

RECENT ADVANCES IN POLYNOMIAL CHAOS EXPANSION: THEORY, APPLICATIONS AND SOFTWARE

Lukáš Novák¹, Drahomír Novák¹

¹Institute of Structural Mechanics, Faculty of Civil Engineering, Brno University of Technology,
Veveří 331/95, Brno, Czech Republic

novak.l@fce.vutbr.cz, novak.d@fce.vutbr.cz

DOI: 10.35181/tces-2023-0015

Abstract. *The paper is focused on recent advances in uncertainty quantification using polynomial chaos expansion (PCE). PCE is a well-known technique for approximation of costly mathematical models with random inputs – surrogate model. Although PCE is a widely used technique and it has several advantages over various surrogate models, it has still several limitations and research gaps. This paper reviews some of the recent theoretical developments in PCE. Specifically a new active learning method optimizing the experimental design and an extension of analytical statistical analysis using PCE will be reviewed. These two topics represent crucial tools for efficient applications: active learning leads generally to a significantly more efficient construction of PCE and improved statistical analysis allows for analytical estimation of higher statistical moments directly from PCE coefficients. Higher statistical moments can be further used for the identification of probability distribution and estimation of design quantiles, which is a crucial task for the probabilistic analysis of structures. Selected applications of the theoretical methods are briefly presented in a context of civil engineering as well as some preliminary results of further research. A part of the paper also presents UQPy package containing state-of-the-art implementation of the PCE theory.*

Keywords

Uncertainty quantification, polynomial chaos expansion, active learning, statistical analysis.

1. Introduction

The realistic analysis of structures is generally based on two aspects: the non-linear behavior of mathematical

models and the uncertainty of model parameters (e.g. material parameters, geometry etc.). Although there are various numerical methods for non-linear deterministic analysis, there is still a lack of efficient methods for general stochastic analysis. In modern structural analysis, uncertainties are represented by random variables described by specific probability distributions, the structural system can then be seen as a mathematical function of a set of random parameters. Evaluations of these functions existing in civil engineering are typically costly since their solutions are obtained numerically by the finite element method and it is necessary to reflect also non-linear behavior of the physical systems. Moreover, due to existing uncertainties in physical systems and/or their mathematical models, the analysis of structures must be enriched by stochastic analysis. The elementary task of stochastic analysis is to propagate uncertainties through a mathematical model and analyse the quantity of interest (QoI), e.g. statistical or sensitivity analysis, generally reference as uncertainty quantification (UQ).

The typical approach for UQ is a well-known crude Monte Carlo simulation based on a large number of repetitive deterministic calculations with randomly generated realizations of an input random vector. Although such an approach leads to accurate results, it is necessary to perform an enormous number of simulations, which is not feasible in industrial applications. Therefore, significant attention was recently given to surrogate models approximating the original mathematical model by simple functions, such as neural networks, support vector machines or polynomial chaos expansion (PCE). This paper is focused on PCE and its recent theoretical developments. Although uncertainty quantification has become one of the most growing research fields in the last two decades, it is still challenging to apply state-of-the-art techniques in practical applications. Therefore it was also important to develop efficient but user-friendly software solutions allowing

fast and accurate analysis of complex mathematical models. The recent theoretical developments and associated state-of-the-art numerical algorithms were thus implemented into software package UQPy (Uncertainty Quantification in Python).

2. Recent Theoretical Developments

The PCE approximates the quantity of interest (QoI) Y representing a result of the original mathematical model $\mathcal{M}(X)$, as a polynomial expansion of another random variable ξ called a germ with a given distribution. A set of polynomials, orthogonal with respect to the probability distribution of the germ, are used as basis functions. The orthogonality condition for all $j \neq k$ is given by the inner product of the Hilbert space defined for any two functions ψ_j and ψ_k with respect to the weight function p_ξ (probability density function of ξ) as:

$$\langle \psi_j, \psi_k \rangle = \int \psi_j(\xi) \psi_k(\xi) p_\xi(\xi) d\xi = 0. \quad (1)$$

Orthogonal polynomials ψ corresponding to common probability distributions p_ξ can be chosen according to the Wiener-Askey scheme [1] or numerically constructed in the case of an arbitrary probability distribution [2]. In the case of \mathbf{X} and ξ being vectors containing M random variables, the polynomial $\Psi(\xi)$ is multivariate and it is built up as a product of univariate orthogonal polynomials.

