Computer Assisted Methods in Engineering and Science

eISSN: 2956-5839, ISSN: 2299-3649

INDEX 324906

Journal published under the auspices of ECCOMAS

Vol. 33
No. 2
2026

Institute of Fundamental Technological Research
Polish Academy of Sciences

COMPUTER ASSISTED METHODS in ENGINEERING and SCIENCE

eISSN 2956-5839 ISSN 2299-3649
(Instytut Podstawowych Problemów Techniki Polskiej Akademii Nauk
– Institute of Fundamental Technological Research, Polish Academy of Sciences)

Aims and Scope

Computer Assisted Methods in Engineering and Science (CAMES) is a refereed international journal, published quarterly, providing a scientific exchange forum and an authoritative source of information in the field of broadly understood computational engineering and applied sciences. The objective of the journal is to support researchers and practitioners by offering them the means facilitating access to the newest research results reported by leading experts in the field, publication of own contributions and dissemination of information relevant to the scope of the journal which includes three main categories:

- Contributions presenting new research methods of mathematical modeling and computer simulations in engineering and applied sciences, including traditional areas such as solid and structural mechanics, material science, fluid dynamics, acoustics and electromagnetics but going beyond them to account for application relevant issues in physics, chemistry, biology and mathematics, scientific computing, large scale optimization, intelligent systems as well as in multi-scale and multi-physics problems.
- Articles describing novel applications of computational techniques supporting engineering practice and education in areas like mechanical, aerospace, civil, naval, software, chemical and architectural engineering, materials science as well as demonstrations of their practical use in solving real life problems.
- State-of-the-art tutorials, providing the readership with a guidance on important research directions as observed in the current world literature on computer assisted methods in engineering and science.

The journal also publishes book reviews and information on activities of the European Community on Computational Methods in Applied Sciences (ECCOMAS).

Editorial Board

Editorial Office

Institute of Fundamental Technological Research, PAS
Pawińskiego 5B, 02-106 Warsaw, Poland
E-mail: cames@ippt.pan.pl, URL: https://cames.ippt.pan.pl

Typesetting in LaTeX: Katarzyna Jezierska Graphic design cover: Mirosław Dziewoński

Institute of Fundamental Technological Research
Polish Academy of Sciences

COMPUTER ASSISTED METHODS in ENGINEERING and SCIENCE

Volume 33
Issue 2

Warsaw 2026

Advisory Editorial Board of CAMES

Advisory Editorial Board consists of members of Managing Board of ECCOMAS with co-opted members and chairmen of Technical Committees of ECCOMAS.

Computer Assisted Methods in Engineering and Science
33(2): 85–99, 2026, https://doi.org/10.24423/cames.2026.1786

Algorithm for Processing Broadband Dielectric Spectroscopy Data of Heterogeneous Materials

Alexander Tonkoshkur[1)], Andrey Shcherbak[1)], Alexander Lyashkov[2)*]

[1)] *Department of Electronic Computing Machinery, Oles Honchar Dnipro National University, Dnipro, Ukraine*
[2)] *Department of Applied Radiophysics, Electronics and Nanomaterials, Oles Honchar Dnipro National University, Dnipro, Ukraine*

[*] *Corresponding Author: alexdnu@ukr.net*

This work focuses on the development of computational techniques for processing dielectric measurement data, particularly those from broadband dielectric spectroscopy. A novel approach is formulated and tested for software implementation, enabling the selection of parameter sets for known dispersion models used in analyzing complex dielectric spectra of heterogeneous materials.
The proposed algorithm divides the wide frequency range of measurements into shorter subranges corresponding to distinct observed dispersion regions. These regions are identified using one (or more known phenomenological relaxation models via least-squares methods (LSMs), deconvolution, and other techniques.
Testing of the algorithm on the frequency-dependent complex dielectric permittivity of varistor ceramic materials demonstrated satisfactory accuracy and physical consistency of the results. These findings highlight the efficiency and potential of the proposed approach.

Keywords: dielectric spectroscopy, data processing, algorithm, complex spectrum, ZnO ceramics.

1. Introduction

In materials science, dielectric spectroscopy has long been a fundamental investigative tool for understanding the electrical and chemical structure, as well as the physical properties, of heterogeneous substances and objects [1, 2]. This method is extensively used in the development and application of various heterogeneous electronic materials (ceramics, glasses, composites) [3–5], the synthesis of new materials based on organic polymers [6–8], and the study of biological and medical objects [9–11].

Modern measurement systems allow for highly accurate investigation of the dielectric properties of various physical media across a wide range of frequencies and temperatures. Fully automated spectrometers have two significant implications [12]:

– the need to handle vast amounts of data,

– the ability to model using sums of relaxation functions.

In broadband dielectric spectroscopy, the complex dielectric permittivity (CDP) is determined at specific measurement frequencies. As the spectra are discrete datasets, only numerical approximations can be used. The accuracy of the selected mathematical methods significantly influences the assessment of the models' suitability [13]. Therefore, the processing and analysis of experimental data remain pressing tasks to this day.

The development of universal algorithms and software for processing and analyzing dielectric measurement data is typically based on well-known models of frequency dispersion of dielectric permittivity, such as the Debye formula or its phenomenological generalizations, including the Cole–Cole, Davidson–Cole, and Havriliak–Negami models [1, 14–16].

A key feature of these models is their versatility, which allows for multipurpose applications, including:

– analytical representation of experimental data for more accurate descriptions,

– interpretation of the parameters of these relationships and their dependencies on various factors, based on specific physical models of polarization processes in various materials and objects,

– derivation of information about general parameters of heterogeneity in dielectric relaxation processes, such as the distribution of relaxation times and the shape of this distribution.

The methodology for processing and analysis based on generalized models using sums of relaxation functions, rooted in the additivity of polarization processes, represents a significant advancement in the dielectric spectroscopy method [1, 2, 17, 18]. Nevertheless, the individual characteristics of the relaxation processes in different materials and objects remain a significant factor hindering the development of universal algorithms and the implementation of computer technologies in this area. A potential solution may involve grouping materials and objects with similar polarization process sets and developing algorithms and software templates tailored to these groups.

It should be noted that the most well-known mathematical procedure for selecting a model of dielectric relaxation processes is the least-squares method (LSM), which enables the determination of all model parameters [19–21]. However, for complex spectra, this entails minimizing a multi-variable function,

which is a sum of several relaxation functions. Each relaxation function may involve two (Debye model) to four (Havriliak–Negami model) adjustable parameters and may not be unimodal. Additional challenges arise when processing overlapping dispersion regions.

To address the complexities of processing and analyzing intricate dielectric spectra, a decomposition-based approach shows potential:

- dividing the initial wide-frequency experimental spectrum into shorter subranges where dispersion dependencies are present. These dependencies might be represented as declining (low-frequency dispersion regions) or ascending (high-frequency dispersion regions) functions, which may not be fully captured in the available data;
- sequentially processing these shorter spectrum segments, beginning with the extremes and considering their impact on adjacent dispersion regions;
- utilizing theoretical insights into relaxation processes in the material or object to guide the selection of these frequency subranges and initial approximations for the parameters of the chosen approximating models.

This approach enables the application of the LSM for analyzing each specific frequency segment, the use of deconvolution procedures to determine relaxation times in overlapping dispersion regions [22–24], and the application of Kramers–Kronig relations to account for through-conductivity (via direct current), among other established methods [25–30].

In this study, we explore the application of this approach to the development of an algorithm for processing and analyzing dielectric spectra of heterogeneous materials. For validation, dielectric spectra of high-voltage varistor zinc oxide (ZnO) ceramics were utilized as experimental data over a broad frequency range.

2. Broadband dielectric spectrum model and data processing algorithm

Based on established concepts of the linear additive nature of contributions from individual dielectric relaxation processes to a complex broadband spectrum [1, 2, 17, 18], the model expression for the entire frequency range can be represented as a sum of corresponding elementary models. Within the approach discussed here – dividing the initial wide frequency range of the dielectric spectrum into shorter subranges – the generalized model for the frequency dispersion of the CDP over the entire range can be expressed as:

$$\epsilon^*\left(f, M\right) = \epsilon_h + \sum_{i=1}^{R}\sum_{j=1}^{N_i}\left[\epsilon^*_{i,j}\left(f, M_i^{\langle j\rangle}\right) - \epsilon_{i.j}^{(h)}\right], \tag{1}$$

where R is the number of subranges; N_i is the number of elementary models in the i-th subrange; $\epsilon^*_{i,j}\left(f, M_i^{\langle j\rangle}\right)$ and $\epsilon^{(h)}_{i.j}$ are the j-th elementary model for the frequency dispersion of CDP in the ith subrange and its high-frequency (post-dispersion) value for the real part, respectively; i ($i = 1, ..., R$) and j ($j = 1, ..., N_i$) are its indices; f is the frequency; $M_i^{\langle j\rangle}$ is the set of parameters for the elementary model (including the relaxation time $\tau_{i,j}$, the difference $\Delta_{i,j} = \epsilon^{(l)}_{i.j} - \epsilon^{(h)}_{i.j}$, where $\epsilon^{(h)}_{i.j}$ is the high-frequency value and $\epsilon^{(l)}_{i.j}$ is the low-frequency value of the real part of the CDP and other parameters. Thus, $M_i^{\langle j\rangle} = \{\tau_{i,j}, \Delta_{i,j}, \alpha_{i,j}, \beta_{i,j}$ (where, depending on the selected mode, the parameters α and / or β may be absent); M is the set of parameters for the generalized model; ϵ_h is the high-frequency value of the real part of the CDP for the highest-frequency dispersion region.

The mathematical implementation of the considered approach for computer processing and preliminary analysis of dielectric spectra can be represented by the following algorithm:

1. Determination of clearly distinguishable separate dispersion regions in the initial experimental frequency dependence of the imaginary part of the CDP, i.e., identification of R frequency subranges between the minima of the dielectric loss coefficient ε''.
2. Selection of models and determination of initial approximate parameter values for the extreme left and right subranges ($i = 1$ and $i = R$) directly from the processed experimental dependencies.

   In particular, using the experimental values of the maximum of the imaginary part of the CDP $\varepsilon''(f_{\max})$, for the considered subrange and the corresponding frequency $f_{\max}$, it is possible to estimate the initial approximate values of the relaxation time $\tau^{(0)} = 1/(2\pi \cdot f_{\max_1})$ and the absolute magnitude of the dispersion amplitude of the real part of the CDP $\Delta^{(0)} =\geq 2 \cdot \varepsilon''(f_{\max_1})$ (the indices of the dispersion region i and j are omitted for simplicity).

   It should be noted that in some cases an appropriate value for $\Delta^{(0)}$ can also be directly determined from the experimental frequency dependence of the real part of the CDP as $\Delta^{(0)} = \varepsilon_{l0} - \varepsilon_{h0}$, where ε_{l0} and ε_{h0} are the low-frequency (static) and high-frequency values of the real part of the CDP, respectively.

   In the case of broadened (overlapping) dispersion regions, it is advisable to apply deconvolution methods [22–24] to separate them and determine the initial values of the dielectric relaxation times. If complete data on a dispersion region are not available, it is permissible to use prior values for the model parameters based on known literature data and

physical concepts of the relaxation process under consideration. The remaining parameters are chosen arbitrarily within the range of permissible values.

3. Approximation of the extreme left ($i = 1$) and right ($i = R-1$) dispersion regions of the selected models using the LSM and determination of more accurate parameter values for the models, which ensure the minimal absolute error of the LSM target functions [31] in these ranges, is performed as follows:

$$\Phi_i\left(f_k, M_i^{\langle 1\rangle}, M_i^{\langle 2\rangle}, ..., M_i^{\langle N_i\rangle}\right) = \sum_{k=1}^{K_i}\left\{\log\varepsilon_k'' - \log\left[\sum_{j=1}^{N_i}\epsilon_{i,j}''\left(f_k, M_i^{\langle j\rangle}\right)\right]\right\}^2 \to \min, \quad (2)$$

where $\epsilon_{i,j}''\left(f_k, M_i^{\langle j\rangle}\right) = -\mathrm{Im}\left[\epsilon_{i,j}^*\left(f_k, M_i^{\langle j\rangle}\right)\right]$; K_i is the number of experimental points in the i-th subrange ($i = 1$ or R). The value $N_i \neq 1$ in cases of overlapping dispersion regions in a single subrange and when the deconvolution procedure is applied for their separation.

4. Correction of the experimental dependencies $\varepsilon_k''(f)$ for internal subranges ($i = 2, 3, ..., R-1$) is performed by subtracting from them the corresponding calculated values of the imaginary part of the CDP, obtained by extrapolating analytical expressions of models already identified in previously processed dispersion regions:

$$\overline{\varepsilon}_k''\big|_{i=2...N-1} = \varepsilon_k'' - \left[\sum_{j=1}^{N_1}\epsilon_{i,j}''\left(f_k, M_i^{\langle j\rangle}\right) + \sum_{j=1}^{N_R}\epsilon_{i,j}''\left(f_k, M_i^{\langle j\rangle}\right)\right]. \quad (3)$$

5. Sequential repetition of the operations described in the previous two steps using the corrected data allows determination of the parameter values of the selected models for internal subranges (i.e., approximation of the extreme left and right subranges $i = 2$ and $i = R-2$, followed by $i = 3$ and $i = R-3$, and so on).

6. Final refinement of the parameters of the generalized additive model (1) for the entire range of the initial experimental broadband dielectric spectrum with K points is performed by minimizing the LSM target function as follows:

$$\Phi\left(f_k, M\right) = \sum_{k=1}^{K}\left\{\log\varepsilon_k'' - \log\left[\epsilon''\left(f_k, M\right)\right]\right\}^2,$$

where $\epsilon''\left(f, M\right) = -\mathrm{Im}\left\{\epsilon^*\left(f, M\right)\right\}$.

7. Error assessment of the approximation for the frequency dependencies of the real and imaginary parts of the CDP can be performed using the ratio of the root-mean-square deviation to the mean value of the experimental dependency (the coefficient of variation of the approximation):

$$v^{(x)} = \sqrt{\left\{\sum_{k=1}^{K}\left[\varepsilon_k^{(x)} - \epsilon^{(x)}\left(f_k, M\right)\right]^2 \Big/ (K-1)\right\}} \Big/ \sum_{k=1}^{K} \varepsilon_k^{(x)}/K, \qquad (4)$$

where (x) corresponds to 'double prime' $(\prime\prime)$ for the imaginary part and 'prime' $(\prime)$ for the real part of the CDP, and $\epsilon'(f) = \mathrm{Re}\left\{\epsilon^*\left(f, M\right)\right\}$.

3. Processing of broadband dielectric spectroscopy data for ZnO-based semiconductor ceramics

This section examines the implementation of the formulated approach for processing the complex broadband dielectric spectrum of a heterogeneous semiconductor material – varistor ZnO-based ceramics – as a demonstration of its applicability.

This material's structure is characterized by a dielectric interlayer [32, 33]. In the frequency dependencies of the real ε' and imaginary ε'' parts of the CDP, several dispersion regions are observed, typical for heterogeneous systems of this type [34–37]:

- a maximum (or decrease), at low and infralow frequencies $f < 10^2$ Hz region (I), associated with the capture of electrons from ZnO crystallites into their 'slow' surface electronic states;
- a broadened maximum ε'' at intermediate frequencies (10^4 to 10^6 Hz), interpreted as the superposition of two relaxation processes: hopping of weakly bound charged particles (electrons) in the intracrystalline phase (II) and recharging of bulk trap levels in ZnO crystallites (III).
- high-frequency behavior: a decrease in ε' and an increase in ε'' in the range $f > 10^7$ Hz region (IV), caused by Maxwell–Wagner separation of free charges in ZnO crystallites.

The frequency dependence of $\varepsilon''(f)$ is shown in Fig. 1.

As seen in Fig. 1, the entire frequency range of the experimental dielectric spectrum can be divided into three subranges: low and infralow frequencies (dispersion region I), intermediate frequencies (regions II and III), and high frequencies (region IV). The procedure for applying the formulated algorithm to each of these subranges is presented subsequent.

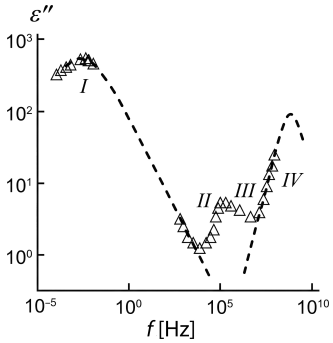

FIG. 1. Experimental frequency dependencies of the imaginary part of the CDP of ZnO-based ceramics (triangles) and the results of approximation for its low- and high-frequency regions using the Cole–Cole and Debye models (dashed lines).

3.1. SUBRANGE OF LOW AND INFRALOW FREQUENCIES

When determining a model for the relatively broad dispersion region *1*, its association with the recharging of surface electronic states, which exhibit significant statistical variability in physical parameters, must be considered. The Cole–Cole model is chosen as the most appropriate to describe the entire range (case – $N_1 = 1$) of region *1*. In the previously established notation, it has the following form:

$$\epsilon^*_{1,1}\left(f, M_1^{\langle 1\rangle}\right) = \epsilon^{(h)}_{1.1} + \frac{\Delta_{1,1}}{1 + (j2\pi f\tau_{1,1})^{\alpha_{1,1}}}, \tag{5}$$

where $M_1^1 = [\Delta_{1,1}, \tau_{1,1}, \alpha_{1,1}]$ are its parameters.

The initial approximations for the parameters of formula (5) were chosen using the maximum value of the experimental frequency dependence of the imaginary part of the CDP $\cdot\, \varepsilon''(f_{\max_1})$ and its corresponding frequency $f_{\max_1}$: $\tau^{(0)}_{1,1} = 1/(2\pi f_{\max_1})$ and $\Delta^{(0)}_{1,1} \geq 2\varepsilon''(f_{\max_1})$, and also accepted $\alpha^{(0)}_{1,1} = 0.5$.

The identification of model (5) for the data of subrange 1 was carried out using the LSM with expression (2). The sequential simplex method (SSM), one of the simplest yet effective optimization techniques, was employed to find the minimum [38–40].

The procedure for implementing the SSM consisted of the following steps:

1. Determination of target function values (2) for $(m+1)$ vertices of a convex figure (simplex) in m-dimensional space (in this case, $m = 3$) is performed by calculating the coordinates (model parameters) of the vertices using the following formula:

$$x_{ij} = x_{0j} + X_{ij} \cdot \Delta x_j, \tag{6}$$

where x_{0j} and Δx_j are the initial value and the variation step of the j-th variable in the natural coordinate system, $i = 1, ..., m$; $j = 1, ..., m+1$; X_{ij} is the coded value of the j-th variable for the i-th vertex, calculated as:

$$\geq X_{ij} = \begin{cases} -\left[\dfrac{1}{2j(j+1)}\right]^{1/2}, & i < j+1, \\ \left[\dfrac{j}{2j(j+1)}\right]^{1/2}, & i = j+1, \\ 0, & i \geq j+1. \end{cases}$$

2. Sequential exclusion of the simplex vertex with the worst target function value, replacing it with a new (reflected) vertex with coordinates:

$$x_{kj}^{(\mathrm{ref})} = \frac{2}{n}\sum_{l}^{n+1} x_{ij} - \left(\frac{2}{n}+1\right)\cdot x_{ij}^{(\mathrm{excl})}. \tag{7}$$

As a result, a new simplex is obtained, where the excluded vertex is replaced by a new vertex obtained by reflecting the excluded vertex relative to the opposite face of the initial simplex. The worst vertex is replaced in the same manner. If the vertex with the minimum target function value remains unchanged, any vertex of the last simplex is selected as the center of the plan, and the variation step is reduced. This process is repeated until the step becomes smaller than the permissible error for determining the coordinates.

The initial approximations and refined parameter values of the model are provided in Table 1, and the visual results of model identification are shown in Fig. 1.

Table 1. Parameters of models approximating the experimental dielectric spectrum.

Subrange	Dispersion region in Fig. 1		Initial approximations of model parameters			Clarified values		
	Number	Model	$\tau^{(0)}$, c	$\Delta^{(0)}$	$\alpha^{(0)}$	τ, c	Δ	α
1	*I*	Cole–Cole	$1.1\cdot10^{3}$	$1.65\cdot10^{3}$	0.5	77.2	$2.53\cdot10^{3}$	0.49
2	*II*	Cole–Cole	$8.4\cdot10^{-6}$	10.1	0.05	$8.04\cdot10^{-7}$	9.85	0.017
2	*III*	Cole–Cole	$1.5\cdot10^{-6}$	8.1	0.05	$8.41\cdot10^{-8}$	4.59	0.029
3	*IV*	Debye	$1\cdot10^{-9}$	$1.84\cdot10^{2}$	–	$2.40\cdot10^{-10}$	$1.84\cdot10^{2}$	–

3.2. High-frequency subrange

The high-frequency rise in the dependence $\varepsilon''(f)$, observed in subrange *3*, corresponds to the initial segment of dielectric dispersion caused by Maxwell

relaxation of free charges in conductive inclusions within a dielectric matrix [35]. This process typically exhibits a narrow distribution in relaxation times and can be described using the Debye model

$$\epsilon^*_{3,1}\left(f, M_3^{\langle 1\rangle}\right) = \epsilon^{(h)}_{3,1} + \frac{\Delta_{3,1}}{1 + j2\pi f\tau_{3,1}}, \tag{8}$$

where $M^{(3)} = [\Delta_{3,1}, \tau_{3,1}]$.

Given the lack of complete experimental data for this dispersion region, it is advisable to rely on physical concepts and literature data about this relaxation process when selecting initial approximations for the parameters in expression (8). This corresponds to the highest-frequency region ($f_{\max_4} \sim 10^9$ and above [35, 37, 41]), and an initial approximation for the relaxation time can be estimated as $\tau^{(0)}_{3,1} = 1/(2\pi f_{\max_4})$. The post-dispersion real part of the CDP $\epsilon^{(h)}_{i.j} = \epsilon_h$ can be estimated as the relative dielectric permittivity of the material with a heterogeneous structure and a larger volume fraction of ZnO ($\varepsilon\prime \approx 10$) [35, 42].

It should be noted that the parameter $\Delta^{(0)}_{3,1}$ can be directly determined using the formula $\Delta^{(0)}_{3,1} = \epsilon^{(l)}_{i.j} - \epsilon^{(h)}_{i.j}$, where $\epsilon^{(l)}_{i.j}$ is the pre-dispersion value of the relative dielectric permittivity, which can be derived directly from the experimental dependence of the real part of the CDP.

The identification of model (8) for the experimental data of subrange 3 was carried out in a similar manner to that described for subrange *1*, using the previously described SSM procedure (see Table 1 and Fig. 1).

3.3. Subrange of intermediate frequencies

When processing the data for the dispersion region in the intermediate frequency subrange, where an asymmetric broadened maximum of the imaginary part of the CDP is observed, the experimental spectral dependence $\varepsilon''(f)$ must first be corrected according to expression (3). Considering the broadened shape of this maximum and the physical concept of two relaxation processes present in this frequency range, the initial values of the relaxation times were determined using the deconvolution method described in [22].

To apply this method, the corrected experimental dependence in the intermediate frequency subrange was represented analytically as $\overline{\varepsilon}''(r)$, where $r = \log(f_0/f)$ and $f_0 = 1\,\text{Hz}$ is a constant reference frequency. The Cole–Cole formula was used as the model for dielectric dispersion in the corresponding regions.

In this case, the distribution function of relaxation times $G(\tau)$ as a function of the generalized coordinate $s = \log(\tau/\tau_0)$, where $\tau_0 = 1/(2\pi f_0)$ is an arbitrary value, has the form:

$$G(s) = \int_{-\infty}^{\infty} \epsilon''(r) T''(s-r) \mathrm{d}r. \tag{9}$$

Here, the function $T''(r)$ is defined as

$$T''(r) = \frac{16k}{\pi} \int_0^{\infty} \widetilde{W}(p) \cosh\left(\frac{\pi^2 p}{k}\right) \cos(2\pi p r)\, \mathrm{d}r,$$

which is derived in [22] based on the general expression for the CDP in terms of a convolution equation with the relaxation time distribution function and its solution via direct (denoted by a tilde) and inverse Fourier transforms, using the Kaiser–Bessel digital filtering window:

$$\widetilde{W}(p) = \begin{cases} \dfrac{I_0\left[\pi a \sqrt{1-(p/p_m)^2}\right]}{I_0[\pi a]}, & |p| \le p_m, \\ 0, & |p| > p_m, \end{cases}$$

where $I_0(x) = \sum_{k=0}^{\infty} \left[\frac{(x/2)^k}{k!}\right]^2$ is the modified Bessel function of the first kind and of order zero.

For the analysis, the recommended window parameters were used: $p_m = 3$ (window width parameter) and $\pi a = 2$ (a is a non-negative real number determining the window shape) [43].

Figure 2 presents the results of applying the described deconvolution procedure to the experimental frequency dependence of the imaginary part of the CDP in the intermediate frequency subrange shown in Fig. 1.

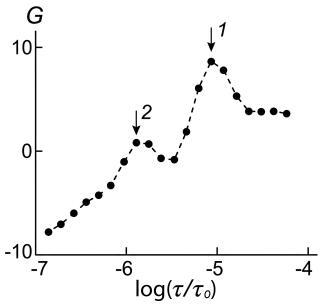

Fig. 2. Relaxation time distribution function obtained for the experimental frequency dependence of the imaginary part of the CDP (Fig. 1) in the intermediate frequency subrange. The identified peaks correspond to dispersion regions *II* (*1*) and *III* (*3*) in Fig. 1.

As seen in Fig. 2, the dependence can be associated with two relaxation processes characterized by relaxation times $\tau_{2,1}^{(0)}$ and $\tau_{2,2}^{(0)}$, which are subsequently

used as their initial values (Table 1). The parameters $\Delta_{2,1}^{(0)}$ and $\Delta_{2,2}^{(0)}$ are estimated using the earlier expression $\tau\Delta_{i,j}^{(0)} \sim 2 \cdot \overline{\varepsilon}''|_{f=1/2\pi_{i,j}^{(0)}}$.

The additional parameters $\alpha_{2,1}^{(0)}$ and $\alpha_{2,2}^{(0)}$ for the Cole–Cole model are chosen arbitrarily (using trial methods aimed at improving calculation efficiency).

Refinement of the parameters of the selected approximation models for the identified dispersion regions is performed by fitting the models to the experimental data using the LSM, which involves minimizing the target function given by expression (2). The results obtained are shown in Fig. 3 and Table 1.

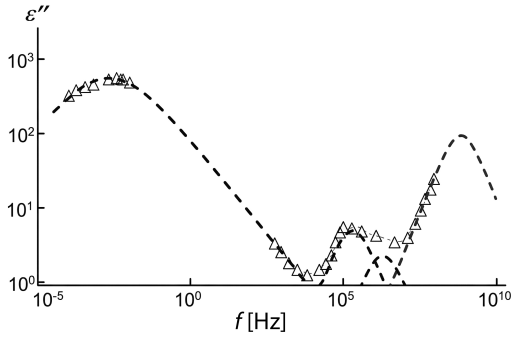

Fig. 3. Experimental frequency dependence of the imaginary part of the CDP of ZnO-based varistor ceramics (triangles) and approximations of its individual dispersion regions using selected phenomenological models (dashed lines).

3.4. Model parameters for the entire frequency range

Based on the analysis conducted, the experimental frequency dependence of the imaginary part of the CDP over the entire investigated range can be described by four relaxation processes. The appropriate analytical mathematical model (1) consists of the sum of three dispersion expressions corresponding to the Cole–Cole model and one Debye formula.

It should be noted that further refinement of the parameters of this additive model for the entire range of the experimental dielectric spectrum, following the general algorithm described earlier (Sec. 2), provided estimates of all its parameters. These values differed only slightly from those obtained earlier when processing each relaxation process individually in its respective frequency subrange. The refined values are presented in Table 1.

The results of applying this processing and preliminary analysis method to several experimental dielectric spectra of ZnO-based varistor ceramics are shown in Fig. 4.

The approximation error estimates for the frequency dependencies of the real and imaginary parts of the CDP, based on formula (4), yielded values $v \leq 10\,\%$.

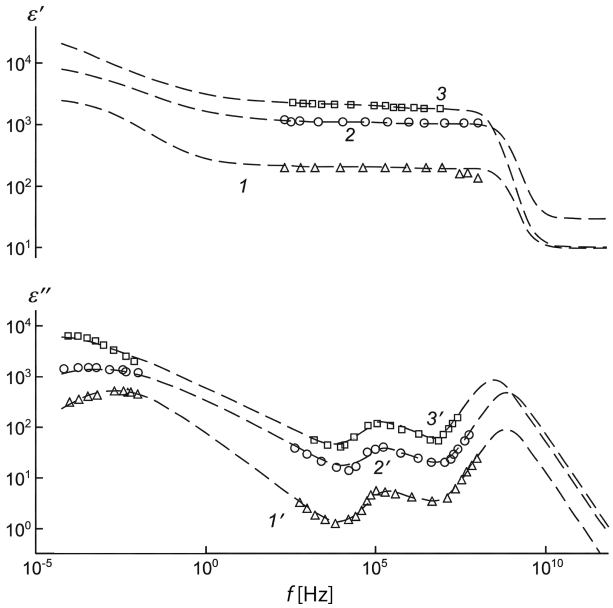

Fig. 4. Frequency dependencies of the real ε' and imaginary ε'' parts of the CDP of ZnO-based varistor ceramics, fabricated at synthesis temperatures T_C, K: *1*, *1′* – 1173; *2*, *2′* – 1423; *3*, *3′* – 1573 (symbols of different shapes), and the results of approximation using the generalized additive model for the entire investigated range (dashed lines).

