Simulation-based Inference for Cardiovascular Models

Antoine Wehenkel
Apple

Jens Behrmann
Apple

Andrew C. Miller
Apple

Guillermo Sapiro
Apple

Ozan Sener
Apple

Marco Cuturi
Apple

Jörn-Henrik Jacobsen
Apple

Abstract

Over the past decades, hemodynamics simulators have become tools of choice for studying cardiovascular systems and are routinely used to simulate whole-body hemodynamics from physiological parameters. Nevertheless, solving the corresponding inverse problem of mapping waveforms back to plausible physiological parameters remains challenging. Motivated by advances in simulation-based inference (SBI), we cast this inverse problem as statistical inference. Our study highlights the potential of estimating new biomarkers from standard-of-care measurements and reveals practically relevant findings that cannot be captured by standard sensitivity analyses, such as the existence of sub-populations for which parameter estimation exhibits distinct uncertainty regimes. In addition, we study how such insights obtained in-silico transfer to in-vivo with the MIMIC-III database.
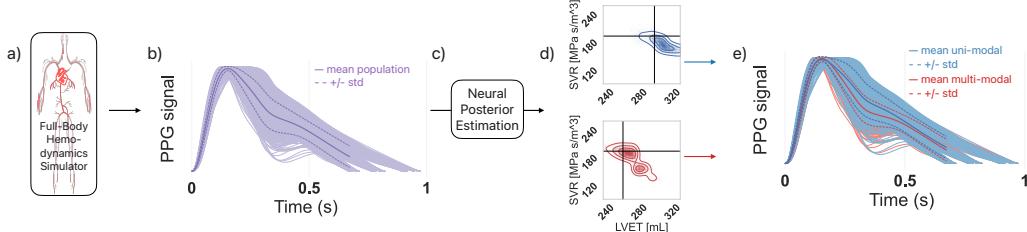
Figure 1: SBI enables uncertainty analysis at the level of individual measurements and reveals two sub-populations with, respectively, uni-modal and bi-modal uncertainty distributions. **a**: Whole-body hemodynamics simulator. **b**: Measurements generated from a simulated population of humans. **c**: Neural posterior estimation provides posterior distributions of the parameters of interest given the digital PPG. **d**: Two posterior distributions respectively corresponding to an individual measurement from each identified sub-population, highlighting the benefit of uncertainty representation at the individual level. **e**: Sub-sets of measurements corresponding to the two identified sub-population, revealing distinct morphological characteristics in each sub-group.

1 Introduction

Nowadays, biophysical models can describe cardiac function with 3D models [6], or even emulate hemodynamics in the entire human arterial system [39, 12, 1]. These models have moved from paper calculations [4, 48] to complex numerical simulations [60, 39, 12, 1] and support the development of personalized monitoring and treatment of cardiovascular (CV) diseases. While whole-body 1D hemodynamics simulators [39, 12] establish a clear path from latent physiological variables to measurable biosignals, their use for scientific inquiry, or precision medicine, necessitates solving the corresponding *inverse* problem of inferring latent biomarkers from measurable biosignals.

Recent works have studied these inverse problems with variance-based sensitivity analysis, highlighting which biomarkers have the most decisive influence on measured biosignals [41, 53, 50]. In

parallel, machine learning approaches, relying on sophisticated patterns for predicting biomarkers from biosignals, have gained popularity [11, 26, 7, 25, 8]. While these approaches provide an essential step towards a better understanding of the inverse problem, they do not address the challenges caused by the non-deterministic and multi-modal nature of inverse solutions. Motivated by breakthroughs in simulation-based inference [SBI, 15, 58], which has addressed similar challenges in other scientific fields, we go beyond producing point-estimates for such inverse problems and consider instead a distributional perspective supported by neural posterior estimation [36]. As a result, the SBI methodology provides a *consistent, multi-dimensional* and, *individualized* representation of uncertainty and naturally handles ambiguous inverse solutions, as showcased in Figure 1.

2 Background on Hemodynamics and SBI

2.1 Inverting Whole-body 1D Cardiovascular Simulations

We define a simulator as a forward generative process $g : \Theta \to \mathcal{X}$ that inputs a vector of parameters $\theta \in \Theta$ and returns a simulation $\mathbf{x} \in \mathcal{X}$. Simulators may depend on a large number of parameters and be stochastic. In practice, we split the parameters $\theta = (\phi, \psi) \in \Theta = \Phi \times \Psi$ into variables of direct interest $\phi \in \Phi$ and nuisance parameters $\psi \in \Psi$ that are necessary to run the simulations but are not of direct interest for the downstream task.

We rely in this work on the simulator from [12], describing the hemodynamics in the 116 largest human arteries. The model's parameters describe the blood out-flowing the left ventricle, uni-dimensional physical properties of each artery, and a lumped-element model of the vascular beds, see Appendix A.1 for more details about this model. We study the identifiability of the parameters $\phi \in \Phi$ of physiological interest (a.k.a. biomarkers), from a given measurement $\mathbf{x} \in \mathcal{X}$ (a.k.a. biosignals). The biomarkers considered are the heart rate [HR, 29], the left ventricular ejection time [LVET, 3], the average diameter of the arteries [Diameter, 47], the pulse wave velocity [PWV, 55], and the systemic vascular resistance [SVR, 14]; which are all relevant to assess CV health as supported by the provided references [29, 3, 47, 55, 14]. We consider biosignals that are commonly collected in intensive care units (ICUs) or in medical studies: the arterial pressure waveform (APW) at the radial artery and the photoplethysmograms (PPGs) at the digital and the radial arteries.

We study a virtual population aged 25 to 75 with several (up to 100s of) free parameters θ that model heterogeneous cardiac and arterial properties. In this context, a consistent representation of the solution's uncertainty is key, in order to capture **1.** the effect of nuisance parameters, responsible for the forward model stochasticity; **2.** the symmetries of the forward model, leading to non-unique inverse solutions; **3.** the lack of sensitivity, magnifying small output uncertainty into high input uncertainty; and **4.** the heterogeneity of the population considered, leading to distinct uncertainty profiles.

2.2 Simulation-based Inference (SBI)

SBI [15] has established itself as an essential tool in various domains of science that rely on complex simulations, e.g., in astrophysics [19, 16, 62], particle physics [10], neuroscience [34, 33], robotics [38], and many others [23, 59, 5]. SBI extends statistical inference to statistical models defined implicitly from a simulator and provides a consistent representation of uncertainty as demanded by the four requirements listed in the previous paragraph.

While possible, applying statistical inference to simulators is challenged by the computational complexity of likelihood evaluation, if not its complete intractability. Here, the CV simulator takes a few minutes to run on a modern CPU, preventing the use of non-amortized inference algorithms, such as Markov Chain Monte Carlo sampling. As a solution, SBI algorithms leverage modern machine learning methods to tackle inference in likelihood-free settings [35, 18, 21], see [15] for a thorough review. Here, we rely on neural posterior estimation [NPE, 36], a *Bayesian* and *amortized* method, which learns a surrogate of the posterior distribution $p(\phi \mid \mathbf{x})$ with conditional density estimation, enabling fast and accurate approximation of the true posterior distributions for any observation.
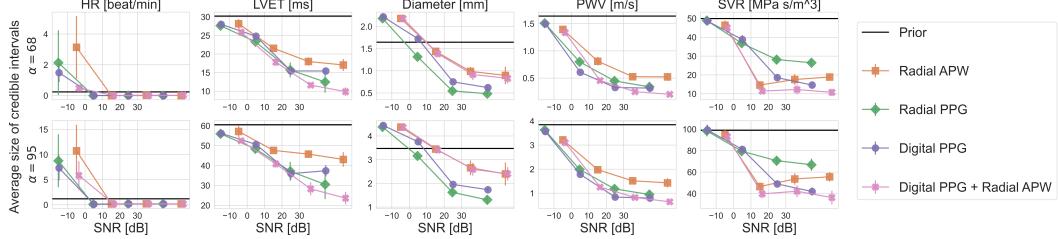
2

Figure 2: Average size of credible intervals over the test population for credibility levels 68% and 95% of the learned posterior distributions. The x-axis denotes the signal-to-noise-ratio (SNR) for different types of measurements. Results are averaged over five training instances, the vertical bars report one standard deviation.

