Check for updates

Received: 9 April 2026 | Revised: 10 September 2026 | Accepted: 23 October 2026

DOI: 10.1002/gamm.70017

ORIGINAL ARTICLE

GAMM Mitteilungen

Conformalized polynomial chaos expansion for uncertainty-aware surrogate modeling

Dimitrios Loukrezis[1] | **Dimitris G. Giovanis**[2]

[1]Scientific Computing, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

[2]Department of Civil & Systems Engineering, Johns Hopkins University, Baltimore, Maryland, USA

Correspondence
Dimitrios Loukrezis, Scientific Computing, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands.
Email: d.loukrezis@cwi.nl

Funding information
U.S. Department of Energy, Grant/Award Number: DE-SC0024162

Abstract

This work introduces a method to equip data-driven polynomial chaos expansion surrogate models with intervals that quantify the predictive uncertainty of the surrogate. To that end, jackknife-based conformal prediction is integrated into regression-based polynomial chaos expansions. The jackknife algorithm uses leave-one-out residuals to generate prediction intervals around the predictions of the polynomial chaos surrogate. The jackknife+ extension additionally requires leave-one-out model predictions. Both methods allow using the entire dataset for model training and do not require a hold-out dataset for prediction interval calibration. The key to efficient implementation is to leverage the linearity of the polynomial chaos regression model, so that leave-one-out residuals and, if necessary, leave-one-out model predictions can be computed with analytical, closed-form expressions. This eliminates the need for repeated model re-training. The conformalized polynomial chaos expansion method is first validated on benchmark models and then applied to an engineering use case. The method produces prediction intervals that provide the target coverages, even for low-accuracy models trained with small datasets. At the same time, training data availability plays a crucial role in improving the empirical coverage and narrowing the prediction interval, as well as in reducing their variability over different training datasets.

KEYWORDS

conformal prediction, jackknife, polynomial chaos expansion, regression, surrogate modeling, uncertainty quantification

1 | INTRODUCTION

Originally developed as an efficient uncertainty quantification (UQ) method [1–3], the polynomial chaos expansion (PCE) is in recent years routinely employed as a machine learning (ML) regression model [4]. Within this context, regression-based PCEs are often used as data-driven surrogate models in various science and engineering

© 2026 Wiley-VCH GmbH.
All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

applications [5–10]. Their use has also been extended to scientific ML tasks like manifold learning [11–14], operator learning [15], and physics-informed ML [16, 17].

In contrast to the use of PCEs for UQ purposes, PCE-based data-driven predictive models are purely deterministic. That is, they yield point predictions that lack any type of uncertainty or confidence metrics. Bayesian regression models [18] natively provide such predictive uncertainty metrics. Examples include Gaussian processes (GPs) [19], Bayesian neural network (NN) [20], but also Bayesian PCEs [21–23]. Considering frequentist approaches, ensemble modeling and bootstrapping are common ways of providing predictive uncertainty intervals to data-driven models, including PCEs [9, 24–25]. Other relevant approaches include combined PCE-GP models, where earlier works embed the PCE in the trend of a GP model [26, 27], whereas more recently suggested hybrid PCE-GP models [28, 29] allow for propagating the GP uncertainty directly to the PCE coefficients. Other alternatives include stochastic PCE formulations [30]. However, these approaches induce either specific modeling limitations (e.g., Gaussianity assumptions in Bayesian approaches) or significant computational overhead (e.g., due to model retraining in bootstrapping and ensemble modeling).

Considering predictive UQ for ML models, conformal prediction (CP) [31, 32] is an established method for producing statistically valid prediction intervals given a point predictor, only under the assumption of data exchangeability. The resulting predictive models are commonly called "conformalized" [33–37]. CP is a distribution-free UQ method, meaning that it does not require any assumptions about the underlying data distribution or the structure of the data-driven model. Additionally, certain CP methods provide statistically rigorous, finite-sample guarantees regarding prediction interval coverage. On the downside, CP does not yield full probability distributions like Bayesian methods do. Moreover, its coverage is marginal instead of conditional, hence possibly under- or overconfident in different regions of the input space. Adaptive CP methods that address the latter issue are actively pursued [38, 39]. Data demand can also be an issue, in particular considering so-called inductive or split CP methods [40, 41], which require that part of the available data is excluded from training and reserved for prediction interval estimation. Remedies have been sought in so-called cross CP approaches leveraging cross-validation techniques [42, 43], including jackknife-based variants that utilize leave-one-out (LOO) residuals [40, 44]. Similar to bootstrapping and ensemble modeling, the downside of these methods is the need for repeated model retraining, which can result in undesired computational overhead.

This work proposes the combination of regression-based PCEs with CP based on the jackknife method [40] and its jackknife+ extension [44], such that the resulting conformalized PCE models are equipped with intervals quantifying their predictive uncertainty. Its main contribution is demonstrating that the linearity of the PCE regression model allows the computation of the LOO residuals and LOO model evaluations used in jackknife-based CP with analytic, closed-form expressions. This bypasses the need to repeatedly re-train the PCE, as is typically necessary for jackknife-based CP, thus effectively eliminating the computational overhead of predictive UQ. Additionally, the suggested approach belongs to the class of cross CP methods, and as such allows the full available dataset to be used for training, which is a crucial advantage in small-data learning regimes. We note that combining CP with data-driven, regression-based surrogate models has been explored before, for example, for GPs [35] or various NNs architectures [45]. To the best knowledge of the authors, this is the first work to develop a conformalized PCE method where PCE regression models are equipped with predictive UQ intervals based on CP.[1]

The method is implemented for total-degree PCEs utilizing ordinary least squares (OLS) for the computation of the PCE coefficients. Other types of bases, for example, corresponding to sparse PCEs, can also be considered; however, for the aforementioned analytical computation of the jackknife-based CP intervals to be possible, the final PCE basis must be held fixed, and the coefficients must be computed via OLS (see Section 2.3). The method is first validated on benchmark models and then applied to an engineering use-case concerning the design of a Stern-Gerlach electromagnet [47]. In all numerical experiments, we assess whether the conformalized PCE surrogate provides predictive uncertainty intervals that result in empirical coverages similar to the required target levels. Additionally, we investigate the impact of different conformalized PCE configurations, with an emphasis on training data availability.

The remainder of this article is organized as follows: Section 2 presents the methodology behind the suggested method, first recalling the necessary prerequisites of PCE and CP methods, and then presenting the conformalized PCE. The numerical experiments are presented in Section 3. Section 4 summarizes the findings of this work and discusses follow-up research opportunities.

[1]In the meantime, another work on conformalized PCEs has appeared in the literature [46]. This work partially overlaps with ours; however, it was developed independently.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | METHODOLOGY

We begin by presenting the PCE regression model in Section 2.1, followed by a presentation of CP in Section 2.2. The latter focuses on the jackknife and jackknife+ CP methods. Section 2.3 shows how LOO residuals and predictions of PCE regression models can be computed with analytical, closed-form expressions, as to enable the cost-effective estimation of prediction intervals based on jackknife and jackknife+ CP.

2.1 | Regression-based polynomial chaos expansion

In the following, we consider a computational model, denoted as $\mu(\mathbf{x})$, which receives an input vector $\mathbf{x} \in \mathbb{R}^N$ and computes a scalar output $y \in \mathbb{R}$, such that $y = \mu(\mathbf{x})$. While scalar outputs are considered throughout this work, vector-valued outputs can also be considered by an element-wise application of all operations described next. The form of the model can range from a simple analytical function to a complex numerical solver. The only assumption is that the model is deterministic, that is, the same output y is observed given the same input $\mathbf{x}$.

Next, we introduce a multivariate random variable (or random vector) $\mathbf{X} = (X_1, \dots, X_N)$, where the individual random variables $X_n, n = 1, \dots, N$, are assumed to be mutually independent. The random vector is defined on the probability space (Ω, Σ, P), where Ω is the sample space containing the outcomes $\omega \in \Omega$, Σ denotes the σ-algebra of events (measurable grouped outcomes), and $P : \Sigma \to [0, 1]$ is a probability measure. The latter is defined based on the assumed probability density function (PDF) $f_{\mathbf{X}} : \Xi \to \mathbb{R}_{\geq 0}$, where $\Xi \subset \mathbb{R}^N$ denotes the state-space of the random vector. Then, the random vector is a map, $\mathbf{X} : \Omega \to \Xi$, with random realizations $\mathbf{x} = \mathbf{X}(\omega) \in \Xi$.

Next, we assume that the random realizations $\mathbf{x} = \mathbf{X}(\omega)$ coincide with the inputs of the computational model $\mu(\mathbf{x})$ and that the Doob-Dynkin lemma is satisfied [48]. Then, by propagation of uncertainty, the model's response is a random variable dependent on the input random vector, that is, $Y = \mu(\mathbf{X})$. Provided that its variance is finite, the random output can be represented in the form of a PCE, such that

$$y = \mu(\mathbf{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}_0^N} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{x}), \tag{1}$$

where $\Psi_{\boldsymbol{\alpha}}$ are multivariate polynomials, $c_{\boldsymbol{\alpha}} \in \mathbb{R}$ are the associated coefficients, and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N)$ are unique multi-indices defined by the polynomial's partial degree in each input dimension. The polynomials are chosen such that they form an orthogonal basis with respect to the input PDF, that is, they satisfy

$$\mathbb{E}\left[\Psi_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\beta}}\right] = \int_{\Xi} \Psi_{\boldsymbol{\alpha}}(\mathbf{x}) \Psi_{\boldsymbol{\beta}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x} = \mathbb{E}\left[\Psi_{\boldsymbol{\alpha}}^2\right] \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}. \tag{2}$$

For simplicity, we assume an orthonormal basis, such that $\mathbb{E}\left[\Psi_{\boldsymbol{\alpha}}^2\right] = 1$.

In practice, the infinite series (1) must be truncated to a finite number of K terms based on a set of selected multi-indices. This truncation yields a PCE approximation of the form

$$y = \mu(\mathbf{x}) \approx \widehat{\mu}(\mathbf{x}) = \sum_{k=1}^{K} c_k \Psi_k(\mathbf{x}) = \sum_{\boldsymbol{\alpha} \in \Lambda} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{x}), \tag{3}$$

where $\Lambda \subset \mathbb{N}_0^N$ is a multi-index set with cardinality $\#\Lambda = K$. Each index $k = 1, \dots, K$ corresponds to a unique multi-index $\boldsymbol{\alpha} \in \Lambda$. One common truncation method is to use a so-called total-degree basis, where $\Lambda = \{\boldsymbol{\alpha} : |\boldsymbol{\alpha}|_1 \leq P\}$ and P denotes the maximum polynomial degree. The size of the total-degree basis is equal to $K = \frac{(N+P)!}{N!P!}$ and becomes intractable for high polynomial degrees or input dimensions. To alleviate this problem, several methods have been suggested for constructing sparse bases [49], for example, by means of dimension-adaptive algorithms [50].

Given a multi-index set Λ and the corresponding polynomial basis, the remaining task is to estimate the PCE coefficients $\mathbf{c} = (c_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \Lambda}^{\top} \in \mathbb{R}^{K \times 1}$. We opt here for a regression-based approach [[51–53]. Alternatives include pseudo-spectral projection [54, 55] or interpolation [56–58]. For the regression-based coefficient estimation, we assume a set of input parameter realizations $\left\{\mathbf{x}^{(m)}\right\}_{m=1}^{M}$, commonly called the experimental design, along with the corresponding model responses $\left\{y^{(m)}\right\}_{m=1}^{M}$. Put in an ML context, the combined dataset $\left\{\mathbf{x}^{(m)}, y^{(m)}\right\}_{m=1}^{M}$ constitutes the training dataset. We then

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

construct the so-called design matrix $\mathbf{D} \in \mathbb{R}^{M\times K}$, such that $d_{mk} = \Psi_k\left(\mathbf{x}^{(m)}\right)$. Collecting the model responses in a vector $\mathbf{y} = \left(y^{(1)}, \ldots, y^{(M)}\right)^\top \in \mathbb{R}^{M\times 1}$, the coefficients can be computed by solving the least squares minimization problem

$$\mathbf{c} = \underset{\hat{\mathbf{c}}\in\mathbb{R}^K}{\arg\min}\left\{\frac{1}{M}\|\mathbf{y} - \mathbf{D}\hat{\mathbf{c}}\|_2^2\right\}, \tag{4}$$

which admits the closed-form solution

$$\mathbf{c} = \left(\mathbf{D}^\top\mathbf{D}\right)^{-1}\mathbf{D}^\top\mathbf{y}. \tag{5}$$

For the regression problem (4) to be well posed, the experimental design must be at minimum equal to the polynomial basis. Larger experimental designs are typically needed for well-determined problems. In fact, the size of the experimental design is crucial for the stability and accuracy of the resulting PCE [52, 53]. This limitation can be overcome with sparse PCE methods, where the problem (4) is complemented with a penalty term that induces sparsity in its solution. Common techniques include the least absolute shrinkage and selection operator (LASSO) and compressive sensing [49].

2.2 | Conformal prediction

Following the notation introduced in Section 2.1, we consider a regression model $\widehat{\mu}(\mathbf{x})$ trained with the dataset $\mathcal{D}_{\text{train}} = \left\{\mathbf{x}^{(m)}, y^{(m)}\right\}_{m=1}^{M}$, where again the experimental design data points $\mathbf{x}^{(m)}$ are assumed to be realizations of random vector $\mathbf{X}$ characterized by a joint probability distribution $f_{\mathbf{X}}(\mathbf{x})$. Then, given a test point $\{\mathbf{x}^*, y^*\} \notin \mathcal{D}_{\text{train}}$, the trained model produces a prediction $\widehat{y}^* = \widehat{\mu}(\mathbf{x}^*)$ for the true value y^*. The goal of CP is to complement the model's prediction with an interval which is likely to contain the true output y^*. More formally, given a significance level $s \in (0, 1)$ corresponding to the target coverage level $(1 - s)$, CP constructs a prediction interval $\widehat{C}_{1-s}(\mathbf{x}^*)$, such that

$$\mathbb{P}\left\{y^* \in \widehat{C}_{1-s}(\mathbf{x}^*)\right\} \geq 1 - s. \tag{6}$$

That is, the probability of the true output being included in the interval is at least equal to the target coverage level. For example, setting $s = 0.05$ corresponds to a target coverage equal to 95%, meaning that the prediction interval has been constructed to capture the true test output values at least 95% of the time. Note that CP assumes that the training and test data are exchangeable, meaning that the joint data distribution is invariant to permutation.