4. Conclusions

An algorithm for processing and analyzing complex dielectric spectra over a wide frequency range has been proposed. The algorithm includes:

- dividing the entire frequency range into several shorter subranges corresponding to separate observable dispersion regions;
- identifying each of these subranges into the experimental spectrum using one of the well-known elementary phenomenological relaxation models, sequentially, starting from the extreme dispersion regions and accounting for their influence on adjacent internal regions;
- refining the parameters of the generalized additive model, which represents the sum of elementary models for the selected subranges, using the complete experimental broadband dielectric spectrum and the LSM.

The developed algorithm was tested by processing the frequency dependencies of the CDP of ZnO-based varistor ceramic materials. The results showed satisfactory agreement with existing physical concepts and demonstrated the efficiency and potential of this approach for developing computer technologies for processing and preliminary analysis of experimental dielectric spectra of a wide range of heterogeneous materials.

References

1. Kremer F., Schönhals A. [Eds], *Broadband Dielectric Spectroscopy*, Springer Science & Business Media, 2002.
2. Poplavko Y., *Dielectric Spectroscopy of Electronic Materials: Applied Physics of Dielectrics*, Woodhead Publishing, 2021.
3. Elbashar Y.H. *et al.*, Electric, dielectric and optical properties of doped metal oxide glasses: A review, *Nonlinear Optics, Quantum Optics: Concepts in Modern Optics*, **52**(1–2): 1–36, 2020.
4. Yang H., Zhang S., Yang H., Li E., Usage of P–V–L bond theory in studying the structural/property regulation of microwave dielectric ceramics: A review, *Inorganic Chemistry Frontiers*, **7**(23): 4711–4753, 2020, https://doi.org/10.1039/D0QI00907E.
5. Shanmugasundram H.P.P.V., Jayamani E., Soon K.H., A comprehensive review on dielectric composites: Classification of dielectric composites, *Renewable and Sustainable Energy Reviews*, **157**: 112075, 2022, https://doi.org/10.1016/j.rser.2022.112075.
6. Popov I., Cheng S., Sokolov A.P., Broadband dielectric spectroscopy and its application in polymeric materials, [in:] *Macromolecular Engineering: From Precise Synthesis to Macroscopic Materials and Applications*, Hadjichristidis N., Gnanou Y., Matyjaszewski K. *et al.* [Eds], Wiley, pp. 1–39, 2022, https://doi.org/10.1002/9783527815562.mme0059.
7. Teyssedre G., Zheng F., Boudou L., Laurent C., Charge trap spectroscopy in polymer dielectrics: A critical review, *Journal of Physics D: Applied Physics*, **54**(26): 263001, 2021, https://doi.org/10.1088/1361-6463/abf44a.
8. Kumbhakar K., Pham T.D., Lee K.K., Kwak K., Cho M., Dielectric relaxation spectroscopy for the characterization of ion transport in solid polymer electrolytes in Li-ion cells, *Electrochimica Acta*, **462**: 142759, 2023, https://doi.org/10.1016/j.electacta.2023.142759.
9. Fahmy H.M. *et al.*, Dielectric spectroscopy signature for cancer diagnosis: A review, *Microwave and Optical Technology Letters*, **62**(12): 3739–3753, 2020, https://doi.org/10.1002/mop.32517.
10. Schreiner T.G., Adam M., Broadband dielectric spectroscopy and its role in the characterization of biological cells, *Bulletin of the Polytechnic Institute of Iaşi. Electrical Engineering, Power Engineering, Electronics Section*, **67**(1): 9–20, 2021, https://doi.org/10.2478/bipie-2021-0001.
11. Mertens M., Chavoshi M., Peytral-Rieu O., Grenier K., Schreurs D., Dielectric spectroscopy: Revealing the true colors of biological matter, *IEEE Microwave Magazine*, **24**(4): 49–62, 2023, https://doi.org/10.1109/MMM.2022.3233510.
12. Wübbenhorst M., van Turnhout J., Analysis of complex dielectric spectra. I. One-dimensional derivative techniques and three-dimensional modelling, *Journal of Non-Crystalline Solids*, **305**(1–3): 40–49, 2002, https://doi.org/10.1016/S0022-3093(02)01086-4.
13. Haspel H., Kukovecz Á., Kónya Z., Kiricsi I., Numerical differentiation methods for the logarithmic derivative technique used in dielectric spectroscopy, *Processing and Application of Ceramics*, **4**(2): 87–93, 2010, https://doi.org/10.2298/PAC1002087H.
14. Koposov G.D., Volkov A.S., Tyagunin A.V., Perfiliev R.O., Numerical simulation method for identification of experimental results according to frequency dispersion of dielectric permittivity by Gavrylyak-Negami, [in:] *IOP Conference Series: Earth and Environmental Science*, **263**(1): 012059, 2019, https://doi.org/10.1088/1755-1315/263/1/012059.

15. Rosa C.F.A.E., Capelas de Oliveira E., Relaxation equations: Fractional models, *Journal of Physical Mathematics*, **6**(2): 1–7, 2015, https://doi.org/10.4172/2090-0902.1000146.

16. Barelli E., *Dielectric Relaxation in Biological Materials*, Master Thesis, Scuola di Scienze Corso di Laurea in Fisica, Universita di Bologna, 2015, https://amslaurea.unibo.it/9102/1/Eleonora_Barelli_tesi.pdf.

17. Odinaev S., Makhmadbegov R.S., Frequency dispersion of dielectric permittivity and dielectric losses in aqueous KCl and CsCl solutions depending on their state parameters, *Ukrainian Journal of Physics*, **60**(12): 1211–1211, 2015, https://doi.org/10.15407/ujpe60.12.1211.

18. Moon Y.I., Jung J.K., Chung K.S., Dielectric relaxation spectroscopy in synthetic rubber polymers: Nitrile butadiene rubber and ethylene propylene diene monomer, *Advances in Materials Science and Engineering*, **1**: 8406059, 2020, https://doi.org/10.1155/2020/8406059.

19. Bello A., Laredo E., Grimau M., Comparison of analysis of dielectric spectra of PCL in the ε^* and the M^* formalism, *Journal of Non-Crystalline Solids*, **353**(47–51): 4283-4287, 2007, https://doi.org/10.1016/j.jnoncrysol.2007.08.041.

20. Boukamp B.A., Distribution (function) of relaxation times, successor to complex nonlinear least squares analysis of electrochemical impedance spectroscopy?, *Journal of Physics: Energy*, **2**(4): 042001, 2020, https://doi.org/10.1088/2515-7655/aba9e0.

21. Sabnis S.M., Rander D.N., Kanse K.S., Joshi Y.S., Kumbharkhane A.C., Spectroscopic measurement and dielectric relaxation study of vegetable oils, *Information Processing in Agriculture*, **11**(3): 397–408, 2024, https://doi.org/10.1016/j.inpa.2023.04.002.

22. Salefran J.L., Dutuit Y., The use of a discriminative window in deconvolution method applied to dielectric data, *The Journal of Chemical Physics*, **74**(5): 3056–3063, 1981, https://doi.org/10.1063/1.441430.

23. Nicholson D.J., Kell D.B., Davey C.L., Deconvolution of the dielectric spectra of microbial cell suspensions using multivariate calibration and artificial neural networks, *Bioelectrochemistry and Bioenergetics*, **39**(2): 185–193, 1996, https://doi.org/10.1016/0302-4598(95)01890-5.

24. Ganea C.P., Zgura I., Frunza L., Numerical deconvolution approaches for dielectric characteristics of complex composite materials based on liquid crystals and oxide nanopowders, *Materials Chemistry and Physics*, **309**: 128372, 2023, https://doi.org/10.1016/j.matchemphys.2023.128372.

25. Steeman P.A.M., van Turnhout J., Fine structure in the parameters of dielectric and viscoelastic relaxations, *Macromolecules*, **27**(19): 5421–5427, 1994, https://doi.org/10.1021/ma00097a023.

26. van Turnhout J., Wübbenhorst M., Analysis of complex dielectric spectra. II: Evaluation of the activation energy landscape by differential sampling, *Journal of Non-Crystalline Solids*, **305**(1–3): 50–58, 2002, https://doi.org/10.1016/S0022-3093(02)01120-1.

27. Schäfer H., Sternin E., Stannarius R., Arndt M., Kremer F., Novel approach to the analysis of broadband dielectric spectra, *Physical Review Letters*, **76**(12): 2177, 1996, https://doi.org/10.1103/PhysRevLett.76.2177.

28. van Turnhout J., Better resolved dielectric dispersions by the apt use of Kramers–Kronig relations, differential operators, and all-in-1 modeling, [in:] *Current Topics on Chemistry and Biochemistry*, Vol. 5, pp. 40–82, 2022, https://doi.org/10.9734/bpi/ctcb/v5/3055C.

29. Güneşer M.T., Artificial intelligence solution to extract the dielectric properties of materials at sub-THz frequencies, *IET Science, Measurement & Technology*, **13**(4): 523–528, 2019, https://doi.org/10.1049/iet-smt.2018.5356.

30. Saadaoui Y., Zegnini B., Seghier T., Analysis of dielectric relaxation spectra by diffusive representations method: Case of organic dielectrics, *Przeglad Elektrotechniczny*, **97**(8): 52–58, 2021, https://doi.org/10.15199/48.2021.08.10.

31. Shoup T.E., *A Practical Guide to Computer Methods for Engineers*, Prentice-Hall, Englewood Cliffs, New York, 1979.

32. Gupta T.K., Application of zinc oxide varistors, *Journal of the American Ceramic Society*, **73**(7): 1817–1840, 1990, https://doi.org/10.1111/j.1151-2916.1990.tb05232.x.

33. Ganesh K.S., A review of zinc oxide varistors for surge arrester, [in:] *2018 4th International Conference on Electrical Energy Systems (ICEES)*, pp. 470–474, 2018.

34. Tonkoshkur A.S., Glot A.B., Ivanchenko A.V., Basic models in dielectric spectroscopy of heterogeneous materials with semiconductor inclusions, *Multidiscipline Modeling in Materials and Structures*, **13**(1): 36–57, 2017, https://doi.org/10.1108/MMMS-08-2016-0037.

35. Tonkoshkur A.S., Effect of charge exchange of local centers on dielectric phenomena in disperse semiconductor structures, *Ukrainskii Fizicheskii Zhurnal*, **23**: 2030–2038, 1978.

36. Avdeenko B.K., Tonkoshkur A.S., Chernenko I.M., Dielectric losses in oxide-zinc ceramics, *Inorganic Materials*, **18**: 1028, 1982.

37. Fernandez-Hevia D., Peiteado M., de Frutos J., Caballero A.C., Fernandez J.F., Wide range dielectric spectroscopy of ZnO-based varistors as a function of sintering time, *Journal of the European Ceramic Society*, **24**(6): 1205–1208, 2004, https://doi.org/10.1016/S0955-2219(03)00411-4.

38. Corriou J.P., *Numerical Methods of Optimization. Theory and Practice for Engineers*, Springer International Publishing, 2021.

39. Michałowska-Kaczmarczyk A.M., Michałowski T., Simplex optimization and its applicability for solving analytical problems, *Journal of Applied Mathematics and Physics*, **2**(7): 723–736, 2014, https://doi.org/10.4236/jamp.2014.27080.

40. Cerdà V., Cerdà J.L., Idris A.M., Optimization using the gradient and simplex methods, *Talanta*, **148**: 641–648, 2016, https://doi.org/10.1016/j.talanta.2015.05.061.

41. Levinson L.M., Philipp H.R., High-frequency and high-current studies of metal oxide varistors, *Journal of Applied Physics*, **47**(7): 3116–3121, 1976, https://doi.org/10.1063/1.323059.

42. Samsonov G.V., *Physicochemical Properties of Oxides* [in Russian], Metallurgy, Moscow, 1976.

43. Nuttall A., Some windows with very good sidelobe behavior, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **29**(1): 84–91, 1981, https://doi.org/10.1109/TASSP.1981.1163506.

Received January 7, 2025; revised February 25, 2026; accepted March 3, 2026; available online March 10, 2026; version of record April 20, 2026; published issue June 17, 2026.

Computer Assisted Methods in Engineering and Science
33(2): 101–124, 2026, https://doi.org/10.24423/cames.2025.1862

Multi-Input CNN for Vision-Based *In Situ* Analysis of Extraterrestrial Surface Composition

Kacper Marciniak[1)*], Michał Grzesiak[2)], Igor Zaworski[3)], Dominik Pawliszewski[3)], Dominika Zygarlicka[3)], Michał Wnuk[3)], Adrian Zakrzewski[1)]

[1)] *Faculty of Mechanical Engineering,*
[2)] *Faculty of Geoengineering, Mining and Geology,*
[3)] *Student Association of Unconventional Vehicles OFF-ROAD,*
Wrocław University of Science and Technology, Wrocław, Poland

[*] *Corresponding Author: kacper.marciniak@pwr.edu.pl*

Earth's natural resources are finite, which is why engineers and scientists are increasingly directing their efforts toward the extraction of materials from celestial bodies. Beyond the extraction itself, however, a major challenge lies in the localization of valuable substances and the assessment of their quality *in situ*. In this article, we present a flexible and robust method for estimating the content of selected components in heterogeneous mixtures using RGB image processing. The proposed multi-input deep learning approach, equipped with a shared lightweight convolutional neural network (CNN) backbone, achieves high prediction accuracy, with a root mean squared error (RMSE) of $(0.190 \pm 0.024)\,\%$. The framework supports a variety of backbone architectures, including lightweight models, making it suitable for deployment on edge devices such as planetary rovers. Furthermore, the method is inherently adaptable, enabling straightforward extension to other task, for example, the analysis of more complex mixtures or inference based on multi- or hyperspectral imagery.

Keywords: machine vision, machine learning, edge computing, *in situ* resource utilization, planetary rover.

1. Introduction

Earth's resources are finite. Despite increasing awareness of responsible mining and improved resource exploitation, in the near future, humanity will face the challenge of how to keep up with the growing demand for metals and rare earth elements (REE) and where to find new deposits [1].

A natural direction for development and mining exploration is therefore space. Given current technology capabilities, the most likely candidate for the new chapter in mining is the Moon. Its crust contains many valuable elements that, due to their properties and economic importance, will be one of the most important targets of extraterrestrial exploitation [2]. One such compound is ilmenite, a titanium-iron oxide found in igneous rocks. On the Moon, it occurs in euhedral to anhedral, blocky, tabular, platy, and skeletal forms in the lunar maria [3], where its abundance is estimated to be as high as 20 % [4]. This is confirmed by the results of Apollo 11 and 17 missions, where ilmenite was identified as the third most common mineral in the samples collected [5]. On Earth, ilmenite is used as the main ore of titanium and is used in the aerospace or medical industries. On the Moon, ilmenite, however, can also be a source of oxygen used for life support systems or construction materials. Currently used models for estimating demand and extraction of metals, indicate that under growth rates higher than those observed today (about 7 % per decade), shortages of some metals, including titanium, could occur in about 50 years [6]. Therefore, increasingly better classification of deposits (including extraterrestrial sources) will be critical for sustaining the metals market.

Regolith containing ilmenite serves as an example of a binary mixture. Such a mixture consists of two – ideally distinguishable – components with one present in lower abundance than the other. A key characteristic of binary mixtures is their non-homogeneous nature, meaning their components are not evenly distributed throughout the entire volume. Instead, local maxima and minima occur, where the density of one material significantly deviates from the overall average. This type of non-homogeneous binary mixture is commonly present in natural environments, both on Earth and on extraterrestrial bodies such as the Moon. Notable examples include olivine grains in volcanic sands in Hawaii, garnet-bearing sands in India, and ilmenite particles dispersed within lunar regolith.

The demand for ilmenite is high, prompting scientific interest in locating and extracting this mineral from the Moon's surface. In addition to numerous scientific publications on the subject, several competitions have been organized to explore how robotic lunar rover simulants can be utilized to study the distribution of ilmenite within the regolith. One such competition was hosted by the University of Adelaide, with participation from the Scientific Association of Unconventional Vehicles OFF-ROAD and its flagship initiative – the Project SCORPIO Mars rover prototype. These initiatives, often organized or supported by major space agencies, are intended to foster the development and validation of emerging space technologies for future missions focused on the exploration and utilization of extraterrestrial resources.

The utilization of lunar resources was first proposed during the development of the Apollo program due to considerations regarding the establishment

of a longer human presence on the Moon [9]. To take full advantage of the raw materials found on the lunar surface, comprehensive knowledge of their composition and distribution is essential. Thanks to missions such as the Lunar Reconnaissance Orbiter (LRO) and Chandrayaan-1, the lunar surface has been well explored, and the data obtained from these missions have made it possible to determine the presence of useful raw materials in lunar regions. Due to their orbital nature, both spacecraft were equipped with advanced remote sensing instruments capable of mapping the surface across multiple spectral ranges [8, 9]. This broad spectral coverage enabled the detection and analysis of surface minerals, including ilmenite [10].

Historically, numerous approaches have been employed to detect ilmenite using spectrometric techniques. Early efforts focused on ultraviolet-visible (UV-Vis) spectrometers, which were favored due to the distinctive spectral signatures produced by reflected light [11]. While UV-Vis analysis proved reliable for ilmenite detection in lunar maria, it lacked universality. Variations in surface coloration near young craters and in highland lunar regions significantly reduced the reliability of UV-Vis-based methods for estimating ilmenite content [12]. Alternative methods were developed and deployed during subsequent missions to improve detection reliability. Notably, the Moon Mineralogy Mapper (M^3) instrument onboard Chandrayaan-1 was specifically designed for high-resolution mineralogical analysis [9] and played a key role in advancing ilmenite detection capabilities.

Imaging in the visible light spectrum is gaining traction as a cost-effective method for analyzing the surfaces of extraterrestrial bodies, most often for terrain classification tasks. The processing of such imagery is increasingly supported by machine learning-based approaches, with convolutional neural networks (CNNs) being particularly prominent in planetary surface analysis. For example, the MarsNet architecture was developed for the detection of geological landforms in images captured by the Mars Reconnaissance Orbiter [13]. In another study, rock imagery from the Mars Science Laboratory (Curiosity) dataset was used to fine-tune a pretrained VGG-16 network, which outperformed other CNN-based models in classification tasks [14]. Additionally, a deep ensemble CNN was proposed for lunar terrain classification using images obtained from a lunar rover or its simulator; the network architecture preserved high-resolution feature maps and modeled inter-channel dependencies, enabling the detection of subtle terrain variations [15].

Complementary to CNN-based approaches, classical machine learning methods have also demonstrated effectiveness in extraterrestrial terrain classification tasks. For instance, in the analysis of Martian surface imagery using the MSLNet dataset (comprising 256×256 RGB patches from the Mastcam instrument), researchers extracted multiscale image features – including gradient-based, edge-strength, and frequency-domain descriptors – specifically tailored to

Martian geology. These hand-crafted features were subsequently evaluated using k-nearest neighbors (k-NN), support vector machines (SVM), and random forest (RF) classifiers, with the latter showing competitive performance [16].

Several rover missions have been equipped with microscopes or other close-up imaging systems [17–19]. Using these existing optical capabilities to estimate ilmenite content in regolith samples offers significant advantages. Most notably, it eliminates the need for additional specialized hardware, thereby reducing development and manufacturing costs. Furthermore, it improves mass efficiency, as these optical systems already serve multiple scientific and navigational purposes [20, 21]. By utilizing existing hardware, current and future rover platforms [22] could be adapted to perform *in situ* mineral content analysis with minimal modification.

This article presents the outcome of a collaborative effort aimed at developing a fully functional machine learning model capable of estimating ilmenite content in regolith samples collected using the Universal Land Exploration Platform (ULEP) mounted on the SCORPIO Infinity rover (Fig. 1).

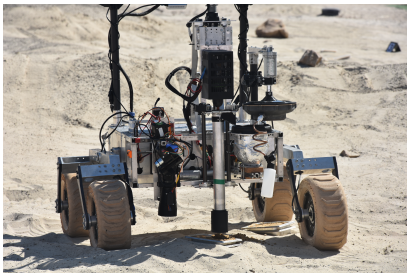

Fig. 1. The SCORPIO Infinity rover, equipped with the ULEP, analyzes the composition of a regolith sample.

The authors propose a multi-input model utilizing an existing lightweight CNN architecture as an encoder, enabling rapid analysis of images acquired by the onboard acquisition system. A key benefit of the proposed method is its low computational complexity. The algorithmic workload is modest sufficiently to be executed on standard onboard central processing units (CPUs), eliminating the need for high-performance processors or external computation resources. This is particularly advantageous in extraterrestrial missions, where bandwidth and communication delays with Earth are severely limited. As a result, the entire analysis pipeline can be conducted autonomously onboard the rover, enabling rapid decision-making and more efficient exploration workflows.

The developed method is highly flexible and scalable, as it supports the use of different encoding backbones and can be conveniently adapted to the requirements of a given task. Beyond ilmenite detection, it can be extended to other material types and to multi-component mixtures, formulated as multi-output

regression problems. Most importantly, the architecture is naturally suited for adaptation to multi- or hyperspectral imaging.

2. Materials and methods

2.1. Problem definition

Although orbital spectroscopic instruments provide data that enable the mapping of ilmenite-rich regions on the Moon, this information alone is insufficient for the effective selection of an optimal mining site. The main problem remains the accurate identification of locations with the highest concentration of ilmenite within the lunar regolith. This parameter is critical because the success of such a lunar *in situ* resource use (ISRU) mission depends on minimizing risks and failure factors.

Given the high cost and complexity of extraterrestrial missions, the trial-and-error approach for ilmenite-rich site identifications is not only unprofessional but also costly and highly inefficient. Without precise data on the concentration of ilmenite, the risk of performing the mission in suboptimal areas increases, which can compromise the viability of the mission.

The matter is further complicated by the fact that the lunar ilmenite is present in a form of non-homogeneous binary mixture with regolith. A simulant of this mixture, used in this study (ilmenite and regolith from the highlands), may be found on the back of basaltic marias with high titanium content and in primary crusts. The Apollo missions studied these sites, where the albedo of light was lower than in the highlands, but higher than in the maria [23]. In addition, for such mixtures to form, phenomena capable of transferring the material with high titanium and iron content are needed. Due to the lack of an atmosphere on the Moon, thermal erosion phenomena, or meteorite impacts and ejecta may be the processes that enable such mixing of these materials [24]. Due to the non-homogeneous nature of the described mixture, standard vision-based analysis cannot accurately determine the abundance of ilmenite at a given location.

Therefore, the development of a reliable method for the accurate determination of the ilmenite content in the regolith is essential. Such a tool would be a major advancement not only for lunar exploration and ilmenite extraction but also a groundbreaking solution for the broader field of space mining.

2.2. Data acquisition

The dataset used in this study was constructed from simulant regolith samples, prepared through a controlled, multi-step process. The initial stage involved mixing sand and ilmenite using the conical mixing method. This method was

based on the formation of a cone with both sand and ilmenite, followed by quartering the mixture. Each of the quarters was then mixed separately for about five minutes to later be recombined into a single well-mixed probe. Samples containing ilmenite at concentrations ranging from 1 % to 15 %, in 1 % increments, were prepared.

Although the simulant inevitably represents a simplified version of lunar regolith, its design was guided by several key physical similarities. Quartz sand with up to 95 % quartz content and a dominant grain-size fraction of 0.1 mm to 1 mm was used, ensuring that the bulk grain-size distribution is reasonably consistent with Apollo mission observations. While natural lunar regolith also contains minor components such as volcanic glass, meteoritic fragments, and agglutinates (typically only a few percent of the total), our mixture includes aluminium oxides and plagioclases, which can serve as analogous phases in microscopic analyses. Most importantly, the simulant reproduces the low albedo (∼0.06), comparable to that of lunar mare soils, making it suitable for systematic assessment of ilmenite-related optical effects. At the same time, finer-scale features of true regolith, such as glass-rich particles and space-weathering effects, are not represented.

Images of the prepared samples were subsequently acquired using a CNC (computer numerical control) machine equipped with a DLT-Cam PRO 5 MP camera (Delta Optical, Poland), a 200× zoom microscope lens, model 074F-97376 (Techrebal, Poland), and a ring illuminator with 16 RGB LEDs (Fig. 2). The 8-bit RGB values for the LEDs were set to R = 160, G = 176, B = 150 with a 5V input voltage. This setup ensured consistent white balance and lighting conditions across all images. To maximize image clarity, the Z-axis position of the microscope was manually adjusted for each sample. During the data acquisition phase, a G-code program controlled the spiral movement of the microscope across the surface of each sample, ensuring comprehensive coverage of the specimen. The resulting recordings were captured as 9-minute video sequences, pro-

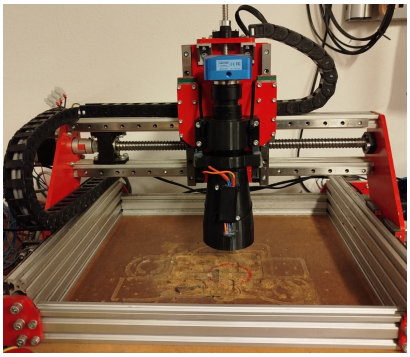

Fig. 2. The experimental setup for data acquisition.

viding high-resolution visual data of the heterogeneous regolith surface. These sequences were subsequently sampled to extract non-overlapping frames, which were then downsized to a resolution of 256×256 pixels.

As a result, a total of 22 176 labeled images were prepared, representing 16 classes corresponding to ilmenite content levels ranging from 0 % to 15 %. A randomly selected sample of 5000 images was chosen while ensuring class balance. This sample was used to create five-fold cross-validation fold, each consisting of 4000 images in the training subset and 1000 in the test subset.

Above procedure enabled the generation of a rich and diverse dataset, representative of a wide range of ilmenite concentrations, as illustrated in Fig. 3.

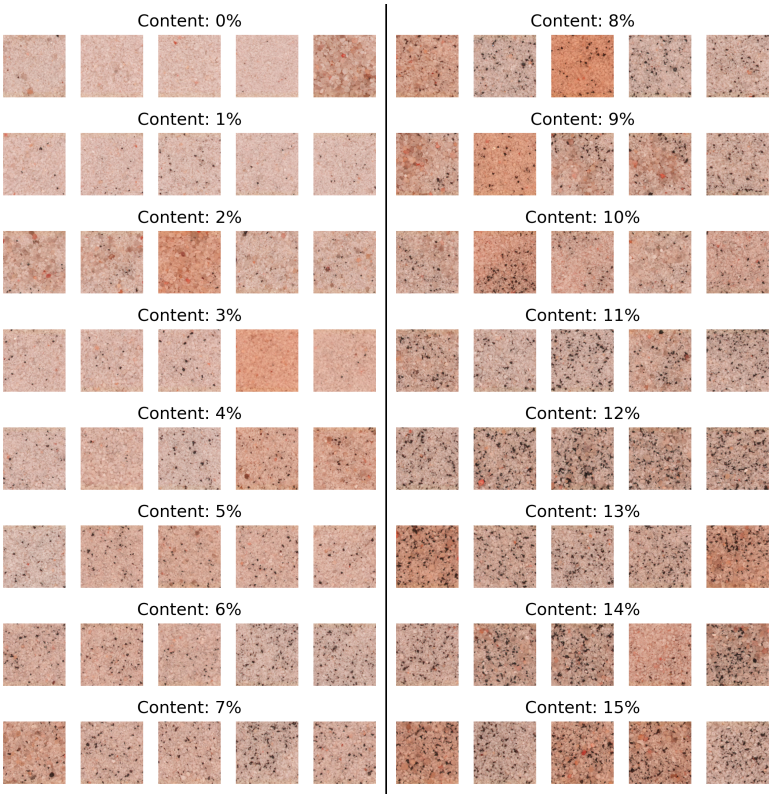

Fig. 3. Representative image samples (five per class) illustrating varying ilmenite content levels.

2.3. Data analysis using a reference method

As a baseline for comparison in the conducted study and to analyze the available data, a classical computer vision approach was proposed. In this method, ilmenite content is estimated by analyzing the ratio of dark to bright regions within the input image, as outlined in Algorithm 1.

Algorithm 1. Reference method: measurement of dark area content in the input image.

Require: Input image I in RGB colorspace

Ensure: Percentage of dark area in the image

1: $KERNEL \leftarrow 3 \times 3$ elliptical structuring element

2: $THRESHOLD \leftarrow 128$

3: Resize image: $I_{\text{resized}} \leftarrow \text{resize}(I, 256 \times 256)$

4: Convert to HSV and extract V channel: $I_V \leftarrow \text{HSV}(I_{\text{resized}})[V]$

5: Apply thresholding: $I_{\text{thresh}} \leftarrow \text{threshold}(I_V, THRESHOLD)$

6: Negate image: $I_{\text{negated}} \leftarrow \text{bitwise\_not}(I_{\text{thresh}})$

7: Morphological closing: $I_{\text{final}} \leftarrow \text{morphology\_close}(I_{\text{negated}}, KERNEL)$

8: Compute the percentage of white pixels:

$$\text{value} = \frac{\text{count\_nonzero}(I_{\text{final}})}{\text{size}(I_{\text{final}})} \times 100$$

9: **return** value

The prepared dataset was analyzed using the previously described reference method. The results – representing the number of dark pixels as a function of actual ilmenite content – are presented as a boxplot in Fig. 4. As expected, a general increasing trend can be observed. However, substantial intra-class variability was also noted, indicating high dispersion within individual ilmenite concentration levels. This variability significantly limits the effectiveness of classical computer vision techniques for accurately estimating material composition in the mixture.

