

Statistical calibration of pattern formation models

Heikki Haario¹, Alexey Kazarnikov², Robert Scheichl², and Anna Marciniak-Czochra²

¹LUT University, Lappeenranta, Finland

²University of Heidelberg, Heidelberg, Germany

Oxford

24.05.2022

Motivation and objective of the research

- Chemical and mechanical pattern formation is fundamental during embryogenesis and multicellular structures development.
- Several competing theories exist, which should be validated against data
- In many experimental situations, it is possible to observe only limit stationary regimes of the pattern formation process, without the detailed knowledge on the transient behaviour and initial data.
- In such case, initial data may be assumed to be random perturbations of some base state of the pattern formation. Then fixed values of model parameters correspond to a family of patterns rather than a fixed solution.
- Measuring the fit (residuals) between model output and pattern data in the standard (LSQ) way is no more available.

Pattern formation in reaction-diffusion systems

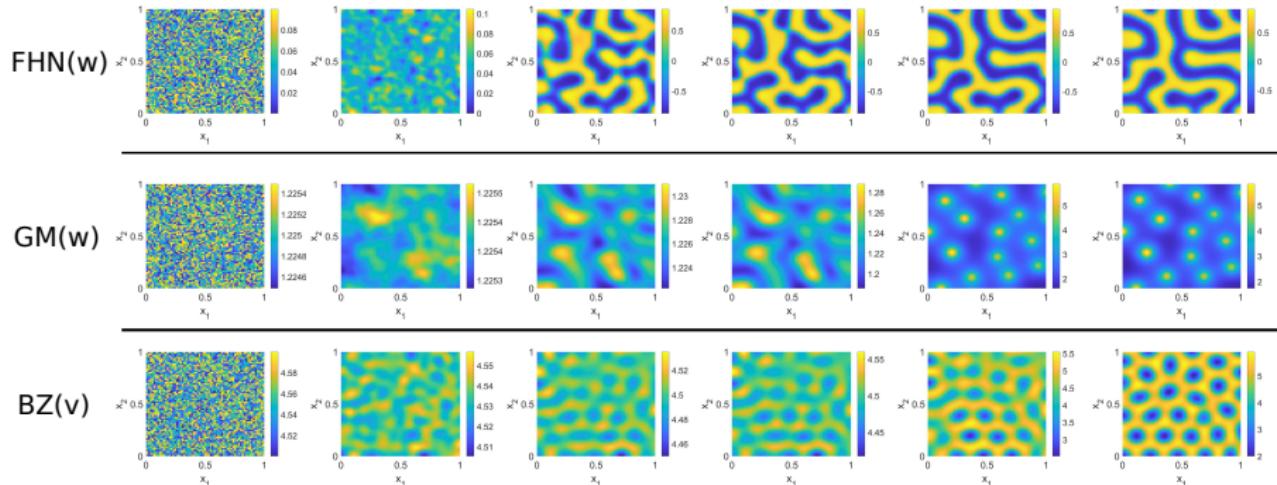


Figure 1: The formation of stationary patterns (Turing patterns) in reaction-diffusion systems. FHN, GM and BRS denote the FitzHugh-Nagumo model, Gierer-Meinhardt system and Brusselator reaction-diffusion system respectively.

Pattern formation (fixed model parameters)

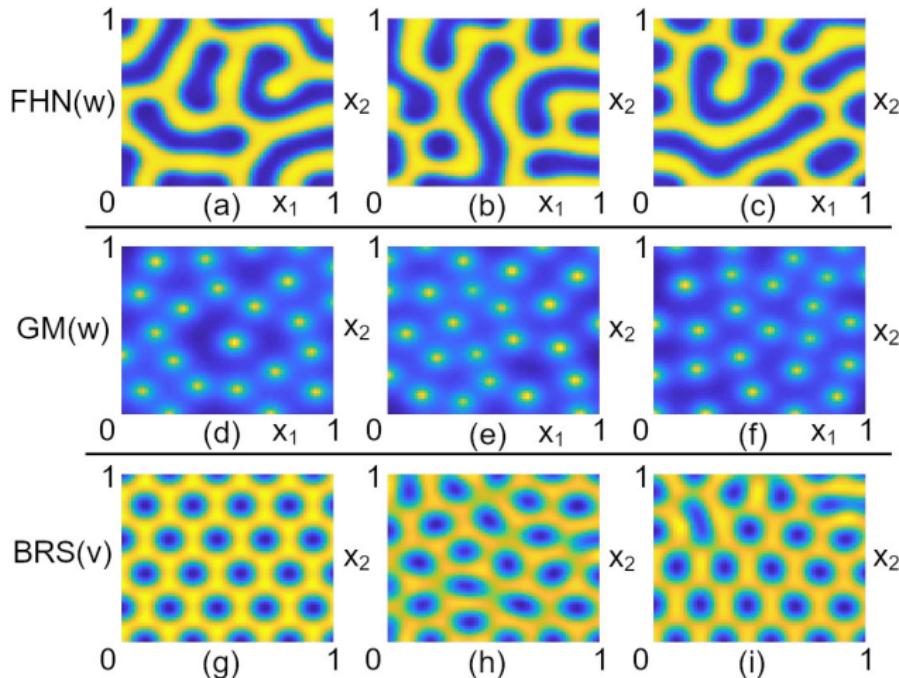


Figure 2: Turing patterns obtained from direct simulations with fixed parameter values and different initial conditions, taken as small random perturbations of the homogeneous steady state.