A central concept of CP is the so-called non-conformity score, which measures how well a given data point conforms to the fitted model. To that end, a nonconformity function $\alpha\left(\mathbf{x}, y; \widehat{\mu}\right)$ is evaluated on a set of calibration points $\left\{\mathbf{x}^{(j)}, y^{(j)}\right\}_{j=1}^{J}$, thus producing the nonconformity scores $\alpha_j = \alpha\left(\mathbf{x}^{(j)}, y^{(j)}; \widehat{\mu}\right), j = 1, \ldots, J$. We denote the $(1 - s)$-quantile of the empirical distribution of the nonconformity scores $\left\{\alpha_j\right\}_{j=1}^{J}$ as $\widehat{q}_{1-s}\left\{\alpha_j\right\}$. The prediction interval for a test input $\mathbf{x}^*$ is then defined in general by

$$\widehat{C}_{1-s}(\mathbf{x}^*) = \left\{\, y : \ \alpha\left(\mathbf{x}^*, y; \widehat{\mu}\right) \leq \widehat{q}_{1-s}\left\{\alpha_j\right\}\right\}, \tag{7}$$

such that $y^* \in \widehat{C}_{1-s}(\mathbf{x}^*)$ implies that $\alpha\left(\mathbf{x}^*, y^*; \widehat{\mu}\right) \leq \widehat{q}_{1-s}\left\{\alpha_j\right\}$ and vice versa. A commonly used nonconformity score is the absolute residual, such that $\alpha_j = R_j = \left|r_j\right| = \left|y^{(j)} - \widehat{\mu}\left(\mathbf{x}^{(j)}\right)\right|$. This choice results in the construction of the symmetric prediction interval

$$\widehat{C}_{1-s}(\mathbf{x}^*) = \left[\widehat{\mu}(\mathbf{x}^*) - \widehat{q}_{1-s}\left\{R_j\right\}, \widehat{\mu}(\mathbf{x}^*) + \widehat{q}_{1-s}\left\{R_j\right\}\right]. \tag{8}$$

Note that this prediction interval does not generally adapt to possible data heteroskedasticity.

A naive approach to estimate the prediction interval (8) would be to use the residuals on the training data $\mathcal{D}_{\text{train}} = \left\{\mathbf{x}^{(m)}, y^{(m)}\right\}_{m=1}^{M}$. However, training data residuals are expected to be smaller compared to residuals obtained from previously unseen test data points. Hence, the resulting prediction interval will most probably not provide the required coverage. Split or inductive CP approaches [40, 41] use a separate validation dataset $\mathcal{D}_{\text{val}} \neq \mathcal{D}_{\text{train}}$ to compute the nonconformity scores and construct the prediction interval (8). Split CP guarantees the required finite-sample coverage (6),

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

however, this comes at the cost of a larger data demand. To avoid the need for additional data, cross CP [42, 43] utilizes K-fold cross-validation, such that the training dataset $\mathcal{D}_{\text{train}}$ is split into K disjoint subsets $\mathcal{D}_1, \dots, \mathcal{D}_K$ of (approximately) equal size. Then, the regression models $\widehat{\mu}_{\sim k}(\mathbf{x})$ are trained using the datasets $\mathcal{D}_{\sim k} = \mathcal{D}_{\text{train}} \setminus \mathcal{D}_k$, $k = 1, \dots, K$, and the absolute cross-validation residuals

$$R_m^{\text{cv}} = \left|r_m^{\text{cv}}\right| = \left|y^{(m)} - \widehat{\mu}_{\sim k(m)}\left(\mathbf{x}^{(m)}\right)\right|, \tag{9}$$

are used as nonconformity scores, where the index $k(m)$ denotes that $\left(\mathbf{x}^{(m)}, y^{(m)}\right) \in \mathcal{D}_k$. Cross CP typically yields an empirical coverage close to the target level $(1 - s)$, but does not provide formal coverage guarantees. A major downside of cross CP is that K different regression models must be trained, which can lead to an undesired computational cost.

A special case of cross-validation is the jackknife method [59]. Using the jackknife in the context of CP [40], the nonconformity scores coincide with the absolute LOO residuals

$$R_m^{\text{LOO}} = \left|r_m^{\text{LOO}}\right| = \left|y^{(m)} - \widehat{\mu}_{\sim m}\left(\mathbf{x}^{(m)}\right)\right|, \tag{10}$$

where the regression model $\widehat{\mu}_{\sim m}(\mathbf{x})$ is trained using the dataset $\mathcal{D}_{\sim m} = \mathcal{D}_{\text{train}} \setminus \left\{\left(\mathbf{x}^{(m)}, y^{(m)}\right)\right\}$. Then, the prediction interval (8) takes two equivalent forms, that is,

$$\widehat{C}_{1-s}^{\text{jack}}(\mathbf{x}^*) = \left[\widehat{\mu}(\mathbf{x}^*) - \widehat{q}_{1-s}\left\{R_m^{\text{LOO}}\right\}, \widehat{\mu}(\mathbf{x}^*) + \widehat{q}_{1-s}\left\{R_m^{\text{LOO}}\right\}\right] \tag{11a}$$

$$= \left[\widehat{q}_s\left\{\widehat{\mu}(\mathbf{x}^*) - R_m^{\text{LOO}}\right\}, \widehat{q}_{1-s}\left\{\widehat{\mu}(\mathbf{x}^*) + R_m^{\text{LOO}}\right\}\right]. \tag{11b}$$

Being a cross CP variant, jackknife CP also yields an empirical coverage close to the target level, without formal coverage guarantees. Barber et al. [44] suggested an extension of jackknife CP, called jackknife+, where the prediction interval is constructed as

$$\widehat{C}_{1-s}^{\text{jack+}}(\mathbf{x}^*) = \left[\widehat{q}_s\left\{\widehat{\mu}_{\sim m}(\mathbf{x}^*) - R_m^{\text{LOO}}\right\}, \widehat{q}_{1-s}\left\{\widehat{\mu}_{\sim m}(\mathbf{x}^*) + R_m^{\text{LOO}}\right\}\right]. \tag{12}$$

In contrast to the jackknife interval (11) which is centered at the prediction $\widehat{\mu}(\mathbf{x}^*)$ of the fully trained model, the jackknife+ interval (12) is based on the ensemble of the LOO predictions $\widehat{\mu}_{\sim m}(\mathbf{x}^*)$, $m = 1, \dots, M$. Hence, the jackknife+ interval (12) is asymmetric. This slight modification results in a formal coverage guarantee of $(1 - 2s)$, while the empirical coverage remains close to the target level $(1 - s)$, similar to cross CP and jackknife CP.

In practice, both the jackknife and jackknife+ CP methods achieve an empirical coverage close to the target level $(1 - s)$ and produce almost identical prediction intervals [44]. Moreover, both approaches entail training M regression models, thus being extreme cases of cross CP in terms of computational cost. However, for linear regression models like the PCE, the LOO residuals r_m^{LOO} and the LOO predictions $\widehat{\mu}_{\sim m}(\mathbf{x}^*)$ can be obtained without model retraining, as shown next.

2.3 | Conformalized polynomial chaos regression

We now assume that $\widehat{\mu}(\mathbf{x})$ is a regression-based PCE in the form of (3) and trained with a dataset $\mathcal{D}_{\text{train}} = \left\{\mathbf{x}^{(m)}, y^{(m)}\right\}_{m=1}^{M}$. Then, the prediction intervals (11) (for jackknife) and (12) (for jackknife+) can be estimated without the need to retrain the PCE M times.

Starting with the jackknife method, only the LOO residuals defined in (10) are needed to estimate the prediction interval (11). Recalling from Section 2.1 the design matrix $\mathbf{D} \in \mathbb{R}^{M \times K}$, the so-called hat matrix $\mathbf{H} \in \mathbb{R}^{M \times M}$ is defined as

$$\mathbf{H} = \mathbf{D}\left(\mathbf{D}^\top \mathbf{D}\right)^{-1} \mathbf{D}^\top. \tag{13}$$

The LOO residuals can then be computed analytically via the closed-form expression

$$r_m^{\text{LOO}} = \frac{y^{(m)} - \widehat{\mu}\left(\mathbf{x}^{(m)}\right)}{1 - h_{mm}}, \tag{14}$$

where $h_{mm} \in \mathbb{R}$ is the m-th diagonal term of $\mathbf{H}$.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

For the prediction interval (12) based on the jackknife+, the LOO predictions $\{\widehat{\mu}_{\sim m}(\mathbf{x}^*)\}_{m=1}^{M}$ for the test input $\mathbf{x}^*$ are additionally needed. Let us denote with $\mathbf{d}_m^\top \in \mathbb{R}^{1\times K}$ the m-th row of the design matrix $\mathbf{D} \in \mathbb{R}^{M\times K}$, which contains the evaluation of the PCE basis on the training input $\mathbf{x}^{(m)}$, such that $\mathbf{d}_m^\top = \left(\Psi_1\left(\mathbf{x}^{(m)}\right), \dots, \Psi_K\left(\mathbf{x}^{(m)}\right)\right)$. Accordingly, let us denote with $\mathbf{d}_*^\top \in \mathbb{R}^{1\times K}$ the vector of basis evaluations at the test input $\mathbf{x}^*$, such that $\mathbf{d}_*^\top = (\Psi_1(\mathbf{x}^*), \dots, \Psi_K(\mathbf{x}^*))$. Then, the LOO predictions can be analytically computed as

$$\begin{aligned}\widehat{\mu}_{\sim m}(\mathbf{x}^*) &= \widehat{\mu}(\mathbf{x}^*) - \mathbf{d}_*^\top\left(\mathbf{D}^\top\mathbf{D}\right)^{-1}\mathbf{d}_m\frac{y^{(m)} - \widehat{\mu}\left(\mathbf{x}^{(m)}\right)}{1-h_{mm}}\\ &= \widehat{\mu}(\mathbf{x}^*) - \mathbf{d}_*^\top\left(\mathbf{D}^\top\mathbf{D}\right)^{-1}\mathbf{d}_m r_m^{\mathrm{LOO}}.\end{aligned} \tag{15}$$

The derivations of the analytical expressions (14) and (15) are provided in Appendix A and Appendix B, respectively. We note that these analytical expressions and their derivation rely on standard linear regression theory [60] and are already known [61]. They are given again using the notation of this work strictly for the reader's convenience.

Remark: As also noted in Section 2.1, sparse PCEs are often employed to overcome computational limitations in problem settings featuring high-dimensional inputs or limited data. These models are typically computed using L_1-minimization and sparsity-promoting algorithms [49, 50]. In this case, explicit model refitting is required during the basis selection procedure, as the closed-form expressions presented above are no longer valid. However, once the final sparse basis has been identified, it is common practice to compute the final PCE coefficients by solving the OLS problem (4) [62]. Therefore, once the final basis is fixed, the analytical, closed-form expressions for the LOO residuals and model evaluations are again applicable. This approach should, however, be used with caution, as the robustness of the identified basis (or lack thereof) is critical to the statistical efficiency of the resulting CP intervals [46].

3 | NUMERICAL EXPERIMENTS

The conformalized PCE method is first applied to benchmark models that are commonly used for the validation of surrogate modeling methods (Section 3.1). Then, the method is applied to an electrical engineering use-case concerning the design of a Stern-Gerlach electromagnet (Section 3.2). Different conformalized PCE configurations are examined, depending on jackknife method, maximum polynomial degree, and, most importantly, training data availability.

The conformalized PCE models are assessed with respect to their ability to construct prediction intervals that provide the required target coverage. To that end, separate test datasets are used to calculate the empirical coverage as the fraction of the test samples for which the true model output successfully falls within the prediction interval. The empirical coverage is then compared against the predefined target coverage to establish the statistical efficiency of the provided CP intervals.

Each experiment is repeated for 100 random seeds to estimate coverage and prediction interval statistics. Note that the use of different random seeds is performed strictly to validate the statistical robustness of the empirical coverage, and is not a part of the methodology itself.

Last, note that the abbreviation "PI" used in the parity plots refers to the prediction intervals provided by the conformalized PCEs.

3.1 | Validation on benchmark models

A univariate model is used in Section 3.1.1, while a model with ten inputs is considered in Section 3.1.2. A discussion based on the numerical results obtained from these benchmark models is available in Section 3.1.3. The discussed trends and observations are further supported by the numerical results corresponding to two further multivariate benchmark models, which are presented in Appendix C. The multivariate models in Section 3.1.2 and Appendix C are taken from the "Virtual Library of Simulation Experiments: Test Functions and Datasets" by Surjanovic and Bingham [63].

For all examined benchmark models, the experimental design used to train the PCE consists of samples drawn randomly from the distribution of the input parameters. The target coverage is fixed at 95% and the test datasets consist of 10^4 random samples.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

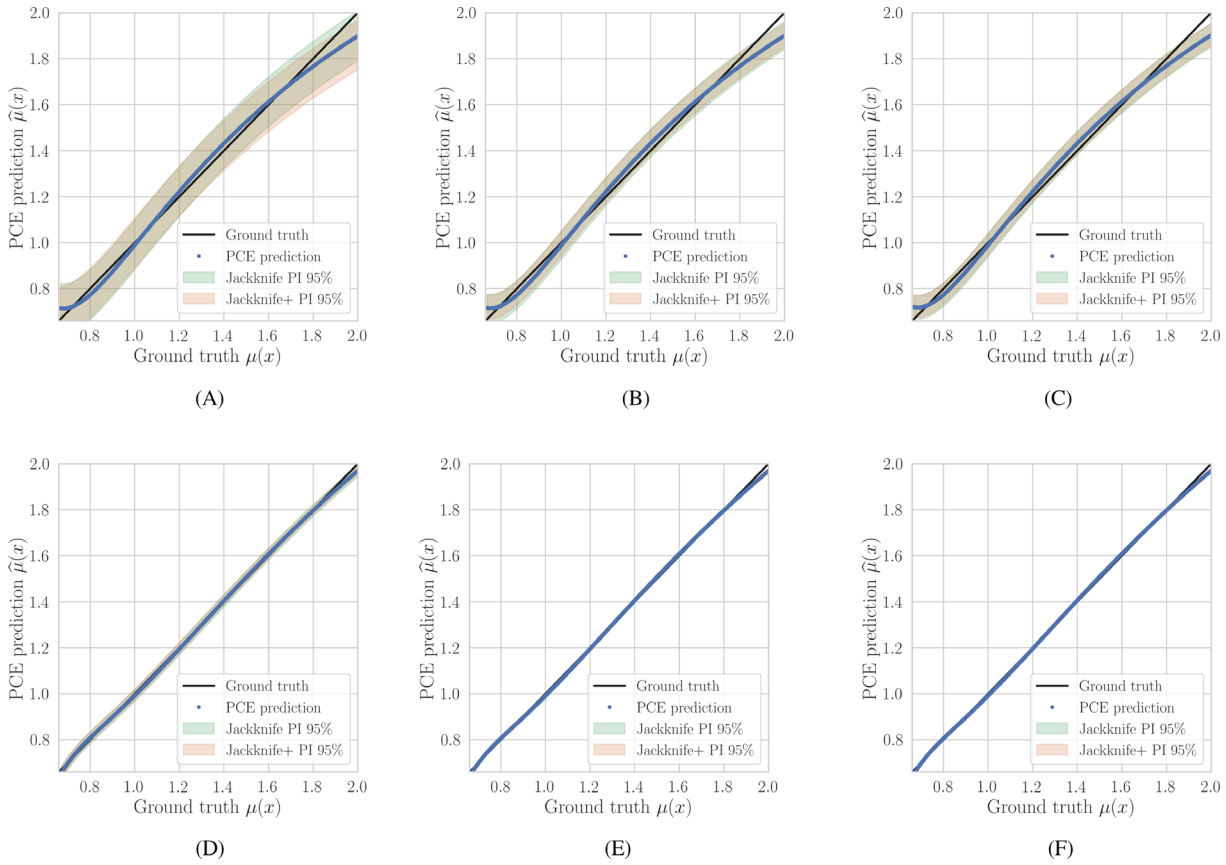

FIGURE 1 Parity plots comparing ground truth values of the meromorphic function against conformalized PCE predictions for different combinations of polynomial degree P and oversampling coefficient C. The results correspond to a single random seed. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 2$, $C = 2$. (B) $P = 2$, $C = 5$. (C) $P = 2$, $C = 10$. (D) $P = 3$, $C = 2$. (E) $P = 3$, $C = 5$. (F) $P = 3$, $C = 10$.

