Proceedings of Machine Learning Research 266:1–3, 2025 Conformal and Probabilistic Prediction with Applications

# A Bayesian framework for calibrating Gaussian process predictive distributions

**Aurélien Pion** AURELIEN.PION@CENTRALESUPELEC.FR
*Transvalor S.A., Biot, France*
*Université Paris-Saclay, CNRS, CentraleSupélec, L2S, Gif-sur-Yvette, France*
**Emmanuel Vazquez** EMMANUEL.VAZQUEZ@CENTRALESUPELEC.FR
*Université Paris-Saclay, CNRS, CentraleSupélec, L2S, Gif-sur-Yvette, France*

**Editor:** Khuong An Nguyen, Zhiyuan Luo, Harris Papadopoulos, Tuwe Löfström, Lars Carlsson and Henrik Boström

Gaussian processes (GPs) are Bayesian models widely used for interpolating an unknown deterministic function $f : \mathcal{X} \subset \mathbb{R}^d \to \mathbb{R}$ from observed data. A GP model places a Gaussian-process prior on $f$, written as $\xi \sim \mathrm{GP}(m, k)$, where $m : \mathcal{X} \to \mathbb{R}$ is the mean function and $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a positive definite covariance kernel. Given noiseless observations $\mathcal{D}_n = \{(x_i, f(x_i))\}_{i=1}^n$, the GP posterior defines a predictive distribution at any $x \in \mathcal{X}$, which is Gaussian with posterior mean $m_n(x)$ and posterior variance $\sigma_n^2(x)$, both computable in closed form.

In the noiseless setting, GPs are prominent tools for sequential prediction, as their posterior distributions provide both predictions and uncertainty quantification. Two common examples of sequential predictions are Bayesian optimization (see, e.g., Jones et al., 1998) and estimation of excursion probabilities (see, e.g., Bect et al., 2012).

In practice, however, GP predictive distributions are frequently miscalibrated: the actual frequency with which $f(x)$ falls within nominal confidence intervals (with respect to a distribution over $\mathcal{X}$) may substantially deviate from the intended level, as shown by Pion and Vazquez (2025). This can lead to overconfident predictions that underestimate uncertainty, or overly conservative ones that overstate it. In practice, it is generally preferable to have conservative predictive distributions.

Among available approaches to improve calibration, conformal prediction (CP) is particularly attractive because it is model-agnostic and provides distribution-free guarantees on marginal coverage. It has been adapted to GPs through full conformal prediction (FCP) method by Papadopoulos (2024), or via J+GP, a variant of Jackknife+ (Barber et al., 2021) proposed by Jaber et al. (2024), providing post-hoc correction at user-specified coverage levels. In another line of work, Vovk et al. (2017b) introduce the Conformal Predictive Systems (CPS), based on CP to build a cumulative distribution function (CDF) at each test point for an unknown label. CPS have been adapted to kernel methods by Vovk et al. (2017a) and thus naturally apply to GP. However, when few points are available, CP methods may produce prediction intervals of infinite length. Moreover, CP methods guarantee only marginal coverage, not coverage for a given dataset.

For these reasons, we propose a Bayesian approach to constructing predictive distributions for GPs and introduce a novel method named *calGP*. The method retains the GP posterior mean as regression estimates, but models normalized prediction errors using a generalized normal distribution. The shape and scale parameters of this distribution are selected using a Bayesian strategy inspired by tolerance intervals (Meeker et al., 2017). The resulting predictive distribution remains centered on the GP posterior mean, while allowing for improved calibration—particularly in the tails—and supports continuous inference at arbitrary confidence levels. By adjusting the variance of the predictive distribution independently of its mean, calGP makes it possible to control the level of conservativeness in uncertainty quantification.

Figure 1 presents a comparison of three calibration methods: the proposed calGP, along with J+GP and FCP. Predictive intervals from the predictive distributions of the GP are also presented. Intervals from J+GP and FCP can become infinite when the confidence level exceeds $1/(n+1)$, failing to capture local variations in uncertainty. calGP captures regions of high deviation ("excursions"), demonstrating improved calibration performance.