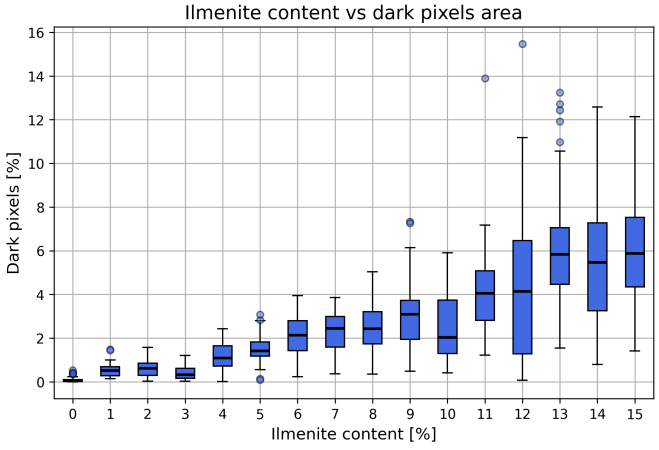

Fig. 4. Image samples with different ilmenite content.

2.4. Proposed solution

Given the high heterogeneity of the input data, it was decided to perform inference on a batch of N input images instead of a single image. For this purpose, three methods were proposed:

- **DL regressor + mean** – utilizing an existing deep learning (DL) model with a final regression layer, followed by averaging the results for the batch of input images,
- **DL regressor + k-NN** – using the same DL regressor in combination with a classical machine learning method – k-nearest neighbors regression (k-NNR),
- **multi-input regressor** – developing a model capable of handling multiple inputs, allowing simultaneous processing of the entire batch of input images.

To adapt an existing DL classification model for the regression task, the classification head was replaced with a linear layer producing a single output. Additionally, a rectified linear unit (ReLU) activation function was applied to the model output to enforce non-negative predictions, consistent with the physical constraint that ilmenite concentration cannot take negative values. The resulting regression outputs for a batch of input images were either averaged, as in the first approach, or used as input to a subsequent k-NNR model.

To enable simultaneous inference on the entire batch of input images and incorporate all available information into the regression process, the authors proposed a custom architecture, as illustrated in Fig. 5. The model consists of an encoding block that processes N input images using a shared backbone, embedding each image into a deep feature vector of length 512. These feature vectors

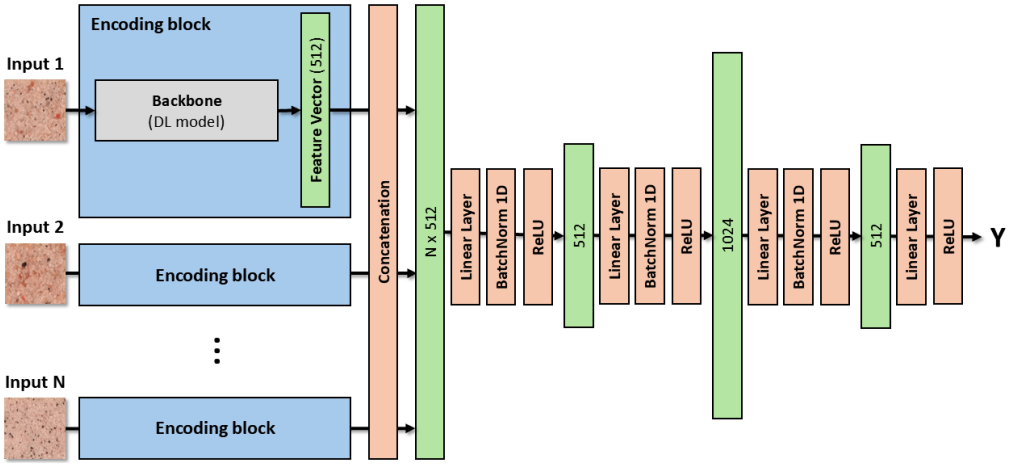

Fig. 5. Overview of the proposed multi-input DL architecture.

are then concatenated to form a single vector of size $N \times 512$, which is subsequently passed to the regression head composed of four fully connected (linear) layers. To mitigate overfitting during training, batch normalization was applied after each linear layer, except the final one. The ReLU activation function was used throughout the model to introduce non-linearity and enforce non-negative outputs.

2.5. Encoder architecture selection

The proposed approach, based on a multi-input model with a shared encoder, enables the use of different deep architectures as well as their straightforward replacement and adaptation to the problem under consideration. In the experiments presented in this study, three representative deep architectures were evaluated:

- **ResNet-18** – the smallest architecture in the ResNet family of residual neural networks, widely used as a lightweight yet effective baseline in many computer vision tasks,
- **MobileNetV2** – an efficient architecture originally designed for mobile and embedded applications, employing depthwise separable convolutions and inverted residual blocks to significantly reduce computational cost while maintaining competitive accuracy,
- **EfficientNet-B0** – a baseline model from the EfficientNet family, which uniformly scales depth, width, and resolution using a compound scaling method, achieving a favorable balance between accuracy and efficiency.

These models were selected to cover a spectrum of trade-offs between model size, computational efficiency, and predictive accuracy: ResNet-18 serving as a well-established baseline, MobileNetV2 representing resource-efficient designs, and EfficientNet-B0 offering state-of-the-art performance under constrained model complexity. The evaluation of these diverse architectures allowed for a fair assessment of which model is best suited for the application under study.

2.6. Evaluation measures

The evaluation of measurement accuracy was based on the following metrics: root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2), as defined in Eqs. (1)–(3).

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}, \tag{1}$$

$$\mathrm{MAE} = \sum_{i=1}^{N} |y_i - \widehat{y}_i|, \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2}. \tag{3}$$

Additionally, the authors introduced an error rate metric, defined as the proportion of predictions for which the absolute error exceeded 0.5 %. This threshold was selected to identify measurements considered significantly inaccurate by the system.

All experiments, including model training and validation, were conducted on a workstation equipped with an Intel Xeon Silver 4110 CPU (Intel Corporation, USA), 64 GB RAM and an NVIDIA RTX 4080 graphics card with 16 GB VRAM (NVIDIA Corporation, USA). The implementation was developed in Python using the PyTorch framework. Additional inference time measurements were conducted on an NVIDIA Jetson AGX Orin (NVIDIA Corporation, USA).

2.7. Model preparation

The input images for the model underwent a standardization process. The mean values and standard deviations for the red, green, and blue (R, G, and B) channels were determined empirically through statistical analysis of a sample consisting of 100 randomly selected images from the prepared dataset. The obtained values are as follows: mean = $[0.536, 0.606, 0.782]$, std = $[0.114, 0.107, 0.103]$.

During the training of the base DL regression models, pre-trained weights from the ImageNet-1K dataset were employed to leverage transfer learning and accelerate convergence. In the case of the multi-input model, a pre-trained backbone was used, initialized with the best-performing weights obtained from training the base DL regressors. To ensure strict separation between training and validation data, each instance of the multi-input model was initialized using weights from a regressor trained on the corresponding training subset. This approach preserved data independence across cross-validation folds and prevented information leakage during evaluation.

The hyperparameters used during the training of DL-based regressors are presented in Table 1. In the case of the k-NN regressor, the hyperparameters were selected using the GridSearchCV method and were as follows:

- distance metric: Manhattan,
- number of neighbors: 11,
- weighting method: distance.

Table 1. Hyperparameters for training the regression models – ResNet-18 (base DL regressor) and the proposed multi-input model.

Parameter	Base DL regressor	Multi-input regressor
Image size	256	256
Epochs	150	60
Batch size	96	96
Base learning rate	1e-3	1e-3
Learning rate scheduler	linear ($lr_f = 0.01$)	linear ($lr_f = 0.05$)
Weight decay	1e-4	1e-4
Optimizer	AdamW	AdamW
Beta 1	0.900	0.600
Beta 2	0.999	0.999

All models were trained using the same data augmentation techniques to improve generalization and robustness. The applied augmentation methods, designed to simulate variations in real-world conditions, are detailed in Table 2.

Table 2. Augmentation methods used during model training.

Augmentation type	Parameters
Hue modification	± 0.015
Saturation modification	± 0.7
Brightness modification	± 0.4
Contrast modification	± 0.3
Horizontal flip	probability = 0.5
Vertical flip	probability = 0.5
Erase	size = $[0.15, 0.35]$, probability = 0.20
Gaussian blur	kernel size = $[3, 7]$, probability = 0.25
Rotation	$\pm 15^\circ$
Translation	± 0.10
Scale	± 0.10

2.8. Evaluation of developed methods

As part of the experiments, the following methods were evaluated:

1. **Reference CV method (single image)** – estimation based on classical computer vision (CV) techniques applied to a single input image;
2. **Reference CV method (mean of N images)** – average of results obtained from classical methods applied to a batch of N images;

3. **DL regressor (single image)** – inference based on a single image using the deep learning regression model;
4. **DL regressor (mean of N images)** – mean prediction over N individual image inferences from the ResNet-based regressor;
5. **DL regressor + k-NN** – predictions from the ResNet model on N images passed as input features to a k-NN regressor;
6. **Multi-input regressor** – a custom model that jointly processes all N input images to produce a single regression result.

The experiment was conducted using a 5-fold cross-validation scheme. For each of the five validation subsets, the RMSE, MAE, coefficient of determination (R^2), and error rate were computed. The final results were reported as the mean values of these metrics across all folds, along with their corresponding standard deviations to reflect performance variability.

The input batch size was constrained by technical limitations. Increasing the number of input images per batch significantly impacts both data acquisition and processing time. Therefore, in the experiments described, the batch size was fixed at $N = 5$.

2.9. Evaluation of robustness to local anomalies

The developed solution is ultimately intended to analyze images of heterogeneous samples. Therefore, it must demonstrate robustness to potential disturbances, such as the presence of a localized region with elevated content of the analyzed component. To assess this capability, the selected processing methods were subjected to a robustness test involving a simulated local anomaly.

The robustness evaluation was conducted for DL methods analyzing batches of N input images, as described in Subsec. 2.8, specifically: (1) DL regressor + mean (N), (2) DL regressor + k-NN (N), (3) multi-input regressor (N).

A local anomaly was introduced by perturbing a single image within the input batch with one of the following anomalies: salt-and-pepper noise, simulated dust accumulation on the lens, varying lighting conditions, or reflection artifacts (Fig. 6). Salt-and-pepper noise is widely used to assess the robustness of machine learning models in computer vision tasks. It enables the simulation of sensor and data transmission errors [25], reveals overfitting to pristine data [26], and – most importantly for the architectures considered in this work – challenges the local feature sensitivity of CNNs [27]. For these reasons, salt-and-pepper noise was selected to model anomalies of varying severity. In addition, two additional types of local anomalies were introduced. Dust accumulation on the camera lens was simulated by overlaying a mask with randomly distributed blobs, where the controlled parameter was the proportion of pixels occluded by the blobs. Reflection artifacts were modeled by adding intensity values along edges detected

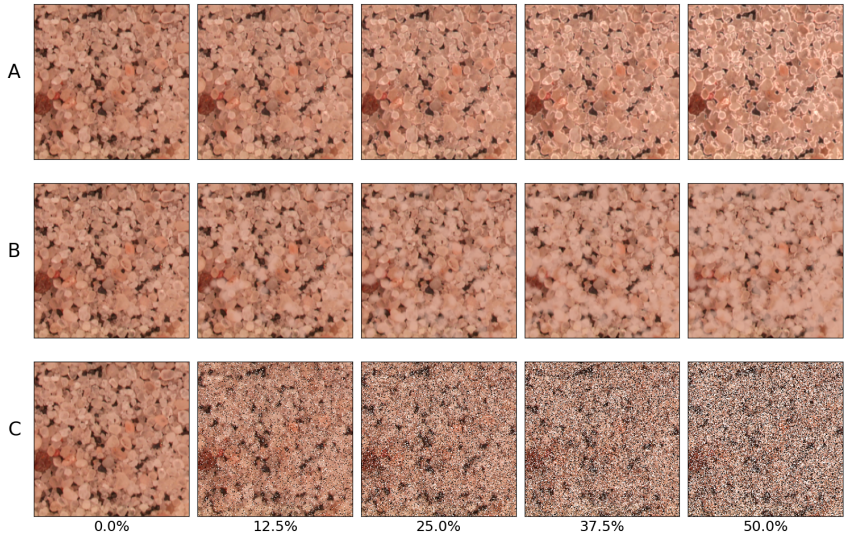

Fig. 6. An example input batch illustrating different types of local anomalies used in the robustness evaluation: reflection artifacts (A), dust accumulation (B), and salt-and-pepper noise (C) at varying intensity levels.

using the Canny algorithm, resulting in images exhibiting reflections of varying intensity along grain boundaries.

The experiment was conducted with varying anomaly intensities, ranging from 0 % (no modification) to 50 %. To to enable comparative assessment of robustness across models, the area under the MAE curve was computed over the investigated range of anomaly intensities. Each deep learning-based method was evaluated using 5-fold cross-validation. For each fold, robustness tests were performed independently, and the results were subsequently averaged. In addition, standard deviations were computed to quantify the variability of the results.

2.10. Evaluation of robustness to global anomalies

In addition to the local modification within the input batch, designed to evaluate the model's robustness to data inhomogeneity, an additional experiment was conducted to assess the model's resilience to variations in the entire input dataset. For this purpose, the anomalies described in Subsec. 2.9 were applied to all input images of the model. The experiment was repeated with anomaly intensities ranging from 0 % (no modification) to 50 %, and the area under the MAE curve was used as the evaluation metric. The robustness evaluation was conducted for two DL methods analyzing batches of N input images (multi-input regressor, DL regressor + k-NN), and, for comparison, a method analyzing a single image (DL regressor).

2.11. Model interpretation and contrast robustness

To explain and visualize the operating principles of the developed DL model, a gradient-based feature analysis approach was employed. For this purpose, deep feature maps were extracted from each layer of the encoder. Each feature tensor of shape $B \times C_i \times H_i \times W_i$ (batch size, number of channels, height, width) was subsequently averaged along the channel dimension, thereby reducing its size to $B \times H_i \times W_i$. This operation yields a two-dimensional spatial activation map for every input image and encoder layer, highlighting the regions most influential for feature representation.

Gradients of these activation maps with respect to the input images were then computed, providing insight into how variations in pixel intensity propagate through the network. The resulting gradient maps allow identification of the image regions the model relies on most strongly when forming its predictions. By examining these maps across consecutive layers, one can trace the progressive transformation of low-level contrast patterns into higher-level representations, thereby offering an interpretable view of the model's internal decision-making process.

Subsequently, an experiment was designed to evaluate the impact of contrast scaling in input images on the performance of the model. The methodology of this experiment follows the procedure described in Subsec. 2.10. The contrast scaling factor was varied in the range from 0 % (contrast suppression) to 200 % (contrast enhancement). For each scaling level, the MAE was computed, and the area under the MAE curve (AUC-MAE) was determined to provide a cumulative measure of robustness. In addition, gradient maps of deep features were visualized to qualitatively assess how contrast variations affect the internal feature representations of the network.

3. Results and discussion

3.1. Evaluation of encoder architectures

During the experiment, three baseline architectures were evaluated (Table 3). Among them, the ResNet-18 model achieved the highest performance, with

Table 3. Summary of results for the evaluated base architectures. Mean values and standard deviations obtained during cross-validation are given.

Method	RMSE	MAE	R^2	Error rate [%]	Inference time [ms]
ResNet-18	0.791 (0.058)	0.536 (0.035)	0.970 (0.004)	38.02 (2.38)	5.60 (3.30)
MobileNetV2	1.574 (0.036)	1.197 (0.020)	0.884 (0.006)	71.14 (2.05)	17.59 (4.30)
EfficientNet-B0	2.059 (0.102)	1.599 (0.086)	0.800 (0.020)	80.32 (1.10)	11.89 (3.52)

an RMSE of 0.791 (0.058) and an error rate of 38.02 (2.38) %. This architecture also demonstrated the shortest inference time, approximately 5.60 ms. Consequently, ResNet-18 was selected as the primary regression model for further experiments and subsequent development.

3.2. Evaluation of developed methods

The results obtained in the conducted experiments are summarized in Table 4. The simple reference method, based on classical computer vision techniques, yielded the weakest performance across all evaluation metrics. This was consistent both for single-image inference and when averaging predictions over batches of five images, with an RMSE of 2.573 (0.109), MAE of 1.929 (0.083), and an error rate of 78.16 (0.68) %. These findings underscore the inadequacy of relying solely on dark pixel ratio analysis in images for estimating ilmenite content. The poor performance can be attributed not only to the inherent heterogeneity of the samples but also to substantial grain overlapping which hinders accurate interpretation of image features.

Table 4. Summary of results for the developed method and reference methods. Mean values and standard deviations obtained during cross-validation are given.

Method	RMSE	MAE	R^2	Error rate [%]
Ref. CV method (1)	4.234 (0.147)	3.146 (0.102)	0.158 (0.059)	84.50 (1.30)
Ref. CV method + mean (N)	2.573 (0.109)	1.929 (0.083)	0.689 (0.026)	78.16 (0.68)
ResNet regressor (1)	0.836 (0.081)	0.564 (0.049)	0.966 (0.006)	39.52 (3.01)
ResNet regressor + mean (N)	0.444 (0.056)	0.319 (0.039)	0.990 (0.003)	21.38 (3.10)
ResNet regressor + k-NN (N)	0.392 (0.015)	0.187 (0.007)	0.993 (0.001)	16.82 (0.62)
Multi-input regressor (N)	0.190 (0.024)	0.107 (0.016)	0.998 (0.001)	3.06 (1.00)

In contrast, the use of a DL-based ResNet regressor significantly improved inference quality. Notably, the two-stage pipeline, where individual predictions from five input images were post-processed using a k-NN regressor, yielded markedly better results, with an average RMSE of 0.392 (0.015), MAE of 0.187 (0.007), and R^2 approaching 1. Nevertheless, despite the low mean error values, these average error rate remained at 16.82 (0.62) %, indicating that approximately one in five predictions deviated from the ground truth by more than 0.5 point. This level of error may still be unacceptable in practical, non-laboratory scenarios.

These limitations were effectively addressed through the adoption of the proposed multi-input regressor. This approach achieved the highest overall performance, with an RMSE of 0.190 (0.024), MAE of 0.107 (0.016), and a sig-

nificantly reduced average error rate of 3.06 (1.00) %. These results confirm the effectiveness of leveraging multi-image batches for robust and precise estimation of mineral content in complex, heterogeneous samples.

During inference on an edge device (NVIDIA Jetson AGX Orin), the model achieved an average execution time of 141.5 ms per sample. This performance indicates that the proposed approach is suitable for quasi-real-time applications, providing sufficiently fast predictions for practical deployment scenarios while maintaining portability to embedded hardware platforms.

3.3. Evaluation of robustness to local anomalies

The results of the robustness evaluation are presented in Fig. 7. The weakest performance was observed for the model that averaged predictions from a batch of input images. This method yielded the highest AUC-MAE, with average values of 0.219 (0.009), 0.219 (0.015), and 0.474 (0.060), indicating limited robustness to all localized disturbances.

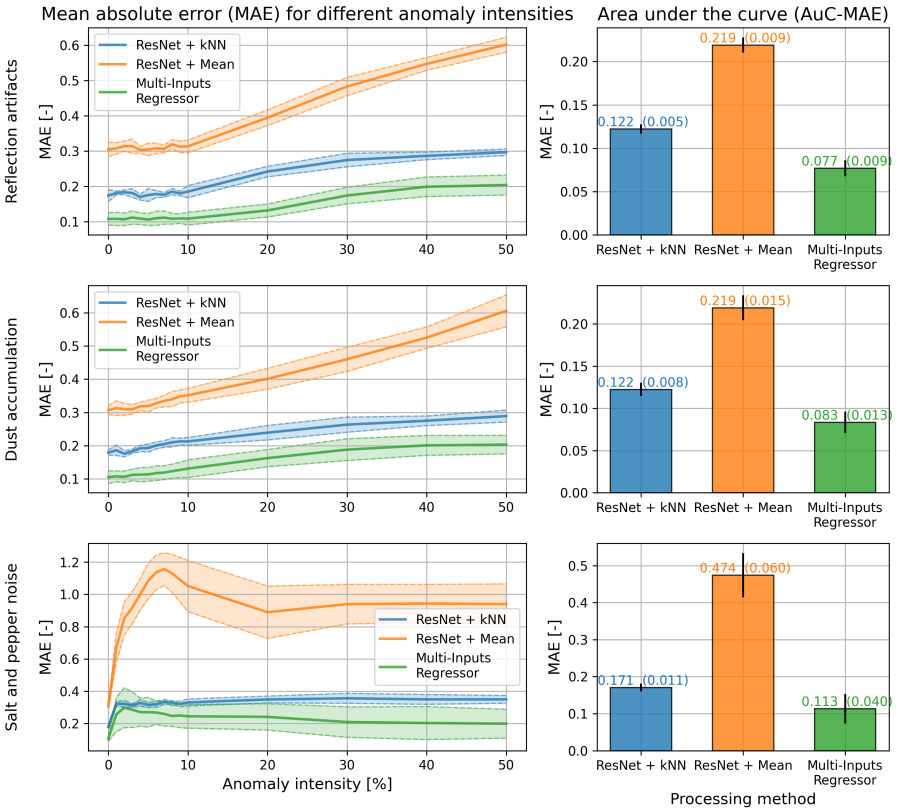

Fig. 7. MAE values for varying local anomaly intensities applied to a single image within the batch.

In contrast, the approach utilizing the k-NN regressor exhibited much higher resilience to the three simulated local anomalies, with MAE values of 0.122 (0.005), 0.122 (0.008), and 0.171 (0.011). This robustness can be attributed to the operating principle of nearest neighbor algorithms, which rely on consensus among multiple nearby data points, thereby mitigating the influence of outliers.

The multi-input model achieved error values even lower than those of the ResNet + k-NN variant, with AUC-MAE values of 0.077 (0.009), 0.083 (0.013), and 0.113 (0.040). These results confirm that explicitly modeling all input images jointly allows the architecture to better capture the aggregate context and suppress the impact of corrupted inputs.

3.4. Evaluation of robustness to global anomalies

The results of the robustness evaluation are presented in Fig. 8. In this case, the outcomes are less straightforward than for the local anomalies. For the 'Reflection' and 'Dust' perturbations, the baseline ResNet model (processing only a single image) achieved the weakest performance, with MAE values of

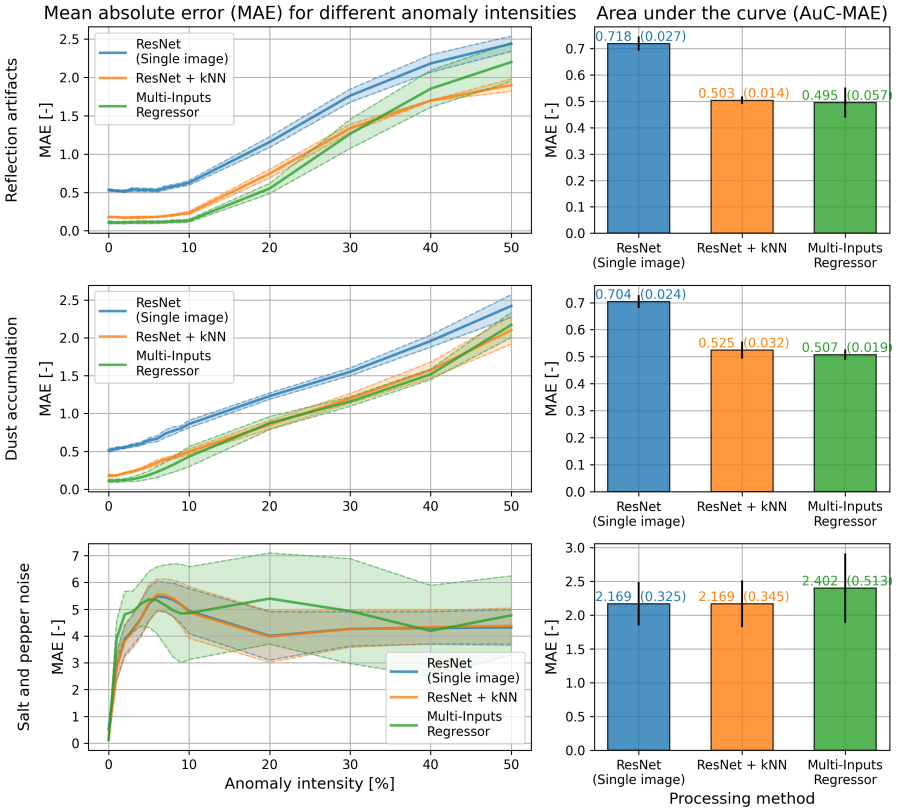

Fig. 8. MAE values for varying global anomaly intensities applied to all images in the batch.

0.718 (0.027) and 0.704 (0.024), respectively. The multi-input model performed noticeably better, obtaining significantly lower MAE values of 0.495 (0.057) and 0.507 (0.019), which are close to those achieved by the ResNet + k-NN method, namely 0.503 (0.014) and 0.525 (0.032). For the 'Noise' perturbation, all models yielded comparable results, with high MAE values above 2 and substantial standard deviations exceeding 0.3.

Considering the overall results, it can be concluded that approaches operating on multi-image input batches exhibit higher, or at least comparable, robustness to input perturbations. In particular, for the 'Dust' anomaly, the multi-input model reduced the MAE by nearly 40 % compared to the baseline ResNet, underscoring the effectiveness of exploiting joint information from multiple images.

3.5. Model interpretation and contrast robustness

The visualization of deep feature gradients is presented in Fig. 9. During inference, the model predominantly focuses on characteristic high-contrast blobs corresponding to ilmenite grains, including small grains that are partially covered. A reduction of contrast below 50 % has a strong negative impact on model performance, leading to the omission of a substantial portion of relevant regions. Conversely, when the contrast is increased to around 200 %, oversaturation effects become apparent, with activations extending to areas unrelated to ilmenite grains, but rather to interstitial spaces between particles of the secondary material.

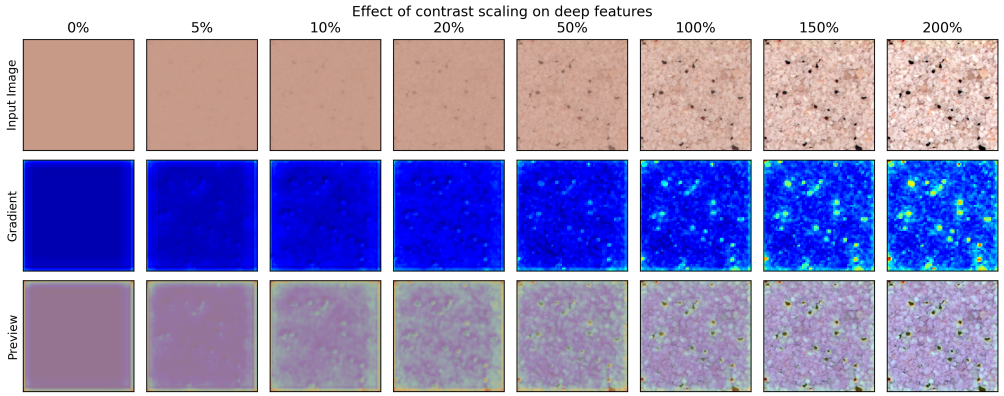

Fig. 9. Gradient-based visualization of deep features for different contrast scaling factors applied to the input images.

This trend is confirmed by the quantitative results in Fig. 10, which reveal distinct regions where the error remains lower. Notably, these regions approxi-

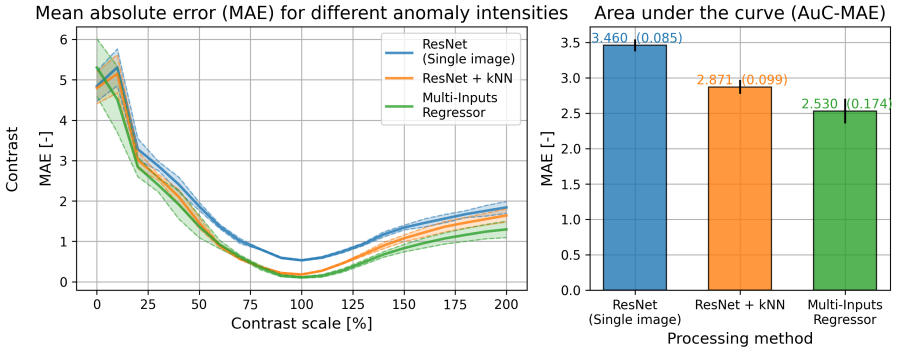

Fig. 10. MAE values for varying contrast scaling applied to all images in the batch.

mately overlap with the augmentation range applied to the training set during model development (Table 2, contrast ±30 %). A further observation concerns the comparative effect of contrast scaling across different processing methods: the single-image approach yielded a high AUC-MAE of 3.450 (0.085), whereas the multi-input method achieved a considerably lower value of 2.530 (0.174).

Overall, these findings underscore the importance of maintaining appropriate contrast conditions to ensure reliable and interpretable predictions, while also demonstrating the robustness advantage of leveraging multiple input images under varying contrast settings. It is worth emphasizing that the observed sensitivity to contrast variations can be mitigated by employing advanced data augmentation strategies and by extending the training dataset, thereby improving the model's generalization to diverse imaging conditions.