3.1.1 | Meromorphic function

We consider the one-dimensional meromorphic function

$$\mu(x) = \frac{1}{a + bx}, \tag{16}$$

with $a = 1$ and $b = 0.5$. We assume that the input x takes values as uniformly distributed in $[-1, 1]$. The conformalized PCE is constructed with maximum polynomial degrees $P \in \{2, 3\}$. Since the input is one-dimensional, the polynomial basis consists of $P + 1$ terms. The experimental design has size $M = C(P + 1)^2$, where $C \in \{2, 3, 5, 10\}$ is an oversampling coefficient. Note that the quadratic scaling of the experimental design in relation to the polynomial basis is necessary to ensure a well conditioned least squares problem (4) in the one-dimensional input case [53].

Figure 1 compares the different configurations of the conformalized PCE against the true meromorphic function for a single random seed, where the results for $C = 3$ have been omitted for better readability. Figure 2 shows the empirical coverage results, where the box plots summarize the distribution of the empirical coverage over 100 random seeds. Similarly, Figure 3 shows box plots regarding the corresponding prediction intervals.

3.1.2 | Wing weight function

We consider a function that models the weight of light aircraft wing, given as

$$\mu(\mathbf{x}) = 0.036 S_{\mathrm{w}}^{0.758} W_{\mathrm{fw}}^{0.0035} \left(\frac{A}{\cos^2(\Lambda)} \right)^{0.6} q^{0.006} \lambda^{0.04} \left(\frac{100 t_{\mathrm{c}}}{\cos(\Lambda)} \right)^{-0.3} (N_{\mathrm{z}}\ W_{\mathrm{dg}})^{0.49} + S_{\mathrm{w}} W_{\mathrm{p}}, \tag{17}$$

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

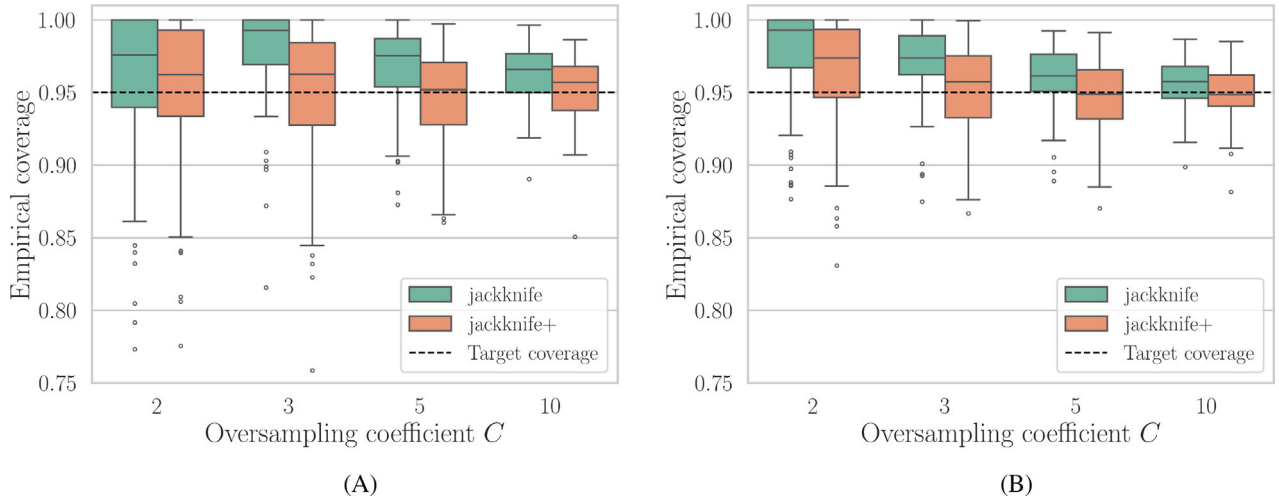

FIGURE 2 Box plots of the empirical coverage provided by conformalized PCE surrogates of the meromorphic function, for different combinations of polynomial degree P and oversampling coefficient C. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 2$. (B) $P = 3$.

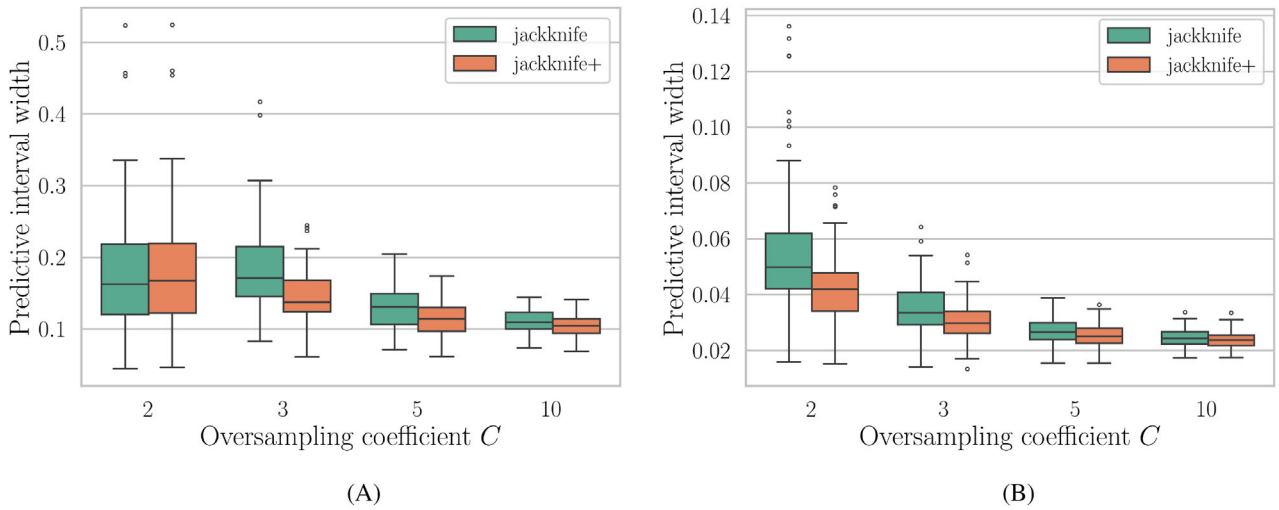

FIGURE 3 Box plots of the prediction interval widths provided by conformalized PCE surrogates of the meromorphic function, for different combinations of polynomial degree P and oversampling coefficient C. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 2$. (B) $P = 3$.

where the input vector $\mathbf{x} \in \mathbb{R}^{10}$ consists of the parameters listed in Table 1, along with their value ranges. All parameters are assumed to be uniformly distributed within their ranges.

Total-degree conformalized PCEs with maximum polynomial degrees $P \in \{1, 2\}$ and oversampling coefficients $C \in \{2, 3, 5, 10\}$ are employed. The experimental design scales linearly with the size of the PCE basis, that is, $M = CK = C\#\Lambda$. Figure 4 compares the different conformalized PCE configurations against the true wing weight function for a single random seed, where the results for $C = 3$ have been omitted for better readability. Figures 5 and 6 present coverage and prediction interval statistics over 100 random seeds.

3.1.3 | Discussion on benchmarks' results

The numerical results presented in Sections 3.1.1 and 3.1.2, but also those available in Appendix C, show a number of patterns and similarities. First, the conformalized PCE is in general capable of providing empirical coverages close to the

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Input parameters of the wing weight function.

Parameter	Description	Units	Range
S_w	wing area	ft^2	[150, 200]
W_{fw}	wing fuel weight	lb	[220, 300]
A	aspect ratio	–	[6, 10]
Λ	quarter-chord sweep	(^{o})	[−10, 10]
q	dynamic pressure at cruise	lb/ft^2	[16, 45]
λ	taper ratio	–	[0.5, 1]
t_c	airfoil thickness to chord ratio	–	[0.08, 0.18]
N_z	ultimate load factor	–	[2.5, 6]
W_{dg}	flight design gross weight	lb	[1700, 2500]
W_p	paint weight	lb/ft^2	[0.025, 0.08]

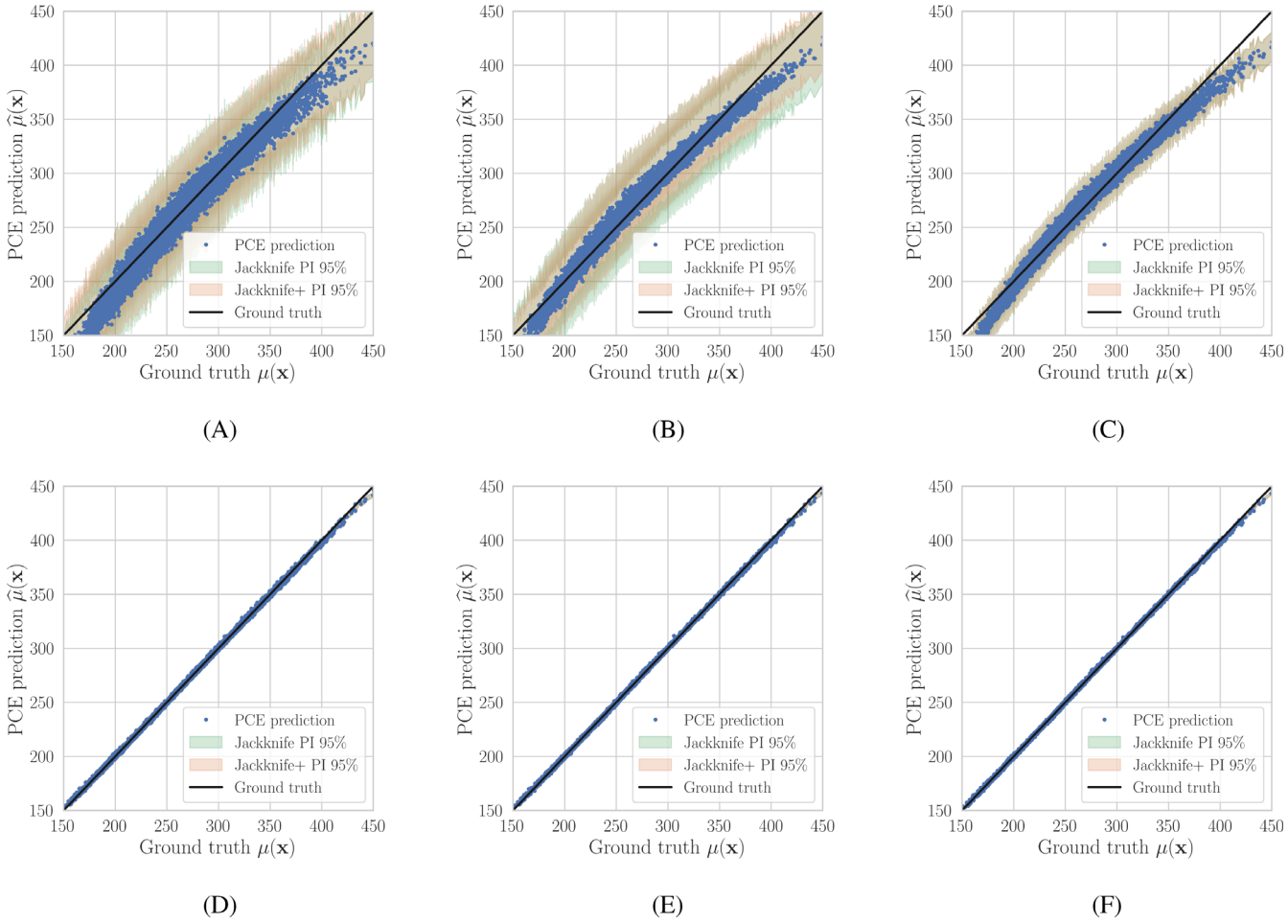

FIGURE 4 Parity plots comparing ground truth values of the wing weight function against conformalized PCE predictions for different combinations of polynomial degree P and oversampling coefficient C. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 1$, $C = 2$. (B) $P = 1$, $C = 5$. (C) $P = 1$, $C = 10$. (D) $P = 2$, $C = 2$. (E) $P = 2$, $C = 5$. (F) $P = 2$, $C = 10$.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

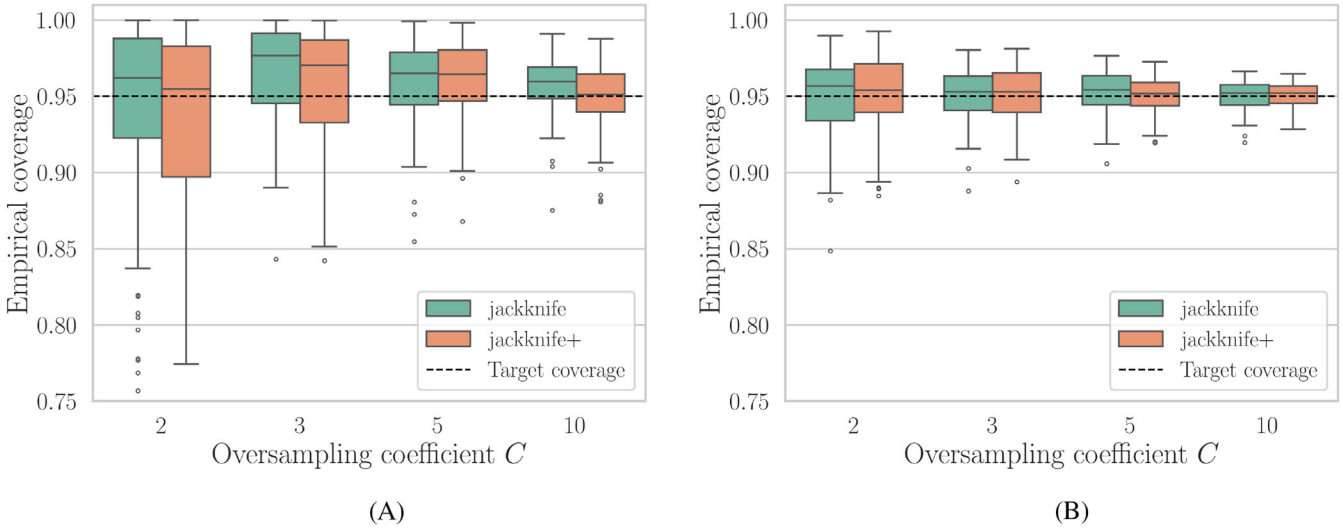

FIGURE 5 Box plots of the empirical coverage provided by conformalized PCE surrogates of the wing weight function, for different combinations of polynomial degree P, and oversampling coefficient C. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 1$. (B) $P = 2$.

