Simulation-based inference with approximately correct parameters via maximum entropy

Rainier Barrett
Department of Chemical Engineering
University of Rochester
Rochester, NY 14627
rbarret8@ur.rochester.edu

Mehrad Ansari
Department of Chemical Engineering
University of Rochester
Rochester, NY 14627
mehrad.ansari@rochester.edu

Gourab Ghoshal
Department of Physics and Astronomy
University of Rochester
Rochester, NY 14627
gghoshal@pas.rochester.edu

Andrew White
Department of Chemical Engineering
University of Rochester
Rochester, NY 14627
andrew.white@rochester.edu

Abstract

Inferring the input parameters of simulators from observations is a crucial challenge with applications from epidemiology to molecular dynamics. Here we show a simple approach in the regime of sparse data and approximately correct models, which is common when trying to use an existing model to infer latent variables with observed data. This approach is based on the principle of maximum entropy and provably makes the smallest change in the latent joint distribution to accommodate new data. This simple method requires no likelihood or simulator derivatives and its fit is insensitive to prior strength, removing the need to balance observed data fit with prior belief. We demonstrate this MaxEnt approach and compare with other likelihood-free inference methods across three example systems.

1 Introduction

Simulation-based inference (SBI) is a class of methods that infer the input parameters and unobservable latent variables in a simulator from observational data. SBI is different than traditional statistical inference or machine learning because simulators are typically not differentiable and their likelihoods are intractable. There have been great strides in methods for SBI and a recent review may be found in [1]. Most SBI methods are concerned with finding a few simulator parameters from a rich set of observations[2, 3, 4]. Here, we consider biasing a simulator with many trusted parameters to match a sparse set of observations. The motivating example for this line of research is in molecular dynamics simulations of proteins. These simulations require thousands of parameters and the observed data is on the order of 10 data points. An approach that has emerged successfully in molecular dynamics simulations is maximum entropy (MaxEnt) biasing. MaxEnt biasing minimally modifies the simulator to match observations. The premise of MaxEnt is that the original model is approximately correct and observations should be matched in expectation, which is different than other methods. These two assumptions lead to a unique bias to the simulator that is independent of the parameters and can be implemented as a simple reweighting procedure. The method's run-time scales only with sample number, rather than the number of model parameters which is atypical of most SBI methods because they require joint sampling.

The idea of using MaxEnt for biasing simulators goes back to Jayne's pioneering work on deriving statistical physics from MaxEnt[5]. It was shown, for example, that the Boltzmann distribution

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

could be derived by simply adding a restraint on average energy that must be satisfied in expectation, analogous to matching an observation. A similar method of incorporating observations in expectation returned 50 years later in determining how to match protein molecular dynamics simulations to observations[6]. This method was then recast as an approximation to MaxEnt[7] and Pitera and Chodera[8] showed how to directly match observations in molecular dynamics with MaxEnt. This was followed by rapid progress to create practical methods for use in simulations[9, 10, 11, 12]. A recent review of these "minimal biasing" methods can be found in Bonomi et. al[13]. The MaxEnt method derived here, which is about reweighting, has been presented in many forms over the years[14, 15, 16, 17, 18, 19, 20, 21, 22, 23].

Our contribution here is deriving a general MaxEnt framework that is applicable to arbitrary simulators, demonstrating its application to areas outside of molecular dynamics, and showing one method of improving the support (sampling) of the posterior, which is important when the simulator is far from the observations. In the remainder of this work, we develop the theory, discuss sampling issues, and compare the MaxEnt method to approximate Bayesian computing (ABC)[3, 24, 25, 26], Sequential neural likelihood (SNL)[27], and direct Bayesian inference when the likelihood is tractable.

2 Theory

Given a simulator $f(\vec{\theta})$ with a set of parameters $\vec{\theta}$, we have a prior distribution of parameters $P(\vec{\theta})$. For example, the function $f(\vec{\theta})$ could be propagating a system of ODEs for some set number of timesteps or a molecular dynamics simulation with intrinsic noise.

