Marginal Percentile Intervals in Bayesian Inference are Overconfident

Sebastian Höpfl * **Hans-Michael Tautenhahn** **
Vincent Wagner * **Nicole Erika Radde** *

* *Institute for Stochastics and Applications, University of Stuttgart, Germany (e-mail: nicole.radde@simtech.uni-stuttgart.de).*
** *Clinic for Visceral, Transplantation, Thoracic and Vascular Surgery, Leipzig University Hospital, Germany*

Abstract: In Bayesian statistics, Highest Posterior Density Regions (HPDR) measure the uncertainty of parameter estimates at a given credibility level. HPDRs are the Bayesian version of Highest Density Intervals (HDI) based on the posterior probability. The calculation of Percentile Intervals (PI) is a common method for approximating HDIs in many popular Python toolboxes. The PI calculation differs from the HPDR calculation as it ignores the posterior probability by simply cutting the lower and upper values of the marginalized posterior distribution symmetrically. Here, we use the phenomenological retarded transient functions to infer the posterior distribution of a clinical dataset. The one-dimensional HPDR projections were compared to the PIs for all inferred parameters. The direct comparison revealed that the one-dimensional HPDR projections of all inferred parameter posteriors exceeded the percentile-based intervals, demonstrating that the PIs were overconfident. Overall, we argue that only HPDRs can be interpreted in terms of probability.

Data availability: Data and code is available on DaRUS (`https://doi.org/10.18419/darus-4068`).

Keywords: Bayesian methods, Bayesian Inference, Systems biology, Highest Density Interval (HDI), Highest Posterior Density Region (HPDR)

1. INTRODUCTION

The combination of randomness, inter-individual differences, and typically sparse datasets complicates accurate inference in biomedicine. Uncertainty quantification considers these factors and enables reliable decisions based on limited data. For example, the recommendation for RNA sequencing is to have between six and 12 biological replicates (Schurch et al., 2016). In practice, many molecular biological experiments are conducted with only three replicates (Ma et al., 2004; Mishra et al., 2011; KB et al., 2017). Ill-posed inverse problems result from this sparse data, i.e., the data does not contain enough information to estimate model parameters accurately. To make decisions in this sparse data setting, the uncertainty has to be quantified reliably.

Summary statistics, such as the Highest Density Interval (HDI) of inferred model parameters, are one established way to evaluate uncertainties. A HDI has two properties: (i) the probability density inside the interval is at least as high as outside of the interval, and (ii) it has the shortest length for a given cumulative probability of $\alpha\%$ inside the interval (Box and Tiao, 1992). However, different approximations of these statistics can result in quite different uncertainty estimates. Here, we focus on the Highest Posterior Density Regions (HPDR) and Percentile Intervals (PI) as Bayesian uncertainty summary statistics examples.

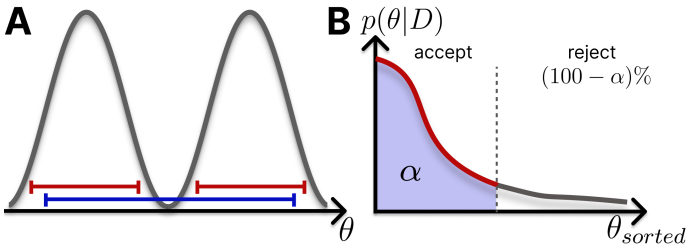

Fig. 1. **Characterization and calculation of the HPDR.** A. HPDR (red) and PI (blue) for a one-dimensional bimodal distribution. Only the HPDR is split between the two modes, while the PI is bound symmetrically. B. The HPDR can be estimated from posterior samples according to their posterior values. The $(100-\alpha)\%$ samples with the lowest posterior probability are cut.

In Bayesian inference, random variables describe the parameters θ of simulation models $s(\theta)$. A posterior parameter distribution $p(\theta|D)$ is inferred, given a prior parameter distribution $p(\theta)$, the observed data D, and the simulation model. Bayes' Theorem is employed to obtain the posterior distribution

$$p(\theta|D) \propto p(D|\theta) \cdot p(\theta), \tag{1}$$

where $p(D|\theta)$ is the likelihood of observing the data D given a parameter θ.

