Continuous time modelling of autocatalytic chemical reaction dynamics from sporadic inputs

Edward De Brouwer
ESAT-STADIUS
KU Leuven
Leuven, Belgium
edward.debrouwer@esat.kuleuven.be

Adam Arany
ESAT-STADIUS
KU Leuven
Leuven, Belgium
adam.arany@esat.kuleuven.be

Jaak Simm
ESAT-STADIUS
KU Leuven
Leuven, Belgium
jaak.simm@esat.kuleuven.be

Yves Moreau
ESAT-STADIUS
KU Leuven
Leuven, Belgium
moreau@esat.kuleuven.be

***Abstract*—Real-world time series are usually *sporadically* observed when collected outside of rigorous scientific experiments. The sampling is irregular across time and across dimensions, which breaks the typical assumption of most recurrent networks and dynamical process models. In this work, we propose to use our GRU-ODE-Bayes [1], a recently proposed model that posits a continuous latent process whose dynamics are driven by an ordinary differential equation parametrized with neural networks [2]. We apply this framework in the estimation of reaction dynamics from observational data. In particular, we show that our method is able to learn unstable autocatalytic chemical reactions within the large class of reactions driven by the Brusselator dynamical model [3], such as the Belousov–Zhabotinsky reaction in the presence of very few sporadic observation points.**

***Index Terms*—Neural Networks, Time Series, ODE, Chemical Reactions Dynamics**

I. Introduction and Motivation

Scientific research produces large quantity of time series data from various domains such as chemistry, astronomy, healthcare or climate science. Even though most statistical methods for analyzing this type of data assumes that signals are measured systematically at fixed constant time intervals, this assumption is often not hold in practice.

Indeed, due to some practical contingencies, much real-world data is *sporadic* (*i.e.*, the signals are sampled irregularly and not all signals are measured each time). A typical example is patient measurements, which are taken when the patient comes for a visit (*e.g.,* sometimes skipping an appointment) and where not every measurement is taken at every visit. Another example would be running lab experiments spread over several days when sensor errors can occur.

Modeling then becomes challenging as such data violates the main assumptions underlying traditional machine learning methods (such as recurrent neural networks). The most straightforward way to address this issue is to perform imputation, but this is often unsatisfactory as it results in biased estimates.

Recently, the GRU-ODE-Bayes model [1] proposed a novel way to address this type of sporadic time series for forecasting and classification. Based on the seminal Neural-ODE idea [2], it proposes a filtering approach that posits a *continuous* latent process that generates the observations. By framing the dynamics as an ordinary differential equation (ODE) parametrized by neural networks, the model is able to model complex dynamics and integrate it over arbitrary time intervals, therefore addressing the sporadicity issue in a natural way. Furthermore, it was shown that the continuity prior was providing extra performance when the continuity assumption is verified in practice, such as in patients clinical data prediction or weather forecasting.

This continuity assumption is also valid in chemical reaction dynamics. Indeed, chemical reactions are intrinsically continuous and can most of the time be expressed analytically as ODEs. However, reactions in multiple component mixtures are very complex and the true dynamic is often unknown. In this case, chemists try to approximate the true dynamic of the process by running some temporal experiments and learning the underlying dynamics of the reaction. What is more, as mentioned above, scientific practice sometimes lead to sporadic time series, as some concentration or variables are not measured continuously.

In this work, we propose to use GRU-ODE-Bayes to learn complex chemical reactions dynamics from sporadic data. We show that using very few data points, our model is able to learn the true chemical dynamics with high accuracy. The rest of

Edward De Brouwer is funded by a FWO-SB grant. Yves Moreau is funded by (1) Research Council KU Leuven: C14/18/092 SymBioSys3; CELSA-HIDUCTION, (2) Innovative Medicines Initiative: MELLODY, (3) Flemish Government (ELIXIR Belgium, IWT, FWO 06260) and (4) Impulsfonds AI: VR 2019 2203 DOC.0318/1QUATER Kenniscentrum Data en Maatschappij. Computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government – department EWI. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. The research has also been supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.2-16-2017-00013, Thematic Fundamental Research Collaborations Grounding Innovation in Informatics and Infocommunications).

the paper is organized as follows. In section II, we review the related works relevant to ours. Section III presents succinctly the GRU-ODE-Bayes method and the main assumptions behind the model. Finally, results of our method on the broad class of Brusselator reactions are shown in Section IV.

