



## Data Article

# Validated dataset combining simulations and measurements for emission analysis of naturally ventilated dairy barns



Julian Hartje\*, Abu Zar Shafiullah

*Thünen Institute for Agricultural Technology, Bundesallee 47, 38116 Braunschweig, Germany*

## ARTICLE INFO

*Article history:*

Received 27 June 2025

Revised 15 August 2025

Accepted 25 August 2025

Available online 1 September 2025

Dataset link: [Application of digital twins to extend the emission data of naturally ventilated dairy cattle barns \(Original data\)](#)

*Keywords:*

Computational fluid dynamics

Livestock building

Wind speed

Wind direction

Large opening

Ammonia

## ABSTRACT

Quantifying emissions from naturally ventilated livestock buildings is challenging due to the large side wall openings. In addition, measurement campaigns are expensive and time consuming and are therefore limited to a few short measurement weeks during the year. However, emission factors or annual averages are extrapolated from these data sets. Simulations can complement this data set by extending it and thus broadening the basis for the extrapolation of emission factors or evaluation of the barn and management system. The dataset presented consists of solution data from computational fluid dynamics (CFD) simulations of naturally ventilated cattle barns and the corresponding simulation and geometry files. The simulations were validated using data sets from measurement campaigns in three naturally ventilated cattle barns in Germany. Together with weather data from the German Weather Service (DWD), weather situations that occurred outside the measurement weeks could be investigated. With the presented data set further investigations are possible. Together with the measured data, simulation techniques, data aggregation and the development of new numerical modelling approaches can be investigated in detail.

© 2025 The Author(s). Published by Elsevier Inc.  
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

---

\* Corresponding author.

E-mail address: [julian.hartje@thuenen.de](mailto:julian.hartje@thuenen.de) (J. Hartje).

## Specifications Table

Subject	Computer Sciences
Specific subject area	Simulation data for evaluating and monitoring emissions in naturally ventilated dairy cattle barns
Type of data	Tables (.xlsx) and simulation files (.sim, .stl), figures and plots (.png), R-script (.txt) Processed
Data collection	Simulated data were generated using computational fluid dynamics (CFD), based on two sets of boundary conditions derived from: (i) measured data from the EmiDaT project, and (ii) meteorological data provided by the German Weather Service. All simulations were conducted using StarCCM+ (version 2020.3.1, Siemens PLM)
Data source location	Thünen Institute for Agricultural Technology Bundesallee 47 38116 Braunschweig, Germany
Data accessibility	Repository name: zenodo Data identification number: <a href="https://doi.org/10.5281/zenodo.16881506">https://doi.org/10.5281/zenodo.16881506</a> . Direct URL to data: <a href="https://zenodo.org/records/16881506">https://zenodo.org/records/16881506</a> .
Related research article	–

## 1. Value of the Data

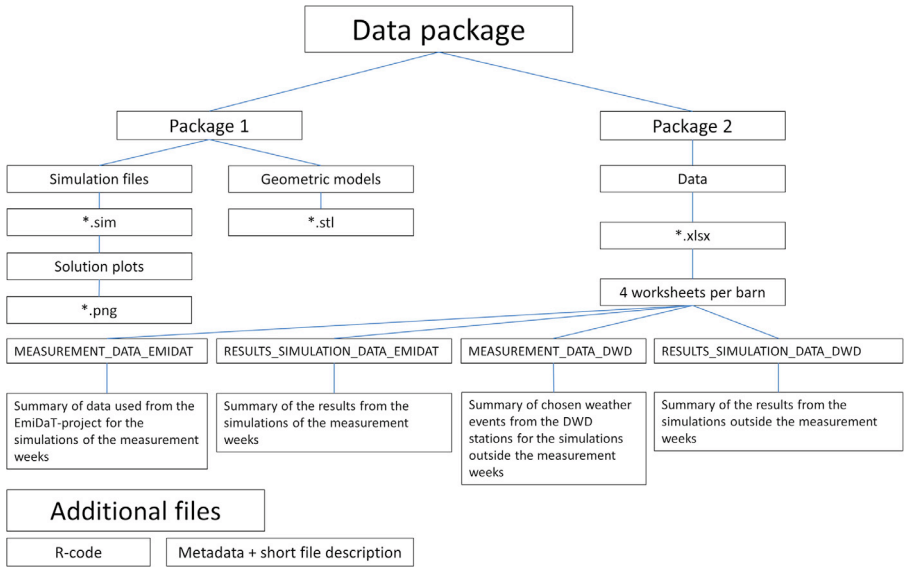
- **Advancing simulation studies in livestock buildings:** the integration of numerical models with measurement data from real-world barn facilities enables in-depth investigations that yield valuable insights for a wide range of simulation studies related to naturally ventilated livestock buildings. This includes areas such as advanced turbulence modelling techniques, representation of animal heat and mass transfer, extended statistical analysis of measurement data and its influence on simulation accuracy, development of models to represent vegetation elements such as hedges and trees, including their impact on flow and emissions and enhanced modelling of structural features like adjustable curtains
- **Extending measurement campaigns through CFD simulations:** The linkage of empirical data with CFD simulations offers unique opportunities to extend the utility of measurement campaigns through validated numerical modeling, enabling improved assessments of emission dynamics in naturally ventilated barns and exploration of environmental conditions beyond the limited measurement periods, such as varying wind or temperature scenarios
- **Cross-platform CFD comparisons:** The defined boundary conditions used in the simulations support reproducibility and facilitate application in other CFD solvers, promoting transparency and comparative analysis across research efforts

## 2. Background

The data were generated within the context of a study that combined empirical measurements with computational fluid dynamics (CFD) simulations to better represent the emission behavior of dairy barns under varying environmental conditions. The dataset includes boundary condition inputs, simulation outputs, and the complete simulation and geometry files used in the modeling process.

Measuring emissions in naturally ventilated dairy barns presents challenges due to large side-wall openings, which makes it harder to determine the airflow. Additionally, measurement campaigns are resource-intensive and typically limited to short durations, which may not capture the full range of meteorological conditions relevant to annual emission estimates [1–6]. This dataset was compiled in response to these limitations, with the aim of supplementing short-term measurements using simulation-based approaches.