In this study, we focus on comparing empirical coverages of prediction intervals provided by calGP, FCP, and J+GP, evaluated on multiple functions using a test grid.

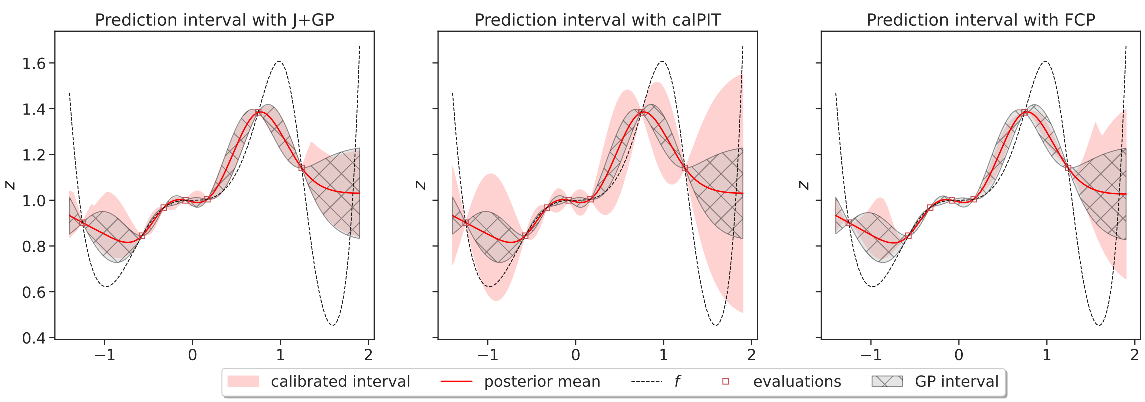


Figure 1: Prediction intervals constructed by J+GP (left) and calGP (middle) and FCP (right) at confidence level $1-\alpha = 0.75$. The parameter $\delta$ for calGP is set to 0.03. J+GP and FCP intervals may become unbounded when $\alpha > 1/(n+1)$, and fail to adapt to excursions. calGP yields more informative, location-sensitive intervals.

## Acknowledgments

This work was funded by Transvalor S.A.

## References

R. F. Barber, E. J. Candès, A. Ramdas, and R. J. Tibshirani. Predictive inference with the jackknife+. *Ann. Stat.*, 49(1), 2021.

J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Stat. Comput.*, 22:773–793, 2012.

E. Jaber, V. Blot, N. Brunel, V. Chabridon, E. Remy, B. Iooss, D. Lucor, M. Mougeot, and A. Leite. Conformal approach to gaussian process surrogate evaluation with coverage guarantees, 2024. hal-04389163 (preprint submitted on 11 January 2024).

Donald Jones, Matthias Schonlau, and William Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 12 1998. doi: 10.1023/A:1008306431147.

W. Q. Meeker, G. J. Hahn, and L. A. Escobar. *Statistical Intervals: A Guide for Practitioners and Researchers*. John Wiley & Sons, Hoboken, New Jersey, second edition, 2017. ISBN 978-0-471-68717-7.

H. Papadopoulos. Guaranteed coverage prediction intervals with gaussian process regression. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12):9072–9083, dec 2024.

A. Pion and E. Vazquez. Gaussian process interpolation with conformal prediction: Methods and comparative analysis. In G. Nicosia, V. Ojha, S. Giesselbach, M. P. Pardalos, and R. Umeton, editors, *Mach. Learn. Optim. Data Sci.*, pages 218–228, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-82484-5.

Vladimir Vovk, Ilia Nouretdinov, Valery Manokhin, and Alex Gammerman. Conformal predictive distributions with kernels, 2017a. URL https://arxiv.org/abs/1710.08894.

Vladimir Vovk, Jieli Shen, Valery Manokhin, and Min-ge Xie. Nonparametric predictive distributions based on conformal prediction. In Alex Gammerman, Vladimir Vovk, Zhiyuan Luo, and Harris Papadopoulos, editors, *Proceedings of the Sixth Workshop on Conformal and Probabilistic Prediction and Applications*, volume 60 of *Proceedings of Machine Learning Research*, pages 82–102. PMLR, 13–16 Jun 2017b.