II. Related works

Our work naturally builds upon the machine learning and statistics literature on (sporadic) time series and on the learning of dynamical systems in chemistry and life sciences in general.

A. Temporal modelling of sporadic time series

The most straightforward approach to handle sporadic time series consists in data imputation. One then feed the observation mask and times of observations jointly to the recurrent network [4]–[6]. Despite some promising experimental results, this approach relies too much on the ability of the model to distinguish imputed and true data points. Some works have tried to address this limitation by introducing more meaningful data representation for sporadic time series [7], [8], like tensors [9], [10].

Gaussian processes (GP) are the most popular generative approach to deal with sporadic observations. They have been used for smart imputation before a RNN or CNN architecture [11], [12], or to derive informative representations of time series [8]. However, they usually suffer from high uncertainty outside the observation support, are computationally intensive, and learning the optimal kernel is tricky.

Most recently, the seminal work of Neural ODEs [2], [13] suggested a continuous version of neural networks that overcomes the limits imposed by discrete-time recurrent neural networks. Coupled with a variational auto-encoder (VAE) architecture, it proposed a natural way of generating irregularly sampled data. However, the VAE nature of the limits its expressivity when it comes to forecasting.

Our method also has connections to the Extended Kalman Filter (EKF) that models the dynamics of the distribution of processes in continuous time. However, the practical applicability of the EKF is limited because of the linearization of the state update and the difficulties involved in identifying its parameters.

B. Learning dynamics in chemistry and life sciences

Mathematical modeling with ordinary differential equations (ODEs) has a very long tradition in biology and ecology. Efforts to apply ODEs to understand population dynamics started already in the 18th century (see, e.g., Malthus's growth model) [14]. More recently, several optimization methods have been proposed to fit ODE models to observational data in domains such as metabolic models and more general biological pathways [15]–[17]. However, those methods require a heavily parametrized formulation of the ODE, motivated by a lot of prior assumptions and expert knowledge. In contrast, the method we propose here allows very flexible representation of the ODE, which alleviates the risk of misparametrization, at the expense of more data samples required for training the model. To the best of our knowledge, this is the first attempt to use ODE parametrized by neural networks to fit biological/chemical processes from observational data.

III. Methods

In this work we are interested in uncovering the dynamics of a D dimensional process $\mathbf{Y}(t) \in \mathbb{R}^D$ driven by an unknown (stochastic) differential equation (SDE) which is observed sporadically :

$$d\mathbf{Y}(t) = \mu(\mathbf{Y}(t))dt + \sigma(\mathbf{Y}(t))d\mathbf{W}(t), \tag{1}$$

where $d\mathbf{W}(t)$ is a Wiener process. The distribution of $\mathbf{Y}(t)$ then evolves according to the Fokker-Planck equation. We refer to the mean and covariance parameters of its probability density function (PDF) as $\mu_{\mathbf{Y}}(t)$ and $\Sigma_{\mathbf{Y}}(t)$. Our goal is to estimate those time evolving parameters conditioned on previous observations.

In practice we have access to N realizations of the SDE but the parameters of the SDE can change from one realization to another. Furthermore, for each realization $\mathbf{Y}(t)_i$ with $(i = 0, ..., N-1)$, we only observed sporadic measurements. That is, we have a vector $\mathbf{t_i^*} \in \mathbb{R}^{T_i}$ of times where the process is observed along with a mask $\mathbf{m} \in \{0,1\}^{T_i \times D}$ that indicates which dimension is observed at each observation time. We then have missingness both across time and across dimensions.

A. GRU-ODE-Bayes

GRU-ODE-Bayes [1] was recently proposed as a new filtering method to deal with *sporadic* time series. It assumes a continuous latent process $\mathbf{h}(t)$ that generates the observations $\mathbf{Y}(t)$ through some mapping $\mathbf{Y}(t) \sim f_{obs}(h(t))$.