It supports the associated research article by providing full access to the modeling framework, enabling reproducibility and offering a basis for further method development or comparison. The dataset is intended to assist other researchers in applying or adapting similar approaches in related contexts.



**Fig. 1.** File-Structure of the data package.

### 3. Data Description

The available data [7] consists of two main packages. Package 1 contains the simulations as sim-files, solution plots of these simulations as png-files stored in a zip-folder and the geometric models of the simulated barns as stl-files. Package 2 contains excel-files (Microsoft Excel (Version 365, Microsoft Corporation, Redmond, WA, USA)) with the results of the first simulations and the corresponding boundary conditions. As additional files, the used R-code (txt-file) and metadata with a short file description (pdf-file) have been added to the package.

An overview of the structure of the presented dataset is given in Fig. 1.

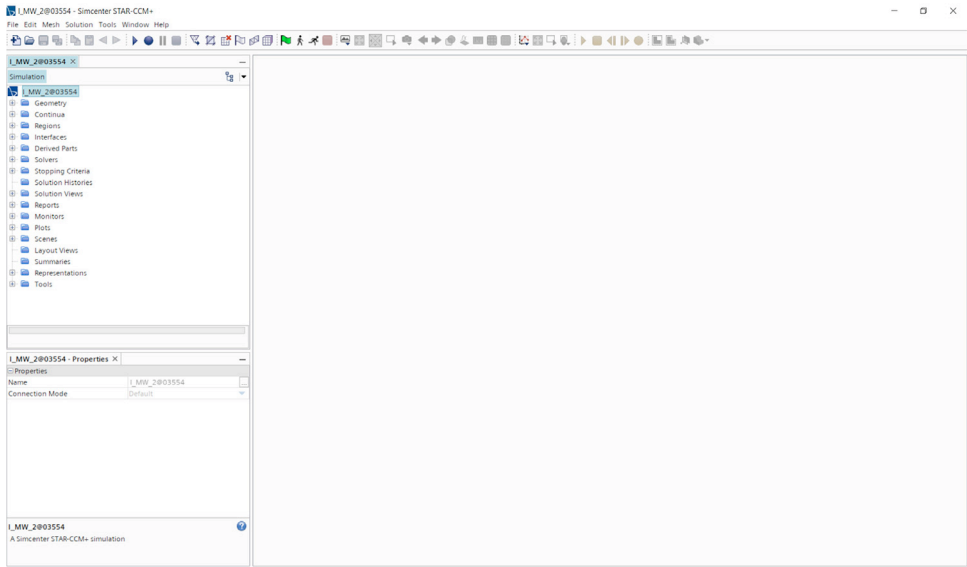
#### 3.1. Simulation files

Due to storing capacities in the repository, only one simulation per barn is saved there. The other cases can be reproduced by changing the boundary conditions as explained in section “EXPERIMENTAL DESIGN, MATERIALS AND METHODS”.

All simulations were carried out using StarCCM+ version 2020.3.1 by Siemens PLM. They are stored in the software-specific file format \*.sim and can be reused using the same software. StarCCM+ offers solutions for creating the geometric models, choosing solvers and numerical models for running the simulations as well as using a variety of tools for postprocessing the solution of the simulation. Fig. 2 shows a snapshot of the graphical user interface (GUI).

Table 1 gives an overview of the used models, boundary conditions, field functions and solution data and where they can be found within the software. Some parameters are stored in multiple places, i. e. the outside temperature. It serves as the reference temperature for all temperature related operations within the solver but also as the input temperature on the outer boundaries so that the incoming air has the correct temperature for the certain simulation. In these cases, multiple lines per variable indicate the multiple locations for them to be found.

The solution data from Table 1 is stored as plots in the file format \*.png on the repository for every barn simulation and every case in a zip-folder called solution\_plots.zip. The nomenclature



**Fig. 2.** Snapshot of the GUI from StarCCM+.

**Table 1**

Overview of location of parameters and solution data in StarCCM+.

Models/boundary condition	Location in StarCCM+
All numerical models	Continua -> Physics 1 -> Models
Outside temperature	Continua -> Physics 1 -> Reference Values -> Reference temperature Continua -> Physics 1 -> Initial conditions -> Static temperature Regions -> air -> Boundaries -> all <i>velocity inlet</i> and <i>pressure outlet</i> boundaries -> Physics Values -> Static temperature
Wind direction	Tools -> Parameters -> inlet_angle
Wind speed	Tools -> Parameters -> v_10
NH <sub>3</sub> background concentration	Continua -> Physics 1 -> Initial conditions -> Passive Scalar Regions -> air -> Boundaries -> all <i>velocity inlet</i> and <i>pressure outlet</i> boundaries -> Physics Values -> Passive scalar
NH <sub>3</sub> surface concentration	Tools -> Field functions -> 38.NH <sub>3</sub> _surface_concentration
<b>Solution data</b>	
Volume flow through the barn	Plots -> sum_V Monitor Plot
Inside temperature	Plots -> temp_points -> Temp_point_inside Monitor
NH <sub>3</sub> concentration inside the barn	Plots -> NH <sub>3</sub> _lines -> NH <sub>3</sub> _lines_inside_mean Monitor
Global normalized Residuals	Plots -> Residuals

for naming them follows the structure seen in [Table 2](#). The different name-sections are separated by an underscore.

Thus, for example, the volume flow through barn I for event 4 of the measurement weeks can be seen in the file I\_MW\_4\_V.png.