HPDRs are specific to posterior distributions, covering $\alpha\%$ of the posterior probability mass. Property (i) of HDIs is adjusted for HPDRs to the samples' posterior probability, which also corresponds to the density of the posterior distribution: The posterior probability of each point inside the HPDR must be higher than anywhere outside the interval. In contrast, PIs of one-dimensional distributions are symmetrically bounded by cutting $(\frac{100-\alpha}{2})\%$ of the samples, resulting in the $[z_{\frac{100-\alpha}{2}}, z_{100-\frac{100-\alpha}{2}}]$ percentiles. PIs and HPDRs are equal for one-dimensional unimodal and symmetric distributions. However, the bounds of PIs and HPDRs differ when applied to bimodal or skewed distributions. In particular, PIs will never be split for the two modes (Fig. 1A), while this might be the case for HPDRs if the posterior probability between the two modes is too low. For bimodal distributions, PIs violate HDI property (i) because regions of high probability density are excluded, while low-density regions are included in the PI. This violation can disguise valuable information about the underlying process and model properties. Further, the PI calculation can be inaccurate under non-ideal conditions. In practice, these conditions might be a not wholly converged Markov Chain Monte Carlo (MCMC) chain, the limitation of drawing only a finite number of samples, or boundary artifacts. Still, PI calculation is often applied as it is computationally and mathematically simple.

Both PIs and HPDR are used to approximate HDIs of single parameters, in practice. While PIs are always continuous, HPDR can be split several times. On the marginal parameter level, we report the maximum uncertainty of the HPDR (HPDR_{max}), corresponding to the outer bounds of the HPDR. The HPDR_{max}, therefore, violates the HDI properties whenever the actual HPDR is split into several subintervals and is only a summary statistic of the outer or maximum uncertainty.

In this study, we analyze a clinical dataset on liver regeneration after hepatectomy (Fig. 2A). Hepatectomy is performed to remove benign and malignant lesions. The liver is the organ with the greatest regeneration capabilities, allowing the regeneration of the pre-operative liver volume within weeks. The dataset contains 24 patients with extended partial liver resection, varying in age, with and without comorbidities, postoperative complications, and medications (ZeLeR-study, ethical vote: 2018-1246-Material). Liver volumes were obtained from computed tomography (CT) measurements and normalized according to the preoperative liver volume for each patient.

In clinical settings, it is crucial to report accurate uncertainties and prevent overconfidence when treatment strategies or research are based on model predictions. The risk judgment for a distinct treatment can only be accurate if the clinician's information about this and possible alternative procedures is accurate (Chen and Asch, 2017). Ultimately, this information is crucial for people's well-being (Kappen et al., 2018).

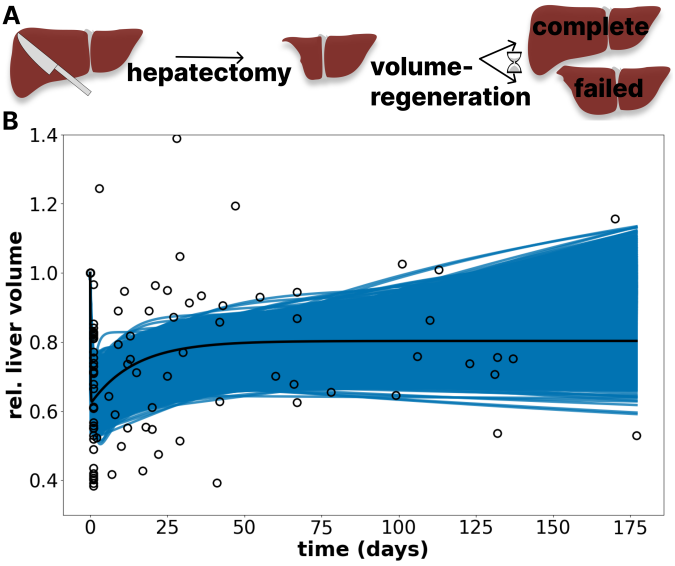

Fig. 2. **Regeneration after partial hepatectomy.** A. In hepatectomy, a part of the liver is resected. The liver remnant completely restores its preoperative volume in a successful regeneration. B. Posterior ensemble prediction of 10,000 90% Highest Posterior Density Region samples (blue) with the maximum a posteriori estimate (black line) and the normalized liver volume data (black circles).