The QoI $Y = \mathcal{M}(\mathbf{X})$, can then be represented as [3]:

$$Y = \mathcal{M}(\mathbf{X}) = \sum_{\alpha \in \mathbb{N}^M} \beta_\alpha \Psi_\alpha(\xi), \quad (2)$$

where $\alpha \in \mathbb{N}^M$ is a set of integers called the multi-index, β_α are deterministic coefficients and Ψ_α are multivariate orthogonal polynomials. For practical computation, PCE expressed in Eq. (2) must be truncated to a finite number of terms P . The truncation is commonly achieved by retaining only the terms whose total degree $|\alpha|$ is less than or equal to a given p .

The truncated PCE is a simple linear regression model with intercept. Therefore, it is possible to use ordinary least squares (OLS) regression as simple non-intrusive solution. In order to use OLS for β estimation, it is necessary to first obtain n_{sim} realizations of the \mathbf{X} and the corresponding results \mathcal{Y} , together called the experimental design (ED). Then, the vector of deterministic coefficients β is calculated using data matrix Ψ as

$$\beta = (\Psi^T \Psi)^{-1} \Psi^T \mathcal{Y}. \quad (3)$$

The optimal size of ED is clearly affected by the number of terms P dependent on M and p . Therefore, it is typically useful to find a sparse solution using advanced model selection algorithms such as Least Angle Regression (LAR) [4, 5], orthogonal matching pursuit [6] or Bayesian compressive sensing [7] to find an optimal set of basis functions further denoted by \mathcal{A} .

2.1. Higher Statistical Moments

Once the PCE is constructed, the first statistical moment (the mean value) is simply the first deterministic coefficient of the expansion $\mu_Y = \langle Y^1 \rangle = \beta_0$. The second raw statistical moment, $\langle Y^2 \rangle$, can be obtained as

$$\begin{aligned} \langle Y^2 \rangle &= \int \left[\sum_{\alpha \in \mathcal{A}} \beta_\alpha \Psi_\alpha(\xi) \right]^2 p_\xi(\xi) d\xi \quad (4) \\ &= \sum_{\alpha_1 \in \mathcal{A}} \sum_{\alpha_2 \in \mathcal{A}} \beta_{\alpha_1} \beta_{\alpha_2} \int \Psi_{\alpha_1}(\xi) \Psi_{\alpha_2}(\xi) p_\xi(\xi) d\xi \\ &= \sum_{\alpha \in \mathcal{A}} \beta_\alpha^2 \int \Psi_\alpha(\xi)^2 p_\xi(\xi) d\xi = \sum_{\alpha \in \mathcal{A}} \beta_\alpha^2 \langle \Psi_\alpha, \Psi_\alpha \rangle \end{aligned}$$

Considering the orthonormality of the polynomials, it is possible to obtain the variance as the sum of all squared deterministic coefficients except the intercept (which represents the mean value).

Higher statistical central moments, skewness γ_Y (3rd moment) and kurtosis κ_Y (4th moment), require triple and quad products of basis functions. These can be obtained analytically only for certain polynomial families, e.g. formulas for Hermite and Legendre polynomials (and their combination) can be found in recent publication [8]. The formulas are based on standard linearization problem for certain polynomial families and Neumann-Adams formula. Although, the final expressions use only deterministic coefficients, they can be computationally expensive for growing P .

Once the first four statistical moments are available, it is possible to analytically approximate the probability density function by the Gram-Charlier expansion as depicted in Fig. 1. The additional information in form of higher moments can be thus used for the construction of arbitrary distribution corresponding to QoI. Moreover, it was shown in recent publication [8], that it is also possible to derive analytical formula for cumulative distribution function and use it further for moment-independent sensitivity analysis.