### 3.2. Geometric models

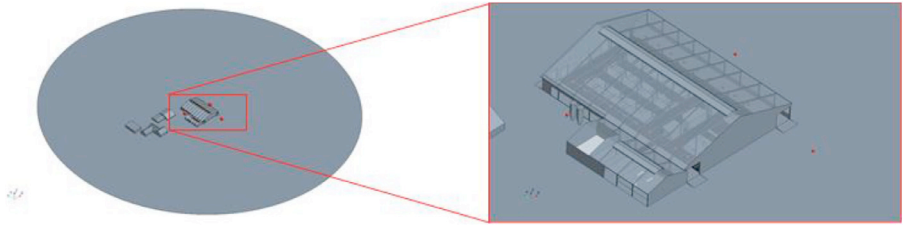
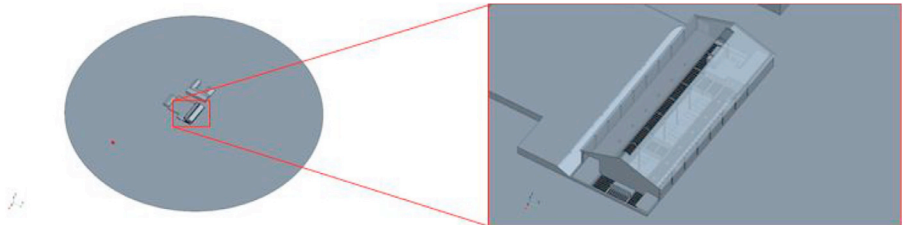
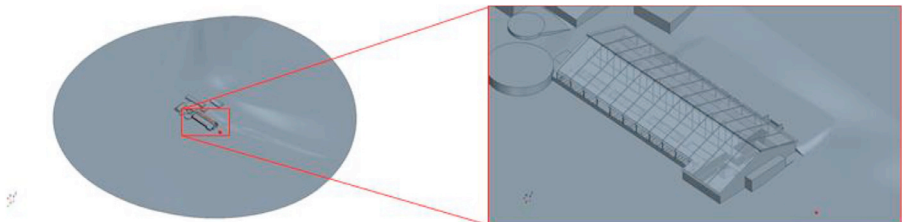
The geometries for each barn are stored as \*.stl files. They were created in the StarCCM+ CAD editor. The models were built using floor plans and own measurements on site. Not only

**Table 2**

Nomenclature for naming the solution plots of the simulations.

Barn identifier	Underlying dataset	Event number	Content
I	MW	1	Residuals
II	DWD	2	V
III		3	NH <sub>3</sub>
		4	T
		5	
		6	

MW = measuring week, DWD = German Weather Service, V = volume flow, NH<sub>3</sub> = NH<sub>3</sub>-concentration inside the barn, T = temperature inside the barn (only barn I).

**Fig. 3.** Geometric model of barn I.**Fig. 4.** Geometric model of barn II.**Fig. 5.** Geometric model of barn III.

the barns, but also the surrounding buildings and important features of the wider environment (mountain/valley) were taken into account. The model was constructed on a 1:1 scale. In order to make it easier to vary the wind direction at a later date, the entire simulation environment was constructed as a cylinder with the barn model on its base. Figs. 3–5 show the geometric models of the respective barns including surrounding buildings and a view of the simulation domain.

**Table 3**

Overview of column entries in the worksheet MEASUREMENT\_DATA\_EMIDAT.

Name of the column	Meaning of the content of the column
Measurement week	Indicates the measurement week in which the selected period is located
Period	Indicates the month of the year in order to obtain a classification in the season
Number of animals	Number of animals that were housed at this time
Averaging of the measurement data Start (UTC)	Start measured value for averaging over the selected period
Averaging of the measurement data Finish (UTC)	Final measured value for averaging over the selected period
Duration	Duration of the selected period in hours and minutes
number of measurement points	Number of available measured values within this period
wind direction in °	Averaged wind direction within this period and CoV of the data
windspeed in m·s <sup>-1</sup>	Averaged wind speed within this time period and CoV of the data
outside temperature in °C	Averaged ambient temperature within this period and CoV of the data
background concentration of NH <sub>3</sub> outside the barn in mg·m <sup>-3</sup> or ppm	Averaged background concentration within this period and CoV of the data
temperature inside the barn in °C	Averaged barn temperature within the period (only barn I) and CoV of the data
measured concentration of NH <sub>3</sub> inside the barn in mg·m <sup>-3</sup> or ppm	Averaged measured ammonia concentration within the period and CoV of the data
volume flow through the barn in m <sup>3</sup> ·h <sup>-1</sup>	Averaged measured air volume flow through the barn within the period and CoV of the data

A	B	C	D	E	F	G	H	I	J	K			L	M	N
Measurement week	Period	number of animals	Averaging of the measurement data		duration	number of measurement points	wind direction in °	windspeed in m·s <sup>-1</sup>	outside temperature in °C	background concentration of NH <sub>3</sub> outside the barn in mg·m <sup>-3</sup>	temperature inside the barn in °C	measured concentration of NH <sub>3</sub> inside the barn in mg·m <sup>-3</sup>	volume flow through the barn in m <sup>3</sup> ·h <sup>-1</sup>		
1	1 November 2017	134	22.11.17 10:39 PM	23.11.17 12:27 AM	2 h 48 min	14	124	2.2	8.2	0.2	8.0	0.1	980.482		
2	2 January 2018	138	15.1.18 10:23 AM	15.1.18 10:24 AM	1 h 00 min	19	139	2.2	8.9	0.2	9.8	0.4	799.097		
3	3 March 2018	127	18.3.18 4:36 AM	18.3.18 6:49 AM	4 h 13 min	22	90	6.7	-4.1	0.1	-5.1	0.3	781.832		
4	4 May 2018	124	15.5.18 1:23 PM	15.5.18 1:57 PM	2 h 36 min	16	67	1.4	18.8	0.2	23.7	0.5	695.767		
5	5 July 2018	127	7.7.18 10:43 AM	7.7.18 12:07 PM	1 h 24 min	8	281	1.9	20.5	0.3	20.5	0.5	751.465		
6	6 August 2018	128	29.8.18 3:53 PM	29.8.18 6:36 PM	2 h 43 min	14	296	1.5	24.0	0.4	25.3	0.7	860.751		
7															
8															
9															
10	1						0.0558	0.1485	0.0004	0.1308	0.0006	0.1483	0.1400		
11	2						0.0589	0.1462	0.0003	0.4915	0.0007	0.2122	0.3500		
12	3						0.0927	0.1474	0.0009	0.4848	0.0006	0.1475	0.1300		
13	4						0.1090	0.1843	0.0009	0.3398	0.0013	0.3964	0.2800		
14	5						0.0774	0.2482	0.0017	0.1404	0.0015	0.2196	0.3300		
15	6						0.0413	0.1762	0.0018	0.1114	0.0012	0.1239	0.3100		

**Fig. 6.** Snapshot of excel worksheet 1 for barn I.

3.3. Data

An individual excel-file was created for each of the three barns. Each contains four worksheets, which are described in more detail below.