2. RESULTS

In practice, Bayesian posteriors $p(D|\theta)$ are difficult to compute analytically. MCMC sampling is frequently used to sample from the posterior, resulting in a set of accepted parameter samples $\{\theta_i^k\}$, with $k = 1, ..., K$ parameters and $i = 1, ..., I$ samples. The accepted parameter samples are proportional to the posterior probability $p(\theta_i|D) \propto p(D|\theta_i) \cdot p(\theta_i)$.

This paper focuses on the sample-based calculation of HPDRs and PIs. The HPDRs are calculated by first sorting the converged joint samples θ_i^k according to their posterior probability. Subsequently, the $(100-\alpha)\%$ samples with the lowest posterior probability are cut, resulting in an $\alpha\%$ HPDR (Algorithm 1 and Fig. 1B).

Sample-based PIs are calculated by symmetrical cutting $(\frac{100-\alpha}{2})\%$ from the empirical cumulative distribution of each parameter marginal separately (Algorithm 2).

2.1 Bayesian Inference of the clinical data using retarded transient functions

To describe liver regeneration, we use the Retarded Transient Functions (RTF) of Kreutz (2020) with identical response times of the transient and sustained response (t_1):

$$s_{RTF}(t) = \underbrace{A_{\text{sus}}\left(1 - e^{-\frac{t}{t_1}}\right)}_{\text{sustained response}} + \underbrace{A_{\text{trans}}\left(1 - e^{-\frac{t}{t_1}}\right)e^{-\frac{t}{t_2}}}_{\text{transient response}} + p_0 \tag{2}$$

and transformed time

$$t = \log_{10}\left(10^{t_{\text{real}} \cdot \frac{10}{T_{\text{range}}}} + 10^{T_{\text{shift}}}\right) - \log_{10}\left(1 + 10^{T_{\text{shift}}}\right). \tag{3}$$

Algorithm 1 Marginal maximum HPDR calculation from posterior samples

Given a converged joint posterior parameter sample θ_i^k with $k = 1, ..., K$ parameters and $i = 1, ..., I$ samples.

1: Sort samples $\theta_i^{\cdot}$ descending by their posterior probability $p(\theta_i^{\cdot}|D)$ to get the sorted samples $\tilde{\theta}$.
2: Remove samples with index i higher than $\lceil \alpha\% \cdot I \rceil$: $\tilde{\tilde{\theta}} = \{\tilde{\theta}_i^{\cdot}, i \leq \lceil \alpha\% \cdot I \rceil\}$
3: **for** $k = 1, ..., K$: **do**
4: Calculate lower and upper uncertainty bounds:
5: $k_l = \min_i(\tilde{\tilde{\theta}}_i^k)$
6: $k_u = \max_i(\tilde{\tilde{\theta}}_i^k)$
7: The $\alpha\%$ $\text{HPDR}_{\max}$ of parameter k is $[k_l, k_u]$
8: **end for**

Here, $\lceil x \rceil$ denotes rounding to the nearest integer higher than x.

Algorithm 2 Marginal PI calculation from posterior samples

Given a converged joint posterior parameter sample θ_i^k with $k = 1, ..., K$ parameters and $i = 1, ..., I$ samples.

1: **for** $k = 1, ..., K$: **do**
2: Sort marginal vector $\theta_{\cdot}^k$ by value to get the sorted marginal samples $\hat{\theta}_{\cdot}^k$
3: Calculate the lower and upper percentile bounds:
4: $k_l = \hat{\theta}^k \left[\lfloor \frac{100-\alpha}{200} \cdot I \rfloor \right]$
5: $k_u = \hat{\theta}^k \left[\lceil \frac{100+\alpha}{200} \cdot I \rceil \right]$
6: The PI of k is $[k_l, k_u]$
7: **end for**

Here, $\lceil x \rceil$ denotes rounding to the nearest integer greater than x and $\lfloor x \rfloor$ denotes rounding to the nearest integer smaller than x.

RTFs can describe various biological processes without requiring a priori knowledge about the underlying metabolics. Computationally, RTFs are advantageous over ordinary differential equations as they can be evaluated directly without numerical integration. For a general process description, a sustained and a transient response function are added with amplitude parameters A_{sus} and A_{trans}, response time t_1, decay time t_2, and an initial offset p. The time transformation allows for delayed responses, for example, starting after several postoperative days. We assume an immediate response as the hepatectomy is performed directly after preoperative liver volume assessment ($T_{shift} = -2$). The observed data has a range of 177 days ($T_{range} = 177$) and shows huge variability as discussed in Höpfl et al. (2024). As the liver volume per patient was normalized to 1 preoperatively, the offset parameter p was set to 1.