4. Conclusions

The architecture developed by the authors enables the processing of batches of multiple input images. As a result, it achieves not only high precision but also substantial robustness to the significant non-homogeneity of the analyzed data, with an RMSE of 0.190 (0.024) and an error rate of 3.06 (1.00) %. By employing the smallest model in the ResNet family as the encoder, the overall computational complexity remains low, making the solution well-suited for on-edge deployment on an autonomous platforms such as Martian rovers or planetary probes. This enables the use of multiple small, low-cost exploratory platforms equipped with *in situ* analysis capabilities during extraterrestrial missions, allowing for the rapid exploration and assessment of large planetary areas. Furthermore, the use of standard RGB imagery significantly reduces the cost of the acquisition system. The proposed solution can utilize simpler and more affordable cameras operating in the visible spectrum, which are often already inte-

grated into exploration platforms. It is important to emphasize that the proposed solution enables seamless replacement of the encoder with either larger or smaller deep models, thereby allowing the architecture to be tailored to the specific requirements and constraints of the target problem. This flexibility facilitates deployment across a wide range of scenarios, enhancing scalability in terms of both computational cost and predictive accuracy, and thus ensuring adaptability to diverse industrial applications.

While the proposed microscopic, machine-learning-based analysis demonstrates promising performance, it also exhibits several acknowledged limitations. Its non-invasive nature confines the analysis to surface-level observations only. Consequently, regions containing significant ilmenite concentrations beneath the lunar surface may remain undetected, posing a risk to the effective planning and execution of lunar mining operations. Furthermore, although the ilmenite–regolith binary mixture is generally distinguishable due to the strong color contrast between ilmenite grains and the surrounding material, this advantage is not universal. In scenarios where both components exhibit similar visual characteristics within the visible light spectrum, the proposed method may encounter difficulties in distinguishing between them, potentially resulting in inaccurate or failed estimations of elemental content. Additionally, when considering deployment of the proposed approach beyond controlled laboratory setting, several important limitations must be acknowledged. The accuracy of the method may degrade substantially under variations in the optical or acquisition setup that alter the appearance of input images. These challenges underscore the importance of careful calibration and, where necessary, domain adaptation when transferring the method to different instruments or operating conditions.

The above mentioned limitations highlight opportunities for further advancement in machine learning-based visual mineral recognition. A logical progression would involve extending the analysis to heterogeneous mixtures. As most regions on celestial bodies contain more than two mineral components, implementing a multi-output regression framework to estimate the abundance of each constituent could be crucial. The proposed architecture could be easily adapted to a multi-target regression task by a simple modification of the final fully-connected layer.

Another promising direction involves the use of hyperspectral imaging. While standard microscopic imaging may suffice for some binary or moderately heterogeneous mixtures, it becomes inadequate when components share similar characteristics within the visible spectrum. Hyperspectral cameras overcome this limitation by capturing data across a wide range of wavelengths, thus significantly enhancing the system's ability to distinguish visually similar minerals. Due to the high flexibility of the proposed architecture, it can be readily adapted for the processing of multi- or hyperspectral data. The use of such input data

would address the aforementioned challenges related to low distinguishability between mixture components. An architecture extended in this way could be easily adapted to other regression tasks, including multi-output regression, offering a wide range of potential applications beyond the analysis of extraterrestrial materials.

Acknowledgements

We would like to express our sincere gratitude to Jakub Mazur from Wrocław University of Science and Technology for his valuable support and assistance to the Project SCORPIO team during both the preparation for and participation in the Australian Rover Challenge.

Conflict of interest

The authors declare no conflicts of interest.

Data availability

The data that support the findings of this study are openly available in RepOD at https://doi.org/10.18150/9RPOUT.

Author contribution

K. Marciniak: conceptualization, methodology, software, validation, formal analysis, investigation, writing – original draft, visualization; M. Grzesiak: conceptualization, methodology, validation, formal analysis, investigation, writing – original draft; I. Zaworski: conceptualization, software, resources, writing – original draft; D. Pawliszewski: investigation, resources, data curation, writing – original draft; D. Zygarlicka: investigation, resources, data curation, writing – original draft; M. Wnuk: investigation, resources, data curation, writing – original draft; A. Zakrzewski: writing – review & editing, supervision, project administration, funding acquisition.

References

1. Henley S., Allington R., PERC, CRIRSCO, and UNFC: Minerals reporting standards and classifications, *European Geologist*, **36**: 49–54, 2013.
2. Sommariva A., Gori L., Chizzolini B., Pianorsi M., The economics of Moon mining, *Acta Astronautica*, **170**: 712–718, 2020, https://doi.org/10.1016/j.actaastro.2020.01.042.
3. Levinson A., Taylor R., *Moon Rocks and Minerals*, Pergamon Press, New York, 1989.

4. McKay D., Williams R., A geologic assessment of potential lunar ores, [in:] *Space Resources and Space Settlements*, J. Billingham, W. Gilbreath, B. O'Leary [Eds], NASA SP-428, pp. 243–255, NASA Ames Research Center, 1979.

5. Hutson M., *Notes of Lunar Ilmenite*, Lunar and Planetary Laboratory, The University of Arizona, pp. A1–A6, 1989.

6. Sverdrup H.U., Sverdrup A.E., An assessment of the global supply, recycling, stocks in use and market price for titanium using the WORLD7 model, *Sustainable Horizons*, **7**: 100067, 2023, https://doi.org/10.1016/j.horiz.2023.100067.

7. Carr B.B., Recovery of water or oxygen by reduction of lunar rock, *AIAA Journal*, **1**(4): 921–924, 1963.

8. Paige D.A. *et al.*, The lunar reconnaissance orbiter diviner lunar radiometer experiment, *Space Science Review*, **150**: 125–160, 2009, https://doi.org/10.1007/s11214-009-9529-2.

9. Pieters C.M. *et al.*, The Moon Mineralogy Mapper (M^3) on Chandrayaan-1, *Current Science*, **96**(4): 500–505, 2008.

10. Surkov Y., Shkuratov Y., Kaydash V., Korokhin V., Videen G., Lunar ilmenite content as assessed by improved Chandrayaan-1 M^3 data, *Icarus*, **341**: 113661, 2020, https://doi.org/10.1016/j.icarus.2020.113661.

11. Charette M.P., McCord T.B., Pieters C., Adams J.B., Application of remote spectral reflectance measurements to lunar geology classification and determination of titanium content of lunar soils, *Solid Earth and Planets*, **79**(11): 1605–1613, 1974, https://doi.org/10.1029/JB079i011p01605.

12. Lucey P.G., Blewett D.T., Jolliff B.L., Lunar iron and titanium abundance algorithms based on final processing Clementine ultraviolet-visible images, *Journal of Geophysical Research*, **105**: 20297–20305, 2000, https://doi.org/10.1029/1999JE001117.

13. Palafox L.F., Hamilton C.W., Scheidt S.P., Alvarez A.M., Automated detection of geological landforms on Mars using Convolutional Neural Networks, *Computers & Geosciences*, **101**: 48–56, 2017, https://doi.org/10.1016/j.cageo.2016.12.015.

14. Li J. *et al.*, Autonomous Martian rock image classification based on transfer deep learning methods, *Earth Science Informatics*, **13**: 951–963, 2020, https://doi.org/10.1007/s12145-019-00433-9.

15. Zhou L., Liu Z., Wang W., Terrain classification algorithm for Lunar rover using a deep ensemble network with high-resolution features and interdependencies between channels, *Wireless Communications and Mobile Computing*, **2020**: 8842227, 2020, https://doi.org/10.1155/2020/8842227.

16. Lv F. *et al.*, Highly accurate visual method of Mars terrain classification for rovers based on novel image features, *Entropy*, **24**(9): 1304, 2022, https://doi.org/10.3390/e24091304.

17. Razzell Hollis J. *et al.*, The power of paired proximity science observations: Co-located data from SHERLOC and PIXL on Mars, *Icarus*, **387**: 115179, 2022, https://doi.org/10.1016/j.icarus.2022.115179.

18. Edgett K.S. *et al.*, *Curiosity's robotic arm-mounted Mars Hand Lens Imager (MAHLI): Characterization and calibration status*, MSL MAHLI Technical Report 0001, Mars Science Laboratory, 2015, https://doi.org/10.13140/RG.2.1.3798.5447.

19. Herkenhoff K.E. *et al.*, Evidence from opportunity's microscopic imager for Water on Meridiani Planum, *Science*, **306**(5702): 1727–1730, 2004, https://doi.org/10.1126/science.1105286.

20. Xu T.Y., Hapke B., Zhang X.P., Wu Y.Z., Lu X.P., Micro-scale photometry of the Moon using Chang'E-3 Panoramic Camera (PCAM), *Astronomy & Astrophysics*, **665**: A15, 2022, https://doi.org/10.1051/0004-6361/202143012.

21. Zhong J., Yan J., Li M., Barriot J.-P., A deep learning-based local feature extraction method for improved image matching and surface reconstruction from Yutu-2 PCAM images on the Moon, *ISPRS Journal of Photogrammetry and Remote Sensing*, **206**: 16–29, 2023, https://doi.org/10.1016/j.isprsjprs.2023.10.021.

22. Els S. *et al.*, The microscope camera CAM-M on-board the Rashid-1 Lunar rover, *Space Science Reviews*, **220**: 81, 2024, https://doi.org/10.1007/s11214-024-01117-7.

23. Oberbeck V.R., Hoerz F., Quaide W., Morrison R.H., *Emplacement of the Caley Formation*, Technical Report, NASA, 1973.

24. Head J.W. *et al.*, Lunar mare basaltic Volcanism: Volcanic features and emplacement processes, *Reviews in Mineralogy and Geochemistry*, **89**(1): 453–507, 2023, https://doi.org/10.2138/rmg.2023.89.11.

25. Steffens C.R., Messias L.R.V., Drews P.L.J., Botelho S.S.d.C., Can exposure, noise and compression affect image recognition? An assessment of the impacts on state-of-the-art ConvNets, [in:] *2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE)*, Rio Grande, Brazil, pp. 61–66, 2019, https://doi.org/10.1109/LARS-SBR-WRE48964.2019.00019.

26. Seals M., *On the Robustness of Object Detection Based Deep Learning Models*, MA thesis, University of Tennessee, 2019.

27. Dai W., Berleant D., Benchmarking robustness of deep learning classifiers using two-factor perturbation, [in:] *2021 IEEE International Conference on Big Data (Big Data)*, Orlando, FL, USA, pp. 5085–5094, 2021, https://doi.org/10.1109/BigData52589.2021.9671976.

Received June 2, 2025; revised September 22, 2025; accepted November 24, 2025; version of record December 16, 2025; published issue June 17, 2026.

Computer Assisted Methods in Engineering and Science
33(2): 125–145, 2026, https://doi.org/10.24423/cames.2026.1874

Isogeometric Topology Optimization Based on Topological Derivatives*)

Guilherme Henrique TEIXEIRA[1)*], Nepomuk KRENN[2)], Peter GANGL[2)], Benjamin MARUSSIG[1)]

[1)] *Graz University of Technology, Institute of Applied Mechanics, Graz, Austria*
[2)] *Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria*

[*] *Corresponding Author: teixeira@tugraz.at*

Topology optimization is a valuable tool in engineering, facilitating the design of optimized structures. However, topological changes often require a remeshing step, which can become challenging. In this work, we propose an isogeometric approach to topology optimization driven by topological derivatives. The combination of a level-set method together with an immersed isogeometric framework allows seamless geometry updates without the necessity of remeshing. At the same time, topological derivatives provide topological modifications without the need to define initial holes [7]. We investigate the influence of higher-degree basis functions in both the level-set representation and the approximation of the solution. Two numerical examples demonstrate the proposed approach, showing that employing higher-degree basis functions for approximating the solution improves accuracy, while linear basis functions remain sufficient for the level-set function representation.

Keywords: topology optimization, isogeometric analysis, topological derivatives, level-set method, immersed methods, higher-degree basis function.

1. Introduction

Design optimization describes an iterative process to define the optimal geometry of a structure given some constraints. This problem can be approached in several ways, including the optimization of geometric parameters, such as radius, length, or width (Fig. 1a), the boundaries of the shape (Fig. 1b) [46], or the material distribution of the structure (Fig. 1c) [26]. The last one is known as topology optimization, and since it was first introduced in [9], several methods

[*)] The results of this study were presented at GAMM 2025 – 95th Annual Meeting of the International Association of Applied Mathematics and Mechanics, April 7–11, 2025, Poznań, Poland.

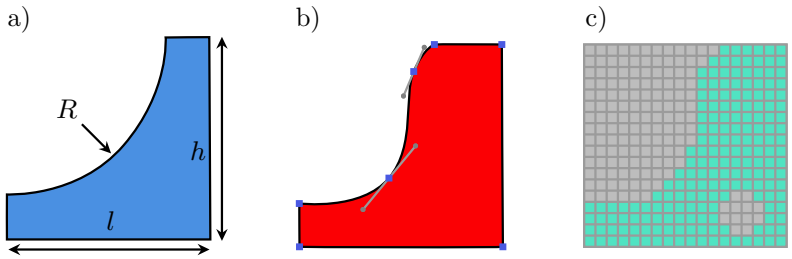

Fig. 1. Different types of optimization: a) parameter optimization, b) shape optimization, c) topology optimization.

have been developed to approach the problem, and as a consequence, various challenges have been addressed [36]. The most popular classes of topology optimization methods are based on design representations by means of density or level-set functions.

Density-based approaches, widely used in topology optimization [36], represent the design through a density variable equal to 1 for material and 0 for void. These methods can be used on a fixed background mesh on which the density variable is defined element-wise, identifying whether an element is solid or void, thereby avoiding the need for remeshing during the optimization process. However, a well-known challenge of density-based topology optimization is the presence of unphysical intermediate materials. This issue is overcome through penalization techniques, such as the solid isotropic material with penalization (SIMP), removing gray areas and resulting in nearly black-and-white designs [13].

In level-set-based approaches [29], the material distribution is represented by the sign of a continuous level-set function which evolves in the course of the optimization process. Level- set methods have been extensively applied to shape and topology optimization, either coupled with a remeshing strategy [15] or applied on a fixed background mesh [2, 3, 44]. The evolution of the level-set function is guided either by shape derivatives via a Hamilton–Jacobi transport equation [2] or by topological derivatives [7]. In the former case, the resulting design heavily depends on the initial topology. New holes cannot be directly introduced, as the method relies on the holes in the initial geometry, which can then be merged or cancelled in the optimization process. In order to allow for the nucleation of new holes, the method has been coupled with topological derivatives in [1, 10]. Another possible approach combining the level-set method and topological derivatives without solving the Hamilton–Jacobi is proposed in [7], with many applications reported in the literature [8, 11, 16, 25, 27, 28, 47] and has also been extended to multi-material problems in [17]. In this algorithm, the optimization is guided only by the topological derivative.

Isogeometric analysis (IGA), first introduced in [21], presents the concept of connecting design and analysis using the same B-splines representing the geometry as basis functions. The straightforward control over the degree and smoothness of a B-spline basis is quite valuable for numerical simulations. Several research studies have been conducted using the isogeometric concept in various topology optimization approaches. On the one hand, applications of the density method in the context of IGA have been reported in several studies in the literature [20], where a connection between the optimization and the CAD environments [34], as well as the benefits of refinement schemes, are straightforwardly targeted into the approach [38, 42]. On the other hand, the conventional level-set method has been used with different discretizations for the level-set function, such as radial basis functions [4, 35], B-splines [22], or piecewise constant functions [23]. To the best of our knowledge, the combination of IGA and topological-derivative-based level-set optimization has only been considered in [30], where the conventional shape-derivative-based level-set method extended by a topological derivative term was used.

In this work, we apply the approach of [7] within the isogeometric framework, using B-splines both for the level-set function discretization and as basis functions to approximate the solution. Compared to standard level-set methods, this approach has the following advantages. Since it uses a fixed background mesh, it eliminates the need for remeshing, and, by using the topological derivative to guide the evolution of the level-set function, it also removes the necessity of defining initial holes and does not require solving the Hamilton–Jacobi equation, relying only on the topological derivative to guide the evolution of the level-set function. In addition, compared to standard density-based methods, this approach does not introduce artificial intermediate materials, since it uses a level-set function to describe the design.

Our contribution focuses on the combination of the level-set method, topological derivatives, and IGA, providing initial configurations free from geometric approximation errors and a simplified fixed high-degree background mesh defined by the knot vector and control points, which also allows straightforward higher-degree simulations. One challenge to this approach is the handling of the material interface, since we use a fixed mesh that does not adapt to the level-set function. We study the sensitivity of the level-set representation in the optimized topology due to different polynomial degrees used for approximating the level-set function and the solution. To accomplish this, we investigate two different settings: one with the same polynomial degree for both approximations, and another one with a linear level-set function discretization and a higher-degree approximation of the solution.

Therefore, in Sec. 2, we present the linear elasticity problem investigated in the topology optimization approach and the main considerations for applying

IGA to it. Then, the definition of the level-set function and the derived topological derivative are provided in Secs. 3 and 4, respectively. In addition, Sec. 5 addresses the procedures for dealing with the cut elements, and two numerical results applied to linear elasticity problems are shown in Sec. 6. Finally, in the conclusion, we summarize the main findings of the numerical examples.

2. Problem description

In this section, we present an overview of the problem to which our approach is applied. The problem studied is the compliance minimization in two-dimensional linear elasticity, and here we present the governing equation, the formulation of the minimization problem, and the key considerations for applying IGA using an immersed approach based on the level-set method.

2.1. Linear elasticity problem

In this work, we consider linear elasticity problems defined on a domain Ω, where the boundary $\partial\Omega$ is divided into three disjoint parts, such that $\Gamma_D \cap \Gamma_N = \emptyset$, $\Gamma_D \cap \Gamma_0 = \emptyset$, $\Gamma_N \cap \Gamma_0 = \emptyset$ and $\Gamma_D \cap \Gamma_N \cap \Gamma_0 = \emptyset$, as shown in Fig. 2a. On these three parts of the boundary, Dirichlet boundary conditions are applied on Γ_D, Neumann boundary conditions in Γ_N, and zero Neumann boundary conditions are applied on Γ_0. Therefore, the governing equations for the problem are given by:

$$\begin{cases} -\nabla \cdot (\boldsymbol{\sigma}(\mathbf{u})) = \mathbf{0} & \text{in } \Omega, \\ \mathbf{u} = \mathbf{0} & \text{on } \Gamma_D, \\ \boldsymbol{\sigma}(\mathbf{u}) \cdot \mathbf{n} = \boldsymbol{\tau} & \text{on } \Gamma_N, \\ \boldsymbol{\sigma}(\mathbf{u}) \cdot \mathbf{n} = \mathbf{0} & \text{on } \Gamma_0. \end{cases} \tag{1}$$

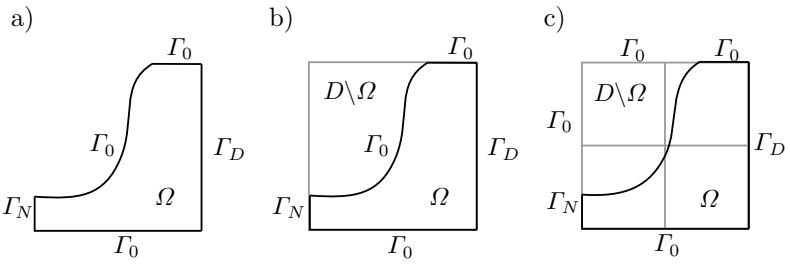

Fig. 2. Representation of the domain problem: a) domain Ω defined by the level-set function, b) domain Ω as a subset of domain D, c) domain Ω inside a B-spline background mesh defined by the knot vector of the geometry D.

Here, $\boldsymbol{\tau}$ is the load applied on the boundaries and $\mathbf{n}$ is the normal vector on them. In addition, $\mathbf{u}$ represents the displacement field and $\boldsymbol{\sigma}$ is the stress tensor, which, for linear elasticity and isotropic materials, is defined as

$$\boldsymbol{\sigma}(\mathbf{u}) = 2\mu\boldsymbol{\epsilon}(\mathbf{u}) + \lambda \operatorname{tr}(\boldsymbol{\epsilon}(\mathbf{u}))\mathbf{I},$$

where

$$\boldsymbol{\epsilon}(\mathbf{u}) = \frac{1}{2}\left(\nabla\mathbf{u} + (\nabla\mathbf{u})^T\right)$$

is the strain tensor and, for 2D plane strain,

$$\mu = \frac{E}{2(1+\nu)}, \qquad \lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}$$

are the Lamé parameters, written in terms of Young's modulus E and Poisson's ratio ν.

In this scenario, the domain Ω represents the material distribution of the geometry and is a subset of a larger domain D. This situation is graphically represented by Fig. 2b.

2.2. Immersed isogeometric approach

The goal of the topology optimization is to find an optimal material distribution Ω under given constraints, such as boundary conditions or area penalization. This can be formulated as a minimization problem. In this scenario, the domain Ω changes during the optimization process, and solving the problem numerically would require redefining the mesh at each iteration. To avoid the necessity of remeshing, we formulate the problem on the fixed domain D, as shown in Fig. 2b, instead of Ω, shown in Fig. 2a, and we introduce a material property alpha, which is equal to α_{in} if it is inside Ω and a small value α_{out} for outside. This approach is based on immersed methods, which are extensively applied in fluid mechanics, solid mechanics, interface problems, and several other areas. An extensive explanation of immersed methods and their aspects can be found in [12, 41, 45]. In this way, this approach allows us to define which part of D represents Ω, so that the governing equations for the linear elasticity problem can be rewritten as:

$$\begin{cases} -\nabla\cdot(\alpha_\Omega\boldsymbol{\sigma}(\mathbf{u})) = \mathbf{0} & \text{in } D, \\ \mathbf{u} = \mathbf{0} & \text{on } \Gamma_D, \\ \boldsymbol{\sigma}(\mathbf{u})\cdot\mathbf{n} = \boldsymbol{\tau} & \text{on } \Gamma_N, \\ \boldsymbol{\sigma}(\mathbf{u})\cdot\mathbf{n} = \mathbf{0} & \text{on } \Gamma_0, \end{cases} \quad \text{with} \quad \alpha_\Omega = \begin{cases} \alpha_{\text{in}} & \text{in } \Omega, \\ \alpha_{\text{out}} & \text{on } D\setminus\overline{\Omega}, \end{cases} \tag{2}$$

where $\alpha_{\text{out}} \ll 1$ is a penalization parameter for the void, small enough to neglect the basis functions located outside the domain Ω, but not too small to result in an ill-conditioned stiffness matrix [33].

Therefore, to finally solve the problem numerically, we discretize the domain D using a background mesh, as shown in Fig. 2c, where the basis functions used to approximate the solution field are B-splines of degree p, refined from the geometry D. These basis functions are constructed from a non-decreasing set of coordinates called knot vector

$$\Xi = \{\xi_1, \xi_2, ..., \xi_{n+p}, \xi_{n+p+1}\} \tag{3}$$

defined in a parameter space $\mathbb{P} = [\xi_1, \xi_{n+p+1}]$ of B-splines, where n is the number of basis functions and p is the polynomial degree. This construction is defined recursively starting from piecewise constant functions for $p = 0$:

$$B_{i,0} = \begin{cases} 1 & \text{if } \xi_i \leq \xi < \xi_{i+1}, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

and extended for higher degrees $p > 0$ by applying the Cox–de Boor formula [21]

$$B_{i,p} = \frac{\xi - \xi_i}{\xi_{i+p} - \xi_i} B_{i,p-1}(\xi) + \frac{\xi_{i+p+1} - \xi}{\xi_{i+p+1} - \xi_{i+1}} B_{i+1,p-1}(\xi). \tag{5}$$

In sequence, having the basis functions in each parametric direction, the geometry mapping from the parametric domain $\mathbb{P}^2 = [\xi_1, \xi_{n+p+1}] \times [\eta_1, \eta_{m+p+1}]$ to the physical domain $\mathbb{R}^2$ is then defined as

$$\mathbf{x}(\xi, \eta) = \sum_{i=1}^{n} \sum_{j=1}^{m} B_{i,p}(\xi) B_{j,p}(\eta) \mathbf{C}_{i,j}, \tag{6}$$

where $\mathbf{C}_{i,j}$ are the control points that define the geometry.

Note that the basis functions used to approximate the solution field may have a different polynomial degree p than those used to construct the geometry D. However, the geometric mapping remains based on the B-splines defined from the geometry D. Finally, the implementation of the problem is made within an open-source IGA code [43], which provides the necessary features.

2.3. Minimization problem

The minimization problem mentioned in the previous subsection, which defines the topology optimization process, has aims to find the optimal domain $\Omega \subset D$ that minimizes a cost function J. This expression can be written as:

$$\min_{\Omega \in \mathcal{E}} J(\Omega, \mathbf{u}), \tag{7}$$

where $\mathcal{E}$ is a set of admissible subsets of D, and

$$J(\Omega, \mathbf{u}) = \int_D \alpha_\Omega \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\epsilon}(\mathbf{u}) \, dD + l \int_\Omega \, d\Omega. \tag{8}$$

Note that the area constraint, which prevents the solution from being $\Omega = D$, is addressed by the second term of the objective function and is controlled by the parameter l [24].

3. Discretized level-set representation

The domain Ω is represented by a continuous level-set function. This means that the interface that divides the material region Ω from the void region $D \setminus \Omega$ is defined by the zero set of the level-set function. Specifically, all points where the level-set function is smaller than 0 belong to Ω, points where it is higher than zero belong to the void $D \setminus \Omega$, and points where the level-set function is equal to zero lie on the interface between the two regions:

$$\begin{cases} \phi(\mathbf{x}) < 0 & \iff \mathbf{x} \in \Omega, \\ \phi(\mathbf{x}) = 0 & \iff \mathbf{x} \in \partial\Omega, \\ \phi(\mathbf{x}) > 0 & \iff \mathbf{x} \in D \setminus \overline{\Omega}. \end{cases} \tag{9}$$

As the background mesh D is discretized, we also discretize the domain Ω. Therefore, the level-set discretization is made using B-spline basis functions of degree d, which might be the same or different from the degree p of the basis functions used for approximating the solution,

$$\phi(\xi, \eta) = \sum_{i=1}^{n} \sum_{j=1}^{m} B_{i,d}(\xi) B_{j,d}(\eta) c_{i,j}. \tag{10}$$

In sequence, the coefficients $c_{i,j}$ are obtained by solving a collocation problem, which enforces that the discretized level-set function (10) matches the initial level-set function at the Greville abscissas. The positions of the Greville abscissas, shown in Fig. 3, serve as anchors for the B-splines and are computed as

$$\widetilde{\xi}_i = \frac{\xi_{i+1} + \xi_{i+2} + \ldots + \xi_{i+d}}{d}, \qquad i = 1, \ldots, n. \tag{11}$$

During the optimization process, the evaluations of the level-set function at the Greville abscissas are updated, and the coefficients $c_{i,j}$ for the new level-set function discretization are obtained in the same process.

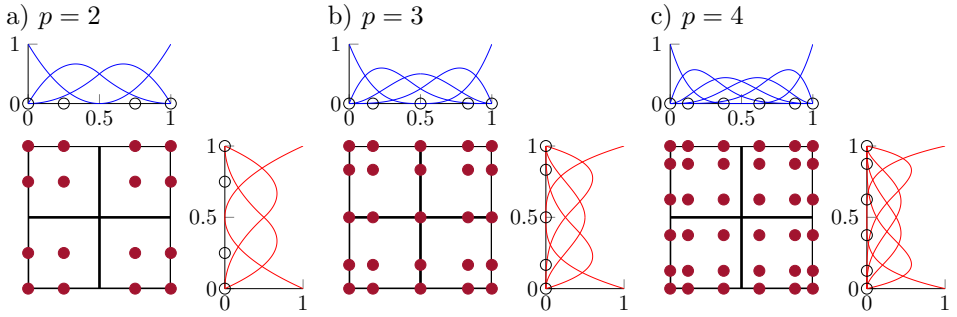

Fig. 3. Distribution of the Greville abscissae on the elements for different polynomial degrees, with basis functions defined as: a) $\Xi = \{0\,0\,0\,0.5\,1\,1\,1\}$, b) $\Xi = \{0\,0\,0\,0\,0.5\,1\,1\,1\,1\}$, c) $\Xi = \{0\,0\,0\,0\,0\,0.5\,1\,1\,1\,1\,1\,1\}$.

In Fig. 4, we observe the identification process of the region where the element is located. This is achieved by evaluating the level-set function at a set of points in each element. Then, based on the signs of these evaluations, we can identify the region of the element and assign the corresponding material property. For cut elements, the material property is computed as the average based on the cut ratio of the element, given by

$$\alpha|_T = \alpha_{\text{out}} + \frac{|T \cap \Omega|}{|T|}(\alpha_{\text{in}} - \alpha_{\text{out}}). \tag{12}$$

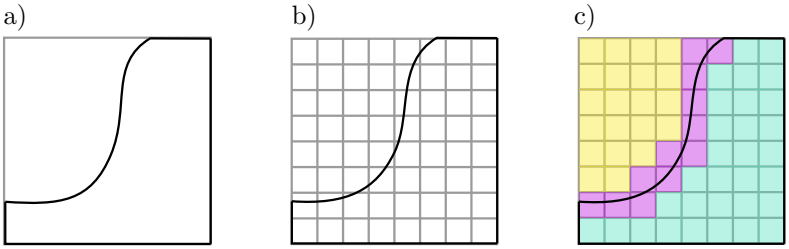

Fig. 4. Type identification of the elements for assembling the material property α: a) domain D divided into two regions by a level-set function, b) domains D discretized as the background mesh, c) identification of the elements; the elements outside Ω are in yellow, inside Ω in blue, and the cut elements in pink.

4. Topological derivatives

In the previous sections, we defined the representation of a fixed domain D and a level-set function ϕ used to represent the domain Ω, both discretized by B-splines. In the following, we discuss how the topological derivative is computed and how the level-set function is updated, focusing on minimizing the cost function.