MEASUREMENT\_DATA\_EMIDAT

This worksheet contains an overview of the data used for the simulation from the EmiDaT project. The worksheet has 14 columns for barn I and 13 columns for barns II and III, as the temperature inside these barns was not measured. A detailed description is given in Table 3.

Fig. 6 shows a snapshot of the presented data for barn I.

RESULTS\_SIMULATION\_DATA\_EMIDAT

The averaged values from the previous worksheet were used as boundary conditions and target values (measured ammonia concentration in the barn). The resulting flow through the barn at the boundaries between the barn building and the surroundings was then evaluated. This worksheet provides an overview of these results. It has four columns for barn I and three columns for barn II and III. The columns are described in Table 4.

MEASUREMENT\_DATA\_DWD

Similar to worksheet 1, the selected events outside the measurement weeks are listed here. The worksheet has nine columns. Table 5 gives an overview of the columns.

RESULTS\_SIMULATION\_DATA\_DWD

**Table 4**

Overview of column entries in the worksheet RESULTS\_SIMULATION\_DATA\_EMIDAT.

Name of the column	Meaning of the content of the column
Measurement week	Data from the respective measurement week from the previous worksheet
Simulated temperature inside the barn in °C	Results of the barn temperature inside the barn at the same measuring points as in the real measurement
Simulated concentration of NH <sub>3</sub> on the surfaces of the barn in ppm	Set NH <sub>3</sub> concentration on the surfaces to achieve the concentration value at the measuring points from the previous worksheet in the simulation
Simulated volume flow through the barn in m <sup>3</sup> .h <sup>-1</sup>	Air volume flow through the barn determined in the simulation

**Table 5**

Overview of column entries in the worksheet MEASUREMENT\_DATA\_DWD.

Name of the column	Meaning of the content of the column
Event	Numbering of the selected events
Period	Specifies the month of the year in order to obtain a classification in the season
ID of weather station	Indicates the ID of the weather station from the DWD network from which the data was recorded
Averaging of the measurement data Start (UTC)	Start measured value for averaging over the selected period
Averaging of the measurement data Finish (UTC)	Final measured value for averaging over the selected period
Duration	Duration of the selected time period in hours and minutes
wind direction in °	Averaged wind direction within this time period
windspeed in m.s <sup>-1</sup>	Averaged wind speed within this time period
outside temperature in °C	Averaged ambient temperature within this period

**Table 6**

Overview of column entries in the worksheet RESULTS\_SIMULATION\_DATA\_DWD.

Name of the column	Meaning of the content of the column
Event	Referencing the selected event from the DWD measurement data
simulated temperature inside the barn in °C	Results of the evaluation at the same measuring points as in the real measurement
simulated concentration of NH <sub>3</sub> on the surfaces of the barn in ppm	NH <sub>3</sub> -concentration set on the surfaces to achieve the concentration value at the measurement points from the previous worksheet in the simulation
simulated NH <sub>3</sub> -concentration inside the barn in ppm	Resulting NH <sub>3</sub> -concentration inside the barn in the simulation at the real measuring points due to the flow situation and the set NH <sub>3</sub> -concentration on the surfaces
simulated volume flow through the barn in m <sup>3</sup> .h <sup>-1</sup>	Air volume flow through the barn determined in the simulation

The results of the simulations with the weather data from the DWD as boundary conditions are listed here. This worksheet has five columns for barn I and four columns for barns II and III. [Table 6](#) lists the columns and their meaning.

### 3.4. R-Code

To investigate the relationships between environmental boundary conditions and simulated ammonia (NH<sub>3</sub>) surface concentrations, structured datasets were created for each barn, containing variables such as temperature (T), wind speed (V), wind direction (D), and the modeled NH<sub>3</sub> concentration (y). A generic R function, *plot\_relationships()*, facilitates exploratory analysis by visualizing both linear and nonlinear (quadratic) bivariate relationships between NH<sub>3</sub> and each

```

# Barn 1 data: T = Temperature, V = Wind speed, D = Wind direction,
# y = Simulated NH3 concentration

datal <- data.frame(
  T = c(8.2, 0.0, -4.1, 23.8, 20.5, 26.0),
  V = c(2.2, 2.2, 6.7, 1.4, 1.9, 1.5),
  D = c(124, 129, 90, 67, 281, 296),
  y = c(6.5, 3.8, 4.0, 14.0, 13.0, 15.0)
)

# Bivariate analysis using generic function (source prior to use)
plot_relationships(datal, y_var = "y", panelttl = "Barn 1")

# Multivariate regression model
modell <- lm(y ~ T + I(V^2), data = datal)

# Variance inflation factor (VIF) to check multicollinearity
vif(modell)

# Add predictions to dataset
datal$y_predicted <- predict(modell)

# Extract and display uncertainty metrics
UMetrics1 <- extract_lm_uncertainty(modell)
UMetrics1$coefficients      # Coefficient-level uncertainty
UMetrics1$uncertaintyv metrics # Model-level diagnostics

```

**Fig. 7.** Extract of the R-script.

predictor variable. For each predictor-response pair, scatterplots with fitted regression lines are generated alongside annotated  $R^2$  values to quantify explanatory power.

Complementing these visual diagnostics, the function `extract_lm_uncertainty()` computes a comprehensive suite of uncertainty metrics from a fitted linear model. These metrics include coefficient standard errors, confidence intervals, p-values, and model-level statistics such as RMSE,  $R^2$ , adjusted  $R^2$ , and the F-statistic with its significance level.

For illustration, after installing packages and sourcing the functions, the following R code block in Fig. 7 generates both graphical outputs of bivariate exploratory analyses and numerical summaries from multivariate regression models using the barn I dataset.