Suppose we have some set of N observations, $\{\bar{g}\}_k$, $k \in [1, \ldots, N]$, which we would like to match with our model. We would like to constrain our model such that

$$\int \mathrm{d}\vec{\theta}\,\mathrm{d}\vec{\epsilon} P'(\vec{\theta}) g_k[f(\vec{\theta})] = E[g_k] = \bar{g}_k \forall k \tag{1}$$

This means that we want the average over the distribution of our biased models ($P'(\vec{\theta})$) to match the observations data. This is an unusual constraint and is weaker than most simulation inference methods. It reflects the strong belief in our prior model in this setting.

The unique maximum entropy modification to the prior distribution $P(\vec{\theta})$ to satisfy the N constraints is[7, 8, 12]

$$P'(\vec{\theta}) = \frac{1}{Z'} P(\vec{\theta}) \prod_k^N e^{-\lambda_k g_k[f(\vec{\theta})]} \tag{2}$$

where Z' is a normalization constant and λ_k are chosen such that $E[g_k] = \bar{g}_k$. The challenge of using MaxEnt is sampling from $P'(\vec{\theta})$. Our assumption is that our prior $P(\vec{\theta})$ is approximately correct, so that samples from $P(\vec{\theta})$ should be similar to $P'(\vec{\theta})$. This is the ideal case and our algorithm is simply a matter of reweighting. You sample $\vec{\theta}_i$, compute $f(\vec{\theta}_i)$, compute weights proportional to $w_i[P'] = \prod_k^N e^{-\lambda_k g_k[f(\vec{\theta})]}$ consistent with the experimental data, and then any other property is reweighted. In the non-ideal case (if e.g. sampling is expensive, the space is high-dimensional, or the model is far from correct), there can be insufficient support to agree with the constraints. To treat insufficient support, we take a simple approach and use gradient descent to modify the sampling distribution $P^j(\vec{\theta})$ to minimize the cross-entropy with $P'(\vec{\theta})$:

$$P^{j+1}(\vec{\theta}) = P^j(\vec{\theta}) - \eta \frac{\delta}{\delta P^j(\vec{\theta}')} \sum_i w_i[P'] \ln P(\vec{\theta}_i) \tag{3}$$

where $w_i[P']$ depends on $P^j(\vec{\theta}')$ via the expectation function and $\frac{\delta}{\delta P^j(\vec{\theta}')}$ indicates a functional derivative. This approach can include uncertainty in experimental data by explicitly accounting for bias[28], though for brevity we do not consider this here.

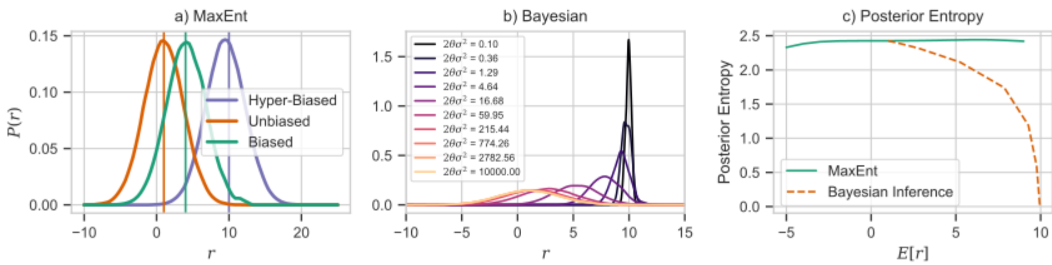

Figure 1: Comparison of Bayesian inference and MaxEnt reweighting. Panel a shows the unbiased simulator distribution in orange and the two versions of MaxEnt biased to observations of 5 and 10. Panel b shows the interplay between strength of prior and assigned uncertainty to the observation at 10. Panel c compares the posterior entropy of the two as a function of the observation location.

