Evaluation of Fire Models by Using Local and Global Metrics and Experimental Uncertainty Estimates: Application to OECD/NEA Prisme Door Tests

O. Riese *, *Division of Fire Safety, Institut Für Baustoffe, Massivbau und Brandschutz, TU Braunschweig, Beethovenstrasse 52, 38106 Braunschweig, Germany*

M. Meyer, *Institut für Mathematische Stochastik, TU Braunschweig, Universitätsplatz 2, 38106 Braunschweig, Germany*

A. Leucht, *Professur für Mathematik in den Wirtschaftswissenschaften, Otto-Friedrich-Universität Bamberg, Feldkirchenstr. 21, 96052 Bamberg, Germany*

Abstract. The use of numerical methods in fire safety investigations for civil buildings and nuclear facilities has received enormous attention in recent years. To evaluate quantities—such as gas temperatures—in fire models, *local* metrics using single points (e.g. comparing maximum or minimum peak value of two time series) are well-established. Experimental (measurement and model input) uncertainty estimates can be used to quantify the model uncertainty. Although the peak value is a relevant and well-defined quantity, *global* metrics comparing the entire course of two time series can often provide additional information for the validation of fire models. A comparative methodology COMET for evaluating the predictive power of fire models is developed and presented in this paper. In the methodology, both local and global metrics are combined to incorporate the explanatory power of both quantities in the validation process. While uncertainty analysis is well established for peak values, to the best of our knowledge, there are no analytic results on quantifying the uncertainty of the global metric in the literature. We address the latter based on experimental measurements and derive confidence regions for both metrics. Finally, this paper summarizes the results using COMET to validate the Fire Dynamics Simulator (FDS) version 6 for a room fire scenario. Validation examples are tests 3, 4 and 5 of the DOOR series of the international OECD/NEA PRISME project, in which the transport of heat and flue gases through a door between two rooms was examined. Using COMET, we can easily identify sensors with high level of agreement between model and experimental results with respect to the local and/or the global metric.

Keywords: Evaluation, Experiment, Uncertainty, Fire modelling, Validation

*Correspondence should be addressed to: O. Riese, E-mail: o.riese@ibmb.tu-bs.de

1. Introduction

In fire safety engineering and many other fields, a variety of issues requires a comparison of different outcome time series of relevant quantities. These can be design issues as well as validation purposes. Both have in common that there is a demand of comparing not only a single time point of a time series but mostly the whole course of two curves (in a cumulative sense). In spite of this common and widespread problem, no standard procedure has been established as a solution yet, even though appropriate mathematical methods are available. An overview of the activities in this field is given in [1]. Documents to assess the predictive capability of fire models, e.g. from ASTM (American Society for Testing and Materials) and ISO (International Standards Organization), can be found in [2, 3], among others. The mathematical foundation is given in [4]. In the nuclear sector in particular, there are extensive approaches towards quantification of the uncertainty in the evaluation process, e.g. in [5–7].

As a solution strategy for the described issue the comparison methodology COMET is developed and presented in this paper. As already intensively discussed in [7], the analysis of single points and the investigation of the overall distance of experimental and simulated curves are of parallel interest. COMET evaluates both criteria simultaneously. Exemplarily, Fig. 1 shows the course of the gas temperature measured in a specific experiment and simulated with a computational fluid dynamics (CFD) model.

Obviously, the peak values of both curves are almost identical (but not at the same time), while the overall behaviour differs significantly. As a metric for local effects the peak distance (PEAK), and for global effects the normalized Euclidean
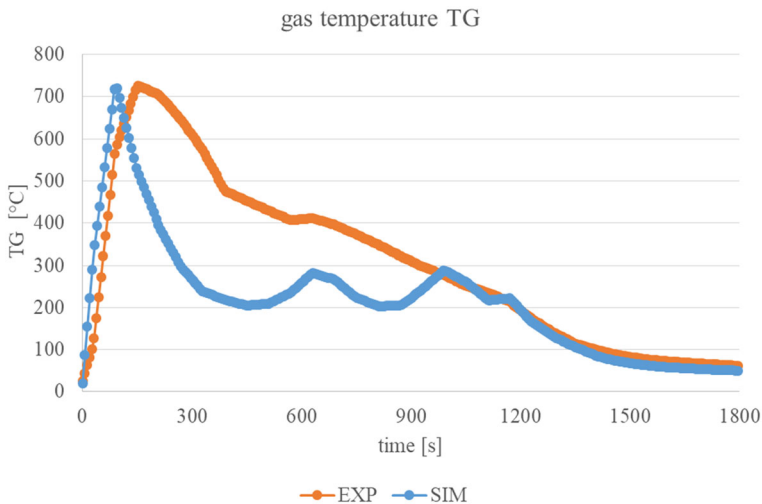

Figure 1. Gas temperature (TG), experiment (EXP) and simulation (SIM) with a CFD fire model, time series representing sensor ''TG_FP_240'' from OECD/NEA PRISME DOOR test 5 [7].

distance-based measure (NED) is used in COMET. The latter global evaluation parameter is considered amongst others by Peacock et al. in [4]. As [7] points out, PEAK (of temperature) is the relevant parameter when e.g. determining the maximum level of stress on building materials during a closed compartment fire. Even if two temperature curves are generally rather close to each other, their PEAK values may differ which may have a significant effect on material stability.

Although for almost all relevant quantities the PEAK criterion is an important information, the course of a quantity over time is an equally important aspect of the impact on the building and for human life in the event of a fire. Hence, the comparison of the overall courses of the entire time series should be considered as well [7]. Conclude that a parallel evaluation of the PEAK and NED metrics is generally advisable.

We introduce the two-dimensional graphical tool COMET for the comparison of model prediction and experiment based on uncertainty quantification of PEAK and NED criteria. While confidence regions for PEAK have already been established by the United States Nuclear Regulatory Commission (NUREG 1824) [5, 6] and the National Institute of Standards and Technology [8], additional effort is required to derive confidence intervals for NED in a similar manner.

In the literature, approximate ranges of uncertainty for NED are therefore sometimes constructed by using the overall experimental uncertainty; see for example [7]. However, as those authors point out, this approach does not take the temporal dependence of time series observations into account (which generally leads to significantly larger confidence regions). The contribution in this paper proposes an improved estimation of the NED uncertainty, which takes the temporal dependencies into account, for the purpose of confidence interval construction. These intervals then serve as ranges of uncertainty (RoU) in the model evaluation process. It is worth noting that the result is a formula that depends exclusively on experimental observations and on the quantities commonly used for PEAK in the literature, and is therefore easily applicable.

Previous studies with respect to local and global effects have already been used to validate simulations concerning results coming from the international OECD[1]/NEA[2] PRISME[3] Project [7, 9, 10]. The acronym PRISME comes from the French phrase *propagation d'un incendie pour des scénarios multi-locaux élémentaires*, which in English can be translated as "fire propagation in elementary, multi-room scenarios". In particular, the PRISME LEAK [11] and PRISME DOOR [12, 13] test series have been already investigated to validate the CFD fire simulation code FDS Version 6 within the OECD/NEA PRISME project. As an example, the PRISME DOOR test 3 (PRS_D3), PRISME DOOR test 4 (PRS_D4), and PRISME DOOR test 5 (PRS_D5) have been chosen for this publication in order to demonstrate the methodology and the results gained by its application for FDS Version 6.7.0 [14].

---

[1] Organisation for Economic Co-operation and Development.
[2] Nuclear Energy Agency.
[3] https://www.oecd-nea.org/jcms/pl_24955/fire-propagation-in-elementary-multi-room-scenarios-prisme-project.

The remainder of this paper is structured as follows. In Sect. 2 the two metrics PEAK and NED used in our novel tool COMET are reviewed and discussed. COMET is then introduced in Sect. 3. As an example for the application of COMET we use results from the PRISME DOOR experiment series described in Sect. 4. Section 5 describes the CFD simulations to be validated. The application of COMET in the setting of Sects. 4 and 5 is described in Sect. 6. Section 7 summarizes our results. Mathematical derivations for the uncertainty estimates are deferred to the Appendix.

2. Evaluation Criteria

The methodology proposed in this paper for the analysis of differences between two time series—one from the experiment and one from a simulated model—includes two criteria. We combine the well-established PEAK method according to [3] with the normalized Euclidean distance (NED) introduced in [2].

The local comparison criterion PEAK describes the relative difference of model peak and experimental peak, and is given by

$$PEAK = \frac{\Delta M_p - \Delta E_p}{\Delta E_p} = \frac{(M_p - M_0) - (E_p - E_0)}{E_p - E_0} \tag{1}$$

where ΔM_p is the difference between the peak value M_p of the model prediction and its baseline value M_0, and ΔE_p is the difference between the peak value E_p of the time series of experimental measurements and its baseline value E_0. This method allows a very fast and easy evaluation of the deviation—with regard to the extreme values—of the time series from each other. However, it does not give information about global similarities of the paths of the two time series. In particular, it lacks information about the time points of the extreme events ΔM_p and ΔE_p. Moreover, the distance of model prediction and experimental measurements at any other time point is not incorporated. This may be a challenge for model evaluation in practice, cf. Figure 1 for an example with small PEAK value where the curves differ significantly on a global scale.