3 Results

Our experiments consider both in-silico and in-vivo scenarios. We split the simulation dataset from [12] into train (70%), validation (10%), and test sets (20%) at random. All results reported are on the test set for the NPE models that maximize the validation likelihood, error bars report the standard deviation over five training instances. As a prerequisite for a meaningful analysis, we have checked that the posterior distributions learned are well statistically calibrated, see Appendix A.4.1.

3.1 In-silico analysis

SBI enables comprehensive population-level uncertainty analyses. Figure 2 shows the average size of credible intervals for the parameters of interest. We consider an additive Gaussian noise with five amplitudes and observe the size of intervals as a function of SNR. Comparing these to the intervals of the prior distribution enables to quantify the information content of a measurement about the biomarkers. Unsurprisingly, the HR is easily identified from all measurements, except for very high levels of noise. Overall, uncertainty about all parameters reduces significantly as the noise level decreases. This observation indicates that the measurements carry information about all parameters considered, as highlighted by other studies [41, 12, 13]. The results also highlight that each measurement has its unique information content. For instance, the digital PPG reveals more about SVR and PWV than the radial PPG. However, it is the opposite for the Diameter for which the Radial PPG is the most informative measurement. These results highlight that, similarly to standard sensitivity analyses, SBI enables interpretable assessment of the predictability of biomarkers from biosignals, in-silico, while having additional properties exemplified in subsequent experiments.
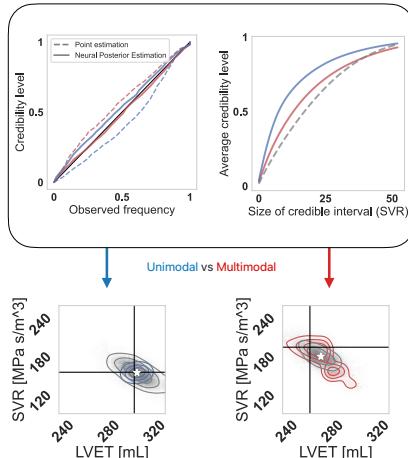

Figure 3: NPE (top: plain line, bottom: in red and blue) vs. Laplace's approximation (top: dashed lines, bottom: gray). Colors denote the different populations considered (cf. Figure 1), black lines denote the true value of the parameter, and the white star is the point estimate. NPE is better calibrated than the Laplace's approximation (left top plot) and yields tighter credible intervals (right top plot).

SBI enables per-individual uncertainty quantification. Figure 3 compares the estimation of uncertainty provided by NPE and Laplace's approximation [37] around the expectation of the posterior distribution, which is representative of the underlying assumptions made in variance-based sensitivity analyses (VBSAs). Similarly to VBSAs, Laplace's approximation models uncertainty through a second-order statistic over the population considered. NPE's credible intervals are tighter and better calibrated than Laplace's ones, which are either overconfident for measurements that lead to multi-modal posterior distributions or under-confident otherwise. We argue that an inconsistent quantification of uncertainty, e.g., as obtained under over-simplified assumptions made by VBSAs, may be misleading.

Figure 1 sketches the use of SBI to study the relationship between the digital PPG and the SVR and LVET. The figure highlights distinctive aspects of posterior distributions within the

3

population studied for which we tested multi-modality [22]. While the uncertainty about the value of SVR and LVET can be reduced substantially for approximately half of the test population, for the other half, the posterior is multi-modal. Although a point estimator is reasonable for the first sub-population, it is a poor guess for the multi-modal sub-population. Together, Figure 3 and Figure 1 demonstrate that the consistent, multi-dimensional, and individualized representation of uncertainty, provided by NPE, yields essential insights from CV models that are left unnoticed by VBSAs.

3.2 In-vivo analysis

Models are never a perfect representation of real-world data [9]. Misspecification, as it becomes more significant, may hamper the practical relevance of insights extracted from a model [66]. Nevertheless, conclusions that are independent of the most critical sources of misspecification are often valid both in-silico and in-vivo. We now consider real-world data from the MIMIC-III dataset [28] and evaluate how observations made in-silico can translate to real-world insights.

MIMIC-III results. In Figure 4, we assess the performance of surrogate posterior distributions learned on simulations in predicting HR and LVET using 8-second waveforms from the MIMIC-III dataset [28]. Examples of such waveforms are showcased in Appendix A.3. As the posterior distributions are uni-modal for the LVET and HR, we focus on point estimates obtained by taking the expectation of the posterior distributions. While we can accurately determine HR by counting the number of beats, we only get noisy labels for LVET [17, 3]. Although these labels are not very accurate, they serve as a baseline for comparison. We evaluate the mean absolute error (MAE) and correlation between the point estimates and the labels. We observe successful transfer of posterior distributions to real-world data for HR but not for LVET. The MAE of LVET approaches that of the prior distribution, indicating limited improvement. However, the remaining correlation between the predicted and real LVET values suggest a partial transfer of information.

On the one hand, in-silico and in-vivo results consistently show that HR estimation performs steadily well if SNR is higher than 5dB. On the other hand, they mismatch for the LVET, suggesting that the LVET effect is significantly misrepresented. Investigating and alleviating this misspecification with the appropriate modifications to the model might be crucial to successfully transferring findings from in-silico to in-vivo. This iterative process of **1.** model analysis, **2.** real-world experimentation, **3.** comparison with observations, and **4.** model refinement; exemplifies the scientific method. Our results demonstrate that SBI facilitates more scrutiny in applying the scientific loop to numerical CV models, in extracting scientific hypotheses from the model (step **1.**); and comparing theoretical predictions and real-world data (step **3.**).
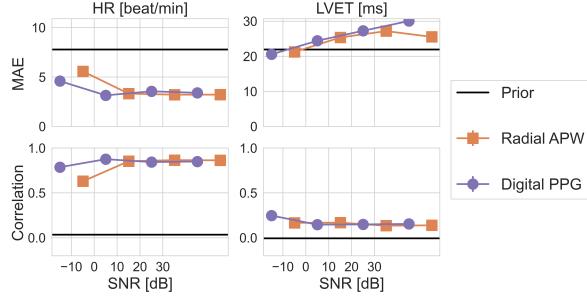

Figure 4: Mean absolute error (MAE) and correlation between the labels and point estimates extracted from the posterior distributions trained for different SNR values. The LVET's performance is compared to the predictions of a prior distribution conditioned on age and HR. *The features predicting HR generalizes better than the one for LVET. The HR MAE decreases with decreasing SNR, indicating the posterior gains robustness to misspecification with decreasing SNR. The features extracted for LVET do not generalize to real-world data but seems to inform more than only age and HR as the posterior's correlation is higher than the prior one.*

4 Conclusion

We have introduced a simulation-based inference methodology to analyze complex cardiovascular simulators. Our results show that our simulation-based inference method yields additional insights about 1D hemodynamics models, beyond the standard sensitivity analyses. This is done by considering the complete posterior distribution, which provides a consistent and multi-dimensional quantification of uncertainty for individual measurements. This uncertainty representation enables us to recognize ambiguous inverse solutions, study the heterogeneity of sensitivity in the population considered, and understand dependencies between biomarkers in the inverse problem. Supported by results on real-world data, we have illustrated and discussed the challenge of model misspecification

in scientific inquiry. In summary, simulation-based inference enables scientists to address inverse problems in CV models, accounting for complex forward model dynamics and individualized uncertainty. This extended abstract provides foundations for a more effective use of CV simulations for scientific inquiry and personalized medicine.

