Approximation of Intractable Likelihood Functions in Systems Biology via Normalizing Flows

Vincent D. Zaballa
Department of Biomedical Engineering
University of California, Irvine
Irvine, CA 92697
vzaballa@uci.edu

Elliot E. Hui
Department of Biomedical Engineering
University of California, Irvine
Irvine, CA 92697
eehui@uci.edu

Abstract

Systems biology relies on mathematical models that often involve complex and intractable likelihood functions, posing challenges for efficient inference and model selection. Generative models, such as normalizing flows, have shown remarkable ability in approximating complex distributions in various domains. However, their application in systems biology for approximating intractable likelihood functions remains unexplored. Here, we elucidate a framework for leveraging normalizing flows to approximate complex likelihood functions inherent to systems biology models. By using normalizing flows in the Simulation-based inference setting, we demonstrate a method that not only approximates a likelihood function but also allows for model inference in the model selection setting. We showcase the effectiveness of this approach on real-world systems biology problems, providing practical guidance for implementation and highlighting its advantages over traditional computational methods.

1 Introduction

Systems biology aims to understand the complex interactions within biological systems by modeling them as integrated networks of genes, proteins, and biochemical reactions. Analogous to ensemble methods in machine learning, where individual models are integrated to improve predictive accuracy, systems biology aims to model complex biological systems as integrated wholes to provide deeper insights into their function and behavior. Leveraging computational approaches, including machine learning algorithms for network analysis and high-dimensional data interpretation, systems biology offers a paradigm shift from reductionist methods that examine individual components, such as proteins, in isolation.

A pressing challenge in systems biology is the problem of inference in models characterized by intractable likelihood functions. These complex models often encapsulate the high dimensionality and non-linearity innate to biological systems but come at the cost of computational intractability for standard statistical methods. Conventional approaches, such as maximum likelihood estimation via least squares regression, are generally ill-suited for these models, failing to provide robust parameter estimates that can be reliably used for further analysis or uncertainty quantification. This creates a bottleneck that limits our ability to draw meaningful insights from biological networks. Examples of biological networks with intractable likelihoods are those governed by mass action kinetics, including protein dynamics and metabolic systems [5].

Simulation-based inference (SBI) methods [3], particularly those employing normalizing flows, have begun to bridge the gap in tackling the challenging problem of approximating intractable likelihoods in systems biology. Normalizing flows transform a complex or unknown distribution into a simple, tractable one, such as a Gaussian, effectively learning the data's underlying distribution without

NeurIPS 2023 Generative AI and Biology Workshop.

requiring an explicit likelihood function. This offers a powerful tool for probabilistic modeling of complex systems, allowing for both robust inference via posterior predictive checks and uncertainty quantification for each model parameter, as well as model comparison.

A salient challenge within systems biology is the inference of network parameters pertaining to the Bone Morphogenetic Protein (BMP) signaling pathway [1]. This pathway assumes a pivotal role in both pathological and developmental contexts, being implicated in an array of malignancies, such as cancer, as well as critical developmental processes [21]. Given its dichotomous functions, understanding the intricacies of BMP and other signaling pathways becomes imperative for predicting their responses to external modulations, such as pharmacological perturbations.

Prior studies modeled the BMP pathway using mass action kinetics that qualitatively described observed experimental outputs [1]. However, traditional parameter fitting methods such as least square regression yield an expansive array of plausible parameter values without a probabilistic interpretation, thereby aggravating the issue of model uncertainty. Additionally, the BMP pathway can be captured by multiple competing models [23], introducing the need for effective model discrimination. Traditional criteria for model selection, such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), offer point estimates of model probabilities. While these methods are generally efficacious for models with tractable likelihoods, they encounter limitations when applied to models characterized by intractable likelihood functions [15]. Hence, there is a need for new computational techniques that can conduct reliable inference and identify the most probable fitting model in the context of observed data, circumventing the limitations imposed by intractable likelihoods.

