_{(dB)}}) \quad (1)$$

## 2.3. Packet loss in inter-network interference via frame collisions

A LoRAWAN transmission contains a frame that transports a packet containing the payload from a device to a gateway. A transmission can be interrupted in three major ways. First, transmissions can be absorbed by the environment, preventing frames from arriving at the destination. Second, transmissions can be interrupted by another destructive concurrent transmission that is taking place on the same SF, the same BW, and the same CF and that overlap in time. Third, transmissions cannot be received when the gateway is unavailable for technical reasons.

To account for environmental interruptions, previous work has focused on creating models that represent the physical environment to the degree that explains the loss of data. Most of this work has focused on outdoor LoRAWAN collisions and limits [37]. For these outdoor applications, the free space Path Loss Models are the simplest class of models [49,50] and mimic the physical process of transmission in an open space. However, they require strong assumptions regarding

**Table 1**  
LoRAWAN indoor studies. Expanded from [32] with permission.

	[72]	[43]	[73]	[66]	[67]	[74]	[71]	[75]	[69]	[76]	[65]	[45]	[77]	[68]	[78]	[31]	[70]
Region	Europe	Europe	Europe	Europe	Europe	Europe	Europe	Asia	Asia	Asia	Europe	Europe	Europe	Asia	Europe	Europe	North America
#Floors	1	19	4	9	8	3	1	11	1	5	4	6	-	-	1	1	7
Distances	103 m	-	110 m	ca. 20 m	ca. 100 m	45 m	190 m	ca. 30 m	120 m	-	60 m	420 m	580 m	150 m	-	ca. 250 m	-
Data rate (SF) <sup>†</sup>	all	auto	12	all	12	-	7/12	-	-	all	all	-	-	11	9/10/12	all	7-10
Gateways	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2
Devices	4	32	1	1	1	7	2	11	1	5	1	1	1	1	1	331	2
Locations	4	32	70	13	-*	7+ ‡	46	11	89	5	13	23	15+‡	4	4	331	7
Transmissions	50	-	1400	1200	-	72077	50	11000	712	-	80 h	600	1500+‡	60	-	23.2 × 10 <sup>6</sup>	4200

<sup>†</sup> All DR used BW125 or unknown.\* Only a heat map was shown. ‡ Minimal confirmed value.

environmental conditions, including a completely free Fresnel zone [51]. Thus, any meaningful application requires a gateway to be mounted on an antenna tower to keep the Fresnel zone free. In reality, that is often insufficient, and models including a ground plane such as the Irregular Terrain Model [52,53] have been produced. However, if anything enters the Fresnel zone indoors, propagation is dramatically changed [51] by the ceiling as well as the ground and walls. Consequently, path loss modelling indoors had not been investigated until recently [54]. There are several models that tediously find special parameters that are environment-dependent [55]. Because the simulation of signal propagation is too computationally expensive and noisy, strong model simplifications are required. However, these simplifications reduce generalisability. A major difficulty is the requirement of a site-specific model and the calculation of specific attenuation factors [54]. Furthermore, all indoor models assume that signal decay is proportional to log-distances between transmitters and receivers [51,54] and that these models are then modified by some additional terms to represent obstacles.

Another common approach across models in previous work is least squares regression, but this technique may neglect the impact that the transmitter has on the transmission. Researchers often assume that the transmitter is invariant under the studied signal. Since most setups use a single device to measure signal quality, it is also intractable to delineate device-specific effects, so these effects are often ignored. However, with a large number of sensors, the idiosyncratic qualities of sensors may impact overall transmission. In contrast, mixed models [56] allow researchers to model the effects separately for single devices and across all devices [57]. Mixed models offer an opportunity to better understand the error term that is produced across a set of transmissions instead of just accepting them as an intractable byproduct.

Recent research has steered away from deterministic and empirical models in favour of Artificial Neural Networks (ANNs) because shadowing and multi-path fading make classicist models intractable [58]. While ANNs cannot be interpreted as easily as explanatory statistical models based on some form of regression, they have been shown to continuously outperform these models at the cost of interpretability [58–60]. Data regarding the transmission is often correlated, and Principal Component Analysis (PCA) offers the ability to decompose the data into orthogonal components that can be used for models [61,62]. Research on ANNs has often been limited to multilayer perceptrons [58–64], but which ANN is best suited for predictive purposes remains an open question.

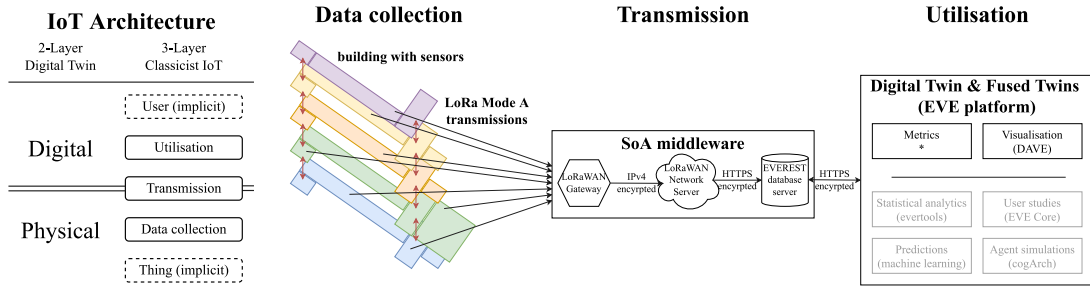
#### 2.4. Indoor Long Range Wide Area Network

Despite the original application of LoRAWAN outdoors, researchers have investigated LoRAWAN indoors (see Table 1) using a minimal viable design (e.g., one gateway, one device, and one network server) and therefore could not measure performance with multiple sensors [45,65–69]. In these studies, the device was moved to multiple locations within the building, mostly on the same floor [45,65,66,69,70]. In addition, many of these models were based on simulations and were never implemented with multiple sensors [71].

More recent work has included multiple devices [31,43,70,74,76] and very rarely even multiple gateways [70,74]. Gateways were usually placed in or on the test building but sometimes further away to test indoor depth penetration of the signal [77] which was lost beyond 300 m. In most cases, the analysis was descriptive rather than predictive, but some model-driven work has been conducted [73]. A move towards testing real-world systems can be observed [31,74], but the literature is primarily reviewing testbeds without real-world use. Despite these limitations, research suggests that higher DRs are useful indoors without sacrificing connectivity coverage [65].

Previous studies have found that a single gateway has reasonable coverage indoors and that the signal can remain stable for buildings between 3 to 7 floors [65,66,70,73,74]. In addition, these studies have not found any discernible effect of placing the sensor nodes on different floors with the exception of the basement where the signal deteriorated quickly in all accounts [45,65–67,70,73–77]. Indeed, in reinforced concrete buildings the signal attenuates quickly covering 8 floors and approximately 100 m [67,69,75,78]. Together, these studies suggest that a higher SF is critical for traversing denser materials and that there are probably limitations how much a single gateway can cover.

In another scenario, a gateway was installed above low-rise buildings without reinforced materials, which allowed for transmissions up to 420 m [45]. With the same setup, a subsequent study found that the seasons have an impact on outdoor and indoor transmission quality [31]. Furthermore, a DISN with a single gateway has been installed on a single floor and was severely influenced by adversarial signals in the environment [31]. In contrast, gateways installed in an open floor could receive transmissions up to 190 m [71]. Only a few studies have investigated distances longer



(a) Abstract system architecture. The digital twin [31] (left) maps to the implementation model [16] (right). Implicit layers (*Things* that sense and *Users* that operate) are also shown in dashed boxes.

(b) Implemented system architecture. LoRaWAN sensors in the building configured in LoRa mode A (left). Gateways forward data to a network server which pushes them to the database EVEREST (center). In the DT, data is loaded from the SoA middleware EVEREST to view metrics and visualisations. Planned features (shown in grey) include statistical analysis, user studies based on EVE [33], predictions, and agent-based simulations [29] (right).

**Fig. 5.** System overview.  
 Source: Image is taken and modified from [32] with permission.

than 100 m [31,45,68,69,71,73,76], and it remains unknown whether interference within buildings reaches a threshold above which an additional gateway must be placed. Also, the studies mostly have not established the long-term quality of the signal due to the low number of transmissions measured but some studies relied on a larger data collection procedure [31,65]. To our knowledge, no study has determined the distance, wall, and floor limits of indoor transmissions that a single gateway can cover, but some investigations have looked into the effect of walls [69].

Overall, previous research indicates that indoor LoRaWAN is feasible for minimal viable setups. Some limitations that have been observed include reinforced material and dense basement walls that reduce signal quality. Only one research project, conducted in parallel to this work, has investigated a DISN, indicating that with a single gateway environmental factors may severely reduce transmissions. The feasibility of LoRaWAN-based DISN remains understudied.

### 3. System design & implementation

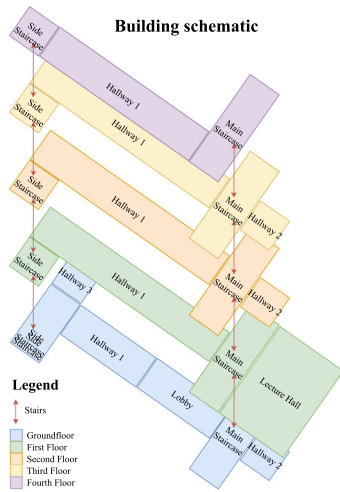
To investigate human presence in the long term based on passive data, it is important to reliably collect environmental data and store them for further analysis. This goal requires reliable transmission, low energy consumption and, low data loss. Our LoRaWAN DISN forms the basis for a DT to represent these physically-based higher-order processes [21]. We map the DT to a classicist three-layer IoT architecture [79] (see Fig. 5(a)), including (1) data collection, (2) transmission, and (3) utilisation (see Fig. 5(b)). To ensure the most effective use of the CF and consequently the most stable data transmission, we use ADR for LoRaWAN. While this choice complicates the analysis by having to take into consideration the impact of ADR, it also ensures that data is appropriately collected because this system is not a testbed but used for actual data collection.

#### 3.1. Data collection

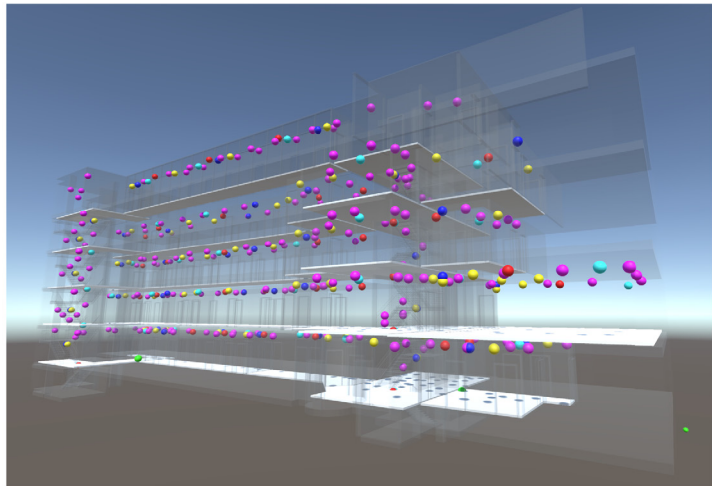
The data collection layer collects environmental characteristics containing passively observable data on human presence in a public building. We tested our LoRaWAN-based DT on an 8 floor office building at ETH Zürich (ETHZ). The public part of the building consists of 5 floors with a small lecture hall on the 1st floor (see Fig. 6(a)). We placed 390 sensors in 19 rooms of varying sizes that may exhibit different levels of human presence (see Fig. 6(b)).

Each of the five types of battery-powered sensor nodes uses LoRa PHY for communication and was selected to cover typical indicators of human presence, including temperature, CO<sub>2</sub>-levels [80], noise [81], humidity, and motion [82] (see Table 2). We focused on sensors that covered CO<sub>2</sub> (40 nodes) and Volatile Organic Compounds (VOC; 40 nodes) as lingering effects of human presence. We also employed noise sensors (40 nodes) and motion sensors (250 nodes) that immediately indicate presence but are susceptible to false positives. In addition, we collected environmental factors (e.g., temperature, humidity, brightness) that can be driven both by human presence and other factors such as weather, heating, time of day, and whether outside windows and doors are open.

The placement of sensors needs to conform to several requirements and was stored for analysis. First, sensors can only collect data in public spaces including staircases, hallways, and auditoriums. The privacy of building users was maintained as sensors were not placed in offices, bathrooms, or other private areas. Second, sensors should be well-distributed to support data fusion. Sensors of the same type were densely and uniformly placed throughout each room. Third, sensors








(a) A schematic layout of the public part of the building with stairs as red arrows. The 19 public rooms link to offices on the left and right along Hallway 1 and around Hallway 2.



(b) The 390 sensors shown in DAVE with white rectangles for the involved rooms. The gateways are green (one is in the non-public area of the building, 3 floors down in the bottom right corner). Different types of sensors are represented as spheres of different colours (noise: blue; CO<sub>2</sub>: red; VOC: cyan; PIR: yellow and magenta). The model is available on zenodo [35].

**Fig. 6.** System implementation. Source: Images is taken from [32] with permission.

**Table 2**  
Selected sensors for the prototype at the university.  
Source: Taken from [32] with permission.

	ERS CO2	ERS	Orbiwise Sampol	Browan Tabs IAQ	Browan Tabs
					
CO <sub>2</sub>	✓			✓	
VOC				✓	
PIR Motion	✓	✓			✓
Brightness	✓	✓		✓	✓
Noise			✓		
Humidity	✓	✓	✓	✓	✓
Temperature	✓	✓	✓	✓	✓
Focus	Motion	CO <sub>2</sub>	Sound	VOC	Motion
Sample rate	10 min	10 min	5 min	5 min	Event
BW	125 kHz	125 kHz	125 kHz	125 kHz	125 kHz
Count	60	40	40	40	210

must be placed within the limitations of the existing building. The sensors are mounted on the ceiling beyond the reach and attention of building users while accounting for the other hardware, including cable trays and WiFi-routers.

### 3.2. Transmission

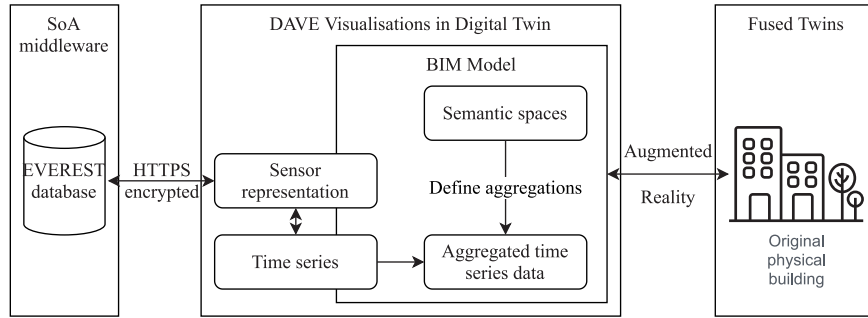
The sensor data is made available with a SoA design by which each processing step is performed independently to increase interoperability. The data from the sensors are sent to the network server, stored persistently on a database server, and made accessible for further analysis in the DT (see Fig. 5(b)). All sensors are configured as Class A LoRaWAN devices [10] that collect data every 5 to 10 min or in an event-based manner (see Table 2). The sensors transmit their payload to the gateways on a CF with a BW of 125 kHz secured with AES-128 [10], see Fig. 1. Re-transmission was disabled to prolong the battery life at the cost of missing some transmissions due to collisions.

The original LoRaWAN network we used had three gateways [32], and a fourth gateway was installed in November 2020 to improve signal quality on the network level. The first two gateways were placed on the ground floor of the public

**Table 3**  
The SNR thresholds for the ADR algorithm.  
Source: Taken from [32] with permission.

SF	Decrease Threshold	Increase Threshold
SF7	*	-7.5
SF8	-5	-10
SF9	-7.5	-12.5
SF10	-10	-15
SF11	-12.5	-17.5
SF12	-15	*

Cannot increase or decrease further. Margins are not included.



**Fig. 7.** The processing pipeline of sensor data in our DT implemented in the EVE platform. Sensor data is delivered by the SoA middleware EVEREST to DAVE. The Fused Twins paradigm [27] is applied to display the data in its spatial context in the original building [26].  
Source: Image is modified from [32] with permission.

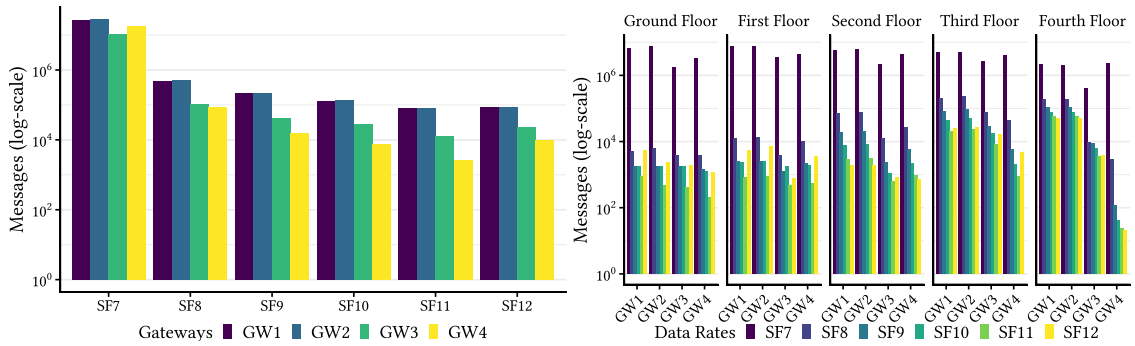
part of the building. The third gateway was placed several floors below on the ground floor of the adjacent non-public part of the building (shown in green in Fig. 6(b)). The newly added fourth gateway was located on the top floor to improve signal reception further away from the ground floor. We used OrbiLINK v4 indoor LoRAWAN nano-gateways (max. 27 dBm; class A & C devices; 8 LoRa CF) [83]. The gateway converts the IEEE 64-bit Extended Unique Identifier (EUI) [10] to a IPv4 address and forwards it to a proprietary LoRAWAN network server.

The network server sets the ADR, enforces the Link Adaptation Policy (LAP), forwards the payload, and synchronises the gateway clocks (delay <20 ms). Our network server performs a two-step optimisation based on ADR and LAP to optimise the available BW, comply with duty cycles, save battery life, and reduce Time-on-Air (ToA). For the ADR step, the SF is adjusted whenever the average of the last 15 transmissions crosses the SF threshold within a 12 dB margin, see Table 3. Switching the SF resets transmission power to the maximal output to ensure that a link can be established. For the LAP step, the transmission power is reduced such that transfers occur barely above the noise level. In practice, LAP is only applied on SF7 because there is usually no power budget available for higher SF. The payload is decrypted with an application key [10], combined with meta data from each gateway (e.g., ESP, SNR, and RSSI), and converted into JavaScript Object Notation (JSON). The network server provides a webhook (HTTP push API) secured with basic access authentication (HTTP Auth) and immediately pushes the JSONs to our persistent storage facility.

The persistent storage facility *Experiments in Virtual Environments Representational State Transfer* (EVEREST) extends the larger open-source platform EVE [84] and forms a core feature of the DT. EVEREST is a flask-based python server that implements Representational State Transfer Hypertext Application Language (REST-HAL) [85]. The webhook consumes all messages from the network server and inserts the data into a local PostgreSQL database extended with TimescaleDB for time-series and Postgis for spatial data. Finally, the EVEREST server is hosted on a virtual machine with 4 virtual processors and 32 GB RAM at ETHZ.

### 3.3. Utilisation

The utilisation layer consists of a prototype of the visual DT and FT which are also integrated into the EVE platform [84]. The software Data Analytics in Virtual Environments (DAVE; see Fig. 6(b)) can be used to explore either the DT on a screen or using an AR HoloLens to access the FT representation of the data *in situ* [26,27]. Because readings from any single sensor would be difficult to interpret, a Building Information Model (BIM) in DAVE provides a spatial context for the densely placed sensors. A stand-alone version is available on zenodo [86]. Our DT also includes features for visualisation and analysis (see Fig. 7). The data are stored on EVEREST in near-real time and immediately displayed in the DT and the FT. The difference to real time display is due to a tradeoff between power usage and transmission delay in the transmission settings of LoRAWAN because of the limited battery life of the sensors.



(a) The SF of all received transmissions at each gateway (GW) on a log-scale. Note that GW4 was installed in November 2020 and Higher SF are more often used on higher floors. On some floors, therefore has 8 months less of received transmissions. In the most SF12 is used more often than lower SFs, indicating that ADR common transmission DR, GW4 still overtakes GW3, indicating increased the SF to improve signal quality. GW4 is located on a much better reception.  
 (b) Received transmissions per GW, floor, and SF on a log-scale. Higher SF are more often used on higher floors. On some floors, therefore has 8 months less of received transmissions. In the most SF12 is used more often than lower SFs, indicating that ADR common transmission DR, GW4 still overtakes GW3, indicating increased the SF to improve signal quality. GW4 is located on the fourth floor and thus receives less high SF transmissions.

Fig. 8. Transmission overview. Each transmission can arrive at all GWs in principle.

## 4. Results

We present three sets of results. First, we show descriptive statistics to provide insight into the data. Second, we produce explanatory models based on different methodologies to understand factors contributing to data loss such as distance and walls and how to best model them indoors. Third, we compare predictive models based on different ANN approaches to understand which are best suited to predict new sensor data and gateway locations.