We assume independent and identically distributed noise with normally distributed measurement errors arising from the measurement method. The noise was calculated as the standard deviation of the complete dataset with $\sigma = 0.226$. For T observed days and N replicates, we obtain the following likelihood function $p(D|\theta)$:

$$p(D|\theta) = \prod_{k=1}^{T} \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(s(t_k, \theta) - m^{(j)}(t_k)\right)^2}{2\sigma^2} \right). \tag{4}$$

Equation (4) is a standard likelihood, penalizing the squared distance between the simulated value $s(t_k, \theta)$ at time t_k and the measurement of the j-th replicate at that time $m^{(j)}(t_k)$.

A uniform prior distribution with wide bounds according to Kreutz (2020) was used for the amplitude of the sustained response A_{sus} and the response time of the transient and sustained responses t_1 as these parameters are specific to the size of the resection performed in this study. For the amplitude of the transient response A_{trans}, a normally distributed prior with the average resected liver volume as mean was used. Further, it is known that the human liver can restore its functional mass within two weeks after up to 66% partial hepatectomy (Koniaris et al., 2003). Therefore, a normal distribution with a mean of 14 days was used for the prior of the transient response decay time t_2. Negative response times were excluded as they are unphysiological. A reproducible version of the parameter estimation problem in the PEtab format (Schmiester et al., 2021) can be found on DaRUS (`https://doi.org/10.18419/darus-4068`), including all results and a validation with a broader normal prior distribution.

For Bayesian inference, ten million MCMC samples were drawn with the adaptive metropolis sampler of pyPESTO (Schälte et al., 2021), leading to an effective sample size of 151,069. The posterior was estimated for the parameters

$$\theta = (A_{\text{sus}}, A_{\text{trans}}, t_1, t_2). \tag{5}$$

Geweke's diagnostic was used to remove burn in samples. The Posterior ensemble prediction of the 90% HPDR shows that the RTF captures the data and uncertainty with good accuracy (Fig. 2B).

2.2 Connection of the posterior samples in the joint space is cut by percentile interval calculation

Using HPDR-based posterior samples ensures the samples are connected in the joint posterior space. In contrast, PIs are cut on the one-dimensional marginals, removing the joint connection of posterior samples whenever only one marginal parameter value exceeds a PI bound. This loss of the joint connection is demonstrated on a two-variate normal distribution (Fig. 3A). The elliptical HPDR represents the covariance of the 2D normal distribution. On the marginals, this covariance structure is not visible. The PIs cut $\alpha\%$ of the samples in each marginal. On the joint space, this affects more than $\alpha\%$ of the samples as sometimes the value for one parameter dimension is cut while it persists for the other. In total, 16 out of 100 samples (Fig. 3A, samples with red edge color) were excluded by at least one marginal bound, while only 10 samples in total were excluded by the HPDR (Fig. 3A, grey samples). This difference in affected samples explains why the estimated maximum bounds of the HPDRs exceed the PI bounds.

While PIs only use the parameter value of each sample during computation, HPDRs use the additional dimension of