We demonstrate the first application of SBI techniques that use normalizing flows to address the task of approximating likelihood functions for the BMP signaling pathway. Using experimental data [11], we address the challenges associated with inferring these likelihood functions and demonstrate how to infer likelihoods of high-dimensional observed data *and* high dimensional simulator parameters using statistical features of i.i.d. likelihoods to infer a data space $\mathcal{D}$ and parameter space $\mathcal{P}$, such that $\mathcal{D} \in \mathbb{R}^{940}, \mathcal{P} \in \mathbb{R}^m$ where $m \leq 70$. Using these approximated likelihoods, we not only scrutinize the uncertainty enveloping each model's parameters but also offer indications regarding which model is more plausible given the observed data. Our findings underscore the efficacy of deep generative models, such as normalizing flows, in the context of biological model parameter inference and model selection. Finally, these surrogate likelihoods hold the potential to be integrated into Bayesian Optimal Experimental Design (BOED) frameworks [6, 10], thereby minimizing both the temporal and experimental resources required to achieve a model with robust predictive accuracy in biology.

2 Background

The Intractable Likelihood of the BMP Model. The BMP signaling pathway is commonly described by mass action kinetics and conservation laws, capturing the dynamics of downstream genetic expression signals reaching a steady-state. Two principal models exist for describing this system: the "twostep" model which involves equations $A_i + L_j \leftrightarrow D_{ij}$ and $D_{ij} + B_k \leftrightarrow T_{ijk}$, and the "onestep" model represented by $A_i + B_k + L_j \leftrightarrow T_{ijk}$, where the binding affinity K represents the transition from one state to another. Additionally, i, j, k represent the ith type A receptor, the jth protein ligand complex, and k^{th} type B receptor, D represents a dimeric protein complex, and T represents a trimeric complex. Both models then describe the phosphorylation efficiency, ε, of a trimeric complex T sending a downstream gene expression signal, S, as $S = \varepsilon T$. While these models are originally formulated as ordinary differential equations (ODEs), they become differential algebraic equations under certain conditions such as fixed volume and large ligand concentration. These models can be solved via convex optimization, however, this results in an intractable likelihood functions. The lack of a likelihood function for the BMP pathway requires novel computational methods, such as SBI, for further refinement and understanding of models' parameters and probabilities given observed data, x_o.

Normalizing Flows. Normalizing flows constitute a specialized subset of invertible and differentiable neural networks engineered to represent a series of monotonic transformations. These transformations are designed to minimize the divergence between a base distribution, commonly Gaussian, denoted as $p_u(\mathbf{u})$, and the target data distribution, $p_x(\mathbf{x})$. This is formulated using the change of variables

formula and a composition of monotonic diffeomorphic functions, $\mathbf{f}_\phi$, which can be parameterized via neural networks. The formal transformation from the base distribution to the data distribution is:

$$p_x(\mathbf{x}) = p_u(f_\phi^{-1}(\mathbf{x})) \left| \det \frac{\partial \mathbf{f}_\phi^{-1}}{\partial \mathbf{x}} \right|. \tag{1}$$

A thorough review on the theoretical framework and practical application of normalizing flows can be found in [12] and [19]. In this work, we use normalizing flows to model the intractable likelihood $p_\phi(x|\theta)$, where ϕ are parameters of the normalizing flow and θ represent biological model parameters of interest, such as binding affinity, K, and phosphorylation efficiency, ε, such that $K, \varepsilon \in \theta$ and x is simulated data from the simulator that we want to get close to the observed data, x_o. Using Bayes Theorem we can then infer parameters' posterior distribution as $p(\theta|x_o) \propto p_\phi(x_o|\theta)p(\theta)$.

Simulation-based Inference. Concurrently with advancements in normalizing flow architectures, significant advancements have been made in SBI algorithms for sequential posterior approximation [8, 18, 14, 16, 4, 9]. These SBI methodologies estimate various elements, such as the posterior distribution, likelihood, and the ratio of posterior to prior probabilities, to gauge the posterior of a specific model given observed data. SBI techniques are valuable in disciplines with complicated simulations based on mathematical models like particle physics [2], where functions are readily simulated but challenging to evaluate analytically.

These sequential SBI methods take inspiration from importance sampling to refine a posterior, likelihood, or likelihood ratio given an observed sample of interest, x_o. There are a variety of sequential SBI techniques, each with their own benefits, and reviewed in [13]. We employ the Sequential Neural Likelihood (SNL) approach [20], which utilizes a normalizing flow to approximate the likelihood represented by a simulator and conditioned over multiple rounds on observed data. We chose to use likelihoods as the likelihood ratio can be used as a form of model comparison, which is another important desiderata of this paper.