Pattern formation (different model parameters)

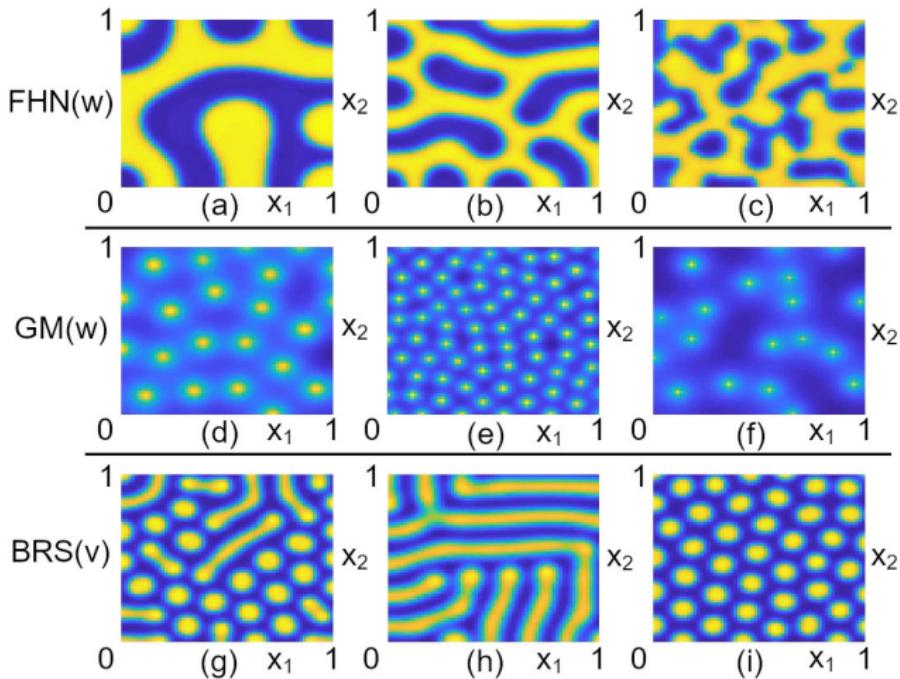


Figure 3: Turing patterns obtained from direct simulations with different parameter values and fixed initial conditions.

The approach: Gaussian feature vectors

- The problem can be approached by a stochastic likelihood construction [1, 2]. Both cases - chaotic dynamics and random patterns - share the same characteristic: unpredictability.
- In [2], the approach has been tested with three classical reaction-diffusion models, showing the formation of Turing patterns.
- The suggested method, however, was successful with at least 50 data patterns available.
- Here: increase the accuracy of the method, allowing estimation by limited data.

¹Kazarnikov A., H.H. Statistical approach for parameter identification by Turing patterns // Journal of Theoretical Biology.—2020.—Vol. 501.—P. 110319.

²H. H., Kalachev L., Hakkarainen J. Generalized correlation integral vectors: A distance concept for chaotic dynamical systems // Chaos.—2015.—Vol. 25, No 6.

Synthetic Likelihood

Idea: in cases where the results of model simulation are non-deterministic

- Create the likelihood by repeated simulations, for any given model parameter value.
- Test the data against the likelihood (accept/reject)
- S.N. Wood: *Statistical inference for noisy nonlinear ecological dynamic systems*, Nature 2010
- L.Price, C.Drovandi, A. Lee, D. Nott *Bayesian synthetic likelihood*, BSL, JCGS 2018
- Several papers since 2014

Issues:

- Several simulations at every step, high CPU
- The statistics of the likelihood remains unknown, typically **assumed** to be Gaussian.

Which summary statistics to use?

Donsker Theorem

Let F_n be the empirical distribution function of the sequence of i.i.d. random variables X_1, X_2, X_3, \dots with distribution function F . Define the centered and scaled version of F_n by

$$G_n(x) = \sqrt{n}(F_n(x) - F(x))$$

Theorem (Donsker, Skorokhod, Kolmogorov) The sequence of $G_n(x)$ converges in distribution to a Gaussian process G with zero mean and covariance given by

$$\text{cov}[G(s), G(t)] = E[G(s)G(t)] = \min\{F(s), F(t)\} - F(s)F(t)$$

Donsker Theorem, finite sample size

Assuming i.i.d scalar data with sample size N :

- Create the empirical distribution function (eCDF) F by data, at bin values $x_i, i = 1, 2, \dots, M$.
- Use the Donsker formula to get the covariance
 $C_{ij} = \min(F_i, F_j) - F_i F_j)/N$ for the eCDF vector, with sample size N .

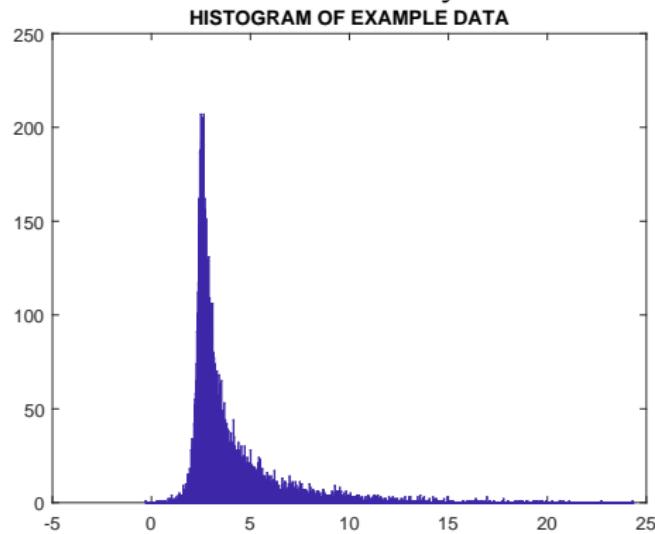
Get a M -dimensional likelihood $N(F, C)$, whose Gaussianity may be numerically verified, e.g., by the χ^2 test.

Note:

- The sample size needs to be large enough.
- The statistics of the likelihood is known, **assured** to be Gaussian.