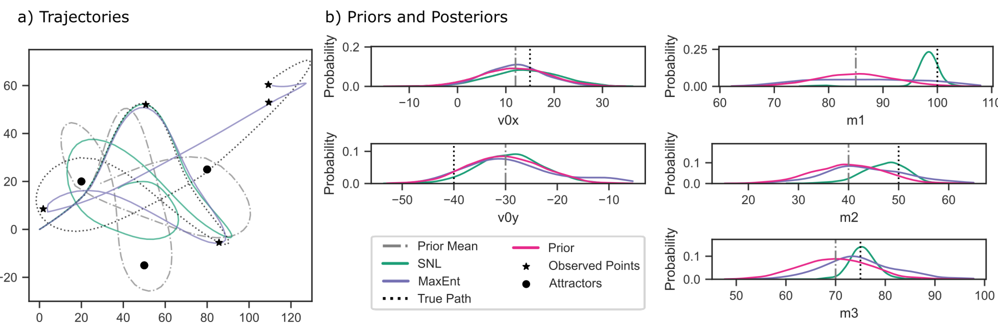

Figure 2: Comparison of SNL and MaxEnt methods on a gravitational field simulation of a particle moving through a fixed field with three attractors. **a)**: weighted mean paths generated by SBI with SNL (green) and MaxEnt (purple), alongside the path generated by the mean of the prior distribution (dash-dotted grey), and the true path used to generate observations (dashed black). Target points appear as black stars, and the attractors are black circles. **b)**: Kernel density estimate of the posterior distribution of parameters after fitting, alongside their respective priors.

3 Results and Discussion

The first example is a simulator f that is a linear function with Gaussian noise: given input θ, $f(\theta) = r\theta + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma)$ and parameter r with uniform prior. This example serves to compare the MaxEnt approach with Bayesian inference. Our observation is a single point and we treat it as an average constraint in the MaxEnt. In the Bayesian inference method, we must provide some uncertainty with this point to create a probability distribution[29]. Bayesian inference balances evidence with the prior distribution and here that is the ratio of our prior certainty (θ) with the certainty of the observation $(1/2\sigma)$. Figure 1 contrasts these two methods. Panel a shows how the MaxEnt method leaves the variance of the simulator unchanged as we move the observed value. The Hyper-Biased method updates hyperparameters to improve sampling as shown in Equation 3. Panel b shows how the Bayesian inference case requires explicit choice between prior belief and experimental uncertainty to match the observation at 10. Panel c shows how the MaxEnt method keeps the distribution entropy maximized regardless of the observation value (x-axis).

For a second example, Figure 2 shows a comparison of SNL and MaxEnt reweighting on a unit mass particle in a gravitational field of three attractors. The parameters for this simulation were m_1, m_2, m_3, v_{0x} and v_{0y}, the masses of the three attractors (fixed positions), and the initial velocity of the particle, respectively. The particle's path over time is shown in Figure 2a. The prior parameter distribution was taken as a multivariate normal distribution centered at $\{m_1 = 85, m_2 = 40, m3 = 70, v_{0x} = 12, v_{0y} = -30\}$, with covariance matrix $\mathbf{I} \times 50$. This wide prior was chosen because MaxEnt needs parameter support that overlaps with the observations we would like to fit. Fitting was done using the SBI package for Python[30] with the SNL method,[27] and a custom implementation of MaxEnt reweighting using Keras[31, 32]. Both methods used 2048 prior samples for fitting. SNL used default parameters from the SBI package[30] and MaxEnt used the Adam optimizer with a learning rate of 0.0001 with mean squared error for 30000 epochs and batch size 2048. The final posterior distributions are shown in Figure 2b. In order to compare these results, we computed the cross-entropy of the prior and posterior produced by each method. These values were 5.09 for SBI with SNL, and 3.43 for MaxEnt reweighting. This demonstrates how MaxEnt minimally alters the prior distribution while still matching observations in expectation – the average path followed by the MaxEnt-biased particle matches all target points, while matching the posterior to the prior's shape more closely than SNL.