Model selection. The Bayes Factor (BF) is another term for the ratio of model probabilities, expressed as $BF = \frac{p(\mathcal{M}_0)}{p(\mathcal{M}_1)}$. It serves as a criterion for selecting between models. A $BF > 10$ indicates strong support for $\mathcal{M}_0$, while a $BF < \frac{1}{10}$ reveals strong support for $\mathcal{M}_1$. The BF inherently favor simpler models due to the Bayesian version of Occam's razor. This, however, depends on unbiased computation of the model's marginal likelihood. For more details on assorted model selection methods, refer to [17].

To engage in model selection, we need an estimate of the marginal likelihood for each model to compute the BF. One approach to achieve this is by sampling from a likelihood using posterior samples $p(x|\mathcal{M}) = \int_\Theta p(x|\theta, \mathcal{M})p(\theta|\mathcal{M})d\theta \approx \frac{1}{N}\sum_{i=1}^{N} p(x|\theta_i, \mathcal{M})$. When each model's prior probabilities are equal, the BF reduces to a comparison of two model likelihoods such that $\frac{p(\mathcal{M}_1)}{p(\mathcal{M}_0)} = \frac{p(x|\mathcal{M}_1)}{p(x|\mathcal{M}_0)}$. We use this formulation in this paper given that both models evaluated are equally likely apriori.

3 Experimental Evaluation

We employed two separate normalizing flow architectures for training both the onestep and twostep BMP models, utilizing the SNL method within the larger SBI framework. Specifically, we used the neural spline flow model [4], wherein each model was constructed with five flow layers and four bins in each layer to approximate invertible polynomial functions. Each layer was conditioned on the model parameters, θ, as well as on static experimental design parameters previously employed, denoted as ξ, which represents the concentration of BMP protein ligand administered (L) and the cell type, each with its own respective pseudo-concentrations of Type A and B receptors. This facilitated the creation of an approximate likelihood $p_\phi(x|\theta|\xi)$ for incorporation into the SNL process. The training regimen extended over 9 rounds of SBI, during which we collected posterior parameters, gauged the median distance between simulated and observed data, and logged probabilities. In each round, 1,000 data points were simulated based on the posterior distribution from the previous round, starting with a uniform prior. Mini-batches comprised 100 data points each were used to train a normalizing flow on both the onestep and twostep models. All priors and subsequently simulated data were combined to form an aggregate training dataset for the normalizing flows.