As an additional criterion for a global comparison, the normalized sum of squared differences—the so-called normalized Euclidian distance NED (also known as standardized L2 norm) is used in COMET which is defined via

$$\mathbf{NED} = \sqrt{\frac{\sum_{t=1}^{T} (\Delta M_t - \Delta E_t)^2}{\sum_{t=1}^{T} (\Delta E_t)^2}} \mathbf{with\ T = number\ of\ data\ points}, \tag{2}$$

where t runs from 1 until T and represents the time points at which each quantity is measured (E_t) or modelled (M_t). As before, $\Delta M_t = M_t - M_0$ and $\Delta E_t = E_t - E_0$. It was initially presented by Peacock in [2] and is easily interpretable as highlighted in [7]. In the present context, NED describes a quantification for the deviation of the experimental and the model prediction curve during the whole temporal

course of the time series and has some key benefits. Note that squaring of the deviation at the single data points ensures that positive and negative deviations cannot compensate each other. Moreover, the consideration of *squared* distances implies a stronger penalization of large distances compared to small distances.

Larger values (e.g. the temperatures in the hot gas layer) are expected to have larger absolute deviations compared to those with smaller values (e.g. the temperatures in the lower, cooler gas layer). A comparison based on absolute values and without normalization is generally feasible, but might assign large NED values to certain time series although their relative differences may be small. This can be avoided by normalisation with respect to the measurements. On this basis, it is also possible to compare the metrics of different physical quantities; see Fig. 2 for an illustration example. There, for a large number of different sensors in a particular PRISME DOOR experiment, both the PEAK and the NED deviation between model simulation and experimental measurement are calculated for each sensor, and the (NED, PEAK) results of the different sensors are depicted as points in a two-dimensional scatterplot. In Fig. 2 σ, μ and m denote empirical standard deviations, means and medians of PEAK and NED, respectively.

Comparing PEAK and NED, first note that PEAK has neither an upper nor a lower bound and yields the value 0 as an optimal result, in terms of an exact congruence of the extreme values in experiment and simulation. The value -1 describes a significant limit. For values smaller than -1, the signs of the extrema
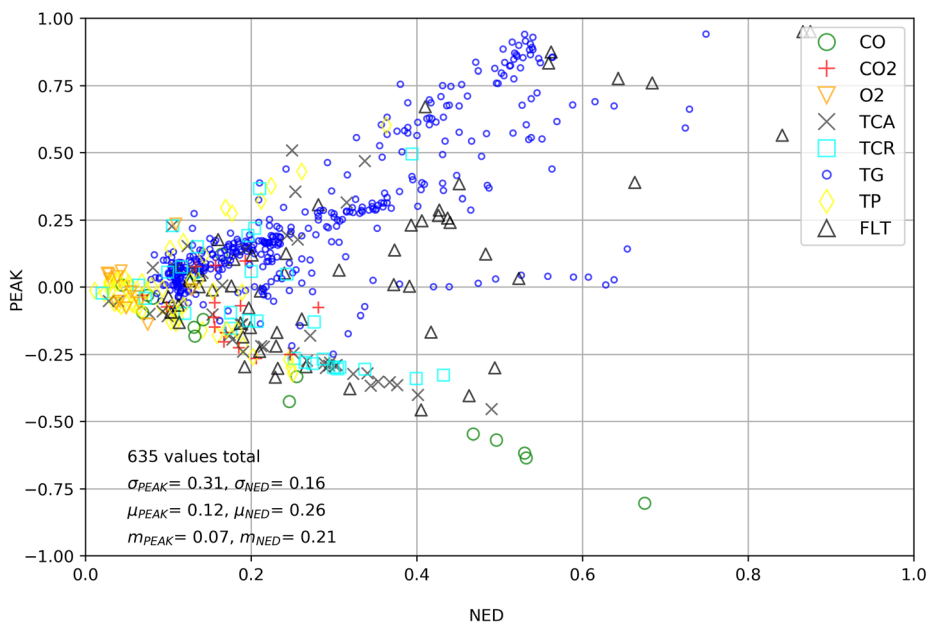

Figure 2. **PEAK and NED results for different sensors (for abbreviations see Table 5 below), PRISME DOOR, test 3, test 4 and test 5, model FDS 6.7.0.**

of the experiment and simulation are different. A detailed analysis of the time series is advisable in this case. The NED value in Eq. (2) merely has a lower bound, the value 0, which, as for PEAK, describes an optimal result. Finally, it can be seen from Fig. 2 that there is a significant difference between (absolute) PEAK and NED values for the majority of the experiments. Thus, in order to decide whether the model is suitable for the experiment at hand, it is indispensable to incorporate confidence regions for PEAK and NED in the validation process.

3. The Tool COMET

COMET is a two-dimensional graphical tool for the comparison of model prediction and experiment based on uncertainty quantification of NED and PEAK criteria. Figure 2 shows the basis for this tool: A scatterplot of the observations from sensors from one experiment, where for each sensor the NED value between model and experiment is depicted on the abscissa, and the PEAK value on the ordinate. To get a benchmark for the magnitude of these values, COMET also contains approximate 95% confidence ranges for both quantities. The result can be seen in Fig. 3; the red area gives the confidence range for NED, the green one the confidence range for PEAK. At the intersection of the red and green area both metrics are within their respective confidence regions. To assess the boundary of these ranges, we rely on [5] (Sect. 1.4.2) for the PEAK criterion. There, the 95% confidence interval is approximated by $[-U_{\text{PEAK}}, +U_{\text{PEAK}}]$, where
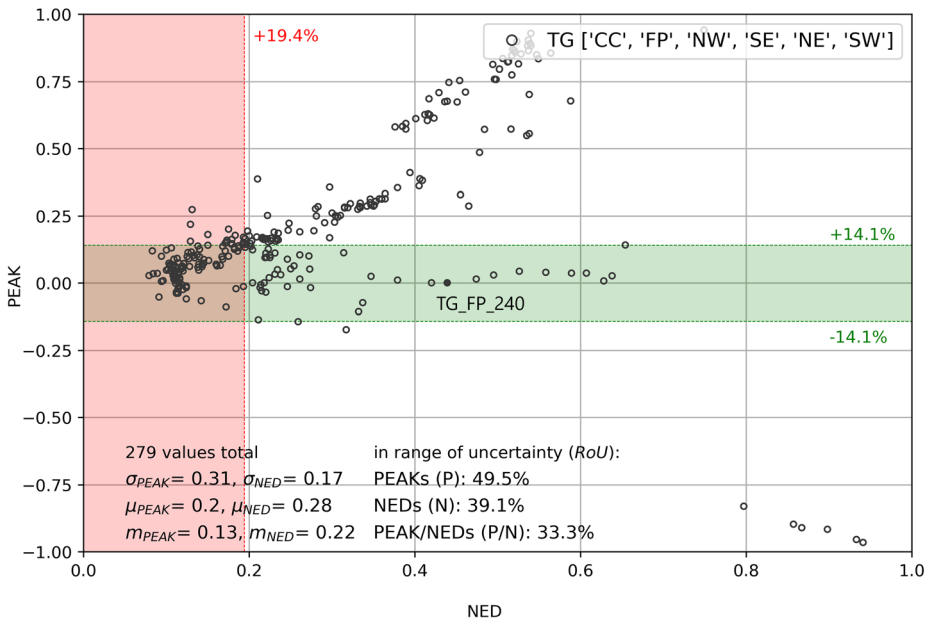

Figure 3. Illustration of the two-dimensional graphical tool (COMET) for the comparison of model prediction and experiment based on uncertainty quantification of NED and PEAK criteria.

Table 1
Measurement (E) and Model Input (M) Uncertainty $\tilde{U} = 2\tilde{u}$ According to [6]

Measured variable in [5]	$\tilde{U}_E$	$\tilde{U}_M$	Reference to measured variables for this study (sensor group), see Table 5
Hot gas layer temperature	0.10	0.10	TG_L1_YY and TG_L2_YY; YY = NW, SE, NE, SW, and CC (only for L2)
Ceiling jet temperature	0.10	0.10	TG (height = 390 cm) not evaluated in this study
Plume temperature	0.10	0.10	TG_L1_FP
Gas concentrations	0.01	0.15	O_2, CO, CO_2
Smoke concentration	0.28	0.26	Not evaluated in this study
Room pressure rise	0.20	0.42	Not evaluated in this study
Surface/target temperature	0.10	0.10	TP, TCR, TCA
Heat flux density	0.10	0.20	FLT

See Table 5 for abbreviations and Fig. 6 and Fig. 7 for sensor locations

$$U_{\text{PEAK}} = \sqrt{\tilde{U}_E^2 + \tilde{U}_M^2}. \tag{3}$$

Here, $\tilde{U}_E$ denotes a measure of relative experimental (measurement) uncertainty and $\tilde{U}_M$ denotes a measure of relative numerical (model input) uncertainty, see [5] (Sect. 1.4.2) for details. For the present study the relative uncertainties $\tilde{U}_E = 2\tilde{u}_E$ and $\tilde{U}_M = 2\tilde{u}_M$ to determine U_{PEAK} are taken from [6] (see Sect. 3.3.3) and are shown in Table 1. This approach will be used as the basis for evaluating the PEAK results of PRISME DOOR.