### 4.1. Descriptive statistics

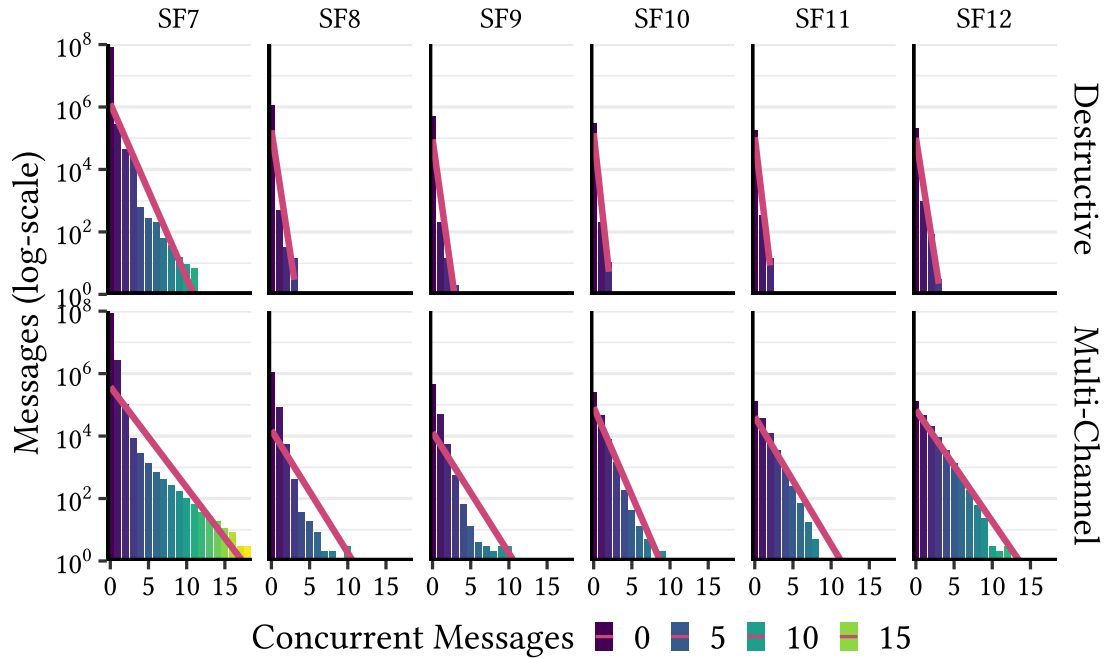
#### 4.1.1. LoRaWAN metrics

Between March 2020 and February 2022, our setup collected 86 million transmissions over a range of up to 64 m and 8 floors. The raw data are available on zenodo [87]. Adaptive Data Rate (ADR) allowed us to analyse the preferred transmission mode in our physical setup. The majority of messages were sent on SF7 (96.3%; see Fig. 8(a)). Given the short maximal distance, we expected all gateways (GWs) to receive a similar amount of transmissions except GW4 due to the later installation. The two GWs located in the public part of the building received a similar amount of messages (see Fig. 8(b)). The GW located in the non-public part of the building received up to two magnitudes fewer transmissions, depending on the originating floor. In addition, as the number of floors increased, ADR increased the SF to compensate for the loss of messages. No device reported a duty cycle usage higher than 0.004%. Regularly occurring colliding concurrent transmissions involved up to 14 transmission on the same SF and CF. We also recorded rare singular events in which 20, 38, and 39 transmissions were transmitted at the same time, on different CF. However, we only recorded singular events of at most 14 transmissions that were concurrent and destructive on the same CF. We still have more than 99.58% of messages reporting no collision, although 3.6% of transmissions were sent concurrently without collision (see Fig. 9).

We calculated the FER from the sequence number of the received JSON such that any jump in the sequence indicates lost frames. We split the FER analysis into two parts (see Table 4). In a first phase, we analysed the data from the first 8 months with three GWs, and in the second phase, we analysed the data from the next 8 months with four GWs. We used a balanced amount of months and not the full data set. In the first phase, the first two ground floor gateways reported a FER of 25.51% and 17.46%, respectively, while the third remote GW reported an FER of 80.55% (see Fig. 10(a) and Table 4). At the same time, the average FER reported at the network server across all sensor nodes was 8.06%. This average FER is considerably lower than any individual GW's FER because some frames only arrived at one GW and appeared lost for the other gateways but could be restored from the network perspective (see Fig. 10 for details). In the second phase, the first two ground floor GWs reported a higher FER of 28.6% and 22.3%, respectively. The third GW slightly improved its FER to 73.6%. The newly added fourth GW on the top floor reports a similar FER to the ground floor GWs at 25.6%. In contrast to the first phase, the average FER reported at the network server across all sensor nodes was four times smaller at 2.12%. Note that our FER included frames lost due to GWs being disconnected from the network and that we observed a downtime of around 15 min per day or 1% of possible transmission time.

#### 4.1.2. Signal strength

We analysed the signal quality of LoRaWAN transmissions in terms of ESP, SNR, and RSSI (for. First, we observed the ESP, SNR, and RSSI grouped by DR and the number of floors between the sensor nodes and the gateway. Next, we plotted these grouped measures as functions of the distance between sensor node and gateway (see Fig. 11). We observed a clear



**Fig. 9.** Summary of concurrent transmissions being received within the same time window. Destructive concurrent transmissions are received on the same SF and CF at different GWs (top) whereas multi-channel transmissions are received on all SFs and CFs at all GWs (bottom). Trendlines (magenta) have been included for all facets. All trends are linear in the log-scale and thus show an exponential reduction in higher order concurrent events. Comparing the slope of the trend line between multi-channel transmissions and destructive concurrent transmissions indicates that destructive concurrent transmissions were still very unlikely and accounted for a minor part of the lost transmissions.

**Table 4**  
Summary statistics on FER across all gateways.

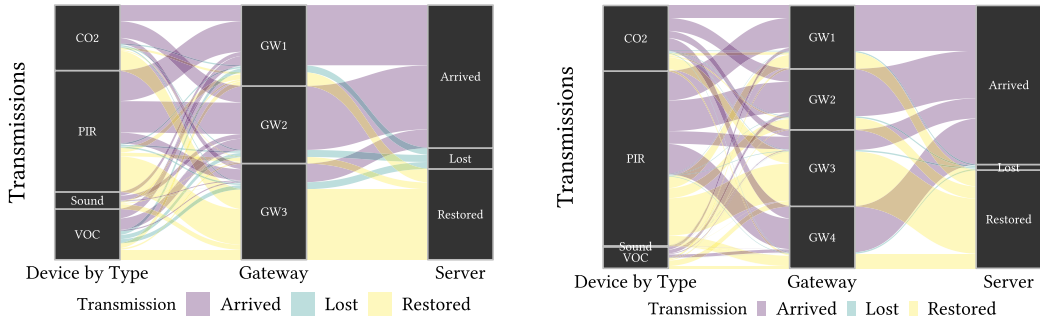
	SNR				
Gateway	GW1	GW2	GW3	GW4	Network
Location	Ground floor	Ground floor	Other building	Top floor	-
Phase 1 FER	25.5%	17.5%	80.6%	-	8.06%
Phase 2 FER	28.6%	22.3%	73.6%	25.6%	2.12%

Note: \* GW was installed in Phase 2.

decay of the signal with increasing distance and number of floors across all DR, indicating attenuation that occurs beyond physical path loss models for open space [88]. However, the effect of floors on signal quality decreases with distance. Because most data were sent on SF7, SF7 produced the most stable curve. Higher SF seemed to offer some signal stability for very short distances but quickly deteriorated and approached the noise floor of LoRa at 30 m. Note that higher SF started with a higher ESP, SNR, and RSSI for short distances because ADR resets the power output. While higher SFs increase Time-on-Air (ToA) and allow the signal to be integrated, leading to higher ESP, SNR, and RSSI [72], they are also more prone to signal shadowing and multi-path fading indoors [72]. This may explain why the average signal strength in our study quickly reached the noise floor as the transmission distance increases.

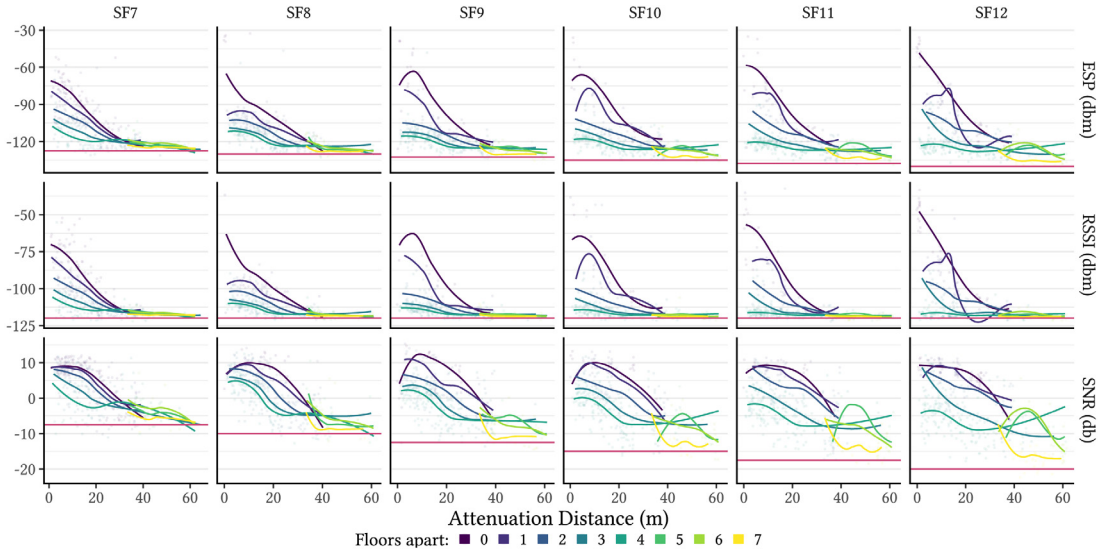
#### 4.2. Explanatory models

Explanatory models enable “introspection” [89] by design and to identify variables that have a meaningful and statistically significant relationship with an outcome. These kind of models for signal transmission allow us to understand the environment’s impact on signal quality. Based on explanatory models, we can build heuristics that help us to quickly assess issues when deploying IoT infrastructure in buildings. Typical explanatory models in signal quality assessment rely on linear regression. But with signal shadowing and multi-path fading in indoor environments, linear regression may not sufficiently explain signal quality. In this paper, we use dimension reduction approaches to remove noise and mixed models to structure noise to gain a better understanding of signal propagation indoors. We compare our new models with a classicist linear model in the model comparison section in Table 7.



(a) Transmissions received between March 2020 and November 2020. On average, 8.06% of transmissions were lost. The remote gateway (GW3) improves reception of transmissions by 1.31%. Taken with permission from [102].  
 (b) Transmissions received between November 2020 and July 2021. On average, 2.12% of transmissions were lost. The newly added gateway (GW4) produces strong reduction in lost messages of 280% compared to the previous phase despite a slightly worse reception for each individual gateway (see Tab. 4).

**Fig. 10.** FER by gateways and devices. Transmissions are aggregated by device type (left), gateways (middle) and the network server (right). Transmissions are represented by different colours (purple: arrived; green: lost; yellow: restored). Restored messages arrived at least at one gateway and could be restored for other gateways, substantially reducing lost transmissions. Each transmission is counted three/four times to account for it passing through each gateway. Within each image, the overall proportion of arrived, lost, and restored messages remains unchanged. Source: Taken with permission from [32].



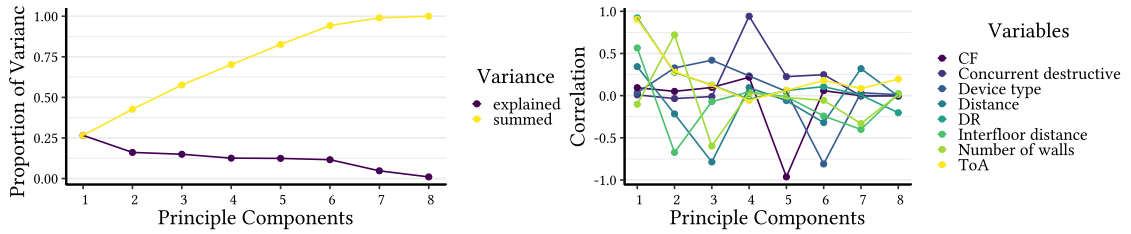
**Fig. 11.** The ESP, SNR, and RSSI as a function of distance given the DR. The red horizontal line indicates the noise floor threshold of LoRAWAN. For SNR, the threshold depends on the DR [35], and for RSSI, the  $-121$  dBm threshold depends on the gateway. For ESP, the threshold is the combined threshold of both SNR and RSSI for each DR. The curves are aggregated ESP, SNR, and RSSI measures based on locally estimated scatterplot smoothing (LOESS) regression. The y-axis has different scale per measurement types. The points represent a random sample of real measurements stratified for DR to display the underlying distribution. The noise floor threshold is reached at about 30 m for RSSI across all DR but is less pronounced for measures of SNR where the noise floor is only partially approached. For the ESP measurements, the signal strength is more differentiated beyond 30 m, which may explain the continued reception of the signal and the gradual approach to the noise floor.

#### 4.2.1. Principal component regression

If PCA indicates that a low-level representation of the data is possible [90], then Principal Component Regression (PCR) is a robust approach for regression [91] because it handles noisy, missing, and mixed (discrete and continuous) valued variables [92]. PCR is a two-step approach to solve a linear regression (see Eq. (2)). First, PCA is performed (see Eq. (3)). Second, a linear least square regression is performed on a subset of the Principal Components (PCs) producing a new data matrix  $\mathbf{Z}$  (see Eq. (4)). The coefficients in the original variables can be reconstructed (see Eq. (5)).

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E} \tag{2}$$

$$\mathbf{X}^T \mathbf{X} = \mathbf{P}\mathbf{D}\mathbf{P}^T = \mathbf{Z}^T \mathbf{Z} \tag{3}$$



(a) The explained variance plot shows a slow decline in the proportion of the variance explained indicating a low potential for relations with the original variables. While no variable uniquely loads on any principal component, each PC is dominated by another variable.

Fig. 12. PCA.

$$\hat{\mathbf{A}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y} = \mathbf{D}^{-1} \mathbf{Z}^T \mathbf{Y} \tag{4}$$

$$\mathbf{B} = \mathbf{P} \mathbf{A} \tag{5}$$

If all components are used, PCR equal linear least squares regression. However, some research has indicated more robust and less biased results [91,93] when dropping some principal components, especially under noise. While common approaches include focusing only on high-variance principal components [90], alternative selection criteria like best subset can result in better models [94].

We conducted a PCA on the variables available from the network meta data (i.e., device type, distance, CF, SF, ToA, interfloor distance, number of walls, and number of concurrent destructive transmissions) to explore how the data varies. Specifically, we hope to understand the relations among the input parameters for our models. We found that the first 6 components accounted for 94.2% of the variation, and we discarded the last 2 components as noise (see Fig. 12(a) and the appendix for bidimensional plots Figs. 21–23). Several components loaded strongly on original variables. The information on transmission time based on SF and ToA loaded on PC1, suggesting a high collinearity between the variables with near equal shape (see Fig. 12(b)). Concurrent destructive transmissions almost exclusively defined PC4, whereas CF almost exclusively defined PC5. Similarly, device type almost exclusively defined PC6. Spatial variables for the number of floors, the number of walls (defining PC2), and distance (defining PC3) have weaker loadings (see also biplots in Appendix B). We can also explain the noise in the dropped components due to loading on the main components. The ToA did not seem to be independent of the SF and indeed was constructed by the transmission time required by the SF and the size of the payload, which was constant over time. The spatial variables were defined by three interrelated concepts, and two components appear sufficient to represent all the information. This pattern can probably be explained via the arbitrary differentiation between walls and floors we made. In practice, we expected these two components to be different from each other because walls are generally thinner and more easily allow transmissions to pass compared to the floor. The PCA indicated that some components may be carriers of noise, so we applied PCR as an alternative model to understand the data.

#### 4.2.2. Partial least squares regression

Partial Least Squares Regression (PLSR) [95] bears some resemblance to PCR in that it can be used to reduce dimensionality but works by projecting both the dependent and independent variables into a new space [96,97]. More accurately, PLSR produces a projection to latent structures [98] in the data (see Eqs. (6) and (7)).

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T + \mathbf{E} \tag{6}$$

$$\mathbf{Y} = \mathbf{U} \mathbf{Q}^T + \mathbf{F} \tag{7}$$

If the covariance between  $\mathbf{T}$  and  $\mathbf{U}$  is maximised, then both  $\mathbf{X}$  and  $\mathbf{Y}$  are projected into the same space [98]. In practice,  $\mathbf{U}$  is substituted for  $\mathbf{T}$  in algorithms such as PLS1 [98]. The iterative construction of  $\mathbf{T}$  allows us to remove dimensions similar to PCR.

To compare PLSR and PCR, we reduce both to the same dimensionality and take the PCA variance explained as a guide. We look at the result from PLSR through the Loading Factors (LFs) rather than the weights or scores [98]. The signal quality indicators load on LF1 negatively and LF2 positively. Their strong inter-relatedness can be identified from the near identical loading on all other LFs. CF loads strongly on LF3 and LF4. Concurrent destructive transmission loads mostly on LF5 and LF6. Device type loads equally on LF4, LF5, and LF6 and in opposite direction on LF3. The ToA loads nearly exclusively on LF6. Distance barely loads on LF6. DR loads strongly on LF1 and LF2, indicating a strong relation with the outcome variables ( $\mathbf{Y}$ ), but DR also loads on LF3 and somewhat on LF5. The interfloor distance and number of walls load nearly equally across all LF except for LF6. They also load strongest on LF1 and LF2, again indicating a strong relation with the outcome variables ( $\mathbf{Y}$ ).

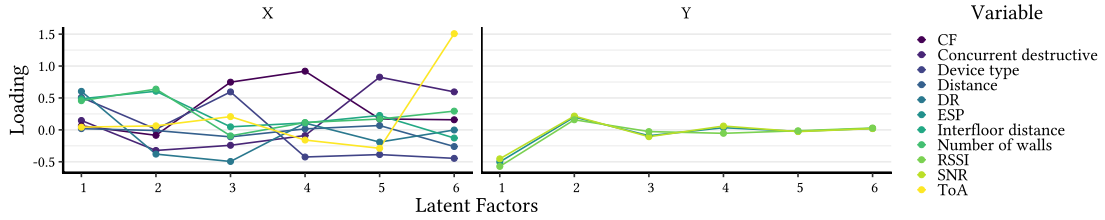


Fig. 13. PLSR as a latent structure. The signal quality indicators (Y) are mapped very closely to each other in the latent space, whereas the variables (X) are widely dispersed around the latent space.

#### 4.2.3. Structured error terms with mixed models

We modelled attenuation as linear signal decay to understand the relations among distance, building structure, indoor obstacles, and network. Because each of the 390 devices regularly sent transmissions, signal strength was not independent. Furthermore, we have five types of devices which each bring a unique configuration per device type and possibly have an impact on the transmission. We used a mixed model [56] to quantify the impact of the structural setup on ESP, SNR, and RSSI. Suppose that a dependent level variable  $y$  is measured at a unit level (level 1) that is nested within units of type A (level 2), and that units of type A are nested within levels of type B (level 3). We attempt to model the dependence between both multiple transmissions (level 1 in the hierarchy) from each device (level 2 in the hierarchy) and multiple devices from each of the device types (level 3 in the hierarchy). In the mixed model, we nested transmissions within devices and devices within device types. To better understand the impact of DRs on ESP, SNR, and RSSI, we created independent models for each SF because a joint model would require higher order interaction terms to tease apart the differences between DRs.

Our models have fixed effects for environmental criteria (e.g., distance and number of floors) and network criteria (e.g., colliding concurrent messages) and random intercepts for each sensor (see Eq. (8)). We computed ESP, SNR, and RSSI as the output  $y$  for a regression on fixed effects at every transmission  $j$ , device  $i$ , and device type  $k$ . The coefficients  $b$  grouped the transmissions  $q$  that each device  $i$  sent and represented the nested random intercept. The coefficients  $c$  grouped the devices  $p$  that belonged to each device type  $k$  and represented the random intercept. For the destructive\_concurrent message variable, we counted the number of other transmissions that were “on air”. We subtracted the ToA from the gateway timestamp to define the interval in which arriving messages overlapped at different gateways on the same SF and CF. In addition, this interval was expanded by 25% on each side to produce a time window that accounts for the gateway synchronisation delay.