Consider a point $\mathbf{x}_0 \in D \setminus \partial\Omega$, and define $\omega_\varepsilon(\mathbf{x}_o) = \{\mathbf{x} \in \mathbb{R}^2 : ||\mathbf{x}_0 - \mathbf{x}|| < \varepsilon\}$ as a circular perturbation of radius ε centered at $\mathbf{x}_0$. The introduction of the perturbation $\omega_\varepsilon(\mathbf{x}_o)$ in the domain D results in a perturbed domain:

$$\Omega_\varepsilon = \begin{cases} \Omega \setminus \overline{\omega_\varepsilon(\mathbf{x}_0)} & \text{if } \mathbf{x}_0 \in \Omega, \\ \Omega \cup \omega_\varepsilon(\mathbf{x}_0) & \text{if } \mathbf{x}_0 \in D \setminus \overline{\Omega}. \end{cases} \tag{13}$$

Let $\mathcal{J}(\Omega) := J(\Omega, \mathbf{u}(\Omega))$ denote the reduced cost function where $\mathbf{u}(\Omega)$ denotes the unique solution to (1) for a given subdomain Ω. In this scenario, to measure the change in the cost function J when a new hole around the point x_0 is introduced, the topological derivative is defined as

$$d\mathcal{J}(\Omega)(\mathbf{x}_0) := \lim_{\varepsilon \to 0} \frac{\mathcal{J}(\Omega_\varepsilon) - \mathcal{J}(\Omega)}{|\omega_\varepsilon|} = \lim_{\varepsilon \to 0} \frac{J(\Omega_\varepsilon, \mathbf{u}_\varepsilon) - J(\Omega, \mathbf{u})}{|\omega_\varepsilon|},$$

where $\mathbf{u}_\varepsilon$ is the solution of (1) with Ω replaced by Ω_ε.

To evaluate this expression, we adopt the approach proposed in [18], which introduces the Lagrangian

$$L(\Omega, \mathbf{u}, \boldsymbol{\lambda}) = J(\Omega, \mathbf{u}) + \boldsymbol{\lambda} E(\Omega, \mathbf{u}),$$

where $E(\Omega, \mathbf{u}) = 0$ represents the weak form of the governing equation. This implies that $L(\Omega, \mathbf{u}, \boldsymbol{\lambda}) = J(\Omega, \mathbf{u})$ at the solution for all $\boldsymbol{\lambda}$. Consequently, the topological derivative can be rewritten as

$$d\mathcal{J}(\Omega)(\mathbf{x}_0) = \lim_{\varepsilon \to 0} \frac{L(\Omega_\varepsilon, \mathbf{u}_\varepsilon, \boldsymbol{\lambda}) - L(\Omega, \mathbf{u}, \boldsymbol{\lambda})}{|\omega_\varepsilon|}. \tag{14}$$

By introducing in the adjoint state $\boldsymbol{\lambda}$ defined as the solution of $\partial_{\mathbf{u}} L(\Omega, \mathbf{u}, \boldsymbol{\lambda}) = 0$, and noting that $\boldsymbol{\lambda} = -\frac{1}{2}\mathbf{u}$, after solving this limit for the linear elasticity problem, as shown in [6], an analytical expression is obtained, which depends only on the solution $\mathbf{u}$ and the material coefficient α:

$$d\mathcal{J}(\Omega)(\mathbf{x}_0) = \begin{cases} d\mathcal{J}_{\text{in}}(\Omega)(\mathbf{x}_0) = -3\alpha_{\text{in}} \left(\dfrac{\alpha_{\text{out}} - \alpha_{\text{in}}}{2\alpha_{\text{out}} + \alpha_{\text{in}}} \right) \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\epsilon}(\mathbf{u}) - l \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \mathbf{x}_0 \in \Omega, \\ d\mathcal{J}_{\text{out}}(\Omega)(\mathbf{x}_0) = -3\alpha_{\text{out}} \left(\dfrac{\alpha_{\text{in}} - \alpha_{\text{out}}}{2\alpha_{\text{in}} + \alpha_{\text{out}}} \right) \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\epsilon}(\mathbf{u}) + l \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \mathbf{x}_0 \in D \setminus \overline{\Omega}. \end{cases} \tag{15}$$

From this, the generalized topological derivative is then defined as:

$$g_\Omega(\mathbf{x}) = \begin{cases} -d\mathcal{J}(\Omega)(\mathbf{x}) & \text{if } \mathbf{x} \in \Omega, \\ d\mathcal{J}(\Omega)(\mathbf{x}) & \text{if } \mathbf{x} \in D \setminus \overline{\Omega}, \end{cases} \tag{16}$$

and is used to update the level-set function, guiding the evolution of the domain Ω. Algorithm 1 shows this update process. The update of the level-set is guided under a spherical linear interpolation, which uses the angle θ_i, in the L^2-sense, between the current level-set ϕ_i and the topological derivative g_i, as a parameter to define the next domain Ω_{i+1}. Note that the stopping criterion is controlled by the same angle θ_i, and this quantity works as a comparison between the current topological derivative g_i and the level-set function ϕ_i. Then, if $\theta_i = 0$, the domain Ω_{i+1} is optimal and the topological derivative g_i can be used as the level-set function ϕ [7]. During this process, we apply a line search to define the parameter κ used to update the level-set in the spherical linear interpolation. Additional filtering processes are applied, described in Sec. 5, to smooth the generalized topological derivative g, working similarly to sensitivity filtering in density-based optimization [19].

Algorithm 1: Level-set update.

Initialize the level-set function ϕ_1;
for $i \leftarrow 1$ ***to*** $n_{\max}$ **do**
  Compute $g_{\Omega_i}(\mathbf{x}) = \begin{cases} -d\mathcal{J}(\Omega_i)(\mathbf{x}) & \text{if } \mathbf{x} \in \Omega_i \\ d\mathcal{J}(\Omega_i)(\mathbf{x}) & \text{if } \mathbf{x} \in D \setminus \overline{\Omega_i} \end{cases}$;
  Compute $\theta_i = \arccos\left(\dfrac{\langle \phi_i, g_{\Omega_i} \rangle}{||g_{\Omega_i}||_{L^2(D)} ||\phi_i||_{L^2(D)}}\right)$;
  if $\theta_i < \varepsilon_\theta$ **then**
    break;
  else
    $\phi_{i+1} = \dfrac{1}{\sin\theta_i}(\sin((1-\kappa_i)\theta_i)\phi_i + \sin(\kappa_i\theta_i)g_{\Omega_i})$,
    where $\kappa = \max\left\{1, \dfrac{1}{2}, \dfrac{1}{4}, ...\right\}$ such that $\mathcal{J}(\Omega_{i+1}) < \mathcal{J}(\Omega_i)$;
  end
  Update c_{ij} in the discretization $\phi(\xi,\eta) = \sum_{i=1}^{n}\sum_{j=1}^{m} B_{i,d}(\xi)B_{j,d}(\eta)c_{ij}$
end
return ϕ;

In the discretized setting, when an element e of the background mesh D is cut by the interface of the level-set function ϕ, the generalized topological derivative is computed using a linear interpolation between the values computed inside and outside the domain Ω, given by

$$g_\Omega|_e(\mathbf{x}) = d\mathcal{J}_{\text{out}}(\Omega)(\mathbf{x})|_e + \frac{|e \cap \Omega|}{|e|}(-d\mathcal{J}_{\text{in}}(\Omega)(\mathbf{x})|_e - d\mathcal{J}_{\text{out}}(\Omega)(\mathbf{x})|_e). \quad (17)$$

5. Cut elements

The correct integration of elements intersected by the level-set function as well as the precise assignment of the material parameter in these elements plays an important role in the quality of the results. Therefore, in this section, we present the procedures used to deal with the cut elements.

In the computation of the cut ratio $|e \cap \Omega|/|e|$, necessary to obtain the α property for the cut elements, we apply a quadrature library for implicitly defined geometries [31, 32]. This library precisely follows the level-set function and provides quadrature points conforming to its zero isoline, as we can see in Fig. 5a. This quadrature allows the integration of the regions defined by the zero level-set with high precision, even for complex geometries with high-degree representations of the interfaces. Therefore, with this quadrature, we can precisely capture the transition between the material property α in the inner and outer parts of the domain Ω. Some examples for area computation, moving geometries, and linear elasticity can be found in [39, 40].

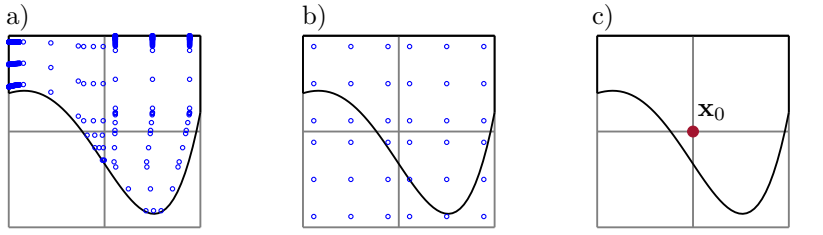

Fig. 5. Approaches to treat the cut elements: a) cut ratio computation, b) integration for assembly, c) average filtering.

However, when applying this precise quadrature rule for assembly, we obtain some instabilities in the shape due to the integration of small parts of basis functions located in regions defined by the level-set function, as well as to the jump between material and void regions, which results in a discontinuous field $\mathbf{u}$ and is approximated with higher-degree basis functions. To smooth these results, in the assembly process, we use the standard Gauss quadrature over the whole cut element and scale the local contribution by the corresponding material property α, computed using the aforementioned procedure, shown in Fig. 5b. Therefore, this approach smooths out the discontinuity in the transition between the elements and results in a solution with less numerical noise.

Another smoothing step may be applied to the generalized topological derivative. In particular, since we define an α property at each element, it occurs that, when evaluating the derivative at a Greville point, shared by multiple elements, as we can see in Fig. 5c, we have more than one α at the point. And to solve

this, we consider the average of the derivative around that point. This procedure effectively creates a smooth transition where the material property changes and works well as a sensitivity filter [17].

A second filtering step to smooth the generalized topological derivative g is performed by replacing g in the spherical linear interpolation, as described in Algorithm 1, using the solution $\widetilde{g}$ of the PDE:

$$\begin{aligned} -\gamma\Delta\widetilde{g}_\Omega + \widetilde{g}_\Omega &= g_\Omega \quad \text{in } D, \\ \nabla\widetilde{g}_\Omega \cdot \mathbf{n} &= 0 \qquad \text{on } \partial D. \end{aligned} \tag{18}$$

6. Numerical results

In this section, two numerical results for different geometries are presented. In both examples, the level-set is initialized as $\phi(\mathbf{x}) = -1$, which results in a full material background mesh, and the background mesh is discretized with 128×128 elements. The material coefficients used are $\alpha_{\text{in}} = 1$ for the material region Ω and $\alpha_{\text{out}} = 10^{-4}$ for the void region $D \setminus \overline{\Omega}$. Additionally, the parameter for area control is $l = 5$, the size control coefficient for the filtering is $\gamma = 10^{-4}$, and the Young modulus and the Poisson ratio are 1 and $1/3$, respectively. In these examples, we investigate the sensitivity of the level-set function representation by applying different polynomial degrees d for the level-set function discretization and p for approximating the solution. This is done considering two settings: one with the level-set function and the solution both approximated with the same polynomial degree, and another with a linear level-set function representation combined with a higher-degree approximation of the solution. For the optimization algorithm, we consider a tolerance of $\varepsilon_\theta = 1$ and a maximum number of iterations of 200.

6.1. Cantilever

The first example is the cantilever problem, a benchmark example for topology optimization present in a large number of research papers (see, e.g., [2–4, 7, 22–24, 30, 35]). In our case, the domain D is represented by the mapping from the parameter domain $\hat{D} = [0,1] \times [0,1]$ to a rectangle of size 2×1 with homogeneous Dirichlet boundary conditions on the left and a concentrated load on the right, as shown in Fig. 6a. Figure 6b shows the final design for $p = d = 2$, and Fig. 7 shows the evolution of the cost function, angle, and area for the two settings. Additionally, the final shapes for different configurations of polynomial degree for the solution and the level-set function discretization are shown in Fig. 8. A mesh sensitivity study with 32×32, 64×64, 128×128 and 256×256 elements is presented in Figs. 9 and 10.

These results show that higher polynomial degrees p provide better convergence behavior compared to $p = 1$, with a faster drop in the middle region of

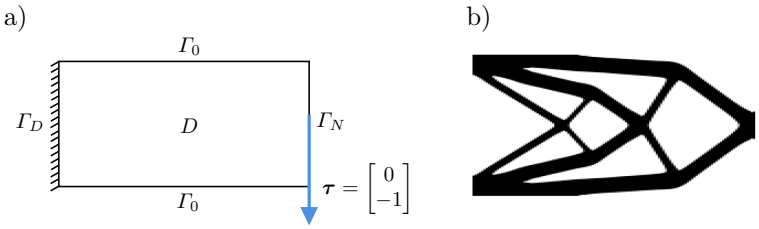

FIG. 6. Cantilever problem: a) initial problem, b) final shape.

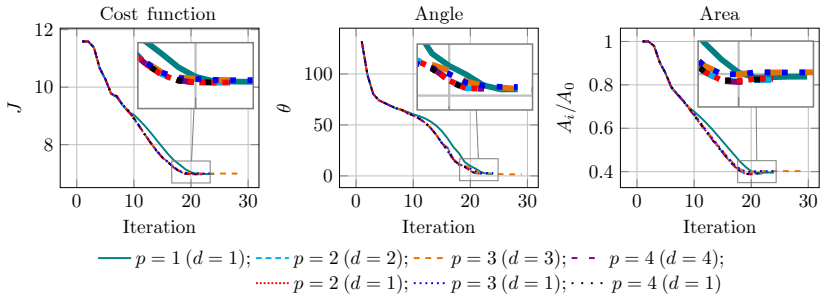

FIG. 7. Comparison of the evolution of the cost function, angle, and area for different polynomial degrees of solution approximation, with a linear level-set function representation (dotted), and with a level-set function representation with the same polynomial degree as the solution approximation (dashed).

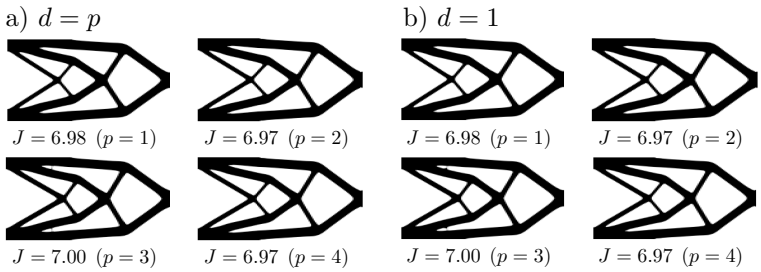

FIG. 8. Final shape for different basis function degrees used for approximating the solution: a) level-set function discretized with the same polynomial degree as the solution, b) level-set function discretized with linear basis functions.

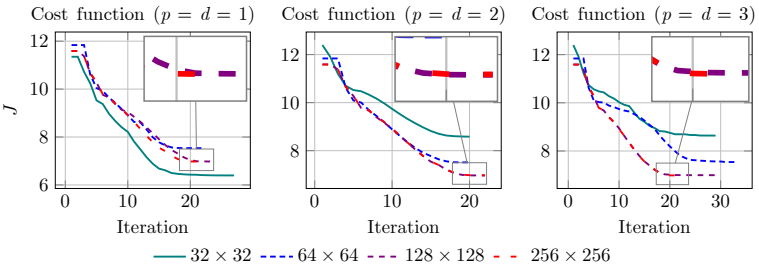

FIG. 9. Comparison of the evolution of the cost function for different mesh resolutions, using the same polynomial degrees for solution approximation and level-set function representation.

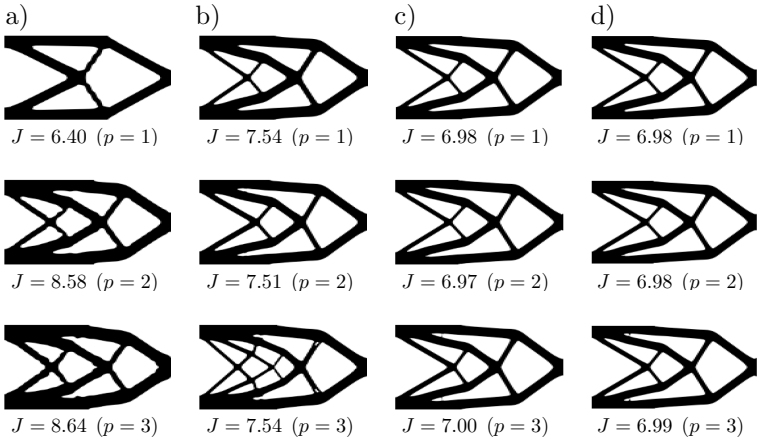

Fig. 10. Final shapes for $p = d$ and meshes using different numbers of elements per direction: a) 32×32, b) 64×64, c) 128×128, d) 256×256.

the graph in Fig. 7. However, since we use a fine mesh (128×128) to accurately represent the topology of the shape, increasing the polynomial degree p does not necessarily improve the accuracy of the solution. In addition, in all cases for this fine mesh, we converge to the same final solution (shown in Fig. 8), with similar convergence paths for higher degrees p, and the same number of steps is required for $p = 2$ and $p = 4$ in both level-set discretization settings, see Fig. 7. However, when using $p = 3$, 29 steps are required for the higher-degree level-set discretization ($d = p$), compared to 24 steps for the linear level-set discretization ($d = 1$). We also observe a slight difference in the final part of the area graph between different solution discretizations using odd and even B-spline degrees. This difference happens because, in the Greville abscissae for odd degrees, some points are shared between elements. As a result, averaging is performed around these positions when evaluating the solution, effectively acting as a filtering process [17]. While this helps to smooth the solution, it introduces a small difference compared to cases where such averaging is not required.

Figures 9 and 10 show the impact of the background mesh resolution on the final shape. As our optimization problem is not convex, it is reasonable to expect different topologies for different mesh sizes (e.g., Fig. 10a for $p = d = 1$, and Fig. 10b for $p = d = 3$). For $p = d = 2$, however, similar final topologies are obtained for all different mesh sizes. In addition, since we can represent the design more accurately with decreasing mesh size (i.e., the design space increases), we are able to find better local minima with decreasing function values from 8.58 (Fig. 10a, $p = d = 2$) to 6.97 (Fig. 10c, $p = d = 2$). Comparing the results using our standard mesh (128×128, Fig. 10c) to the finest one (256×256, Fig. 10d), we observe the same designs with minor variations in the objective value for all polynomial degrees. In Fig. 9, the evolutions of the cost functional for $p = d = 2$

and $p = d = 3$ on the 128×128 and 256×256 meshes are identical. This demonstrates that our method is mesh-independent, i.e., stable with respect to further mesh refinements, with the minimal feature size determined by the filtering of the topological derivative (18).

6.2. Quarter ring

This example has also been considered in several research papers [4, 5, 14, 22], where the same geometry is applied under different approaches and loading configurations. In our example, the domain D is defined by a mapping from the parameter domain $\widehat{D} = [0,1] \times [0,1]$ to a quarter ring of inner radius $R_{\text{in}} = 1$ and outer radius $R_{\text{out}} = 2$. Homogeneous Dirichlet boundary conditions are imposed on the bottom boundary, and a concentrated load is applied at $(0, 2)$, as illustrated in Fig. 11a. The final design for $p = d = 3$ is shown in Fig. 11b, while the evolution of the level-set function and the corresponding final shapes are presented in Figs. 12 and 13, respectively.

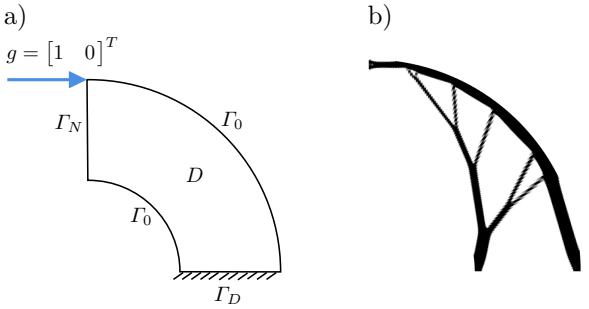

Fig. 11. Curved cantilever problem: a) initial problem, b) final shape.

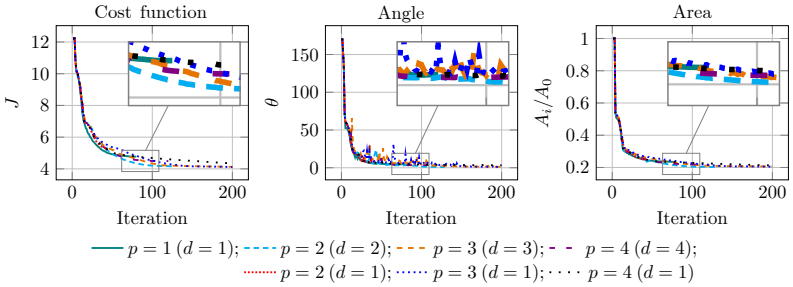

Fig. 12. Comparison of the evolution of the cost function, angle, and area for different polynomial degrees of solution approximation, with a linear level-set function representation (dotted), and with a level-set function representation with the same polynomial degree as the solution approximation (dashed).

From these results, we observe that for $p = 1$ $(d = 1)$ and $p = 2$ $(d = 1)$ the optimization stops after 80 and 43 iterations, with angles θ equal to 4.72 and 9.33, respectively, while in all of the remaining simulations an angle $\theta < 4$

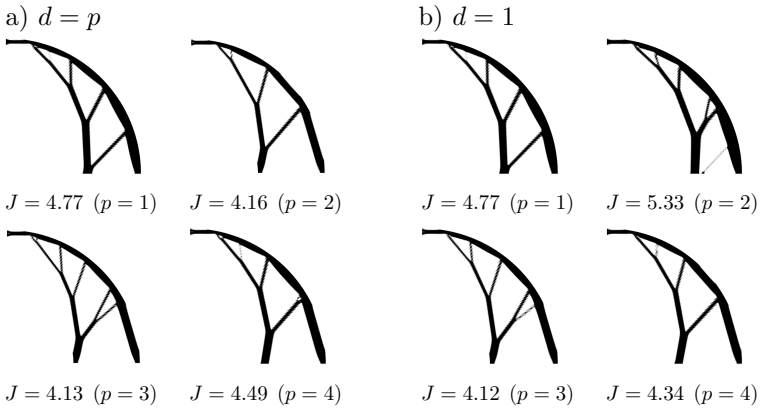

Fig. 13. Final shape for different basis function degrees used for approximating the solution: a) level-set function discretized with the same polynomial degree as the solution, b) level-set function discretized with linear basis functions.

is reached, which is a reasonable value for a numerical solutions [7]. We also notice that, although the cost function converges to a similar minimum value, all the shapes present different solutions with different configurations of features. Regarding the minimization, the lowest cost function values are obtained for $p = 3$ $(d = 1)$ with 4.12 and $p = 3$ $(d = 3)$ with 4.13, both achieved in 200 iterations. While, for $p = 2$ $(d = 2)$, a slightly higher value of 4.15 is obtained, but it requires only 129 iterations.

In topology optimization, the presence of a local minimum is a well-known challenge and depends on the choice of initial parameters [37]. Small variations in the definition of the initial setting of parameters can lead to different solutions. As a result, even if the initialization is too far from the global minimum, we still can achieve a solution that converges to a local minimum [3]. Therefore, while strategies such as refining the mesh in a coarse-to-fine approach can partially cure this issue, they do not eliminate it [1]. In this example, we notice that the possibility of setting different configurations of polynomial degrees for the level-set discretization and solution approximation, does not overcome this phenomenon completely. However, depending on the choice of setting, we can obtain a solution that satisfies the condition of having a small θ. Another parameter that can be changed and opens the possibility to find different solutions is the parameter γ, which limits the size of features in the filtering process, overcoming some noise in the final shape.

7. Conclusion

In this work, we developed an immersed isogeometric approach for level-set-based topology optimization guided only by the topological derivative. The

isogeometric approach within this framework provides seamless geometry update through a simplified meshing process, defined by a knot vector and control points. It also facilitates straightforward higher-order simulations, yielding results that are comparable to or slightly better than those obtained by standard approaches. In addition, the elements cut by the level-set function are treated using a quadrature library for implicitly defined geometries to compute a material property used to neglect the contributions from outside the domain Ω. A filtering process is also applied to material values to smooth changes between elements.

This study investigated the impact of using higher-degree basis functions for both the approximation of the solution and the discretization of the level-set function. In this investigation, we observed that being able to perform higher-order simulations can be beneficial in terms of iterations or final cost function values. However, regarding the level-set discretization, we observe that using linear basis functions yields results comparable to those obtained with higher-degree polynomials. Therefore, although the level-set is continuous, there is a discontinuity in the material property between the material region and the void. Our results indicate that using higher-degree basis functions does not directly imply a better representation of the jump across the interface. We will further investigate this aspect in future research. In addition, the proposed approach is currently limited to single-patch background geometries, and extending it to multi-patch structures is also a subject for future research.

Acknowledgements

This work wais supported by the joint DFG/FWF Collaborative Research Centre CREATOR (CRC – TRR361/F90; Grant-DOI 10.55776/F90) at TU Darmstadt, TU Graz, RICAM and JKU Linz. Moreover, P. G. is partially supported by the State of Upper Austria.

References

1. Allaire G., Jouve F., Coupling the level set method and the topological gradient in structural optimization, [in:] *Proceedings IUTAM Symposium on Topological Design Optimization of Structures, Machines and Materials*, pp. 3–12, Dordrecht, 2006, https://doi.org/10.1007/1-4020-4752-5_1.

2. Allaire G., Jouve F., Toader A., A level-set method for shape optimization, *Comptes Rendus. Mathématique*, **334**(12): 1125–1130, 2002, https://doi.org/10.1016/S1631-073X(02)02412-3.

3. Allaire G., Jouve F., Toader A., Structural optimization using sensitivity analysis and a level-set method, *Journal of Computational Physics*, **194**(1): 363–393, 2004, https://doi.org/10.1016/j.jcp.2003.09.032.

4. Aminzadeh M., Tavakkoli S.M., A parameter space approach for isogeometrical level set topology optimization, *International Journal for Numerical Methods in Engineering*, **123**(15): 3485–3506, 2022, https://doi.org/10.1002/nme.6976.

5. Aminzadeh M., Tavakkoli S.M., Multiscale topology optimization of structures by using isogeometrical level set approach, *Finite Elements in Analysis and Design*, **235**: 104167, 2024, https://doi.org/10.1016/j.finel.2024.104167.

6. Amstutz S., Sensitivity analysis with respect to a local perturbation of the material property, *Asymptotic Analysis*, **49**(1–2): 87–108, 2006, https://doi.org/10.3233/ASY-2006-778.

7. Amstutz S., Andrä H., A new algorithm for topology optimization using a level-set method, *Journal of Computational Physics*, **216**(2): 573–588, 2006, https://doi.org/10.1016/j.jcp.2005.12.015.

8. Amstutz S., Novotny A.A., de Souza Neto E.A., Topological derivative-based topology optimization of structures subject to Drucker–Prager stress constraints, *Computer Methods in Applied Mechanics and Engineering*, **233–236**: 123–136, 2012, https://doi.org/10.1016/j.cma.2012.04.004.

9. Bends?e M.P., Kikuchi N., Generating optimal topologies in structural design using a homogenization method, *Computer Methods in Applied Mechanics and Engineering*, **71**(2): 197–224, 1988, https://doi.org/10.1016/0045-7825(88)90086-2.

10. Burger M., Hackl B., Ring W., Incorporating topological derivatives into level set methods, *Journal of Computational Physics*, **194**(1): 344–362, 2004, https://doi.org/10.1016/j.jcp.2003.09.033.

11. Carvalho F.S., Ruscheinsky D., Giusti S.M., Anflor C.T.M., Novotny A.A., Topological derivative-based topology optimization of plate structures under bending effects, *Structural and Multidisciplinary Optimization*, **63**: 617–630, 2021, https://doi.org/10.1007/s00158-020-02710-4.

12. de Prenter F., Verhoosel C.V., van Brummelen E.H., Larson M.G., Badia S., Stability and conditioning of immersed finite element methods: Analysis and remedies, *Archives of Computational Methods in Engineering*, **30**: 3617–3656, 2023, https://doi.org/10.1007/s11831-023-09913-0.

13. Deaton J.D., Grandhi R.V., A survey of structural and multidisciplinary continuum topology optimization: Post 2000, *Structural and Multidisciplinary Optimization*, **49**: 1–38, 2014, https://doi.org/10.1007/s00158-013-0956-z.

14. Dedé L., Borden M.J., Hughes T.J.R., Isogeometric analysis for topology optimization with a phase field model, *Archives of Computational Methods in Engineering*, **19**: 427–465, 2012, https://doi.org/10.1007/s11831-012-9075-z.

15. Feppon F., Allaire G., Dapogny C., Jolivet P., Body-fitted topology optimization of 2D and 3D fluid-to-fluid heat exchangers, *Computer Methods in Applied Mechanics and Engineering*, **376**: 113638, 2021, https://doi.org/10.1016/j.cma.2020.113638.

16. Luz Filho J.M.M., Mattoso R., Fernandez L., A freeFEM code for topological derivative-based structural optimization, *Structural and Multidisciplinary Optimization*, **66**: 74, 2023, https://doi.org/10.1007/s00158-023-03529-5.

17. Gangl P., A multi-material topology optimization algorithm based on the topological derivative, *Computer Methods in Applied Mechanics and Engineering*, **366**: 113090, 2020, https://doi.org/10.1016/j.cma.2020.113090.