To the best of our knowledge, there are no results on uncertainty quantification based on NED in the literature. For that reason [7] used $\tilde{U}_E$ as a rough approximation. However, it turns out that one can follow the ideas presented in [5] (Sect. 1.4.2) to derive an analogue to U_{PEAK} which is in general substantially larger than $\tilde{U}_E$. Still, the design of the underlying time series model is much more delicate since we have to consider temporal dependencies within the experimental and within the simulated curves. A detailed description of our approach is given in the Appendix. Finally, we will obtain an approximation for the variance of NED:

$$\widehat{\text{Var}}(\text{NED}) = \frac{\left(\sum_{t=1}^{T} \Delta E_t\right)^2}{\sum_{t=1}^{T-1} \Delta E_t \Delta E_{t-1}} \cdot \frac{\tilde{u}_M^2}{T} + \frac{\tilde{u}_E^2}{T}. \tag{4}$$

In line with the PEAK approach, the latter quantity is a function of the uncertainties $\tilde{U}_E = 2\tilde{u}_E$ and $\tilde{U}_M = 2\tilde{u}_M$ given in Table 1. As the NED values are always positive, the accuracy of numerical predictions concerning the NED criteria is given with the confidence interval $[0, + U_{\text{NED}}]$ with

$$U_{\mathrm{NED}} = 2 \cdot \widehat{Var}(\mathrm{NED})^{\frac{1}{2}} \tag{5}$$

In Fig. 3 an illustration of the two-dimensional graphical tool (COMET) for the comparison of model prediction and experiment based on uncertainty quantification of NED and PEAK criteria is given. Every sensor is marked by a point in the plane with its NED value on the abscissa and its PEAK value on the ordinate. The *RoU*s for PEAK (P) and for NED (N) resulting from the confidence intervals $[-U_{\mathrm{PEAK}}, + U_{\mathrm{PEAK}}]$ and $[0, + U_{\mathrm{NED}}]$, respectively, are highlighted as green and red bands. The uncertainty measures for PEAK and NED are denoted in percent and are averaged values for the compared sensors. In Fig. 3 PEAKs (P), NEDs (N) and PEAK/NEDs (P/N) denote the proportions of sensors lying in the green *RoU* of PEAK, in the red *RoU* of NED and the intersection of the green and red *RoU*s, respectively. Additionally, for statistical evaluation, standard deviation (σ), mean (μ), and median (m) for PEAK and NED have been computed for all data and for individual sensors. In this sense, Fig. 3 delivers a rather compact but yet informative presentation of our model evaluation. It can be summarized as follows:

- For all sensors (points) in this graphic lying in the intersection of the green and red area, model and experiment fit well with respect to both criteria; NED and PEAK.
- For all sensors (points) in this graphic lying within the green but outside red area, model and experiment fit well with respect their PEAK values but differ significantly with respect to their overall structure (and vice versa for points within the red but outside the green area).
- If the sensor points are located in the white area, there are doubts on the validity of the model.
- For sensors behaving as described in the latter two bullet points, a deeper analysis is advisable, e.g. using the actual plots of modeled and measured results as displayed in our Fig. 1.

In contrast to PEAK, NED is capable of detecting deviations between the curves. As an example, the (NED, PEAK) point resulting from sensor TG_FP_240 illustrated in Fig. 1 is highlighted (black filled). While PEAK is nearly zero (in particular, lying within the green *RoU* of PEAK), NED is remarkably large and, hence, outside the red *RoU* of NED. This can be interpreted as follows: While the maximum values of the experiment are reasonably simulated by the model, this is not the case for the overall gas temperature curves over time. In Sect. 6.2, this behavior is discussed in further detail.

4. OECD/NEA PRISME DOOR Experiments

4.1. *Experimental Setup*

The tests of the first test series PRISME DOOR within the OECD/NEA PRISME project [12, 13] were carried out by the IRSN (Institute de Radioprotection et de Sûreté Nucléaire) in the test facility DIVA in Cadarache (France). The experi-
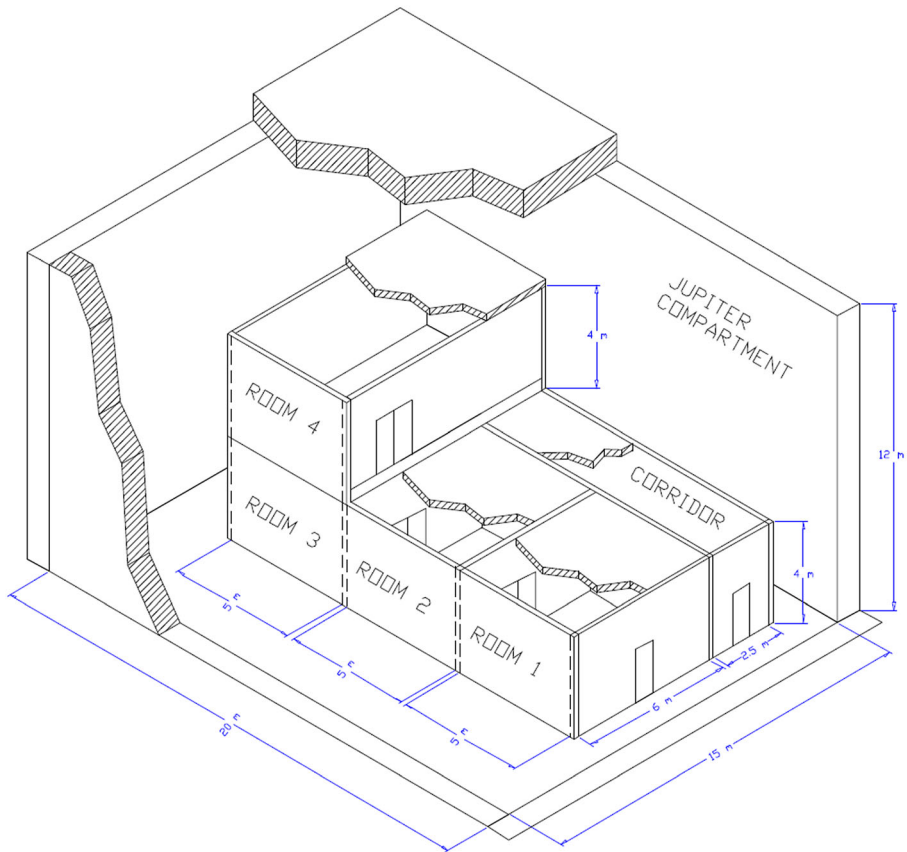
Figure 4. View of the experimental facility DIVA, housed in the JUPITER facility.

ments on the DOOR series (1 to 5) of the OECD/NEA PRISME project were carried out in the rooms "Local 1" (L1 or Room 1) and "Local 2" (L2 or Room 2) of DIVA, cf. Figure 4 for an overview of the spatial conditions. The DIVA compartments are located in the JUPITER facility, which has a volume of total 3600 m^3 and a net volume of approximately 2700 m^3 considering the DIVA internals.

4.2. Room Geometry and Ventilation

Each of the lower cuboid-shaped rooms (room 1 to 3) has a volume of 120 m^3 with a clear dimension of 6 m $\times$ 5 m $\times$ 4 m and is connected to a complex ventilation system, which controls the rooms' air exchange via inlet and outlet ducts (see Fig. 5). The walls of these rooms are made of 30 cm thick concrete. During the tests, the ceiling and the walls of the fire room (room 1) were provided with a 5 cm thick insulation layer of rock wool, in order to avoid spalling of the concrete
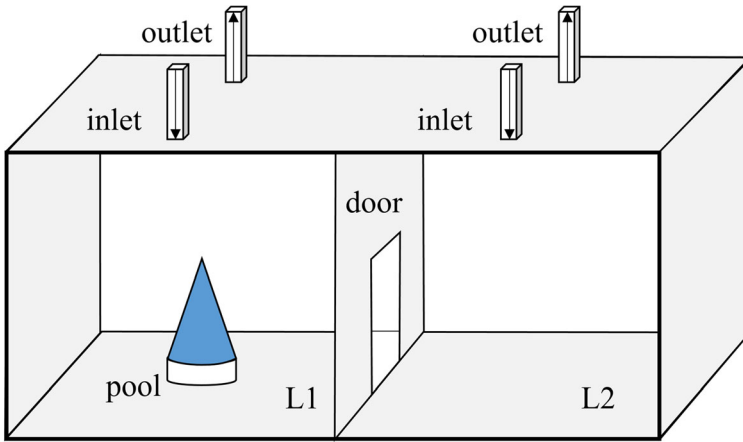
Figure 5. Schematic representation of the ventilation system, L1 (fire room left side) and L2 (target room right side).

Table 2
Pool Area and Air Exchange Rate During the PRISME DOOR Tests PRS_D3, PRS_D4 and PRS_D5

Test	Pool area	Air exchange rate
PRS_D3	0.4 m^2	4.7 1/h or 560 m^3/h
PRS_D4	0.4 m^2	8.4 1/h or 1000 m^3/h
PRS_D5	1 m^2	4.7 1/h or 560 m^3/h

and thus damage to the test facility. For the DOOR series, the door between the two rooms (L1 and L2) was open. The door opening has a dimension of 0.8 m × 2 m and is located in the middle of the partition (see Fig. 5).

The air exchange rate in the tests PRS_D3 and PRS_D5 was 4.7 1/h or 560 m^3/h for both rooms (fire and target room); in test PRS_D4 the air exchange rate was 8.4 1/h or 1000 m^3/h, see Table 2.

The inlet and outlet ports during these tests were located at the top of the compartments, which means 75 cm below the ceiling.

4.3. Fire Source

The fire source (liquid pool, see Fig. 5) was modelled as a rectangular area roughly corresponding to the pool size used in the experiment, which was 0.4 m^2 for PRS_D3 and PRS_D4 and 1 m^2 for PRS_D5, respectively. The required reaction parameters for the fuel (n-dodecane) used in all experiments (C12H26) can be specified directly in the CFD program and are given in Table 3. Soot yield (SOOT_YIELD), carbon monoxide yield (CO_YIELD) and the heat released per unit mass O_2 (EPUMO2) were taken from the SFPE Handbook [15].