Because complex physical simulations would be necessary to account for signal shadowing and multi-path fading, the distance variable (meters from gateway to sensor nodes) accounts for physical “path loss” as a black-box linear factor that covers different covarying environmental aspects. This approach substantially differs from classical LoRAWAN models based only on free space path loss [88], which usually applies the log-distance. We demonstrate that, for modelling signal quality in a DISN, the log-distance factor is not useful (see Appendix C) because it predicts very similar on test data to the linear distances but with a slightly larger Root Mean Square Error (RMSE). Therefore, we replaced the log-distance factor with a linear distance factor. This simplifies the interpretability of the model because we can read the possible signal decay per meter rather than having to take the logarithm.

$$\begin{aligned}
 y_{ij} = & \beta_0 + \beta_1 \text{distance}_{ij} + \beta_2 \text{floor}_{ij} + \beta_3 \text{wall\_count}_{ij} + \beta_3 \text{carrier\_frequency}_{ij} + \\
 & \beta_3 \text{time\_on\_air}_{ij} + \beta_4 \text{destructive\_concurrent}_{ij} + \\
 & \beta_5 \text{distance}_{ij} * \text{floor}_{ij} + \beta_6 \text{distance}_{ij} * \text{wall\_count}_{ij} + \beta_7 \text{floor}_{ij} * \text{wall\_count}_{ij} + \\
 & \beta_8 \text{distance}_{ij} * \text{floor}_{ij} * \text{wall\_count}_{ij} + \\
 & b_{i1} z_{1ij} + \dots + b_{iq} z_{qij} + \epsilon_{ij} \\
 & c_{ik1} w_{1ijk} + \dots + c_{ip} w_{pijk} + \epsilon_{ijk}
 \end{aligned} \tag{8}$$

#### 4.3. Predictive models

In contrast to explanatory models, predictive models are not intended to delineate the factors that impact the result. Their goal is to associate predictor variables with the outcome to generate good predictions such as for the deployment of additional devices and gateways. Variables that are used in a predictive model are based on association, not statistical significance or scientific meaning. Predictive models often rely on non linear transformations and are solely optimised to reduce the error of a prediction at all costs, especially comprehension. ANN are particularly suited for the task because they work when an exact analytical formula is unavailable [99]. Indoor signal transmissions are a natural fit for ANN because the complexity of signal shadowing and multi-path fading cannot be captured easily in simple models. A predictive model may therefore allow us to assess the placement of sensors and gateways in a building more accurately.

Because we have no theory that allows us to describe the impact of signal shadowing and multi-path fading in a model, we require an ANN to investigate the impact of the wall features on signal quality and to account for non-linearities. We also require an ANN model that can take into account imbalanced feature input such as a variable number of walls to be traversed. Varying features can be represented either recurrently in a Gated Recurrent Unit (GRU) or a Long Short-Term Memory (LSTM), or as a full feature in a MLP or Transformer ANN. The latter two models require us to know the total number of features limiting the transferability of the model to setups with an equal or lower number of walls. We compare these machine learning models to test the capability of forecasting sensor signal quality from existing transmission data while taking into consideration the spatial configuration of the building. The main idea behind these models is to use a simplified problem in which we consider only the direct propagation of the data through a straight line between the sensor and the gateway and exploit the sequential nature of this simplified signal propagation and obstacle clearing process. In other words, we limit the model to consider the data on the straight line between each sensor and gateway.

The features used to train the ANNs are the 8 device parameters (i.e., number of walls, floor difference, device type, distance, CF, DR, ToA, and concurrent destructive transmission count) and the 3 characteristics of a maximum of 16 obstacles that the straight line component of the signal would have to cross in its path to the gateway (i.e., angle of incidence, thickness, distance). We use the BIM in the DT to determine these characteristics along the direct line between the sensor and gateway. We also determine the number of walls traversed, the angle at which the line hits the walls, and the thickness of the traversed walls (see Algo. Section 4.3).

---

**Algorithm 1:** Wall properties calculation
 

---

**Input** : The position of a gateway and a device. A building information model.

**Output** : A list of tuples containing the distance, thickness, and angle of incidence.

**Ray** : Returns a directed line segment between a start point and end point.

**Raycast**: Returns the intersections of a line segment with geometry in order from start to end.

```

ray ← Ray(gateway.position,device.position);
intersections ← Raycast(ray,bim);
free_space ← True;
result ← list();
properties ← tuple();
for i ← 0 to length(intersections) do
  if free_space then
    properties.angle ← intersections[i].angle;
    properties.distance ← intersections[i].distance;
    intersection.position ← intersections[i].position ;
    free_space ← False;
  else
    thickness ← distance(intersections[i].distance,intersection.position);
    free_space ← True;
    result.append(properties);
    properties ← tuple();
  end
end

```

---

#### 4.3.1. Artificial neural networks

Multi-Layer Perceptrons (MLPs) have been used for signal quality prediction since the very early days of ANN application in the 2000s [64]. A main advantage is that MLP can distinguish data that are not linearly separable because it is a universal function approximator [100]. Their classical design is limited to a single hidden, densely connected layer, but modern interpretations also allow for multiple dense layers. The input parameters must be fixed, which makes feeding environmental information about the signal transmission difficult if the environment is unknown.

The limitation of a fixed input can be overcome by a Long Short-Term Memory (LSTM) [101]. As a form of a recurrent ANN [102], it differs from MLPs by having feedback connections that allow it to process sequences of data and remember states of the sequence. Based on our wall data format, we can feed in an arbitrary sequence of walls without the limitations on the number of walls imposed by the MLP. A possible alternative candidate is the Gated Recurrent Unit (GRU) [103], which has a simpler design than the LSTM but also tends to forget quicker. However, for short sequences, GRUs have been known to outperform LSTMs [104,105].

Lastly, the deep learning revolution [106] produced Transformer architectures [107] that differ from recurrent ANN in that they consume a sequence as a whole. In contrast to MLPs, they use self-attention to determine points of interest in the same sequence but not necessarily in the same sequence order. Attention can be understood as a soft weight that is changed dynamically at runtime based on the input. Differing input lengths are managed by padded input.

**Table 5**  
Results of various models.

Model	Explanatory																		Predictive					
	Perfect estimator		Mixed model*				Linear model*		PLSR		PCR		LSTM		GRU		Transformer		MLP					
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE						
ESP	-107.43	6.30	-107.87	<b>9.06</b>	-110.26	10.47	-108.80	11.46	<b>-108.77</b>	11.47	<b>-107.08</b>	<b>8.09</b>	-106.19	8.46	-108.32	8.98	-105.88	8.42						
SNR	2.58	2.82	<b>2.54</b>	<b>3.84</b>	1.64	4.26	2.20	4.24	2.22	4.23	<b>2.99</b>	<b>3.63</b>	3.04	3.59	3.23	3.87	3.33	3.72						
RSSI	-104.44	5.59	<b>-104.92</b>	<b>8.34</b>	-106.92	9.47	-105.69	10.57	-105.75	10.59	-106.01	<b>14.37</b>	<b>-104.44</b>	7.48	-105.69	8.09	-103.48	7.45						

Note: \* For model specification see [Appendix C](#).

#### 4.4. Model comparison

Throughout the previous sections, eight models, both explanatory and predictive, have been introduced. We compare the models on a balanced subsample of transmissions and compare them to a perfect estimator as a baseline for the best possible performance (see appendix for [Table 17](#)). We subsampled the dataset to obtain approximately 1 million transmissions. When possible, we balanced the number of samples for each of the 1560 sensor-gateway pairs, resulting in 909,414 samples. From the data, we selected 70% of the sensor-gateway pairs for training, 15% for validation, and 15% for testing. We evaluate the models' performance on the test set using the RMSE for ESP, RSSI, and SNR (see [Table 5](#)) and by inspecting the mean. As a benchmark, we use a "perfect estimator" that is created from the ground truth of the signal quality of the test set and provides the expected value (mean) of each individual sensor-gateway pair distribution (see [Table 5](#)). This perfect estimator provides a lower bound on the achievable quality of forecasting given the training features as a RMSE describing the difference between the test data distribution and the training data distribution. Interestingly, the best explanatory model (mixed model) and the best predictive model (LSTM) perform on a similar level compared to the perfect estimator. The mean is well estimated by all models except the linear model which overestimate and the MLP which underestimates. However, generally the predictive models achieve a better RMSE in line with their design to optimise exactly for that purpose.

Unsurprisingly, the mixed model performs best in the group of explanatory models because it accounts for the repeated transmissions and device types. Both PLSR and PCR fall behind in terms of RMSE because they approximate least squares linear regression with projected variables and a reduced set of principal components, respectively. While PCR produces the closest mean to the perfect estimator, the difference in means between all explanatory models is too small for us to prefer either outcome. However, both PLSR and PCR estimate the mean more accurately than a linear model. This is consistent with previous work [93] that found that these estimates are less biased than a linear model.

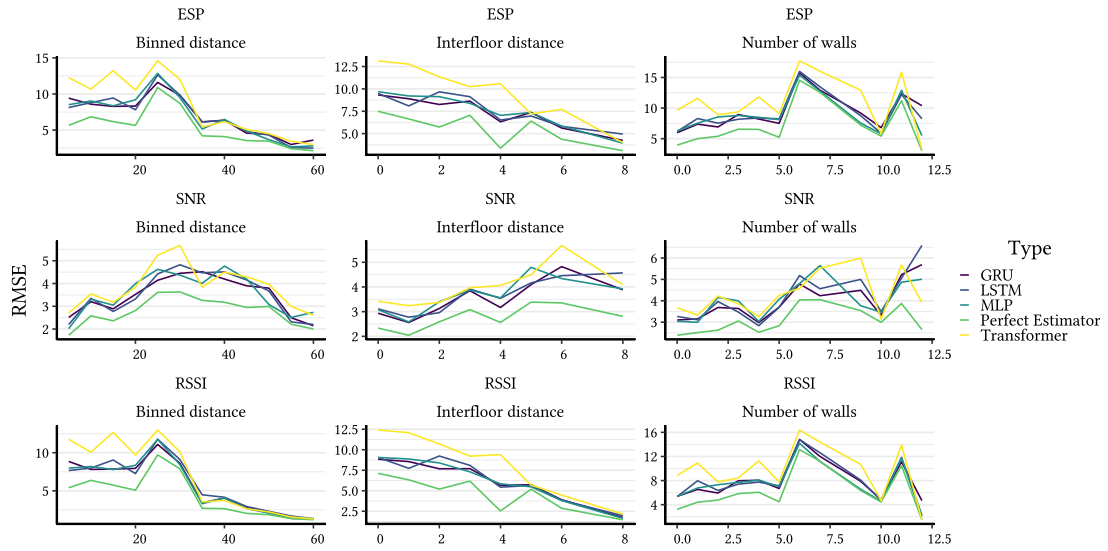
Among the predictive models, the LSTM model obtains a mean prediction closest to that of the perfect estimator on the test set. Furthermore, when comparing the non-linear LSTM model to the linear mixed model (see [Table 5](#) and [Appendix C](#) for model specification), we can see an improved in prediction of 10.7% (ESP), 5.5% (SNR), and 13.7% (RSSI). Nonetheless, the means for both the mixed model and the LSTM model are very close to the perfect predictor.

Interestingly, the GRU and Transformer are nearly on par with the LSTM, and hyper-parameter tuning may favour one or the other architecture. However, the MLP performs worst in this comparison with a mean that is far off from the perfect estimator with a relatively high RMSE.

However, the RMSE among all ANN is still larger than the RMSE of the perfect estimator. Therefore, we explore how the forecasting quality distributes on the different features of the data to understand where our models deviate from the perfect estimator (see [Fig. 14](#)). We find a close match in terms of the RMSE among most curves, and the trend for RMSE in the data is clearly mirrored by the predictive models. This match is strongest for RSSI and less pronounced for SNR. For ESP, this match for RMSE is slightly worse compared to RSSI given the influence of SNR (see Eq. (1)). The LSTM, GRU and MLP perform very similarly and usually close to the "perfect estimator". An exception is formed by for the inter-floor distance under SNR suggesting that the simplified direct line model is a better approximation for a sensor further away from the gateway with more walls, distance, and floors in between. The Transformer underperforms on nearly every category but has the best performance for single values and mostly on RSSI. Therefore, despite the worse performance it has a lower RMSE than the other ANN.

##### 4.4.1. Final explanatory model

Given the models' comparative performance, the Mixed Model is the best choice for analysing the contributing factors for signal decay. Our mixed models for the different signal quality metrics (see [Appendix D](#) for ESP [Table 8](#), SNR [Table 9](#), and RSSI [Table 10](#)) confirmed the expectations derived from the descriptive data. The models' constants indicated average signal strengths at the sensor node of around  $-79 \pm 19$  dBm for ESP,  $9 \pm 5$  dB for SNR, and  $-79 \pm 19$  dBm for RSSI. Here, we discuss SF7 in detail because it represented most messages. Due to attenuation, our models across DR predicted a signal loss per meter of  $-1.72$  dBm for ESP,  $-0.26$  dB for SNR, and  $-1.67$  dBm for RSSI. Given that the height of each floor in the building was 3.36 m, we estimated that each floor affected the signal strength by  $-14.26$  dBm (ESP),  $-1.59$  dB (SNR), and  $-15.43$  dBm (RSSI). Similarly, walls were on average 0.1 m thick and therefore affected the signal strength by  $-4.93$  dBm (ESP),  $-0.27$  dB (SNR), and  $-4.54$  dBm (RSSI). The interactions between attenuation distance, the number of floors, and the number of walls were all significant across ESP, SNR, and RSSI, except for the SNR wall-attenuation interaction.



**Fig. 14.** RMSE of the predictive models LSTM, GRU, Transformer, and MLP and the perfect estimator on data subgroups for ESP, RSSI, and SNR. The RMSE impact is shown for the variables for binned distance (in 5 m steps), interfloor distance (counted in floors), and number of walls.

**Table 6**  
ESP limits for transmission across floors in our system based on mixed models.

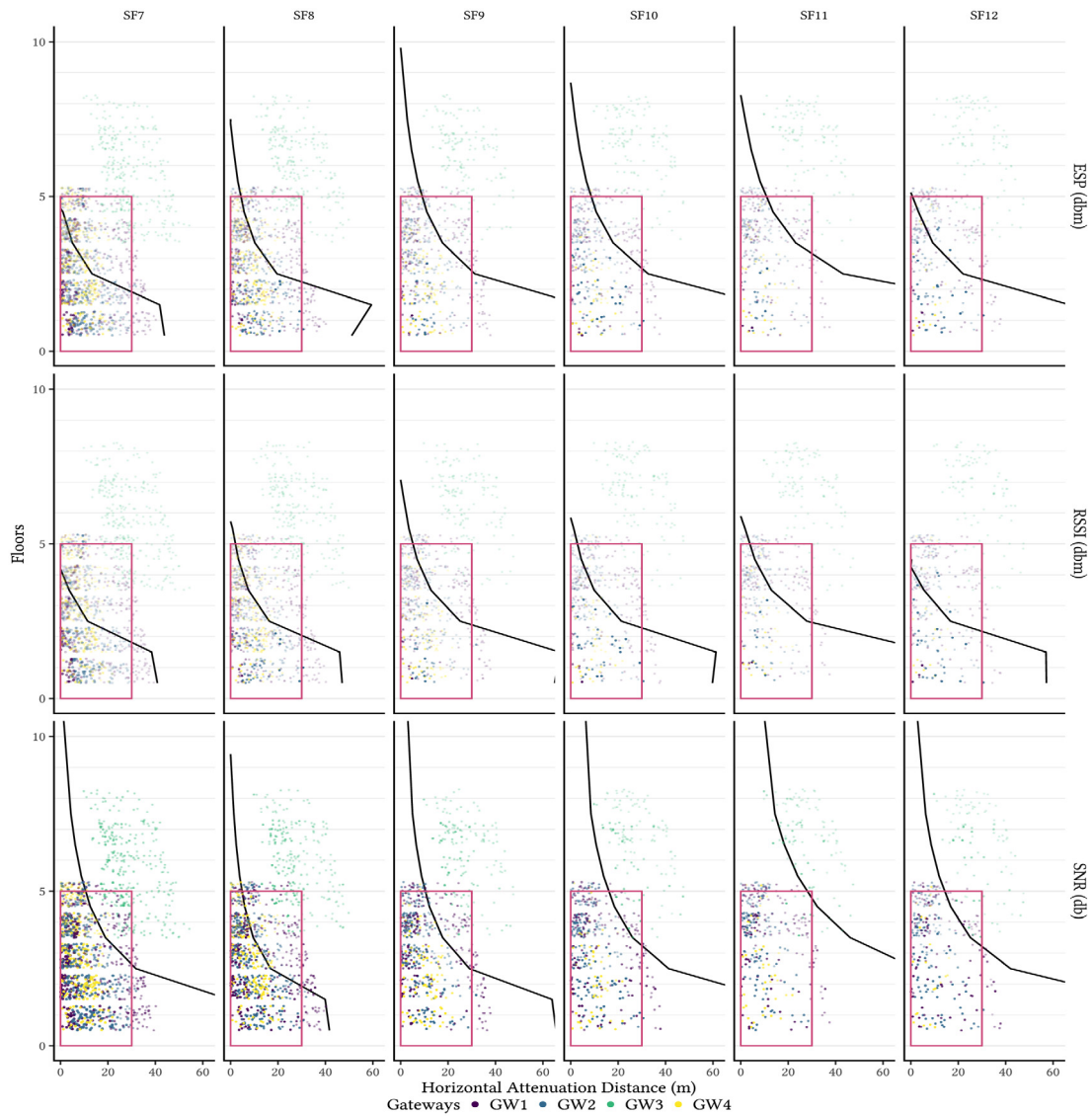
Max. reachable	ESP					
	SF7	SF8	SF9	SF10	SF11	SF12
Max. floor	4.61	7.36	9.81	8.68	8.28	5.13
Max. distance 0 floors up	43.8	51.08	74.46	79.71	104.41	65.68
Max. distance 1 floor up	41.81	50.17	75.63	81.53	111.42	66.93
Max. distance 2 floors up	13.28	19.58	31.26	32.65	43.09	21.99
Max. distance 3 floors up	4.99	10.23	17.65	17.89	23.15	9.26
Max. distance 4 floors up	1.03	5.7	11.06	10.77	13.64	3.25
Max. distance 5 floors up	*	3.03	7.16	6.58	8.08	*
Max. walls 0 floors up	13.27	75.84	65.48	1836.83‡	60.31	9.35
Max. walls 1 floor up	10.51	30.36	45.6	176.45	48.9	8.77
Max. walls 2 floors up	7.47	16.94	32.02	53.14	37.94	7.6
Max. walls 3 floors up	4.54	11.09	24.04	31.32	29.88	5.86
Max. walls 4 floors up	1.69	7.63	18.67	21.01	23.48	3.46
Max. walls 5 floors up	*	5.13	14.67	14.62	17.99	*
Max. concurrent messages†	141.73	198.57	61.18	134.51	422.05	738.42

Note: \* No meaningful (negative) output. † No significant result. ‡ Model deficiency: Positive wall co-efficient. that is only countered by large distances and multiple floors.

While these interactions for the SNR metric model were small and negative, they were substantial and mostly positive for the ESP metric and RSSI metric models, indicating that the signal was becoming weaker at a slower rate as the number of floors or walls increased. The three-way interaction term was negative and significant across all models but had a very small coefficient. Concurrent transmissions occurred so seldom that, in our models, they were not significant, but a negative trend is visible across all models.

To compare all SF, it is necessary to interpret the models' coefficients in terms of limits for transmission (see Table 6, Table 11, and Table 12). Here, we indicate the maximal distances that our system would support according to our models. Under ADR, sensor nodes on SF7 communicate with less transmission power than would be possible otherwise (to conserve battery), which may explain why SF7 underperforms in terms of RSSI compared to higher SF. The negative impact of additional floors on signal quality causes LAP to transmit at maximal power and thus allows for longer distances to be covered on the first floor. Distances between 40.85 m and 166.89 m can be observed within a single floor for all SF but quickly lose signal strength for each extra floor by nearly halving the reachable distance per additional floor. Depending on the SF, the maximal number of reachable floors varies from 4 to 19. However, we observe that after 5 floors the area covered in a floor by the signal is substantially reduced under any SF.

Fig. 15 compares our model output with the received signal data at each gateway. We find that our model accurately identifies strong signal strength but underestimates how long a viable but weak signal can be received. The remote third gateway remained outside of our model's boundary with its reception of transmissions and exhibited generally lower ESP,



**Fig. 15.** Received data quality by floor (3.36 m per floor), distance, and gateway. Different colours for each graph represent different gateways (purple: 1; blue:2; jade: 3; yellow: 4). The black curves show the boundary where the models predict that transmissions occur below the noise floor. Each point represents a sensor-gateway connection, the opacity represents the average signal quality, and the size depends on the number of transmissions received at a gateway compared to the number of transmissions received by the other gateways.

SNR, and RSSI. In general, we observed that increasing the DR (SF8-SF12) allowed for longer ranges similar to outdoor applications, but the signal quickly diminished across distance and with every additional floor. An exception was SF7 which overperformed in SNR but underperformed in RSSI and ESP possibly due to ADR. Overall, we observe high signal quality for transmissions below 30 m and within 5 floors of a GW which we name the 30 × 5 rule.

### 5. Discussion

While the use of indoor LoRAWAN has been previously advocated, research has failed to deliver large scale applications, leaving a gap between the expectation and its realisation. To our knowledge, this is the first project to deploy a LoRAWAN DISN and demonstrate its feasibility in a multi-level building beyond theoretical models and ns-3 simulations [108–112] and provide models of the data transmission that can be used for future research. However, at the time of the conference submission, the Finnish study on DISN with 331 devices and 23.2 million transmission has not yet been published [31]. The study explores a single-floor setup for 2 years but mostly reports descriptive statistics. In contrast, we collected a large dataset with more than 86 million transmissions from 390 sensors across 5 floors over approximately two years

an performed in-depth explanatory and predictive analysis in a complex environment. Other previous work usually only reported at most a couple thousand transmissions (see Table 1). Consequently, we were able to assess the specific real-world factors that affected our LoRAWAN DISN from both explanatory and predictive perspectives beyond what any of the previous literature has offered.