It consists of two modules: GRU-ODE, responsible for learning the continuous dynamics of the latent process that generates the observations and GRU-Bayes, responsible for dealing with incoming observations and update the conditional current estimate of the latent process.

1) GRU-ODE: The GRU-ODE module parametrizes the dynamics of the latent process $\mathbf{h}(t)$ with an Neural-ODE inspired from the classical GRU module. We use the following parametric ODE:

$$\frac{d\mathbf{h}(t)}{dt} = (1 - \mathbf{z}(t)) \odot (\mathbf{g}(t) - \mathbf{h}(t)), \tag{2}$$

Where $\mathbf{z}(t)$ and $\mathbf{g}(t)$ are given as in the GRU equations :

$$\begin{aligned} \mathbf{r}_t &= \sigma(W_r \mathbf{x}_t + U_r \mathbf{h}_{t-1} + \mathbf{b}_r) \\ \mathbf{z}_t &= \sigma(W_z \mathbf{x}_t + U_z \mathbf{h}_{t-1} + \mathbf{b}_z) \\ \mathbf{g}_t &= \tanh(W_h \mathbf{x}_t + U_h(\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h), \end{aligned} \tag{3}$$

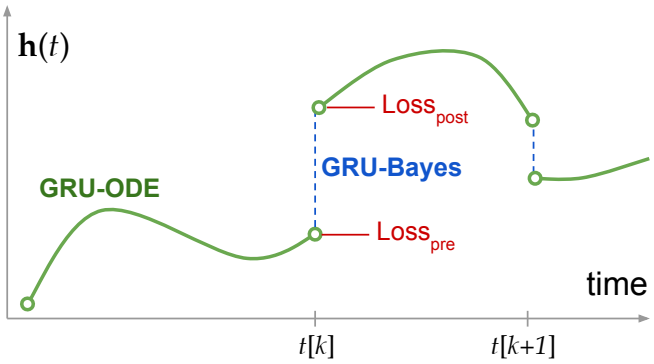

Fig. 1. GRU-ODE-Bayes uses GRU-ODE to evolve the hidden state between two observation times $t[k]$ and $t[k+1]$. GRU-Bayes processes the observations and updates the hidden vector $\mathbf{h}$ in a discrete fashion, reflecting the additional information brought in by the observed data.

2) GRU-Bayes: GRU-Bayes module is responsible for the update of the hidden state when new measurements are observed. As data comes in in packets, we allow the hidden process to jump to a new point in hidden space where it reflects more the newly observed data point as shown on Figure 1.

This update is performed by using a GRU cell that takes as input the previous hidden state and the current observation and then mimics a Bayesian update to set the hidden to a new value that matches the current observations :

$$\mathbf{h}(t_+) = \mathbf{GRU}(\mathbf{h}(t_-), f_{\text{prep}}(\mathbf{y}[k], \mathbf{m}[k], \mathbf{h}(t_-))), \quad (4)$$

3) End-to-end: At test time, the model performs predictions as suggested on Figure 1. We integrate the hidden process according to the GRU-ODE dynamics until the next observation (done with numerical integration). When an observation is reached, we process it with GRU-Bayes and update the hidden state. We then resume to GRU-ODE integration from the new initial point and continue until a next observation is reached. At each point in time, we can use $f_{obs}(.)$ to predict the distribution of the measurements.

B. The Brusselator

In order to demonstrate the capabilities of our approach, we choose to learn the dynamics of a range of reactions that can be modelled by the *Brusselator*. The Brusselator is a theoretical model for a type of autocatalytic chemical reactions. The type of chemical reactions it's modeling consists of two initial reactants A and B whose products X and Y are also catalysts of coupled reactions as shown below:

$$\begin{aligned} \mathrm{A} &\longrightarrow \mathrm{X} \\ 2\,\mathrm{X} + \mathrm{Y} &\longrightarrow 3\,\mathrm{X} \\ \mathrm{B} + \mathrm{X} &\longrightarrow \mathrm{Y} + \mathrm{D} \\ \mathrm{X} &\longrightarrow \mathrm{E} \end{aligned}$$

In the case of A and B being in large excess compared to products of above reactions, their concentration can be assumed constant in the solution and the following dynamics are obtained :

$$\frac{d}{dt}[X] = [A] + [X]^2[Y] - [B][X] - [Y] \quad (5)$$

$$\frac{d}{dt}[Y] = [B][X] - [X]^2[Y] \quad (6)$$

where the brackets stand for the concentration of the given chemical in the solution (*e.g.* in moles/l).