Example: Synthetic likelihood with eCDF summary statistics

The 'g-and-k' distribution is a complex a complex distribution depending on four parameters, often used in comparing likelihood-free methods (ABC, BSL/Bayesian synthetic likelihood: *A Comparison of Likelihood-Free Methods With and Without Summary Statistics* C. Drovandi, D. Frazier, ArXiv 2021)

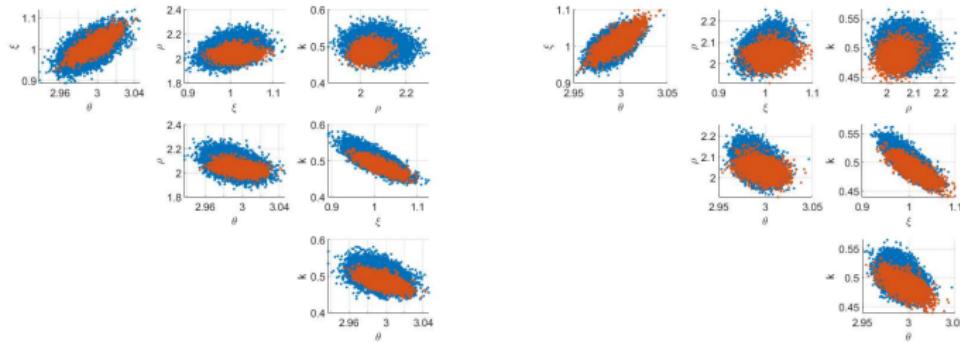


Example: Synthetic likelihood with eCDF summary statistics

BSL and MCMC, using the eCDF statistics, with Donsker formula for covariance estimates:

- For each proposed $\theta_1, \theta_2, \theta_3, \theta_4$ create n repetitions of sample size N , as in data. (Here $n = 10, N = 10000$)
- Create the 'synthetic likelihood' by simulated data, using the eCDF and Donsker formula
- Check if the eCDF of data fits the likelihood (accept/reject)

Results (in red), compared to competing BSL and ABC posteriors:



Beyond Donsker: Gaussian Subset Likelihoods (GSL)

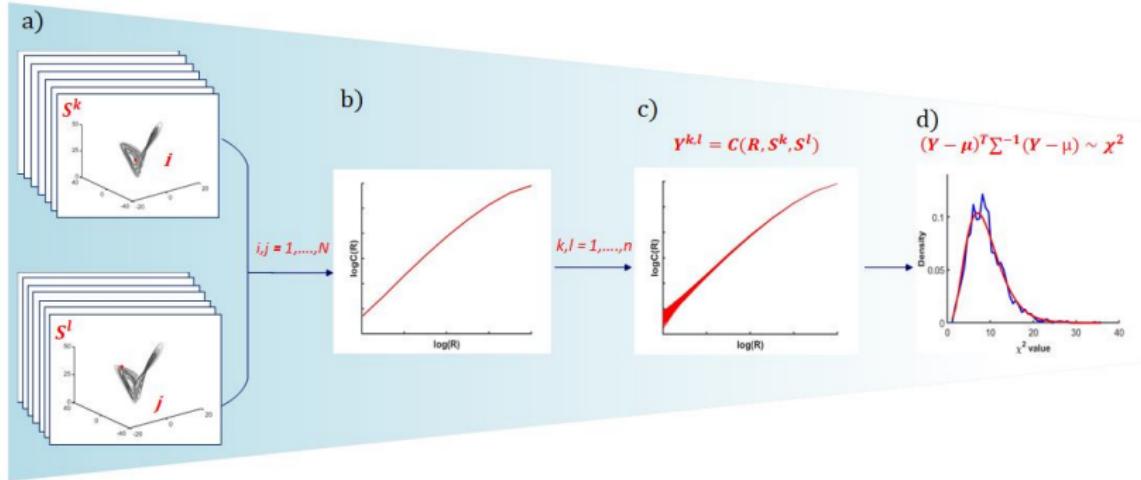
BSL (Bayesian Synthetic Likelihood) works fine for 'small' data sets, but may require excessive calculations for 'large' data sets and high CPU models.

In such settings:

- Avoid high CPU demand by creating likelihood for a **subset size** of data.
- Data generally not i.i.d. But normality holds also for 'weakly dependent' data instead of i.i.d (theorems in U-statistic literature)
- Vector valued data: employ an 'informative' mapping to scalar values

Likelihood construction. Example: using distance as the scalar mapping

Steps to create the likelihood for one 'epoch' of chaotic dynamics by calculating **pairwise distances** between all points of different data trajectories ('Correlation Integral Likelihood', CIL: modify the correlation fractal dimension)



2D reaction-diffusion systems

A general two-component reaction-diffusion system can be written in the form:

$$v_t = \nu_1 \Delta v + f(v, w); \quad w_t = \nu_2 \Delta w + g(v, w), \quad (1)$$

with Neumann boundary conditions:

$$\frac{\partial v}{\partial n}|_{\partial\Omega} = 0, \quad \frac{\partial w}{\partial n}|_{\partial\Omega} = 0, \quad (2)$$

where variables $v = v(\mathbf{x}, t)$ and $w = w(\mathbf{x}, t)$ describe spatio-temporal dynamics of unknown chemical concentrations, Δ is the Laplace operator, and $\nu_1, \nu_2 > 0$ are fixed diffusion coefficients. The non-linear functions $f(v, w)$ and $g(v, w)$ represent local chemical reactions and $t \geq 0$. Here $\mathbf{x} \in \Omega$, where $\Omega = (0, 1)^2 \subset \mathbb{R}^2$.

2D reaction-diffusion systems

We focus on three well-known types of kinetics. For simplicity, assume that all parameters are fixed except the pair of varying control parameters $\theta = (\theta_1, \theta_2)$.