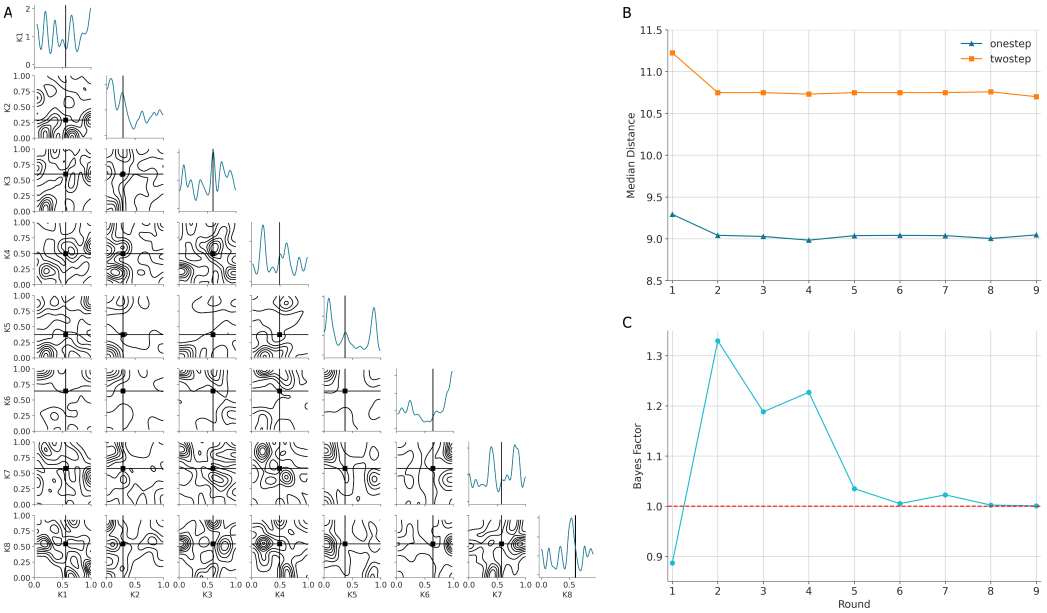

Figure 1: (A) Example of pairplots generated of the first 8 binding affinities, K, for the onestep model. Crosshairs indicate median values and occasionally align with a maximum a posteriori estimate of parameters' marginal distribution, but other times do not, reflecting uncertainty in the parameter values. (B) Median distance between x_o and simulated values $x_i \sim p_i(x|\theta_{i-1})$ from the simulator using priors drawn from round $i-1$. (C) Bayes Factor of the onestep to twostep model using each model's respective density estimator in round i and the final posterior returned after the last round. The dotted red line indicates the boundary between preference for the onestep model (>1) and the twostep model (<1). We see early indication that model 1 is slightly more likely, supported by the median distance, but diminishing probability with subsequent trained likelihoods, indicating ambiguity in model or overfitting of the likelihoods.

Concurrently, 5% of the data was reserved in each round as a validation set for early stopping, which took place if the validation error did not decrease for more than 20 epochs. After each round, the approximate likelihood was used to infer a new posterior using $p_i(\theta|x_o) \propto p_\phi(x_o|\theta)p(\theta)$, where $p_i(\theta|x_o)$ represents the posterior after round i using the trained likelihood from the previous round.

Given that the observed data point, x_o, was a 940-dimensional output, modeling of the joint likelihood is computationally impractical. To address this issue, we exploited the property that the joint distribution could be factorized into a product of individual likelihoods, $p(x_1, \ldots, x_{940}) = \prod_{i=1}^{940} p(x_i)$, thus simplifying the computational requirements for both training and the architecture.

Figure 1 shows (left) a subset example of pairplots for the binding affinity parameters from the onestep model, (top right) median distance from x_o, and (bottom right) BF calculation using normalizing flows from each round. The final round's posterior is qualitatively different than the flat prior used in the initial round. The median distance serves as a validation metric for the normalizing flow's approximation of the true likelihood. The BF plot seems to agree with the median distance in earlier rounds that the onestep model seems to be a slightly better model for the data, but the convergence to essentially no difference in the later rounds indicates either that it is still ambiguous as to which model better fits the data, or, that the normalizing flows may not be expressive enough to accurately represent each likelihood. Pairplots of all parameters for both models can be found in Appendix A.

4 Discussion

We demonstrated how to use normalizing flows to estimate the likelihood of systems biology models, overcoming their known limitations with high-dimensional data [7] by simplifying input dimensions and demonstrated utility in model selection. Future work will explore deeper flow networks and alternative methods like diffusion models, which are better at handling high-dimensional datasets while permitting a likelihood approximation via Probability Flow ODE [22].

5 References

[1] Yaron E Antebi, James M Linton, Heidi Klumpe, Bogdan Bintu, Mengsha Gong, Christina Su, Reed McCardell, and Michael B Elowitz. "Combinatorial signal perception in the BMP pathway". In: *Cell* 170.6 (2017), pp. 1184–1196.

[2] Johann Brehmer. "Simulation-based inference in particle physics". In: *Nature Reviews Physics* 3.5 (2021), pp. 305–305.

[3] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. "The frontier of simulation-based inference". In: *Proceedings of the National Academy of Sciences* 117.48 (Nov. 2020). arXiv: 1911.01429 Publisher: Proceedings of the National Academy of Sciences, pp. 30055–30062. ISSN: 0027-8424. DOI: 10.1073/pnas.1912789117. URL: http://arxiv.org/abs/1911.01429 (visited on 09/13/2020).

[4] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. "Neural spline flows". In: *arXiv* (June 10, 2019). Publisher: arXiv. ISSN: 23318422. arXiv: 1906.04032. URL: http://arxiv.org/abs/1906.04032 (visited on 11/28/2020).

[5] Martin Feinberg. "Foundations of chemical reaction network theory". In: (2019).

[6] Adam Foster, Martin Jankowiak, Matthew O'Meara, Yee Whye Teh, and Tom Rainforth. "A unified stochastic gradient approach to designing Bayesian-optimal experiments". In: *arXiv* (Nov. 2019). arXiv: 1911.00294 Publisher: arXiv. ISSN: 23318422. URL: http://arxiv.org/abs/1911.00294 (visited on 02/04/2021).