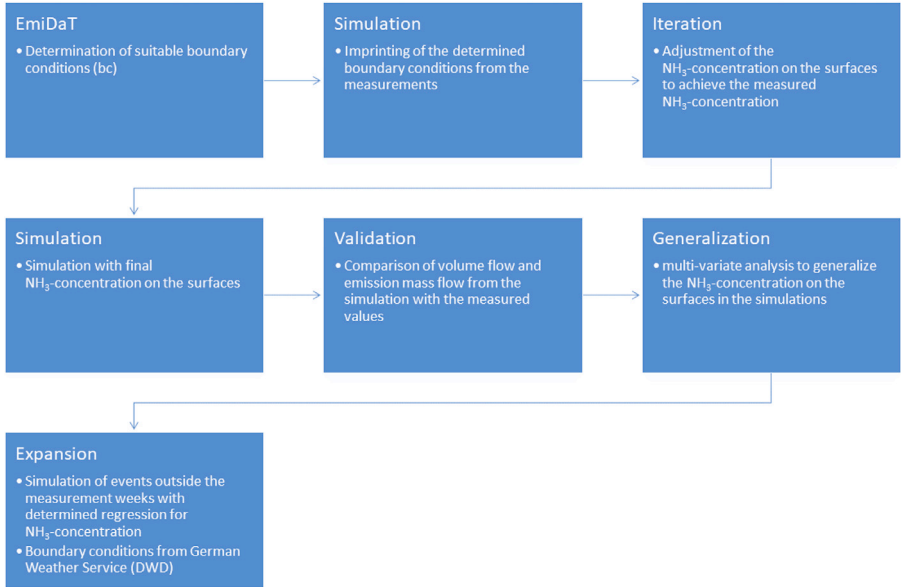
Prior to use, users should review the R script and install the required packages as specified in the accompanying file (*R scripts\_Bivariate and Multivariate Analysis.txt*). The script is designed for ease of use, enabling implementation of the statistical analyses with minimal prior R experience by sequentially running code blocks in the R console.

## 4. Experimental Design, Materials and Methods

The generation and usage of data followed a path which can be seen in the following Fig. 8. Each of these boxes is explained in detail in the following paragraphs.

### 4.1. EmiDaT

The basis of the data set used in this project are the measurement results of the project “Determination of emission data for the assessment of the environmental impact of livestock housing” (*EmiDaT*) [8]. In this project, different dairy barns were measured regarding the emission mass flow of ammonia and other parameters. The measurements were conducted for 6 measuring weeks starting in late 2017 and ending in late 2018. The measurements followed the VERA protocol [9]. The simulations were carried out for three of the investigated dairy barns. They differed in size, animal number and slurry management. Barn I housed an average of 126 cows and had a slurry cellar with slatted floor. Barn II was significantly smaller with 78 animals on



**Fig. 8.** Path of data generation and further usage.

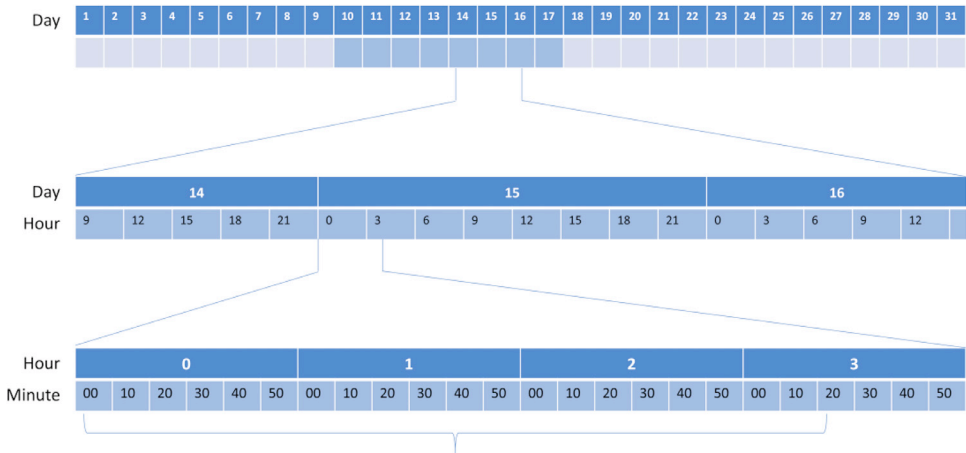
average and an external slurry storage. In barn III, 199 cows were in there on average and the slurry on the solid floor was removed with a scraper. Barns II and III are located in the south-eastern part of Germany and barn I in the north-western area. A detailed description of the measurements conducted and the barn managements can be found in [8].

In order to later validate the numerical models with the measured results, the boundary conditions for these simulations had to be replicated in the simulation for the certain results to compare with. The parameters needed for the simulations as boundary conditions were the ambient temperature, the local wind speed and the local wind direction. The latter two were measured at a weather station in 10 m height near the barn. The temperature was measured close to the barn. The variables were usually measured at an interval of once per minute.

For each of the six measurement weeks for each barn, one event was extracted for simulation. Since the simulations are steady (i.e., time-independent), it was necessary to identify situations where the boundary conditions were as stable as possible, so the simulation would closely approximate reality. First, the evaluation was done visually by looking at the time series of three variables: ambient temperature, wind speed, and wind direction. Here, periods were picked that showed a small “visual variance” in at least two of the variables. These periods were evaluated further with two main tools. Box plots, generated with Microsoft Excel (Version 2021, Microsoft Corporation, Redmond, WA, USA), were used to visualize the variance of the three parameters across the periods. The interquartile range (Q3–Q1), mean, median, and data dispersion (whiskers) provided a good indication of the suitability as boundary conditions for the simulation. Furthermore, for quantitatively assessing the data consistency, the Coefficient of Variation (CoV  $\equiv \lambda$ ) was used, which is described by [3] and can be seen in Eq. (1) (Eq. (1)). It is the ratio of the root mean square deviation  $\sigma(k)$  and the mean  $\bar{\phi}(k)$  for a variable  $k$ .

$$\text{CoV} = \lambda(k) = \frac{\sigma(k)}{\bar{\phi}(k)}, \text{ where } \sigma(k) = \sqrt{\frac{1}{n} \sum_{j=k-n+1}^k [\phi(j) - \bar{\phi}(j)]^2}, \bar{\phi}(k) = \frac{1}{n} \sum_{j=k-n+1}^k \phi(j) \quad (1)$$

This process led to the identification of one period (event) per measurement week per barn which results in 18 simulations for the validation set. The final value for each parameter is given



Chosen timeframe of the measurement week for evaluating parameters for simulation’s boundary conditions:  
**15th of January, 12:13 am – 3:22 am**

**Fig. 9.** Exemplary visualization of determining the simulation event within measurement week 2 of barn I.

as the mean over the chosen period. The CoV values are also presented in the excel-sheets for each barn.