The **FitzHugh-Nagumo** model involves the following reaction terms:

$$f(v, w) = \varepsilon(w - \alpha v); \quad g(v, w) = -v + \mu w - w^3; \quad \theta = (\mu, \varepsilon), \quad (3)$$

where $\alpha \geq 0$ is the fixed parameter. The kinetics of the **Gierer-Meinhardt** activator-inhibitor system are given by equations:

$$f(v, w) = -\mu_v v + \frac{v^2}{w}; \quad g(v, w) = -\mu_w w + v^2; \quad \theta = (\mu_v, \mu_w). \quad (4)$$

The **Brusselator** reaction-diffusion system is given by

$$f(v, w) = A - (B+1)v + v^2 w; \quad g(v, w) = Bv - v^2 w; \quad \theta = (A, B). \quad (5)$$

Parameter estimation ($N_{set} = 500$)

After construction the likelihood by data, find the MAP estimate of the unknown parameter:

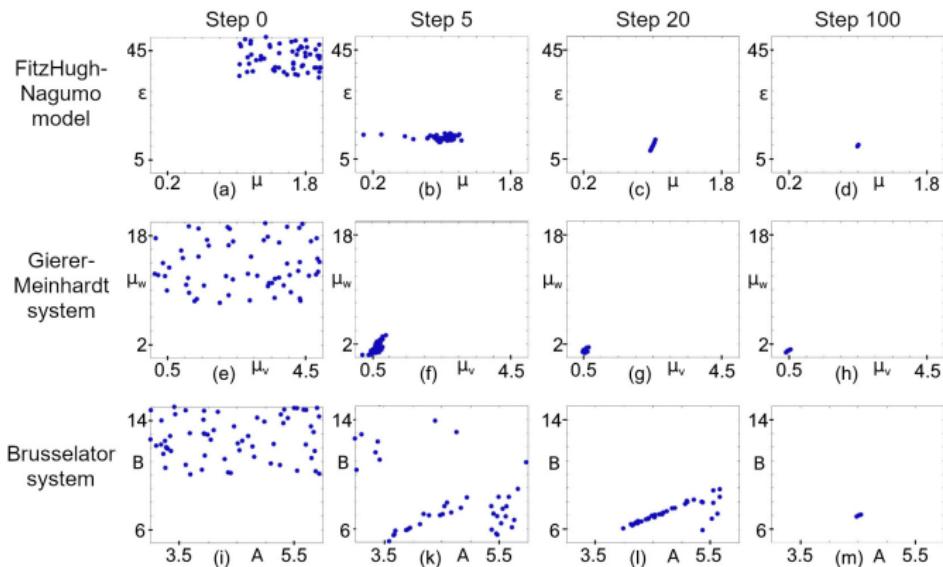


Figure 4: Parameter identification of the reaction-diffusion systems. The optimisation of stochastic cost function $f(\theta)$ by the **Differential Evolution algorithm**.

Parameter posterior distribution ($N_{set} = 500$)

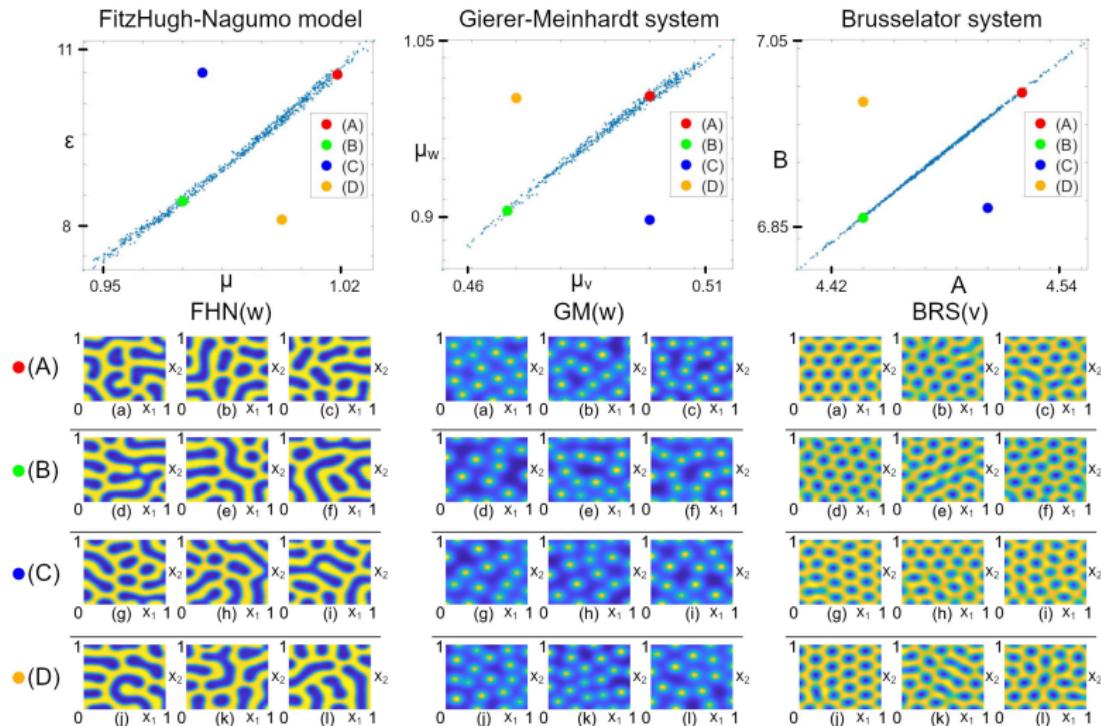


Figure 5: Posterior distributions of model parameters and Turing patterns obtained for the verification values.

Multi-feature identification

- In the GSL approach, empirical cumulative distribution functions (eCDF) evaluated at given bin values provide multidimensional Gaussian likelihoods for parameter identification.
- Several feature vectors (by scalar valued mappings from pattern data) can be used, highlighting different features of the patterns.
- Here, we define the mappings by different norms.