[7] Pedro J Gonçalves, Jan-Matthis Lueckmann, Michael Deistler, Marcel Nonnenmacher, Kaan Öcal, Giacomo Bassetto, Chaitanya Chintaluri, William F Podlaski, Sara A Haddad, Tim P Vogels, David S Greenberg, and Jakob H Macke. "Training deep neural density estimators to identify mechanistic models of neural dynamics". In: *eLife* 9 (Sept. 2020). Ed. by John R Huguenard, Timothy O'Leary, and Mark S Goldman, e56261. ISSN: 2050-084X. DOI: 10.7554/eLife.56261. URL: https://doi.org/10.7554/eLife.56261.

[8] David Greenberg, Marcel Nonnenmacher, and Jakob Macke. "Automatic posterior transformation for likelihood-free inference". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 2404–2414.

[9] Joeri Hermans, Volodimir Begy, and Gilles Louppe. *Likelihood-free MCMC with Amortized Approximate Ratio Estimators*. 2020. arXiv: 1903.04057 [stat.ML].

[10] Steven Kleinegesse and Michael U Gutmann. "Efficient Bayesian experimental design for implicit models". In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, pp. 476–485.

[11] Heidi E Klumpe, Matthew A Langley, James M Linton, Christina J Su, Yaron E Antebi, and Michael B Elowitz. "The context-dependent, combinatorial logic of BMP signaling". In: *Cell systems* 13.5 (2022), pp. 388–407.

[12] Ivan Kobyzev, Simon J.D. Prince, and Marcus A. Brubaker. "Normalizing Flows: An Introduction and Review of Current Methods". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.11 (Nov. 2021), pp. 3964–3979. DOI: 10.1109/tpami.2020.2992934. URL: https://doi.org/10.1109%2Ftpami.2020.2992934.

[13] Jan-Matthis Lueckmann, Jan Boelts, David Greenberg, Pedro Goncalves, and Jakob Macke. "Benchmarking simulation-based inference". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2021, pp. 343–351.

[14] JM Lueckmann, G Bassetto, T Karaletsos, and JH Macke. "Likelihood-free inference with emulator networks. arXiv e-prints". In: *arXiv preprint arXiv:1805.09294* (2018).

[15] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[16] Benjamin Kurt Miller, Christoph Weniger, and Patrick ForrÃ©. *Contrastive Neural Ratio Estimation*. arXiv:2210.06170 [astro-ph, physics:hep-ph, stat]. Oct. 2022. DOI: 10.48550/arXiv.2210.06170. URL: http://arxiv.org/abs/2210.06170 (visited on 01/10/2023).

[17] Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. URL: probml.ai.

[18] George Papamakarios and Iain Murray. "Fast e-free inference of simulation models with Bayesian conditional density estimation". In: *Advances in Neural Information Processing Systems*. arXiv: 1605.06376 Issue: Nips ISSN: 10495258. May 2016, pp. 1036–1044. URL: http://arxiv.org/abs/1605.06376 (visited on 11/07/2020).

[19] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. "Normalizing Flows for Probabilistic Modeling and Inference". In: (Dec. 2019). arXiv: 1912.02762. URL: http://arxiv.org/abs/1912.02762.

[20] George Papamakarios, David C. Sterratt, and Iain Murray. *Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows*. arXiv: 1805.07226 Publication Title: arXiv. May 2018. URL: http://arxiv.org/abs/1805.07226 (visited on 11/26/2020).

[21] Valerie S Salazar, Laura W Gamer, and Vicki Rosen. "BMP signalling in skeletal development, disease and repair". In: *Nature Reviews Endocrinology* 12.4 (2016), pp. 203–221.

[22] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. *Score-Based Generative Modeling through Stochastic Differential Equations*. 2021. arXiv: 2011.13456 [cs.LG].

[23] Christina J Su, Arvind Murugan, James M Linton, Akshay Yeluri, Justin Bois, Heidi Klumpe, Matthew A Langley, Yaron E Antebi, and Michael B Elowitz. "Ligand-receptor promiscuity enables cellular addressing". In: *Cell systems* 13.5 (2022), pp. 408–425.

A Full Posteriors of Onestep and Twostep Models

For qualitative reference, we present the entire posteriors for both binding affinities K, and phosphorylation efficiency ε, in this section.