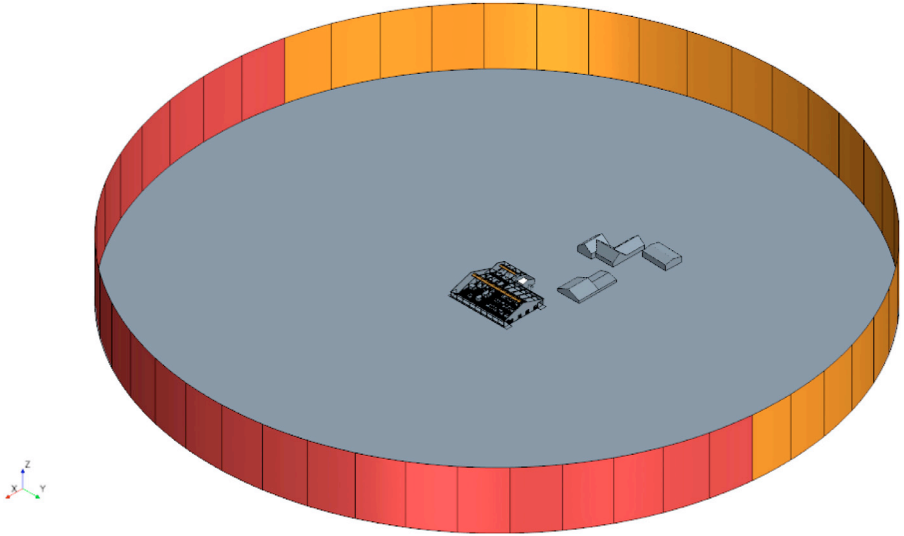
For example, Fig. 9 illustrates the selected timeframe of measurement data from week 2 in barn I.

#### 4.2. Simulation I

The three-dimensional, steady-state simulations were carried out with *Simcenter StarCCM+* in version 2020.3.1 by Siemens PLM. The walls of the buildings had the no-slip boundary condition. The top boundary of the simulation domain was modelled as a slip-wall. The standard  $k-\epsilon$  model was chosen for resolving the turbulence [10]. The animals were modelled as porous blocks using a single animal approach to eliminate the need for extensive pre-simulations [11]. Constant density and the Boussinesq approximation were further simulation settings due to the small thermal gradient between the ambient air and animal body temperature [12]. Inlets were modelled as velocity inlets and outlets as pressure outlets with 0 Pa pressure. Their positioning depends on the inlet angle. Fig. 10 shows the inlet and outlet distribution for barn I with the inlets being marked in red, and the outlets in orange.

Due to modelling the simulation domain as a cylinder, the outer surfaces can be divided into  $n$  equally sized parts. For barn I,  $n = 48$ , so that each inlet or outlet represents  $7.5^\circ$  of the entire lateral surface. For barns II and III,  $n = 24$  was chosen, meaning each segment corresponds to a  $15^\circ$  section of the lateral surface. The simulations showed that for a better overall convergence, there should always be  $(n/2 + 1)$  inlets. When assigning the inlets and outlets to the surfaces accordingly, the inlet-section that contains the inlet angle, should be in the middle of the inlet chain. For example, if the inlet angle is  $\alpha = 127^\circ$ , the first velocity inlet to assign is on the surface which represents the angular section  $120^\circ \leq \alpha < 127.5^\circ$  for barn I. From there,  $n/4$  surfaces to both sides should be velocity inlets, while the remaining surfaces are designated as pressure outlets. In the simulation files, the surfaces are named according to their smallest angular value they contain. For instance, for barn I, the section covering  $22.5^\circ \leq \alpha < 30.0^\circ$  is labeled “22\_5”.

The release of ammonia was not directly modelled (in a chemical way). It was imprinted on the surfaces (slurry surface, alleys, slatted floor) as a source for a passive scalar. The spatial resolution of this source was neglected and set to a constant value for all surfaces.



**Fig. 10.** Visualization of the inlets (red) and outlets (orange) of the simulation domain for barn I.

Convergence was considered achieved when the global residuals for continuity, momentum, energy, and turbulence dropped below  $10^{-3}$ . Additionally, the monitored values for ammonia concentration at the measurement points, volume flow through the barn, and climate parameters were required to show no significant variance.

With the identified boundary conditions from above and an initial ammonia surface concentration, the first simulation for each event was carried out and processed to the next step after meeting the convergence criteria.

#### 4.3. Iteration and simulation II

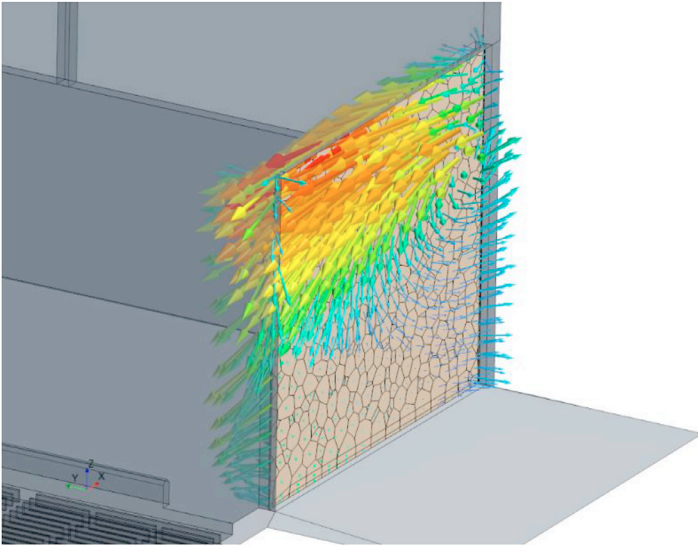
In this step, the ammonia concentration at the simulated measurement points was first evaluated. To do this, the average was formed over all measurement points, analogous to the measurement procedure (see Eq. (2)).

$$\bar{c}_{NH_3} = \frac{1}{N} \sum_{i=1}^N c_{NH_3,i} \quad (2)$$

If the ammonia concentration was too low, the surface concentration of ammonia in the simulation was manually increased. If it was too high, it was reduced accordingly. The simulation was then restarted and allowed to run until the ammonia concentration converged again. The described comparison between simulated and measured ammonia concentrations was then repeated. This iterative process continued until the values in the simulation equalled the measured concentration. The adjusted ammonia concentration from the simulation is an important outcome variable, used in the statistical analysis as outlined in the R-Code and Generalization sections of this paper.