Norms to compute distances of 2D patterns

We consider the following norms:

$$\|f(\mathbf{x})\|_{L_2} = \left(\int_{\Omega} |f(\mathbf{x})|^2 dx \right)^{\frac{1}{2}}, \quad (6)$$

$$\|f(\mathbf{x})\|_{L_\infty} = \sup_{\mathbf{x} \in \Omega} |f(\mathbf{x})|, \quad (7)$$

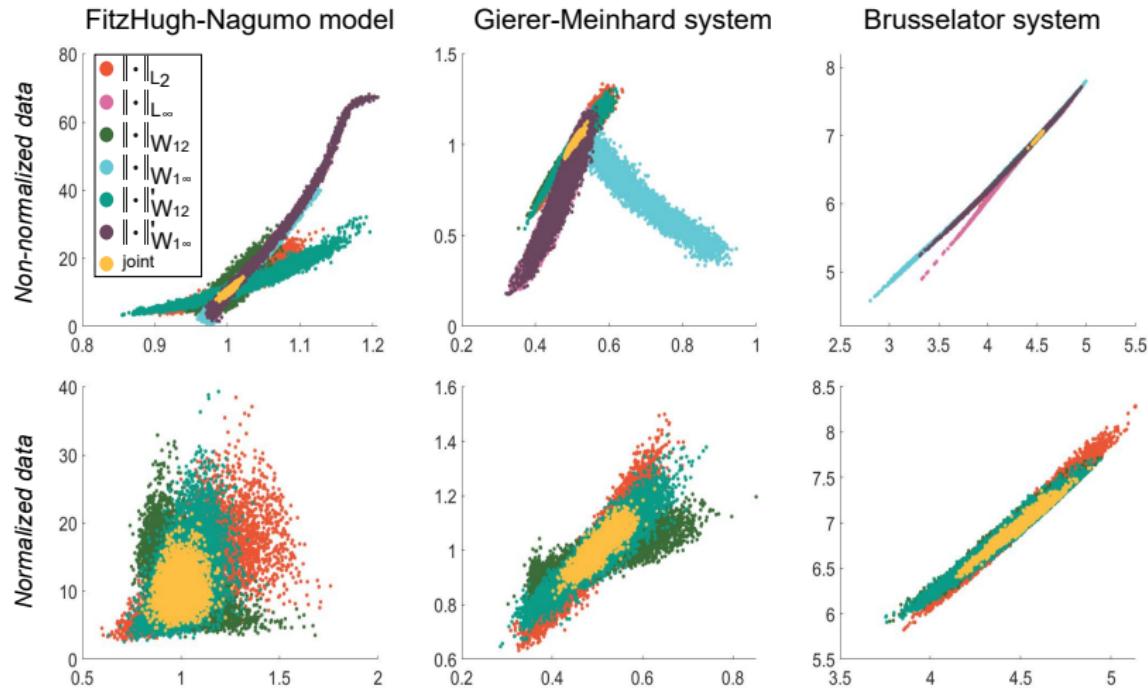
$$\|f(\mathbf{x})\|_{W^{1,\infty}} = \max_{|\alpha| \leq 1} \|D^\alpha(f(\mathbf{x}))\|_{L^\infty}, \quad (8)$$

$$\|f(\mathbf{x})\|_{W^{1,2}} = \left(\sum_{|\alpha| \leq 1} \|D^\alpha(f(\mathbf{x}))\|_{L^2}^2 \right)^{\frac{1}{2}}, \quad (9)$$

$$\|f(\mathbf{x})\|'_{W^{1,\infty}} = \sum_{|\alpha| \leq 1} \|D^\alpha(f(\mathbf{x}))\|_{L^\infty}, \quad (10)$$

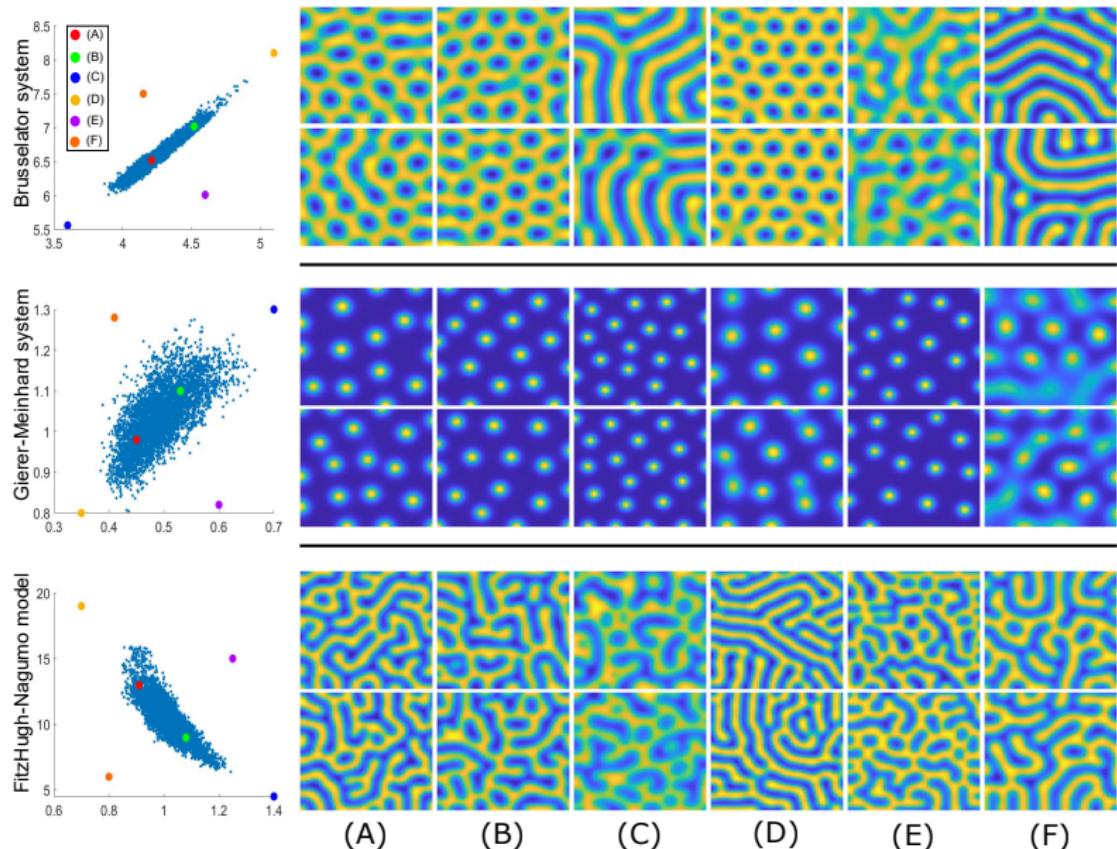
$$\|f(\mathbf{x})\|'_{W^{1,2}} = \sum_{|\alpha| \leq 1} \|D^\alpha(f(\mathbf{x}))\|_{L^2}, \quad (11)$$