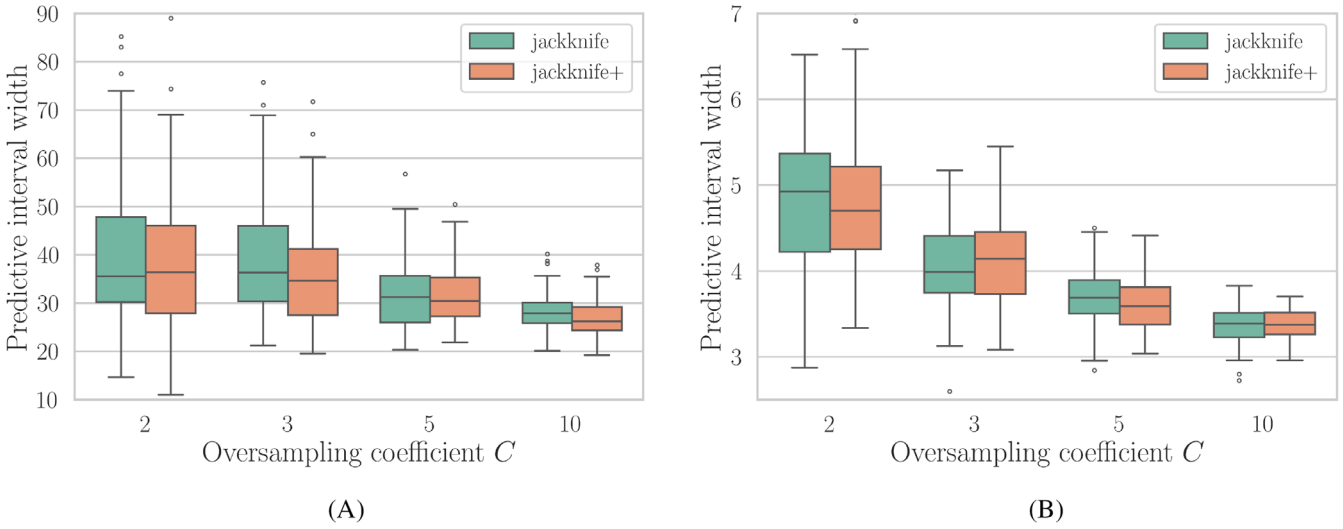

FIGURE 6 Box plots of the prediction interval widths provided by conformalized PCE surrogates of the wing weight function, for different combinations of polynomial degree P, and oversampling coefficient C. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 1$. (B) $P = 2$.

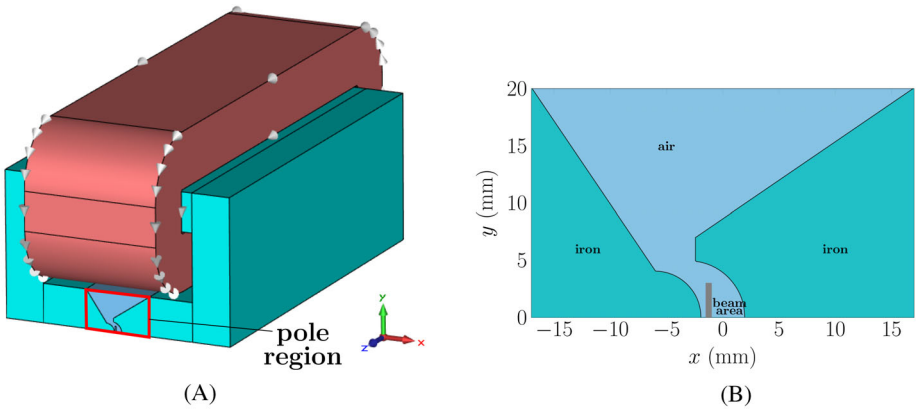

FIGURE 7 Illustration of the Stern-Gerlach magnet model. (A) Stern–Gerlach magnet model (half geometry). (B) Pole region geometry.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Input parameters of the Stern-Gerlach magnet model.

Parameter	Description	Unit	Range
x_1	x-coordinate, control point 1	mm	$[-3.38, -1.38]$
y_1	y-coordinate, control point 1	mm	$[5.96, 7.96]$
x_2	x-coordinate, control point 2	mm	$[-3.38, -1.38]$
y_2	y-coordinate, control point 2	mm	$[3.96, 5.96]$
x_3	x-coordinate, control point 3	mm	$[16, 18]$
y_3	y-coordinate, control point 3	mm	$[19, 21]$
x_4	x-coordinate, control point 4	mm	$[-18, -16]$
y_4	y-coordinate, control point 4	mm	$[19, 21]$
x_5	x-coordinate, control point 5	mm	$[-7, -5]$
y_5	y-coordinate, control point 5	mm	$[3, 2]$
w_1	weight, control point 1	—	$[0, 1]$
w_2	weight, control point 2	—	$[0, 1]$
w_3	weight, control point 3	—	$[0, 1]$
w_4	weight, control point 4	—	$[0, 1]$

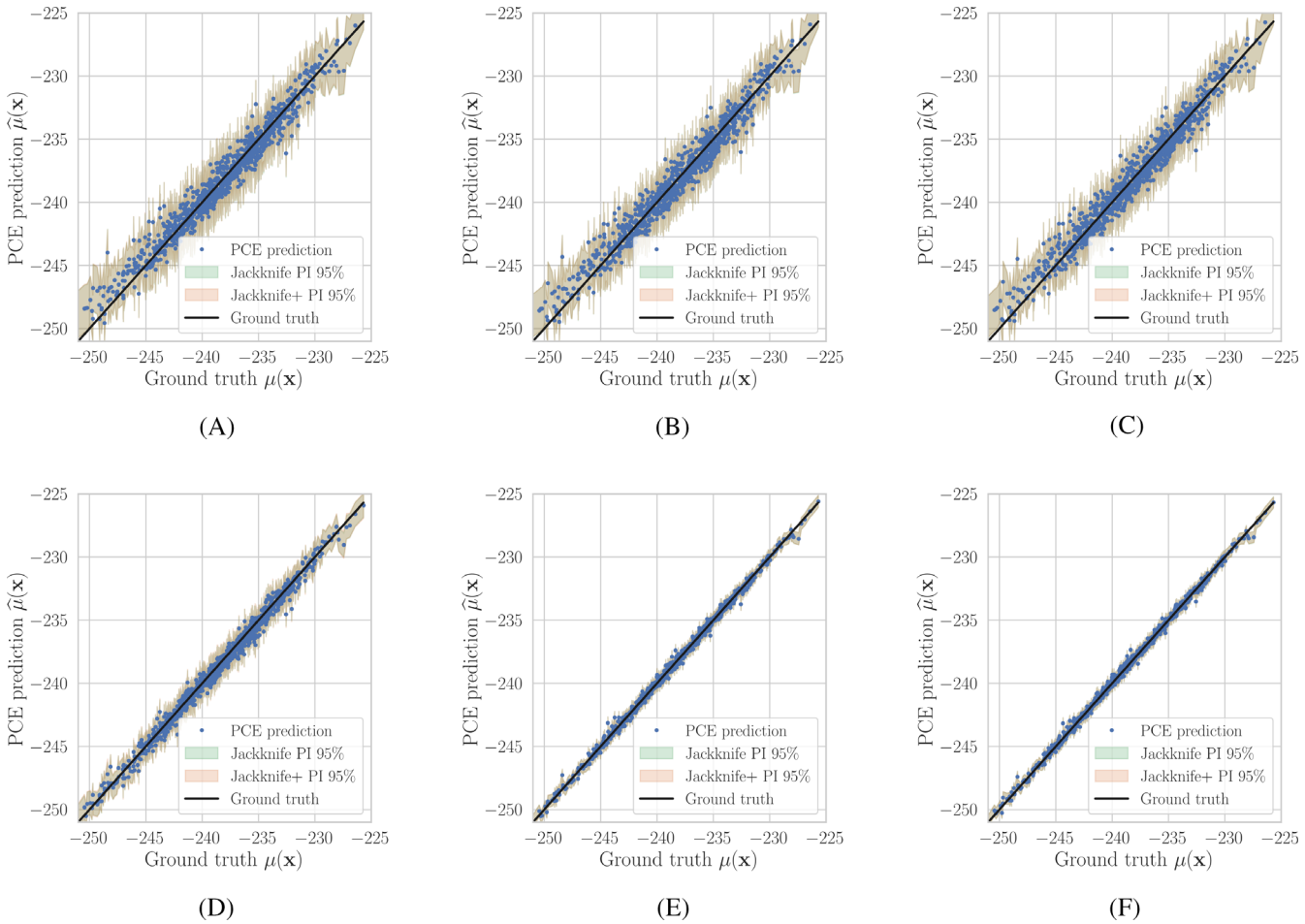

FIGURE 8 Parity plots comparing ground truth $\overline{\tau}_b$ values given by the Stern-Gerlach magnet model against conformalized PCE predictions for different combinations of polynomial degree P, training dataset size M and significance level s. The results correspond to a single partition of the available design dataset. The target coverage is set to 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 1$, $M = 150$. (B) $P = 1$, $M = 300$. (C) $P = 1$, $M = 450$. (D) $P = 2$, $M = 150$. (E) $P = 2$, $M = 300$. (F) $P = 2$, $M = 450$.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

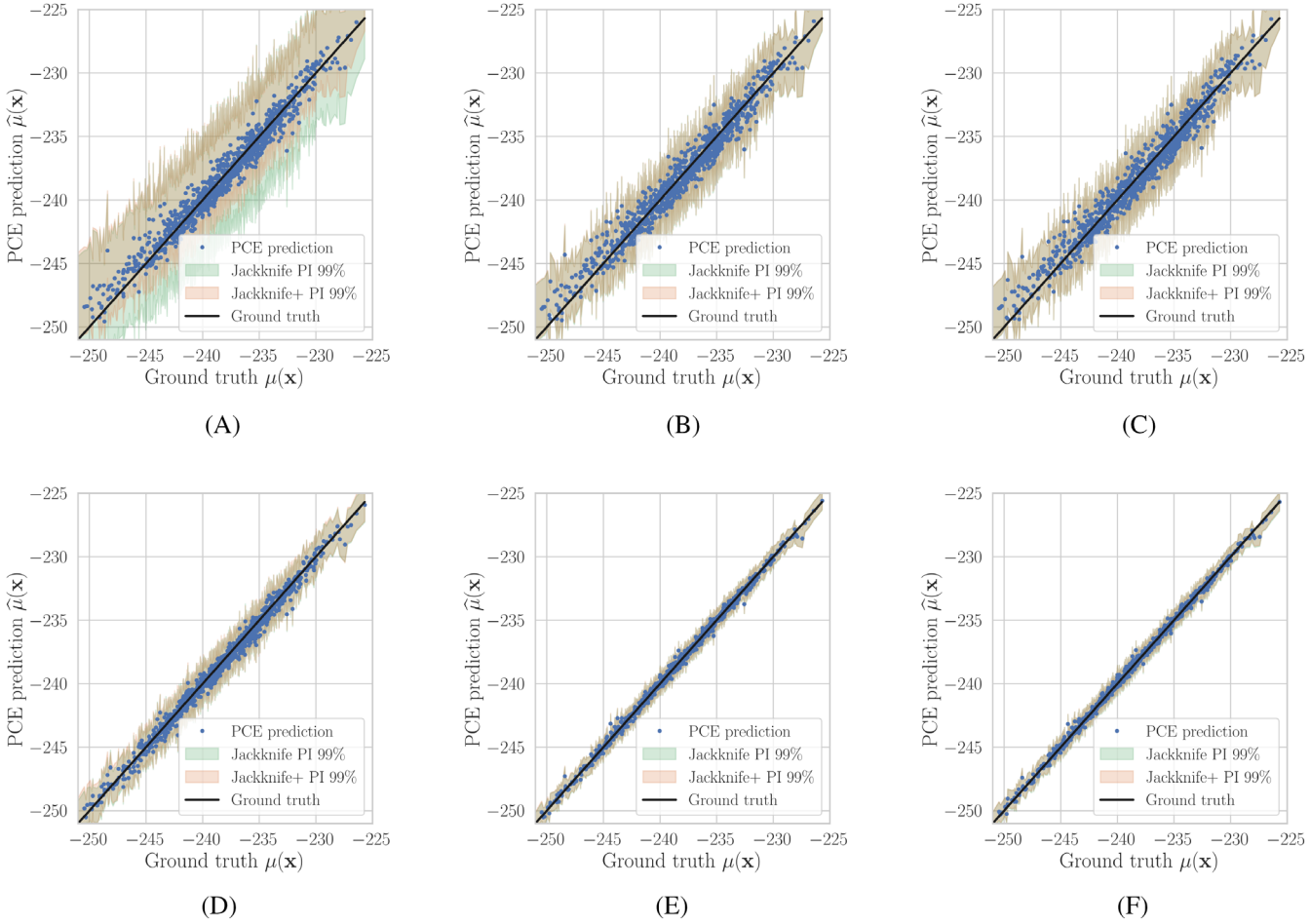

FIGURE 9 Parity plots comparing ground truth $\overline{\tau}_b$ values given by the Stern-Gerlach magnet model against conformalized PCE predictions for different combinations of polynomial degree P, training dataset size M and significance level s. The results correspond to a single partition of the available design dataset. The target coverage is set to 99%, corresponding to a significance level of $s = 0.01$. (A) $P = 1$, $M = 150$. (B) $P = 1$, $M = 300$. (C) $P = 1$, $M = 450$. (D) $P = 2$, $M = 150$. (E) $P = 2$, $M = 300$. (F) $P = 2$, $M = 450$.

target level, more commonly above it. For less accurate PCEs, for example, ones corresponding to low polynomial degrees and/or small experimental designs, the coverage over 100 random seeds varies significantly. As the accuracy of the PCE increases with larger experimental designs and higher polynomial degrees, the variability in coverage is reduced and the coverage moves closer to the target level. The size of the experimental design is particularly important, as a sufficiently large training dataset enables even low-accuracy conformalized PCEs to reach the target coverages.