#### 4.4. Validation

The simulations were then validated by comparing the resulting volume flow and emission mass flow in the simulation with the measured values of the EmiDaT project.



**Fig. 11.** Exemplary visualization of flow through a sidewall opening of a barn as a vector plot.

#### 4.4.1. Evaluation of volume flow

To get the volume flow through the barn  $\dot{V}$  in  $\text{m}^3 \cdot \text{s}^{-1}$ , the area of the sidewall openings  $A_i$  in  $\text{m}^2$  of the barn and the velocity component  $v_{\perp}$  perpendicular to that area in  $\text{m} \cdot \text{s}^{-1}$  need to be evaluated. This was done via field functions in the solver for each element on the openings. The resulting volume flow for the barn is calculated via the sum over all elements in the openings generally following Eq. (3).

$$V_{barn} = \sum_{i=1}^N v_{\perp,i} \cdot A_i \quad (3)$$

In detail, this was evaluated for each coordination direction. Fig. 11 shows an exemplary visualization of a flow through a sidewall opening of a barn. It can be seen that some areas of the opening function as inlets where the air is moving into the barn and some as outlets where the air is coming out of the barn at the same time. For the volume flow through the barn, only the part of the flow that points outward of the barn is considered as there are no sources of flow inside the barn.

In order to evaluate the volume flow coming out of the barn, the area of elements with a velocity vector pointing outwards and the velocity component perpendicular to that surface are needed. As the barn is built up perpendicular to the coordinate axis, the two parameters can be directly addressed via field functions. In the given example, the area is detected depending on the velocity component in negative y-direction via an if condition. The syntax of StarCCM+ follows the scheme “condition ? if true: if not true”. For the certain case, this results in “ $\$Velocity[1] < 0 ? \text{mag}(\$Area) : 0$ ” (if the velocity in y-direction is smaller than zero, return the area of that element and if not, return zero). In the same way, the velocity itself is evaluated: “ $\$Velocity[1] < 0 ? \text{abs}(\$Velocity[1]) : 0$ ”. It is important to use the absolute value here as the volume flow through each element should add up to the final volume flow of the barn. Finally, the product of the two field functions returns a positive value for vectors having a velocity component in negative y-direction and 0 for the other elements. This evaluation is supposed to be done for every direction and every sidewall opening of the barn.

#### 4.4.2. Estimation of ammonia emission mass flow

The product of the volume flow through the barn  $V$  in  $\text{m}^3 \cdot \text{s}^{-1}$  and the measured concentration of ammonia  $c$  in  $\text{g} \cdot \text{m}^{-3}$  gives the emission mass flow  $m$  (Eq. (4)) in  $\text{g} \cdot \text{s}^{-1}$ . The ammonia concentration  $c$  is determined as an average over all measuring points [8].

$$m_{\text{NH}_3} = V_{\text{barn}} \cdot \bar{c}_{\text{NH}_3} \quad (4)$$

#### 4.4.3. Evaluation of model performance

For validating the simulation, an appropriate measure is required. In our case, the Normalized Mean Square Error (NMSE) was chosen [13]. It is defined in Eq. (5) as follows:

$$\text{NMSE} = \frac{1}{N} \sum_{i=1}^N \frac{(P_i - M_i)^2}{\bar{P} \cdot \bar{M}}. \quad (5)$$

Here,  $P_i$  and  $M_i$  are the predicted and measured values from the  $i^{\text{th}}$  sample ( $i = 1, 2, \dots, N$ ) respectively, and  $\bar{P}$  and  $\bar{M}$  are their arithmetic means. With this definition, the closer the NMSE gets to zero, the better is the agreement. The investigations were done for the measured and simulated volume flow through the barn. For further evaluations, box plots were also used for illustrating the variance between the simulation results and the measured values.

#### 4.5. Generalization

To elucidate generalized relationships between these climatic variables and ammonia surface concentrations, we conducted both bivariate and multivariate regression analyses using R (version 4.4.2) [14]. The bivariate analyses separately examined linear and nonlinear effects of temperature, wind speed, and wind direction, while the multivariate regression assessed their combined predictive capabilities and potential interactions.

Ensuring full transparency and reproducibility, all related R scripts are publicly accessible via the repository [7]. These scripts enable users to:

- Reproduce the complete statistical analyses across all barns, facilitating validation and comparative studies;
- Explore individual and combined effects of meteorological variables on ammonia concentrations through straightforward functions;
- Compute detailed uncertainty metrics at the coefficient level, including confidence intervals, standard errors, and p-values;
- Obtain comprehensive model diagnostics such as RMSE, R-squared, adjusted R-squared, F-statistics, and variance inflation factors (VIF) for assessing model fit and multicollinearity.

By providing these resources, we enhance the transparency, usability, and adaptability of our dataset and statistical analysis. Researchers with minimal R experience can implement the analyses by sequentially running annotated code blocks, making this workflow broadly accessible. Moreover, the modularity of the scripts facilitates easy extension or adaptation to new datasets, barns, or environmental conditions, thereby supporting ongoing scientific validation and broader methodological generalization.

#### 4.6. Expansion

In order to simulate events which took place outside the measurement periods, the two main values needed are again the boundary conditions (temperature, wind speed, wind direction) and the ammonia concentration on the surfaces for the simulation. The latter one was derived via the analysis in the generalization step. The climatic boundary conditions were derived using climate data from a nearby weather station of the German Weather Service. The values for the

required parameters are publicly accessible and stored as hourly averages. When selecting the new boundary conditions, care was taken to choose those that differ from the previously simulated ones. Simulations were then carried out with these values following the same setup as described in section Simulation I. The results regarding emission mass flow can be compared to the other simulation results. As a further comparison measure, the literature value for the emission behavior provided by the German technical instruction on air quality control (TA-Luft) [15] was considered as well.

## Limitations

The simulation files including the solution (.sim) can only be opened with StarCCM+. If the files are loaded into a different solver a different file type is required. An attempt can be made to provide this data type after consultation with the corresponding author.