Equations 5 and 6 suggest that a fixed point is located at $[X] = A$ and $[Y] = [B]/[A]$. Importantly, this fixed point becomes unstable when:

Unstability condition :

$$B > 1 + A^2 \quad (7)$$

When the instability condition is satisfied, the concentrations X and Y start oscillating over time. An example of such oscillation is shown on figure 2.

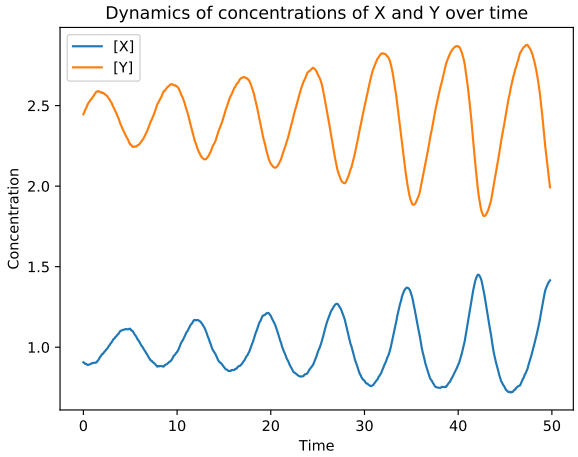

Fig. 2. Example of an unstable dynamic evolution of concentrations of products X and Y over time when A = 0.7 and B = 1.7.

The most well known example of this type of unstable autocatalytic reactions is the Belousov-Zhabotinsky reaction which can be created using a reducing agent (malonic acid) an oxidiser (bromate), and appropriate redox catalyst (eg.: manganese ion, cerium ion, or ferroin).

C. Dataset Generation

Because there is usually some measurement error and because of the uncertainty about the true nature of the hidden driving ODE, we generate trajectories of concentrations $[X]$ and $[Y]$ from a modified stochastic differential equation:

$$\begin{aligned} \frac{dx}{dt} &= 1 + (b+1)x + ax^2y + \sigma dW_1(t) \\ \frac{dy}{dt} &= bx - ax^2y + \sigma dW_2(t) \end{aligned} \quad (8)$$

Where $dW_1(t)$ and $dW_2(t)$ are correlated Brownian motions with correlation coefficient ρ. We simulate 1,000 trajectories

driven by the dynamics given in Eq. 8 with parameters $a = 0.3$ and $b = 1.4$ such that the ODE is unstable.

We simulate those trajectories over a time period of 50 seconds from which we sample sporadically N_{samp} observations with $N_{samp} \sim \mathcal{U}niform(0.38, 0.42) * T$ and $T = 50$. For each of those observations, we then randomly sample across dimensions to generate missing value dimension wise. We select both dimensions with probability 0.8 and only one of them with probability 0.2. The resulting dataset is then a sporadic time series of the the generative SDE process from Equation 8.

IV. Results

We train the GRU-ODE-Bayes model on the data described in Section III. We use 700 trajectories for training, 200 for validation and leave 100 out for testing. We used DOPRI-5 for the numerical integration of the ODE. Prediction of our model on a test trajectory is displayed on Figure 3.

We observe that at the start of the trajectory, as no measurement is observed, the model is uncertain about the future of the trajectory, which is illustrated by the very large confidence intervals before the first observations (shaded areas). However, as soon as an observation is reached, the model can update its latent state more accurately. The uncertainty remains high until more observations are reached, when the model can condition the expected trajectory on more information. Furthermore, we observe that the model is able to capture correlations between both dimensions of the process as indicated by the red arrow on the figure, where dimension 2 is updated even if only information about dimension 1 is gathered.