The coverage provided by the conformalized PCE is naturally connected to the corresponding prediction intervals. For less accurate PCEs, the required coverage is accomplished by larger prediction intervals, which also show a large variability over 100 random seeds. As expected, the prediction intervals become narrower as the accuracy of the PCE increases. Importantly, even for low-accuracy PCEs, increasing the size of the experimental design leads not only to a reduced variability in the prediction interval, but also to a reduced interval width. It is worth noting that interval widths for different polynomial orders can differ significantly and do not necessarily converge to the same median value, even for large training datasets. This behavior is expected and reflects the irreducible structural approximation bias of PCE models, which is directly affected by the choice of polynomial order. That is, the residuals remain bounded below by the structural approximation error of the model, regardless of training data availability. Higher-degree models capture the underlying function with higher fidelity, resulting in smaller residuals and correspondingly narrower, more efficient prediction intervals.

Comparing the jackknife and jackknife+ methods, their prediction intervals are in general very similar, especially for larger experimental designs and more accurate PCEs. The jackknife method is typically more conservative and results in comparatively larger intervals and over-coverage, especially for smaller experimental designs. On the other hand,

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

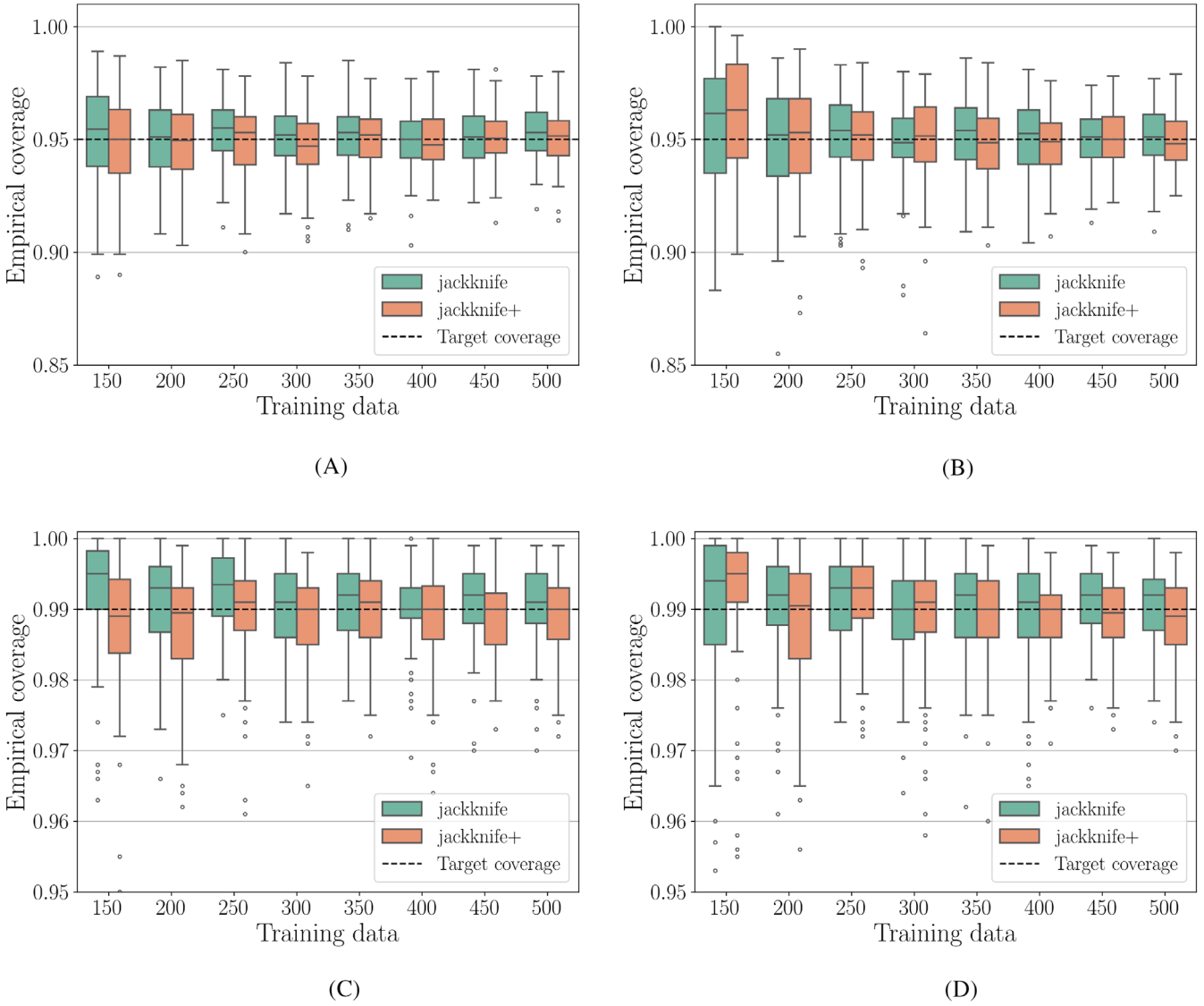

FIGURE 10 Box plots of the empirical coverage provided by conformalized PCE surrogates of the Stern–Gerlach magnet model, for different combinations of polynomial degree P, training dataset size M, and significance level s. (A) $P = 1$, $s = 0.05$. (B) $P = 2$, $s = 0.05$. (C) $P = 1$, $s = 0.01$. (D) $P = 2$, $s = 0.01$.

jackknife+ produces empirical coverages closer to the target level, in most cases. For both jackknife and jackknife+, it is clear that training data availability plays a significant role.

3.2 | Application to electromagnet design

As an engineering design use-case, we consider a so-called Stern-Gerlach electromagnet, typically employed for the magnetic separation of atom beams [47]. A 3D model of the magnet is depicted in Figure 7A, where half of the geometry is shown. This model is generated with the CST Studio Suite® commercial software.[2] A key design requirement of such magnets is a magnetic field with a homogeneous and strong gradient. Field homogeneity is particularly important in the magnet's beam area, which is included in the pole region of the magnet, as shown in Figure 7B.

[2]www.3ds.com/products/simulia/cst-studio-suite.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

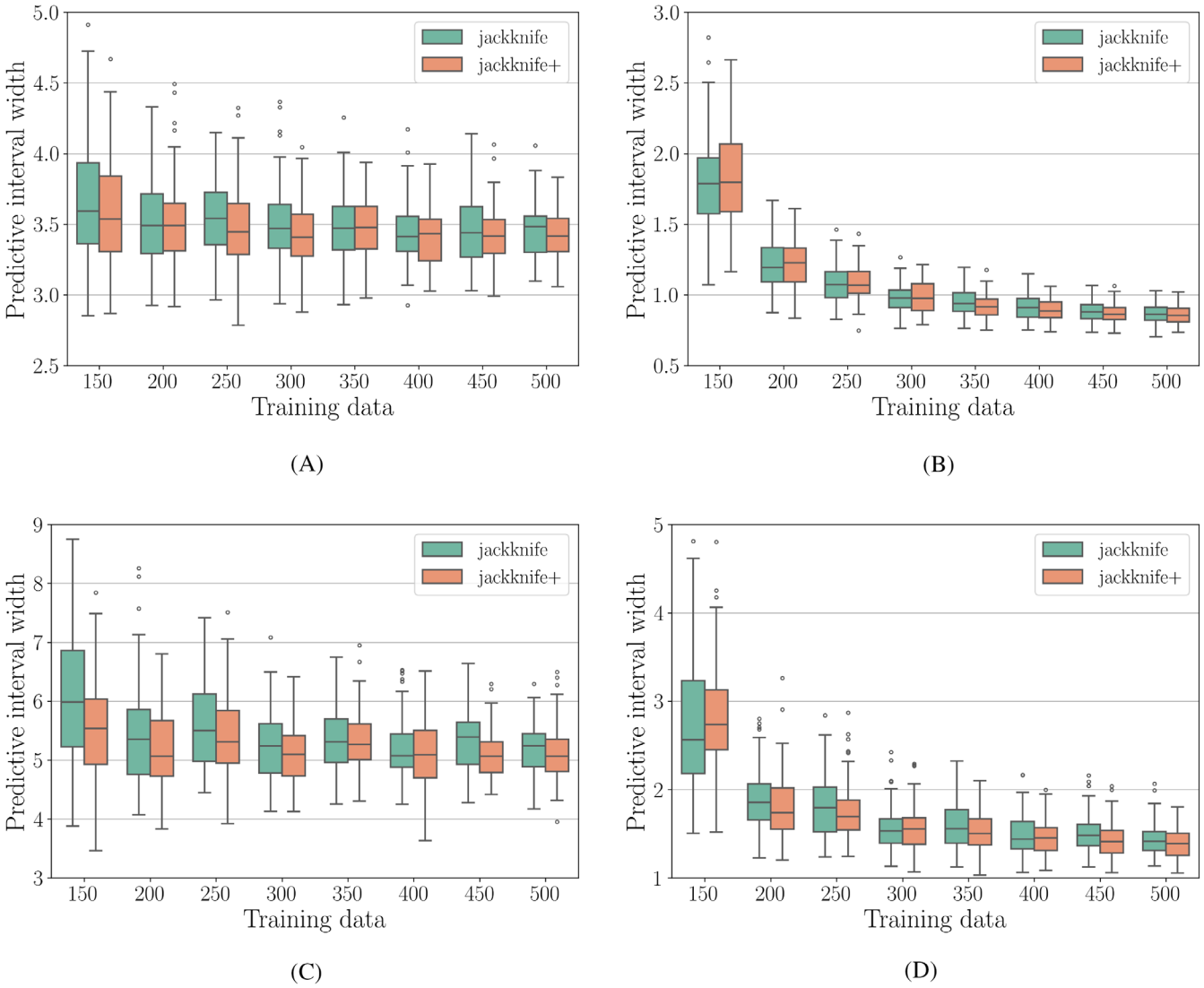

FIGURE 11 Box plots of the prediction interval widths provided by conformalized PCE surrogates of the Stern–Gerlach magnet model, for different combinations of polynomial degree P, training dataset size M, and significance level s. (A) $P = 1$, $s = 0.05$. (B) $P = 2$, $s = 0.05$. (C) $P = 1$, $s = 0.01$. (D) $P = 2$, $s = 0.01$.

The target model output is the average magnetic field gradient in the beam area, denoted as $\overline{\tau}_\mathrm{b}$ and measured in Tm^{-1}. For computational efficiency, $\overline{\tau}_\mathrm{b}$ is computed with a 2D isogeometric analysis (IGA) model of the pole region, while the coils and yoke of the original 3D model are replaced by equivalent circuit models [64]. Then, $\overline{\tau}_\mathrm{b}$ is estimated as

$$\overline{\tau}_\mathrm{b} = \frac{1}{\Omega_\mathrm{b}} \int_{\Omega_\mathrm{b}} \tau(x,y)\mathrm{d}\Omega, \tag{18}$$

where Ω denotes the full 2D computational domain, Ω_b corresponds to the beam area, and $\tau(x,y) = \frac{\partial|\vec{B}|}{\partial x}$ is the magnetic field gradient in the x-direction, where $\vec{B}$ denotes the magnetic flux density. The latter is obtained by the IGA model's field solution.

While much more efficient compared to the original 3D model, the circuit-coupled IGA model is still too computationally expensive for design optimization [47, 64] or UQ [65] studies, hence the need for a surrogate model. Here, the surrogate model is used to approximate $\overline{\tau}_\mathrm{b}$ under random shape deformations in the pole region. The geometry of the pole region is described using non-uniform B-splines (NURBS) $\mathbf{R}(\xi)$, $\xi \in [0,1]$, defined by control points $\mathbf{P}_i$, $i = 1, \ldots, N_\mathrm{cp}$,

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

given as

$$\mathbf{R}(\xi) = \sum_{i=1}^{N_{\text{cp}}} \mathbf{P}_i \frac{w_i B_i^p(\xi)}{\sum_{j=1}^{N_{\text{cp}}} w_j B_j^p(\xi)}, \tag{19}$$

where $B_i^p(\xi)$ are B-spline basis functions of order p and w_i are their corresponding weights. Random shape deformations are introduced by modeling the coordinates of five control points and four of the corresponding weights as uniform random variables. All input parameters and their ranges are listed in Table 2.

Two significance levels $s = 0.05$ and $s = 0.01$ are considered regarding the predictive uncertainty of the surrogate model. The conformalized PCE is applied with maximum polynomial degrees $P \in \{1, 2\}$. A dataset with 1000 different shape deformations of the pole region is available, along with the corresponding $\overline{\tau}_{\text{b}}$ values. Training datasets of increasing size $M \in \{150, 200, \ldots, 500\}$ are employed. The remaining $M' = 500$ data points are used for validation. The dataset is shuffled 100 times to estimate coverage and prediction interval statistics.

Figures 8 and 9 compare ground truth $\overline{\tau}_{\text{b}}$ values provided by the IGA model against conformalized PCE predictions for the original dataset without reshuffling, for the significance levels $s = 0.05$ and $s = 0.01$, respectively. Training dataset sizes $M \in \{150, 300, 450\}$ are considered in these figures. Figures 10 and 11 show boxplots that summarize the statistics concerning empirical coverage and prediction interval width over the 100 dataset reshuffles and for all considered training dataset sizes. The conformalized PCE provides the required coverage, even for the inaccurate $P = 1$ models and the more demanding significance level $s = 0.01$. As would be expected, the prediction intervals become narrower for $P = 2$, while still preserving the required coverage. Similar to the previous test-cases, the empirical coverages and the prediction interval widths present high variations for smaller experimental designs that progressively stabilize as more training data become available. The jackknife method is once again the more conservative option.

4 | SUMMARY, CONCLUSIONS, AND OUTLOOK

This work presented a new method that combines regression-based PCEs with jackknife-based CP for constructing uncertainty-aware, data-driven surrogate models. Using this method, data-driven PCEs are complemented with prediction intervals that quantify the uncertainty in their predictions. The prediction intervals are estimated with very low computational overhead by leveraging the linearity of the PCE regression model. In particular, analytical, closed-form expressions can be used for computing the LOO residuals and LOO predictions employed by the jackknife and jackknife+ CP methods for prediction interval estimation.

The conformalized PCE method was demonstrated on benchmark models and one electrical engineering application concerning the design of a high-gradient-field electromagnet. Different conformalized PCE configurations were examined, based on maximum polynomial degree, experimental design size, and CP method (jackknife/jackknife+). Based on the numerical results, the conformalized PCE is in general able to provide empirical coverages close to the required target level. Training data availability, equivalently, experimental design size, is the most important factor for providing the required coverage with the minimum possible interval width. Moreover, considering PCE models trained with different datasets, larger training datasets result in reduced variations in coverage and prediction interval width. The jackknife method typically results in more conservative predictive uncertainty estimates, with wider intervals and empirical coverages above the target level. The jackknife+ method yields coverages closer to the target level.