2.2. Active Learning

The size of ED for accurate PCE is affected not only by P but also by a position of each data point in a de-

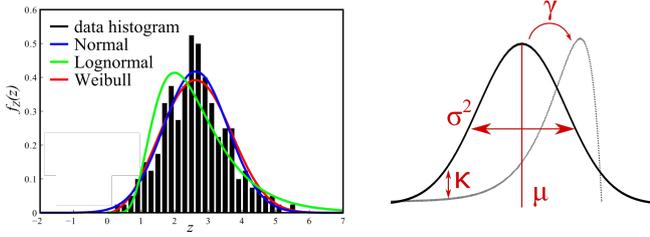


Fig. 1: Role of higher statistical moments in estimation of probability distribution.

sign domain, i.e. amount of information associated to each data point. Therefore, it is important to find the optimal position of realizations in the design domain extracting the highest possible amount of information about the original mathematical model. This task is generally referenced as active learning. Active learning should reflect two aspects: *exploration* of the design domain leading to identification of important locations, and *exploitation* of the available information about the mathematical model. An ideal active learning scheme should combine both aspects as recently proposed Θ -criterion designed specifically for PCE [9] resulting to optimal sampling schemes for given PCE and mathematical model (see Fig. 2). This approach selects the best candidate for extension of existing ED associated to high variance density (local contribution to the variance of QoI) as well as to large contribution to the exploration part of the criterion (distance from the existing data points in ED):

$$\Theta(\xi^{(c)}) \equiv \Theta^c = \frac{\sqrt{\sigma_A^2(\xi^{(c)}) \cdot \sigma_A^2(\xi^{(s)})} l_{c,s}^M}{\text{ave variance density} \cdot \text{vol.}} \quad (5)$$

The *exploration* aspect is maintained by accounting for the distance $l_{c,s}$ between a candidate $\xi^{(c)}$ and its nearest neighboring realization from the existing ED, $\xi^{(s)}$ as

$$l_{c,s} = \sqrt{\sum_{i=1}^M |\xi_i^{(c)} - \xi_i^{(s)}|^2}. \quad (6)$$

The *exploitation* component aims to sample points in regions with the greatest contributions to the total variance of the QoI σ_Y^2 , i.e. at points with the highest *variance density* (see Eq. 4) defined as

$$\sigma_A^2(\xi) = \left[\sum_{\alpha \in \mathcal{A}, \alpha \neq 0} \beta_\alpha \Psi_\alpha(\xi) \right]^2 p_\xi(\xi). \quad (7)$$

2.3. Uncertainty Quantification in Python

The theory of PCE together with presented recent advances was implemented into the new version of open-

source UQPy package [10]. UQPy represents multi-purpose complex software package for python containing recently developed techniques for uncertainty quantification including PCE, more details can be found on the official website, see QR codes in Fig. 3. UQPy contains several modules associated to common techniques for uncertainty quantification: probability distributions, statistical sampling, probabilistic transformation, stochastic processes, dimension reduction, inference, reliability analysis, surrogate modeling and sensitivity analysis. State-of-the-art techniques are present in each of the modules. The module for surrogate modeling contains various types of approximation including also presented PCE. The PCE module in UQPy contains techniques developed for advanced statistical sampling, efficient construction of the approximation (e.g. truncation schemes, sparse solvers [11]) and its post-processing (e.g. Sobol indices and complex statistical information derived from PCE). UQPy can be easily used for practical applications as well as for research, since it is an open-source package and anyone can contribute to the UQPy code once their contributions pass the quality checks. Such environment is ideal for further development of UQ methods as well as for routine applications of the theoretical methods in industry.

3. Selected Practical Applications

The theoretical background and recent advances have been applied already in several real-life examples from civil engineering and obtained results were partly published. Here we summarize two selected applications: a) Approximation of mathematical model representing a prestressed roof girder (see fig 4), which was further used for sensitivity analysis and optimization, b) statistical and sensitivity analysis of existing concrete bridge. Both approximations were constructed using advanced numerical algorithms and developed theoretical methods later implemented to UQPy.