Previous research has focused on descriptive data for the distance limits of specific scenarios that provide some indication of the limitations of such systems [31,45,65]. In some cases, theoretical models were developed to qualify these descriptions, but it is difficult to extend these models to other building structures and obstacles [66,71,73,88]. Distance to the gateway is a key metric for LoRAWAN networks performance in both indoor and outdoor applications. Outdoor applications with minimal interference are typically evaluated with respect to signal attenuation in different environments (e.g., rural versus urban) at the scale of kilometres. In contrast, signal attenuation indoors is at the scale of hundreds of meters because of building structures (e.g., walls and floors) causing signal shadowing and multi-path fading [72,73]. Whereas, in general, log-distances are used for modelling signal decay, we have shown that linear distance works better indoors (see Appendix C) because it is a proxy for complex attenuation. We have also shown that, for DISN with many devices, a linear model or similar approaches such as PCR and PLSR are insufficient and that a mixed model is required (see Table 5) to capture repeated transmissions by different devices in different device classes.

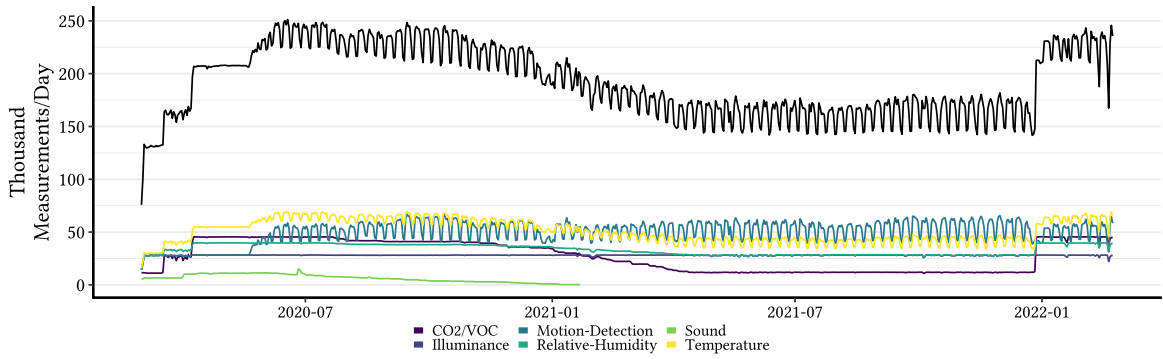
Our mixed model indicates that the signal could be transmitted between 43.8 m and 104.41 m (depending on SF) within the same floor (see Table 6, Table 11, and Table 12) with the caveat that we had no transmissions beyond 64 m. Notably, this estimate is considerably lower than what has been previously reported [6,45,65,77]. The maximal distance further decreased when floor levels and walls are also considered. Here, our mixed model indicates a maximum coverage of 4 to 19 floors (depending on SF) directly above the gateway or at most 65.05 m upwards with the caveat that we had no transmissions beyond 8 floors. Each additional floor upwards roughly halves the signal strength so that coverage on the fifth floor is already below 24 m on any SF and in most cases even below 10 m. Our findings contrast with previous reports of a negligible effect of floors on signal strength [65] and extrapolations that a coverage of one gateway every 10 floors is manageable [43,75]. When comparing our mixed model with the network data, we noticed that the third remote gateway is outside of our area of high quality transmission (jade in Fig. 15) and also has a FER of 73.6%, indicating a useful correlation between our model's predictions and the ability to receive transmissions. Adding a fourth gateway had the biggest impact reducing the overall FER reported on the network side by 73.69%.

Furthermore, we may have underestimated the impact of construction materials on signal strength because many messages were sent barely above the noise floor of ESP, SNR, and RSSI due to the way in which ADR manages SF by selecting only higher SF on short distances when signal strength is weak. This could result in a biased sample of higher SF, but the increased data transmission reliability was required for the real-world application of this DISN to collect data for passively detecting human presence. These results also need to be interpreted carefully because linear modelling of the attenuation in a building may not be sufficient. As construction materials for floors and walls vary from building to building, they impose different constraints on the network. These constraints are difficult to generalise without computationally expensive simulations or large studies on the effects of different complex structures on signal quality.

We used ANNs to investigate an aspect of generalisability based on number of walls, number of floors, and distance which we derived from the BIM in the DT. For the analysis, we excluded several sensor gateway-pairs from the analysis and used them as a test set. We found that these models accurately predicted signal quality (see Fig. 14), despite being presented with a new configuration. All predictive models, LSTM, GRU, Transformer, and MLP, outperformed the explanatory models in terms of RMSE but not necessarily in terms of the mean. Also, all models had a similar success rate except for MLP which deviated more strongly in terms of the mean. While we cannot test scenarios with more complexity, our expectation would be that, as complexity increases, GRU becomes weaker and Transformer should have an advantage for remembering complex relations. The complexity of an indoor environment given for any DISN is too complex to handle for an MLP. While MLPs were successfully used for outdoor predictions, indoor predictions were more complex and ANN may be required.

When comparing the explanatory and the predictive models, we note that the ANN models outperformed the mixed model by 5% to 10%. Despite not being a predictive model, the mixed model performs reasonably well compared to the ANN and even the perfect predictor based on the true data distribution. However, we have no data to check whether such generalisations would hold across buildings. Despite these caveats, we believe to have found a conservative boundary for the deployment of DISNs and strongly recommend the use of multiple gateways placed every 30 m and 5 floors to ensure that as many transmissions as possible are received.

Another major source of discussion for the usability of LoRAWAN DISN is the loss of transmissions. While LoRAWAN is more robust than many other communication technologies, there is a limit on how many messages can be transmitted at the same time. With our setup (1.5 transmissions per second), we observed only 1 to 11 colliding concurrent messages, accounting for 0.42% of all transmissions. Overall, we lost 2.12% of transmissions. We found that multiple gateways have mitigated the loss of transmissions significantly by reducing the impact from environmental factors and network issues such as concurrency. Notably, our use of the third remote GW and newly added fourth gateway provided several important lessons. First, our use of the third GW demonstrated how distances longer than 30 m across multiple floors led to a severe loss in capacity, as indicated by considerably worse FER (73.6%). Second, despite the weak contribution to overall transmission, this third GW still reduced the overall FER per sensor node by 1.31% compared to having only two GWs in the first phase. Third, adding the fourth GW according to the  $30 \times 5$  rule significantly improved the reception of



**Fig. 16.** Measurements per sensor type and day. The total is shown in black. Lower number of measurements before mid May are due to the COVID-19-related lockdown delaying sensor installation. The slow decline in overall transmissions is explained by devices that power down due to an empty battery for the energy-hungry sound and CO<sub>2</sub> sensors. Transmission levels have been restored after exchanging the batteries.

transmissions at the network level. The loss of transmission in the network dropped from 8.06% to 2.12%, reducing the loss by a factor of 4. A well-placed GW can have a significant impact on received signal quality while reusing externally placed gateways for indoor transmissions has only marginal benefits as demonstrated with our third GW.

SF is usually an important factor to consider for the quality of a LoRAWAN network. However, in our setup, the ADR compels our network to use SF7 for 96.3% of transmissions because SF7 provides the highest data throughput at the lowest battery cost. This finding aligns with theoretical expectations that, for short distances, lower SFs are more effective. Contrary to expectations, larger SFs can actually improve signal quality over shorter distances. Interestingly, the larger SFs only provide up to twice the distance coverage of SF7 indoors which contrasts starkly with the performance gain in outdoor scenarios. For indoor applications, we would recommend to develop an ADR that also switches to a higher SF as long as the required data throughput can be maintained. If only standard ADR is available, our distance and floor recommendations ( $30 \times 5$  rule) ensure a high quality signal throughout a building regardless of the chosen SF. Whereas it would be interesting to study the system without ADR, this project is part of a data collection effort that required the best configuration to ensure data availability from the sensors in the devices for the research delineated in the next section.

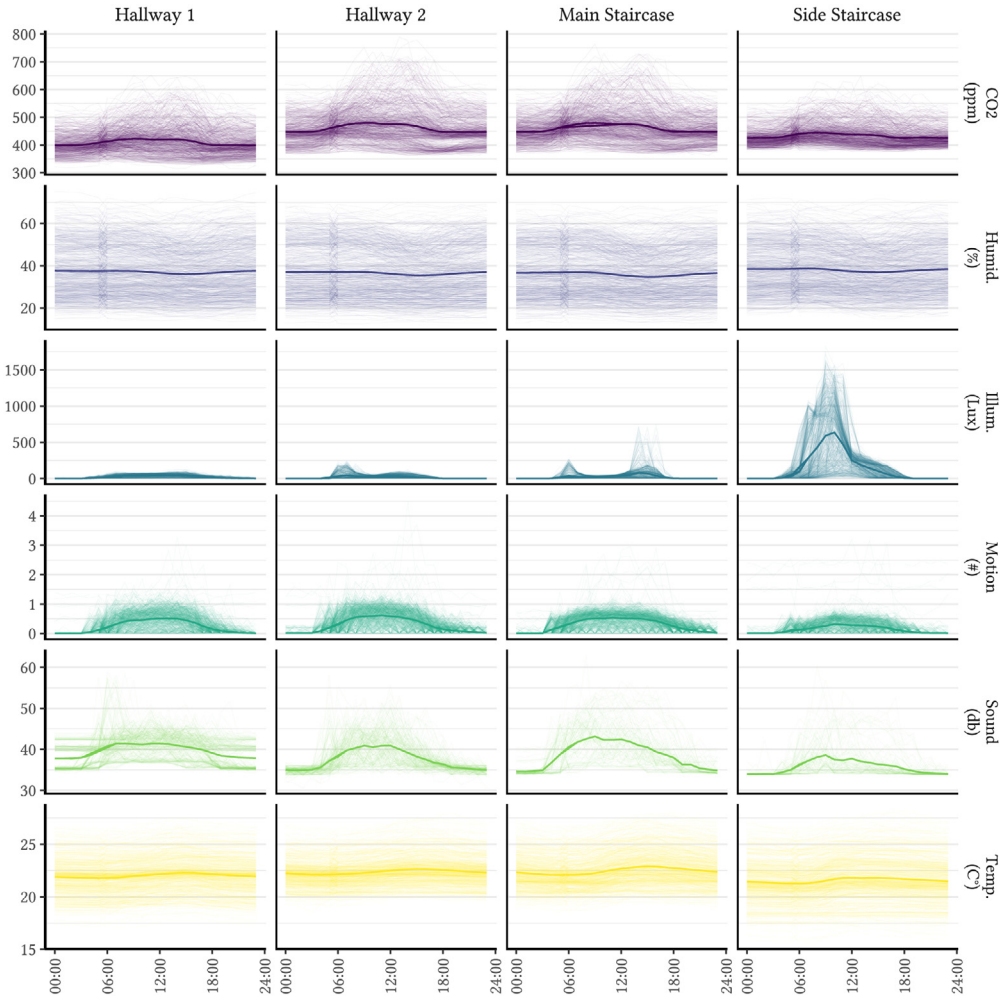
## 6. Case study–Human presence during COVID19 lockdowns

The core focus of this paper was on the feasibility of the network infrastructure. However, while testing the network, we also collected a large data set that continues to increase by an average 55,000 data points every day. Between March 8th 2020 and February 25th 2022, our prototype collected 139.5 million data points (see Fig. 16). Before data collection, 137 sensors were installed, but due to COVID-19, the installation of the remaining 253 sensors was delayed until mid-May 2020. Some devices had a higher than expected battery consumption requiring an exchange of batteries in December 2021.

An overview over the data variation is presented in Fig. 17. The data is shown in 24-hour cycles aggregated by room over the whole period for the second floor of the building. Several daily rhythms in the building are visible when looking at the average across days of hourly averages. First, during the first half of the day, CO<sub>2</sub> concentrations are higher and drop during the second half of the day. There are two peaks, one in the morning hours and one in the afternoon hours, in most rooms. Humidity generally drops during the night and rises throughout the day. Peaks in illumination are connected with daytime hours and probably indicate sunlight reaching through the windows. The side-staircase has mostly glass walls and receives the highest illumination. Motion is detected throughout the day with slight peaks before and after noon. Sound is distributed similarly to motion. Temperature slightly rises throughout the morning and decreases thereafter. Several of these variations intuitively are connected to human activity and have been shown to correlate with human activity in previous research [80–82].

While a ground truth dataset (i.e., number of people counted) was not yet available to validate human presence models, we used the COVID-19-related lockdown to investigate whether variations in the sensor data could indicate human presence. We take two different approaches to derive human presence from the data with temporal and spatial analyses. First, we look at the fine-grained variation in the temporal signal and collapse the spatial dimension. Second, we look at spatial variation when we collapse the time signal into discrete blocks corresponding to key events during the COVID-19-related lockdown.

For the temporal analysis, we create a variance indicator per room across time that describes how much a value differs from the overall average of the room. We summarise the sensor data at the room level as variance indicators  $\mathbb{V}_{r,t}$  (see Eq. (10)), which aggregate data by room  $r$  and time slot  $t$  (hourly) based on the sensor type  $i$  and the individual sensor device  $j$ . Because of the different scales of these measurements, we use z-scores by sensor type  $i$  so that each measurement  $v_{r,i,j,t_k}$  is aggregated over all  $j$  devices in a room  $r$  and measurements  $k$  in time slot  $t$  (see Eq. (9)). As some extraordinary



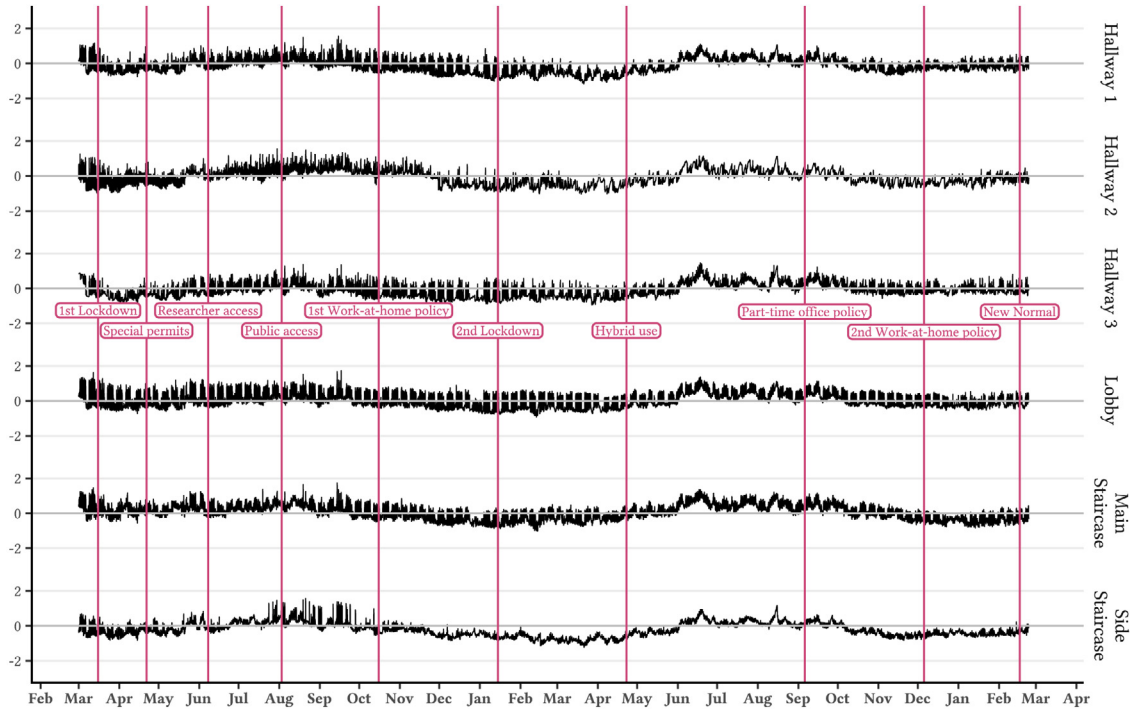
**Fig. 17.** Sensor data on the second floor. The soft lines indicate the values for each day. The strong line indicates the average over all days. The shaded areas represent the range between the minimal and maximal values measured.

events (e.g., construction with noise up to 14 times louder than average) have resulted in extreme outliers, we eliminated data points beyond 3 standard deviations and then aggregated with respect to sensor type  $i$  (see Eq. (10)).

$$\tilde{V}_{r,i,t} = \frac{1}{n_{r,i}} \sum_{j,k} \frac{v_{r,i,j,t_k} - \bar{v}_i}{\sigma_{v_i}} \tag{9}$$

$$V_{r,t} = \frac{1}{n_r} \sum_i \max(-3, \min(3, \tilde{V}_{r,i,t})) \tag{10}$$

Our variance indicator allows us to observe strong periodic variation in each room and to tentatively attribute portions of the variation to human presence. See Fig. 18 for our analysis of the ground floor. Because of the COVID-19 lockdown at the beginning of the study, we were able to attribute some of the later variations to returning human presence by marking key events that occurred inside the building. On March 13th 2020, the lockdown was enacted, and until April 22nd 2020, only essential personnel (e.g., building management and reception) were still present. From April 22nd 2020 to June 8th 2020, a slight easing of the lockdown was introduced that allowed some researchers to return with special permits. From June 8th 2020 to August 3rd 2020, all researchers were allowed back in the building. From August 3rd 2020 to October 16th 2020, the building was again opened to the public with students and visitors allowed to enter without special permits. From October 16th 2020 to January 15th 2021, a first work-at-home order was issued recommending personnel to work from home. From January 15th 2021 to April 23rd 2021, a second lockdown was enacted. From April 23rd 2021 to September 6th 2021, the building was opened to the public again. From September 6th 2021 until December 6th 2021, a part-time office policy was in place. From December 6th 2021 until February 17th 2022, a second work-at-home order was issued recommending personnel to work from home. From February 17th 2022 onwards, a new normal



**Fig. 18.** Sensor data aggregated by hour and room for the ground floor from March 2020 until February 2022. The variance indicator represents variation up to  $\pm 3$  standard deviations. The magenta lines indicate key events related to the COVID-19 lockdown. The grey line at 0 shows the standard variance level across the whole observed time period. The lockdowns and work-at-home policy periods show strongly reduced variance whereas the public access and hybrid use periods show increased variance.

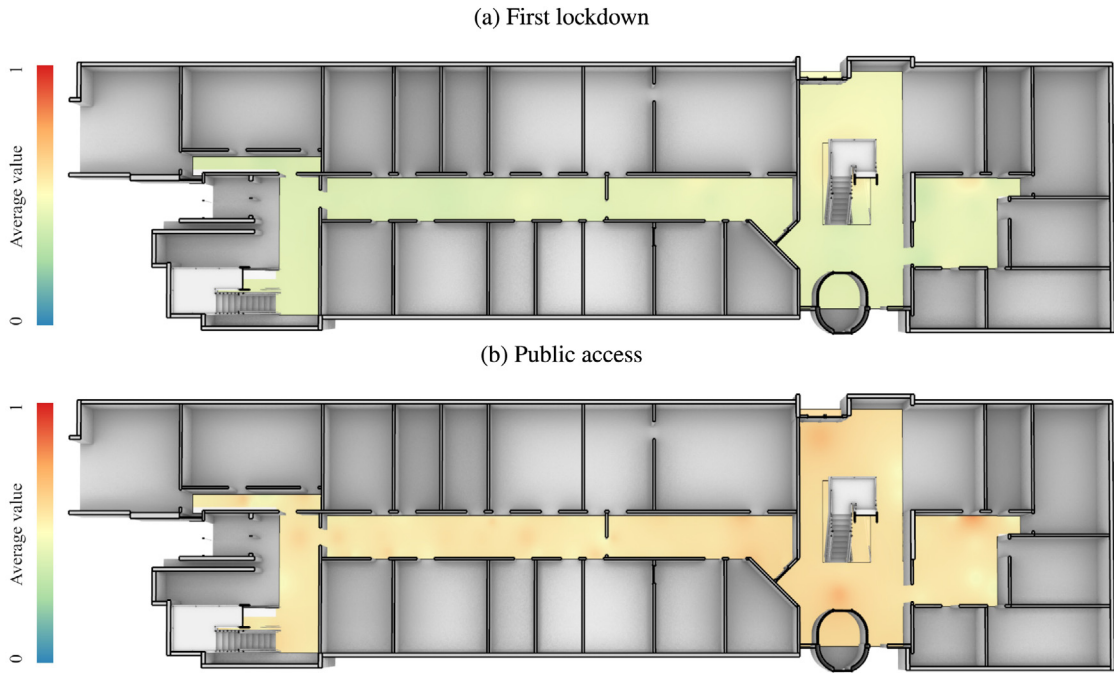
was declared. The strongest impact on variations in the data was when presence was strongly reduced during the first and second lockdown periods. This is most visible in the Side Staircase, Hallways 1 and 3, and to a lesser degree in Hallway 2. Two major exceptions were the Main Staircase and the Lobby where the offices of essential personnel are located. After the special permits were issued, some presence returned to the building, producing additional variation. When full access was restored for researchers, a regular pattern in variation returned.