The conformalized PCE method developed in this work shows definite promise as an uncertainty-aware, data-driven, surrogate modeling method. Nonetheless, it can be improved in several aspects. First, this work has considered total-degree PCEs computed by means of OLS. Therefore, the selected benchmark models and engineering use-cases were limited to low-to-moderate dimensions and polynomial degrees. The integration of CP within sparse PCE algorithms [49, 50] is a natural extension that would allow addressing more challenging problem settings. In fact, first attempts toward this goal have already emerged [46]. Second, in its current form, the conformalized PCE method is applicable to scalar outputs only. In principle, it can be applied elementwise to address multidimensional outputs. However, this would be a suboptimal approach, as it does not account for correlations among the output's elements and would incur high computational costs with increasing output dimension. The development of a conformalized PCE method suitable for multivariate outputs seems to be a natural follow-up, for example, based on the combination of multi-output PCEs [66–68] and multi-output CP [69, 70]. Third, the current approach provides symmetric prediction intervals that do not

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

adapt to heteroskedastic data. Combinations with locally weighted nonconformity measures [40] or with conformalized quantile regression [37] could offer possible remedies. We aim to address these limitations in dedicated follow-up works.

Finally, a rigorous comparative study assessing and benchmarking the currently available uncertainty-aware, PCE-based surrogate modeling approaches, for example, conformalized PCEs, Bayesian PCEs [21–23], and combined PCE-GP models [26–29], could provide valuable insights.

AUTHOR CONTRIBUTIONS

Dimitrios Loukrezis: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Writing – review and editing. **Dimitris G. Giovanis:** Validation, Data curation, Writing – original draft, Writing – review and editing.

FUNDING INFORMATION

Dimitris G. Giovanis was supported by the Department of Energy (DOE) under Grant No. DE-SC0024162.

CONFLICT OF INTEREST STATEMENT

The authors have nothing to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Dimitrios Loukrezis https://orcid.org/0000-0003-1264-1182
Dimitris G. Giovanis https://orcid.org/0000-0003-2272-2584

REFERENCES

[1] R. G. Ghanem and P. D. Spanos, *Stochastic finite elements: A spectral approach*, Courier Corporation, Mineola, New York, USA, 2003.
[2] P. D. Spanos and R. G. Ghanem, Stochastic finite element expansion for random media, J. Eng. Mech. **115** (1989), no. 5, 1035–1053.
[3] D. Xiu and G. E. Karniadakis, The wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. **24** (2002), no. 2, 619–644.
[4] E. Torre, S. Marelli, P. Embrechts, and B. Sudret, Data-driven polynomial chaos expansion for machine learning regression, J. Comput. Phys. **388** (2019), 601–623.
[5] J. O. Campos, R. M. Guedes, Y. B. Werneck, L. P. S. Barra, R. W. Santos, and B. M. Rocha, Polynomial chaos expansion surrogate modeling of passive cardiac mechanics using the Holzapfel–Ogden constitutive model, J. Comput. Sci. **71** (2023), 102039.
[6] W. Dai, D. Li, H. Liu, and Y. Liu, A cost surrogate model for TSO-DSO coordination based on polynomial chaos expansion, IEEE Trans. Power Syst. **40** (2025), 3580–3583.
[7] R. Haghi and C. Crawford, Surrogate models for the blade element momentum aerodynamic model using non-intrusive polynomial chaos expansions, Wind Energy Sci. **7** (2022), no. 3, 1289–1304.
[8] J. Lee, L. Chan, T. Zahtila, W. Lu, G. Iaccarino, and A. Ooi, Surrogate models for multiregime flow problems, Phys. Rev. Fluids **10** (2025), no. 2, 024703.
[9] D. Loukrezis and H. De Gersem, Power module heat sink design optimization with ensembles of data-driven polynomial chaos surrogate models, E-Prime-Adv. Electr. Eng. Electron. Energy **2** (2022), 100059.
[10] P. Manfredi and R. Trinchero, Nonparametric formulation of polynomial chaos expansion based on least-square support-vector machines, Eng. Appl. Artif. Intell. **133** (2024), 108182.
[11] D. G. Giovanis, D. Loukrezis, I. G. Kevrekidis, and M. D. Shields, Polynomial chaos expansions on principal geodesic Grassmannian sub manifolds for surrogate modeling and uncertainty quantification, J. Comput. Phys. **519** (2024), 113443.
[12] K. Kontolati, D. Loukrezis, D. G. Giovanis, L. Vandanapu, and M. D. Shields, A survey of unsupervised learning methods for high-dimensional uncertainty quantification in black-box-type problems, J. Comput. Phys. **464** (2022), 111313.
[13] K. Kontolati, D. Loukrezis, K. R. M. Santos, D. G. Giovanis, and M. D. Shields, Manifold learning-based polynomial chaos expansions for high-dimensional surrogate models, Int. J. Uncertain. Quantif. **12** (2022), no. 4, 39–64.
[14] C. Soize and R. G. Ghanem, Polynomial chaos representation of databases on manifolds, J. Comput. Phys. **335** (2017), 201–221.
[15] H. Sharma, L. Novák, and M. D. Shields. Polynomial chaos expansion for operator learning. *arXiv preprint arXiv:2508.20886* 2025.
[16] L. Novák, H. Sharma, and M. D. Shields, Physics-informed polynomial chaos expansions, J. Comput. Phys. **506** (2024), 112926.
[17] H. Sharma, L. Novák, and M. Shields, Physics-constrained polynomial chaos expansion for scientific machine learning and uncertainty quantification, Comput. Methods Appl. Mech. Eng. **431** (2024), 117314.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[18] S. Ranftl, W. von der Linden, and MaxEnt 2021 Scientific Committee, Bayesian sur rogate analysis and uncertainty propagation, Phys. Sci. Forum **3** (2021), no. 1, 6, MDPI.
[19] C. K. I. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, The MIT Press, Cambridge, MA, USA, 2006 ISBN: 026218253X.
[20] L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun, Hands-on Bayesian neural networks—A tutorial for deep learning users, IEEE Comput. Intell. Mag. **17** (2022), no. 2, 29–48.
[21] Y. Li, Y. Luo, Z. Zhong, J. Yu, and R. Wang, A new Bayesian sparse polynomial chaos expansion based on Gaussian scale mixture prior model, Mech. Syst. Signal Process. **229** (2025), 112511.
[22] K. N. Rumsey, D. Francom, G. Gibson, J. Derek Tucker, and G. Huerta, Bayesian adaptive polynomial chaos expansions, Underst. Stat. **15** (2026), no. 1, e70151.
[23] Q. Shao, A. Younes, M. Fahs, and T. A. Mara, Bayesian sparse polynomial chaos expansion for global sensitivity analysis, Comput. Methods Appl. Mech. Eng. **318** (2017), 474–496.
[24] S. Marelli and B. Sudret, An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis, Struct. Saf. **75** (2018), 67–74.
[25] X. Shang et al., Active learning of ensemble polynomial chaos expansion method for global sensitivity analysis, Reliab. Eng. Syst. Saf. **249** (2024), 110226.
[26] P. Kersaudy, B. Sudret, N. Varsier, O. Picon, and J. Wiart, A new surrogate modeling technique combining kriging and polynomial chaos expansions–application to uncertainty analysis in computational dosimetry, J. Comput. Phys. **286** (2015), 103–117.
[27] R. Schobi, B. Sudret, and J. Wiart, Polynomial-chaos-based kriging, Int. J. Uncertain. Quantif. **5** (2015), no. 2, 171–193.
[28] P. Manfredi, A hybrid polynomial chaos expansion–Gaussian process regression method for Bayesian uncertainty quantification and sensitivity analysis, Comput. Methods Appl. Mech. Eng. **436** (2025), 117693.
[29] P. Manfredi, Polynomial chaos vs kernel machine learning methods for uncertainty quantification: A comparative study and benchmarking with the hybrid PCE-GPR approach, Comput. Methods Appl. Mech. Eng. **449** (2026), 118523.
[30] X. Zhu and B. Sudret, Stochastic polynomial chaos expansions to emulate stochastic simulators, Int. J. Uncertain. Quantif. **13** (2023), no. 2, 31–52.
[31] M. Fontana, G. Zeni, and S. Vantini, Conformal prediction: A unified review of theory and new challenges, Bernoulli **29** (2023), no. 1, 1–23.
[32] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic learning in a random world*, Springer, New York, NY, USA, 2005.
[33] E. Burnaev and I. Nazarov, "*Conformalized kernel ridge regression*," *2016 15th IEEE international conference on machine learning and applications (ICMLA)*, IEEE, Anaheim, California, USA, 2016, pp. 45–52.
[34] K. Huang, Y. Jin, E. Candes, and J. Leskovec, Uncertainty quantification over graph with conformalized graph neural net works, Adv. Neural Inf. Process. Syst. **36** (2023), 26699–26721.
[35] E. Jaber, V. Blot, N. J. B. Brunel, V. Chabridon, E. Remy, B. Iooss, D. Lucor, M. Mougeot, and A. Leite, Conformal approach to gaussian process surrogate evaluation with marginal coverage guarantees, J. Mach. Learn. Model. Comput. **6** (2025), no. 3, 37–68.
[36] L. Podina, M. T. Rad, and M. Kohandel. Conformalizedphysics-informed neural networks. *arXiv preprint arXiv:2405.08111* 2024.
[37] Y. Romano, E. Patterson, and E. Candes, Conformalized quantile regression, Adv. Neural Inf. Process. Syst. **32** (2019), 3543–3553.
[38] N. Deutschmann, M. Rigotti, and M. Rodriguez Martinez, Adaptive conformal regression with split-jackknife+ scores, Trans. Mach. Learn. Res. **2024** (2024).
[39] M. Zaffran et al., "*Adaptive conformal predictions for time series*," *International conference on machine learning*, PMLR, 2022, pp. 25834–25866.
[40] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, Distribution-free predictive inference for regression, J. Am. Stat. Assoc. **113** (2018), no. 523, 1094–1111.
[41] H. Papadopoulos et al., "*Inductive confidence machines for regression*," *European conference on machine learning*, Springer, Berlin, Germany, 2002, pp. 345–356.
[42] V. Vovk, Cross-conformal predictors, Ann. Math. Artif. Intell. **74** (2015), no. 1, 9–28.
[43] V. Vovk et al., "*Cross-conformal predictive distributions*," *Conformal and probabilistic prediction and applications*, PMLR, 2018, pp. 37–51.
[44] R. F. Barber, E. J. Candès, A. Ramdas, and R. J. Tibshirani, Predictive inference with the jackknife+, Ann. Stat. **49** (2021), no. 1, 486–507.
[45] V. Gopakumar, A. Gray, J. Oskarsson, L. Zanisi, D. Giles, M. J. Kusner, S. Pamela, and M. Peter Deisenroth, Uncertainty quantification of surrogate models using conformal prediction, Mach. Learn.: Sci. Technol. **7** (2024), 015025.
[46] A. Hatstatt, X. Zhu, and B. Sudret. Conformal prediction for full and sparse polynomial chaos expansions. *arXiv Preprint arXiv:2601.16636* 2026.
[47] B. Masschaele, T. Roggen, H. de Gersem, E. Jannsens, and T. T. Nguyen, Design of a strong gradient magnet for the deflection of nanoclusters, IEEE Trans. Appl. Supercond. **22** (2011), no. 3, 3700604.
[48] M. M. Rao and R. J. Swift, *Probability theory with applications*, Springer, New York, NY, USA, 2006.
[49] N. Lüthen, S. Marelli, and B. Sudret, Sparse polynomial chaos expansions: Literature survey and benchmark, SIAM/ASA J. Uncertain. Quantif. **9** (2021), no. 2, 593–649.
[50] N. Lüthen, S. Marelli, and B. Sudret, Automatic selection of basis-adaptive sparse polynomial chaos expansions for engineering applications, Int. J. Uncertain. Quantif. **12** (2022), no. 3, 49–74.
[51] M. Hadigol and A. Doostan, Least squares polynomial chaos expansion: A review of sampling strategies, Comput. Methods Appl. Mech. Eng. **332** (2018), 382–407.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[52] G. Migliorati, F. Nobile, E. Schwerin, and R. Tempone, Approximation of quantities of interest in stochastic PDEs by the random discrete L2 projection on polynomial spaces, SIAM J. Sci. Comput. **35** (2013), no. 3, A1440–A1460.
[53] G. Migliorati, F. Nobile, E. Von Schwerin, and R. Tempone, Analysis of discrete L2 projection on polynomial spaces with random evaluations, Found. Comput. Math. **14** (2014), no. 3, 419–456.
[54] P. R. Conrad and Y. M. Marzouk, Adaptive Smolyak pseudospectral approximations, SIAM J. Sci. Comput. **35** (2013), no. 6, A2643–A2670.
[55] P. G. Constantine, M. S. Eldred, and E. T. Phipps, Sparse pseudospectral approximation method, Comput. Methods Appl. Mech. Eng. **229** (2012), 1–12.
[56] G. T. Buzzard, Global sensitivity analysis using sparse grid interpolation and polynomial chaos, Reliab. Eng. Syst. Saf. **107** (2012), 82–89.
[57] A. Galetzka, D. Loukrezis, N. Georg, H. de Gersem, and U. Römer, An hp-adaptive multi-element stochastic collocation method for surrogate modeling with information re-use, Int. J. Numer. Methods Eng. **124** (2023), no. 12, 2902–2930.
[58] G. Porta, L. Tamellini, V. Lever, and M. Riva, Inverse modeling of geochemical and mechanical compaction in sedimentary basins through polynomial chaos expansion, Water Resour. Res. **50** (2014), no. 12, 9414–9431.
[59] M. R. Chernick, The jackknife: A resampling method with connections to the bootstrap, Wiley Interdiscip. Rev. Comput. Stat. **4** (2012), no. 2, 224–226.
[60] G. A. F. Seber and A. J. Lee, *Linear regression analysis*, John Wiley Sons, Hoboken, New Jersey, USA, 2003.
[61] G. Blatman, "*Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis*," *PhD thesis*, Universite Blaise Pascal-Clermont-Ferrand II, Aubière, France, 2009.
[62] G. Blatman and B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, J. Comput. Phys. **230** (2011), no. 6, 2345–2367.
[63] S. Surjanovic and D. Bingham, *Virtual library of simulation experiments: Test functions and datasets*, Simon Fraser University, Burnaby, British Columbia, Canada, http://www.sfu.ca/ ssurjano, 2013.
[64] A. Pels, Z. Bontinck, J. Corno, H. de Gersem, and S. Schops, Optimization of a Stern–Gerlach magnet by magnetic field–circuit coupling and isogeometric analysis, IEEE Trans. Magn. **51** (2015), no. 12, 1–7.
[65] D. Loukrezis, U. Römer, and H. De Gersem, Assessing the performance of Leja and Clenshaw-Curtis collocation for computational electromagnetics with random input data, Int. J. Uncertain. Quantif. **9** (2019), no. 1, 33–57.
[66] B. Bhattacharyya, E. Jacquelin, and D. Brizard, Uncertainty quantification of stochastic impact dynamic oscillator using a proper orthogonal decomposition-polynomial chaos expansion technique, J. Vib. Acoust. **142** (2020), no. 6, 061013.
[67] L. Hawchar, C.-P. El Soueidy, and F. Schoefs, Principal component analysis and polynomial chaos expansion for time-variant reliability problems, Reliab. Eng. Syst. Saf. **167** (2017), 406–416.
[68] D. Loukrezis, E. Diehl, and H. De Gersem, Multivariate sensitivity-adaptive polynomial chaos expansion for high-dimensional surrogate modeling and uncertainty quantification, Appl. Math. Model. **137** (2025), 115746.
[69] X. Chen and Y. Xie, Conformal prediction for time series, IEEE Trans. Pattern Anal. Mach. Intell. **45** (2023), no. 10, 11575–11587.
[70] R. Zhang, P. Zhou, and T. Chai, Improved copula-based conformal prediction for uncertainty quantification of multi-output regression, J. Process Control **129** (2023), 103036.