Table 3
Reaction Parameters

Parameter abbreviation	Parameter	Value
MW_FUEL	Molecular weight of fuel [g/mol]	170
NU_O2	Stoichiometric coefficient for O_2	18.5
NU_CO2	Stoichiometric coefficient for CO_2	12
NU_H2O	Stoichiometric coefficient for H_2O	13
RADIATIVE_FRACTION	Amount of heat emitted by flames as thermal radiation	0.35
EPUMO2	Heat released per unit mass O_2 [kJ/kg]	12 700
CO_YIELD	Carbon monoxide yield	0.011
SOOT_YIELD	Soot yield	0.041

Table 4
Material Properties

Material	Conductivity λ [W/mK]	Density ρ [kg/m^3]	Specific heat c_p [kJ/kgK]	Thickness d [m]	Emissivity ε
Concrete (room L1 and L2)	1.78–0.80	2 240	0.870–0.317	0.30	0.70
Stone wool (ceiling room L1 and L2)	0.036–0.096	140	0.840	0.05	0.95
PVC-cable (analytical cable)	0.143–0.151	1 380	0.933–1.548	0.025	0.90
PVC-cable (real cable)	0.290–0.255	1 190	1.014–1.499	0.0277	0.80
Steel (pan)	75	7 850	0.484	0.005	0.90
Steel (air channel: inlet, outlet)				0.0013	0.90

4.4. Materials

Table 4 contains the properties of the materials for walls, ceilings, fuel pan and ventilation ducts used in the experiments. The properties for PVC analytical and real cable can be found as well. In FDS a simple 1D heat conduction calculation is carried out across the thickness d of the material and, as a result, the temperatures and the gradient inside are determined. In addition, as Table 4 shows, the thermal conductivity λ and the specific heat capacity c_p can be set as a function of the temperature.

4.5. Instrumentation

In Fig. 6 a schematic of the two rooms, noted L1 and L2, and selected instrumentation are given. For more information concerning the legend see Table 5.

4.6. Target Objects

In order to investigate the effects of the fire on safety-related objects (so called targets) in the experiments on the PRISME DOOR series two types of objects are used: On the one hand, PVC rods, so-called "analytical cables", and on the other
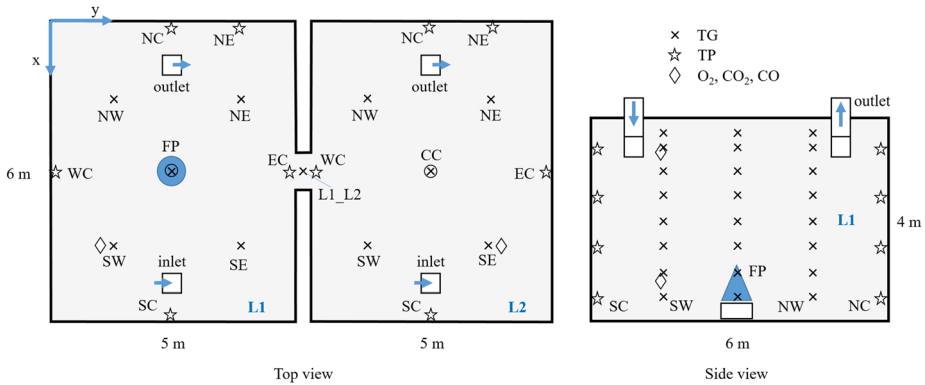
Figure 6. Schematic of the fire room L1, and target room L2, and illustration of selected instrumentation, NW (north west), SE (south east), NE (north east), SW (south west), CC (center), and FP (plume).

Table 5
Investigated Quantities and Abbreviations for Different Measurement Locations

Abbreviations (sensor group)	Quantities	Unit
CO	Carbon monoxide	mol/mol
CO_2	Carbon dioxide	mol/mol
O_2	Oxygen	mol/mol
TCA; TCR	Temperature analytical cable; temperature real cable	°C
TG	Gas temperature	°C
TP	Temperature wall surface	°C
FLT	Total heat flux	kW/m^2

real PVC cables (used in the tests PRS_D4 and PRS_D5). These objects were placed on horizontal steel ladders on the walls at the top and bottom of both rooms Li (i = 1, 2), as shown in Fig. 7 in detail and later in Fig. 9. In addition, the total heat flux density (FLT) at two locations (_UPW and _DWN) and gas temperatures at the steel ladders (TG) nearby the target (_TA) were measured, and analytical (TCA) and real cable temperatures (TCR) at different positions of the cables (_SURF, _INTER, and _CENTER) each at three sections (not shown) in each target object.

5. Simulations with Numerical Model

We illustrate the COMET approach with simulations based on the Fire Dynamics Simulator (FDS) which is well-established in the fire community and frequently used internationally. However, COMET is not restricted to this particular model but can be applied to other numerical fire models as well. Extensive information about the FDS model is given in [14]. FDS was used without changing the default
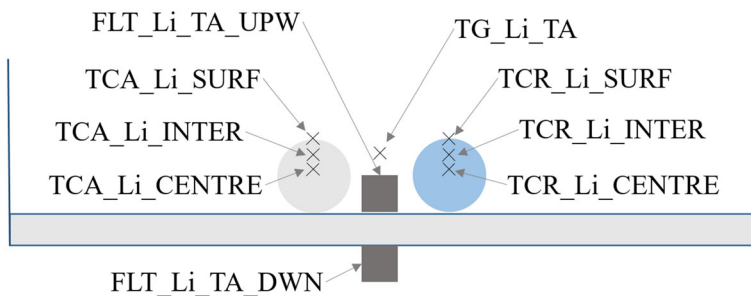
Figure 7. Arrangement of analytical (TCA) and real cables (TCR), PRISME DOOR experiments, i = 1, 2 (room number), UPW (upward), DWN (downward).

settings of the respective versions. For the calculations, however, the starting temperatures measured in the experiment were used as initial conditions.

"Open" simulations are evaluated in this paper. In open simulations, there is essentially a specification of the scenario with regard to the geometry of the room, the physical parameters of the enclosure components and the course of the experimentally determined heat release rate of the fire source, which is known as the input parameter. The specification of the heat release rate corresponds to the state of the art in fire modelling purposes.

The simulations for PRS_D3, PRS_D4 and PRS_D5 were performed with FDS, version 6.7.0 [14]. The results are presented and discussed in the following section.

For the simulation, time courses of the heat release rate (HRR) must be specified. For this study, open simulations for validation were used as a basis, that is, the course of the HRR which was measured (see Fig. 8) during the individual experiments in the test facility DIVA. As heat of combustion a value of 46 MJ/kg was used for n-dodecane. Volume flows (inlet/outlet) were specified as boundary conditions.

A 10 cm grid was used for the simulation of the gas phase. Finer grating checks have not revealed any significant change in the calculated sizes. The calculation of heat conduction in solids was carried out independently of the gas phase on a much finer grid.

Modelling suppression of a fire due to the exhaustion of oxygen within a closed compartment is challenging because the relevant physical mechanisms typically occur at subgrid-scale. Flames are extinguished due to lowered temperatures and dilution of the fuel or oxygen supply [14]. FDS with default settings uses a simple suppression model. There must be sufficient energy released to raise the cell temperature above the critical flame temperature for combustion to occur. This is the case when SUPPRESSION = TRUE is set in the FDS input file. To illustrate the advantages of using COMET, FDS version 6.7.0 was also used without any flame suppression (i.e. SUPPRESSION = FALSE in the FDS input file). For results concerning different settings from FDS SUPPRESSION parameter consult Sect. 6.2.
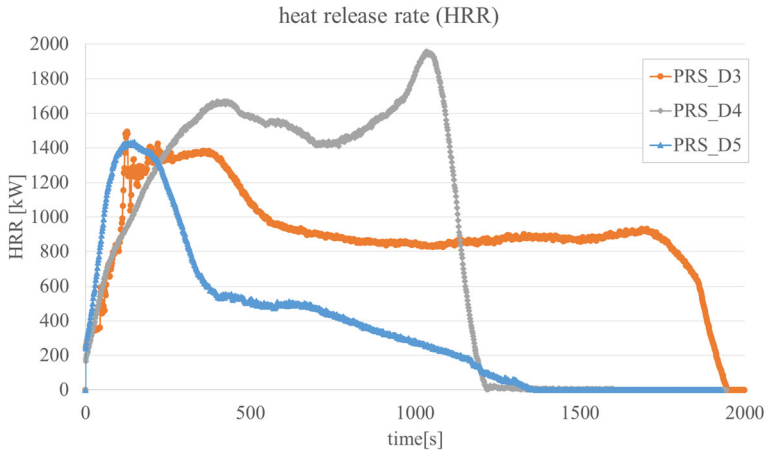
Figure 8. Heat release rate (HRR) for the tests PRS_D3, PRS_D4 and PRS_D5 (experimental measurement and model input).
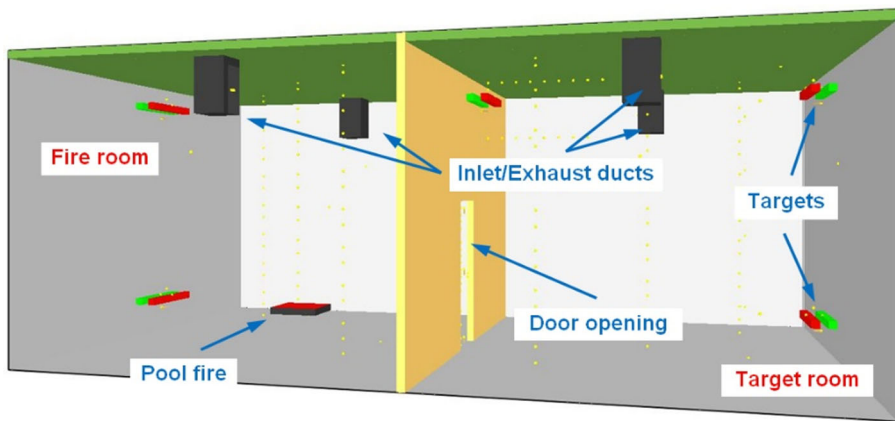

Figure 9. Input model for the numerical simulations with FDS.