Parameter posterior distributions (reaction-diffusion systems, $N_{set} = 50$)



- Consider the situation when training data is limited so that resampling techniques cannot be applied to approximate the CIL likelihood.
- Approximate the likelihood by using the model-generated synthetic set of data and comparing the training data with it.
- Requires additional computational costs. The overhead, however, can be efficiently handled if optimal parallel algorithms are available.

Verification (reaction-diffusion models, $N_{set} = 1$)



Mechanochemical models for pattern formation in biological tissues

- Traditionally models for biological patterns make the assumption that mechanical patterns are formed by chemical processes (Turing patterns)
- Experimental data show that there is a connection between chemical and mechanical processes
- As a result, new mechanochemical models have been developed (see [1-3])

¹M. Mercker, A. Kothe, and A. Marciniak-Czochra, Mechanochemical symmetry breaking in Hydra aggregates, Biophysical Journal, vol. 108, pp. 2396-2407, 2015.

²M. Mercker, D. Hartmann, and A. Marciniak-Czochra, A Mechanochemical Model for Embryonic Pattern Formation: Coupling Tissue Mechanics and Morphogen Expression, PLOS ONE, vol. 8, p. e82617, 2013.

³F. Brinkmann, M. Mercker, T. Richter, and A. Marciniak-Czochra, Post-Turing tissue pattern formation: Advent of mechanochemistry, PLOS Computational Biology, vol. 14, p. e1006259, 2018.

Mechanochemical models for pattern formation in biological tissues

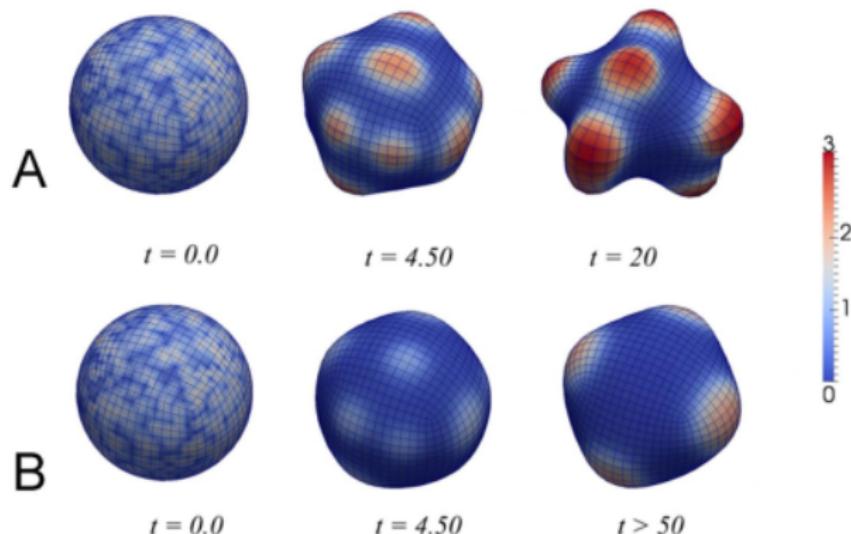


Figure 6: Simulations of spontaneous tissue pattern formation starting from stochastically distributed morphogen on a sphere at different time steps for a strong (A) and weak (B) coupling between curvature and morphogen. Reproduced from [2].

Mechano-chemical models

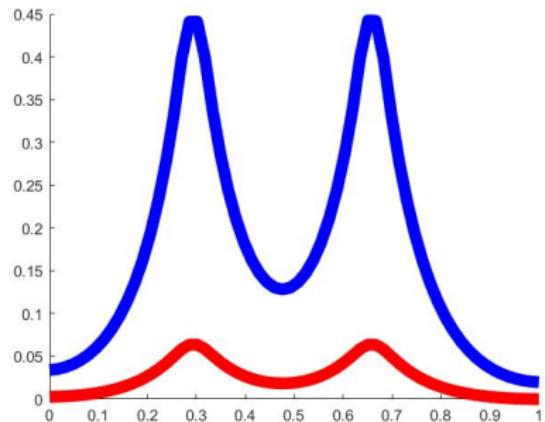
- Mechano-chemical models explain the spontaneous pattern formation by a feedback loop between tissue curvature and morphogen production.
- In this context, tissue curvature plays a role of a long-range inhibitor, thus making possible the formation of spatial concentration patterns with at least one morphogen on the surface.
- Here we consider one-dimensional mechano-chemical models, simplified versions of more realistic equations studied in [4,5,6].

⁴M. Mercker, A. Kothe, and A. Marciniak-Czochra, Mechanocellular symmetry breaking in Hydra aggregates, *Biophysical Journal*, vol. 108, pp. 2396-2407, 2015.