References

[1] J. Alastruey, P. H. Charlton, V. Bikia, B. Paliakaitė, B. Hametner, R. M. Bruno, M. P. Mulder, S. Vennin, S. Piskin, A. W. Khir, et al. Arterial pulse wave modelling and analysis for vascular age studies: a review from vascagenet. *American Journal of Physiology-Heart and Circulatory Physiology*, 2023.

[2] J. Alastruey, K. H. Parker, and S. J. Sherwin. Arterial pulse wave haemodynamics. In *11th international conference on pressure surges*, pages 401–443. Virtual PiE Led t/a BHR Group, 2012.

[3] A. S. Alhakak, J. R. Teerlink, J. Lindenfeld, M. Böhm, G. M. Rosano, and T. Biering-Sørensen. The significance of left ventricular ejection time in heart failure with reduced ejection fraction. *European Journal of Heart Failure*, 23(4):541–551, 2021.

[4] M. D. Altschule, D. R. Gilligan, et al. The effects on the cardiovascular system of fluids administered intravenously in man. ii. the dynamics of the circulation. *The Journal of Clinical Investigation*, 17(4):401–411, 1938.

[5] G. Avecilla, J. N. Chuong, F. Li, G. Sherlock, D. Gresham, and Y. Ram. Neural networks enable efficient and accurate simulation-based inference of evolutionary parameters from adaptation dynamics. *PLoS biology*, 20(5):e3001633, 2022.

[6] B. Baillargeon, N. Rebelo, D. D. Fox, R. L. Taylor, and E. Kuhl. The living heart project: a robust and integrative simulator for human heart function. *European Journal of Mechanics-A/Solids*, 48:38–47, 2014.

[7] V. Bikia, M. Lazaroska, D. Scherrer Ma, M. Zhao, G. Rovas, S. Pagoulatou, and N. Stergiopulos. Estimation of left ventricular end-systolic elastance from brachial pressure waveform via deep learning. *Frontiers in Bioengineering and Biotechnology*, 9:754003, 2021.

[8] J. Bonnemain, M. Zeller, L. Pegolotti, S. Deparis, and L. Liaudet. Deep neural network to accurately predict left ventricular systolic function under mechanical assistance. *Frontiers in Cardiovascular Medicine*, 8:752088, 2021.

[9] G. E. Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.

[10] J. Brehmer. Simulation-based inference in particle physics. *Nature Reviews Physics*, 3(5):305–305, 2021.

[11] N. K. Chakshu, I. Sazonov, and P. Nithiarasu. Towards enabling a cardiovascular digital twin for human systemic circulation using inverse analysis. *Biomechanics and modeling in mechanobiology*, 20(2):449–465, 2021.

[12] P. H. Charlton, J. Mariscal Harana, S. Vennin, Y. Li, P. Chowienczyk, and J. Alastruey. Modeling arterial pulse waves in healthy aging: a database for in silico evaluation of hemodynamics and pulse wave indexes. *American Journal of Physiology-Heart and Circulatory Physiology*, 317(5):H1062–H1085, 2019.

[13] P. H. Charlton, B. Paliakaitė, K. Pilt, M. Bachler, S. Zanelli, D. Kulin, J. Allen, M. Hallab, E. Bianchini, C. C. Mayer, et al. Assessing hemodynamics from the photoplethysmogram to gain insights into vascular age: A review from vascagenet. *American Journal of Physiology-Heart and Circulatory Physiology*, 322(4):H493–H522, 2022.

[14] G. Cotter, Y. Moshkovitz, E. Kaluski, O. Milo, Y. Nobikov, A. Schneeweiss, R. Krakover, and Z. Vered. The role of cardiac power and systemic vascular resistance in the pathophysiology and diagnosis of patients with acute congestive heart failure. *European journal of heart failure*, 5(4):443–451, 2003.

[15] K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

[16] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf. Real-time gravitational wave science with neural posterior estimation. *Physical review letters*, 127(24):241103, 2021.

[17] P. Dehkordi, F. Khosrow-Khavar, M. Di Rienzo, O. T. Inan, S. E. Schmidt, A. P. Blaber, K. Sørensen, J. J. Struijk, V. Zakeri, P. Lombardi, et al. Comparison of different methods for estimating cardiac timings: a comprehensive multimodal echocardiography investigation. *Frontiers in physiology*, 10:1057, 2019.

[18] A. Delaunoy, J. Hermans, F. Rozet, A. Wehenkel, and G. Louppe. Towards reliable simulation-based inference with balanced neural ratio estimation. In *Advances in Neural Information Processing Systems 2022*.

[19] A. Delaunoy, A. Wehenkel, T. Hinderer, S. Nissanke, C. Weniger, A. Williamson, and G. Louppe. Lightning-fast gravitational wave parameter inference through neural amortization. In *Machine Learning and the Physical Sciences. Workshop at the 34th Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

[20] J. Gao, Y. Hua, G. Hu, C. Wang, and N. M. Robertson. Reducing distributional uncertainty by mutual information maximisation and transferable feature learning. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, pages 587–605. Springer, 2020.

[21] M. Glöckler, M. Deistler, and J. H. Macke. Variational methods for simulation-based inference. In *International Conference on Learning Representations 2022*.

[22] J. A. Hartigan and P. M. Hartigan. The dip test of unimodality. *The annals of Statistics*, pages 70–84, 1985.

[23] M. Hashemi, A. N. Vattikonda, J. Jha, V. Sip, M. M. Woodman, F. Bartolomei, and V. K. Jirsa. Simulation-based inference for whole-brain network modeling of epilepsy using deep neural density estimators. *medRxiv*, pages 2022–06, 2022.

[24] L. Hespanhol, C. S. Vallio, L. M. Costa, and B. T. Saragiotto. Understanding and interpreting confidence and credible intervals around effect estimates. *Brazilian journal of physical therapy*, 23(4):290–301, 2019.

[25] E. Ipar, N. A. Aguirre, L. J. Cymberknop, and R. L. Armentano. Blood pressure morphology as a fingerprint of cardiovascular health: A machine learning based approach. In *Applied Informatics: Fourth International Conference, ICAI 2021, Buenos Aires, Argentina, October 28–30, 2021, Proceedings 4*, pages 253–265. Springer, 2021.

[26] W. Jin, P. Chowienczyk, and J. Alastruey. Estimating pulse wave velocity from the radial pressure wave using machine learning algorithms. *Plos one*, 16(6):e0245026, 2021.

[27] L. John. Forward electrical transmission line model of the human arterial system. *Medical and Biological Engineering and Computing*, 42:312–321, 2004.

[28] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[29] W. B. Kannel, C. Kannel, R. S. Paffenbarger Jr, and L. A. Cupples. Heart rate and cardiovascular mortality: the framingham study. *American heart journal*, 113(6):1489–1494, 1987.

[30] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[31] I. Kobyzev, S. J. Prince, and M. A. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):3964–3979, 2020.

[32] S. B. Lasrado, R. Vinoth, and S. Balaji. In silico blood flow simulation using finite volume modeling in a generic 3d aneurysm model. In *2022 6th International Conference on Devices, Circuits and Systems (ICDCS)*, pages 210–214. IEEE, 2022.

[33] J. Linhart, P. L. C. Rodrigues, T. Moreau, G. Louppe, and A. Gramfort. Neural posterior estimation of hierarchical models in neuroscience. In *GRETSI 2022-XXVIIIème Colloque Francophone de Traitement du Signal et des Images*, 2022.

[34] J.-M. Lückmann. *Simulation-Based Inference for Neuroscience and Beyond*. PhD thesis, Universität Tübingen, 2022.

[35] J.-M. Lueckmann, J. Boelts, D. Greenberg, P. Goncalves, and J. Macke. Benchmarking simulation-based inference. In *International Conference on Artificial Intelligence and Statistics*, pages 343–351. PMLR, 2021.