For the spatial analysis, we used the key events of the COVID-19 lockdown to define time spans across which to aggregate the data per sensor (see Eq. (11)). In principle, the aggregation follows the pattern for the temporal analysis but aggregates across different indices. Eq. (11) is similar to Eq. (9) and (10) as it follows the same form of aggregation including the limits on outliers to 3 standard deviations. Instead of aggregating by room, each sensor  $j$  is kept on its own. However, the time slot  $t$  is expanded to cover the whole time span defined by the key events and therefore uses a larger number of measurements  $k$ .

$$\mathbb{V}_{j,t} = \max \left( -3, \min \left( 3, \frac{1}{n_i} \sum_k \frac{v_{i,j,t_k} - \bar{v}_i}{\sigma_{v_i}} \right) \right) \quad (11)$$

In order to visualise the sensor data, the physical positioning of each sensor (x,y,z) is used to represent each sensor spatially in a BIM of the building. Using Inverse Distance Weighted (IDW) interpolation [113], a density map showing the spread and variability in different sensor measurements (e.g., average value, minimum and maximum values) is visualised (see Fig. 19 and Appendix I). Whereas this visualisation cannot yet be used to describe human presence directly, a visual inspection of it may be useful to estimate how changes in human presence may be captured by the DISN. Towards this end, we generate two visualisations of the sensor measurements from two events: (1) the first lockdown in Spring 2020 (see Fig. 19(a)) and (2) the restored public access in Summer 2020 (see Fig. 19(b)). In both events, human presence was strictly regulated by organisational policies, providing an unique opportunity to estimate the system's ability to capture human presence. A visual inspection of Fig. 19(a) clearly denotes that the sensor measurements during restored public access in Summer 2020 is different in intensity and spread when compared to the first lockdown in Spring 2020. Specifically, the measurement intensity during the restored public access in Summer 2020 seems higher, corresponding to employees' and students' returns to the office and regular activities after the first lockdown.

The focus of this case study is delineating the ability of the sensor network to detect human presence. However, to explore the data, we also made use of the FT to prototype a visualisation of the density maps *in situ* (see Fig. 20) for human-building interaction [25,114,115]. We walked the hallways of the building and were able to immerse ourselves



**Fig. 19.** Spatial visualisations in sensor data comparing variation according to Eq. (11) during the first lockdown and restored public access. The data is interpolated with IDW [113]. The yellow represents the average variation across the whole time period. Increased variation is indicated in red, and decreased variation is indicated in blue.



**(a)** A hallway as seen through the HoloLens. The ground is covered in a visualisation of the variation during restored public access. **(b)** The same hallway seen through the HoloLens from the other end and rotated by 180 degrees. The side staircase is made visible in AR in the distance.

**Fig. 20.** A prototype visualisation of density maps in the FT [26]. The footage was recorded from a HoloLens. The coloured volume on the ground is based on the density map from Fig. 19(b). Without the depth cues from the HoloLens, the volumetric density map covering the floor of the building would be difficult to differ from the carpet in the images.

in the density maps (see Fig. 19). This allows us to spatially explore the data distribution and hopefully reduce cognitive load by providing spatial context [116]. While these first visualisations are crude, we plan to advance the theory and application of *in situ* visualisations in future work to enhance immersive analytics [117].

Analysing human presence indoors can be a difficult but useful endeavour for researchers, designers, and planners. This new and rich data set of environmental characteristics from a real building will allow us to develop novel applications for analysing the impact of human presence in an enclosed public space. While we lacked ground truth measurements to validate any kind of occupancy model, we benefited from the natural experiment resulting from the COVID-19 lockdowns. This allowed us to observe patterns of variation in the data that could only be explained by the sudden absence of human presence in the building over an extended period of time. Our immediate next steps will be to compile a ground truth dataset for human presence against which we will be able to run more complex models that can spatially describe human

activity within a building. Towards this end, we will use automated visitor counts and localisation derived from videos, while at the same time address the privacy concerns that are related to applying such technology.

While accuracy is expected to be much lower with passive tracking compared to active tracking, our approach benefits from the preservation of privacy, measures that are not biased by sampling, and the observation of more natural user behaviour. Passive tracking elegantly circumvents the privacy issue by estimating changes in environmental measurements and “naturally aggregating” data from stationary sensors without reference to the individuals’ identities. Our approach can also reduce bias in modelling human activity in a building because we account for the whole population of a building rather than convenience samples of people who allow active tracking. Finally, the passive nature of the system also allows us to observe more natural behaviour from building users, reducing the observer effects that can occur with active tracking.

## 7. Conclusion

LoRAWAN indoors has been investigated with a long series of proof-of-concept systems that indicated the potential for fine-grained data indoors. Nonetheless, most previous work has relied on few sensor nodes or simulations without a dense network implementation. We built the first LoRAWAN DISN that extends LoRAWAN’s appeal beyond its initial technical scope of long-range, low-power applications. Indoor environments attenuate the LoRAWAN signal at a quicker rate compared to outdoor scenarios, and previous estimates for indoor ranges of 200 m to 600 m seem optimistic [6,31,45,77]. This may be due to the fact that the formulation of outdoor signal propagation models is inappropriate for indoor scenarios. We have shown that linear models are not sufficient to analyse DISN scenarios and that mixed models are required. Furthermore, the prospect of signal shadowing and multi-path fading breaks the assumptions behind the log-distance in signal propagation models. We have shown that for dense indoor scenarios, a linear attenuation factor based on distance is more effective at capturing the impact of signal shadowing and multi-path fading. We also opt for a new signal quality indicator called ESP [47,48] because it combines the information in SNR and RSSI and therefore has more range, especially close to the noise floor. Our PLSR analysis has shown that all three measures are highly related (see Fig. 13).

Based on our mixed model and the ESP, our system achieved good coverage up to 30 m with an upper bound around 111 m at SF11. Higher SFs improve signal quality but only marginally increase distance coverage to up to twice that of SF7 due to signal shadowing and multi-path fading. Similarly, the impact of floors on signal quality resulted in nearly halving of reachable distance per additional floor. Walls also attenuate the signal but at a slower rate than floors, generally limiting the maximal room coverage to tens of rooms. However, the complexity of signal shadowing and multi-path fading can likely not be captured with an explanatory model. We used our DT to evaluate obstacles along transmission lines in the BIM to produce highly accurate predictive models based on modern ANN such as LSTM, GRU, and Transformers. Whereas these predictive models have no theoretical basis how to incorporate this information, they predicted the signal quality indicators more accurately. Interestingly, the widely used MLP performed worst. For future work, we recommend to switch to LSTM, GRU, or Transformers, all of which can be easily implemented with modern ANN libraries.

To our knowledge, this is the first validation of previous simulation research using multiple gateways [109]. We demonstrated the positive impact of an additional fourth gateway by recovering up to four times more lost messages in real-world conditions compared to three gateways. While there are only marginal gains with every added gateway, having several gateways improves the reliability of the system. Our recommendation focuses on obtaining the strongest possible signal while minimising energy costs at the nodes and data loss. Specifically, we recommend the placement of one LoRAWAN gateway every 30 m and 5 floors to make maximal use of the effective coverage a gateway can provide. A dense gateway placement is critical for the effective deployment of LoRAWAN DISNs. Placing LoRAWAN gateways indoors is therefore more similar to placing WiFi-routers than outdoor LoRAWAN gateways. While LoRAWAN’s performance indoors on average may approximate technologies such as WiFi under optimal conditions (at 45 m), the energy consumption of LoRAWAN devices is lower, and they do not require wires. We believe that, for pure sensing applications, LoRAWAN offers many advantages and can help to increase sensing density in buildings while keeping maintenance costs low.

Our next steps will be to extend the research on LoRAWAN DISNs and passively sensing human presence with a DT. We will also develop variations of ADR and apply the models to new environments to test their generalisability. Explanatory models could be improved by combining PLSR and Mixed Models to account both for the latent structure in the data and the error. The application of predictive models based on modern ANN architectures requires more research. We could not select the best ANN, except dethroning the MLP as the de facto standard. We will investigate a larger set of ANN for their suitability for signal modelling as well as performing effective hyperparametrisation to compare the strongest possible candidates. Furthermore, the data from the Finish study [31] will be requested to perform out of sample predictions and see whether the combination of a DT for transmission meta data information (e.g. wall meta data) is enough to produce effective predictions. For the analysis of human presence, we will generate ground truth datasets of human activity in the building and explore the sensor data in more detail. Furthermore, we will expand this prototype DT into a fully-fledged DT platform for human building interaction including advanced visualisation through FT [26], which together can enable predictions, statistical modelling, user studies [84], and agent simulations [118] and offer both researchers and analysts a new toolbox to study human behaviour in the build environment.

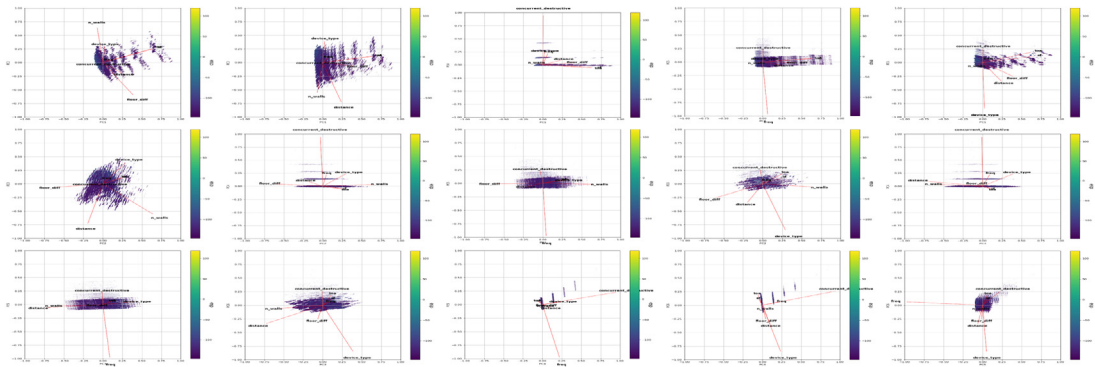


Fig. 21. Comparison of PCA dimensions on ESP.

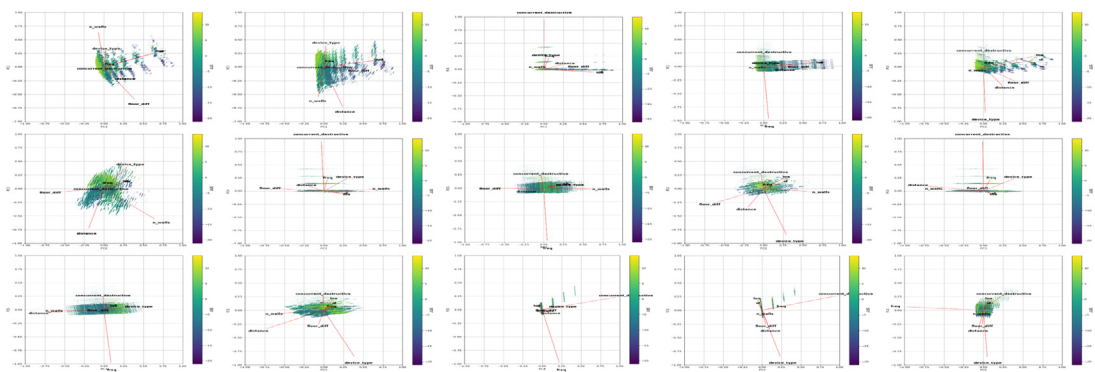


Fig. 22. Comparison of PCA dimensions on SNR.

**CRedit authorship contribution statement**

**Jascha Grübel:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Tyler Thrash:** Funding acquisition, Methodology, Writing – review & editing. **Leonel Aguilar:** Formal analysis, Methodology, Software, Validation, Visualization, Writing – review & editing. **Michal Gath-Morad:** Formal analysis, Methodology, Software, Visualization, Writing – review & editing. **Didier Héjal:** Methodology, Resources, Writing – review & editing. **Robert W. Sumner:** Project administration, Supervision, Writing – review & editing. **Christph Hölscher:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Victor R. Schinazi:** Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgements**

We thank Axel Beckert for feedback and for the invitation by Amy Lynn Murphy in the name of the PerCom’21 organisers to submit this extended version.

**Appendix A. Acronyms**

ADR Adaptive Data Rate

ANN Artificial Neural Network

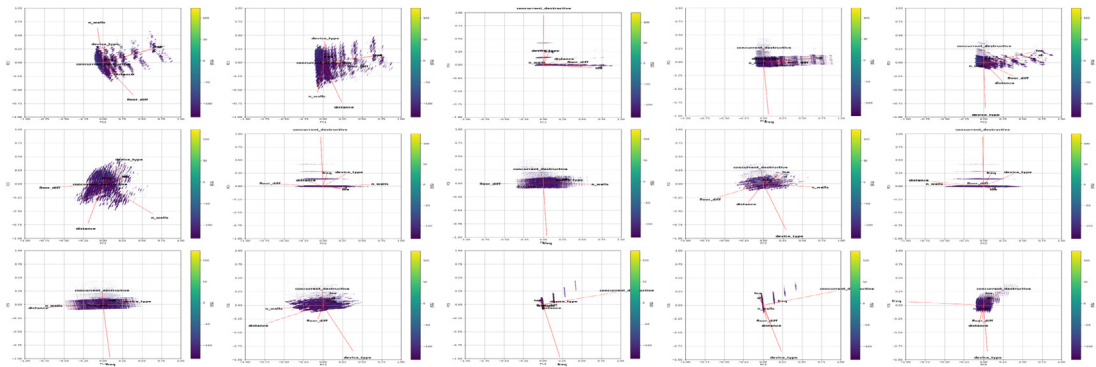


Fig. 23. Comparison of PCA dimensions on RSSI.

Table 7  
Comparison for linear and mixed models with linear and log-distance model.

Model	Perfect estimator		Mixed models				Linear models			
	Mean	RMSE	Linear distance		Log distance		Linear distance		Log distance	
			Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
	ESP	-107.43	6.30	<b>-107.87</b>	<b>9.06</b>	-107.92	9.82	-110.26	10.47	-110.28
SNR	2.58	2.82	2.54	<b>3.84</b>	<b>2.57</b>	4.19	2.58	6.64	2.58	6.64
RSSI	-104.44	5.59	<b>-104.92</b>	<b>8.34</b>	-104.99	8.89	-106.92	9.47	-106.96	9.85

Note: Mixed models specification see [Appendix G](#).

AR Augmented Reality

BIM Building Information Model

BW Bandwidth

CF Carrier Frequency

DISN Dense Indoor Sensor Network

DR Data Rate

DT Digital Twin

ESP Effective Signal Power

EUI Extended Unique Identifier

FEC Forward Error Correction

FER Frame Error Rate

FT Fused Twins

GRU Gated Recurrent Unit

GW gateway

IDW Inverse Distance Weighted

IoT Internet of Things

ISM Industrial, Scientific, and Medical Frequency Bands

JSON JavaScript Object Notation

**Table 8**  
Mixed models to determine ESP for each DR.

	ESP					
	SF7	SF8	SF9	SF10	SF11	SF12
Average signal strength at node (constant)	-60.944* (3.977)	-102.940* (2.053)	-102.679* (1.454)	-104.841* (2.174)	-94.914* (2.233)	-68.505* (3.182)
Attenuation in m	-1.604* (0.007)	-0.655* (0.007)	-0.525* (0.007)	-0.531* (0.007)	-0.533* (0.009)	-1.333* (0.009)
One floor between	-10.836* (0.073)	-2.608* (0.051)	-2.422* (0.059)	-3.329* (0.065)	-5.250* (0.084)	-13.536* (0.054)
One wall between	-0.160 (0.050)	1.661* (0.044)	1.085* (0.050)	1.679* (0.057)	0.786* (0.080)	-5.113* (0.083)
Colliding message	-0.496 (0.280)	-0.169 (0.514)	-0.638 (0.526)	-0.314 (0.458)	-0.132 (0.278)	-0.119 (0.202)
CF 867.3 MHz	0.562* (0.103)	0.718* (0.051)	0.647* (0.049)	0.622* (0.058)	0.655* (0.063)	0.424* (0.083)
CF 867.5 MHz	0.942* (0.104)	1.042* (0.052)	0.921* (0.050)	0.861* (0.059)	1.074* (0.063)	0.778* (0.084)
CF 867.7 MHz	-0.079 (0.102)	-0.187* (0.051)	-0.170* (0.049)	-0.311* (0.058)	-0.260* (0.063)	-0.352* (0.082)
CF 867.9 MHz	-0.570* (0.103)	-0.432* (0.052)	-0.514* (0.049)	-0.495* (0.058)	-0.443* (0.063)	-0.682* (0.083)
CF 868.1 MHz	0.339* (0.102)	0.273* (0.051)	0.312* (0.049)	0.419* (0.054)	0.402* (0.055)	0.097 (0.070)
CF 868.3 MHz	0.408* (0.102)	0.289* (0.051)	0.392* (0.049)	0.488* (0.054)	0.489* (0.055)	0.094 (0.070)
CF 868.5 MHz	-0.434* (0.102)	-0.596* (0.051)	-0.405* (0.049)	-0.107 (0.054)	-0.128 (0.055)	-0.420* (0.070)
ToA	-0.033 (0.046)	0.081* (0.006)	0.011* (0.003)	0.012* (0.002)	-0.002 (0.001)	0.002 (0.001)
Floor-attenuation interaction	0.288* (0.002)	0.079* (0.001)	0.061* (0.001)	0.072* (0.002)	0.094* (0.002)	0.280* (0.002)
Wall-attenuation interaction	-0.006* (0.002)	-0.054* (0.001)	-0.033* (0.002)	-0.041* (0.002)	-0.029* (0.002)	0.123* (0.003)
Wall-floor interaction	-0.109* (0.017)	-0.663* (0.013)	-0.262* (0.014)	-0.394* (0.015)	-0.164* (0.021)	1.293* (0.021)
Wall-floor-attenuation interaction	0.002* (0.001)	0.019* (0.0004)	0.007* (0.0004)	0.008* (0.0005)	0.005* (0.001)	-0.028* (0.001)
Random effects						
# Devices	390	330	229	184	149	139
Devices Std. Dev.	5.90	8.43	7.55	11.94	7.67	12.26
# Device Types	5	5	5	5	5	5
Device Type Std. Dev.	5.72	4.06	2.46	3.60	4.20	4.46
Observations	145,932	149,763	149,684	149,349	149,568	147,694

Note: \*p<0.001; \*\*p<[0.\*\*]; \*\*\*p<[0.\*\*\*].

LAP Link Adaptation Policy

LF Loading Factor

LoRA PHY LoRa Transceiver

LoRAWAN Long Range Wide Area Network

LPWAN Low Power Wide Area Network

LSTM Long Short-Term Memory

MLP Multi-Layer Perceptron

PC Principal Component

PCA Principal Component Analysis

PCR Principal Component Regression

PLSR Partial Least Squares Regression

**Table 9**  
Mixed models to determine SNR for each DR.

	SNR					
	SF7	SF8	SF9	SF10	SF11	SF12
Average signal strength at node (constant)	11.087* (1.209) (0.122)	7.051* (0.812) (0.174)	8.069* (1.031) (0.266)	2.025 (1.086) (0.333)	6.865* (1.316) (0.376)	2.021 (1.890) (0.475)
Attenuation in m	-0.246* (0.002)	-0.409* (0.005)	-0.314* (0.006)	-0.183* (0.006)	-0.146* (0.007)	-0.229* (0.006)
One floor between	-0.763* (0.024)	-0.524* (0.037)	-0.332* (0.049)	-0.249* (0.052)	-0.886* (0.066)	-0.966* (0.036)
One wall between	0.965* (0.017)	1.374* (0.033)	1.207* (0.042)	1.981* (0.046)	2.171* (0.064)	1.112* (0.055)
Colliding message	-0.287 (0.093)	-0.612 (0.380)	-0.727 (0.441)	-0.248 (0.367)	-0.042 (0.222)	-0.093 (0.134)
CF 867.3 MHz	-0.091 (0.034)	0.001 (0.038)	-0.021 (0.041)	-0.086 (0.046)	-0.020 (0.050)	-0.194* (0.055)
CF 867.5 MHz	-0.550* (0.035)	-1.234* (0.038)	-1.412* (0.042)	-1.428* (0.047)	-1.404* (0.050)	-1.192* (0.056)
CF 867.7 MHz	-0.046 (0.034)	0.018 (0.038)	-0.054 (0.041)	-0.101 (0.046)	-0.115 (0.050)	-0.270* (0.055)
CF 867.9 MHz	-0.166* (0.034)	-0.125 (0.038)	-0.202* (0.041)	-0.119 (0.047)	-0.109 (0.050)	-0.388* (0.055)
CF 868.1 MHz	-0.080 (0.034)	-0.205* (0.038)	-0.155* (0.041)	-0.085 (0.043)	-0.045 (0.044)	-0.290* (0.047)
CF 868.3 MHz	-0.122* (0.034)	-0.212* (0.038)	-0.169* (0.041)	-0.047 (0.043)	-0.003 (0.044)	-0.314* (0.047)
CF 868.5 MHz	-0.110 (0.034)	-0.322* (0.038)	-0.224* (0.041)	0.051 (0.043)	0.081 (0.044)	-0.237* (0.047)
ToA	-0.014 (0.015)	0.031* (0.005)	-0.005 (0.003)	0.003 (0.002)	-0.004* (0.001)	0.001 (0.001)
Floor-attenuation interaction	0.006* (0.001)	0.027* (0.001)	0.013* (0.001)	-0.005* (0.001)	-0.0004 (0.001)	0.015* (0.001)
Wall-attenuation interaction	-0.031* (0.001)	-0.038* (0.001)	-0.032* (0.001)	-0.054* (0.001)	-0.067* (0.002)	-0.038* (0.002)
Wall-floor interaction	-0.296* (0.006)	-0.505* (0.009)	-0.260* (0.011)	-0.446* (0.012)	-0.461* (0.016)	-0.260* (0.014)
Wall-floor-attenuation interaction	0.006* (0.0002)	0.013* (0.0003)	0.005* (0.0003)	0.010* (0.0004)	0.013* (0.0005)	0.010* (0.0004)
Random effects						
# Devices	390	330	229	184	149	139
Devices Std. Dev.	1.71	1.86	2.24	2.14	2.24	3.49
# Device Types	5	5	5	5	5	5
Device Type Std. Dev.	1.50	1.26	1.80	1.65	2.39	2.74
Observations	145,932	149,763	149,684	149,349	149,568	147,694

Note: \*p<0.001; \*\*p<[0.\*\*]; \*\*\*p<[0.\*\*\*].