3.1. Prestressed Roof Girders

The prestressed reinforced concrete roof girder is extensively described in [12]. Non-linear computational model of pre-stressed girder (see fig. 5) was created using ATENA Science software environment [13]. Geometry of beam and reinforcement was modelled exactly according to drawings provided by manufacturer. Two linear supports were used for the model, since rollers were used during test. Regular hexagonal FE mesh composed of 16728 finite elements was generated in the program GID preprocessor. The mesh has been

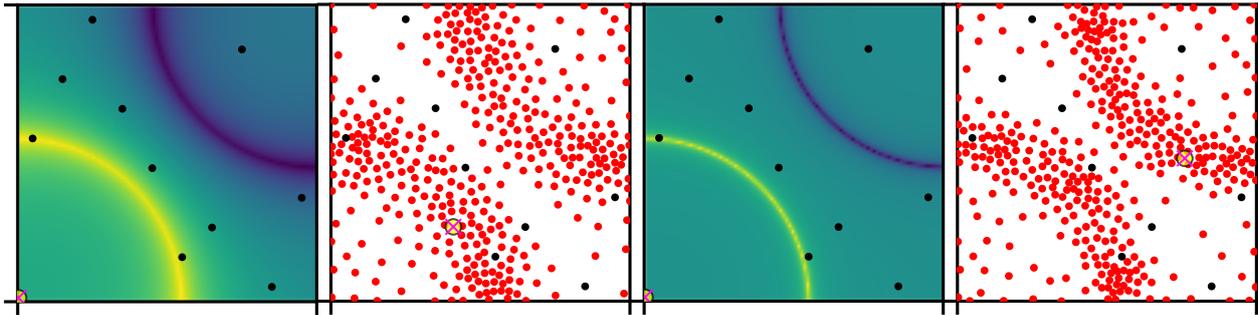


Fig. 2: Different shapes of mathematical models and associated optimal experimental design obtained by an active learning.

condensed in the area of assumed shear failure. Prestressing is applied as initial strain for reinforcement line. This application ensured that loss of prestressing due to elastic deformation of concrete was calculated explicitly. Material characteristics of concrete were obtained by laboratory experiments, including compressive strength of concrete f_c , tensile strength of concrete f_{ct} , fracture energy G_f and Young's modulus E . Statistical parameters of tendons and steel were assumed according to JCSS [14]. Full stochastic model contains 14 correlated random variables and correlation matrix of concrete material parameters was estimated from experimental results [12].

Latin Hypercube Sampling method was performed to generate 100 realizations in uncorrelated space. Because of the assumption of correlation among random variables, they were transformed to correlated space assuming Gaussian Copula using Nataf transformation. Evaluation of the original NLFEM was performed once the samples in correlated space were prepared and PCE was further constructed by non-intrusive approach. Specifically, a truncated set of basis functions was obtained by total-order scheme with $p = 8$ and number of basis functions was further reduced by LAR. Obtained accuracy of PCE was measured by leave-one-out-error $Q = 0.01$ and it shows suitable accuracy for statistical or sensitivity analysis, though it could lead to significant errors in a reliability analysis and an estimation of extreme quantiles of PDF. More details about the process of surrogate modeling can be found in previous studies of authors [15, 16].

The main task was sensitivity analysis of the QoI, since the stochastic model contains a several uncertain parameters with complicated correlation structure. Therefore, various types of sensitivity analysis were performed [17]: Spearman rank-order correlation, Sobol indices see in Fig. 6 and finally analysis of covariance (generalization of Sobol indices). On the one hand, Spearman rank-order correlation measures the strength and direction of association among variables and on the other hand, ANCOVA measures contribution of input variable to the output variance. Moreover, sensitivity analysis also directly quantified the role of correlation among random variables. Obtained results further served for reduction of stochastic model and optimization of girders [18].