[36] J.-M. Lueckmann, P. J. Goncalves, G. Bassetto, K. Öcal, M. Nonnenmacher, and J. H. Macke. Flexible statistical inference for mechanistic models of neural dynamics. *Advances in neural information processing systems*, 30, 2017.

[37] D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. 2003.

[38] N. Marlier, O. Brüls, and G. Louppe. Simulation-based bayesian inference for robotic grasping. *arXiv preprint arXiv:2303.05873*, 2023.

[39] A. Melis. *Gaussian process emulators for 1D vascular models*. PhD thesis, University of Sheffield, 2017.

[40] A. Melis. Gaussian process emulators for 1d vascular models. 2017.

[41] A. Melis, R. H. Clayton, and A. Marzo. Bayesian sensitivity analysis of a 1d vascular model with gaussian process emulators. *International Journal for Numerical Methods in Biomedical Engineering*, 33(12):e2882, 2017.

[42] B. Moody, G. Moody, M. Villarroel, G. Clifford, and I. Silva III. Mimic-iii waveform database (version 1.0). *PhysioNet*, 3, 2020.

[43] G. Papamakarios and I. Murray. Fast ε-free inference of simulation models with bayesian conditional density estimation. *Advances in neural information processing systems*, 29, 2016.

[44] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *The Journal of Machine Learning Research*, 22(1):2617–2680, 2021.

[45] G. Papamakarios, T. Pavlakou, and I. Murray. Masked autoregressive flow for density estimation. *Advances in neural information processing systems*, 30, 2017.

[46] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[47] A. S. Patel, R. H. Mackey, R. P. Wildman, T. Thompson, K. Matthews, L. Kuller, and K. Sutton-Tyrrell. Cardiovascular risk factors associated with enlarged diameter of the abdominal aortic and iliac arteries in healthy women. *Atherosclerosis*, 178(2):311–317, 2005.

[48] S. W. Patterson, H. Piper, and E. Starling. The regulation of the heart beat. *The Journal of physiology*, 48(6):465, 1914.

[49] L. Pegolotti, M. R. Pfaller, N. L. Rubio, K. Ding, R. B. Brufau, E. Darve, and A. L. Marsden. Learning reduced-order models for cardiovascular simulations with graph neural networks. *arXiv preprint arXiv:2303.07310*, 2023.

[50] F. Piccioli, A. Valiani, J. Alastruey, and V. Caleffi. The effect of cardiac properties on arterial pulse waves: An in-silico study. *International Journal for Numerical Methods in Biomedical Engineering*, 38(12):e3658, 2022.

[51] D. Rezende and S. Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

[52] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical science*, 4(4):409–423, 1989.

[53] F. Schäfer, J. Sturdy, M. Mesek, A. Sinek, R. Białecki, Z. Ostrowski, B. Melka, M. Nowak, and L. R. Hellevik. Uncertainty quantification and sensitivity analysis during the development and validation of numerical artery models. *Scandinavian Simulation Society*, pages 256–262, 2022.

[54] Y. Shi, P. Lawford, and R. Hose. Review of zero-d and 1-d models of blood flow in the cardiovascular system. *Biomedical engineering online*, 10:1–38, 2011.

[55] K. Sutton-Tyrrell, S. S. Najjar, R. M. Boudreau, L. Venkitachalam, V. Kupelian, E. M. Simonsick, R. Havlik, E. G. Lakatta, H. Spurgeon, S. Kritchevsky, et al. Elevated aortic pulse wave velocity, a marker of arterial stiffness, predicts cardiovascular events in well-functioning older adults. *Circulation*, 111(25):3384–3390, 2005.

[56] E. G. Tabak and C. V. Turner. A family of nonparametric density estimation algorithms. *Communications on Pure and Applied Mathematics*, 66(2):145–164, 2013.

[57] E. G. Tabak and E. Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences*, 8(1):217–233, 2010.

[58] A. Tejero-Cantero, J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. J. Gonçalves, D. S. Greenberg, and J. H. Macke. sbi: A toolkit for simulation-based inference. *The Journal of Open Source Software*, 5(52), 2020.

[59] N. Tolley, P. L. Rodrigues, A. Gramfort, and S. R. Jones. Methods and considerations for estimating parameters in biophysically detailed neural models with simulation based inference. *bioRxiv*, pages 2023–04, 2023.

[60] A. Updegrove, N. M. Wilson, J. Merkow, H. Lan, A. L. Marsden, and S. C. Shadden. Simvascular: an open source pipeline for cardiovascular simulation. *Annals of biomedical engineering*, 45:525–541, 2017.

[61] A. Verine, B. Negrevergne, Y. Chevaleyre, and F. Rossi. On the expressivity of bi-lipschitz normalizing flows. In *Asian Conference on Machine Learning*, pages 1054–1069. PMLR, 2023.

[62] S. Wagner-Carena, J. Aalbers, S. Birrer, E. O. Nadler, E. Darragh-Ford, P. J. Marshall, and R. H. Wechsler. From images to dark matter: End-to-end inference of substructure from hundreds of strong gravitational lenses. *The Astrophysical Journal*, 942(2):75, 2023.

[63] W. Wang, P. Mohseni, K. L. Kilgore, and L. Najafizadeh. Pulsedb: A large, cleaned dataset for benchmarking cuff-less blood pressure estimation methods. *Frontiers in Digital Health*, 4:277, 2023.

[64] A. Wehenkel and G. Louppe. You say normalizing flows i see bayesian networks. In *ICML2020 Workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models*, 2020.

[65] A. Wehenkel and G. Louppe. Graphical normalizing flows. In *International Conference on Artificial Intelligence and Statistics*, pages 37–45. PMLR, 2021.

[66] H. White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the econometric society*, pages 1–25, 1982.

[67] N. Xiao, J. Alastruey, and C. Alberto Figueroa. A systematic comparison between 1-d and 3-d hemodynamics in compliant arterial models. *International journal for numerical methods in biomedical engineering*, 30(2):204–231, 2014.

A   Materials & Methods

A.1   1D hemodynamics of the human arterial network

The full-body arterial model introduced in [2], on which [12] relies, describes the arterial pulse wave propagation into 116 arterial segments, making up the largest arteries of the thorax, limbs, and head. This model is a good compromise between faithfulness to the real-world system and complexity [1]. It enables forward simulating APWs and PPGs at multiple locations, given a set of physiological parameters describing the geometrical and physical properties of the cardiovascular system. Running a simulation takes a few minutes on any standard CPU [41], allowing [12] to release a dataset of 4374 simulated healthy individuals aged 25 to 75.

Compared to 3D and 0D models, 1D models offer a better balance between expressivity and efficiency. While 1D simulations may be less accurate than 3D models (e.g., they cannot model atherosclerosis as they do not consider wall shear stress), they trade a modest and well-studied decrease in accuracy against much lighter simulation costs [67, 1]. Furthermore, the tractable parameterization and efficient simulation of 3D whole-body hemodynamics remain two open research questions [49, 32]. On the other side of the CV modeling spectrum, 0D simulations [27, 54] rely on a lumped-element model to describe the relationship between blood flow at one location (e.g., left ventricle outflow) and blood pressure and flow at other locations. In addition to ignoring significant physical effects such as wave propagation and reflection, 0D models are partially parameterized by non-physiological quantities. Generating a representative population, such as the one considered in [12], can thus be challenging with these models.

Model description.   In [2], the authors consider the compartmentalized arterial model made of the following sub-models: **1.** the heart function; **2.** the arterial system; **3.** the geometry of arterial segments; **4.** the blood flow; and **5.** the vascular beds. *The heart function* describes the blood volume along time at the aorta as a five-parameter function. *The arterial system* is described as a graph, the heart is the parent root, and then arteries branch out into the body. Every branch of the network represents an arterial segment. Segments are coupled so that the conservation of mass and momentum hold in the complete system. Additionally, the heart function defines the boundary condition on the parent root of the arterial network. The vascular bed describes the boundary condition on the leaf nodes. *The geometry of arterial segments* assumes the segments are axial-symmetric and tapered tubes. Hence, the geometry of each arterial segment can be described using 1D parameters such as radius and thickness of the arterial wall. *The blood flow* in the 1D segments follows fluid dynamics, which depends on the geometry and visco-elastic properties of the arterial wall. *The vascular beds* are modeled using 0D approximations, i.e., the geometrical description is being lumped into a space-independent parametric transfer function.