18. Gangl P., Sturm K., A simplified derivation technique of topological derivatives for quasi-linear transmission problems, *ESAIM – Control, Optimisation and Calculus of Variations*, **26**: 106, 2020, https://doi.org/10.1051/cocv/2020035.

19. Gangl P., Komann T., Krenn N., Ulbrich S., Robust topology optimization of electric machines using topological derivatives, *Arxiv*, 2025, http://arxiv.org/abs/2504.05070.

20. Gao J., Xiao M., Zhang Y., Gao L., A comprehensive review of isogeometric topology optimization: Methods, applications and prospects, *Chinese Journal of Mechanical Engineering*, **33**: 87, 2020, https://doi.org/10.1186/s10033-020-00503-w.

21. Hughes T.J.R., Cottrell J.A., Bazilevs Y., Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement, *Computer Methods in Applied Mechanics and Engineering*, **194**(39–41): 4135–4195, 2005, https://doi.org/10.1016/j.cma.2004.10.008.

22. Jahangiry H.A., Tavakkoli S.M., An isogeometrical approach to structural level set topology optimization, *Computer Methods in Applied Mechanics and Engineering*, **319**: 240–257, 2017, https://doi.org/10.1016/j.cma.2017.02.005.

23. Khatibinia M., Roodsarabi M., Structural topology optimization based on hybrid of piecewise constant level set method and isogeometric analysis, *International Journal of Optimization in Civil Engineering*, **10**(3): 493–512, 2020, https://www.researchgate.net/publication/342924423.

24. Krenn N., *Multi-Material Topology Optimization Subject to Pointwise Stress Constraints for Additive Manufacturing*, Master's Thesis, Graz University of Technology, 2021.

25. Lopes C.G., dos Santos R.B., Novotny A.A., Topological derivative-based topology optimization of structures subject to multiple load-cases, *Latin American Journal of Solids and Structures*, **12**(5): 834–860, 2015, https://doi.org/10.1590/1679-78251252.

26. Ma B., Zheng J., Lei G., Zhu J., Jin P., Guo Y., Topology optimization of ferromagnetic components in electrical machines, *IEEE Transactions on Energy Conversion*, **35**(2): 786–798, 2020, https://doi.org/10.1109/TEC.2019.2960519.

27. de Souza Neto E.A., Amstutz S., Giusti S.M., Novotny A.A., Topological derivative-based optimization of micro-structures considering different multi-scale models, *Computer Modeling in Engineering & Sciences*, **62**(1): 23–54, 2010, https://doi.org/10.3970/cmes.2010.062.023.

28. Novotny A.A., Lopes C.G., Santos R.B., Topological derivative-based topology optimization of structures subject to self-weight loading, *Structural and Multidisciplinary Optimization*, **63**: 1853–1861, 2021, https://doi.org/10.1007/s00158-020-02780-4.

29. Osher S., Sethian J.A., Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations, *Journal of Computational Physics*, **79**(1): 12–49, 1988, https://doi.org/10.1016/0021-9991(88)90002-2.

30. Roodsarabi M., Khatibinia M., Sarafrazi S.R., Hybrid of topological derivative-based level set method and isogeometric analysis for structural topology optimization, *Steel and Composite Structures*, **21**(6): 1389–1410, 2016, https://doi.org/10.12989/scs.2016.21.6.1389.

31. Saye R.I., High-order quadrature methods for implicitly defined surfaces and volumes in hyperrectangles, *SIAM Journal on Scientific Computing*, **37**(2): A993–A1019, 2015, https://doi.org/10.1137/140966290.

32. Saye R.I., High-order quadrature on multi-component domains implicitly defined by multivariate polynomials, *Journal of Computational Physics*, **448**: 110720, 2022, https://doi.org/10.1016/j.jcp.2021.110720.

33. Schillinger D., Ruess M., The finite cell method: A review in the context of higher-order structural analysis of CAD and image-based geometric models, *Archives of Computational Methods in Engineering*, **22**: 391–455, 2015, https://doi.org/10.1007/s11831-014-9115-y.

34. Shakour E., Amir O., Topology optimization with precise evolving boundaries based on IGA and untrimming techniques, *Computer Methods in Applied Mechanics and Engineering*, **374**: 113564, 2021, https://doi.org/10.1016/j.cma.2020.113564.

35. Shojaee S., Mohamadian M., Valizadeh N., Composition of isogeometric analysis with level set method for structural topology optimization, *International Journal of Optimization in Civil Engineering*, **2**(1): 47–70, 2012, https://www.researchgate.net/publication/259593893.

36. Sigmund O., Maute K., Topology optimization approaches: A comparative review, *Structural and Multidisciplinary Optimization*, **48**: 1031–1055, 2013, https://doi.org/10.1007/s00158-013-0978-6.

37. Sigmund O., Petersson J., Numerical instabilities in topology optimization: A survey on procedures dealing with checkerboards, mesh-dependencies and local minima, *Structural Optimization*, **16**: 68–75, 1998, https://doi.org/10.1007/BF01214002.

38. Suttakul P., Ngo H.T., Nguyen M.N., Bui T.Q., Rungamornrat J., Vo D., Isogeometric proportional topology optimization (IGA-PTO) for multi-material problems, *Mechanics of Advanced Materials and Structures*, **32**(20): 5006–5025, 2025, https://doi.org/10.1080/15376494.2024.2418352.

39. Teixeira G.H., Loibl M., Marussig B., Comparison of integration methods for cut elements, [in:] *Proceedings The 9th European Congress on Computational Methods in Applied Sciences and Engineering*, pp. 1–9, Lisboa, Portugal, 2024, https://doi.org/10.23967/eccomas.2024.098.

40. Toprak T. *et al.*, Employing continuous integration inspired workflows for benchmarking of scientific software – a use case on numerical cut cell quadrature, *Advances in Engineering Software*, **213**: 104087, 2026, https://doi.org/10.1016/j.advengsoft.2025.104087.

41. Verzicco R., Immersed boundary methods: Historical perspective and future outlook, *Annual Review of Fluid Mechanics*, **55**: 129–155, 2023, https://doi.org/10.1146/annurev-fluid-120720-022129.

42. Vo D., Nguyen M.N., Bui T.Q., Suttakul P., Rungamornrat J., Isogeometric gradient-free proportional topology optimization (IGA-PTO) for compliance problem, *International Journal for Numerical Methods in Engineering*, **124**(19): 4275–4310, 2023, https://doi.org/10.1002/nme.7315.

43. Vázquez R., A new design for the implementation of isogeometric analysis in Octave and MATLAB: Geopdes 3.0, *Computers and Mathematics with Applications*, **72**(3): 523–554, 2016, https://doi.org/10.1016/j.camwa.2016.05.010.

44. Wang M.Y., Wang X., Guo D., A level set method for structural topology optimization, *Computer Methods in Applied Mechanics and Engineering*, **192**(1–2): 227–246, 2003, https://doi.org/10.1016/S0045-7825(02)00559-5.

45. Wang X.S., Zhang L.T., Liu W.K., On computational issues of immersed finite element methods, *Journal of Computational Physics*, **228**(7): 2535–2551, 2009, https://doi.org/10.1016/j.jcp.2008.12.012.

46. Wiesheu M., Komann T., Merkel M., Schöps S., Ulbrich S., Cortes Garcia I., Combined parameter and shape optimization of electric machines with isogeometric analysis, *Optimization and Engineering*, **26**: 1011–1038, 2025, https://doi.org/10.1007/s11081-024-09925-0.

47. Xavier M., Novotny A.A., Topological derivative-based topology optimization of structures subject to design-dependent hydrostatic pressure loading, *Structural and Multidisciplinary Optimization*, **56**: 47–57, 2017, https://doi.org/10.1007/s00158-016-1646-4.

Received June 25, 2025; revised January 12, 2026; accepted January 17, 2026; available online January 20, 2026; version of record March 23, 2026; published issue June 17, 2026.

Computer Assisted Methods in Engineering and Science
33(2): 147–170, 2026, https://doi.org/10.24423/cames.2026.1887

Goal-Oriented Optimal Sensor Placement for PDE-Constrained Inverse Problems in Crisis Management*)

Marco Mattuschka[1)*], Noah An der Lan[1)], Max von Danwitz[1)], Daniel Wolff[2)], Alexander Popp[1),2)]

[1)] *Institute for the Protection of Terrestrial Infrastructures, German Aerospace Center (DLR), Sankt Augustin, Germany*
[2)] *Institute for Mathematics and Computer-Based Simulation (IMCS), University of the Bundeswehr Munich, Neubiberg, Germany*

[*] *Corresponding Author: marco.mattuschka@dlr.de*

This paper presents a novel framework for goal-oriented optimal static sensor placement and dynamic sensor steering in PDE-constrained inverse problems, utilizing a Bayesian approach accelerated by low-rank approximations. The framework is applied to airborne contaminant tracking, extending recent dynamic sensor steering methods to complex geometries to improve computational efficiency. A C-optimal design criterion is employed to strategically place sensors, minimizing uncertainty in predictions. Numerical experiments validate the approach's effectiveness for source identification and monitoring, highlighting its potential for real-time decision-making in crisis management scenarios.

Keywords: airborne contaminant transport, advection-diffusion equation, large-scale inverse problems, optimal experimental design, dynamic sensor steering.

1. Introduction

The growing capabilities of unmanned systems (UxS), such as unmanned aerial vehicles (UAVs, drones), autonomous underwater vehicles (AUVs), and unmanned ground vehicles (e.g., mobile robots) have made these systems indispensable tools in crisis management situations (see Fig. 1 for examples). Such autonomous sensor platforms facilitate the collection of valuable information in regions where manned missions would be too dangerous or simply impossible due

[*)] The results of this study were presented at GAMM 2025 – 95th Annual Meeting of the International Association of Applied Mathematics and Mechanics, April 7–11, 2025, Poznań, Poland.

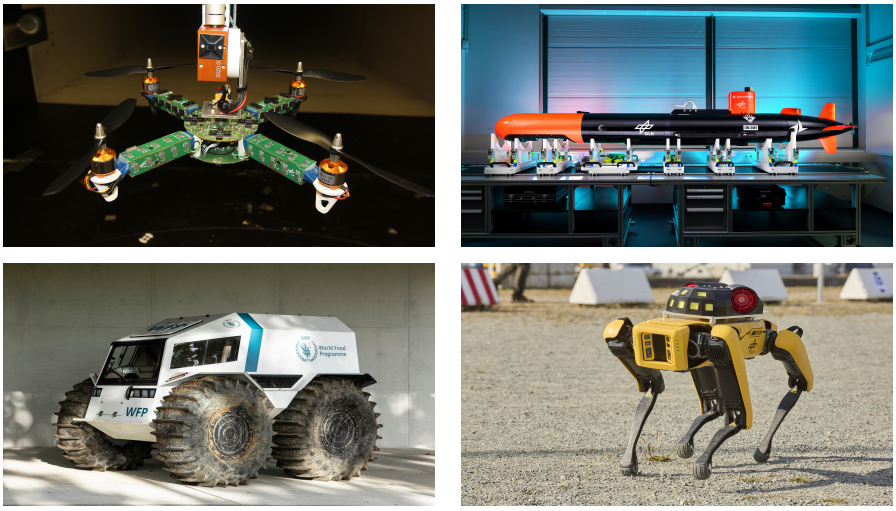

Fig. 1. Unmanned systems operated as autonomous sensor platforms at the German Aerospace Center (DLR). Images: DLR, CC BY-NC-ND 3.0.

to inaccessibility for humans. However, during the routing of a sensor platform in the complex environment of a crisis situation, the question arises as to at which locations (additional) measurements provide information gain and thereby constitute added value for decision-makers. This very question leads to the need for goal-oriented optimal sensor placement and, in the dynamic case, to a sensor steering problem. The task of selecting optimal measurement locations to infer system knowledge falls within the classical field of optimal experimental design (OED). Goal-oriented optimal experimental design generalizes this concept by determining the optimal design not for the parameters themselves, but for a derived quantity of interest (QoI). This approach offers two major advantages. First, it enables a more application-specific formulation, as the optimal design targets the QoI directly rather than the entire parameter space of the inverse problem. Second, it substantially reduces computational cost, since the QoI typically has far fewer degrees of freedom in its discretized representation.

This contribution addresses the problem in the mathematical setting of inverse problems constrained by partial differential equations (PDEs). A selected sensor steering strategy is developed and applied to the specific challenge of mapping airborne contaminant dispersion in the region of interest using discrete sensor measurements. By specifying the region of interest in both space and time, the problem naturally falls within the framework of goal-oriented optimal sensor placement.

Current methods for contaminant source identification and spread prediction rely heavily on an informative sensor placement. The selection of measuring points (sensor positions) is crucial; however, many existing studies focus solely

on stationary sensors. This work bridges this gap by incorporating recent advances in sensor selection and experimental design to derive an algorithm for optimal sensor steering. Our research focuses on developing a systematic approach to select sensor positions that maximize the accuracy of contaminant source identification and prediction. By integrating current methodological advancements, our aim is to provide a practical solution for rescue workers and first responders, enabling informed decision-making in high-stakes situations. Whereas the numerical examples presented in this work focus on the specific application of airborne contaminant transport, the goal-oriented optimal sensor placement strategy is independent of the underlying physical model and can therefore be easily transferred to other crisis management applications.

The starting point for this work is the recent publication by Wogrin *et al.* [23], which pioneers a dynamic sensor steering method in the context of airborne contaminant transport. In the present work, we extend this approach to a significantly more complex geometry. Moreover, the inverse problem solution follows a more advanced approach that uses a Laplacian-like trace class operator as prior information within a Bayesian inverse problem framework [16, 20]. To achieve approximate real-time capability, low-rank approximations of the Hessian matrix are precomputed in an offline phase [13, 15], enabling efficient problem solving in the online phase using a preconditioned inexact Newton-conjugate gradient (CG) solver [19].

In the calculation of an optimal (stationary) sensor layout, Alexanderian *et al.* [3] used a reduced model for the contaminant transport to determine an A-optimal design that minimizes the average point-wise posterior variance of the inferred parameter vector. Following extensions proposed in [5] and [18], we focus on a goal-oriented design, i.e., the uncertainty associated with the prediction of the contaminant concentration in a specific region and time is minimized. To demonstrate the feasibility of sensor steering, a relaxed optimality criterion compared to the A-optimal criterion is chosen for this test case. Specifically, we use the C-optimal criterion, which focuses on minimizing the posterior variance of a particular linear combination of the inversion parameters. This approach eliminates the need to estimate the trace of the full covariance matrix and allows us to directly assess the impact of the covariance matrix on the parameter of interest. Alternatively, a D-optimal goal-oriented design in infinite dimensions maximizes the expected information gain [3, 24]. For a broader perspective on optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs, the interested reader is referred to [1]. The remainder of this paper is organized as follows. Section 2 provides background and mathematical formulations of the forward problem of contaminant transport, the inverse problem of source identification, as well as sensor positioning strategies and goal-oriented optimization. The combination of methodological

developments into an algorithm for goal-oriented optimal sensor placement and steering is described in Sec. 3. Section 4 introduces dynamic sensor steering, showing how a mobile sensor can be guided in real time to maximize information about the contaminated field. Numerical results are presented in Sec. 5 for three test cases of goal-oriented optimal sensor placement: (a) identifying an instantaneous contaminant source in a user-defined area of interest, (b) monitoring an area of special interest over a predefined time window, and (c) steering a moving sensor. Finally, Sec. 6 offers a conclusion and outlook.

2. Background

2.1. Forward problem: contaminant distribution evolution

A mathematical description of the transport of a substance (contaminant) concentration u in a bounded open domain $\Omega \subseteq \mathbb{R}^n$ for $n \in \{2, 3\}$, as shown in Fig. 2, is given by the following equation:

$$\begin{aligned} r_{\mathcal{K}}(u) := u_t - \kappa \Delta u + \mathbf{v} \cdot \nabla u &= 0 && \text{in } (0,T) \times \Omega, \\ \kappa \nabla u \cdot \mathbf{n} &= 0 && \text{in } (0,T) \times (\Gamma_+ \cup \Gamma_0), \\ u &= 0 && \text{in } (0,T) \times \Gamma_-, \\ u(0,\cdot) &= m && \text{in } \Omega. \end{aligned} \tag{1}$$

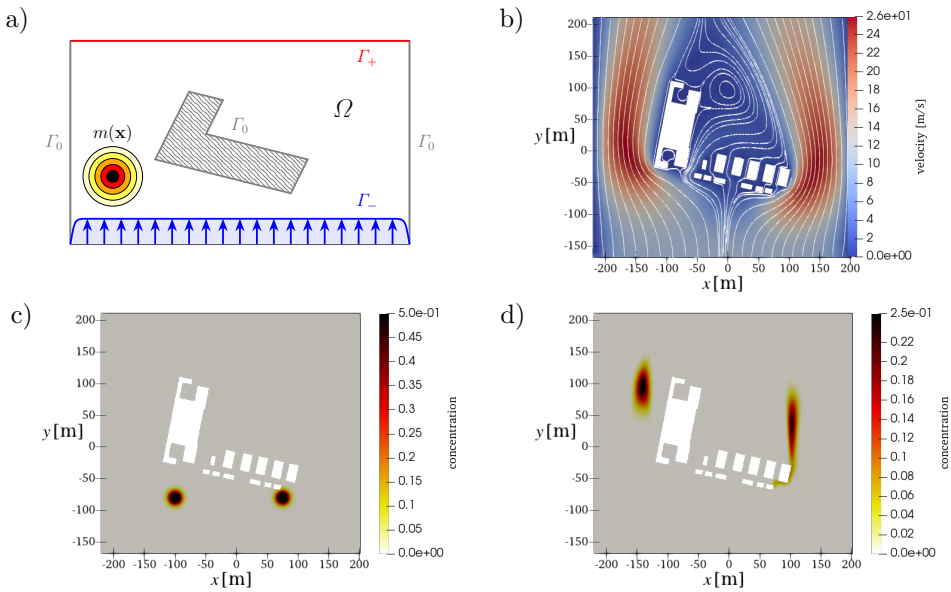

Fig. 2. Forward simulation of airborne contaminant transport on a campus geometry. Computational domain Ω with highlighted inflow (Γ_-), outflow (Γ_+), and characteristic (Γ_0) boundaries as well as the initial contaminant source $u(0,\cdot) = m(\cdot)$ (a). Estimated wind vector field $\mathbf{v}$ (b), initial condition (c), and simulated concentration at $t = 10\,\mathrm{s}$ (d).

The parameter-dependent forward problem shown in Eq. (1) is formulated for realizations of the parameter m. A visualization of the contaminant dispersion is provided in Fig. 2.

The underlying transport process is governed by a wind vector field $\mathbf{v}$, which is assumed to be sufficiently smooth, bounded – i.e., $\mathbf{v} \in L^\infty(\Omega, \mathbb{R}^n)$ – and divergence-free – i.e., $\nabla \cdot \mathbf{v} = 0$. The example wind vector field used hereinafter is shown in Fig. 2b. Based on the orientation of the wind vector relative to the outward-pointing boundary normal $\mathbf{n}$, the domain boundary $\partial\Omega$ is partitioned into three disjoint subsets: the outflow boundary $\Gamma_+ \subset \partial\Omega$, where $\mathbf{v} \cdot \mathbf{n} > 0$; the characteristic (or tangential) boundary $\Gamma_0 \subset \partial\Omega$, where $\mathbf{v} \cdot \mathbf{n} = 0$; and the inflow boundary $\Gamma_- \subset \partial\Omega$, where $\mathbf{v} \cdot \mathbf{n} < 0$, following the convention in [10].

2.2. Inverse problem: source identification

Whenever measurements of the concentration at discrete locations and times are available, an obvious question is whether the initial condition can be reconstructed on the basis of the given measurements. The respective inverse problem is illustrated in Fig. 3, see also [20]. As the function space for the initial condition, we consider an admissible subset of square-integrable functions. For instance, we define $D := H^{1,2}_{\Gamma_-}(\Omega) := \left\{ m \in H^{1,2}(\Omega) \;\middle|\; m|_{\Gamma_-} = 0 \right\}$, where $H^{1,2}(\Omega)$ denotes the Sobolev space of square-integrable functions whose weak derivatives $\partial_i u$ exist and are square-integrable. The gradient is denoted as $\nabla u = (\partial_1 u, \partial_2 u)$, and the norm on $H^{1,2}(\Omega)$ is defined as

$$\|u\|_{H^{1,2}(\Omega)} := \left(\int_\Omega u^2 + \|\nabla u\|_2^2 \, d\mathbf{x} \right)^{1/2}.$$

In this setting, the estimation of the initial value leads to an optimization problem, which will be addressed in the following. The first step is to describe sensor measurements within this formulation.

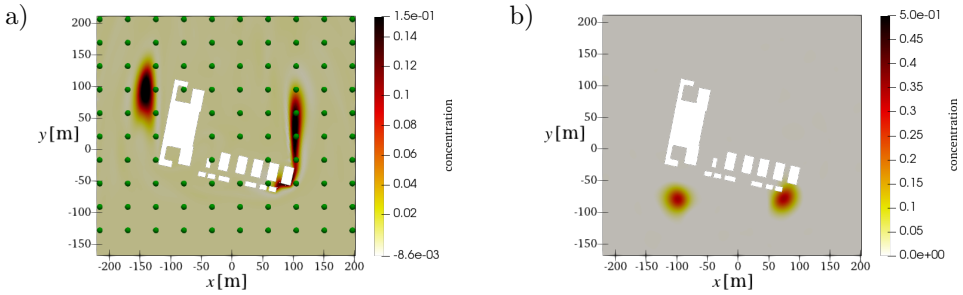

Fig. 3. Inverse problem. Measurements at 96 equidistantly spaced sensor positions (a) and reconstructed initial condition (solution of the inverse problem (b)).

To incorporate sensor measurements, let $T_0 > 0$ be fixed and consider space-time points $(t_i^{\text{obs}}, x_i^{\text{obs}})$, $i = 1, ..., q$, with $t_i^{\text{obs}} \in [T_0, T)$ and interior points $x_i^{\text{obs}} \in \Omega = \overline{\Omega} \setminus \partial\Omega$. For the transient problem in Eq. (1) on a smooth domain Ω, the diffusion term induces strong smoothing on the parameter field m. By standard parabolic regularity (see, e.g., [11]), $u(t, \cdot)$ embeds in the space of continuous functions $C^0(\overline{\Omega})$ on the closure of Ω for any $t > 0$. Since continuity is only required near the observation points, define $\overline{\Omega}_o := \bigcup_{i=1}^{q} B_{r_0}(x_i^{\text{obs}}) \subset \Omega$, where $r_0 > 0$ is chosen such that $u \in C^0([T_0, T] \times \overline{\Omega}_o)$ admits a continuous representation. This ensures the formulation remains valid for less regular domains Ω. Finally, we define a well-posed and bounded space-time observation operator $\mathcal{B} : C^0([T_0, T] \times \overline{\Omega}_o) \to \mathbb{R}^q$ by $u \mapsto \sum_{i=1}^{q} \delta_{(t_i^{\text{obs}}, x_i^{\text{obs}})}(u)\, \mathbf{e}_i = \left(u(t_i^{\text{obs}}, x_i^{\text{obs}})\right)_{i=1}^{q}$, where $\mathbf{e}_i$ is the standard basis of $\mathbb{R}^q$. Using this observation operator, the final parameter-to-observable map $\mathcal{F} : D \to \mathbb{R}^q$ is defined by

$$\mathcal{F}(m) := \mathcal{B} \circ \mathcal{K}(m), \quad \text{with} \quad \mathcal{K}(m) := u \quad \text{such that} \quad r_{\mathcal{K}}(u) = m. \tag{2}$$

Here, $\mathcal{K}$ is the parameter-to-state map, mapping the parameter space D to the state space, often referred to as the 'model' in this context [8]. An example problem in which $\mathcal{K}$ maps the initial condition u_0 to the solution of $u(t = 10\,\text{s}, \cdot)$ is illustrated in Fig. 2.

The next step is to model sensor noise, which is usually present in real-world measurements. To this end, it is assumed that the observations $\mathbf{d} = \mathcal{F}(m) + \boldsymbol{\epsilon}$ contain centered additive Gaussian noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \Gamma_{\text{noise}})$ due to measurement uncertainties. For the sake of simplicity, it is further assumed that the sensor noise at the different sensor positions is uncorrelated and of equal magnitude, represented by the diagonal matrix $\Gamma_{\text{noise}} = \text{diag}(\sigma^2, ..., \sigma^2)$. Moreover, the conformity of the simulation with the measured values, also called misfit, $\mathbf{y} = \mathcal{F}(m) - \mathbf{d}$, is evaluated in the following norm:

$$||\mathcal{F}(m) - \mathbf{d}||^2_{\Gamma^{-1}_{\text{prior}}} = 1/\sigma^2 \sum_{i=1}^{q} \left(u(t_i^{\text{obs}}, x_i^{\text{obs}}) - d_i\right)^2.$$

In summary, the inverse problem consists of using available measurements $\mathbf{d}$ to infer the values of the unknown parameter field m. Alternatively, the maximum a-posteriori point m_{map} can be characterized as the solution of the minimization problem:

$$m_{\text{map}} = \arg\min_{m \in D} J(m) := \frac{1}{2}||\mathcal{F}(m) - \mathbf{d}||^2_{\Gamma^{-1}_{\text{noise}}} + \frac{1}{2}||m - m_{\text{prior}}||^2_{\Gamma^{-1}_{\text{prior}}}, \tag{3}$$

with the prior information encoded as a Tikhonov regularization term.

Unfortunately, the system is heavily under-determined under realistic conditions, as sensor measurements are only available at a few locations, whereas the initial condition is to be reconstructed for the entire domain. In order to transform this into a well-posed problem, some prior knowledge is needed and a formulation as a Bayesian inverse problem provides a suitable framework. In this setting, a Gaussian prior $\mathcal{N}(m_{\text{prior}}, \Gamma_{\text{prior}})$ with mean m_{prior} and covariance Γ_{prior} is chosen for parameter regularization. Then, the posterior density of m satisfies by Bayes' theorem: $\pi_{\text{post}}(m|\mathbf{d}) \propto \pi_{\text{like}}(\mathbf{d}|m)\, \pi_{\text{prior}}(m)$. Here, $\pi_{\text{like}}(\mathbf{d}|m) \propto \exp(\frac{1}{2}||\mathcal{F}(m) - \mathbf{d}||^2_{\Gamma^{-1}_{\text{noise}}})$ is the likelihood function under the observational Gaussian noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \Gamma_{\text{noise}})$. Due to the linearity of $\mathcal{F}$, the posterior distribution is again a Gaussian distribution $\mathcal{N}(m_{\text{map}}, \Gamma_{\text{post}})$ with covariance and mean:

$$\begin{aligned} \Gamma_{\text{post}} &= (\mathcal{F}^* \Gamma^{-1}_{\text{noise}} \mathcal{F} + \Gamma^{-1}_{\text{prior}})^{-1}, \\ m_{\text{map}} &= \Gamma_{\text{post}} (\mathcal{F}^* \Gamma^{-1}_{\text{noise}} \mathbf{d} + \Gamma^{-1}_{\text{prior}} m_{\text{prior}}) \,. \end{aligned} \tag{4}$$

Here, the formal adjoint operator $\mathcal{F}^* : \mathbb{R}^q \to D^*$ is required for this framework. For the mapping $\mathcal{F} : D \to \mathbb{R}^q$ between Hilbert spaces, the formal adjoint operator $\mathcal{F}^*$ is characterized by the relation $\langle \mathcal{F}(m), \mathbf{y} \rangle_{\mathbb{R}^q} = \langle m, \mathcal{F}^*(\mathbf{y}) \rangle_{L^2(\Omega)}$ for all $\mathbf{y} \in \mathbb{R}^q$ and $m \in D$, and its existence follows from Riesz's representation theorem [4]. The posterior covariance or Hessian $\mathcal{H} := \mathcal{F}^* \Gamma^{-1}_{\text{noise}} \mathcal{F} + \Gamma^{-1}_{\text{prior}}$ of the objective function J, see Eq. (3), contains a wealth of information about the system. In line with the Bayesian framework [20], the covariance or inverse Hessian matrix can be employed to predict the uncertainty of the system and is of particular significance for optimal sensor placement in the following section.

The mean value m_{map} is a reliable estimate for the initial value and thus represents the solution of the inverse problem Fig. 3. In order to calculate m_{map} from Eq. (4), a further specification of the adjoint operator $\mathcal{F}^*$ is necessary. By applying integration by parts to the weak form of Eq. (1), the adjoint state p can be derived and satisfies the following equation:

$$\begin{aligned} -p_t - \kappa \Delta p - \operatorname{div}(p\mathbf{v}) &= -\frac{1}{\sigma^2} \sum_{i=1}^{q} y_i \, \delta_{(t_i^{\text{obs}}, x_i^{\text{obs}})} && \text{in } (0,T) \times \Omega, \\ (\mathbf{v}p + \kappa \nabla p) \cdot \mathbf{n} &= 0 && \text{in } (0,T) \times (\Gamma_+ \cup \Gamma_0), \\ p &= 0 && \text{in } (0,T) \times \Gamma_-, \\ p(T, \cdot) &= 0 && \text{in } \Omega, \end{aligned} \tag{5}$$

for given $\mathbf{y} \in \mathbb{R}^q$. Finally, the adjoint operator $\mathcal{F}^*$ can be explicitly determined, resulting in $\mathcal{F}^* \mathbf{y} = p(0, \cdot)$.