Finally, we consider maximum entropy reweighting of an SEAIR epidemiology model. Epidemic spreading in networks can be modeled as a reaction-diffusion process. The reaction corresponds to an infection caused by interactions of subjects within a fully-mixed region or patch of varying granularities (a meta-population), while diffusion corresponds to movement of people (of various infection states) between patches [33]. In our example application, the meta-population system is comprised of three isolated local populations (patches) connected via flows corresponding to migrating individuals. The spreading process is represented through a temporally discretized ODE that includes the spatial distribution of the population as well as their mobility patterns [34]. Populations in each patch can be in any one of Susceptible (**S**), Exposed (**E**), Asymptomatic (**A**), Infected (I), and Recovered (**R**). The choice for this particular flavor of compartments was inspired by its relevance in modeling the evolution of the current COVID-19 pandemic [35, 36]. The connection between the patches is defined based on an $M \times M$ mobility matrix, where M corresponds to the number of patches.

The empirical number of confirmed cases (compartment **I**) is typically noisy. Here, uncertainty in case numbers was accounted for by adding random additive noise to the observations from the reference trajectory. The reference trajectory was obtained from arbitrary epidemiology parameters (see dashed lines in Figure 3.a) In our simulation, the infection begins in patch 1, propagating to other patches according to the mobility matrix. Five randomly-selected data points within the first half of the trajectory of the compartment **I** in patch 1 were considered as observations. Parameters for this simulation were asymptomatic, infected and exposed periods along with the starting values for **I** and **A**. Figure 3.b compares the performance of MaxEnt, least-squares, and ABC in fitting the prior to the observations. The pyABC[37] package was used with default parameters, and the same MaxEnt implementation was used with the SGD optimizer, a learning rate of 0.1, and loss of mean squared error for 1000 epochs with a batch size of 8192. Compared to MaxEnt, the result from the least squares method was a poor fit with high variance, as it over-fitted to observation noise. This was shown by doing a 5-fold leave-one-out cross-validation of the observations and evaluating the standard deviation at times $t = 0, 125$ and 250 for each method (inset in Figure 3.b). Out of all methods evaluated, ABC had the least variance, but was computationally more expensive to run, whereas MaxEnt can include more model parameters without additional computational cost.

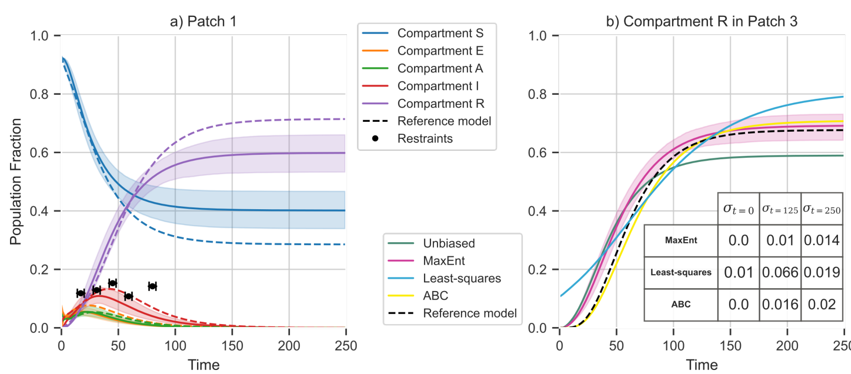

Figure 3: Maximum entropy biasing of disease trajectory in a meta-population SEAIR model. a) Unbiased trajectory in patch 1 for compartments S (blue), E (orange), A (green), I (red) and R (purple) are shown with solid lines and the reference trajectory is in dashed lines. The colored area represents the one-third higher and lower quantiles than average. Restraints (black circles) are selected randomly from compartment I with additive noise. b) Comparing the performance of MaxEnt (pink), Least-squares (blue), ABC (yellow) in fitting to reference model (black dashed line) in patch 3, based on observations in patch 1. Table inset shows standard deviations from 5-fold cross validation of the observations at three times.