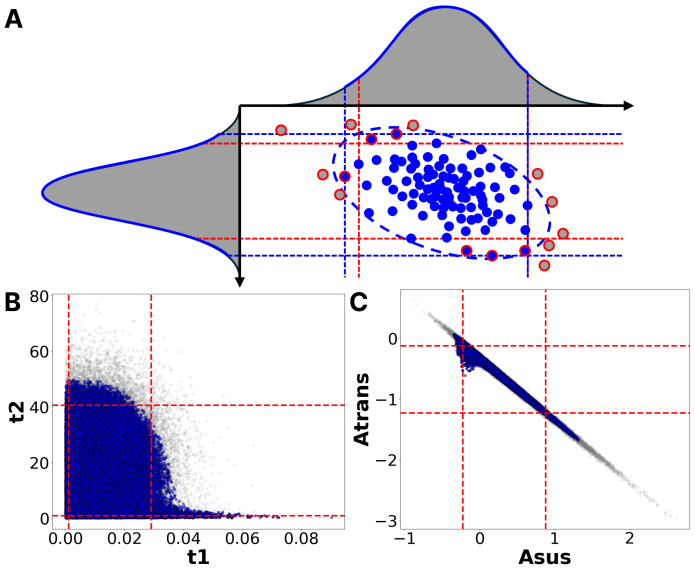

Fig. 3. **2D comparison of PIs and HPDRs** A. PI bounds (red dashed) and HPDR (blue dashed ellipse) of 100 samples of a two-variate normal distribution. The maximum HPDR uncertainty is the projection of each dimension's outer highest posterior density samples. Blue samples are part of the HPDR, the HPDR excludes grey samples, and the PIs exclude samples with red edge color. B, C. Converged full posterior samples (grey), 90% HPDR samples (blue), and marginal PI bounds (red dotted) for the A_{sus}-A_{trans}, and t_1-t_2 scatter of the liver regeneration example.

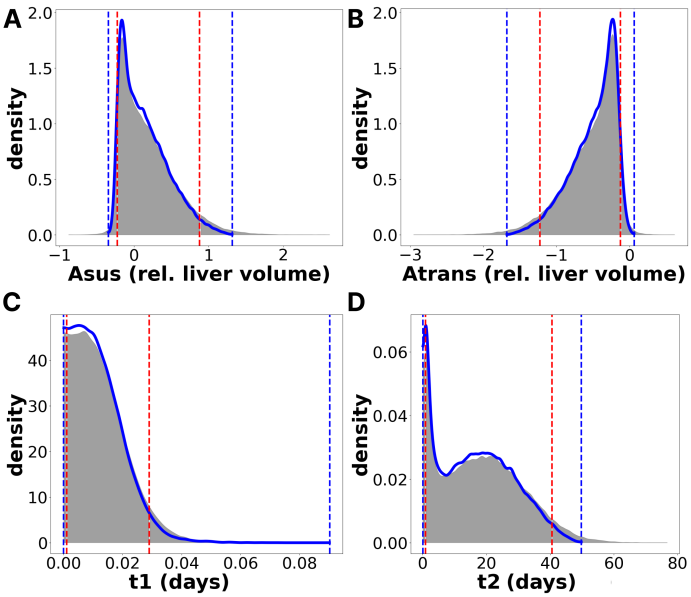

Fig. 4. **Comparison of the PIs and HPDR$_{\max}$ of the liver regeneration example.** Kernel density estimates of the full converged chain (shaded grey), 90% PIs (between red dotted lines), and 90% HPDR$_{\max}$ (between blue dotted lines, blue line) for: A. the sustained response amplitude A_{sus}, B. the transient response amplitude A_{trans}, C. the response time of the transient and sustained response t_1, and E. the decay time of the transient response t_2.

the joint samples' posterior probabilities. HPDRs remove samples based on the posterior probability. Therefore, samples in the center of the posterior distribution can also be removed if their posterior probability is lower than the credibility level (Fig. 3B, C). For our inferred posterior distribution, the t_1 PI, for example, cuts all t_1 marginal values above 0.03 while the t_2 values above this threshold are still included in the t_2 PI between $[0.88, 40.59]$ (Fig. 3B).

2.3 Percentile intervals underestimate the maximum uncertainty

Removing parameter combinations in the joint space can result in split marginal intervals. This split occurs because parameter values in the center of the marginals can be removed. Strictly, one could argue that the HPDR needs to end at every removed parameter sample in the marginal space. Terminating HPDRs at every removed value would lead to many split intervals for the marginal parameter. To calculate the maximum uncertainty, this is practically undesirable, and detecting split intervals increases the needed computational effort. Therefore, we projected the lowest and highest HPDR parameter values in each dimension, giving a maximum uncertainty interval (HPDR$_{\max}$) per parameter to compare HPDRs to the PIs.

The maximum uncertainty, for a parameter k is defined by $[\min_i(\tilde{\tilde{\theta}}_i^k), \max_i(\tilde{\tilde{\theta}}_i^k)]$, i.e., the lower and upper marginal HPDR bounds for parameter k at a given credibility level α (Algorithm 1). For more advanced methods for HPDR estimation from samples, we refer to Held (2004). We are aware that this maximum uncertainty region might comprise several split HPDR subintervals. However, for many clinical applications, decisions are based only on maximum uncertainty. For those decisions, it is most important not to underestimate the uncertainty.