Onestep Model Posteriors. We begin with a subset of posteriors for the onestep model shown in Figure 2. Generally, the posteriors for both types of parameters seem to have dominating peaks for both sets of parameters, which may indicate convergence of the model to those peak estimates. Occasionally, there are bimodal distributions for binding affinities, which physically means either the complex weakly or strongly prefers. Future experiments could help distinguish whether there is strong or weak binding affinity.

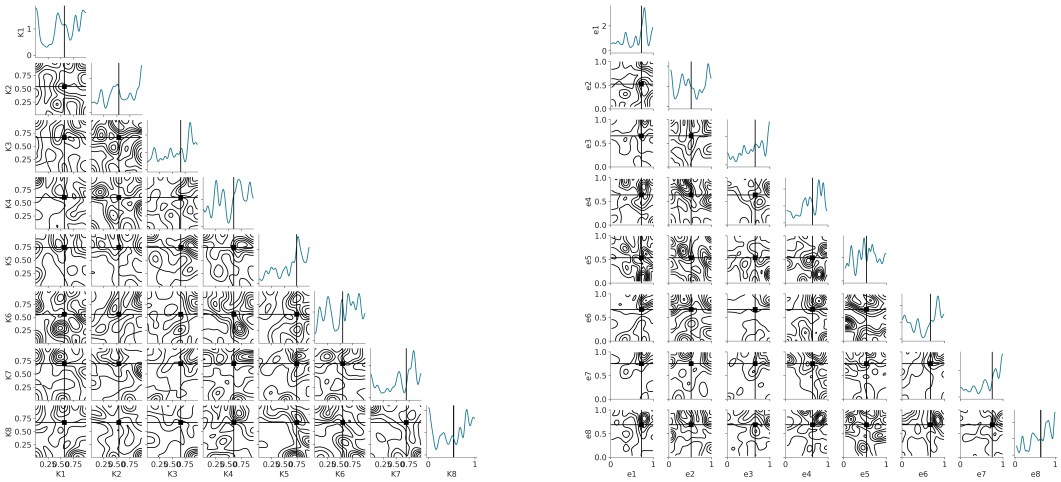

Figure 2: Pairplots for subsets of the K (left) and ε (right) parameters for the onestep model.

Twostep Model Posteriors. Looking at a subset of posteriors for the twostep model, Figure 3 also shows uncertain posteriors but with more "peaks" in parameter values that could be possible. In a BOED context, this would indicate conducting experiments that could help elucidate between these peaks to drive convergence to one set of parameters.

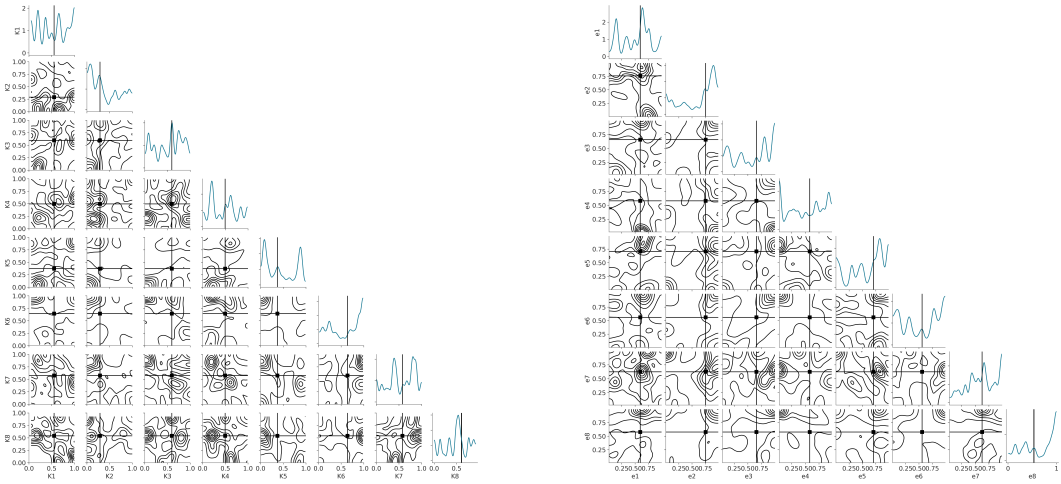

Figure 3: Pairplots for subsets of the K (left) and ε (right) parameters for the twostep model.