Figure 9 shows the used (geometrical) input model applied in the simulations with FDS. The main elements of the input model are the pool fire, the target area with the liquid pool and the supply (inlet) and outlet (exhaust) air ducts connecting the rooms. Furthermore, the safety-relevant objects investigated can be recognized on the walls in the upper and lower area of the rooms. The analytical cables are shown in red and the real cables in green as a cuboid. Analytical cables are simple rods of pure PVC.

The input model contains about 300 measuring points/sensors, which were also used in the experiments. In this way, a comparison of almost all measured quantities between simulation calculation and experiment is possible.

Table 5 lists the analysed and evaluated quantities. Since the measurements were carried out at different heights, the "original" designations may be additionally provided with height information. Although pressure measurements were also carried out, the results were distorted or unrealistic on both the experimental and the simulation side, so this quantity was not analysed.

Objects and housing components were considered thermally and for a one-dimensional heat transfer calculation the thermal conductivities and the specific heat, if present, were considered temperature-dependent.

6. Application of Methodology

The computation results of the open simulations of the test are contrasted with the data measured during the test and analysed by an application of COMET for the time series as described above.

Table 6 summarizes the uncertainties of the analysed sensor groups and the number of evaluated and analysed sensors for this investigation. For the evaluation, a limitation of the data with respect to the evaluation criteria PEAK and NED was assumed: Data that were outside the interval [− 1; 1] for PEAK and outside the interval [0; 1] for NED were excluded from further evaluation. For PEAK or NED values, which lie within these intervals, it is assumed that there were no irregularities in the experimental execution or in the CFD fire simulation model.

6.1. PEAK/NED Analysis for All Sensor Groups (Weight D3, D4, D5)

Figure 10 shows the PEAK/NED analysis for all sensor groups named PEAK/NED—CO, CO_2, FLT, O_2, TCA, TCR, TG, and TP. The results are sampled for PRISME DOOR test 3, test 4 and test 5 (D3, D4, and D5 in the following) and illustrated with COMET.

Table 6
Uncertainties U_{NED} and U_{PEAK} for PRISME DOOR and Number of Used Sensors (using FDS 6.7.0)

| Analysed sensorgroups | U_{NED} [%] | U_{PEAK} [%] | Analysed sensors[a] | | | Data available |
			PRS_D3	PRS_D4	PRS_D5	This study
CO	26.1	15.5	5	5	5	15
CO_2	26.2	15.5	5	5	5	15
O_2	29.1	15.5	5	5	5	15
TCA	19.0	14.1	15	15	15	45
TCR	19.1	14.1	–	15	15	30
TG	18.6	14.1	140 (1)	139 (1)	124 (12)	403
TP	19.6	14.1	17 (5)	17	22	56
FLT	31.8	22.2	18 (4)	21 (1)	17 (5)	56
Sum total			205 (10)	222 (2)	208 (17)	635

[a]Values in brackets indicates the amount of data with inconsistencies (not further analysed values)
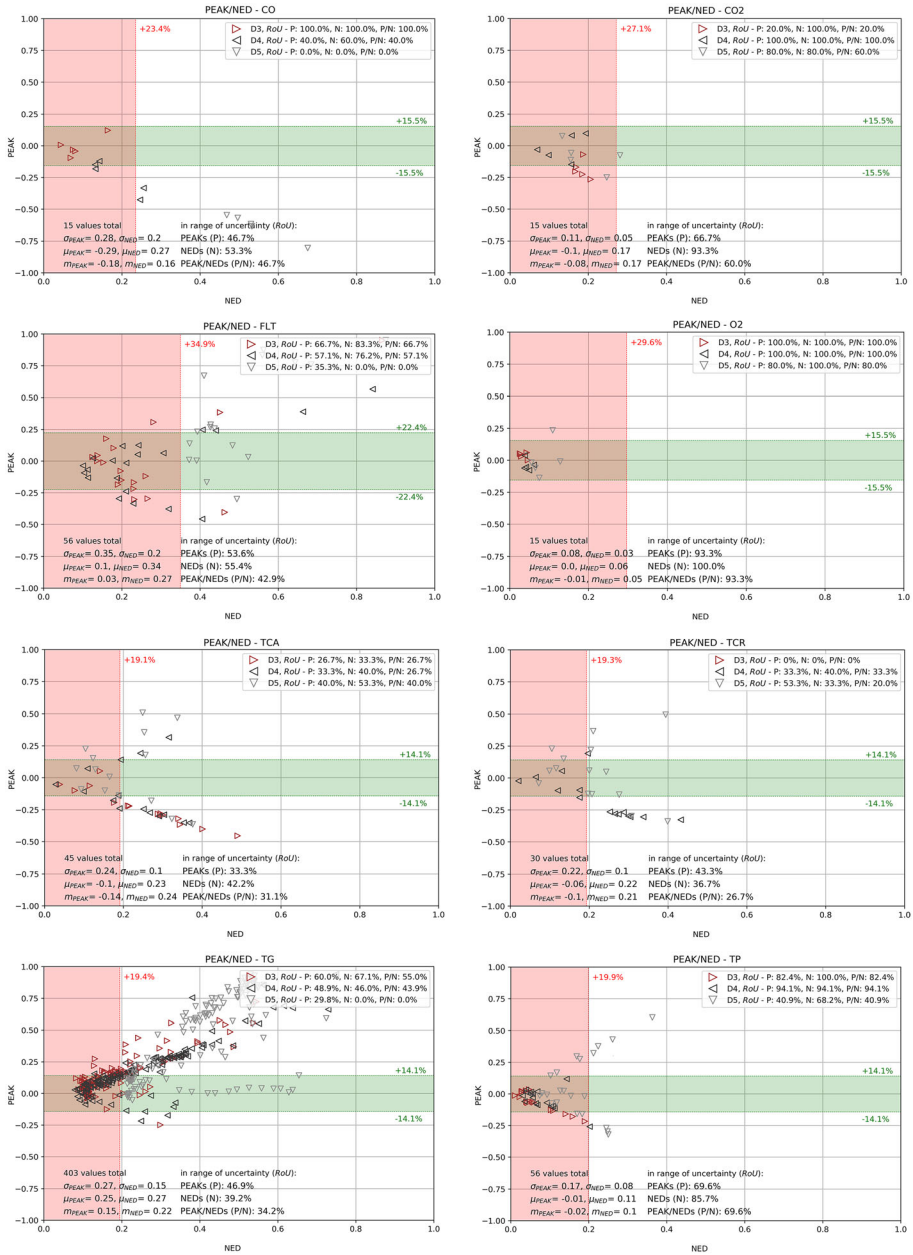
Figure 10. PEAK/NED plots of all sensor groups, sampled for PRISME DOOR test 3 (D3), test 4 (D4) and test 5 (D5).

PEAK values close to zero exist for all sensor groups. It can be seen, however, that NED values close to zero can be observed for certain sensor groups, only

Table 7
Results for PRISME DOOR D3, D4 and D5

Analysed sensor groups	Used sensors each group	Standard deviation		Mean value		In range of uncertainty (*RoU*) [%]		
	N	σ_{PEAK}	σ_{NED}	μ_{PEAK}	μ_{NED}	PEAKs	NEDs	PEAK/ NEDs
CO	15	0.28	0.20	− 0.29	0.27	46.7	53.3	46.7
CO$_2$	15	0.11	0.05	− 0.10	0.17	66.7	93.3	60.0
O$_2$	15	0.08	0.03	0.00	0.06	93.3	100	93.3
TCA	45	0.24	0.10	− 0.10	0.23	33.3	42.2	31.1
TCR	30	0.22	0.10	− 0.06	0.22	43.3	36.7	26.7
TG	403	0.27	0.15	0.25	0.27	46.9	39.2	34.2
TP	56	0.17	0.08	− 0.01	0.11	69.6	85.7	69.6
FLT	56	0.35	0.20	0.10	0.34	53.6	55.4	42.9

(e.g. TCR, TCA, TP), while for other sensor groups the NED values start at values around NED = 0.1 (e.g. TG, FLT).

Table 7 summarizes the results concerning this validation study (see Fig. 10 for details) from FDS version 6.7.0. For each sensor group, the table lists the standard deviation, the mean value and the proportion of PEAKs and NEDs and of PEAKs and NEDs (PEAK/NEDs) in percent [%] that are within the range of uncertainty.

Figure 11 summarizes the results concerning this validation study. For each sensor group the mean (NED, PEAK) value μ is illustrated. The figure also contains horizontal and vertical whiskers for each sensor group. The horizontal whiskers of length U_{NED} are used in a one-sided manner (to the left) since all NED values are larger than the optimal value zero by construction of this quantity. If these whiskers cross the ordinate, the model fits the data well (on average) for the corresponding sensor class in terms of NED value. Since PEAK values of sensors can be smaller or larger than the optimal value zero, we use two-sided whiskers of length U_{PEAK} here. If they cross the abscissa, the model fits the data well (on average) for the corresponding sensor class in terms of PEAK value.

The results illustrated in Fig. 11 are summarized in the following.