3.2. Concrete Bridge

The existing post-tensioned concrete bridge has three spans. The super-structure of the mid-span is 19.98 m long with total width 16.60 m and it is crucial part of the bridge for assessment. In transverse direction, each span is constructed from 16 bridge girders KA-61 commonly used in Czech Republic. Load is applied according to national annex of Eurocode for load-bearing capacity of road bridges by exclusive loading (by six-axial truck).

The non-linear finite element model (NLFEM) is created using software ATENA Science based on theory of non-linear fracture mechanics [13]. In order to reflect complex behavior of the bridge, the numerical model contains three construction phases as illustrated in Fig.7. The NLFEM consists of 13,000 elements of hexahedra type in the major part of the volume and triangular 'PRISM' elements in the part with complicated geometry. Reinforcement and tendons are represented by discrete 1D elements with geometry according to original documentation. The numerical model is further analysed in order to investigate the following three limit states: ultimate limit state, first occurrence of bending cracks, decompression in prestressed concrete.



Fig. 3: UQPy software: left) QR code leading to Git-Hub repository containing the open-source code in python, middle) the graphical logo representing the package, right) QR code leading to documentation of the package.



Fig. 4: Complex long-term research of prestressed concrete roof girders: material experiments, scaled destructive tests and numerical modeling.

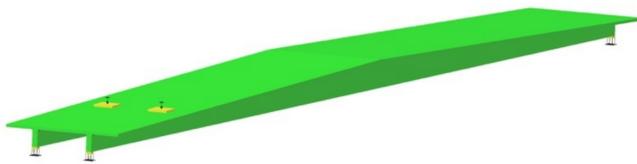


Fig. 5: Non-linear finite element model of prestressed concrete roof girders in software Atena Science.

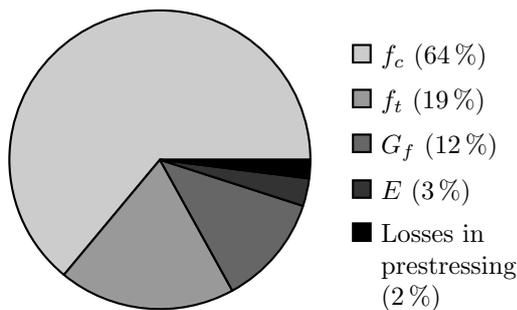


Fig. 6: Sobol indices of the analyzed pre-stressed concrete roof girder.

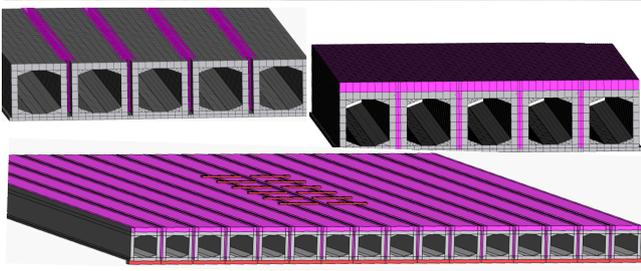


Fig. 7: The analyzed post-tensioned concrete bridge. Top: photography of the bridge; bottom: non-linear finite element model of the bridge.

The stochastic model contains 4 random material parameters of a concrete C50/60: Young's modulus; compressive strength of concrete f_c ; tensile strength of concrete f_{ct} and fracture energy G_f . Characteristic values of E , f_{ct} , G_f were determined from f_c according to formulas implemented in the fib Model Code 2010 [19] (G_f , E) and prEN 1992-1-1: 2021 (f_{ct}). The last random variable P represents prestressing losses with CoV according to JCSS: Probabilistic Model Code [14]. Mean values and coefficients of variation were obtained according to prEN 1992-1-1: 2021 (Annex A) for adjustment of partial factors for materials.

The experimental design (ED) contains 30 numerical simulations generated by Latin Hypercube Sampling (LHS). Note that each simulation takes approximately 24 hours and construction of the whole ED took approx. 1 week of computational time. PCE was constructed using advanced adaptive $p \in [5, 10]$ algorithm together with Least Angle Regression algorithm implemented in UQPpy. Obtained accuracy of PCE was measured by leave-one-out-error $Q = 0.002$. Once the approximation was created, it was possible to perform million of simulations instantly to identify histograms and distributions of analyzed limit states (e.g., see Fig. 8). Note that although the difference is not significant, it leads to dramatic differences in estimated quantiles used for estimation of the design value of resistance.