RMSE Root Mean Square Error

RSSI Received Signal Strength Indicator

SF Spreading Factor

SNR Signal to Noise Ratio

SoA Service-oriented Architecture

ToA Time-on-Air

## Appendix B. PCA biplots

See Figs. 21–23.

## Appendix C. Linear distance vs log-distance indoors

The log-distance models performed worse than the linear distance models in comparison with the perfect estimator. While improvements in the mean are marginal, the RMSE is substantially lower for the linear distance model. Furthermore,

**Table 10**  
Mixed models to determine RSSI for each DR.

	RSSI					
	SF7	SF8	SF9	SF10	SF11	SF12
Average signal strength at node (constant)	-60.617* (3.704)	-102.438* (1.845)	-103.536* (0.969)	-102.634* (1.551)	-94.406* (1.626)	-63.217* (2.458)
Attenuation in m	-1.527* (0.007)	-0.501* (0.005)	-0.392* (0.004)	-0.491* (0.005)	-0.507* (0.006)	-1.272* (0.007)
One floor between	-10.749* (0.069)	-2.687* (0.038)	-2.508* (0.037)	-3.692* (0.041)	-5.224* (0.056)	-13.809* (0.043)
One wall between	-0.563* (0.047)	0.926* (0.033)	0.432* (0.031)	0.452* (0.036)	-0.774* (0.053)	-6.244* (0.065)
Colliding message	-0.405 (0.264)	0.101 (0.379)	-0.256 (0.329)	-0.251 (0.289)	-0.111 (0.185)	-0.047 (0.159)
CF 867.3 MHz	0.599* (0.097)	0.702* (0.038)	0.620* (0.031)	0.660* (0.036)	0.644* (0.042)	0.555* (0.065)
CF 867.5 MHz	1.150* (0.098)	1.534* (0.038)	1.617* (0.031)	1.733* (0.037)	2.037* (0.042)	1.684* (0.066)
CF 867.7 MHz	-0.059 (0.097)	-0.220* (0.038)	-0.192* (0.031)	-0.303* (0.037)	-0.228* (0.042)	-0.194 (0.065)
CF 867.9 MHz	-0.498* (0.097)	-0.408* (0.038)	-0.464* (0.031)	-0.494* (0.037)	-0.413* (0.042)	-0.435* (0.065)
CF 868.1 MHz	0.371* (0.097)	0.331* (0.038)	0.341* (0.030)	0.424* (0.034)	0.394* (0.037)	0.263* (0.055)
CF 868.3 MHz	0.458* (0.096)	0.349* (0.038)	0.422* (0.031)	0.464* (0.034)	0.453* (0.036)	0.283* (0.055)
CF 868.5 MHz	-0.389* (0.097)	-0.477* (0.038)	-0.308* (0.031)	-0.168* (0.034)	-0.207* (0.036)	-0.294* (0.055)
ToA	-0.031 (0.043)	0.078* (0.005)	0.020* (0.002)	0.016* (0.001)	0.001 (0.001)	0.002 (0.001)
Floor-attenuation interaction	0.292* (0.002)	0.078* (0.001)	0.066* (0.001)	0.094* (0.001)	0.113* (0.001)	0.296* (0.001)
Wall-attenuation interaction	0.006* (0.002)	-0.032* (0.001)	-0.014* (0.001)	-0.003 (0.001)	0.022* (0.002)	0.168* (0.002)
Wall-floor interaction	0.015 (0.016)	-0.418* (0.009)	-0.147* (0.008)	-0.131* (0.009)	0.159* (0.014)	1.580* (0.017)
Wall-floor-attenuation interaction	-0.0002 (0.001)	0.012* (0.0003)	0.004* (0.0003)	0.002* (0.0003)	-0.004* (0.0004)	-0.040* (0.001)
Random effects						
# Devices	390	330	229	184	149	139
Devices Std. Dev.	5.69	8.43	7.35	11.86	7.79	11.05
# Device Types	5	5	5	5	5	5
Device Type Std. Dev.	5.23	3.71	1.43	2.21	2.90	3.14
Observations	145,932	149,763	149,684	149,349	149,568	147,694

Note: \*p<0.001; \*\*p<[0.\*\*]; \*\*\*p<[0.\*\*\*].

the linear component can be interpreted more easily than the log-distance component. We assume that the impact of shadowing and multi-path fading is captured more accurately with a linear factor and that the rather short distances indoors make the log-effects of distance vanish. Since the signal decays so quickly indoors, the log-component is not required and may be even detrimental to the model quality in our case. The linear models estimate the mean worse than the log-distance models for ESP and RSSI and have overall a much worse RMSE. This pattern may be expected since the Mixed Models give structure to the error term and allow us to explain variance.

**Appendix D. Mixed model tables**

See Tables 8–10.

**Appendix E. Max transmission distance SNR and RSSI**

See Tables 11–12.

**Appendix F. Neural network specification**

See Tables 13–16.

**Table 11**  
SNR Limits for Transmission across Floors in our System Based on Mixed Models.

Max. reachable	SNR					
	SF7	SF8	SF9	SF10	SF11	SF12
Max. floor	11.85	9.43	15.3	19.36	17.68	13.07
Max. distance 0 floors up	75.61	41.69	65.53	93.23	166.89	96.27
Max. distance 1 floor up	70.98	39.88	63.85	86.13	157.01	95.1
Max. distance 2 floors up	31.77	16.87	29.01	41.29	73.9	42.03
Max. distance 3 floors up	18.93	9.54	17.71	26.08	46.16	25.13
Max. distance 4 floors up	12.54	5.93	12.12	18.43	32.28	16.82
Max. distance 5 floors up	8.72	3.79	8.79	13.83	23.94	11.87
Max. walls 0 floors up	*	*	*	*	*	*
Max. walls 1 floor(s) up	*	46.27	*	*	*	*
Max. walls 2 floor(s) up	53.43	18.42	*	*	*	*
Max. walls 3 floor(s) up	27.75	11.35	44.4	*	*	*
Max. walls 4 floor(s) up	18.67	7.98	29.5	*	*	*
Max. walls 5 floor(s) up	13.83	5.94	22.44	*	*	*
Max. concurrent messages† 64.67	27.84	28.28	68.54	586.47	236.67	

Note: \* No meaningful (negative) output. † No significant result.

**Table 12**  
RSSI Limits for Transmission across Floors in our System Based on Mixed Models.

Max. reachable	RSSI					
	SF7	SF8	SF9	SF10	SF11	SF12
Max. floor	4.19	5.73	7.07	5.84	5.89	4.26
Max. distance 0 floors up	40.85	47.04	64.97	59.78	76.15	57.2
Max. distance 1 floor up	38.43	45.99	66.99	61.28	81.33	57.02
Max. distance 2 floors up	11.44	16.32	25.11	21.38	27.89	16.72
Max. distance 3 floors up	3.66	7.52	12.8	9.89	12.93	5.55
Max. distance 4 floors up	*	3.29	6.92	4.44	5.89	*
Max. distance 5 floors up	*	*	3.46	1.26	1.79	*
Max. walls 0 floors up	11.45	34.66	30.91	26.21	16.12	7.07
Max. walls 1 floor up	8.99	19.46	23.89	20.36	14.58	6.38
Max. walls 2 floors up	6.18	11.83	17.87	14.75	12.27	5.12
Max. walls 3 floors up	3.36	7.45	13.31	10.12	9.55	3.25
Max. walls 4 floors up	*	4.36	9.6	6.17	6.47	*
Max. walls 5 floors up	*	1.8	6.34	2.68	3.12	*
Max. concurrent messages†	154	*	99.64	116.95	347.17	1564.45

Note: \* No meaningful (negative) output. † No significant result.

**Table 13**  
Specification for LSTM model.

	Model	
Input shape:	(19,3)	
Input structure:	(16 hops with 3 features, 8 parameters + 1 padding value arranged in 3 features)	
Layer	Output shape	# Parameters
LSTM	(None, 50)	10800
Dense	(None, 3)	153

Total params: 10,953.

Total Trainable: 10,953.

Non-trainable params: 0.

## Appendix G. Mixed models for LSTM comparison and distance model comparison

To obtain a single distribution in the mixed model similar to the ANN model, it is required to produce a joint model across all DRs. To enable predictions, the device identity was removed as a random effect to allow for new observations from the test set that, by design, did not include any of the train set devices. This also includes the models with the log-distance terms for comparability with the perfect estimator (see [Table 17](#)).

**Table 14**  
Specification for MLP model.

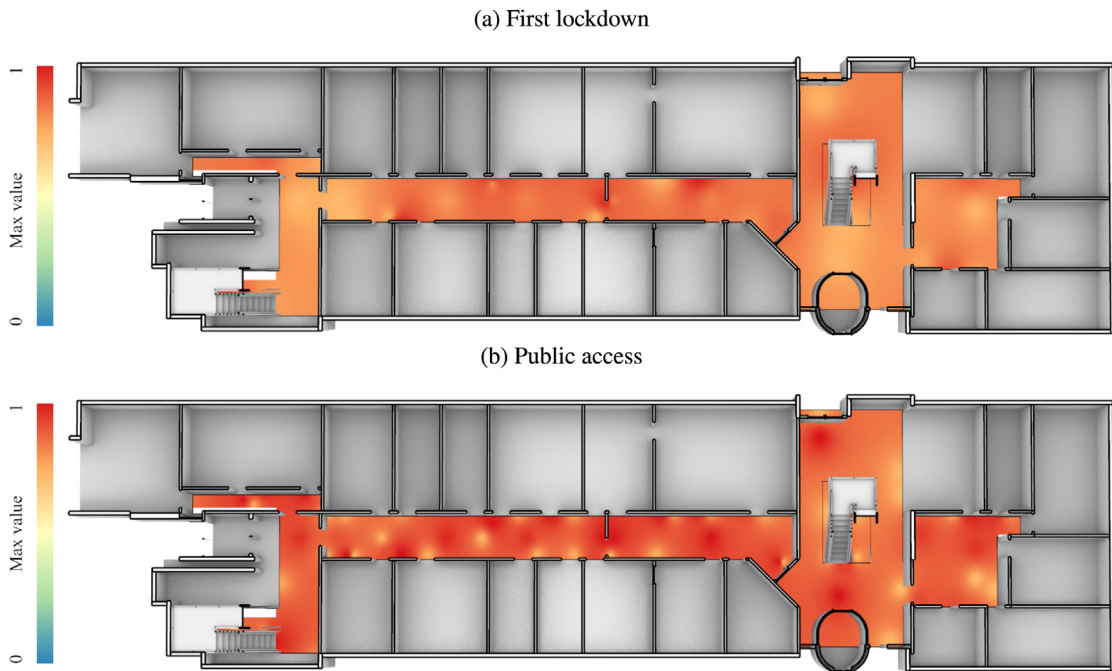
	Model	
Input shape:	(56)	
Input structure:	(16 hops with 3 features, 8 parameters )	
Layer	Output shape	# Parameters
Dense	(None, 350)	19950
Dense	(None, 50)	17550
Dense	(None, 3)	153

Total params: 37,653.  
Total Trainable: 37,653.  
Non-trainable params: 0.

**Table 15**  
Specification for GRU model.

	Model	
Input shape:	(19,3)	
Input structure:	(16 hops with 3 features, 8 parameters + 1 padding value arranged in 3 features)	
Layer	Output shape	# Parameters
LSTM	(None, 50)	8250
Dense	(None, 3)	153

Total params: 8,403.  
Total Trainable: 8,403.  
Non-trainable params: 0.



**Fig. 24.** Spatial visualisations in sensor data comparing maximal variation according to Max Value during the first lockdown and the restored public access. The yellow represents the average maximal variation across the whole time. Increased variation is indicated in red and decreased variation is indicated in blue.

### Appendix H. Lockdown difference for maximal values

This visualisation (see Fig. 24) follows Fig. 19 in style and calculation. Here, instead of averaging the average, we averaged the maximum value for each sensor. A single maximum value is determined by the sensor over a 5 min interval.

**Table 16**  
Specification for Transformer model.

Model				
Input shape:	(19,3)			
Input structure:	head_size=16,(16 hops with 3 features, 8 parameters + 1 padding value arranged in 3 features)			
Parameters:	head_size=16, num_heads=2, ff_dim=4, num_transformer_blocks=4, mlp_units= [128], mlp_dropout=0.4, dropout=0.25			
Layer	Label	Output shape	Connected to	# Parameters
Input	input <sub>1</sub>	(None, 19, 3)	0	
<i>Transformer Block 1</i>				
Normalisation	norm <sub>1</sub>	(None, 19, 3)	6	input <sub>1</sub>
Multi-Head Attention	mha <sub>1</sub>	(None, 19, 3)	483	norm <sub>1</sub> ; norm <sub>1</sub>
Dropout	drop <sub>1</sub>	(None, 19, 3)	0	mha <sub>1</sub>
Lambda	add <sub>1</sub>	(None, 19, 3)	0	drop <sub>1</sub> ; input <sub>1</sub>
Normalisation	norm <sub>2</sub>	(None, 19, 3)	6	add <sub>1</sub>
Conv1D	conv <sub>1</sub>	(None, 19, 4)	16	norm <sub>2</sub>
Dropout	drop <sub>2</sub>	(None, 19, 4)	0	conv <sub>1</sub>
Conv1D	conv <sub>2</sub>	(None, 19, 3)	15	drop <sub>2</sub>
Lambda	add <sub>2</sub>	(None, 19, 3)	0	conv <sub>2</sub> ; add <sub>1</sub>
<i>Transformer Block 2</i>				
Normalisation	norm <sub>3</sub>	(None, 19, 3)	6	add <sub>2</sub>
Multi-Head Attention	mha <sub>2</sub>	(None, 19, 3)	483	norm <sub>3</sub> ; norm <sub>3</sub>
Dropout	drop <sub>3</sub>	(None, 19, 3)	0	mha <sub>2</sub>
Lambda	add <sub>3</sub>	(None, 19, 3)	0	drop <sub>3</sub> ; add <sub>2</sub>
Normalisation	norm <sub>4</sub>	(None, 19, 3)	6	add <sub>3</sub>
Conv1D	conv <sub>3</sub>	(None, 19, 4)	16	norm <sub>4</sub>
Dropout	drop <sub>4</sub>	(None, 19, 4)	0	conv <sub>3</sub>
Conv1D	conv <sub>4</sub>	(None, 19, 3)	15	drop <sub>4</sub>
Lambda	add <sub>4</sub>	(None, 19, 3)	0	conv <sub>4</sub> ; add <sub>3</sub>
<i>Transformer Block 3</i>				
Normalisation	norm <sub>5</sub>	(None, 19, 3)	6	add <sub>4</sub>
Multi-Head Attention	mha <sub>3</sub>	(None, 19, 3)	483	norm <sub>5</sub> ; norm <sub>5</sub>
Dropout	drop <sub>5</sub>	(None, 19, 3)	0	mha <sub>3</sub>
Lambda	add <sub>5</sub>	(None, 19, 3)	0	drop <sub>5</sub> ; add <sub>4</sub>
Normalisation	norm <sub>6</sub>	(None, 19, 3)	6	add <sub>5</sub>
Conv1D	conv <sub>5</sub>	(None, 19, 4)	16	norm <sub>6</sub>
Dropout	drop <sub>6</sub>	(None, 19, 4)	0	conv <sub>5</sub>
Conv1D	conv <sub>6</sub>	(None, 19, 3)	15	drop <sub>6</sub>
Lambda	add <sub>6</sub>	(None, 19, 3)	0	conv <sub>6</sub> ;add <sub>5</sub>
<i>Transformer Block 4</i>				
Normalisation	norm <sub>7</sub>	(None, 19, 3)	6	add <sub>6</sub>
Multi-Head Attention	mha <sub>4</sub>	(None, 19, 3)	483	norm <sub>7</sub> ; norm <sub>7</sub>
Dropout	drop <sub>7</sub>	(None, 19, 3)	0	mha <sub>4</sub>
Lambda	add <sub>7</sub>	(None, 19, 3)	0	drop <sub>7</sub> ;add <sub>6</sub>
Normalisation	norm <sub>8</sub>	(None, 19, 3)	6	add <sub>7</sub>
Conv1D	conv <sub>7</sub>	(None, 19, 4)	16	norm <sub>8</sub>
Dropout	drop <sub>8</sub>	(None, 19, 4)	0	conv <sub>7</sub>
Conv1D	conv <sub>8</sub>	(None, 19, 3)	15	drop <sub>8</sub>
Lambda	add <sub>8</sub>	(None, 19, 3)	0	conv <sub>8</sub> ; add <sub>7</sub>
<i>Dense Block</i>				
Global Average Pooling 1D	gap <sub>1</sub>	(None, 19)	0	add <sub>8</sub>
Dense	dense <sub>1</sub>	(None, 128)	2560	gap <sub>1</sub>
Dropout	drop <sub>9</sub>	(None, 128)	0	dense <sub>1</sub>
Dense	dense <sub>2</sub>	(None, 3)	387	drop <sub>9</sub>

Total params: 5,051 Total Trainable: 5,051. Non-trainable params: 0.

Consequently, this figure captures the maximal activity that was recorded throughout the period. Both in the average and in the maximal value, there is a clear increase in activity after the lockdown.

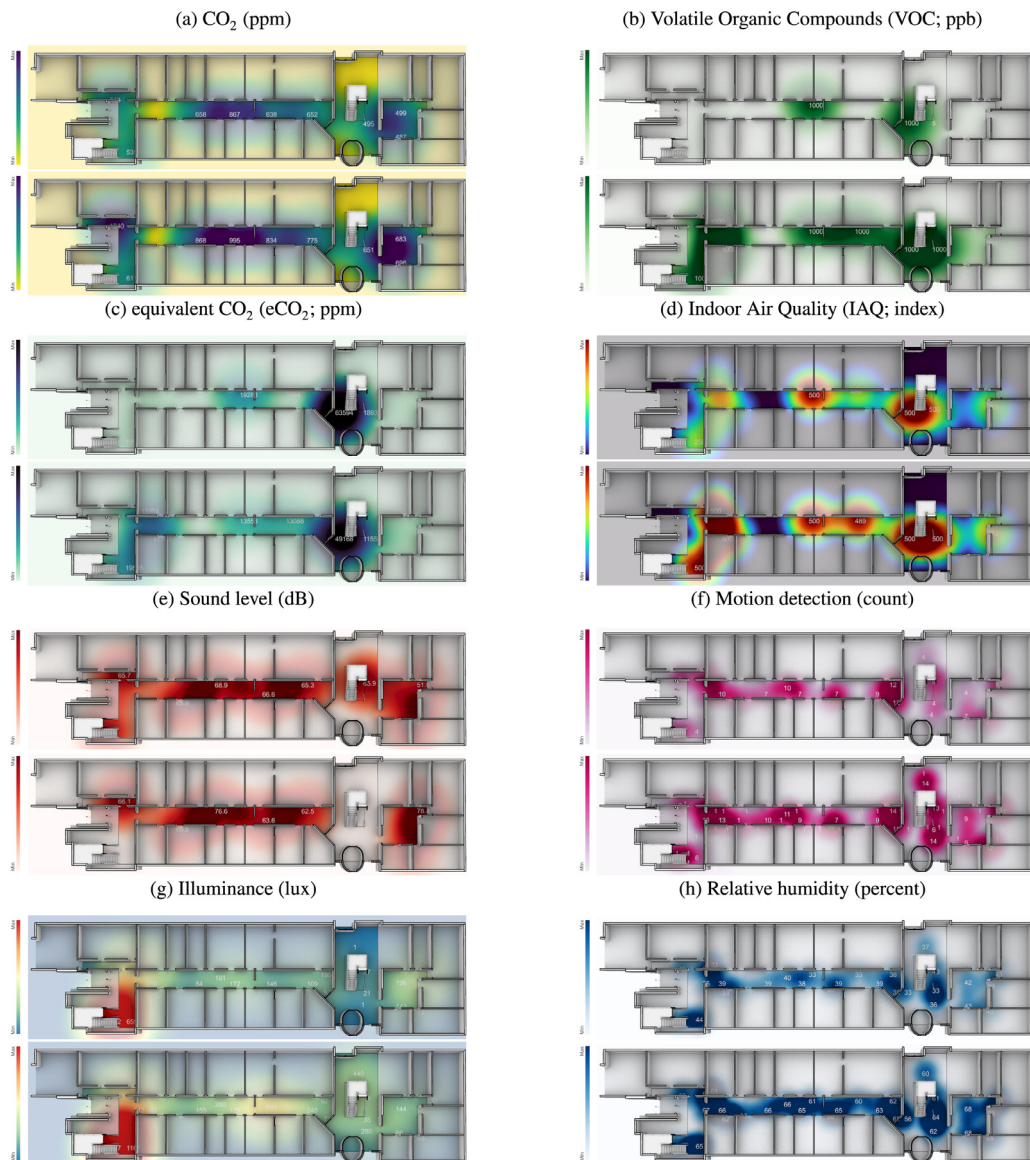
### Appendix I. Spatial sensor representation

See Fig. 25.