Looking at the marginal parameter space, there is a vast difference between the PI bounds and the HPDR$_{\max}$ (Fig. 4A-D and Table 1). All marginalized HPDRs exceed the calculated PIs. The 90% HPDR$_{\max}$ exceeds the 90% PI between 25-245% for the parameters of our on the clinical dataset calibrated RTF (Table 1). The absolute difference in the HPDR and PIs bounds increases with the skewness of the marginal distributions. This increase persists for all marginals, with slightly heavy tails on the upper or lower boundary (Fig. 4A-D). Larger HPDRs show that the marginal PIs underestimate the maximum uncertainty in this real-world setting. Therefore, reporting and drawing conclusions based on percentile HDI approximations tends to be overconfident. In decision-making and uncertainty quantification based on marginal parameter values, only the HPD-based regions are based on posterior probability. Therefore, we argue that only the HPDR can be interpreted in terms of probability.

One example of a critical decision based on this liver resection data is assessing the regeneration time. The regeneration time, i.e., the time in which the liver regrows, can be approximated by the transient response decay contestant t_2. Knowing the uncertainty of this parameter could indicate when regeneration has failed, additional measures to aid regeneration should be taken, or when subsequent procedures could be scheduled. The maximum

Table 1. **90% PI, 90% HPDR_{max}, and length increase of the HPDR compared to the PI of the estimated liver regeneration parameters.**

parameter	90% PI	90% HPDR_{max}	length increase
A_{sus}	$[-0.22, 0.88]$	$[-0.34, 1.32]$	51%
A_{trans}	$[-1.22, -0.12]$	$[-1.67, 0.07]$	58%
t_1	$[0.001, 0.03]$	$[4 \cdot 10^{-8}, 0.1]$	245%
t_2	$[0.88, 40.59]$	$[0.03, 49.81]$	25%

regeneration time with 90% credibility is about 41 days for the PI and about 50 days for the HPDR. This difference of nine days could lead to overconfident decisions and implicate premature follow-up treatments.

2.4 Parameter scatters indicate partly correlated response amplitudes

The amplitude parameters A_{sus} and A_{trans} are negatively correlated (Fig. 5). This correlation holds for a narrow range of relative liver weight in $[-3, 1]$ for A_{trans} and $[-1, 3]$ for A_{sus}. In this range, the sustained part can compensate for the transient part and vice versa. This correlation might indicate that our observed data is insufficient to strictly separate the effect of the transient and sustained response in the first postoperative days. The response time parameters are not correlated but ran to their lower physical bound at zero, indicating a rapid volume decrease after resection, which is expected.

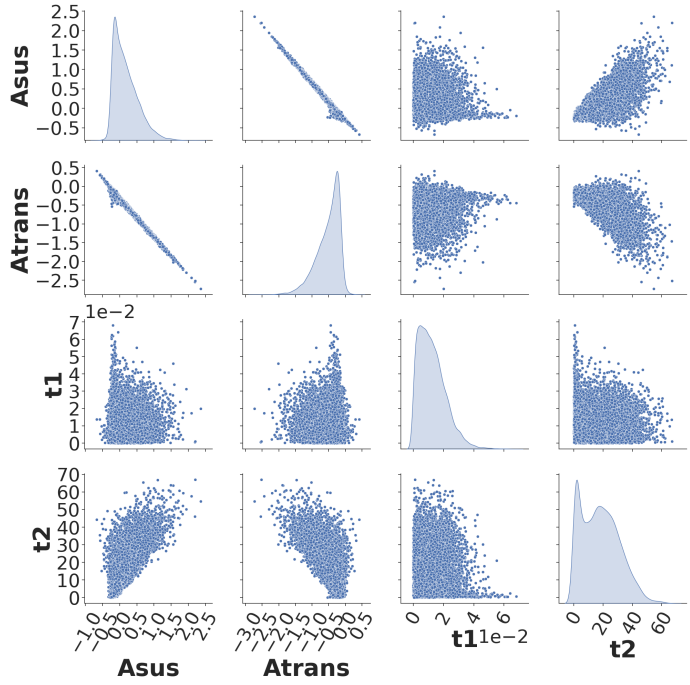

Fig. 5. **Sampling scatter of the estimated liver regeneration parameters.** Ten million MCMC samples were drawn from the posterior distribution with the adaptive metropolis sampler of pyPESTO (Schälte et al., 2021), leading to an effective sample size of 151,069. Geweke's diagnostic was used to remove burn in samples. Every 100th parameter value of the posterior chain is displayed.