- For O2 and TP the forecast capability is best in comparison to all other sensor groups
- CO2 and TCR show very low mean PEAK, but higher NED values, which indicates good forecast capability of local aspects (peak of time series course), but lower forecast capability of global aspects
- Mean NED for CO, TG and FLT have larger values than for other sensor groups
- Mean PEAK values for TG are positive which indicates over-estimation of temperatures in the whole
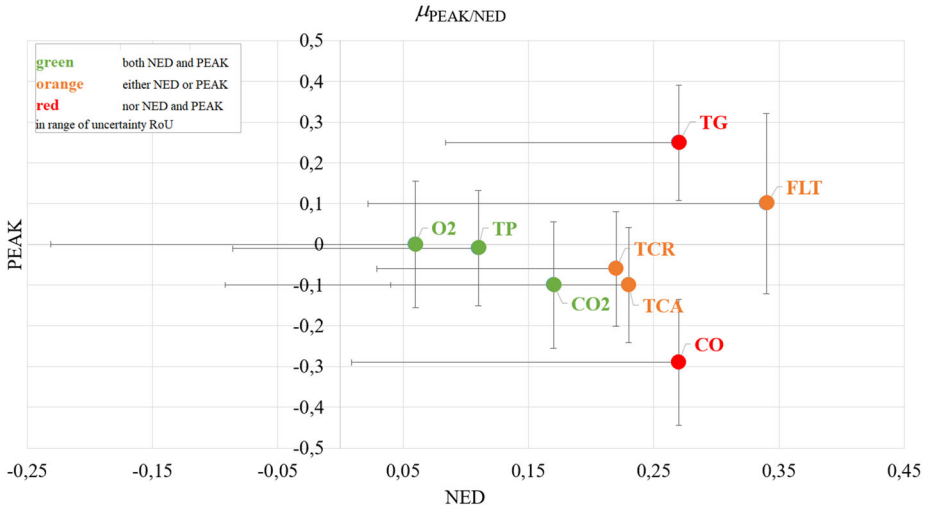- Mean PEAK values for CO are negative which indicates under-estimation of temperatures in the whole

Figure 11. PEAK and NED mean values μ for all sensor groups (FDS version 6.7.0).

6.2. Impact of FDS SUPPRESSION Parameter on Gas Temperature (TG) Modelling

In the next figures PEAK and NED results for individual sensors from sensor group TG (CC; FP; NW; SE; NE; SW see Fig. 6) are given, where SUPPRES-SION = TRUE was used for Fig. 12 in the model (FDS version 6.7.0), while SUPPRESSION = FALSE for Fig. 13 (for more details see [10] and Sect. 5).

RoU for PEAK (meaning fraction of data lying within *RoU* of PEAK) is about 47.1% versus 49.5%, *RoU* for NED is 52.9% versus 39.1% and for PEAK/NEDs the *RoU* is 39.3% versus 33.3%. Switching the model parameter SUPPRESSION from its default value (TRUE) to FALSE strongly improves the model adequacy with respect to NED while the PEAK values do not change significantly which confirms the benefit of the evaluation of the NED criterion in addition to the PEAK value. In detail the forecast capability for sensors NW (TG sensor north west) and SE (TG sensor south east) strongly improves with respect to NED, in case of using SUPPRESSION = FALSE in the model (and in connection with the investigated experiment).

6.3. PEAK/NED Analysis for Different Parameters for All Sensor Groups

In order to obtain information on how much of the observed uncertainty is due to the individual test and how much is due to the respective room, two groups are analyzed separately in the following. The groups are PRISME DOOR test (D3, D4 and D5, Fig. 14) and room (L1 and L2, Fig. 15). Since all sensor groups were evaluated together according to Table 7, a representation of the *RoU*s (green and red areas) for all physical quantities in one diagram is not feasible.

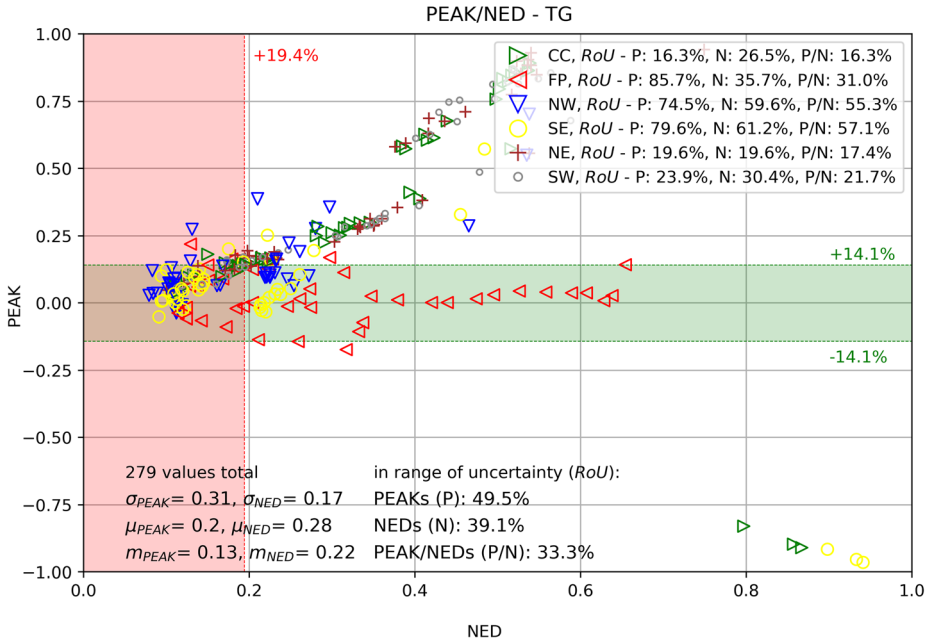The results illustrated in Fig. 14 are summarized in the following.

Figure 12. **PEAK and NED scatterplot, individual sensors TG, (CC; FP; NW; SE; NE; SW see Fig. 6) sampled for FDS version 6.7.0 (SUPPRESSION = TRUE).**

- Standard deviation (σ) mean (μ) and median (m) for PEAK and NED are significantly higher for D5 than for D3 and D4
- Mean (μ), and median (m) for PEAK are positive, which indicates over-estimation of the relevant sensors
- Some sensors show higher NED values (NED > 0.4), but PEAK values close to zero, coming from PRISME DOOR test 5, while the contrary situation does never occur

The results illustrated in Fig. 15 are summarized in the following.

- Standard deviation (σ) mean (μ) and median (m) for PEAK and NED are significantly higher for the door between room L1 and L2 (L1_L2) and room L2 than for room L1
- Mean (μ) and median (m) for PEAK are nearby zero for room L1 which indicates good forecast capability for PEAK values in room L1; in contrast, from higher NED values we can conjecture that the simulation forecast for the overall curve does not reach a convincing level. A deeper investigation using COMET for individual sensors or specific sensor groups as in Sect. 6.1 will provide more insights.
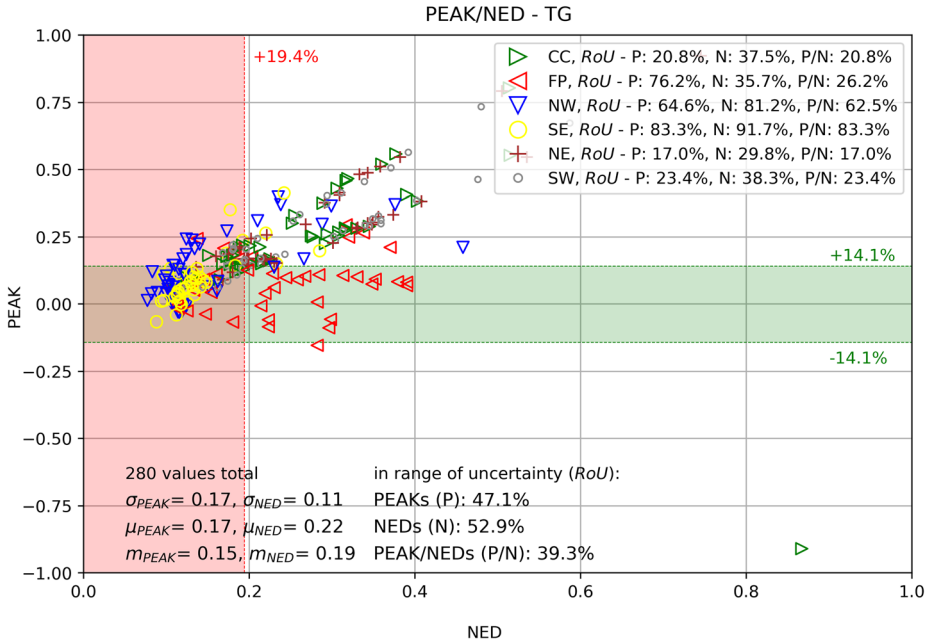
Figure 13. PEAK and NED scatterplot, individual sensors TG, (CC; FP; NW; SE; NE; SW see Fig. 6) sampled for FDS version 6.7.0 (SUPPRESSION = FALSE).

7. Conclusion and Outlook

For many validation purposes, the local PEAK criterion provides far-reaching information about the prognostic capability of CFD models. Nevertheless, the global NED method can provide additional information about the performance of sub-models used in a CFD model for fire simulation.

We introduced the novel two-dimensional tool COMET for uncertainty quantification with respect to both of these frequently used local and global metrics in parallel. To obtain the boundaries of the *RoU* for NED we applied similar techniques as described in NUREG [5, 6] for the corresponding uncertainty quantification for PEAK. A set of three experiments coming from OECD/NEA PRISME DOOR test 3, test 4 and test 5 was used to demonstrate the new approach. The results emphasize that an investigation of both metrics is inevitable.