**Table 17**  
Mixed models and linear models for LSTM comparison and distance model comparison.

	Model											
	Mixed						Linear					
	Linear distance			Log distance			Linear distance			Log distance		
	ESP	SNR	RSSI	ESP	SNR	RSSI	ESP	SNR	RSSI	ESP	SNR	RSSI
Average signal strength at node (constant)	-76.398*	14.966*	-78.753*	-54.101*	18.199*	-57.080*	-79.210*	8.756*	-77.996*	-58.089*	11.514*	-57.329*
	(2.739)	(0.889)	(2.479)	(2.654)	(0.887)	(2.396)	(0.156)	(0.057)	(0.143)	(0.205)	(0.077)	(0.186)
Attenuation in m	-1.378*	-0.376*	-1.254*				-1.334*	-0.362*	-1.216*			
	(0.003)	(0.001)	(0.003)				(0.004)	(0.001)	(0.003)			
Attenuation in log(m)				-16.972*	-3.182*	-16.194*				-16.636*	-3.054*	-15.915*
				(0.053)	(0.020)	(0.048)				(0.056)	(0.021)	(0.051)
One floor between	-8.922*	-1.933*	-8.395*	-10.913*	-1.270*	-10.897*	-8.527*	-1.807*	-8.054*	-10.365*	-1.090*	-10.429*
	(0.023)	(0.009)	(0.021)	(0.043)	(0.016)	(0.039)	(0.025)	(0.009)	(0.023)	(0.046)	(0.017)	(0.042)
One wall between	-4.226*	-0.580*	-4.119*	-3.835*	0.536*	-4.226*	-4.009*	-0.510*	-3.933*	-3.422*	0.684*	-3.878*
	(0.025)	(0.009)	(0.023)	(0.055)	(0.021)	(0.050)	(0.027)	(0.010)	(0.025)	(0.059)	(0.022)	(0.053)
Colliding message	-0.248	-0.389*	-0.090	-0.381	-0.448*	-0.197	-1.113*	-0.657*	-0.840*	-1.186*	-0.695*	-0.896*
	(0.129)	(0.048)	(0.119)	(0.133)	(0.051)	(0.122)	(0.139)	(0.051)	(0.128)	(0.143)	(0.053)	(0.129)
CF 867.3 MHz	0.643*	0.013	0.626*	0.669*	0.019	0.651*	0.638*	0.014	0.621*	0.666*	0.021	0.647*
	(0.054)	(0.020)	(0.050)	(0.056)	(0.021)	(0.051)	(0.059)	(0.022)	(0.054)	(0.060)	(0.022)	(0.055)
CF 867.5 MHz	1.394*	-0.561*	1.613*	1.515*	-0.518*	1.719*	1.596*	-0.493*	1.786*	1.704*	-0.454*	1.879*
	(0.056)	(0.021)	(0.052)	(0.058)	(0.022)	(0.053)	(0.061)	(0.022)	(0.056)	(0.062)	(0.023)	(0.057)
CF 867.7 MHz	-0.134	0.012	-0.151	-0.137	0.008	-0.152	-0.151	0.008	-0.166	-0.152	0.005	-0.165
	(0.054)	(0.020)	(0.050)	(0.056)	(0.021)	(0.051)	(0.059)	(0.022)	(0.054)	(0.060)	(0.022)	(0.055)
CF 867.9 MHz	-0.248*	0.015	-0.264*	-0.218*	0.023	-0.236*	-0.256*	0.014	-0.271*	-0.226*	0.022	-0.243*
	(0.055)	(0.020)	(0.050)	(0.056)	(0.021)	(0.052)	(0.059)	(0.022)	(0.054)	(0.060)	(0.023)	(0.055)
CF 868.1 MHz	0.456*	-0.011	0.444*	0.480*	-0.003	0.465*	0.428*	-0.021	0.421*	0.454*	-0.013	0.443*
	(0.054)	(0.020)	(0.050)	(0.056)	(0.021)	(0.051)	(0.058)	(0.021)	(0.053)	(0.060)	(0.022)	(0.054)
CF 868.3 MHz	0.401*	-0.035	0.396*	0.420*	-0.031	0.414*	0.391*	-0.039	0.387*	0.408*	-0.036	0.404*
	(0.054)	(0.020)	(0.050)	(0.056)	(0.021)	(0.051)	(0.058)	(0.021)	(0.053)	(0.060)	(0.022)	(0.054)
CF 868.5 MHz	-0.373*	-0.067*	-0.362*	-0.339*	-0.056	-0.332*	-0.398*	-0.076*	-0.383*	-0.364*	-0.065	-0.353*
	(0.054)	(0.020)	(0.050)	(0.056)	(0.021)	(0.051)	(0.058)	(0.021)	(0.053)	(0.060)	(0.022)	(0.054)
ToA	0.199*	0.006	0.210*	0.151*	-0.014	0.170*	0.204*	0.089*	0.165*	0.172*	0.076*	0.140*
	(0.018)	(0.007)	(0.017)	(0.019)	(0.007)	(0.017)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
SF8	1.944	0.361	2.348*	1.765	0.189	2.267*	-0.013	-0.408	0.751	0.111	-0.484	0.934
	(0.640)	(0.239)	(0.592)	(0.662)	(0.252)	(0.606)	(0.691)	(0.253)	(0.632)	(0.707)	(0.264)	(0.642)
SF9	3.381*	-1.434*	5.219*	3.407*	-1.529*	5.315*	0.353	-1.814*	2.268	0.823	-1.759*	2.749*
	(0.793)	(0.296)	(0.733)	(0.820)	(0.312)	(0.750)	(0.852)	(0.312)	(0.780)	(0.872)	(0.325)	(0.792)
SF10	-4.888*	-3.447*	-2.123	-7.908*	-4.871*	-4.517*	-2.850	-4.319*	0.475	-6.206*	-5.834*	-2.225
	(1.125)	(0.419)	(1.039)	(1.163)	(0.442)	(1.064)	(1.207)	(0.442)	(1.105)	(1.235)	(0.461)	(1.122)
SF11	-0.495	-4.865*	3.905*	-4.725*	-6.871*	0.593	-1.760	-4.920*	2.618	-5.587*	-6.797*	-0.341
	(0.949)	(0.354)	(0.877)	(0.982)	(0.373)	(0.898)	(1.027)	(0.376)	(0.940)	(1.051)	(0.392)	(0.954)
SF12	-16.038*	-13.182*	-5.681*	-24.413*	-17.124*	-12.210*	-14.254*	-14.782*	-2.925	-22.957*	-18.811*	-9.764*
	(1.133)	(0.422)	(1.047)	(1.171)	(0.445)	(1.071)	(1.197)	(0.438)	(1.096)	(1.224)	(0.457)	(1.112)
Floor-attenuation interaction	0.204*	0.030*	0.203*				0.197*	0.028*	0.197*			
	(0.001)	(0.0003)	(0.001)				(0.001)	(0.0003)	(0.001)			
Floor-log(attenutation) interaction				2.150*	-0.139*	2.398*				2.067*	-0.167*	2.327*
				(0.014)	(0.005)	(0.013)				(0.015)	(0.006)	(0.014)
Wall-attenuation interaction	0.107*	0.006*	0.109*				0.097*	0.003*	0.101*			
	(0.001)	(0.0003)	(0.001)				(0.001)	(0.0003)	(0.001)			
Wall-log(attenutation) interaction				0.749*	-0.355*	0.955*				0.614*	-0.403*	0.842*
				(0.016)	(0.006)	(0.015)				(0.017)	(0.006)	(0.016)
Wall-floor interaction	0.389*	-0.091*	0.443*	-0.100*	-0.306*	0.005	0.278*	-0.129*	0.348*	-0.445*	-0.422*	-0.289*
	(0.009)	(0.003)	(0.008)	(0.020)	(0.008)	(0.018)	(0.009)	(0.003)	(0.009)	(0.021)	(0.008)	(0.019)
SF8-ToA interaction	-0.130*	-0.011	-0.136*	-0.105*	-0.0004	-0.115*	-0.134*	-0.048*	-0.117*	-0.118*	-0.041*	-0.105*
	(0.010)	(0.004)	(0.009)	(0.010)	(0.004)	(0.009)	(0.006)	(0.002)	(0.006)	(0.007)	(0.002)	(0.006)
SF9-ToA interaction	-0.173*	-0.008	-0.183*	-0.139*	0.006	-0.154*	-0.172*	-0.068*	-0.146*	-0.151*	-0.058*	-0.129*
	(0.014)	(0.005)	(0.013)	(0.014)	(0.005)	(0.013)	(0.005)	(0.002)	(0.004)	(0.005)	(0.002)	(0.004)
SF10-ToA interaction	-0.166*	-0.005	-0.176*	-0.118*	0.015	-0.135*	-0.181*	-0.075*	-0.149*	-0.145*	-0.060*	-0.120*
	(0.015)	(0.006)	(0.014)	(0.016)	(0.006)	(0.014)	(0.004)	(0.001)	(0.003)	(0.004)	(0.001)	(0.003)
SF11-ToA interaction	-0.189*	-0.005	-0.201*	-0.139*	0.016	-0.159*	-0.196*	-0.083*	-0.161*	-0.162*	-0.068*	-0.133*
	(0.017)	(0.006)	(0.015)	(0.017)	(0.007)	(0.016)	(0.002)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)
SF12-ToA interaction	-0.185*	-0.001	-0.199*	-0.133*	0.020	-0.156*	-0.192*	-0.081*	-0.159*	-0.156*	-0.065*	-0.129*
	(0.017)	(0.006)	(0.016)	(0.018)	(0.007)	(0.016)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Wall-floor-attenuation interaction	-0.009*	0.004*	-0.011*				-0.004*	0.006*	-0.007*			
	(0.0003)	(0.0001)	(0.0003)				(0.0003)	(0.0001)	(0.0003)			
Wall-floor-log(attenutation) interaction				0.103*	0.129*	0.050*				0.210*	0.165*	0.142*
				(0.006)	(0.002)	(0.006)				(0.006)	(0.002)	(0.006)
Random effects												
# Device Types	5	5	5	5	5	5						
Device Type Std. Dev.	5.51	1.72	4.97	5.25	1.68	4.72						
Adjusted R <sup>2</sup>							0.551	0.590	0.506	0.530	0.553	0.491
Residual Std. Error (df = 629352)							11.686	4.277	10.694	11.954	4.464	10.857
Observations	629,379	629,379	629,379	629,379	629,379	629,379	629,379	629,379	629,379	629,379	629,379	629,379

Note: \* p<0.001; \*\* p<[0.01]; \*\*\* p<[0.001].



**Fig. 25.** Spatial visualisations of several sensor types with IDW interpolation. The top row of each sensor type shows the lockdown in Spring 2020 and the bottom row shows the reopening to public access in Summer 2020. Most of the variance appears to be driven by human presence. Exceptions can be found in illuminance where the side staircase has direct exposure to sunlight which overrides any impact of human presence.

## References

- [1] U. Raza, P. Kulkarni, M. Sooriyabandara, Low power wide area networks: An overview, *IEEE Commun. Surv. Tuts.* 19 (2) (2017) 855–873.
- [2] M.C. Vuran, A. Salam, R. Wong, S. Irmak, Internet of underground things in precision agriculture: Architecture and technology aspects, *Ad Hoc Netw.* 81 (2018) 160–173.
- [3] M. Loriot, A. Aljer, I. Shahrou, Analysis of the use of LoRaWAN technology in a large-scale smart city demonstrator, in: *Sens. Netw. Smart and Emerg. Technol. (SENSET)*, 2017, IEEE, New York, 2017, pp. 1–4.
- [4] M. Saravanan, A. Das, V. Iyer, Smart water grid management using LPWAN IoT technology, in: *2017 Global Internet of Things Summit, GIoTS*, IEEE, 2017, pp. 1–6.
- [5] L. Li, J. Ren, Q. Zhu, On the application of LoRa LPWAN technology in sailing monitoring system, in: *Wireless on-Demand Netw. Syst. and Services (WONS)*, 2017 13th Annu. Conf. on, IEEE, Jackson Hole, Wyoming, 2017, pp. 77–80.
- [6] P.J. Radcliffe, K.G. Chavez, P. Beckett, J. Spangaro, C. Jakob, Usability of LoRaWAN technology in a central business district, in: *85th IEEE Veh. Technol. Conf., VTC*, IEEE, New York, 2017, pp. 1–5.
- [7] N. Varsier, J. Schwoerer, Capacity limits of LoRaWAN technology for smart metering applications, in: *2017 IEEE Int. Conf. Commun., ICC*, IEEE, 2017, pp. 1–6.

- [8] M.T. Lazarescu, Design of a WSN platform for long-term environmental monitoring for IoT applications, *IEEE Trans. Emerg. Sel. Top. Circuits Syst.* 3 (1) (2013) 45–54.
- [9] K. Mekki, E. Bajic, F. Chaxel, F. Meyer, A comparative study of LPWAN technologies for large-scale IoT deployment, *ICT Express* 5 (1) (2019).
- [10] N. Sornin, LoRaWAN 1.1 Specification, Technical Report, LoRa Alliance, 2017, p. 101, URL <https://lora-alliance.org/>.
- [11] K. Mekki, E. Bajic, F. Chaxel, F. Meyer, Overview of cellular LPWAN technologies for IoT deployment: Sigfox, LoRaWAN, and NB-IoT, in: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops, Percom Workshops, IEEE, 2018, pp. 197–202.
- [12] D. Lundell, A. Hedberg, C. Nyberg, E. Fitzgerald, A routing protocol for LoRa mesh networks, in: 2018 IEEE 19th International Symposium on "a World of Wireless, Mobile and Multimedia Networks", WoWMoM, IEEE, 2018, pp. 14–19.
- [13] N.C. Almeida, R.P. Rolle, E.P. Godoy, P. Ferrari, E. Sisinni, Proposal of a hybrid LoRa Mesh/LoRaWAN network, in: 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Ieee, 2020, pp. 702–707.
- [14] K.-C. Lai, B.-H. Ku, C.-Y. Wen, Using cooperative PIR sensing for human indoor localization, in: 27th Wireless and Optical Commun. Conf., WOCC, IEEE, New York, 2018, pp. 1–5.
- [15] M. Moussaïd, V.R. Schinazi, M. Kapadia, T. Thrash, Virtual sensing and virtual reality: How new technologies can boost research on crowd dynamics, *Front. Robot. AI* 5 (2018) <http://dx.doi.org/10.3389/frobt.2018.00082>.
- [16] P. Kiefer, I. Giannopoulos, M. Raubal, Where am I? Investigating map matching during self-localization with mobile eye tracking in an Urban environment, *Trans. GIS* 18 (5) (2014) 660–686.
- [17] W. Wang, J. Chen, T. Hong, Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings, *Autom. Constr.* 94 (2018) 233–243.
- [18] M. Batty, Digital twins, *Environ. Plan. B* 45 (5) (2018) 817–820.
- [19] N. Mohammadi, J.E. Taylor, Smart city digital twins, in: IEEE Symp. Series Int. Computational Intelligence, SSCI, IEEE, Honolulu, Hawaii, 2017, pp. 1–5.
- [20] M. Bell, Service-Oriented Modeling, John Wiley & Sons, Inc, Wiley Online Library, 2008.
- [21] M. Grieves, J. Vickers, Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems, in: *Transdisciplinary Perspectives Complex Systems*, Springer, Heidelberg, 2017, pp. 85–113.
- [22] F. Tao, H. Zhang, A. Liu, A. Nee, Digital twin in industry: State-of-the-art, *IEEE Trans. Ind. Inf.* 15 (2018) 2405–2415.
- [23] J. Jiong, M. Tobia, R. Lawther, D. Branchaud, T. Bednarz, Double vision: Digital twin applications within extended reality, in: *ACM SIGGRAPH 2020 Appy Hour*, 2020, pp. 1–2.
- [24] D.K. Hamilton, D.H. Watkins, Evidence-Based Design for Multiple Building Types, John Wiley & Sons, 2009.
- [25] H.S. Alavi, D. Lalanne, J. Nembrini, E. Churchill, D. Kirk, W. Moncur, Future of human-building interaction, in: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 3408–3414.
- [26] J. Grübel, M. Gath-Morad, L. Aguilar, T. Thrash, R.W. Sumner, C. Hölscher, V.R. Schinazi, Fused twins: A cognitive approach to augmented reality media architecture, in: *MAB '20: Proceedings of the 5th Media Architecture Biennale Conference*, 2021, pp. 1–8, <http://dx.doi.org/10.1145/3469410.3469435>.
- [27] J. Grübel, The Hitchhiker's guide to fused twins—a conceptualization to access digital twins in situ in smart cities, 2022, arXiv preprint [arXiv:2202.07104](https://arxiv.org/abs/2202.07104).
- [28] M. Alenezi, K.K. Chai, Y. Chen, S. Jimaa, Ultra-dense LoRaWAN: Reviews and challenges, *IET Commun.* 14 (9) (2020) 1361–1371.
- [29] T. Polonelli, D. Brunelli, A. Marzocchi, L. Benini, Slotted aloha on lorawan-design, analysis, and deployment, *Sensors* 19 (4) (2019) 838.
- [30] J.C. Liando, A. Gamage, A.W. Tengourtius, M. Li, Known and unknown facts of LoRa: Experiences from a large-scale measurement study, *ACM Trans. Sens. Netw.* 15 (2) (2019) 1–35.
- [31] R. Yasmin, K. Mikhaylov, A. Pouttu, LoRaWAN for smart campus: Deployment and long-term operation analysis, *Sensors* 20 (23) (2020) 6721.
- [32] J. Grübel, T. Thrash, D. Héhal, R.W. Sumner, C. Hölscher, V.R. Schinazi, The feasibility of dense indoor LoRaWAN towards passively sensing human presence, in: 2021 IEEE International Conference on Pervasive Computing and Communications, PerCom, 2021, pp. 1–11, <http://dx.doi.org/10.1109/PERCOM50583.2021.9439137>, **Best Paper Nominee**.
- [33] N. Abramson, The ALOHA System: Another alternative for computer communications, in: *Proc. Fall Joint Comput. Conf., AFIPS*, ACM, New York, 1970, pp. 281–285, <http://dx.doi.org/10.1145/1478462.1478502>.
- [34] A. Berni, W. Gregg, On the utility of chirp modulation for digital signaling, *IEEE Trans. Commun.* 21 (6) (1973) 748–751.
- [35] Semtech Corporation, AN1200.22 LoRa Modulation Basics, Technical Report, Semtech Corporation, Camarillo, CA, 2015, p. 26.
- [36] IEEE Comput Soc, IEEE Standard 802.15.4a-2007, Technical Report, IEEE, New York, 2007, <http://dx.doi.org/10.1109/IEEESTD.2007.4299496>.
- [37] A. Rahmadhani, F. Kuipers, When LoRaWAN frames collide, in: *Proceedings of the 12th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, 2018, pp. 89–97.
- [38] A. Mahmood, E. Sisinni, L. Guntupalli, R. Rondon, S.A. Hassan, M. Gidlund, Scalability analysis of a LoRa network under imperfect orthogonality, *IEEE Trans. Ind. Inf.* 15 (3) (2018) 1425–1436.
- [39] ETSI, ETSI EN 300 220-2 V3.2.1, Technical Report, ETSI, 2018, p. 33.
- [40] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, T. Watteyne, Understanding the limits of LoRaWAN, *IEEE Commun. Mag.* 55 (9) (2017) 34–40.
- [41] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner, Z. Li, Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements, *IEEE Netw.* 32 (2) (2018) 8–15.
- [42] A.J. Wixted, P. Kinnaird, H. Larjani, A. Tait, A. Ahmadi, N. Strachan, Evaluation of LoRa and LoRaWAN for wireless sensor networks, in: *Proc. IEEE Sensors*, IEEE, Orlando, FL, 2016, pp. 1–3.
- [43] M. Centenaro, L. Vangelista, A. Zanella, M. Zorzi, Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios, *IEEE Trans. Wireless Commun.* 23 (5) (2016) 60–67.
- [44] M. Paredes, S. Bertoldo, L. Carosso, C. Lucianaz, E. Marchetta, M. Allegretti, P. Savi, Propagation measurements for a LoRa network in an urban environment, *J. Electromagn. Waves Appl.* 33 (15) (2019) 2022–2036.
- [45] J. Petäjäjärvi, K. Mikhaylov, R. Yasmin, M. Hämäläinen, J. Iinatti, Evaluation of LoRa LPWAN technology for indoor remote health and wellbeing monitoring, *Int. J. Wirel. Inf. Netw.* 24 (2) (2017) 153–165.
- [46] M. Sauter, From GSM to LTE-Advanced: An Introduction to Mobile Networks and Mobile Broadband, John Wiley & Sons, 2014.
- [47] A. Rahmadhani, F. Kuipers, Understanding collisions in a LoRaWAN, 2017, SURF Wiki.
- [48] A. Abdelghany, B. Uguen, C. Moy, D. Lemur, On superior reliability of effective signal power versus RSSI in LoRaWAN, in: 28th International Conference on Telecommunications, ICT 2021, 2021, pp. 1–5.
- [49] T.S. Rappaport, et al., *Wireless Communications: Principles and Practice*, Vol. 2, Prentice Hall PTR New Jersey, 1996.
- [50] A. Aragon-Zavala, J.L. Cuevas-Ruiz, J.A. Delgado-Penín, High-Altitude Platforms for Wireless Communications, John Wiley & Sons, 2008.
- [51] R. He, Z. Zhong, B. Ai, J. Ding, K. Guan, Analysis of the relation between Fresnel zone and path loss exponent based on two-ray model, *IEEE Antennas Wirel. Propag. Lett.* 11 (2012) 208–211.
- [52] A.G. Longley, Prediction of Tropospheric Radio Transmission Loss over Irregular Terrain: A Computer Method-1968, Vol. 67, Institute for Telecommunication Sciences, 1968.