4. Conclusion

The paper presented recent theoretical developments in PCE significantly extending its applicability in uncertainty quantification of structures. Combination of extended statistical analysis with improved efficiency by active learning leads to computationally effective and accurate tool for UQ of costly mathematical models as was shown in selected two industrial examples. A part of the paper presented also a UQPpy, an open-source package for python containing state-of-the-art theoretical techniques and efficient algorithms for UQ. UQPpy offers unique combination of computational efficiency with user-friendly architecture and thus it can be eas-

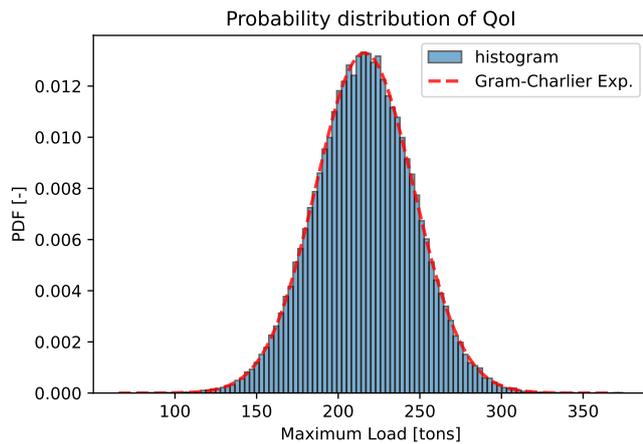


Fig. 8: The histogram of the analyzed limit state: decompression of the prestressed concrete. Red line corresponds to analytical estimation of PDF by Gram-Charlier Expansion.

ily used for practical applications. Note that surrogate models in both practical applications were based on one-shot ED created by LHS. Although the obtained accuracy was sufficient for performed UQ tasks, further work will be focused on application of the recently proposed active learning algorithm in order to improve their accuracy allowing for reliability analysis.

Acknowledgment

The authors acknowledge the financial support provided by the Czech Science Foundation under project number 22-00774S.

References

- XIU, D.; G. E. KARNIADAKIS. The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*. 2002, vol. 24, no. 2, pp. 619–644. Available from DOI: 10.1137/s1064827501387826.
- WAN, X.; G. E. KARNIADAKIS. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *Journal of Computational Physics*. 2005, vol. 209, no. 2, pp. 617–642. ISSN 0021-9991. Available from DOI: 10.1016/j.jcp.2005.03.023.
- GHANEM, R. G.; P. D. SPANOS. *Stochastic Finite Elements: A Spectral Approach*. Springer New York, 1991. ISBN 0387974563. Available from DOI: 10.1007/978-1-4612-3094-6.
- EFRON, B.; T. HASTIE; I. JOHNSTONE; R. TIBSHIRANI. Least angle regression. *The Annals of Statistics*. 2004, vol. 32, no. 2, pp. 407–451. ISSN 00905364. Available from DOI: 10.2307/3448465.
- BLATMAN, G.; B. SUDRET. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*. 2011, vol. 230, no. 6, pp. 2345–2367. Available from DOI: 10.1016/j.jcp.2010.12.021.
- TROPP, J. A.; A. C. GILBERT. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory*. 2007, vol. 53, no. 12, pp. 4655–4666. Available from DOI: 10.1109/tit.2007.909108.
- JI, S.; Y. XUE; L. CARIN. Bayesian Compressive Sensing. *IEEE Transactions on Signal Processing*. 2008, vol. 56, no. 6, pp. 2346–2356. Available from DOI: 10.1109/TSP.2007.914345.
- NOVÁK, L. On distribution-based global sensitivity analysis by polynomial chaos expansion. *Computers & Structures*. 2022, vol. 267, p. 106808. ISSN 0045-7949. Available from DOI: 10.1016/j.compstruc.2022.106808.
- NOVÁK, L.; M. VOŘECHOVSKÝ; V. SADÍLEK; M. D. SHIELDS. Variance-based adaptive sequential sampling for Polynomial Chaos Expansion. *Computer Methods in Applied Mechanics and Engineering*. 2021, vol. 386, p. 114105. ISSN 0045-7825. Available from DOI: <https://doi.org/10.1016/j.cma.2021.114105>.
- TSAPETIS, D.; M. D. SHIELDS; D. G. GIOVANIS; A. OLIVIER; L. NOVAK; P. CHAKROBORTY; H. SHARMA; M. CHAUHAN; K. KONTOLATI; L. VANDANAPU; D. LOUKREZIS; M. GARDNER. UQpy v4.1: Uncertainty quantification with Python. *SoftwareX*. 2023, vol. 24, p. 101561. ISSN 2352-7110. Available from DOI: <https://doi.org/10.1016/j.softx.2023.101561>.
- NOVÁK, L.; D. NOVÁK. Polynomial chaos expansion for surrogate modelling: Theory and software. *Beton- und Stahlbetonbau*. 2018, vol. 113, pp. 27–32. Available from DOI: 10.1002/best.201800048.
- SLOWIK, O.; D. NOVÁK; L. NOVÁK; A. STRAUSS. Stochastic modelling and assessment of long-span precast prestressed concrete elements failing in shear. *Engineering Structures*. 2021, vol. 228, p. 111500. ISSN 0141-0296. Available from DOI: <https://doi.org/10.1016/j.engstruct.2020.111500>.