4 Conclusion

We have presented MaxEnt reweighting as an SBI method for altering an approximately-correct simulator to agree with observations. This black box method can be used on arbitrary simulators with arbitrary numbers of parameters, requiring only sufficient sampling of the prior distribution. We demonstrated this by comparing with other SBI methods with three different simulators. MaxEnt reweighting is effective and robust when data is scarce or expensive, and provably changes the prior minimally to fit observations. Thus, MaxEnt reweighting is well-suited to the regime of nearly-correct simulators when observed data is scarce or expensive.

Broader Impact

MaxEnt reweighting is a straightforward black box method that can be applied to arbitrary simulators with few observations. Its runtime is independent of the number of parameters used by the simulator, and it has been shown analytically[5, 8] to minimally change the prior to agree with observations. This method fills a niche in the small-data, high-complexity regime of SBI parameter inference, because it accurately and minimally biases a prior to match observations and does not scale in runtime with the number of model parameters. The benefits of this biasing method is a stable biasing option in the case when experimental observations of true data is limited and an initial model is well-trusted. This could benefit anyone seeking to improve the accuracy of statistical models in such a situation. We do not foresee anyone being disadvantaged by this research. Should this system fail, predictive models which have been biased with this method will be inaccurate, which could result in negative outcomes for anyone using such models to choose a course of action or make predictions. This method does not make use of biases in data but rather can be used to account for and correct systematic error.

Acknowledgments and Disclosure of Funding

Funding for this research was provided by National Science Foundation under Grant No 2029095.

References

[1] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, page 201912789, may 2020.

[2] Donald B Rubin. Bayesianly justifiable and relevant frequency calculations for the applies statistician. *The Annals of Statistics*, pages 1151–1172, 1984.

[3] Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.

[4] Peter J Diggle and Richard J Gratton. Monte carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(2):193–212, 1984.

[5] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[6] Shahidul M Islam, Richard A Stein, Hassane S Mchaourab, and Benoît Roux. Structural refinement from restrained-ensemble simulations based on epr/deer data: application to t4 lysozyme. *The journal of physical chemistry B*, 117(17):4740–4754, 2013.

[7] Benoît Roux and Jonathan Weare. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *The Journal of chemical physics*, 138(8):02B616, 2013.

[8] Jed W Pitera and John D Chodera. On the Use of Experimental Observations to Bias Simulated Ensembles. *Journal of Chemical Theory and Computation*, 8(10):3445–3451, 2012.

[9] Andrea Cavalli, Carlo Camilloni, and Michele Vendruscolo. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *The Journal of chemical physics*, 138(9):03B603, 2013.

[10] Wouter Boomsma, Jesper Ferkinghoff-Borg, and Kresten Lindorff-Larsen. Combining experiments and simulations using the maximum entropy principle. *PLoS Comput Biol*, 10(2):e1003406, 2014.

[11] Andrew D White and Gregory A Voth. Efficient and minimal method to bias molecular simulations with experimental data. *Journal of chemical theory and computation*, 10(8):3023–3030, 2014.

[12] Dilnoza B Amirkulova and Andrew D White. Recent advances in maximum entropy biasing techniques for molecular dynamics. *Molecular Simulation*, 45(14-15):1285–1294, 2019.

[13] Massimiliano Bonomi, Gabriella T Heller, Carlo Camilloni, and Michele Vendruscolo. Principles of protein structural ensemble determination. *Current opinion in structural biology*, 42:106–116, 2017.

[14] Sabine Reißer, Silvia Zucchelli, Stefano Gustincich, and Giovanni Bussi. Conformational ensembles of an RNA hairpin using molecular dynamics and sparse NMR data. *Nucleic Acids Research*, 48(3):1164–1174, 12 2019.

[15] Kyle A Beauchamp, Vijay S Pande, and Rhiju Das. Bayesian energy landscape tilting: towards concordant models of molecular ensembles. *Biophysical journal*, 106(6):1381–1390, 2014.

[16] Bartosz Różycki, Young C Kim, and Gerhard Hummer. Saxs ensemble refinement of escrt-iii chmp3 conformational transitions. *Structure*, 19(1):109–116, 2011.