2.3. Goal-oriented optimal experimental design: sensor placement

So far, a model for the forward problem and an estimate of the initial condition, given a fixed sensor configuration, have been derived. However, the question of how such a sensor arrangement ought to be chosen remains unanswered. As mentioned in the previous section, the covariance of the posterior, or more specifically, the Hessian matrix, plays a crucial role in developing an indicator for the uncertainty in the system. In classical Bayesian optimal experimental design (OED), a design functional Φ serves as a quantitative criterion for the uncertainty described by the covariance matrix and is used to guide the selection of an optimal experimental setup. Specifically, the optimization problem

$$\min_{\mathbf{w}\in\mathbb{W}}\Big(\Phi[\Gamma_{\text{post}}(\mathbf{w})] + \mathcal{R}(\mathbf{w})\Big) \tag{6}$$

is solved, where $\mathbb{W}$ denotes the set of all valid sensor configurations, and $\mathcal{R}$ is a suitable regularization term. As in [2], a finite set of candidate sensor placements $(t_i^{\text{obs}}, x_i^{\text{obs}}) \in [T_0, T] \times \Omega$ for $1 \le i \le q$ is considered. An example of a spatial grid with 96 sensors can be seen in Fig. 3. For this set of candidate locations, a weight vector $\mathbf{w} \in [0,1]^q$ is defined with the i-th candidate entry corresponding to the i-th location in space and time. In fact, the weight vector determines which measurements are realized or taken into account. In the case of stationary sensors located at positions $\mathbf{x}_s^{\text{obs}}$, measurements collected at these spatial points over the entire time horizon, denoted as $(\cdot, \mathbf{x}_s^{\text{obs}})$, are constantly weighted and included in the misfit. Hence, the number of independent entries in the weight vector reduces to the number of possible stationary sensor positions. If we consider a mobile sensor, we have a trajectory $\gamma : \{t_0^{\text{obs}}, ..., t_s^{\text{obs}}\} \to \{\mathbf{x}_0^{\text{obs}}, ..., \mathbf{x}_s^{\text{obs}}\}$. For points on the trajectory $(t_i^{\text{obs}}, \mathbf{x}_i^{\text{obs}})$ the sensor weight is set to $\mathbf{w}_i = 1$. All weights away from the trajectory are set to 0. To adjust the forward model to the chosen sensor configuration, we consider the diagonal matrix $W \in \mathbb{R}^{q\times q}$ with $W_{ii} = \mathbf{w}_i$. If we denote the parameter-to-observable map for all sensor positions as $\mathcal{F}$, then for each design $\mathbf{w} \in \mathbb{W}$, we have $\mathcal{F}(\mathbf{w}) = W\mathcal{F}$. Taking this further and using W to modify the noise matrix, the influence of a selected sensor layout is also captured in the likelihood function

$$\pi_{\text{like}}(\mathbf{d}|m, \mathbf{w}) \propto \exp\left\{-\frac{1}{2}(\mathcal{F}(m) - \mathbf{d})^T W^{1/2} \Gamma_{\text{noise}}^{-1} W^{1/2} (\mathcal{F}(m) - \mathbf{d})\right\}.$$

In consequence, the posterior covariance and mean also depend on the sensor layout via (cf. Eq. (4)):

$$\begin{aligned} \Gamma_{\text{post}}(\mathbf{w}) &= (\mathcal{F}^* W^{1/2} \Gamma_{\text{noise}}^{-1} W^{1/2} \mathcal{F} + \Gamma_{\text{prior}}^{-1})^{-1}, \\ m_{\text{map}}(\mathbf{w}) &= \Gamma_{\text{post}}(\mathcal{F}^* \Gamma_{\text{noise}}^{-1} W \mathbf{d} + \Gamma_{\text{prior}} m_{\text{prior}})\,. \end{aligned} \tag{7}$$

In Bayesian optimal experimental design, the objective is to construct experiments that minimize the posterior uncertainty in the unknown parameter field m. This approach belongs to the broader class of inverse uncertainty quantification methods. Goal-oriented optimal experimental design extends this framework by additionally incorporating forward uncertainty quantification into the design objective. This is achieved by introducing a goal QoI, denoted by ρ. Following [5, Sec. 3], we consider a QoI defined as the action of a bounded linear operator $\mathcal{P}$ on the parameter field, i.e., $\rho = \mathcal{P}m$. Due to the linearity of $\mathcal{P}$, the prior distribution of ρ is Gaussian. In particular, $\rho \sim \mathcal{N}(\rho_{\text{prior}}, \Sigma_{\text{prior}})$, with mean $\rho_{\text{prior}} = \mathcal{P}m_{\text{prior}}$ and covariance $\Sigma_{\text{prior}} = \mathcal{P}\Gamma_{\text{prior}}\mathcal{P}^*$. The resulting Bayesian inverse problem is well-defined, and the posterior distribution of the QoI given measurements $\mathbf{d}$ is also Gaussian: $\pi_{\text{post}}(\rho \mid \mathbf{d}) \sim \mathcal{N}(\rho_{\text{post}}, \Sigma_{\text{post}})$. The posterior mean and covariance are given by:

$$\Sigma_{\text{post}} = \mathcal{P}\,\Gamma_{\text{post}}\,\mathcal{P}^* \quad \text{and} \quad \rho_{\text{post}} = \mathcal{P}m_{\text{map}},$$

where Γ_{post} and m_{map} are the derived quantities from Eq. (7).

3. Discretization, preconditioning, and sparsification

3.1. Finite element discretization

To solve the PDE problems (Eq. (1) and Eq. (5)) numerically, a finite element discretization is employed using n_{dof} Lagrange basis functions $\mathcal{V}_h = \text{span}\{\phi_1, ..., \phi_{n_{\text{dof}}}\}$. Moreover, we find an identification between a vector in $\mathbb{R}^{n_{\text{dof}}}$ and finite elements $I : \mathbb{R}^{n_{\text{dof}}} \to \mathcal{V}_h$ via $I(a) = \sum_{i=1}^{n_{\text{dof}}} a_i\phi_i$. This leads to discretized versions of the parameter-to-observable map $\mathcal{F}_h : \mathbb{R}^{n_{\text{dof}}} \to \mathbb{R}^q$ defined by $\mathcal{F}_h(m_h) = \mathcal{B}(u_h)$, where u_h solves Eq. (1) in a weak form, and its adjoint operator $\mathcal{F}_h^* : \mathbb{R}^q \to \mathbb{R}^{n_{\text{dof}}}$ given by $\mathcal{F}_h^*\mathbf{y} = p_h(0, \cdot)$, where p_h solves Eq. (5) in a weak form. The identification is an isometry, i.e., $\langle I(a), I(b)\rangle_{L^2(\Omega)} = \langle a, b\rangle_M =: a^T M\, b$, where the mass matrix $M_{ji} := \int_\Omega \phi_i(\mathbf{x})\phi_j(\mathbf{x})\, d\mathbf{x}$, $M \in \mathbb{R}^{n_{\text{dof}} \times n_{\text{dof}}}$ is used to define the corresponding scalar product. For further details of the finite element discretization, we refer to [20] and [21].

3.2. Preconditioning of the discrete inverse problem

To solve the discrete inverse problem, the prior distribution needs to be carefully chosen. A Laplacian-like operator of trace class $\mathcal{A} := (\eta\, I - \gamma\Delta)$, with Robin boundary condition, $\gamma\nabla m \cdot \mathbf{n} + \beta m = 0$ in $(0, T) \times \partial\Omega$, is applied with the constant β as proposed in [9]. This operator serves as a suitable covariance operator for the prior, e.g., $\Gamma_{\text{prior}} = \mathcal{A}^{-2} = (\eta\, I - \gamma\Delta)^{-2}$. In addition,

its discrete counterpart is given by the mapping $\Gamma_{\text{prior,h}} : \mathbb{R}^{n_{\text{dof}}} \to \mathbb{R}^{n_{\text{dof}}}$ via $\Gamma_{\text{prior,h}} = (M^{-1}A)^{-2} = A^{-1}MA^{-1}M := R^{-1}M$, with matrix representation $A_{ij} = \int \phi_i(x)\mathcal{A}\phi_j(x)\,dx$.

Combining this covariance operator with an appropriate prior mean m_{prior} (in our applications, e.g., $m_{\text{prior}} = 0$) renders the inverse problem well-posed and its solution can be found by solving the following equation for m_{map}:

$$\mathcal{H}_h(\mathbf{w})m_{\text{map}} = \mathcal{F}_h^*\Gamma_{\text{noise}}^{-1}(W\mathbf{d}) + {\Gamma_{\text{prior,h}}}^{-1} m_{\text{prior}}, \tag{8}$$

for the discrete version of the Hessian, that is, $\mathcal{H}_h(\mathbf{w}) = \mathcal{F}_h^* W^{1/2}\Gamma_{\text{noise}}^{-1}W^{1/2}\mathcal{F}_h + {\Gamma_{\text{prior,h}}}^{-1}$. Since directly determining the Hessian matrix is computationally expensive for large-scale problems (requiring $O(n_{\text{dof}})$-PDE solutions), an iterative CG method is employed. This approach requires only the action of the Hessian-vector on a given vector $m_k \in \mathbb{R}^{n_{\text{dof}}}$ at each iteration. Specifically, the Hessian action is computed by the following steps: first, solve the forward equation $\mathbf{d} = \mathcal{F}_h(m_k) = \mathcal{B}(u_h)$, then, solve the corresponding adjoint equation $\mathcal{F}_h^*\big(W^{1/2}\Gamma_{\text{noise}}^{-1}W^{1/2}\,\mathbf{d}\big) = p_h(0,\cdot)$, next, compute $\widetilde{m}_k = \Gamma_{\text{prior,h}}m_k$, and finally obtain the Hessian action as $\mathcal{H}_h(\mathbf{w})m_k = p_h(0,\cdot) + \widetilde{m}_k$.

Since two PDE solutions have to be determined per iteration, a reduced model of the Hessian matrix is precomputed in advance so that the inverse problem can be solved quickly. In addition, using the Cholesky decomposition of the prior covariance, ${\Gamma_{\text{prior,h}}}^{-1} = (M^{-1}A)(M^{-1}A)^*$, one obtains the preconditioned Hessian matrix as:

$$(\mathcal{A}_h^{-1})^*\mathcal{H}_h(\mathbf{w})\mathcal{A}_h^{-1} = (\mathcal{F}_h \circ \mathcal{A}_h^{-1})^* W^{1/2}\Gamma_{\text{noise}}^{-1}W^{1/2}(\mathcal{F}_h \circ \mathcal{A}_h^{-1}) + I, \tag{9}$$

where $\mathcal{A}_h^{-1} = (A^{-1}M)$. This preconditioned system $\mathcal{F}\circ\mathcal{A}^{-1}$ has rapidly decaying eigenvalues. Therefore, we follow [14, 20] in constructing a low-rank approximation of the prior-preconditioned misfit part of the Hessian, i.e., $\mathcal{H}_h^{\mathbf{misfit}}(\mathbf{w}) := (\mathcal{F}_h\circ\mathcal{A}_h^{-1})^* W^{1/2}\Gamma_{\text{noise}}^{-1}W^{1/2}(\mathcal{F}_h\circ\mathcal{A}_h^{-1})$ by solving the symmetric eigenvalue problem [13, 15]:

$$\mathcal{H}_h^{\mathbf{misfit}}(\mathbf{w})v_i = \lambda_i\, M\, {\Gamma_{\text{prior,h}}}^{-1} v_i = \lambda_i R v_i$$

for an orthogonal basis $V_r = (v_1, ..., v_r) \in \mathbb{R}^{n_{\text{dof}}\times r}$ and $\lambda_1 \geq ... \geq \lambda_r$ using the scalar product induced by $M{\Gamma_{\text{prior,h}}}^{-1}$, i.e., $\langle a, b\rangle_{M{\Gamma_{\text{prior,h}}}^{-1}} =: a^T M{\Gamma_{\text{prior,h}}}^{-1} b$. Applying the Sherman–Morrison–Woodbury formula, we write:

$$\mathcal{A}_h(\mathcal{H}_h(\mathbf{w}))^{-1}\mathcal{A}_h = (\widetilde{\mathcal{H}}_h^{\mathbf{misfit}}(\mathbf{w}) - I)^{-1} \approx I + V_r D_r(\mathbf{w})V_r^T, \tag{10}$$

where $D_r = \operatorname{diag}(\lambda_1/(1+\lambda_1), ..., \lambda_r/(1+\lambda_r))$ is a low-rank approximation of the Hessian. Detailed information on this approach can be found in [2] and [14]. Using this approximation, the solution of Eq. (8) can be determined with a preconditioned Newton-CG method, see [19, 20].

3.3. Sparsification of sensor layouts and optimality criteria

Obviously, the trace of the posterior covariance is minimized when every sensor weight is set to 1, which corresponds to using all available information to reduce the level of uncertainty. Thus, to derive a sparse sensor configuration, a penalty term must be introduced into the optimization problem (Eq. (6)). A common choice for the regularization term in Eq. (6) is the ℓ_1-norm, which leads to a convex minimization problem with a unique minimizer. Specifically, the regularization term is defined as:

$$\mathcal{R}(\mathbf{w}) := \alpha \|\mathbf{w}\|_1 = \alpha \mathbf{1}^\top \mathbf{w}, \tag{11}$$

where $\alpha > 0$ is the regularization parameter and $\mathbf{1}$ is a vector of ones. Finally, a binary sensor configuration $\{0,1\}^q$ is obtained by considering only sensor locations with weights above a chosen threshold.

It should be noted that the resulting binary vector is not the optimal binary weight vector. Obtaining such an optimal solution requires solving the ℓ_0-regularized problem considered in [2]. The key idea of that approach is to employ a regularization term based on the number of nonzero entries in $\mathbf{w}$, i.e.,

$$\mathcal{R}(\mathbf{w}) := \alpha \sum_{w_i \neq 0} 1.$$

To approximate this discontinuous penalty, a sequence of strictly convex functions $f_j(\mathbf{w})$ is introduced such that $f_j(\mathbf{w}) \longrightarrow \sum_{w_i \neq 0} 1$ as $j \to \infty$. At each iteration, the optimization problem

$$\min_{\mathbf{w} \in \mathbb{W}} \left(\Phi[\Gamma_{\text{post}}(\mathbf{w})] + f_j(\mathbf{w}) \right)$$

is solved and the optimizer obtained at iteration j is used as the initial value for iteration $j+1$. While effective, this procedure entails substantially higher computational cost. Moreover, as observed empirically in [2, Subsec. 6.1.4], the sensor configurations obtained using this ℓ_0-approximation strategy offer only marginal improvement compared with those produced by ℓ_1-regularization followed by a simple thresholding step. Since a highly efficient method is required for the sensor steering presented in Subsec. 5.3, we therefore employ an ℓ_1-regularization approach.

The A-optimality criterion for sensor placement minimizes the integrated point-wise posterior variance in linear Bayesian inverse problems: $\int_\Omega \text{Var}\,[m(x)]\,dx = \text{tr}\,(\Gamma_{\text{post}}(\mathbf{w}))$. To compute the discrete variance field $\sigma^2(x) := \text{Var}\,[m(x)]$, one can extract the diagonal of the inverse Hessian matrix in the finite element basis, assigning each diagonal entry to its corresponding node. However, this exact

calculation is computationally expensive ($O(n_{\text{dof}})$), so the reduced-order model (Eq. (10)) is used for visualization. Figure 4b shows the variance field $\sigma^2(x)$ of m_{map} obtained with a C-optimal sparse sensor layout.

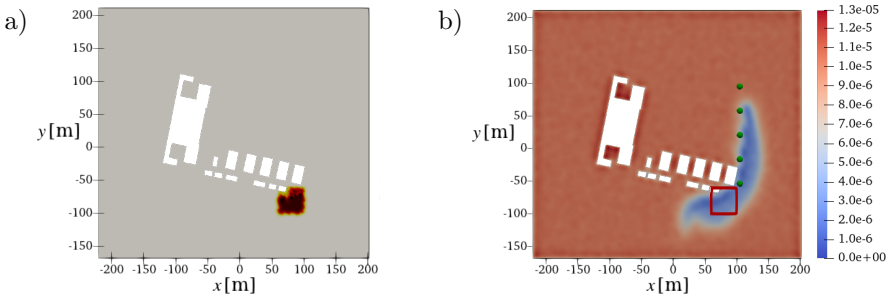

FIG. 4. Region of interest P defining the QoI (a) and point-wise variance $\sigma^2(x)$ of m_{map} obtained with five selected sensors (green spheres (b)).

Moreover, the prior covariance is computed using approximate random sampling [20]. Finally, computing the trace of the inverse Hessian is costly. Therefore, we relax A-optimality to a C-optimal design, which requires only the evaluation of the Hessian's action on a fixed vector $c \in \mathbb{R}^{n_{\text{dof}}}$ [1]. In this case, the design criterion becomes $\Phi(\mathbf{w}) = c^T \Gamma_{\text{post}}(\mathbf{w}) c$.

4. GOAL-ORIENTATION AND SENSOR STEERING

4.1. GOAL-ORIENTED OPTIMAL EXPERIMENTAL DESIGN

In the next step, the presented method is adjusted to achieve goal-oriented C-optimal experimental designs for stationary sensors. To this end, an operator $\mathcal{P} : D \to \mathbb{R}$ is first selected so that the design, namely the sensor placement, is optimized to observe the initial conditions in a region $P \subset \Omega$, as shown in Fig. 4. This simplification via forward uncertainty quantification, defined by the operator $\mathcal{P}$, leads to a C-optimal design for the problem in Eq. (7), as described in [1].

The indicator function of a subset, that is the function $\mathbf{1}_A \colon \Omega \to \{0, 1\}$, which for a given subset A of Ω, takes the value 1 at points in A and the value 0 at points outside of A, is used to define our operator $\mathcal{P}$. If an optimal design for the initial condition m is desired, the operator $\mathcal{P}$ does not depend on the solution u, or more specifically, it does not depend on the operator $\mathcal{F}$. Concretely, the operator for the QoI is given by $\mathcal{P}(m) := \langle m, \mathbf{1}_P \rangle_{L^2(\Omega)} = \int_P m(\mathbf{x})\, d\mathbf{x}$. By identifying the dual space of $L^2(\Omega)$ and defining the adjoint operator, it follows that $\mathcal{P}^* = \mathbf{1}_P$, and the design function for the optimal experimental design takes the form $\text{tr}[\Gamma_{\text{post}}(\mathbf{w})] = \langle \mathbf{1}_P, \Gamma_{\text{post}}(\mathbf{w}) \mathbf{1}_P \rangle_{L^2(\Omega)}$. In a finite element setting,

the function $\mathbf{1}_P$ is represented by a vector $c \in \mathbb{R}^{n_{\text{dof}}}$ via the standard projection, i.e., $\mathbf{1}_P - c_h \perp \mathcal{V}_h$ into the finite element space $\mathcal{V}_h = \{\phi_1, ..., \phi_{n_{\text{dof}}}\}$. According to [1, Subsec. 4.1], the design function is given by $\Phi(\mathbf{w}) = c_h^T M \mathcal{H}_h^{-1}(\mathbf{w}) c_h$. Together with a suitable regularization term, e.g., $\alpha||\mathbf{w}||_{l^1}$, an optimal design can be obtained by minimizing the objective $c_h^T M \mathcal{H}_h^{-1}(\mathbf{w})\, c_h + \alpha||\mathbf{w}||_{l^1}$ which is illustrated in Fig. 4. If this procedure is generalized to determine an optimal sensor placement for the contaminant concentration over a specific spatial region and time interval, the operator $\mathcal{P}$ must be extended accordingly. Specifically, by defining $\mathcal{P}$ over a subset of space-time, e.g., $[T_0^{\text{QoI}}, T_{\text{final}}^{\text{QoI}}] \times P$, the operator $\mathcal{P}$ is then constructed as:

$$\mathcal{P}(m) = \int_0^T \int_\Omega \mathbf{1}_{[T_0^{\text{QoI}}, T_{\text{final}}^{\text{QoI}}] \times P} \mathcal{K}(m)(t, \mathbf{x})\, dt\, d\mathbf{x} = \int_{T_0^{\text{QoI}}}^{T_{\text{final}}^{\text{QoI}}} \int_P u(t, \mathbf{x})\, dt\, d\mathbf{x}.$$

Again, the adjoint of $\mathcal{P}$ is needed. For this, we calculate

$$\begin{aligned} \mathcal{P}^*(1)(\widehat{m}) &= \langle \mathcal{K}(\widehat{m}), \mathbf{1}_{[T_0^{\text{QoI}}, T_{\text{final}}^{\text{QoI}}] \times P} \rangle_{L^2(\Omega)} \\ &= \langle \widehat{m}, \mathcal{K}^* \mathbf{1}_{[T_0^{\text{QoI}}, T_{\text{final}}^{\text{QoI}}] \times P}) \rangle_{L^2(\Omega)}, \end{aligned}$$

and therefore the map c is obtained by $c := \mathcal{P}^*(1) = \mathcal{K}^*(\mathbf{1}_{[T_0^{\text{QoI}}, T_{\text{final}}^{\text{QoI}}] \times P})$, since $\widehat{m}$ was arbitrary. More precisely, the map c satisfies the following PDE:

$$\begin{aligned} -c_t - \kappa \Delta c - \operatorname{div}(c\mathbf{v}) &= \mathbf{1}_{[T_0^{\text{QoI}}, T_{\text{final}}^{\text{QoI}}] \times P} && \text{in } (0, T) \times \Omega, \\ (\mathbf{v}c + \kappa \nabla c) \cdot \mathbf{n} &= 0 && \text{in } (0, T) \times (\Gamma_+ \cup \Gamma_0), \\ c &= 0 && \text{in } (0, T) \times \Gamma_-, \\ c(T, \cdot) &= 0 && \text{in } \Omega. \end{aligned} \tag{12}$$

In Fig. 5a an example of a constant-over-time QoI $\mathbf{1}_{[T_0^{\text{QoI}}, T_{\text{final}}^{\text{QoI}}] \times P}$ is shown. The corresponding map $c(t = 0)$ is visualized in Fig. 5b. Returning to the definition, the objective function for the goal-oriented sensor design reads:

$$\begin{aligned} \Phi(\mathbf{w}) &= \mathcal{P}\mathcal{H}^{-1}(\mathbf{w})\mathcal{P}^*(1) = \mathcal{P}(\mathcal{H}^{-1}(\mathbf{w})c) \\ &= \langle \mathcal{K}(\mathcal{H}^{-1}(\mathbf{w})c), \mathbf{1}_{[T_0^{\text{QoI}}, T_{\text{final}}^{\text{QoI}}] \times P} \rangle_{L^2(\Omega)} \\ &= \langle \mathcal{H}^{-1}(\mathbf{w})c, \mathcal{K}^*(\mathbf{1}_{[T_0^{\text{QoI}}, T_{\text{final}}^{\text{QoI}}] \times P}) \rangle_{L^2(\Omega)} = \langle c, \mathcal{H}^{-1}(\mathbf{w})c \rangle_{L^2(\Omega)}. \end{aligned}$$

Thus, the time-dependent case is reduced to finding an optimal design for the initial conditions of the transported QoI c, which must be found and coincides with the stationary case.

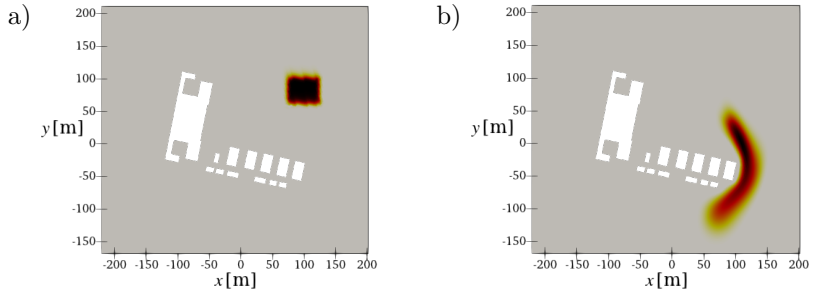

Fig. 5. Illustration of an example QoI $\mathbf{1}[T_0^{\mathrm{QoI}}, T_{\mathrm{final}}^{\mathrm{QoI}}] \times P$ (kept constant over time (a)) and solution $\mathcal{K}^*(\mathbf{1}[T_0^{\mathrm{QoI}}, T_{\mathrm{final}}^{\mathrm{QoI}}] \times P)(t = 0, \cdot)$ of transport problem induced by adjoint operator $\mathcal{K}^*$ (b).

For the numerical evaluation of this objective function and its gradient, which are required to minimize the regularized design function using the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with bound constraints (L-BFGS-B) solver, we start again with the inverse low-rank approximation of the Hessian from Eq. (10) and proceed to compute the design function as follows:

$$\Phi(\mathbf{w}) = \langle c, \mathcal{H}^{-1}(\mathbf{w})c\rangle_{L^2} \approx c_h^T \mathcal{A}_h^{-1}(I - V_r D_r V_r^T)\mathcal{A}_h^{-1} c_h.$$

For a shorter notation, we introduce $\widehat{q}_h := (I - V_r D_r V_r^T)\mathcal{A}_h^{-1} c_h$, $q_h := \mathcal{A}_h^{-1}\widehat{q}_h$ and obtain $\Phi(\mathbf{w}) \approx c_h^T q_h$. Hence, the calculation of the of the trace reduces to a projection in the low-rank subspace and solutions of an elliptic problem $\mathcal{A}$, respectively $\mathcal{A}_h$, for which very fast solving strategies exist. To calculate the derivative, we follow [2] and conclude for this simplified case

$$\frac{\partial}{\partial \mathbf{w}_i}\Phi(\mathbf{w}) = (\mathcal{F}^i(q))^2 \approx (\mathcal{F}_h^i \circ \mathcal{A}_h^{-1}(\widehat{q}))^2. \tag{13}$$

So, this calculation can be replaced by a surrogate model for the preconditioned forward operator. In principle, this procedure can be extended to a stronger optimality criteria, such as A- or D-optimal designs.

4.2. Dynamic sensor steering based on goal-oriented optimal sensor placement

A method will now be presented which is capable of dynamically controlling a sensor in such a way that a greater knowledge of the actual contaminant concentration can be generated. We assume that some knowledge about the concentration is already available from certain stationary sensors, i.e., the true contaminant field already possesses an appreciable concentration at these sensor locations, so that the inverse problem can be solved. This scenario can be seen in Fig. 6. We then define the QoI so that its center point is at the maximum of the reconstructed initial condition. The optimum design is then calculated on

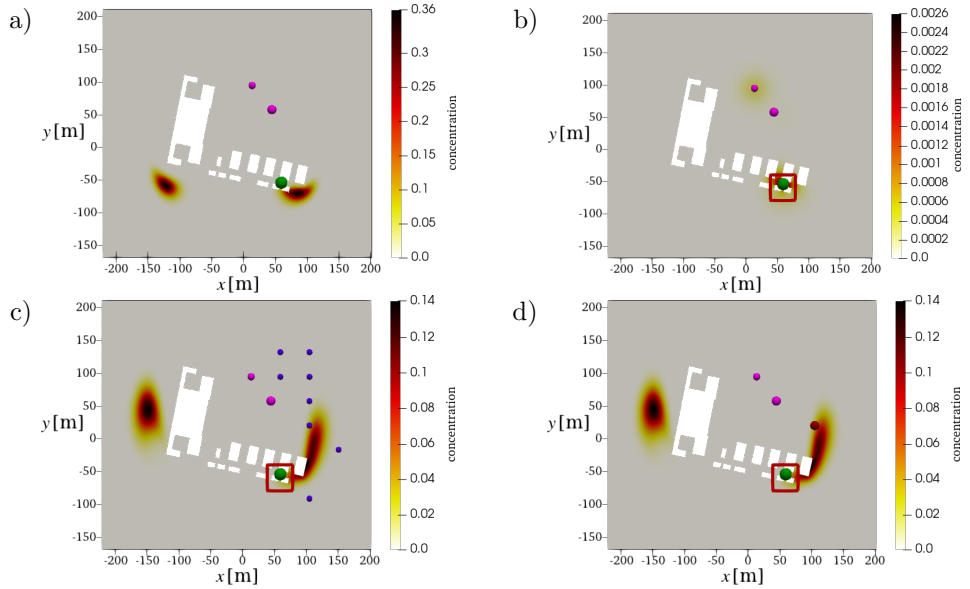

FIG. 6. Dynamic sensor steering test case with stationary sensor (green) and mobile sensor (purple trajectory): a) true concentration field at $u(\cdot, t = 2.2\,\mathrm{s})$ during the second steering time step, b) maximum-a-posteriori estimate m_{map} of the inverse problem and algorithmically selected zone of interest (red square), c) optimal sensor design, d) target position of the mobile sensor (red sphere), true concentration field $u(\cdot, t = 7\,\mathrm{s})$ shown in background of (c) and (d) to visualize transport problem dynamics.

this basis and the sensor is steered to the position with the highest weight $\mathbf{w}$. After the next measurement is collected, the procedure is repeated from the beginning. This method is schematically shown in Fig. 7. In this way, we obtain a trajectory $\gamma : \{t_0^{\mathrm{obs}}, ..., t_s^{\mathrm{obs}}\} \to \{\mathbf{x}_0^{\mathrm{obs}}, ..., \mathbf{x}_s^{\mathrm{obs}}\}$ for the steered sensor.

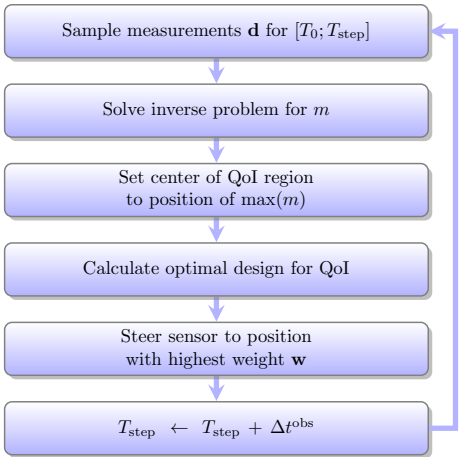

FIG. 7. Algorithm for dynamic sensor steering based on goal-oriented sensor placement [22, 23].