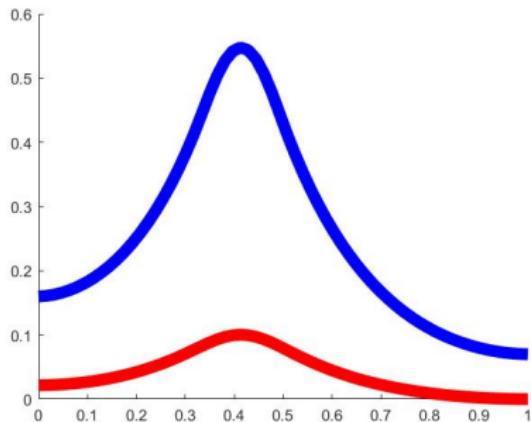
⁵M. Mercker, D. Hartmann, and A. Marciniak-Czochra, A Mechanocellular Model for Embryonic Pattern Formation: Coupling Tissue Mechanics and Morphogen Expression, *PLOS ONE*, vol. 8, p. e82617, 2013.

⁶F. Brinkmann, M. Mercker, T. Richter, and A. Marciniak-Czochra, Post-Turing tissue pattern formation: Advent of mechanochemistry, *PLOS Computational Biology*, vol. 14, p. e1006259, 2018.

Mechano-chemical models, 1D



(a)



(b)

Figure 7: Mechano-chemical patterns, obtained for different values of control parameters: tissue surface (red) and morphogen concentration (blue).

Mechano-chemical models, 1D

One can write a mechano-chemical model for the concentration u and surface ϕ as

$$\tau u_t(x,t) = -L \left[\frac{1}{\sqrt{1+u_x(x,t)^2}} \frac{\partial}{\partial x} \left(\frac{\kappa_X(u) - \bar{\kappa}_X(\phi)}{\sqrt{1+u_x(x,t)^2}} \right) - (\kappa(u) - \bar{\kappa}(\phi)) \kappa^2(u) + \lambda \kappa(u) \right], \quad (12)$$

$$\phi_t(x,t) = \frac{D}{\sqrt{1+u_x(x,t)^2}} \frac{\partial}{\partial x} \left(\frac{\phi_x(x,t)}{\sqrt{1+u_x(x,t)^2}} \right) + f(\kappa); \quad (13)$$

$$S(u) = \int_0^1 \sqrt{1+u_x(x,t)^2} dx = \int_0^1 \sqrt{1+u_x(x,0)^2} dx = \text{const.} \quad (14)$$

where $\kappa(x) = \frac{d}{dx} \left(\frac{u'(x)}{\sqrt{1+u'(x)^2}} \right)$ is the local curvature, $\bar{\kappa}(\phi) = \kappa_0 - \beta \phi$ is the preferred local curvature, $f(\kappa) = \frac{\zeta \kappa_{\geq 0}}{1+\kappa_{\geq 0}}$, $\kappa_{\geq 0} = \max((\kappa_0 - \kappa), 0)$ is the morphogen production function. For this model, we define $\theta = (D, \alpha)$.

Reaction-diffusion-ODE models, 1D

Another option for pattern formation is by coupled ODE-PDE systems [6]:

$$\begin{aligned} u_t &= -u - uw + m_1 \frac{u^2}{1 + ku^2}; \\ w_t &= D \frac{\partial^2}{\partial x^2} w - m_3 w - uw + m_2 \frac{u^2}{1 + ku^2}, \end{aligned} \tag{15}$$

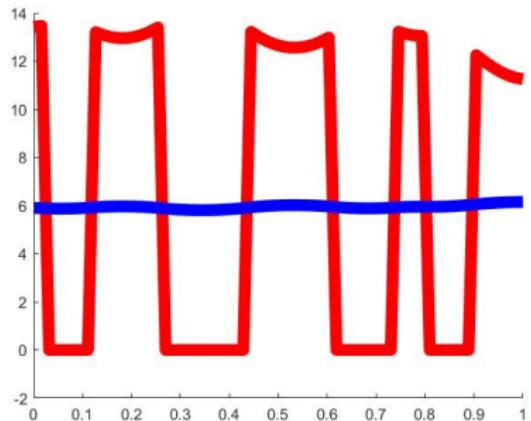
with Neumann boundary conditions:

$$\frac{\partial w}{\partial x} \Big|_{x=0,1} = 0. \tag{16}$$

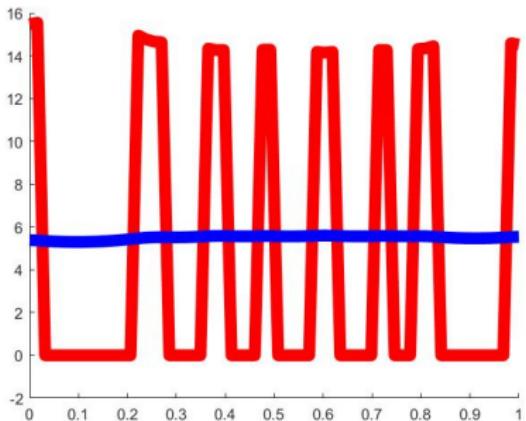
Here $u = u(x, t)$, $w = w(x, t)$, $x \in (0, 1)$, $D > 0$ is diffusion coefficient, $m_1, m_2, m_3 > 0$ and $k > 0$ are reaction parameters.

⁶S. Härting, A. Marciniak-Czochra, I. Takagi. Stable patterns with jump discontinuity in systems with Turing instability and hysteresis, Discrete & Continuous Dynamical Systems, 2017, 37 (2) : 757-800.