- [53] G.A. Hufford, A.G. Longley, W.A. Kissick, et al., A Guide to the Use of the ITS Irregular Terrain Model in the Area Prediction Mode, US Department of Commerce, National Telecommunications and Information, 1982.
- [54] W. Xu, J.Y. Kim, W. Huang, S.S. Kanhere, S.K. Jha, W. Hu, Measurement, characterization, and modeling of lora technology in multifloor buildings, *IEEE Internet Things J.* 7 (1) (2019) 298–310.
- [55] J. Keenan, A. Motley, Radio coverage in buildings, *Br. Telecom Technol. J.* 8 (1) (1990) 19–24.
- [56] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, *J. Stat. Softw.* 67 (1) (2015) 1–48, <http://dx.doi.org/10.18637/jss.v067.i01>.
- [57] V.A. Brown, An introduction to linear mixed-effects modeling in R, *Adv. Methods Pract. Psychol. Sci.* 4 (1) (2021) 2515245920960351.
- [58] N. Faruk, S.I. Popoola, N.T. Surajudeen-Bakinde, A.A. Oloyede, A. Abdulkarim, L.A. Olawoyin, M. Ali, C.T. Calafate, A.A. Atayero, Path loss predictions in the VHF and UHF bands within urban environments: experimental investigation of empirical, heuristics and geospatial models, *IEEE Access* 7 (2019) 77293–77307.
- [59] E. Ostlin, H.-J. Zepernick, H. Suzuki, Macrocell path-loss prediction using artificial neural networks, *IEEE Trans. Veh. Technol.* 59 (6) (2010) 2735–2747.
- [60] S.I. Popoola, A. Jefia, A.A. Atayero, O. Kingsley, N. Faruk, O.F. Oseni, R.O. Abolade, Determination of neural network parameters for path loss prediction in very high frequency wireless channel, *IEEE Access* 7 (2019) 150462–150483.
- [61] L. Wu, D. He, B. Ai, J. Wang, H. Qi, K. Guan, Z. Zhong, Artificial neural network based path loss prediction for wireless communication network, *IEEE Access* 8 (2020) 199523–199538.
- [62] H.-S. Jo, C. Park, E. Lee, H.K. Choi, J. Park, Path loss prediction based on machine learning techniques: Principal component analysis, artificial neural network, and Gaussian process, *Sensors* 20 (7) (2020) 1927.
- [63] C. Nguyen, A.A. Cheema, A deep neural network-based multi-frequency path loss prediction model from 0.8 GHz to 70 GHz, *Sensors* 21 (15) (2021) 5100.
- [64] I. Popescu, I. Nafomita, P. Constantinou, A. Kanatas, N. Moraitis, Neural networks applications for the prediction of propagation path loss in urban environments, in: *IEEE VTS 53rd Vehicular Technology Conference*, Spring 2001. Proceedings, Vol. 1, Cat. No. 01CH37202, IEEE, 2001, pp. 387–391.
- [65] P. Neumann, J. Montavont, T. Noël, Indoor deployment of low-power wide area networks (LPWAN): A LoRaWAN case study, in: *12th IEEE Int. Conf. Wireless and Mobile Comput., Netw. and Commun., WiMob*, IEEE, New York, 2016, pp. 1–8.
- [66] M.M. Erbaty, G. Schiele, G. Batke, Analysis of LoRaWAN technology in an outdoor and an indoor scenario in Duisburg-Germany, in: *2018 3rd Int. Conf. Comput. and Commun. Syst., ICCCS*, IEEE, New York, 2018, pp. 273–277.
- [67] L. Gregora, L. Vojtech, M. Neruda, Indoor signal propagation of LoRa technology, in: *17th Int. Conf. Mechatronics-Mechatronika*, IEEE, New York, 2016, pp. 1–4.
- [68] N. Vatcharatiensakul, P. Tuwanut, C. Pornavalai, Experimental performance evaluation of LoRaWAN: A case study in Bangkok, in: *14th Int. Joint Conf. Comput. Science and Softw. Eng., JCSSE*, IEEE, New York, 2017, pp. 1–4.
- [69] R. Muppala, A. Navnit, S. Poondla, A.M. Hussain, Investigation of indoor LoRaWAN signal propagation for real-world applications, in: *2021 6th International Conference for Convergence in Technology, I2CT*, IEEE, 2021, pp. 1–5.
- [70] A.M. Yousuf, E.M. Rochester, M. Ghaderi, A low-cost LoRaWAN testbed for IoT: Implementation and measurements, in: *2018 IEEE 4th World Forum on Internet of Things, WF-IoT*, IEEE, 2018, pp. 361–366.
- [71] J. Haxhibeqiri, A. Karaagac, F. Van den Abeele, W. Joseph, I. Moerman, J. Hoebeke, LoRa indoor coverage and performance in an industrial environment: Case study, in: *22nd IEEE Int. Conf. Emerg. Technol. Factory Autom., ETFA*, IEEE, New York, 2017, pp. 1–8.
- [72] E.D. Ayele, C. Hakkenberg, J.P. Meijers, K. Zhang, N. Meratnia, P.J. Havinga, Performance analysis of LoRa radio for an indoor IoT applications, in: *2017 Int. Conf. Internet of Things for the Global Community, IoTGC*, IEEE, New York, 2017, pp. 1–8.
- [73] R. El Chall, S. Lahoud, M. El Helou, Lorawan network: Radio propagation models and performance evaluation in various environments in Lebanon, *IEEE Internet Things J.* 6 (2) (2019) 2366–2378.
- [74] E. Harinda, A.J. Wixted, A.-U.-H. Qureshi, H. Larijani, R.M. Gibson, Performance of a live multi-gateway LoRaWAN and interference measurement across indoor and outdoor localities, *Computers* 11 (2) (2022) 25.
- [75] W. Li, G. Shen, J. Zhang, An indoor environmental monitoring system for large buildings based on LoRaWAN, in: *Proceedings of the Conference on Research in Adaptive and Convergent Systems*, 2019, pp. 34–38.
- [76] M.I. Muzammir, H.Z. Abidin, S.A.C. Abdullah, F.H.K. Zaman, Performance analysis of LoRaWAN for indoor application, in: *9th IEEE Symp. Comput. Appl. & Ind. Electron., ISCAIE*, IEEE, New York, 2019, pp. 156–159.
- [77] S. Ruepp, A.C. Mateo, K.M. Malarski, J. Thrane, M.N. Petersen, Internet of things connectivity in deep-indoor environments, in: *2018 9th International Conference on the Network of the Future, NOF*, IEEE, 2018, pp. 96–100.
- [78] T. Wendt, F. Volk, E. Mackensen, A benchmark survey of long range (LoRa) spread-spectrum-communication at 2.45 GHz for safety applications, in: *16th IEEE Annu. Wireless and Microw. Technol. Conf., WAMICON*, IEEE, New York, 2015, pp. 1–4.
- [79] E. Borgia, The internet of things vision: Key features, applications and open issues, *Comput. Commun.* 54 (2014) 1–31.
- [80] C. Jiang, M.K. Masood, Y.C. Soh, H. Li, Indoor occupancy estimation from carbon dioxide concentration, *Energy Build.* 131 (2016) 132–141.
- [81] Z. Yang, B. Becerik-Gerber, Modeling personalized occupancy profiles for representing long term patterns by using ambient context, *Build. Environ.* 78 (2014) 23–35, <http://dx.doi.org/10.1016/j.buildenv.2014.04.003>.
- [82] P. Liu, S. Nguang, A. Partridge, Occupancy inference using pyroelectric infrared sensors through hidden Markov models, *IEEE Sens. J.* 16 (4) (2016) 1062–1068, <http://dx.doi.org/10.1109/JSEN.2015.2496154>.
- [83] Orbiwise, OrbiLINK, Technical Report, Orbiwise, Geneva, 2018, p. 1.
- [84] J. Grübel, R. Weibel, M.H. Jiang, C. Hölscher, D.A. Hackman, V.R. Schinazi, EVE: A framework for experiments in virtual environments, in: *Spatial Cogn. X*, Springer, Heidelberg, 2016, pp. 159–176.
- [85] L. Richardson, M. Amundsen, M. Amundsen, S. Ruby, *RESTful Web APIs: Services for a Changing World*, O'Reilly Media, 2013.
- [86] J. Grübel, M. Gath-Morad, Fused Twin Base (Github), Zenodo, 2022, <http://dx.doi.org/10.5281/zenodo.6345416>.
- [87] J. Grübel, T. Thrash, D. Héhal, R.W. Sumner, C. Hölscher, V.R. Schinazi, LoRaWAN Dense Indoor Sensor Network (DISN) Transmission Meta Data, Zenodo, 2021, <http://dx.doi.org/10.5281/zenodo.4476317>, The accompanying research is presented at IEEE International Conference on Pervasive Computing and Communications 2021 (PerCom'21). The research that produced this data set is funded by ETH Zürich under the grant ETH-15 16-2. We thank Michal Gath-Morad for the BIM used for distance computations.
- [88] J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, M. Pettissalo, On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology, in: *14th Int. Conf. ITS Telecommun., ITST*, IEEE, 2015, pp. 55–59.
- [89] S. Liu, B. Kailkhura, D. Loveland, Y. Han, Generative counterfactual introspection for explainable deep learning, in: *2019 IEEE Global Conference on Signal and Information Processing, GlobalSIP*, IEEE, 2019, pp. 1–5.
- [90] I.T. Jolliffe, A note on the use of principal components in regression, *J. R. Stat. Soc. Ser. C. Appl. Stat.* 31 (3) (1982) 300–303.
- [91] A. Agarwal, D. Shah, D. Shen, D. Song, On robustness of principal component regression, *Adv. Neural Inf. Process. Syst.* 32 (2019).
- [92] G. Chao, Y. Luo, W. Ding, Recent advances in supervised dimension reduction: A survey, *Mach. Learn. Knowl. Extr.* 1 (1) (2019) 341–358.

- [93] Y. Zhang, W. Pan, Principal component regression and linear mixed model in association analysis of structured samples: competitors or complements? *Genet. Epidemiol.* 39 (3) (2015) 149–155.
- [94] J.M. Sutter, J.H. Kalivas, P.M. Lang, Which principal components to utilize for principal component regression, *J. Chemometr.* 6 (4) (1992) 217–225.
- [95] S. Wold, A. Ruhe, H. Wold, W. Dunn, The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, *SIAM J. Sci. Stat. Comput.* 5 (3) (1984) 735–743.
- [96] M. Haenlein, A.M. Kaplan, A beginner's guide to partial least squares analysis, *Underst. Stat.* 3 (4) (2004) 283–297.
- [97] V.E. Vinzi, W.W. Chin, J. Eriksler, H. Wang, et al., *Handbook of Partial Least Squares*, Vol. 201, Springer, 2010.
- [98] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemometr. Intell. Lab. Syst.* 58 (2) (2001) 109–130.
- [99] S. Haykin, *Neural Network, A comprehensive foundation*, *Neural Netw.* 2 (2004) (2004) 41.
- [100] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Systems* 2 (4) (1989) 303–314.
- [101] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [102] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, *Nature* 323 (6088) (1986) 533–536.
- [103] K. Cho, B. van Merriënboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: Encoder–decoder approaches, in: *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 2014, pp. 103–111.
- [104] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014, arXiv preprint arXiv:1412.3555.
- [105] N. Gruber, A. Jockisch, Are GRU cells more specific and LSTM cells more sensitive in motive classification of text? *Front. Artif. Intell.* 3 (2020) 40.
- [106] D. Ciregan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, in: *2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE*, 2012, pp. 3642–3649.
- [107] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [108] D. Magrin, M. Centenaro, L. Vangelista, Performance evaluation of LoRa networks in a smart city scenario, in: *IEEE Int. Conf. Commun., ICC*, 2017, pp. 1–7, <http://dx.doi.org/10.1109/ICC.2017.7996384>.
- [109] F. Van den Abeele, J. Haxhibeqiri, I. Moerman, J. Hoebeke, Scalability analysis of large-scale LoRaWAN networks in ns-3, *IEEE Internet Things J.* 4 (6) (2017) 2186–2198.
- [110] M.S.A. Muthanna, P. Wang, M. Wei, A.A. Ateya, A. Muthanna, Toward an ultra-low latency and energy efficient LoRaWAN, in: *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, Springer, 2019, pp. 233–242.
- [111] N. Matni, J. Moraes, H. Oliveira, D. Rosário, E. Cerqueira, LoRaWAN gateway placement model for dynamic internet of things scenarios, *Sensors* 20 (15) (2020) 4336.
- [112] A. Hoeller, J. Sant'Ana, J. Markkula, K. Mikhaylov, R. Souza, H. Alves, Beyond 5G low-power wide-area networks: A LoRaWAN suitability study, in: *2020 2nd 6G Wireless Summit, 6G SUMMIT, IEEE*, 2020, pp. 1–5.
- [113] D. Shepard, A two-dimensional interpolation function for irregularly-spaced data, in: *Proceedings of the 1968 23rd ACM National Conference*, 1968, pp. 517–524.
- [114] A.M. Malkawi, R.S. Srinivasan, A new paradigm for human-building interaction: the use of CFD and augmented reality, *Autom. Constr.* 14 (1) (2005) 71–84.
- [115] H.S. Alavi, E.F. Churchill, M. Wiberg, D. Lalanne, P. Dalsgaard, A. Fatah gen Schieck, Y. Rogers, Introduction to human-building interaction (hbi) interfacing hci with architecture and urban design, *ACM Trans. Comput.-Hum. Interact.* 26 (2) (2019) 1–10.
- [116] M. Scaife, Y. Rogers, External cognition: how do graphical representations work? *Int. J. Hum. Comput. Stud.* 45 (2) (1996) 185–213.
- [117] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, et al., *Immersive analytics*, in: *2015 Big Data Visual Analytics, BDVA, IEEE*, Hobart, Australia, 2015, pp. 1–8.
- [118] M. Gath-Morad, L. Aguilar, R.C. Dalton, C. Hölscher, cogARCH: Simulating wayfinding by architecture in multilevel buildings, in: *11th Annu. Symp. Simulation for Architecture & Urban Design, SimAUD 2020, ACM*, New York, 2020, pp. 27–34.



**Jascha Grübel** is a Ph.D. student in Computer Science at ETH Zürich and a visiting fellow at Harvard University's Visual Computing Group. He is working at the intersection with Cognitive Science and City Science at the Chair of Cognitive Science at ETH Zürich. In his thesis, he is investigating human behaviour in public spaces with Digital Twins. In addition, his thesis is focusing on the visualisation of data in VR and AR such as through Fused Twins and the analysis of Big Data as presented in this article. Jascha is also maintaining a framework to run VR experiments for scientific purposes called EVE that is intended to be intertwined in the future with this work to enable augmented reality experiments. This summer, Jascha will start work as a senior researcher at the Center for Sustainable Future Mobility at ETH Zürich on a Digital Twin of the Swiss Mobility System.

**Dr. Tyler Thrash** is currently a Ph.D. student in Biology at Saint Louis University. Previously, he was a postdoctoral research in cognitive science and geography at ETH Zürich and the University of Zurich, respectively. His work has focused on spatial cognition and navigation from a largely ecological/Gibsonian perspective. With this approach, he attempts to explain higher-level cognition (e.g., biases in spatial memory) in terms of lower-level, perceptual processes (e.g., visual exposure to environmental structure). Dr. Thrash also develops simulations for interactive virtual environments and mathematical models for understanding human behaviour.



**Leonel Aguilar** is a Lecturer and Senior Researcher in Applied Artificial Intelligence at the Data Science, Systems and Services laboratory (DS3), ETH Zürich. He held postdoctoral appointments at the Cognitive Science (COG) and Computational Social Science (COSS) groups, ETH Zürich, and the Earthquake Research Institute of the University of Tokyo, Japan. Leonel obtained his Ph.D. at the Computational Science and High-Performance Computing Laboratory at the University of Tokyo. His research focuses on modelling, simulating, and analysing social phenomena and the development and deployment of AI/ML-based systems to support this. He has worked on quantifying human behaviour through experiments and has leveraged these insights in AI-driven software for applications such as evacuation planning and education.



**Michal Gath-Morad** is an SNSF postdoctoral fellow at the University of Cambridge and at University College London, as well as a lecturer at the Chair of Cognitive Science at ETH Zürich. Michal is an architect by training and a cognitive scientist with a Ph.D. from ETH Zürich. Michal's interdisciplinary research and teaching seeks to develop, critically evaluate, and creatively experiment with methods of evidence-based co-design in architectural design pedagogy and praxis to promote health and well-being by architecture.



**Didier Héral** is co-founder and Director of Strategic Accounts at OrbiWise, managing the major Accounts in EMEA and APAC. Didier was previously Director of Operational Development for the Silicon System Development division of ST-Ericsson. He has been in the wireless industry for 25 years occupying senior positions from advanced communications research to product development management in STMicroelectronics and ST-Ericsson, in the fields of UWB, WLAN and cellular standards. Didier received his Ph.D. in Electronics, Optics and Microelectronics applied to Radar-Meteorology from Université du Sud-Toulon-Var, France. He is contributing to the LoRa Alliance activities and promotes LPWA IoT as an accelerator to the circular economy and for reaching the SDGs of the United Nations agenda of 2030.



**Prof. Robert W. Sumner** is the Director of Research and Development at the Walt Disney Studios and an Adjunct Professor at ETH Zürich. Robert received a B.S. degree in computer science from the Georgia Institute of Technology and his M.S. and Ph.D. degrees from the Massachusetts Institute of Technology. At ETH, Prof. Sumner teaches a course called the Game Programming Laboratory in which students from ETH and the Zurich University of the Arts work in small teams to design and implement novel video games. In 2015, Prof. Sumner founded the ETH Game Technology Center, which explores the unique way game technology can advance ETH's mission in research, education, and outreach. Prof. Sumner was featured on BBC Click and Ars Technica for his work on Unfolding the 8-Bit Era as well as Reuters for his augmented reality research.



**Christoph Hölscher** is Full Professor of Cognitive Science in the D-GESS at ETH Zürich since 2013, with an emphasis on Applied Cognitive Science. Since 2016 Christoph is a Principal Investigator at the Singapore ETH Center (SEC) Future Cities Laboratory, heading a research group on 'Cognition, Perception and Behaviour in Urban Environments'. Christoph is the Program Director of Future Resilient Systems FRS at the SEC since 2019, leading the current FRS 2 phase (2020–2025). He holds a Ph.D. in Psychology from University of Freiburg, served as honorary senior research fellow at UCL, Bartlett School of Architecture, and as a visiting Professor at Northumbria University Newcastle. Christoph has several years of industry experience in Human-Computer Interaction and usability consulting.



**Assist. Prof. Victor R. Schinazi** is at the Psychology Department of Bond University and PI of the Early Detection of Health Risks and Prevention module in the Future Health Technologies programme in Singapore. Previously, Victor was a Senior Lecturer at the Chair of Cognitive Science and the GeoGazeLab at the ETH Zürich. He also served as co-PI for the Cognition, Perception and Behaviour in Urban Environments project at the Future Cities Laboratory in Singapore. Victor holds a Ph.D. from the Centre for Advanced Spatial Analysis (CASA, UCL). He also was a Postdoctoral Fellow at the Center for Cognitive Neuroscience (UPenn) and served as Chief Science Officer for Strategic Spatial Solutions, Inc. Victor has published high impact articles and book chapters in psychology, cognitive neuroscience, geography and computer science.