In particular, it is shown that the gas temperatures calculated with the model are often higher than those determined experimentally, i.e. they show predominantly positive PEAK values. In comparison to the first room, larger deviations of statistical values such as standard deviations, means and medians of the (NED, PEAK) values are obtained for the studied sensors in the adjacent room. It can be assumed that the transport of the physical quantities from the first room through the door into the second room cannot be reproduced with the same prediction accuracy in the model with the default values set here.
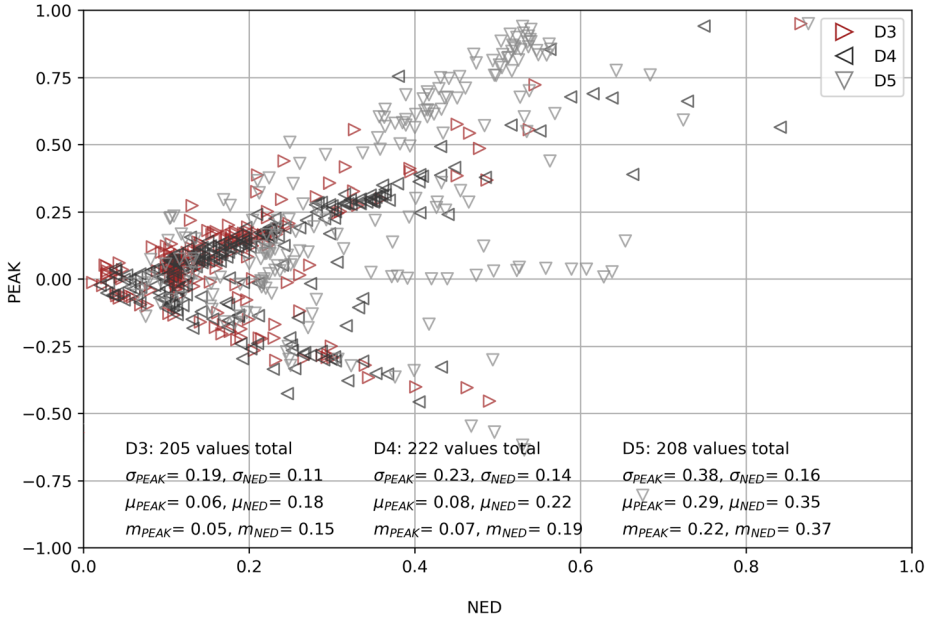
Figure 14. PEAK and NED scatterplot, all sensor groups, sampled for PRISME DOOR test D3, D4 and D5.
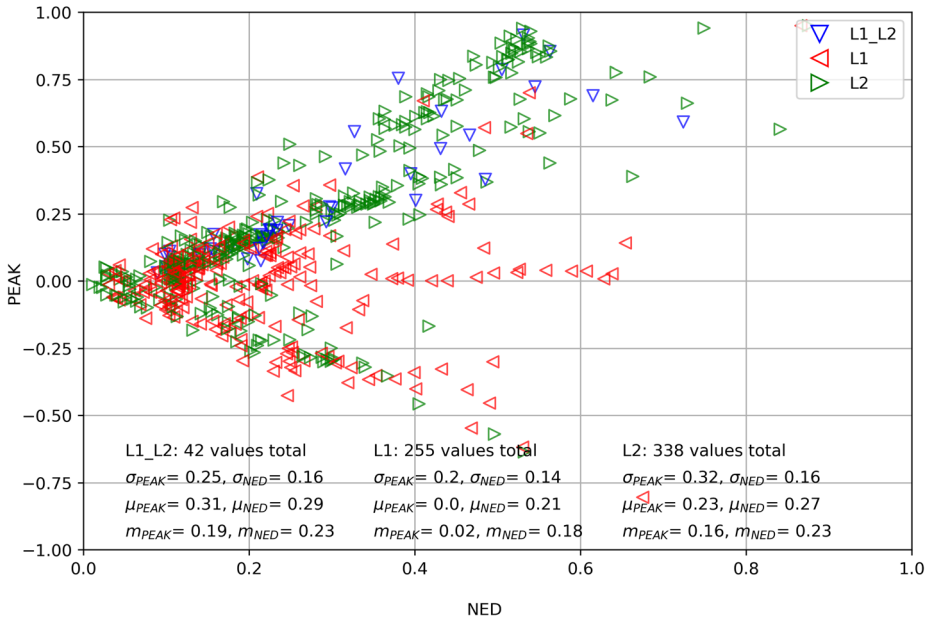

Figure 15. PEAK and NED scatterplot, all sensor groups, sampled for room 1 (L1) and room 2 (L2) and Door (L1_L2) between both rooms.

Carbon dioxide and temperatures of real cables shows very low mean PEAK, but higher NED values, which indicates good forecast capability of local aspects (peak of time series course), but lower forecast capability of global aspects (course at every time point).

Summing up, using COMET instead of simply PEAK can provide substantial additional information about model performances. Still, to check the performance of a model, it may not be useful to evaluate all possible sensors with the COMET method. For example, for the evaluation of temperatures in the plume, it is certainly useful to check the ability of the model for the prediction in the close range of the flames. Moreover, several additional issues have to be tackled before the new methodology can serve as a guide for regulatory compliance. For instance, it might be useful to restrict the time interval for which NED is applied since the ramp-up period of fires is known to be rather volatile, and the influence of this ramp-up phase on intervention measures is rather negligible. All these kinds of considerations go far beyond the scope of the paper and are left for further research.

Acknowledgements

Funding

Open Access

Appendix: Quantifying Model Uncertainties with Comet

In this section we describe the methodology used to compare the model and measurement results for the NED criterion in Sect. 2. In doing so, we intend to establish a model evaluation approach which is comparable to the one described in Sect. 1.4.2 of [3] for the peak criterion. Note that the ordinary L2 metric (without normalizing denominator) can be treated similarly.

Let $M_1, M_2, \ldots, M_T$ denote the model predictions and $E_1, E_2, \ldots, E_T$ the experimental measurements at time points $t = 1, 2, \ldots, T$. We assume baseline values M_0 and E_0 at the start of the experiment and are interested in the deviations from these baseline values over time, denoted by $\Delta M_t = M_t - M_0$ and $\Delta E_t = E_t - E_0$.

Our assumption is that there exists a function $m : [0, 1] \to \mathbb{R}$ which describes the "true" (unobserved) development of the quantity of interest over time, such that at time points $1, 2, \ldots, T$ the true values are $m(\frac{1}{T}), m(\frac{2}{T}), \ldots, m(\frac{T}{T})$. In line with [5], we further assume that the ΔE_t's as well as the ΔM_t's are normally distributed (see Fig. 16).

While only one point in time has to be considered for the PEAK criterion in [3], the situation is more delicate here, since we have to incorporate time dependencies in the NED setup. Suppose for simplicity that the observed values are given by

$$\Delta M_t = m\left(\frac{t}{T}\right)(1 + \mu),$$

$$\Delta E_t = m\left(\frac{t}{T}\right)(1 + \eta_t),$$

for $t = 1, 2, \ldots, T$ where μ and η_t are independent random variables with normal distributions. To be precise we assume $\mu \sim \mathcal{N}(0, \tilde{u}_M^2)$ and $\eta_t \sim \mathcal{N}(0, \tilde{u}_E^2)$ with $\tilde{u}_M^2$ and $\tilde{u}_E^2$ denoting the relative uncertainties considered in [3]. Hence, μ describes the effect of the model input uncertainty, that is, the amount by which the model predictions deviate from the true values. Because of the multiplicative structure we have

$$\Delta M_t = m\left(\frac{t}{T}\right) + m\left(\frac{t}{T}\right)\mu \sim \mathcal{N}\left(m\left(\frac{t}{T}\right), m^2\left(\frac{t}{T}\right)\tilde{u}_M^2\right),$$

so the effect is proportional to the size of the true values. This reflects the fact that in [5] the quantity $\tilde{u}_M = \frac{u_M}{\Delta M}$ also describes uncertainty relative to the size of the model prediction.

Our goal is to approximate the variance of the normalized Euclidean distance between the time series $\Delta M := (\Delta M_1, \ldots, \Delta M_T)$ and $\Delta E := (\Delta E_1, \ldots, \Delta E_T)$, given by
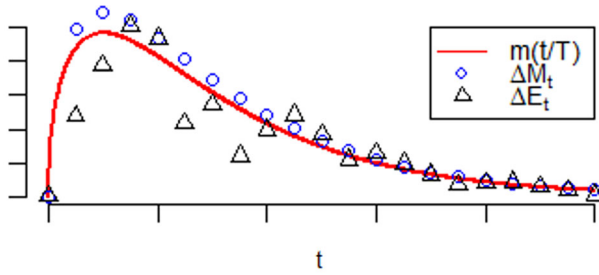
Figure 16. True development of a certain quantity (red solid line), and deviations from the baseline's values, for model prediction ΔM_t and experimental measurements ΔE_t over time.

$$\text{NED} = \sqrt{\frac{\sum_{t=1}^{T}(\Delta M_t - \Delta E_t)^2}{\sum_{t=1}^{T}(\Delta E_t)^2}}.$$

To this end, we interpret NED as a function of random quantities, to be precise.
$\text{NED} = f(\Delta M_1, \ldots, \Delta M_T, \Delta E_1, \ldots, \Delta E_T)$, where

$$f(x_1, \ldots, x_T, y_1, \ldots, y_T) := \sqrt{\frac{\sum_{t=1}^{T}(x_t - y_t)^2}{\sum_{t=1}^{T} y_t^2}}.$$