The main state parameters of whole-body 1D hemodynamics models are the volumetric flow rate $Q(z, t)$, the blood pressure $P(z, t)$, and the vessel cross-sectional area $A(z, t)$ at axial position z and time t, in each artery considered. Based on the conservation of mass and momentum, one can derive the partial differential equations (PDEs)

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial z} = 0 \tag{1}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial z}\left(\alpha \frac{Q^2}{A}\right) + \frac{A}{\rho}\frac{\partial P}{\partial z} = -2\frac{\mu}{\rho}(\gamma_\nu + 2)\frac{Q}{A}, \tag{2}$$

where α is the Coriolis' coefficient, μ is the blood dynamic viscosity, and γ_ν is a parameter defining the shape of the radial velocity profile. A third relationship of the arterial wall mechanics relates pressure and cross-section area as

$$P(A) = P_{ext} + \beta\left(\sqrt{A} - \sqrt{A_0}\right) + \frac{\Gamma}{\sqrt{A}}\frac{\partial A}{\partial t}, \tag{3}$$

$$\text{where } \beta = \frac{4}{3}\frac{\sqrt{\pi}Eh_0}{A_0} \text{ and } \Gamma = \frac{2}{3}\frac{\sqrt{\pi}\varphi h_0}{A_0} \tag{4}$$

respectively denote the elastic and viscous components of the Voigt-type visco-elastic tube law, P_{ext} is the reference pressure at which the geometry is described by the cross-sectional area A_0 and thickness of the arterial wall h_0. The elastic modulus E and wall viscosity φ characterize the

mechanical properties of the wall. In addition to these PDEs, boundary conditions are formulated by coupling each artery segment with the parents and children in the arterial network. For further details, see [40, 12, 2].

The considered 1D hemodynamics model constitutes a complex simulator with many parameters. As described in Section 2, only a subset of these parameters are of direct interest. Other parameters are considered nuisance effects. In addition, we consider a measurement model that generates biosignals similar to the one in MIMIC-III. Appendix A.6 provides additional details on the parameters distributions and the measurement model considered.

A.2   Simulation-based inference

Neural Posterior Estimation (NPE).   As mentioned in Section 2, NPE [43, 36] is a Bayesian and amortized SBI algorithm. It trains a parametric conditional density estimator for the parameters of interest $p_\omega(\phi \mid \mathbf{x})$ on a dataset $\mathcal{D} := \{(\phi_i, \mathbf{x}_i)\}_{i=1}^N$ of samples from the joint distribution $p(\phi, \mathbf{x}) = \int p(\phi, \psi) p(\mathbf{x} \mid \phi, \psi) d\psi$. In this work, we rely on a rich class of neural density estimators called normalizing flows [NF, 57, 56, 51, 31, 44], from which both density evaluation and sampling is possible.

Given an expressive class of neural density estimators $\{p_\omega(\phi \mid \mathbf{x}) : \omega \in \Omega\}$, NPE aims to learn an amortized posterior distribution $p_{\omega^\star}(\phi \mid \mathbf{x})$ that works well for all possible observation $\mathbf{x} \in \mathcal{X}$, by solving

$$\omega^\star \in \arg\min_{\omega \in \Omega} \mathbb{E}_\mathbf{x} \left[\mathbb{KL} \left[p(\phi \mid \mathbf{x}) \parallel p_\omega(\phi \mid \mathbf{x}) \right] \right] \tag{5}$$

$$\Longleftrightarrow \omega^\star \in \arg\min_{\omega \in \Omega} \int p(\mathbf{x}) p(\phi \mid \mathbf{x}) \left[\log \frac{p(\phi \mid \mathbf{x})}{p_\omega(\phi \mid \mathbf{x})} \right] d\mathbf{x} d\phi \tag{6}$$

$$\Longleftrightarrow \omega^\star \in \arg\max_{\omega \in \Omega} \int p(\mathbf{x}) p(\phi \mid \mathbf{x}) \log p_\omega(\phi \mid \mathbf{x}) d\mathbf{x} d\phi \tag{7}$$

$$\Longleftrightarrow \omega^\star \in \arg\max_{\omega \in \Omega} \mathbb{E}_{(\phi, \mathbf{x})} \left[\log p_\omega(\phi \mid \mathbf{x}) \right]. \tag{8}$$

In practice, NPE approximates the expectation in (8) with an empirical average over the training set $\mathcal{D}$ and relies on stochastic gradient descent to solve the corresponding optimization problem. Assuming $\phi \in \mathbb{R}^k$ and unpacking the evaluation of the NF-based conditional density estimator, the training loss is

$$\ell(\mathcal{D}, \omega) = \frac{1}{N} \sum_{i=1}^N \log p_z \left(f_\omega(\phi_i; \mathbf{x}_i) \right) + \log |J_{f_\omega}(\phi_i; \mathbf{x}_i)|, \tag{9}$$

following from the change-of-variables theorem [56]. The symbol p_z denotes the density function of an arbitrary k-dimensional distribution (e.g., an isotropic Gaussian), $f_\omega : \mathbb{R}^k \times \mathcal{X} \to \mathbb{R}^k$ denotes a continuous function invertible for its first argument ϕ, parameterized by a neural network, and $|J_{f_\omega}|$ denotes the absolute value of the Jacobian's determinant of f_ω with respect to its first argument. In addition to density evaluation, as in (9), the NF enables sampling from the modeled distribution by inverting the function f_ω.

In our experiments, we combine a convolutional neural network encoding the observations $\mathbf{x}$ with a three-step autoregressive affine NF [45] which offers a good balance between expressivity and sampling efficiency as demonstrated in [64]. These models have an inductive bias towards simple density functions [61], which support that the multi-modality and diversity of posterior distributions observed in the population is not an artifact of our analysis but follows from the 1D cardiovascular model and prior considered. We provide additional details on the parameterization of f and the sampling algorithm in Appendix A.7.

Uncertainty analysis with SBI.   Uncertainty analysis [52, 24] regards identifiability as a continuous attribute of a model which allows ranking models by how much information the modeled observation process carries about the parameter of interest. We move away from the classical notion of statistical identifiability – convergence in probability of the maximum likelihood estimator to the actual parameter value – because this binary notion is not always relevant in practice and mainly applies to studies in the large sample size regime. In contrast, uncertainty analysis directly relates to the

mutual information between the parameter of interest and the observation as expressed by the model considered. It captures that biased or noisy estimators are informative and may suffice for downstream tasks.

As is standard in Bayesian uncertainty analyses, we look at credible regions $\Phi_\alpha(\mathbf{x})$ at different levels α, which are directly extracted from the posterior distribution $p(\phi \mid \mathbf{x})$. Formally, a credible region is a subset, Φ_α, of the parameter space Φ over which the conditional density integrates to α, i.e., $\Phi_\alpha : \int_{\phi \in \Phi_\alpha(\mathbf{x})} p(\phi \mid \mathbf{x}) \mathrm{d}\phi = \alpha, \Phi_\alpha \subseteq \Phi$. In this paper, we consider the smallest covering union of regions, denoted by Φ_α, which is always unique in our case and in most practical settings.