13. ČERVENKA, J.; V. K. PAPANIKOLAOU. Three dimensional combined fracture-plastic material model for concrete. *International Journal of Plasticity*. 2008, pp. 2192–2220.
14. JCSS. *JCSS Probabilistic Model Code*. Joint Committee on Structural Safety, 2001. ISBN 978-3-909386-79-6.
15. LEHKÝ, D.; D. NOVÁK; L. NOVÁK; M. ŠOMODÍKOVÁ. Prestressed concrete roof girders: Part II – surrogate modeling and sensitivity analysis. In: *Proceedings of the Sixth International Symposium on Life-Cycle Civil Engineering (IALCCE 2018)*. CRC press, Taylor and Francis group, 2018, pp. 2649–2698. ISBN: 978-113862633-1.
16. NOVAK, L.; D. NOVAK. Surrogate modelling in the stochastic analysis of concrete girders failing in shear. In: *Proc. of the Fib Symposium 2019: Concrete - Innovations in Materials, Design and Structures*. 2019.
17. PAN, L.; L. NOVÁK; D. LEHKÝ; D. NOVÁK; M. CAO. Neural network ensemble-based sensitivity analysis in structural engineering: Comparison of selected methods and the influence of statistical correlation. *Computers & Structures*. 2021, vol. 242, p. 106376. ISSN 0045-7949. Available from DOI: <https://doi.org/10.1016/j.compstruc.2020.106376>.
18. SLOWIK, O.; D. LEHKÝ; D. NOVÁK. Reliability-based optimization of a prestressed concrete roof girder using a surrogate model and the double-loop approach. *Structural Concrete*. 2021, vol. 22, no. 4, pp. 2184–2201. Available from DOI: <https://doi.org/10.1002/suco.202000455>.
19. FEDERATION INTERNATIONALE DU BETON. *fib Model Code for Concrete Structures 2010*. Berlin, Heidelberg: John Wiley & Sons, 2013. ISBN 9783433030615.

About Authors

Lukáš NOVÁK was born in Hustopeče, Czech Republic. He received his M.Sc. from Brno University of Technology in 2018 and his Ph.D. in 2021. His research interests include uncertainty quantification, polynomial chaos expansions, semi-probabilistic design of structures and theory of probability.

Drahomír NOVÁK was born in Brno, Czech Republic. He received his M.Sc. from Brno University of Technology in 1984 and his PhD in 1991. Since 2003 he has been engaged as a full professor at Brno University of Technology. His research interests