How to cite this article: D. Loukrezis and D. G. Giovanis, *Conformalized polynomial chaos expansion for uncertainty-aware surrogate modeling*, GAMM-Mitteilungen. **49** (2026), e70017. https://doi.org/10.1002/gamm.70017

APPENDIX A. LEAVE-ONE-OUT RESIDUALS IN CLOSED FORM

Let $\widehat{\mu}(\mathbf{x})$ denote a PCE regression model as in Section 2.1. Removing the m-th sample from the training data, we get the LOO design matrix $\mathbf{D}_{\sim m} \in \mathbb{R}^{(M-1)\times K}$ and the LOO output vector $\mathbf{y}_{\sim m} \in \mathbb{R}^{M-1}$. The omitted m-th row of the original design matrix $\mathbf{D}$ is denoted with $\mathbf{d}_m^\top \in \mathbb{R}^{1\times K}$, where $\mathbf{d}_m^\top = \left(\Psi_1\left(\mathbf{x}^{(m)}\right), \ldots, \Psi_K\left(\mathbf{x}^{(m)}\right)\right)$. Then, the LOO equivalent to the least-squares solution (5) is

$$\mathbf{c}_{\sim m} = \left(\mathbf{D}_{\sim m}^\top \mathbf{D}_{\sim m}\right)^{-1} \mathbf{D}_{\sim m}^\top \mathbf{y}_{\sim m}, \tag{A1}$$

where it also holds that

$$\mathbf{D}_{\sim m}^\top \mathbf{D}_{\sim m} = \mathbf{D}^\top \mathbf{D} - \mathbf{d}_m \mathbf{d}_m^\top, \tag{A2}$$

$$\mathbf{D}_{\sim m}^\top \mathbf{y}_{\sim m} = \mathbf{D}^\top \mathbf{y} - \mathbf{d}_m y^{(m)}. \tag{A3}$$

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Let $\mathbf{A} = \left(\mathbf{D}^\top \mathbf{D}\right)^{-1} \in \mathbb{R}^{K\times K}$. Using the Sherman-Morrison formula we get

$$\left(\mathbf{D}^\top \mathbf{D} - \mathbf{d}_m \mathbf{d}_m^\top\right)^{-1} = \mathbf{A} + \frac{\mathbf{A}\mathbf{d}_m \mathbf{d}_m^\top \mathbf{A}}{1 - \mathbf{d}_m^\top \mathbf{A}\mathbf{d}_m}, \tag{A4}$$

assuming that $1 - \mathbf{d}_m^\top \mathbf{A}\mathbf{d}_m \neq 0$. Note that $\mathbf{d}_m^\top \mathbf{A}\mathbf{d}_m = h_{mm}$, where h_{mm} are the diagonal elements of the hat matrix $\mathbf{H} \in \mathbb{R}^{M\times M}$ defined in (13). Using expressions (A2), (A3), and (A4), expression (A1) can now be written as

$$\begin{aligned}
\mathbf{c}_{\sim m} &= \left(\mathbf{A} + \frac{\mathbf{A}\mathbf{d}_m \mathbf{d}_m^\top \mathbf{A}}{1 - h_{mm}}\right)\left(\mathbf{D}^\top \mathbf{y} - \mathbf{d}_m y^{(m)}\right) \\
&= \mathbf{A}\mathbf{D}^\top \mathbf{y} - \mathbf{A}\mathbf{d}_m y^{(m)} + \frac{\mathbf{A}\mathbf{d}_m \mathbf{d}_m^\top \mathbf{A}\mathbf{D}^\top \mathbf{y}}{1 - h_{mm}} - \frac{\mathbf{A}\mathbf{d}_m \mathbf{d}_m^\top \mathbf{A}\mathbf{d}_m y^{(m)}}{1 - h_{mm}} \\
&= \mathbf{c} - \mathbf{A}\mathbf{d}_m y^{(m)} + \frac{\mathbf{A}\mathbf{d}_m \mathbf{d}_m^\top \mathbf{c}}{1 - h_{mm}} - \frac{\mathbf{A}\mathbf{d}_m h_{mm} y^{(m)}}{1 - h_{mm}} \\
&= \mathbf{c} + \mathbf{A}\mathbf{d}_m \left(\frac{\mathbf{d}_m^\top \mathbf{c} - h_{mm} y^{(m)}}{1 - h_{mm}} - y^{(m)}\right) \\
&= \mathbf{c} + \mathbf{A}\mathbf{d}_m \frac{\mathbf{d}_m^\top \mathbf{c} - y^{(m)}}{1 - h_{mm}} \\
&= \mathbf{c} + \mathbf{A}\mathbf{d}_m \frac{\widehat{y}^{(m)} - y^{(m)}}{1 - h_{mm}} \\
&= \mathbf{c} - \mathbf{A}\mathbf{d}_m \frac{y^{(m)} - \widehat{y}^{(m)}}{1 - h_{mm}},
\end{aligned} \tag{A5}$$

where we have used that $\mathbf{c} = \mathbf{A}\mathbf{D}^\top \mathbf{y}$ and $\widehat{y}^{(m)} = \widehat{\mu}\left(\mathbf{x}^{(m)}\right) = \mathbf{d}_m^\top \mathbf{c}$. The LOO model evaluation at $\mathbf{x}^{(m)}$ is then given by

$$\widehat{\mu}_{\sim m}\left(\mathbf{x}^{(m)}\right) = \mathbf{d}_m^\top \mathbf{c}_{\sim m} = \mathbf{d}_m^\top \mathbf{c} - \mathbf{d}_m^\top \mathbf{A}\mathbf{d}_m \frac{y^{(m)} - \widehat{y}^{(m)}}{1 - h_{mm}} = \widehat{y}^{(m)} - h_{mm} \frac{y^{(m)} - \widehat{y}^{(m)}}{1 - h_{mm}}. \tag{A6}$$

Therefore, the LOO residual defined in (10) is given by

$$\begin{aligned}
r_m^{\mathrm{LOO}} &= y^{(m)} - \widehat{\mu}_{\sim m}\left(\mathbf{x}^{(m)}\right) \\
&= y^{(m)} - \widehat{y}^{(m)} + h_{mm} \frac{\widehat{y}^{(m)} - y^{(m)}}{1 - h_{mm}} \\
&= \frac{y^{(m)} - \widehat{y}^{(m)}}{1 - h_{mm}} \\
&= \frac{y^{(m)} - \widehat{\mu}\left(\mathbf{x}^{(m)}\right)}{1 - h_{mm}}.
\end{aligned} \tag{A7}$$

APPENDIX B. LEAVE-ONE-OUT PREDICTIONS IN CLOSED FORM

Using similar notation as in Appendix A, the LOO prediction at a test input $\mathbf{x}^*$ is given by

$$\widehat{\mu}_{\sim m}(\mathbf{x}^*) = \mathbf{d}_*^\top \mathbf{c}_{\sim m}, \tag{B1}$$

where $\mathbf{d}_*^\top = (\Psi_1(\mathbf{x}^*), \ldots, \Psi_K(\mathbf{x}^*))$. Substituting the LOO PCE coefficients from (A1) into (B1), we get

$$\begin{aligned}
\widehat{\mu}_{\sim m}(\mathbf{x}^*) &= \mathbf{d}_*^\top \left(\mathbf{c} - \mathbf{A}\mathbf{d}_m \frac{y^{(m)} - \widehat{y}^{(m)}}{1 - h_{mm}}\right) \\
&= \mathbf{d}_*^\top \mathbf{c} - \mathbf{d}_*^\top \mathbf{A}\mathbf{d}_m \frac{y^{(m)} - \widehat{y}^{(m)}}{1 - h_{mm}}
\end{aligned}$$

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$
\begin{aligned}
&= \widehat{\mu}(\mathbf{x}^*) - \mathbf{d}_*^\top \mathbf{A}\mathbf{d}_m \frac{y^{(m)} - \widehat{y}^{(m)}}{1 - h_{mm}} \\
&= \widehat{\mu}(\mathbf{x}^*) - \mathbf{d}_*^\top \mathbf{A}\mathbf{d}_m \frac{y^{(m)} - \widehat{\mu}\left(\mathbf{x}^{(m)}\right)}{1 - h_{mm}} \\
&= \widehat{\mu}(\mathbf{x}^*) - \mathbf{d}_*^\top \left(\mathbf{D}^\top \mathbf{D}\right)^{-1} \mathbf{d}_m \frac{y^{(m)} - \widehat{\mu}\left(\mathbf{x}^{(m)}\right)}{1 - h_{mm}} \\
&= \widehat{\mu}(\mathbf{x}^*) - \mathbf{d}_*^\top \left(\mathbf{D}^\top \mathbf{D}\right)^{-1} \mathbf{d}_m r_m^{\mathrm{LOO}},
\end{aligned} \tag{B2}
$$

where we used that $\mathbf{A} = \left(\mathbf{D}^\top \mathbf{D}\right)^{-1}$, $\widehat{\mu}(\mathbf{x}^*) = \mathbf{d}_*^\top \mathbf{c}$, and the analytical expression (A7) for the LOO residual.

APPENDIX C. VALIDATION ON ADDITIONAL BENCHMARK MODELS

C1 OTL circuit function

We consider the response of an output transformerless (OTL) push-pull circuit, given as

$$
\mu(\mathbf{x}) = \frac{\left(\frac{12\ R_{b2}}{R_{b1}+R_{b2}} + 0.74\right)\ \beta\ (R_{c2}+9)}{\beta\ (R_{c2}+9) + R_f} + \frac{11.35\ R_f}{\beta\ (R_{c2}+9) + R_f} + \frac{0.74\ R_f\ \beta\ (R_{c2}+9)}{\left(\beta\ (R_{c2}+9) + R_f\right)\ R_{c1}}, \tag{C1}
$$

where the input vector $\mathbf{x} \in \mathbb{R}^6$ consists of the resistances R_{b1}, R_{b2}, R_f, R_{c1}, R_{c2}, and the current gain β. The response corresponds to the midpoint voltage. The input parameters are assumed to be uniformly distributed within the value ranges given in Table C1.

Total-degree conformalized PCEs with maximum polynomial degrees $P \in \{1, 2, 3\}$ and oversampling coefficients $C \in \{2, 3, 5, 10\}$ are employed. Following Migliorati et al. [53], the experimental design now scales linearly with the size of the PCE basis, such that $M = CK = C\#\Lambda$. The different conformalized PCE configurations are compared against the true OTL circuit function for a single random seed in Figure C1, where the results for $C = 3$ have been omitted for better readability. Figures C2 and C3 show box plots regarding the distributions of the conformalized PCEs' empirical coverages and prediction intervals over 100 random seeds.

C2 Piston simulation function

We consider the circular motion of a piston within a cylinder, the cycle-time of which is given by the function

$$
\begin{aligned}
&\mu(\mathbf{x}) = 2\pi \sqrt{\frac{M}{k + S^2 \frac{P_0 V_0}{T_0} \frac{T_0}{V^2}}}, \\
&\text{where } V = \frac{S}{2k}\left(\sqrt{A^2 + 4k\frac{P_0 V_0}{T_0} T_a} - A\right) \text{ and } A = P_0\ S + 19.62\ M - \frac{k\ V_0}{S}.
\end{aligned} \tag{C2}
$$

The input vector $\mathbf{x} \in \mathbb{R}^7$ consists of the parameters listed in Table C2 along with their value ranges. All parameters are assumed to be uniformly distributed within their ranges.

TABLE C1 Input parameters of the OTL circuit function.

Parameter	Units	Range
R_{b1}	kΩ	[50, 150]
R_{b2}	kΩ	[25, 70]
R_f	kΩ	[0.5, 30]
R_{c1}	kΩ	[1.2, 2.50]
R_{c2}	kΩ	[0.25, 1.2]
β	A	[50, 300]