Size of credible intervals (SCI). We rely on the SCI to shed light on the uncertainty of a parameter given a measurement process. The SCI at a level α is the expected size of the credible region at this level: $\mathbb{E}_\mathbf{x}[\|\tilde{\Phi}_\alpha(\mathbf{x})\|]$, where $\|\cdot\|$ measures the size of a subset of the parameter space. In practice, we split the parameter space into evenly sized cells and count the number of cells belonging to the credible interval, as detailed in Appendix A.5.2. SCI is easier to interpret for domain experts than mutual information, as the former is expressed in the parameter's units. In addition, SCI is robust to multi-modality in contrast to point-estimator-based metrics (e.g., mean squared/absolute error) that cannot discriminate between two posterior distributions if they lead to the same point estimate.

Calibration. Given samples from the joint distribution $p(\phi, \mathbf{x})$, credible intervals are expected to contain the true value of the parameter at a frequency equal to the credibility level α, that is, $\mathbb{E}_{p(\phi,\mathbf{x})}[\mathbb{1}_{\Phi_\alpha}(\phi)] = \alpha$, where $\mathbb{1}$ is the indicator function. In this work, we do not have access to the true posterior but a surrogate $\tilde{p}$ of it. Hence, the coverage property of credible regions, which support the interpretation of uncertainty, may be violated, even when the forward model and the prior accurately describe the data. The calibration $C(\tilde{p}(\phi \mid \mathbf{x}), \mathcal{D})$ of a surrogate posterior $\tilde{p}$ is a metric, computed on a set $\mathcal{D} := \{(\phi_j^\star, \mathbf{x_j})\}_{j=1}^N$, that measures whether the surrogate's credible regions respect coverage. We compute calibration as

$$C(\tilde{p}(\phi \mid \mathbf{x}), \mathcal{D}) = \frac{1}{k} \sum_{i=1}^k \left| \frac{i}{k} - \frac{1}{N} \sum_{j=1}^N \mathbb{1}_{\tilde{\Phi}_{\frac{i}{k}}} (\phi_j^\star(\mathbf{x}_j)) \right|,$$

where $\tilde{\Phi}_{\frac{i}{k}}(\mathbf{x}_j)$ is the credible region at level $\alpha = \frac{i}{k}$ corresponding to the surrogate posterior distribution $\tilde{p}(\phi \mid \mathbf{x}_j)$. The calibration directly relates to how much the surrogate posterior model violates the coverage property over all possible levels $\alpha \in\ ]0, 1]$.

A.3 In-vivo vs in-silico data

In this section we provide an overview of the generation of in-silico data, in Figure 5, and a few examples of the real-world data considered from MIMIC, in Figure 6. We observe that real-world data contains degenerated beats. Moreover, another source of variation between beats comes from physiological parameter dynamics, which can vary from one beat to another. These observations motivate the introduction of noise on top of the deterministic simulation as discussed in the main paper and shown in Figure 5.
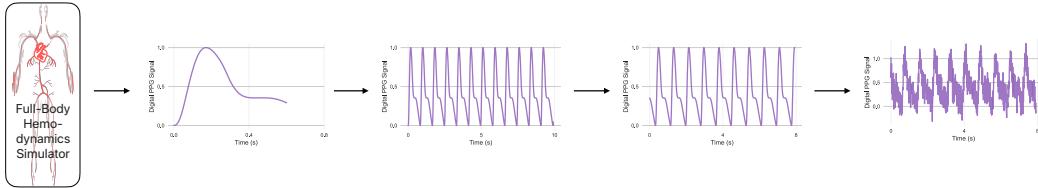

Figure 5: Generation of a digital PPG observation in-silico. From left to right: a PPG signal is extracted from the 1D hemodynamics simulator, the same wave is concatenated to reach a length of 10 seconds, the 10-second segment is cropped randomly by two seconds, additive Gaussian noise is added (SNR $\approx$ 11dB).
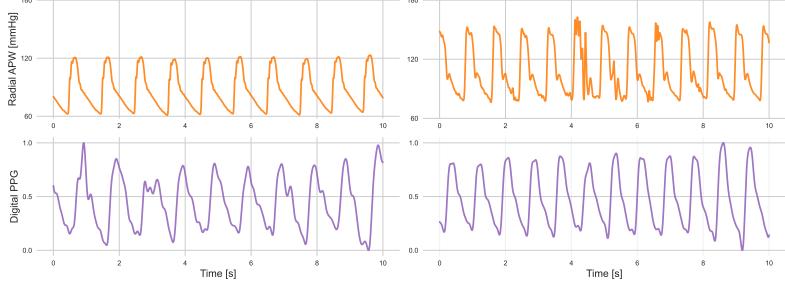
Figure 6: Waveforms reproduced from the MIMIC-III waveform database [42] subset of the PulseDB dataset [63].

A.4 Supplementary results

In this section we provide additional perspectives on the learned surrogate models of the posterior distributions.

A.4.1 Calibration and MAE

Figure 7 presents the mean average precision of a point estimator obtained as the mean of the posterior distribution and the calibration of these posterior distributions. Most surrogate models trained with NPE are well calibrated. However, there remains a risk that a surrogate model is not well calibrated, such as observed for the Digital PPG for inferring some of the parameters for low levels of noise. The MAEs have a similar behavior as the average sizes of credible intervals at size 68% and 95% discussed in the main materials.
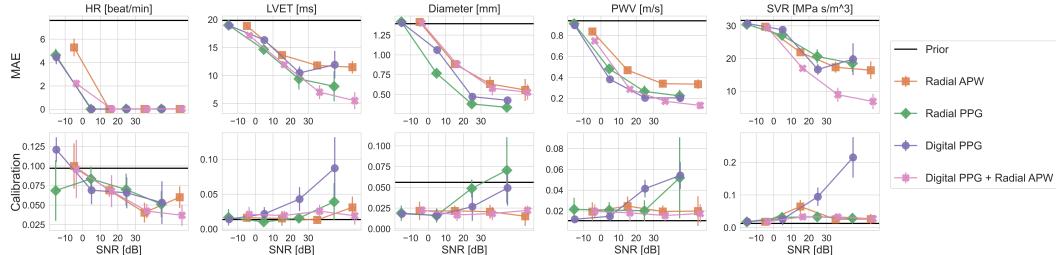

Figure 7: Mean absolute error (MAE) of the expected value of the posterior distributions and calibration of the credible intervals. Except for the Digital PPG, at high level of noise, models are well calibrated. The MAEs follow the trend expected from the analysis of the size of credible intervals in the main materials.

A.4.2 Posterior distributions

One desirable consequence of relying on SBI to analyze hemodynamics models is to provide access to the joint conditional distribution of parameters given an observation. In Figure 8, we show the posterior distributions corresponding to two randomly selected simulated digital PPGs of the test set. These plots reveal how different measurements can lead to very different posterior distributions and highlight the relevance of considering sub-groups rather than the entire population at once. In addition, Figure 9 presents the posterior distributions corresponding to the different measurement types studied. From this figure we observe that different measurements carry different information about the parameters and thus can lead to very different posterior distributions. Such plots may also indicate when multiple measurements should be done and when this is likely useless.