The geometric models of the barns presented here are specific to the barns investigated in the EmiDaT project.

## Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief. We confirm that our work does not involve human subjects, animal experiments, or any data collected from social media platforms.

## Credit Author Statement

**Julian Hartje:** Conceptualization, Data Curation, Formal analysis, Writing – original draft. **Abu Zar Shafiqullah:** Data curation, Formal analysis, Visualization, Writing – review and editing.

## Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work the authors used ChatGPT and DeepL in order to improve the written language. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

## Data Availability

[Application of digital twins to extend the emission data of naturally ventilated dairy cattle barns \(Original data\)](#) (Zenodo).

## Acknowledgements

This work is financially supported by the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the [Federal Republic of Germany](#), granted by the [Federal Office for Agriculture and Food](#) (BLE; grant number 28N206509).

## Declaration of Competing Interest

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

## References

- [1] D. Janke, Q. Yi, L. Thormann, S. Hempel, B. Amon, Š. Nosek, et al., Direct measurements of the volume flow rate and emissions in a large naturally ventilated building, *Sensors*. (Basel) 20 (21) (2020), doi:[10.3390/s20216223](https://doi.org/10.3390/s20216223).
- [2] L. Rong, B. Bjerg, T. Batzanas, G. Zhang, Guoqiang, Mechanisms of natural ventilation in livestock buildings: perspectives on past achievements and future challenges, *Biosyst. Eng.* 151 (2016) 200–217, doi:[10.1016/j.biosystemseng.2016.09.004](https://doi.org/10.1016/j.biosystemseng.2016.09.004).
- [3] S. Calvet, R.S. Gates, G. Zhang, F. Estellés, N.W.M. Ogink, S. Pedersen, D. Berckmans, Measuring gas emissions from livestock buildings: a review on uncertainty analysis and error sources, *Biosyst. Eng.* 116 (3) (2013) 221–231, doi:[10.1016/j.biosystemseng.2012.11.004](https://doi.org/10.1016/j.biosystemseng.2012.11.004).
- [4] W. Wu, J. Zhai, G. Zhang, P.V. Nielsen, Evaluation of methods for determining air exchange rate in a naturally ventilated dairy cattle building with large openings using computational fluid dynamics (CFD), *Atmos. Environ.* 63 (2012) 179–188, doi:[10.1016/j.atmosenv.2012.09.042](https://doi.org/10.1016/j.atmosenv.2012.09.042).
- [5] C.K. Saha, C. Ammon, W. Berg, C. Loebstin, M. Fiedler, R. Brunsch, K. von Bobrutzki, The effect of external wind speed and direction on sampling point concentrations, air change rate and emissions from a naturally ventilated dairy building, *Biosyst. Eng.* 114 (3) (2013) 267–278, doi:[10.1016/j.biosystemseng.2012.12.002](https://doi.org/10.1016/j.biosystemseng.2012.12.002).
- [6] C.K. Saha, Q. Yi, D. Janke, S. Hempel, B. Amon, T. Amon, Opening size effects on airflow pattern and airflow rate of a naturally ventilated dairy building—A CFD study, *Appl. Sci.* 10 (17) (2020) 6054, doi:[10.3390/app10176054](https://doi.org/10.3390/app10176054).
- [7] J. Hartje, A.Z. Shafullah, Application of digital twins to extend the emission data of naturally ventilated dairy cattle barns , zenodo, 2025. <https://doi.org/10.5281/zenodo.15672439>.
- [8] KTBL Ermittlung Von Emissionsdaten für die Beurteilung von Umweltwirkungen der Nutztierhaltung (EmiDaT). Abschlussbericht. Hg. v. Kuratorium für Technik Und Bauwesen in Der Landwirtschaft e. V. (KTBL). Kuratorium für Technik Und Bauwesen in Der Landwirtschaft e. V. (KTBL), 2024 Darmstadt.
- [9] Internationales VERA Sekretariat Vera-Prüfprotokoll für Tierhaltungs- und Management-Systeme. Version 3:2018-09, 2018. Deutsche Fassung accessed 12 February 2025 [https://www.veraverification.eu/app/uploads/sites/9/2019/05/VERA\\_Pruefprotokoll\\_Stallsysteme\\_v3\\_2018.pdf](https://www.veraverification.eu/app/uploads/sites/9/2019/05/VERA_Pruefprotokoll_Stallsysteme_v3_2018.pdf).
- [10] G.K. Ntinias, X. Shen, Y. Wang, G. Zhang, Evaluation of CFD turbulence models for simulating external airflow around varied building roof with wind tunnel experiment, *Build. Simul.* 11 (1) (2018) 115–123, doi:[10.1007/s12273-017-0369-9](https://doi.org/10.1007/s12273-017-0369-9).
- [11] J. Hartje, Development of an anisotropic porous model of a single cow for numerical barn simulations—A numerical study, *Fluids* 10 (2025).
- [12] E.M. Doumbia, D. Janke, Q. Yi, T. Amon, M. Kriegel, S. Hempel, CFD modelling of an animal occupied zone using an anisotropic porous medium model with velocity depended resistance parameters, *Comput. Electron. Agric.* 181 (2021) 105950, doi:[10.1016/j.compag.2020.105950](https://doi.org/10.1016/j.compag.2020.105950).
- [13] L. Xia, H. Montazeri, B. Blocken, On the accuracy of eulerian-lagrangian CFD simulations for spray evaporation in turbulent flow, *Int. J. Thermal Sci.* 208 (2025) S. 109461, doi:[10.1016/j.ijthermalsci.2024.109461](https://doi.org/10.1016/j.ijthermalsci.2024.109461).
- [14] R Core Team: *A Language and Environment for Statistical Computing* [software], 2024 Vienna, Austria.
- [15] Bundesministerium des Innern, für Bau und Heimat, Neufassung der Ersten Allgemeinen Verwaltungsvorschrift zum Bundesmissionsschutzgesetz (Technische Anleitung zur Reinhaltung der Luft – TA Luft), TA-Luft, vom 18.08.2021 Gemeinsames Ministerialblatt, 2021 <https://fragdenstaat.de/dokumente/243256-gmbl-nr-48-54-2021/?page=1> accessed 18 February 2025.