[17] Hoi Tik Alvin Leung, Olivier Bignucolo, Regula Aregger, Sonja A Dames, Adam Mazur, Simon Berneche, and Stephan Grzesiek. A rigorous and efficient method to reweight very large conformational ensembles using average experimental data and to determine their relative information content. *Journal of chemical theory and computation*, 12(1):383–394, 2016.

[18] Wing-Yiu Choy and Julie D Forman-Kay. Calculation of ensembles of structures representing the unfolded state of an sh3 domain. *Journal of molecular biology*, 308(5):1011–1032, 2001.

[19] Pau Bernadó, Efstratios Mylonas, Maxim V Petoukhov, Martin Blackledge, and Dmitri I Svergun. Structural characterization of flexible proteins using small-angle x-ray scattering. *Journal of the American Chemical Society*, 129(17):5656–5664, 2007.

[20] Konstantin Berlin, Carlos A Castaneda, Dina Schneidman-Duhovny, Andrej Sali, Alfredo Nava-Tudela, and David Fushman. Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *Journal of the American Chemical Society*, 135(44):16595–16609, 2013.

[21] Ivano Bertini, Andrea Giachetti, Claudio Luchinat, Giacomo Parigi, Maxim V Petoukhov, Roberta Pierattelli, Enrico Ravera, and Dmitri I Svergun. Conformational space of flexible biological macromolecules from average data. *Journal of the American Chemical Society*, 132(38):13553–13558, 2010.

[22] Martin Pelikan, Greg L Hura, and Michal Hammel. Structure and flexibility within proteins as identified through small angle x-ray scattering. *General physiology and biophysics*, 28(2):174, 2009.

[23] David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.

[24] Michael G B Blum and Chi Tran. HIV with contact tracing: a case study in approximate Bayesian computation. *Biostatistics*, 11(4):644–660, 2010.

[25] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P.H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6(31):187–202, 2009.

[26] Theodore Kypraios, Peter Neal, and Dennis Prangle. A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. *Mathematical Biosciences*, 287:42–53, 2017.

[27] George Papamakarios, David C. Sterratt, and Iain Murray. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. *ArXiv*, 2019.

[28] Andrea Cesari, Alejandro Gil-Ley, and Giovanni Bussi. Combining simulations and solution experiments as a paradigm for rna force field refinement. *Journal of chemical theory and computation*, 12(12):6192–6200, 2016.

[29] Gerhard Hummer and Jürgen Köfinger. Bayesian ensemble refinement by replica simulations and reweighting. *The Journal of chemical physics*, 143(24):12B634_1, 2015.

[30] Alvaro Tejero-Cantero, Jan Boelts, Michael Deistler, Jan-Matthis Lueckmann, Conor Durkan, Pedro J. Gonçalves, David S. Greenberg, and Jakob H. Macke. sbi: A toolkit for simulation-based inference. *Journal of Open Source Software*, 5(52):2505, 2020.

[31] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[32] François Chollet. keras. `https://github.com/fchollet/keras`, 2015.

[33] Jesús Gómez-Gardenes, David Soriano-Panos, and Alex Arenas. Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks. *Nature Physics*, 14(4):391–395, 2018.

[34] Alex Arenas, Wesley Cota, Jesus Gomez-Gardenes, Sergio Gómez, Clara Granell, Joan T Matamalas, David Soriano-Panos, and Benjamin Steinegger. A mathematical model for the spatiotemporal epidemic spreading of covid19. *MedRxiv*, 2020.

[35] Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature*, 579(7798):270–273, 2020.

[36] Fan Wu, Su Zhao, Bin Yu, Yan-Mei Chen, Wen Wang, Zhi-Gang Song, Yi Hu, Zhao-Wu Tao, Jun-Hua Tian, Yuan-Yuan Pei, et al. A new coronavirus associated with human respiratory disease in china. *Nature*, 579(7798):265–269, 2020.

[37] Emmanuel Klinger, Dennis Rickert, and Jan Hasenauer. pyABC: distributed, likelihood-free inference. *Bioinformatics*, 34(20):3591–3593, 05 2018.