3. DISCUSSION

Approximating HDIs for posterior parameters based on marginal PIs can lead to overconfident estimates of the maximum uncertainty. This overconfidence appears whenever parameter combinations in the center of marginal distributions have lower posterior probability than at the margins. Using one-dimensional HPDRs in these cases is more accurate since the calculation is based on the probability of the joint posterior distribution. Using the maximum HPDR instead of PIs increased the outer interval bounds on a model for liver regeneration after hepatectomy calibrated to a clinical dataset. We argue that only using HPDRs allows the interpretation of the intervals in terms of probability, as PIs are not based on the posterior probability.

We showed that using PIs to estimate and assess uncertainties might create a false impression of the maximum uncertainty. Ultimately, underestimated uncertainty might lead to wrong decisions in the clinics, where pharmacokinetic parameters can be decisive for a particular procedure. Here, the decay time of the transient response could be interpreted as an estimate of the time patients need to regenerate. If this time is underestimated, follow-up procedures could be initiated prematurely. Therefore, practitioners should always ensure to report reliable uncertainty estimates that do not underestimate the maximum uncertainty. Our current dataset is sparse and has too many confounding factors to reliably support such decisions. Nevertheless, this example serves as an illustration.

In low dimensional settings, PIs approximate HDIs reasonably whenever the posterior is smooth and unimodal, and parameters are not correlated. Further, PIs can be used to approximate central tendencies. We expect that the overconfidence of PIs increases with the dimensionality of the parameter space as more parameter combinations are possible there. In addition, correlated parameters or non-identifiabilities can increase the difference between the HPDRs and PIs. Here, the HPDR might include the complete range of the correlated or non-identifiable parameter, while the PI is cut by $\frac{100-\alpha}{2}\%$ symmetrically.

State-of-the-art Bayesian Python libraries, including Arviz (Kumar et al., 2019), approximate HDIs with percentiles per default. This might have led to the report of multiple overconfident credibility intervals in the literature, as described by Caspi et al. (2023) in the inference of RNA virus haplotypes mutation rate, selection, and epistasis, for example. Therefore, one should always be aware of the method used to approximate HDIs.

Computationally, the cost of our pragmatic approach to calculating HPDRs is about the same as calculating PIs. While HPDR is calculated by simply cutting samples based on a lower posterior probability threshold, this gives only an approximation of the maximum uncertainty as the algorithm does not so far include the detection of split interval regions. Including the detection of split intervals would increase the computation time and will be part of our future work.

4. CONCLUSION AND OUTLOOK

HPDR in Bayesian inference improves sampling diagnostics and provides a more accurate way of communicating probabilities. In particular, the HPDR is the only measure to get precise estimates of the maximum uncertainty. For a clinical dataset on liver resection, all marginal $\text{HPDR}_{\max}$ exceeded the corresponding PIs, showing the PIs were overconfident. We provide a simple algorithm to calculate the HPDR sample-based and plan to incorporate these improvements in the pyPESTO Python toolbox for Bayesian Inference (Schälte et al., 2021). Ultimately, the aim should be to computationally efficiently detect split HPDR and move from the maximum HPDR to the actual HPDR.

ACKNOWLEDGEMENTS

Funded by the Research Unit Programme FOR 5151 QuaLiPerF (Quantifying Liver Perfusion–Function Relationship in Complex Resection—A Systems Medicine Approach) by grant no. 436883643 and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 - 390740016. We acknowledge the support by the Stuttgart Center for Simulation Science (SimTech). HMT was supported by the Federal Ministry of Education and Research (BMBF, Germany) within ATLAS (grant number 031L0304B) and by the DFG by grant number 465194077 (Priority Programme SPP 2311, Subproject SimLivA). The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG. We acknowledge the use of Grammarly for language editing.

REFERENCES

Box, G.E. and Tiao, G.C. (1992). *Bayesian Inference in Statistical Analysis*. Wiley. doi:10.1002/9781118033197.