Reaction-diffusion-ODE models, 1D



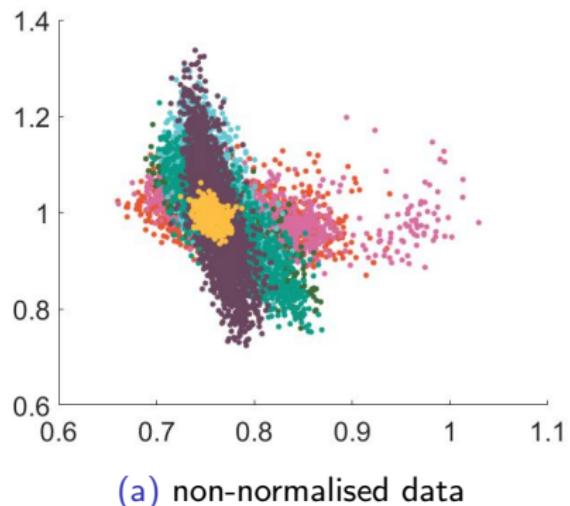
(a)



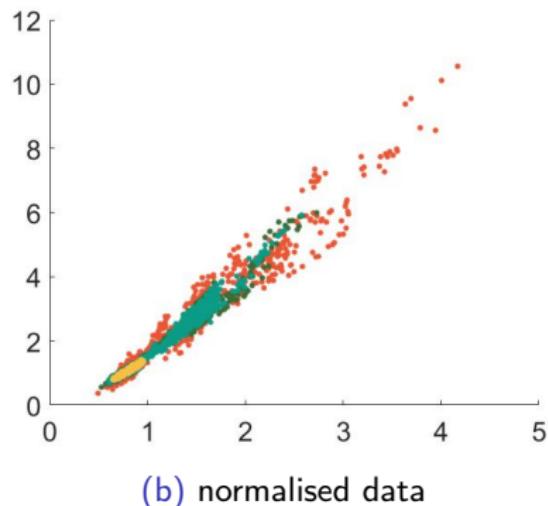
(b)

Figure 8: Far-from-equilibrium patterns, obtained for different values of control parameters: non-diffusive component (red) and diffusive component (blue).

Parameter posterior distributions (mechano-chemical model, $N_{set} = 200$)



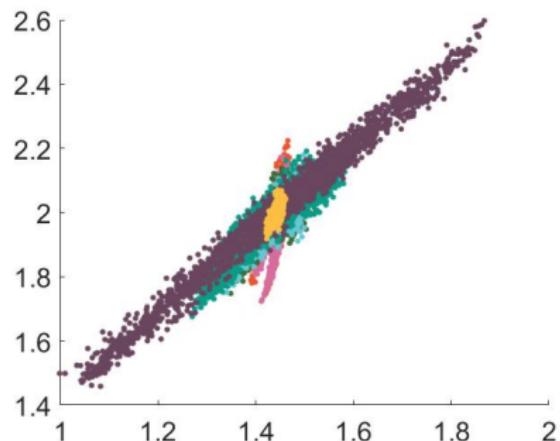
(a) non-normalised data



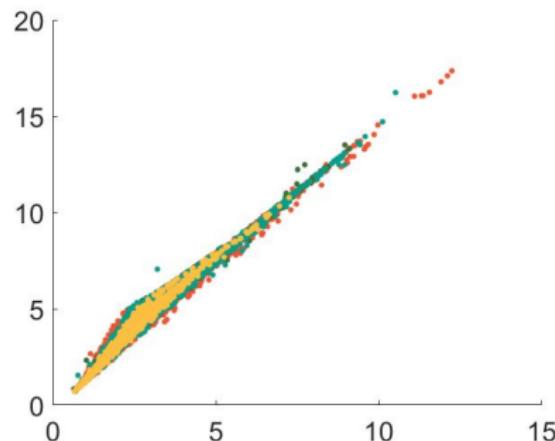
(b) normalised data

Figure 9: Posterior distribution of model parameters for the case of $N_{set} = 200$ data patterns.

Parameter posterior distributions (reaction-diffusion-ODE model, $N_{set} = 50$)



(a) non-normalised data



(b) normalised data

Figure 10: Posterior distribution of model parameters for the case of $N_{set} = 50$ data patterns.

Numerical integration of the models under study

- The usage of the approach requires many repeated simulations of the model under study.
- The performance of a numerical solver for the forward problem is crucial.
- Modern Graphical Processing Units (GPUs) are very effective at processing many independent tasks in parallel, which allows one to compute several model simulations at the same time.
- At the present work, GPU-accelerated algorithms for parallel simulation of the pattern formation models under study were implemented.

Conclusions

- The GSL/BSL approach with eCDF provides Gaussian likelihoods for situations where standard methods are not available.
- Applied to
 - Chaotic dynamics, SDE systems.
 - Deterministic reaction-diffusion systems with unknown/random initial values (Turing, Cahn-Hilliard)
 - Quantification of synchronization.
- Complex, high dimensional chaotic systems possible by
 - Parallel numerics, Parallel (ensemble) simulations, Parallel chains
 - Surrogate sampling for the parameter posterior
- Requires
 - GSL needs 'Large' training data. With limited data use the 'opposite' BSL approach, but with the same eCDF based likelihood construction – and more CPU
 - Scalar-valued mappings of original data needed for vector valued data.
- Next
 - Model discrimination for pattern formation
 - Other applications: Cahn-Hilliard, Cellular automata

References

- Haario H., Kalachev L., Hakkarainen J. Generalized correlation integral vectors: A distance concept for chaotic dynamical systems *Chaos*.—2015.—Vol. 25, No 6.
- Springer, Sebastian; Haario, Heikki; Shemyakin, Vladimir; Kalachev, Leonid; Shchepakin, Denis. Robust parameter estimation of chaotic systems. *Inverse Problems and Imaging*, 13(6), 1189-1212, DOI: 10.3934/ipi.2019053, 2019
- Kazarnikov A., Haario H. Statistical approach for parameter identification by Turing patterns. *Journal of Theoretical Biology*.—2020.—Vol. 501.—P. 110319.
- Ramona Maraia, Sebastian Springer, Heikki Haario, Janne Hakkarainen, Eero Saksman Parameter estimation of stochastic chaotic systems. *International Journal for Uncertainty Quantification*. 11 (2), 49–62, 2021.
- Sebastian Springer, Heikki Haario, Jouni Susiluoto, Aleksandr Bibov, Andrew Davis, and Youssef Marzouk: Efficient Bayesian inference for large chaotic dynamical systems. *Geosci. Model Dev.* 14 (7), 4319-4333,2021
- MCMC toolbox for Matlab: <https://github.com/mjlaine/mcmcstat>