5. Numerical results of optimal experimental designs

In order to simulate scenarios on real-world domains, we use a highly automated process for grid generation. Building imprints as obstacles for two-dimensional contaminant transport are imported directly from OpenStreetMap (OSM), and locally refined triangular meshes are generated for the region of interest [7]. The forward model is implemented with stabilized linear Lagrange finite elements in the software framework `FEniCs` [6]. The `FEniCs` extension `hIPPYlib` (Inverse Problem PYthon library [20]) is used for the implementation of the inverse problem.

In the three following inverse problem OED examples, we use the forward simulation of Eq. (1) illustrated in Fig. 2, as the ground truth. Two radially symmetric functions:

$$m_{\mathbf{x}_s}(\mathbf{x};\mathbf{x}_s,r) = \min\left\{0.5, \exp\left(-\ln(\epsilon)||\mathbf{x}-\mathbf{x}_s||_2^2/r^2\right)\right\}, \quad \epsilon = 0.001,$$

describe the initial concentration field defined as:

$$\begin{aligned} u_0(\mathbf{x}) = {} & m_{\mathbf{x}_s}(\mathbf{x};\mathbf{x}_s = [-100\,\mathrm{m}, -80\,\mathrm{m}], r = 25\,\mathrm{m}) \\ & + m_{\mathbf{x}_s}(\mathbf{x};\mathbf{x}_s = [75\,\mathrm{m}, -80\,\mathrm{m}], r = 25\,\mathrm{m}). \end{aligned} \tag{14}$$

The initial concentration field is transported by the vector field $\mathbf{v}$. For the considered test cases, we estimate the stationary wind vector field as the solution of the incompressible Navier–Stokes equations with wind entering the given geometry from the south at $\mathbf{v} = 10\,\mathrm{m\,s^{-1}}$. This condition is realized using a Dirichlet boundary condition. In the inner boundaries that represent the imprints of the buildings, a no-slip condition is applied. The remaining edges correspond to free boundary conditions. For the chosen Reynolds number of 50, we obtain the laminar wind field visualized in Fig. 2. Moreover, the diffusion coefficient is selected as $\kappa = 1\,\mathrm{m^2\,s^{-1}}$, resulting in a transport problem with a moderate Peclet number. Finally, the time step size for the implicit Euler time-stepping scheme is set to 0.05 s. In the parametrization of the prior, the constants were chosen as $\eta = 8$ and $\gamma = 800$, yielding the operator $\mathcal{A} := 8\,I - 800\,\Delta$.

In order to make this problem computationally feasible, reduced-order models (ROMs) of the forward and adjoint operators are derived. Considering the forward operator $\mathcal{F}_h : \mathbb{R}^{n_{\mathrm{dof}}} \to \mathbb{R}^q$, it is observed that it constitutes a linear mapping from a high-dimensional to a lower-dimensional space. Thus, a singular value decomposition is performed to construct a ROM; see also [13, 15]. The decomposition provides singular values $\lambda_1 \geq \ldots \geq \lambda_r$, an L^2-orthogonal basis $U_r = (u_1, \ldots, u_r) \in \mathbb{R}^{n_{\mathrm{dof}} \times r}$ and an orthogonal basis $V_r = (v_r, \ldots, v_r) \in \mathbb{R}^{q \times r}$. During the online phase, for example when the precomputed ROM is used for sensor steering, the selected initial condition is projected so that only matrix-vector multiplications are required:

$$\mathcal{F}_h(m_h) \approx V_r\, D_r\, U_r\, M\, m_h,$$

where $m_h \in \mathbb{R}^{n_{\text{dof}}}$ and $D_r = \text{diag}(\lambda_1, ..., \lambda_r)$.

As discussed in Eq. (9) and Eq. (13), it is also viable to construct a ROM directly for the preconditioned forward operator. The singular values of the operators $\mathcal{F}_h$ and $\mathcal{F}_h \circ \mathcal{A}$ are compared in Fig. 8a. It is observed that the singular values of the preconditioned operator decay faster and therefore the number of computations that have to be performed to construct a ROM with acceptable accuracy is reduced. Fig. 8b shows that the ROM approximates the forward operator fairly well. Furthermore, the computed singular value decomposition is reused to approximate the adjoint operator $\mathcal{F}_h^* : \mathbb{R}^q \to \mathbb{R}^{n_{\text{dof}}}$ with a ROM as well, namely, $\mathcal{F}_h^*(y) \approx M\, U_r\, D_r\, V_r\, y$, for $y \in \mathbb{R}^q$.

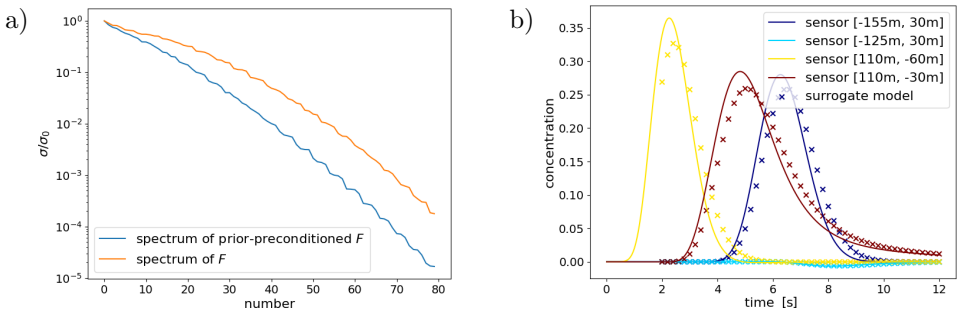

FIG. 8. Reduced-order modeling. Decay of singular values of the $\mathcal{F}_h$ and of preconditioned $\mathcal{F}_h \circ \mathcal{A}$ (a), comparison of the ROM with the forward model $\mathcal{F}_h$ evaluated at the sensor positions (b).

To assess the benefits of the reduced model in the context of forward evaluations, we first performed a single evaluation of the full-order model, which required approximately 1 s on our hardware. To simulate a scenario relevant to optimal sensor placement, the reduced model was used to evaluate the full-order model at 96 spatial positions over 90 time steps, resulting in a total of $q = 8640$ measurements. In total, we computed 200 spectral values. On average, each evaluation of the reduced model took 6.25 ms, yielding a relative speedup of approximately 160× compared to the full-order model.

5.1. OED 1. Sensor configuration to reconstruct initial condition in a critical area

As the first example of a goal-oriented optimal experimental design, we address the problem of identifying an optimal sensor layout to recover the initial condition in a defined subset of the computational domain $P_1 := \{(x, y) \in \Omega \,|\, 75 \leq x \leq 125, -100 \leq y \leq -60\}$. In a practical application, P_1 might represent a critical area of a chemical plant site where hazardous material is stored.

The inverse problem is posed under the assumption that only stationary sensors are used. These sensors sample the concentration at a rate of 5 Hz, beginning at $T_0 = 2\,\mathrm{s}$. Measurements taken after 12 s are not taken into account. A noise variance of $\sigma^2 = (0.005)^2$ is assumed, resulting in a signal-to-noise ratio of approximately SNR ≈ 100. Moreover, a regularization parameter of $\alpha = 0.1$ is applied to obtain a sparse sensor configuration, see Eq. (11). The selected domain where the QoI is inferred is indicated in Fig. 4a, along with the optimal sensor configuration (Fig. 4b). Moreover, the point-wise variance, which represents the uncertainty in the reconstruction, is also illustrated in Fig. 4. The solution to the inverse problem represented by m_{map} is visualized in Fig. 9. The numerical result demonstrates a reconstruction quality in P_1 comparable to that achieved using the full configuration with 96 sensors (Fig. 3), despite using five optimally selected sensors in OED 1.

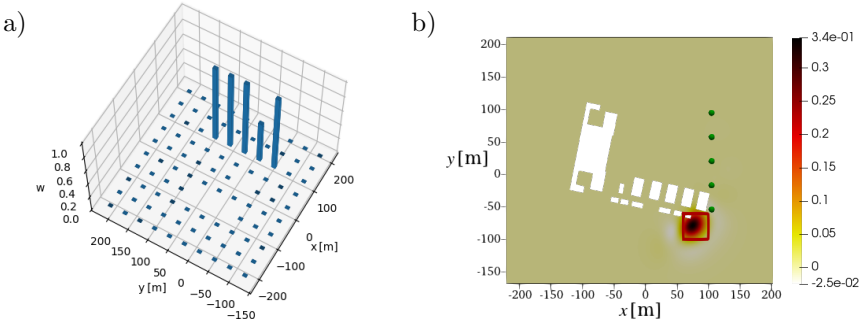

Fig. 9. OED 1. Weights of the optimal sensor configuration for monitoring P_1 (a) and reconstructed initial condition ((b), ground truth shown in Fig. 2).

5.2. OED 2. Configuration to monitor concentration evolution in a critical area

In the second scenario, our aim is to secure a specific area for a given time period. To achieve this, the QoI is defined to depend on the state u. To create a meaningful scenario, we shift $P_2 := \{(x, y) \in \Omega \,|\, 75 \leq x \leq 125,\ 60 \leq y \leq 100\}$ upwards. The goal of OED 2 is to ensure that concentration values can be predicted correctly in region P_2 during the time interval from 5 s to 12 s. This setup results in the following operator:

$$\mathcal{P}_2(m) = \int_{T_0^{\mathrm{QoI}}=5\,\mathrm{s}}^{T_{\mathrm{final}}^{\mathrm{QoI}}=12\,\mathrm{s}} \int_{P_2} \mathcal{K}(m)(t, \mathbf{x})\, dt\, d\mathbf{x} = \int_{T_0^{\mathrm{QoI}}=5\,\mathrm{s}}^{T_{\mathrm{final}}^{\mathrm{QoI}}=12\,\mathrm{s}} \int_{P_2} u(t, \mathbf{x})\, dt\, d\mathbf{x}.$$

The sensor weights $\mathbf{w}$ in Fig. 10 are computed using a regularization parameter of $\alpha = 1.0$. In this case as well, the source relevant to the QoI is reconstructed accurately using the optimized sensor configuration. The reconstructed initial

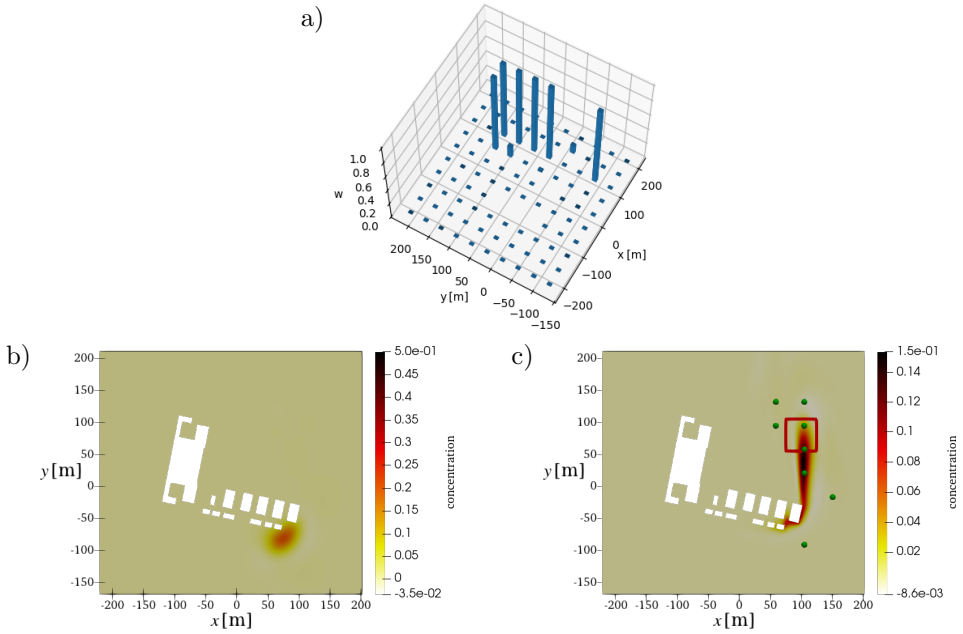

FIG. 10. OED 2. Weights of the optimal configuration for monitoring the concentration evolution in $\mathcal{P}_2$ (a), reconstructed initial condition (b), predicted concentration for $T = 10\,\mathrm{s}$ (c).

condition and the corresponding prediction are shown in Fig. 10. As illustrated in Fig. 11, the reduction in uncertainty is concentrated primarily in the region relevant to QoI compared with the complete sensor configuration.

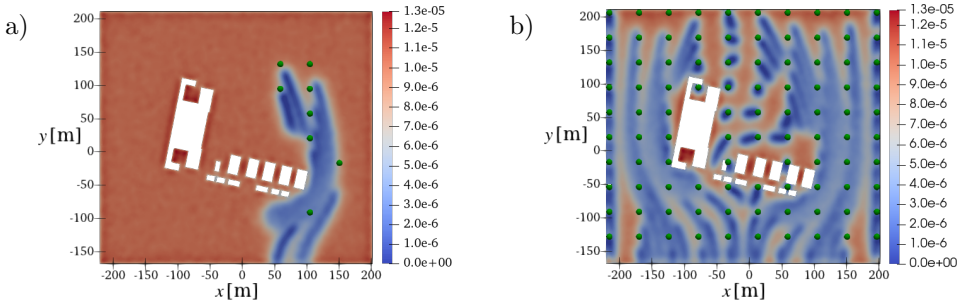

FIG. 11. OED 2. Point-wise variance σ^2 as a measure of uncertainty in the inferred parameter for the optimal configuration for $\mathcal{P}_2$ (a) compared to the full sensor grid (b).

5.3. OED 3. Dynamic sensor steering for source identification

Finally, the sensor steering method described in Subsec. 4.2 is tested in a numerical application case with $\kappa = 10\,\mathrm{m}^2\,\mathrm{s}^{-1}$. To steer the sensor, a much finer sensor grid consisting of 1511 possible sensor locations is used. In this setup, the moving sensor is allowed to take one step on this grid per cycle, which

needs 0.2 s and thus corresponds to a measurement frequency of 5 Hz. This corresponds to a speed of approximately $40\,\mathrm{m\,s^{-1}}$ for the moving sensor. To demonstrate the capabilities of the sensor steering approach, we placed a single stationary sensor just behind one of the obstacles. However, due to the transport characteristics in this region, information solely from this stationary sensor results in an inaccurate reconstruction of the source, which grossly underestimates the degree of contamination further from the buildings. In addition to the stationary sensor, measurements from a mobile sensor are available. The measurement process begins at time $T_0 = 2\,\mathrm{s}$, with data collected at a frequency of 5 Hz. The state u at $T_{\mathrm{step}} = T_0 = 2\,\mathrm{s}$ is shown in Fig. 12a. At this point, the stationary sensor receives very limited information and is unable to provide an accurate source estimate. However, computing the optimal sensor design based

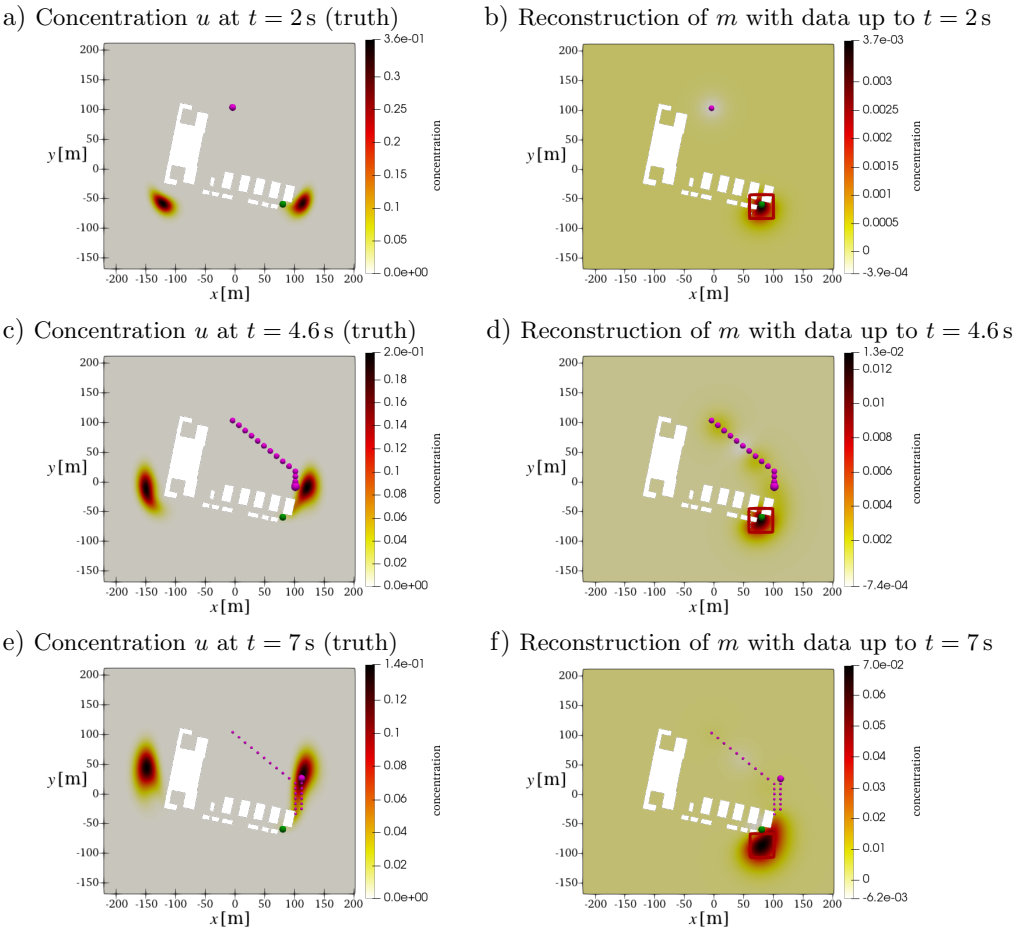

Fig. 12. OED 3. Data fusion of stationary sensor (marked in green) and mobile sensor (marked in purple). Mobile sensor steered to maximize information gain (trajectory marked with small purple spheres).

on the current QoI, defined as the integral over a square measuring 40 m on each side, centered on the maximal point of the reconstructed initial condition, yields favorable estimates for informative measurement positions. The mobile sensor is then directed toward the location associated with the highest weight $\mathbf{w}$, as determined by the C-optimal design criterion, computed over the observation period $[T_{\text{step}}, T_{\text{step}} + 2\,\text{s}]$ with the same sampling rate of 5 Hz, wherein we take as sensor weights for times prior to T_{step}, the actual past locations of the sensor. In the subsequent time steps, illustrated in Fig. 12c, the sensor continues to move toward regions of increasing concentration. In Fig. 12d, at $T_{\text{step}} = 4.6\,\text{s}$, the mobile sensor reaches the core of the contaminant. Finally, Fig. 12e and Fig. 12f demonstrate that the mobile sensor continues to accurately track the contaminant in further time steps. Comparing the performance of the stationary sensor alone with the combination of a stationary sensor and dynamically steered one, we find, as depicted in Fig. 13, that the mobile sensor achieves substantially improved reconstruction accuracy after just 7 s, whereas the stationary sensor fails to produce a reliable estimate even after 12 s.

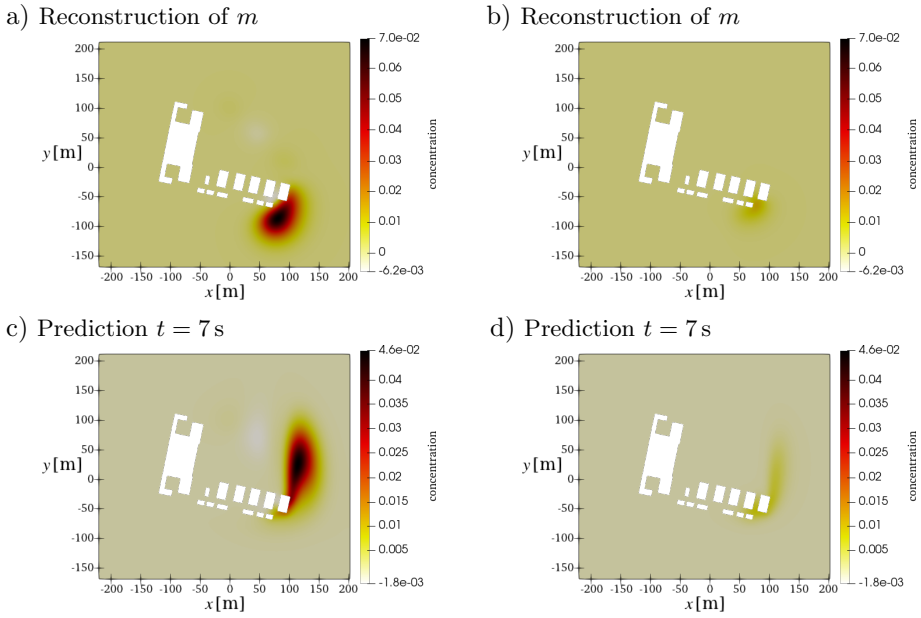

Fig. 13. OED 3. Comparison of predictions with mobile sensor and stationary sensor (a,c) and stationary sensor alone (b,d).

6. Conclusion and outlook

This paper investigated a novel approach for goal-oriented optimal static sensor placement and dynamical sensor steering for PDE-constrained problems. Adopting previous work by Wogrin *et al.* [22, 23] on dynamic sensor steering,

we leveraged a Bayesian approach for the solution of the inverse problem, accelerated by offline low-rank approximations of the Hessian matrix and an online preconditioned inexact Newton-CG method. The resulting framework was then applied to a more complex geometry extracted from real-world map data. We showcased the strengths of the proposed workflow i three application cases in the field of airborne contaminant transport. In the first example, we derived an optimal placement of stationary sensors to recover the initial condition inside a spatially constrained rectangular region. The results showed that our proposed method only required five sensors to reconstruct the initial condition locally with accuracy comparable to the full configuration of 96 sensors. In our second example, we extended the QoI in the sense that a RoI was monitored not only at a specific time instance, but also over a fixed time period. From a practical point of view, this corresponds to the goal of securing a specific area for a given time period. Using only eight optimally placed sensors, the evolution of the concentration was accurately reconstructed and the uncertainty was minimized in the area of interest. Lastly, we investigated a dynamic sensor steering problem. Here, we showed that while we were still able to roughly predict the general shape of the initial condition with an unfavorably placed stationary sensor, adding a mobile sensor significantly improved the reconstruction while simultaneously reducing the required measurement time to one-third compared to the stationary case. This proves that the presented method is able to successfully handle the complexity of moving sensors and steer them to achieve fast and reliable reconstruction of unknown (in practice) initial condition.

While we believe this work to be an important step toward the optimal steering of unmanned sensor platforms in crisis situations, there still remain several points for improvement and further investigation of the proposed algorithm is needed. One potential area for improvement concerns the mathematical formulation and solution of the inverse problem. In many applications, it is reasonable to assume that the initial condition is sparse. Incorporating this additional knowledge into the solution procedure is expected to speed up computations and further improve real-time capabilities of the method [17]. Moreover, we plan to extend the sensor steering methodology to a reinforcement learning approach, where the position and size of the QoI at each step is determined by an agent that was previously trained based on trial-and-error interactions with the forward model [12].

Acknowledgements

Daniel Wolff and Alexander Popp gratefully acknowledge funding by Digitalization and Technology Research Center of the Bundeswehr (dtec.bw) (project RISK.twin). dtec.bw is funded by the European Union – NextGenerationEU.

References

1. Alexanderian A., Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: A review, *Inverse Problems*, **37**(4): 043001, 2021, https://doi.org/10.1088/1361-6420/abe10c.
2. Alexanderian A., Petra N., Stadler G., Ghattas O., A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized ℓ_0-sparsification, *SIAM Journal on Scientific Computing*, **36**(5): A2122–A2148, 2014, https://doi.org/10.1137/130933381.
3. Alexanderian A., Saibaba A.K., Efficient D-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems, *SIAM Journal on Scientific Computing*, **40**(5): A2956–A2985, 2018, https://doi.org/10.1137/17M115712X.
4. Alt H.W., *Lineare Funktionalanalysis*, Springer, Berlin, Heidelberg, 2012, https://doi.org/10.1007/978-3-642-22261-0.
5. Attia A., Alexanderian A., Saibaba A.K., Goal-oriented optimal design of experiments for large-scale Bayesian linear inverse problems, *Inverse Problems*, **34**(9): 095009, 2018, https://doi.org/10.1088/1361-6420/aad210.
6. Baratta I.A *et al.*, DOLFINx: The next generation FEniCS problem solving environment, *Zenodo*, 2023, https://doi.org/10.5281/zenodo.10447666.
7. Bonari J., Kühn L., von Danwitz M., Popp A., Towards real-time urban physics simulations with digital twins, [in:] *2024 28th International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*, Urbino, Italy, pp. 18–25, 2024, https://doi.org/10.1109/DS-RT62209.2024.00013.
8. Cvetković N., Lie H.C., Bansal H., Veroy K., Choosing observation operators to mitigate model error in Bayesian inverse problems, *SIAM/ASA Journal on Uncertainty Quantification*, **12**(3): 723–758, 2024, https://doi.org/10.1137/23M1602140.
9. Daon Y., Stadler G., Mitigating the influence of the boundary on PDE-based covariance operators, *Inverse Problems and Imaging*, **12**(5): 1083–1102, 2018, https://doi.org/10.3934/ipi.2018045.
10. Elman H.C., Su T., A low-rank solver for the stochastic unsteady Navier–Stokes problem, *Computer Methods in Applied Mechanics and Engineering*, **364**: 112948, 2020, https://doi.org/10.1016/j.cma.2020.112948.
11. Evans L.C., *Partial Differential Equations*, Graduate Studies in Mathematics, Vol. 19, American Mathematical Society, Providence, Rhode Island, 2nd ed., 2022.
12. Fricke C., Wolff D., Kemmerling M., Elgeti S., Investigation of reinforcement learning for shape optimization of 2D profile extrusion die geometries, *Advances in Computational Science and Engineering*, **1**(1): 1–35, 2023, https://doi.org/10.3934/acse.2023001.
13. Halko N., Martinsson P.G., Tropp J.A., Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Review*, **53**(2): 217–288, 2011, https://doi.org/10.1137/090771806.
14. Isaac T., Petra N., Stadler G., Ghattas O., Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet, *Journal of Computational Physics*, **296**: 348–368, 2015, https://doi.org/10.1016/j.jcp.2015.04.047.

15. Liberty E., Woolfe F., Martinsson P.-G., Rokhlin V., Tygert M., Randomized algorithms for the low-rank approximation of matrices, *Proceedings of the National Academy of Sciences*, **104**(51): 20167–20172, 2007, https://doi.org/10.1073/pnas.0709640104.

16. Petra N., Stadler G., *Model variational inverse problems governed by partial differential equations*, ICES REPORT 11-05, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, 2011, https://math.nyu.edu/∼stadler/papers/PetraStadler11.pdf.

17. Pieper K., Walter D., Linear convergence of accelerated conditional gradient algorithms in spaces of measures, *ESAIM: Control, Optimisation and Calculus of Variations*, **27**: 38, 2021, https://doi.org/10.1051/cocv/2021042.

18. Spantini A., Cui T., Willcox K., Tenorio L., Marzouk Y., Goal-oriented optimal approximations of Bayesian linear inverse problems, *SIAM Journal on Scientific Computing*, **39**(5): S167–S196, 2017, https://doi.org/10.1137/16M1082123.

19. Steihaug T., Local and superlinear convergence for truncated iterated projections methods, *Mathematical Programming*, **27**(2): 176–190, 1983, https://doi.org/10.1007/BF02591944.

20. Villa U., Petra N., Ghattas O., hIPPYlib: An extensible software framework for large-scale inverse problems governed by PDEs: Part I: Deterministic inversion and linearized Bayesian inference, *ACM Transactions on Mathematical Software*, **47**(2): 16, 2021, https://doi.org/10.1145/3428447.

21. von Danwitz M., Bonari J., Franz P., Kühn L., Mattuschka M., Popp A., Contaminant dispersion simulation in a digital twin framework for critical infrastructure protection, [in:] *9th European Congress on Computational Methods in Applied Sciences and Engineering*, CIMNE, 2024, https://doi.org/10.23967/eccomas.2024.301.

22. Wogrin S., *Model Reduction for Dynamic Sensor Steering: A Bayesian Approach to Inverse Problems*, Master's Thesis, Massachusetts Institute of Technology, Computation for Design and Optimization Program, 2008, http://hdl.handle.net/1721.1/43739.

23. Wogrin S., Singh A., Allaire D., Ghattas O., Willcox K., From data to decisions: A real-time measurement–inversion–prediction–steering framework for hazardous events and health monitoring, [in:] Darema F., Blasch E.P., Ravela S., Aved A.J. [Eds], *Handbook of Dynamic Data Driven Applications Systems*, pp. 195–227, Springer, Cham, 2023, https://doi.org/10.1007/978-3-031-27986-7_8.

24. Wu K., Chen P., Ghattas O., An offline-online decomposition method for efficient linear Bayesian goal-oriented optimal experimental design: Application to optimal sensor placement, *SIAM Journal on Scientific Computing*, **45**(1): B57–B77, 2023, https://doi.org/10.1137/21M1466542.

Received July 2, 2025; revised November 24, 2025; accepted January 18, 2026; available online February 9, 2026; version of record April 20, 2026; published issue June 17, 2026.

Contents