A.5 Metrics

In this section, we provide the algorithms used to compute the calibration, in Algorithm 1, and size of credible intervals from samples of the surrogate posterior distributions, in Algorithm 2.
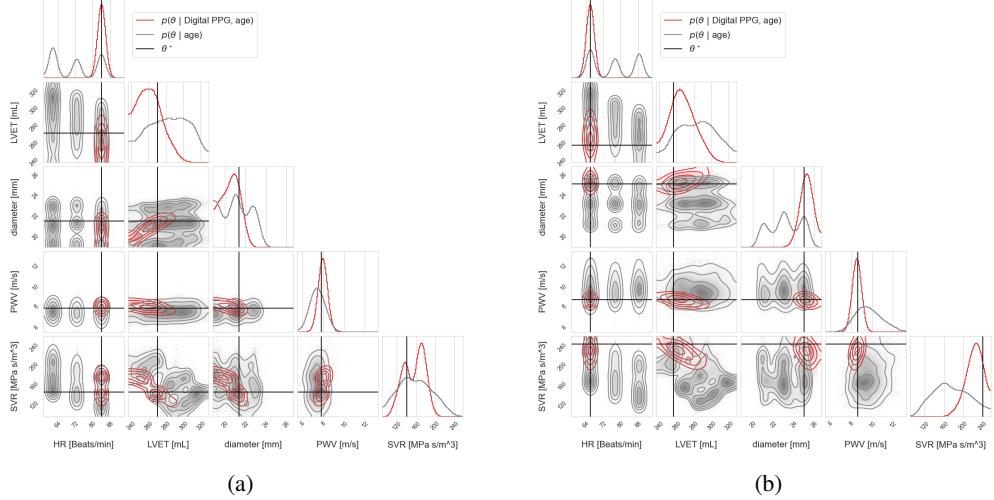
13

Figure 8: Posterior distributions obtained for two different test observation of digital PPGs.
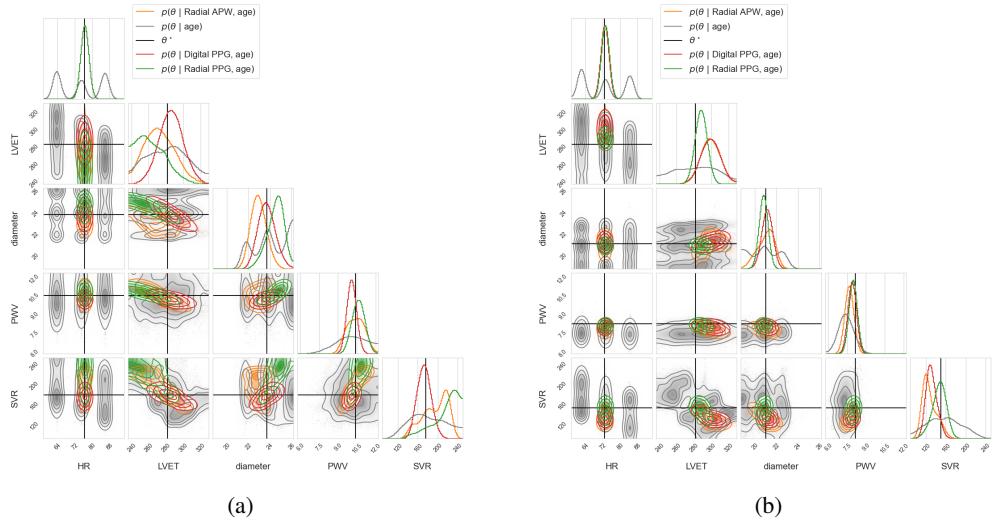

Figure 9: Comparison between the posterior distributions corresponding to different measurements.

A.5.1   Calibration

Algorithm 1 returns the distribution of minimum credibility levels required to not reject the true value of ϕ. Under calibration, these values should be uniformly distributed – we expect to reject falsely the true value with a frequency equal to the credibility level chosen. We report the integral (along credibility levels α) between the observed cumulative distribution function (CDF) of minimum credibility levels and the CDF of a uniform distribution. This metric equals 0 under perfect calibration and is bounded by 0.5. We report the calibration for each dimension independently as the metric does not generalize to multiple dimension.

A.5.2   Size of credible intervals

Algorithm 2 describes a procedure to compute the size of credible intervals. In our experiments, we consider each dimension independently and discretize the space of value in 100 cells. We finally report the average number of cells multiplied by the size of one cell in the physical unit of the parameter.

Algorithm 1 Statistical Calibration of Posterior Distribution

---

Input: Dataset of pairs $\mathcal{D} = \{(\phi_i, x_i)\}$, Posterior distribution $p(\phi|x)$, Number of samples N.
Output: Distribution of minimum credibility levels.

1: Initialize an empty list $CredLevels$
2: **for** $(\phi_i, x_i) \in \mathcal{D}$ **do**
3:     Initialize an empty list $Samples$
4:     **for** $i = 1$ to N **do**
5:         Sample ϕ_i from $p(\phi|x)$
6:         Append ϕ_i to $Samples$
7:     **end for**
8:     Sort $Samples$
9:     Compute the rank (position in ascending order) r of ϕ in $Samples$
10:     Set $CredLevel = \frac{r}{N}$
11:     Append $CredLevel$ to $CredLevels$
12: **end for**
Return: $CredLevels$.

Algorithm 2 Compute Average Size of Credible Intervals

Input: Dataset of observations x, Posterior distribution $p(\phi|x)$, Credibility level α **Output:** Average size of credible intervals

1: Initialize an empty list $CredIntSizes$
2: **for** each observation x in the dataset **do**
3:     Generate samples from the posterior distribution: $\phi_{\text{samples}} = \text{SampleFromPosterior}(p(\phi|x))$
4:     Discretize the parameter space into cells
5:     Initialize an empty list $CellCounts$
6:     **for** each sample ϕ in ϕ_{samples} **do**
7:         Increase by one the count the cell covering ϕ
8:     **end for**
9:     Sort the $CellCounts$ list in descending order
10:     Append the minimum number of cells required to reach the credible level α to $CredIntSizes$.
11: **end for**
12: Compute the average size of credible intervals by taking the mean of the $CredIntSizes$ list
Return: Average size of credible intervals

A.5.3 Mutual information and SCI

In our experiments, we gave the average size of credible (SCI) intervals rather than the mutual information, as the former quantity is expressed in physical units that have a direct interpretation to specialists. We now discuss how the SCI relates to mutual information.
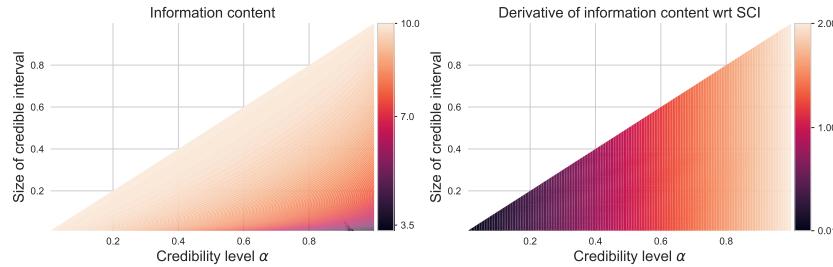

Figure 10: Relationship between the size of credible intervals and the information content present in the signal, for all credibility level. On the left: the plot of (11); on the right: the plot of the derivative of (11) with respect to the SCI. We observe that larger SCI corresponds to larger number of bits required to encode the true value of the parameter of interest given the observation.

We aim to drive this discussion in the context of comparing the quality of two distinct measurement processes for inferring one quantity of interest. Formally, we denote these two measurements by $\mathbf{x}_1 \in \mathcal{X}_1$ and $\mathbf{x}_2 \in \mathcal{X}_2$ and the quantity of interest by $\phi \in \Phi$.

Assuming a fixed marginal distribution $p(\phi)$ over the parameter, the two measurement processes $p(\mathbf{x}_1 \mid \phi)$ and $p(\mathbf{x}_2 \mid \phi)$ define two joint distributions $p(\mathbf{x}_1, \phi)$ and $p(\mathbf{x}_2, \phi)$. Considering a discretized space of parameters ϕ, the mutual information can be written as

$$\mathcal{I}(\phi, \mathbf{x}_i) = \mathcal{H}(\phi) - \mathcal{H}(\phi \mid \mathbf{x}_i), \tag{10}$$

where $\mathcal{I}(\phi, \mathbf{x}_i)$ is the mutual information and $\mathcal{H}$ the entropy. As the marginal entropy of the parameter remains constant, it is clear that only the second term matters for comparing the information content of the two measurement processes. The quantity $\mathcal{H}(\phi \mid \mathbf{x}_i)$ can be interpreted as the average remaining number of bits necessary to encode ϕ if we know $\mathbf{x}_i$. The average is taken over the joint distribution induced by the marginal distribution of ϕ and the measurement process $p(\mathbf{x}_i \mid \phi)$.