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

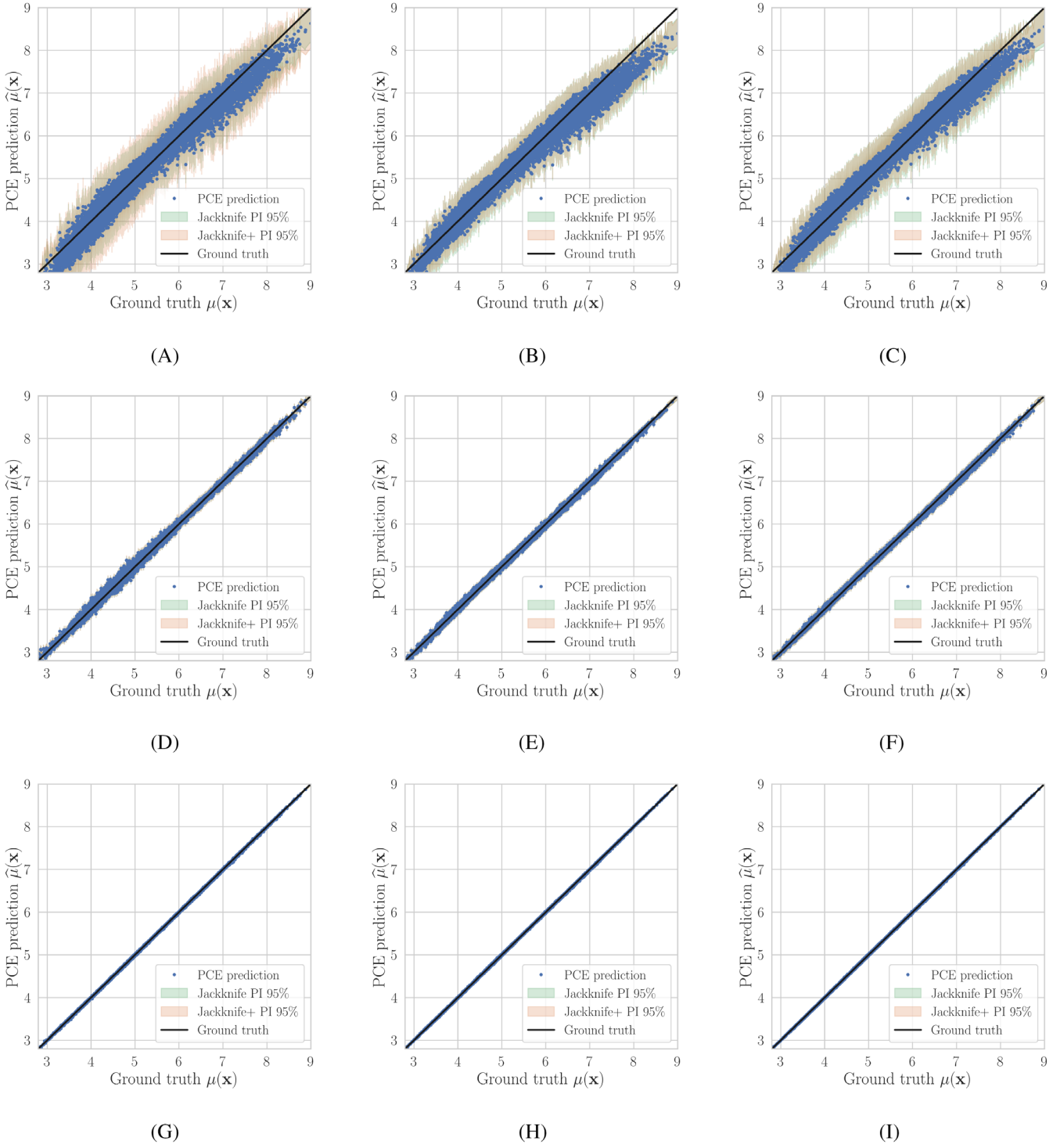

FIGURE C1 Parity plots comparing ground truth values of the OTL circuit function against conformalized PCE predictions for different combinations of polynomial degree P and oversampling coefficient C. The results correspond to a single random seed. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 1$, $C = 2$. (B) $P = 1$, $C = 5$. (C) $P = 1$, $C = 10$. (D) $P = 2$, $C = 2$. (E) $P = 2$, $C = 5$. (F) $P = 2$, $C = 10$. (G) $P = 3$, $C = 2$. (H) $P = 3$, $C = 5$. (I) $P = 3$, $C = 10$.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

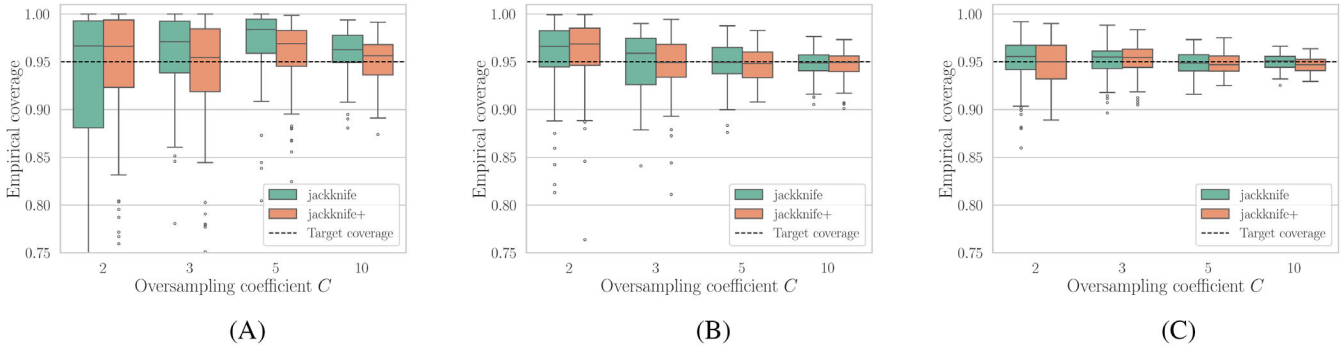

FIGURE C2 Box plots of the empirical coverage provided by conformalized PCE surrogates of the OTL circuit function, for different combinations of polynomial degree P, and oversampling coefficient C. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 1$. (B) $P = 2$. (C) $P = 3$.

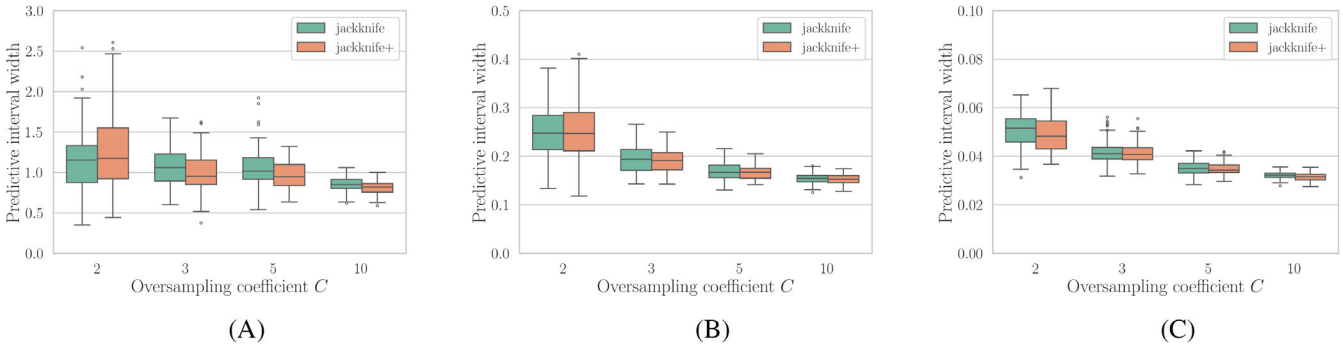

FIGURE C3 Box plots of the prediction interval widths provided by conformalized PCE surrogates of the OTL circuit function, for different combinations of polynomial degree P, and oversampling coefficient C. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 1$. (B) $P = 2$. (C) $P = 3$.

TABLE C2 Input parameters of the piston simulation function.

Parameter	Description	Units	Range
M	piston weight	kg	[30, 60]
S	piston surface area	m^2	[0.005, 0.02]
V_0	initial gas volume	m^3	[0.002, 0.01]
k	spring coefficient	N/m	[1000, 5000]
P_0	atmospheric pressure	N/m^2	[90000, 110000]
T_a	ambient temperature	K	[290, 296]
T_0	filling gas temperature	K	[340, 360]

Total-degree conformalized PCEs with maximum polynomial degrees $P \in \{2, 3, 4\}$ and oversampling coefficients $C \in \{2, 3, 5, 10\}$ are employed. Similar to Section C1, the experimental design scales linearly with the size of the PCE basis, that is, $M = CK = C\#\Lambda$. Figure C4 compares the various configurations of the conformalized PCE against the true piston simulation function for a single random seed, where the results for $C = 3$ have been omitted for better readability. Figures C5 and C6 show coverage and prediction interval statistics over 100 random seeds.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

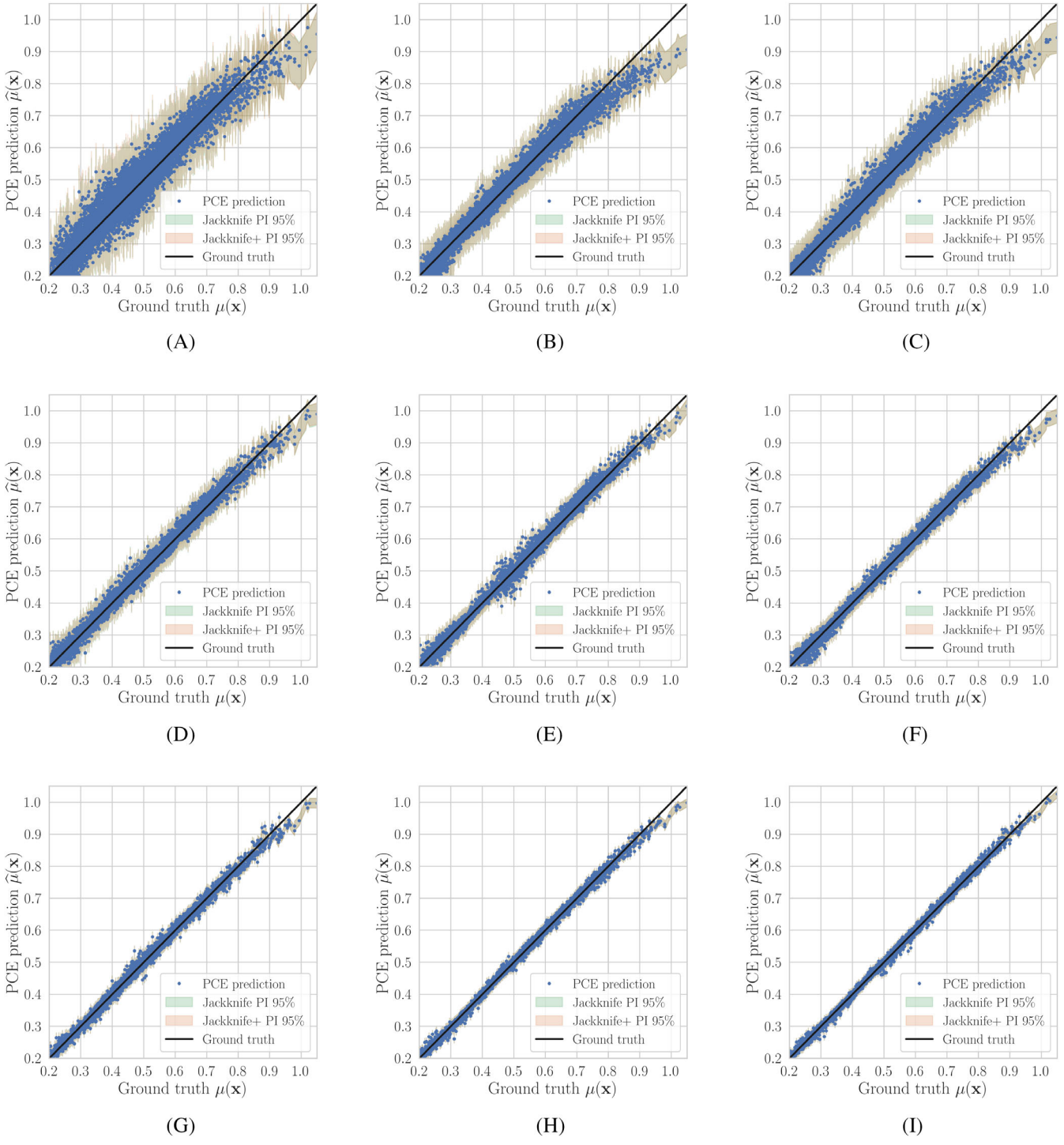

FIGURE C4 Parity plots comparing ground truth values of the piston simulation function against conformalized PCE predictions for different combinations of polynomial degree P and oversampling coefficient C. The results correspond to a single random seed. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 2$, $C = 2$. (B) $P = 2$, $C = 5$. (C) $P = 2$, $C = 10$. (D) $P = 3$, $C = 2$. (E) $P = 3$, $C = 5$. (F) $P = 3$, $C = 10$. (G) $P = 4$, $C = 2$. (H) $P = 4$, $C = 5$. (I) $P = 4$, $C = 10$.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

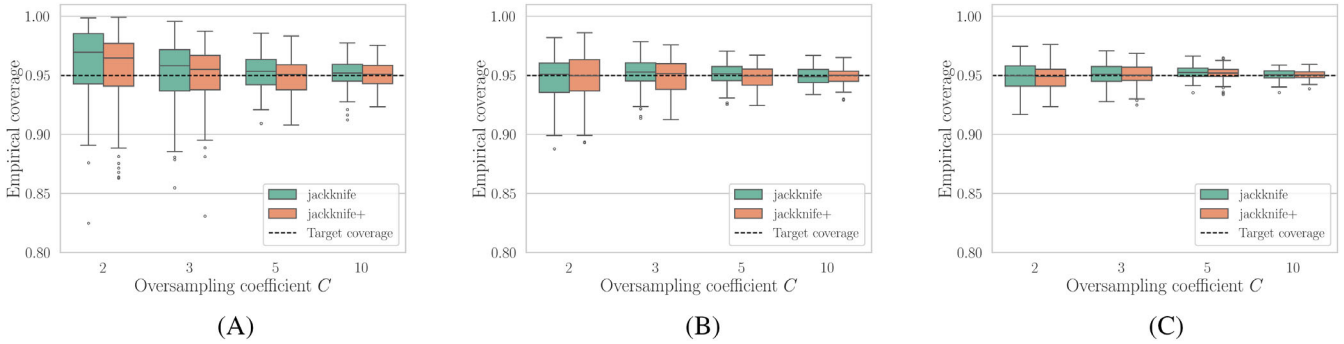

FIGURE C5 Box plots of the empirical coverage provided by conformalized PCE surrogates of the piston simulation function, for different combinations of polynomial degree P, and oversampling coefficient C. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 2$. (B) $P = 3$. (C) $P = 4$.

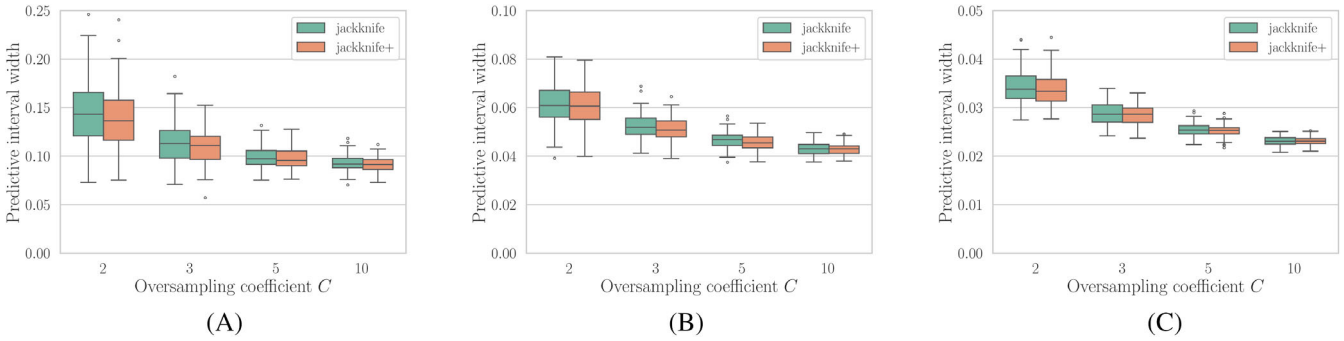

FIGURE C6 Box plots of the prediction interval widths provided by conformalized PCE surrogates of the piston simulation function, for different combinations of polynomial degree P, and oversampling coefficient C. The target coverage is fixed at 95%, corresponding to a significance level of $s = 0.05$. (A) $P = 2$. (B) $P = 3$. (C) $P = 4$.

15222608, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gamm.70017 by Centrum Voor Wiskunde En Info (Cwi), Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License