Caspi, I., Meir, M., Ben Nun, N., Abu Rass, R., Yakhini, U., Stern, A., and Ram, Y. (2023). Mutation rate, selection, and epistasis inferred from RNA virus haplotypes via neural posterior estimation. *Virus evolution*, 9(1), vead033. doi:10.1093/ve/vead033.

Chen, J.H. and Asch, S.M. (2017). Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *The New England journal of medicine*, 376(26), 2507–2509. doi:10.1056/NEJMp1702071.

Held, L. (2004). Simultaneous Posterior Probability Statements From Monte Carlo Output. *Journal of Computational and Graphical Statistics*, 13(1), 20–35. doi:10.1198/1061860043083.

Höpfl, S., Albadry, M., Dahmen, U., Herrmann, K.H., Kindler, E.M., König, M., Reichenbach, J.R., Tautenhahn, H.M., Wei, W., Zhao, W.T., and Radde, N.E. (2024). Bayesian modeling of time series data (BayModTS) - a FAIR workflow to process sparse and highly variable data. *Bioinformatics (Oxford, England)*. doi:10.1093/bioinformatics/btae312.

Kappen, T.H., van Klei, W.A., van Wolfswinkel, L., Kalkman, C.J., Vergouwe, Y., and Moons, K.G.M. (2018). Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagnostic and prognostic research*, 2, 11. doi:10.1186/s41512-018-0033-6.

KB, S., JJA, P., V, V., and K, S. (2017). Analyzing the Efficacy of Phosphate Solubilizing Microorganisms by Enrichment Culture Techniques. *Biochemistry & Molecular Biology Journal*, 03(01). doi:10.21767/2471-8084.100029.

Koniaris, L.G., McKillop, I.H., Schwartz, S.I., and Zimmers, T.A. (2003). Liver regeneration. *Journal of the American College of Surgeons*, 197(4), 634–659. doi:10.1016/S1072-7515(03)00374-0.

Kreutz, C. (2020). A New Approximation Approach for Transient Differential Equation Models. *Frontiers in Physics*, 8. doi:10.3389/fphy.2020.00070.

Kumar, R., Carroll, C., Hartikainen, A., and Martin, O. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software*, 4(33), 1143. doi:10.21105/joss.01143.

Ma, J.F., Mitani, N., Nagao, S., Konishi, S., Tamai, K., Iwashita, T., and Yano, M. (2004). Characterization of the silicon uptake system and molecular mapping of the silicon transporter gene in rice. *Plant physiology*, 136(2), 3284–3289. doi:10.1104/pp.104.047365.

Mishra, A., Kumar, S., Bhargava, A., Sharma, B., and Pandey, A.K. (2011). Studies on in vitro antioxidant and antistaphylococcal activities of some important medicinal plants. *Cellular and Molecular Biology*, 57(1), 16–25. URL `https://mail.cellmolbiol.org/index.php/CMB/article/view/951`.

Schälte, Y., Fröhlich, F., Stapor, P., Vanhoefer, J., Weindl, D., Lakrisenko, P., Raimúndez, E., Pathirana, D., Schmiester, L., Städter, P., Contento, L., Merkt, S., Dudkin, E., Grein, S., and Hasenauer, J. (2021). pyPESTO - Parameter EStimation TOolbox for python. doi:10.5281/zenodo.5647404. URL `https://doi.org/10.5281/zenodo.5647404`.

Schmiester, L., Schälte, Y., Bergmann, F.T., Camba, T., Dudkin, E., Egert, J., Fröhlich, F., Fuhrmann, L., Hauber, A.L., Kemmer, S., Lakrisenko, P., Loos, C., Merkt, S., Müller, W., Pathirana, D., Raimúndez, E., Refisch, L., Rosenblatt, M., Stapor, P.L., Städter, P., Wang, D., Wieland, F.G., Banga, J.R., Timmer, J., Villaverde, A.F., Sahle, S., Kreutz, C., Hasenauer, J., and Weindl, D. (2021). PEtab-Interoperable specification of parameter estimation problems in systems biology. *PLoS computational biology*, 17(1), e1008646. doi:10.1371/journal.pcbi.1008646.

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., Blaxter, M., and Barton, G.J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA (New York, N.Y.)*, 22(6), 839–851. doi:10.1261/rna.053959.115.