From this interpretation, choosing the one with the highest mutual information is a well-motivated criterion for choosing between two measurement processes. Said differently, we are looking for measurement processes with the smallest $\mathcal{H}(\phi \mid \mathbf{x}_i)$, the one leading to small uncertainty about the correct value of ϕ.

We use an information theory point of view to explain why, similarly to mutual information maximization [20], aiming for the measurement process with the smallest SCI is a sound approach. Let us consider the measurement process $p(\mathbf{x}_i \mid \phi)$ leading credible intervals with size $S(\alpha, \mathbf{x}_i)$ for a certain credibility level α and observation $\mathbf{x}_i$. Similarly to what we do in practice to compute the SCI, we discretize the space of parameters into N cells. The SCI is then defined as the minimal number of cells required to cover the credible region at level α. From the SCI, we can say that the true value of ϕ belongs to one of the $S(\alpha, \mathbf{x}_i)$ cells of the credible region with probability α or to one of the $N - S(\alpha, \mathbf{x}_i)$ remaining cells with a probability $1 - \alpha$. From this, we can bound the average number of bits required to encode the true value of ϕ given the observation $\mathbf{x}_i$ as a function of the SCI S and the credibility level α as

$$\text{N bits} \leq -\alpha \log_2 \frac{\alpha}{S(\alpha, \mathbf{x}_i)} - (1 - \alpha) \log_2 \frac{1 - \alpha}{N - S(\alpha, \mathbf{x}_i)}. \tag{11}$$

Figure 10 shows the relationship between this bound and the credibility level α and SCI. We treat SCI and the credibility α as independent quantities, as different measurement processes can lead to different relationships between these two quantities. We must notice that given a credibility level α, smaller SCI corresponds to better bounds. We can conclude that selecting models with the smallest SCI for a given credibility level is a sound approach with a similar interpretation as making this choice based on mutual information.

A.6 Whole-body hemodynamics model

A.6.1 Parameterization

We use a dataset of 4374 simulations from healthy individuals aged 25 to 75 [12]. By deriving APW and simulating PPG waveforms from the blood flow and pressure data, we obtain signals corresponding to a single heartbeat of varying lengths. The parameters of interest ϕ can be parameters of the forward model, such as HR, LVET, Diameter, or quantity that are derived from the simulation, such as PWV and SVR. In general, the parameters of interest and the observation do not constitute a one-to-one mapping. This is especially caused by the presence of additional parameters treated as nuisance ψ. In the dataset from [12], the following parameters vary from one simulation to the other:

- Heart function:
    - Heart Rate (HR).
    - Stroke Volume (SV).
    - Left Ventricular Ejection Time (LVET). *Note:* LVET changes as a deterministic function of HR and SV.

16

- Peak Flow Time (PFT).
  - Reflected Fraction Volume (RFV).
- Arterial properties:
  - $Eh = R_d(k_1 e^{k_2 R_d} + k_3)$ where $k1, k2$ are constant and $k3$ follows a deterministic function of age.
  - Length of proximal aorta
  - Diameter of larger arteries
- Vascular beds
  - Resistance adjusted to achieve mean average pressure (MAP) distribution compatible with real-world studies.
  - Compliance adjusted to achieve realistic peripheral vascular compliance (PVC) compatible with real-world studies.

The interested reader will find further details in [12].

A.6.2 Measurement model

The dataset from [12] is made of individual beats, which differs from real-world data usually made of a fixed-size segments. While pre-processing the real-world data to extract unique beat is feasible, it may pose challenges to ensure this extraction is consistent with the simulated waves. Instead, we add a measurement model that reduce the gap between the real-world and simulated data formats. We first generate segments longer than 10 seconds by concatenating the same beat multiple times. Then, we randomly crop time series into 8-second segments. This ensures that the posterior distributions are defined for 8-second segments and accounts for all possible starting positions within the heartbeat. Finally, we introduce a white Gaussian noise to the waveforms to make our analysis less sensitive to the model misspecification. Appendix A.3 showcases these steps and the resulting waveforms.

A.7 Normalizing flows

We provide additional details on the normalizing flows used to model the posterior distributions. In all our experiments, we apply the same training and model selection procedures. Moreover we use the same neural network architecture for all experiments.

We rely on the open-source libraries PyTorch [46] and Normalizing Flows, a lightweight library to build NFs built upon the abstraction of NFs as Bayesian networks from [65].

A.7.1 Training setup

We randomly divide the complete dataset into 70% train, 10% validation, and 20% test sets. We optimize the parameters of the neural networks with stochastic gradient descent on (9) with the Adam optimizer [30]. We use a batch size equal to 100, a fixed learning rate ($= 10^{-3}$), and a small weight decay ($= 10^{-6}$). We train each model for 500 epochs and evaluate the validation loss after each epoch. The best model based on the lowest validation loss was returned and used to obtain the results presented in the paper. All data are normalized based on their standard deviation and mean on the training set. For the time series, we compute one value across the time dimension.

A.7.2 Neural network architecture

We use the same neural network architecture for all the results reported. It is made of a 3-step autoregressive normalizing flow [45] combined with a convolutional neural network (CNN) to encode the 8-second segments sampled at $125Hz$ ($\in \mathbb{R}^{1000}$). The CNN is made of the following layers:

1. 1D Convolution with no padding, kernel size $= 3$, stride $= 2$, 40 channels, and followed by ReLU;
2. 1D Convolution with no padding, kernel size $= 3$, stride $= 2$, 40 channels, and followed by ReLU;
3. 1D Convolution with no padding, kernel size $= 3$, stride $= 2$, 40 channels, and followed by ReLU;

4. Max pooling with a kernel $= 3$;

5. 1D Convolution with no padding, kernel size $= 3$, stride $= 2$, 20 channels, and followed by ReLU;

6. 1D Convolution with no padding, kernel size $= 3$, stride $= 2$, 10 channels, and followed by ReLU,

leading to a 90 dimensional representation of the input time series. The 90-dimensional embedding is concatenated to the age and denoted $\mathbf{h}$. Then, $\mathbf{h}$ is passed to the NF as an additional input to the autoregressive conditioner [45, 65].

The NF is made of a first autoregressive step that inputs both the 91 conditioning vector $\mathbf{h}$ and the parameter vector and outputs 2 real values $\mu_i(\phi_{<i}, \mathbf{h}), \sigma_i(\phi_{<i}, \mathbf{h}) \in \mathbb{R}$ per parameter in an autoregressive fashion. Then the parameter vector is linearly transformed as $u_i = \phi_i e^{\sigma_i(\phi_{<i}, \mathbf{h})} + \mu_i(\phi_{<i}, \mathbf{h})$. The vector $\mathbf{u} := [u_1, \ldots, u_k]$ is then shuffled and passed through 2 other similar transformations, leading to a vector denoted $\mathbf{z}$, which eventually follows a Gaussian distribution after learning [44]. The 3 autoregressive networks have the same architecture: a simple masked multi-layer perceptron with ReLu activation functions and 3 hidden layers with 350 neurons each. We can easily compute the Jacobian determinant associated with such a sequence of autoregressive affine transformations on the vector ϕ and thus compute (9).

We can easily show that the Jacobian determinant is equal to the product of all scaling factors e^{σ_i}. We also directly see that ensuring these factors are strictly greater than 0 enforce a continuously invertible Jacobian for all value of ϕ and thus continuous bijectivity of the associated transformation.

As mentioned, under perfect training, the mapping from Φ to $\mathcal{Z}$ defines a continuous bijective transformation that transforms samples from $\phi \sim p(\phi \mid \mathbf{h})$ into samples $\mathbf{z} \sim \mathcal{N}(0, I)$. As the transformation is bijective, we can sample from $p(\phi \mid \mathbf{h})$ by inverting the transformation onto samples from $\mathcal{N}(0, I)$. As the transformation is autoregressive, we can invert it by doing the inversion sequentially for all dimensions as detailed in [44, 65, 45].