By a Taylor expansion of first order at point $a_0 = \left(m\left(\frac{1}{T}\right) + \delta, \ldots, m\left(\frac{T}{T}\right) + \delta, m\left(\frac{1}{T}\right), \ldots, m\left(\frac{T}{T}\right)\right)$, for any $\delta > 0$,[4] we have the approximation

$$\text{NED} \approx \sqrt{\frac{T\delta^2}{\sum_{t=1}^{T} m(t/T)^2}} + [(\Delta M, \Delta E) - a_0]\nabla f(a_0), \tag{6}$$

where remainder terms of higher order are neglected. Here, $\nabla f(a_0)$ denotes the gradient of f at point a_0, and derivation of the components yields

$$\nabla f(a_0) = \begin{pmatrix} z \\ \vdots \\ z \\ w_1 \\ \vdots \\ w_T \end{pmatrix},$$

with

---

[4] Note that $\delta \equiv 0$ is not feasible, since the function $x \rightarrow \sqrt{x}$ is not differentiable at 0.

$$z = \frac{1}{\sqrt{T \sum_{t=1}^{T} m(t/T)^2}}, w_t = -\frac{1}{\sqrt{T \sum_{t=1}^{T} m(t/T)^2}} - \frac{\sqrt{T}\delta m\left(\frac{t}{T}\right)}{\left(\sum_{t=1}^{T} m(t/T)^2\right)^{\frac{3}{2}}}.$$

On the other hand, we have

$$(\Delta M, \Delta E) - a_0 = \left(m\left(\frac{1}{T}\right)\mu - \delta, \ldots, m\left(\frac{T}{T}\right)\mu - \delta, m\left(\frac{1}{T}\right)\eta_1, \ldots, m\left(\frac{T}{T}\right)\eta_T\right).$$

Since the first term on the right-hand side of (6) is constant, we have for the variance of NED:

$$Var(\text{NED}) \approx Var([(\Delta M, \Delta E) - a_0]\nabla f(a_0))$$
$$= Var\left(\frac{\sum_{t=1}^{T} m\left(\frac{t}{T}\right)}{\sqrt{T \sum_{t=1}^{T} m(t/T)^2}}\mu\right) + Var\left(\frac{\sum_{t=1}^{T} m\left(\frac{t}{T}\right)\eta_t}{\sqrt{T \sum_{t=1}^{T} m(t/T)^2}} + \frac{\sqrt{T}\delta \sum_{t=1}^{T} m(t/T)^2\eta_t}{\left(\sum_{t=1}^{T} m(t/T)^2\right)^{\frac{3}{2}}}\right).$$

So far, we did not specify $\delta > 0$. Assuming that experimental and numerical data points are not too far away from each other at each time point (both close to the true underlying curve m), the Taylor approximation can be expected to work well for δ close to zero. For simplification, we further approximate

$$Var(\text{NED}) \approx Var\left(\frac{\sum_{t=1}^{T} m\left(\frac{t}{T}\right)}{\sqrt{T \sum_{t=1}^{T} m(t/T)^2}}\mu\right) + Var\left(\frac{\sum_{t=1}^{T} m\left(\frac{t}{T}\right)\eta_t}{\sqrt{T \sum_{t=1}^{T} m(t/T)^2}}\right). \tag{7}$$
$$= \frac{\left(\sum_{t=1}^{T} m(t/T)\right)^2}{\sum_{t=1}^{T} m(t/T)^2} \cdot \frac{\tilde{u}_M^2}{T} + \frac{\tilde{u}_E^2}{T}$$

When approximating *Var*(*N*ED) based on observed data in practice, the factor containing the unknown function m has to be estimated. We propose an estimation of numerator and denominator based on the observed experimental measurements ΔE_t as follows:

$$\hat{Var}(\text{NED}) = \frac{\left(\sum_{t=1}^{T} \Delta E_t\right)^2}{\sum_{t=1}^{T} \Delta E_t \Delta E_{t+1}} \cdot \frac{\tilde{u}_M^2}{T} + \frac{\tilde{u}_E^2}{T}. \tag{8}$$

This is formula (4) from Sect. 3. To show that this estimator is consistent we assume differentiability of the function m. Then the first summand on the right-hand side of (7) can be written via Riemann sums and it holds

$$\frac{\left(\sum_{t=1}^{T} m(t/T)\right)^2}{\sum_{t=1}^{T} m(t/T)^2} \cdot \frac{\tilde{u}_M^2}{T} = \frac{\left(\frac{1}{T}\sum_{t=1}^{T} m(t/T)\right)^2}{\frac{1}{T}\sum_{t=1}^{T} m(t/T)^2} \cdot \tilde{u}_M^2 \rightarrow \frac{\left(\int_0^1 m(x)dx\right)^2}{\int_0^1 m(x)^2 dx} \cdot \tilde{u}_M^2, \tag{9}$$

as $T \to \infty$. On the other hand, the first summand on the right-hand side of (8) can be written as

$$\frac{(\sum_{t=1}^{T} \Delta E_t)^2}{\sum_{t=1}^{T-1} \Delta E_t \Delta E_{t+1}} \cdot \frac{\tilde{u}_M^2}{T} = \frac{(\frac{1}{T}\sum_{t=1}^{T} \Delta E_t)^2 \cdot \mathfrak{u}_M^2}{\frac{1}{T}\sum_{t=1}^{T-1} \Delta E_t \Delta E_{t+1}} =: \frac{Z_T^2 \tilde{u}_M^2}{N_T}. \tag{10}$$

For Z_T it holds due to independence of η_t and η_s for $t \neq s$:

$$EZ_T = \frac{1}{T}\sum_{t=1}^{T} m\left(\frac{t}{T}\right) \to \int_0^1 m(x)dx$$

and

$$Var(Z_T) = \frac{1}{T^2}\sum_{t=1}^{T} m(t/T)^2 \tilde{u}_E^2 \to 0$$

due to boundedness of the function m. Hence Z_T^2 converges towards $\left(\int_0^1 m(x)dx\right)^2$ in probability. For the denominator we have

$$EN_T = \frac{1}{T}\sum_{t=1}^{T} m(t/T)^2 + \frac{1}{T}\sum_{t=1}^{T} m(t/T)\left[m\left(\frac{t+1}{T}\right) - m\left(\frac{t}{T}\right)\right] \to \int_0^1 m(x)^2 dx,$$

as well as

$$Var(N_T) = \frac{1}{T^2}\sum_{t=1}^{T-1} Cov(\Delta E_t \Delta E_{t+1}, \Delta E_t \Delta E_{t+1})$$
$$+ \frac{2}{T^2}\sum_{t=1}^{T-2} Cov(\Delta E_t \Delta E_{t+1}, \Delta E_{t+1} \Delta E_{t+2})$$
$$\to 0,$$

which implies that N_T converges in probability towards $\int_0^1 m(x)^2 dx$ and (10) converges towards the limit in (9). Therefore, (8) is an asymptotically valid approximation for (9).

References

1. Van Hees P (2013) Validation and verification of fire models for fire safety engineering, 9th Asia—Oceania Symposium on fire science and technology. Procedia Eng 82:154–168
2. ISO 16730:2008–07, Fire safety engineering—assessment, verification and validation of calculation methods, International Standard Organization, 2008
3. ASTM E1355–12, Standard Guide for evaluating the predictive capability of deterministic fire models, ASTM International, 2018

4. Peacock RD, Reneke PA, Davis WD, Jones WW (1999) Quantifying fire model evaluation using functional analysis. Fire Saf J 33:167–184
5. NUREG 1824–2: Verification & validation of selected fire models for nuclear power plant applications, Volume 2: experimental uncertainty, NUREG-1824 EPRI 1011999, Final Report, Table 6–8, May 2007
6. NUREG 1824-SUPPLEMENT 1: Verification & validation of selected fire models for nuclear power plant applications, Supplement 1, NUREG-1824 supplement 1, EPRI 3002002182, Final Report, Table 3–4, Nov 2016
7. Audouin L et al (2011) Quantifying differences between computational results and measurements in the case of a large-scale well-confined fire scenario. Nuclear Eng Design 241(1):18–31
8. Mc Grattan K, Toman B (2011) Quantifying the predictive uncertainty of complex numerical models. Metrologia 48:173–180. https://doi.org/10.1088/0026-1394/48/3/011
9. Le Saux W, Pretrel H, Lucchesi C, Guillou P (2008) Experimental study of the fire mass loss rate in confined and mechanically ventilated multi-room scenarios, Fire Safety Science proceedings of the Ninth International Symposium
10. Le Saux W (2008) PRISME DOOR—analysis of the test results, presentation during the 5th meeting of the OECD PRISME Project, Marseille (France)
11. Riese O, Hohm V, Shiping L (2011) Untersuchung der Prognosefähigkeit von deterministischen Brandsimulationsmodellen—Anwendung PRISME LEAK, Ernst & Sohn, ISSN 1437–0980, A 1879, Ausgabe 6, 33. Jahrgang, Bauphysik
12. Hosser D, Hohm V, Riese O (2009) EMVANEMED—a methodology to compare and evaluate numerical results with experimental data—application to OECD PRISME DOOR test PRS_DI_D3. In: Proceedings of SMiRT 20, 11th International Post Conference Seminar on "Fire Safety in nuclear power plants and installations, August 17–19, 2009, Helsinki, Finnland
13. Riese O, Siemon M (2014) Untersuchung der Prognosefähigkeit von deterministischen Brandsimulationsmodellen—Anwendung PRISME DOOR, Ernst & Sohn, ISSN 1437–0980, A 1879, Ausgabe 4, 36. Jahrgang, Bauphysik
14. McGrattan KB, Hostikka S, Floyd J, McDermott R, Vanella M (2021) Fire dynamics simulator (Version 6), User's Guide, NIST Special Publication 1019–1. 6th edn. National Institute of Standards and Technology (NIST), Gaithersburg, Maryland
15. Tewarson T (2002) Generation of heat and chemical compounds in fires. In: SFPE handbook of fire protection engineering, 3rd edn. Section three hazard calculations, National Fire Protection Association (NFPA), p 3–134, Table 3–4.19