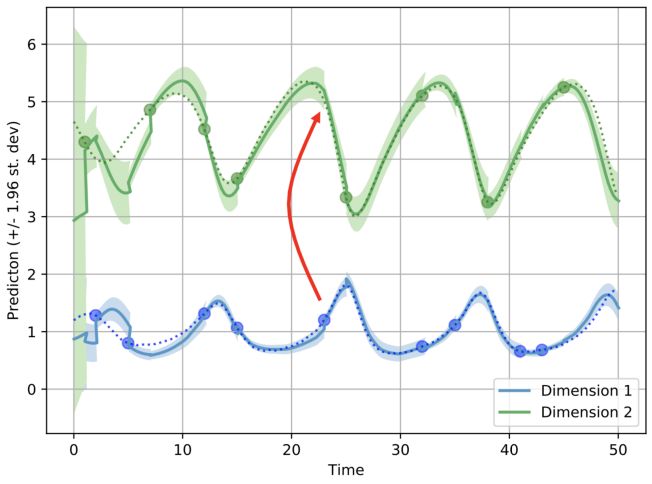

Fig. 3. Predictions of GRU-ODE-Bayes on an unknown realization of the BXLator process of Equation 8. Green and Blue are the two different concentrations we try to model over time. Solid lines represent the predicted means, shaded areas are the confidence intervals (1.96 standard deviations). Red arrow points out the update of dimension 2 when only dimension 1 is observed, showing that the model is able to detect meaningful correlation between both dimensions.

Table I further displays mean square errors (MSE) and negative log-likelihoods obtained on the test sets. They are computed by using the first 35 seconds of the processes as observable data and predictions are made on the remaining hidden observations between $t = 35$ and $t = 50$. We compare it with the Neural-ODE model as well as a Gaussian Process. We observe that our method clearly outperforms the competitors both in terms of MSE and log-likelihood. We motivate this performance by 1) the information bottleneck imposed in the Neural-ODE and 2) the lack of strong extrapolation capabilities in Gaussian processes. Furthermore, compared to a Gaussian process approach that would scale quadratically with the number of observations and dimensions, our model is much more efficient and scales linearly with the number of observations.

TABLE I
Test performance results

Model	MSE	Neg-Loglik
GRU-ODE-Bayes	0.11	−0.62
Neural-ODE	0.28	0.91
Gaussian Process	0.33	1.09

V. Conclusion

In this paper, we showed the fitness of the GRU-ODE-Bayes method to assess complex chemical reactions dynamics from sporadic observational data. We showed that our method was able to accurately learn the dynamics of the broad range of reaction dynamics modeled by the Brusselator equation. In order to make the problem more realistic, we added Browian motion noise to the dynamics, making every realization of the temporal process different from one another. Our approach outperformed a selected subset of baselines and seem to be suited for long term prediction of the concentrations (see Figure 3).

Those encouraging results motivate us to make this method more useful for the chemical community. In particular, an interesting application would be to be able to predict the constants a and b from a single experiment time series, allowing the chemist to infer the concentration of reactants A and B from the temporal measurements of X and Y. From preliminary experiments we expect that more time series realizations would be required to learn this task accurately. This is left for future work.

References

[1] E. De Brouwer, J. Simm, A. Arany, and Y. Moreau, "Gru-ode-bayes: Continuous modeling of sporadically-observed time series," *arXiv preprint arXiv:1905.12374*, 2019.

[2] T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," in *Advances in Neural Information Processing Systems, 2018*, 2018.

[3] I. Prigogine, *From being to becoming*. Freeman, 1982.

[4] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, p. 6085, 2018.

[5] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor ai: Predicting clinical events via recurrent neural networks," in *Machine Learning for Healthcare Conference*, 2016, pp. 301–318.

[6] Z. C. Lipton, D. Kale, and R. Wetzel, "Directly modeling missing data in sequences with rnns: Improved classification of clinical time series," in *Machine Learning for Healthcare Conference*, 2016, pp. 253–270.

[7] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun *et al.*, "Scalable and accurate deep learning with electronic health records," *npj Digital Medicine*, vol. 1, no. 1, p. 18, 2018.

[8] M. Ghassemi, M. A. Pimentel, T. Naumann, T. Brennan, D. A. Clifton, P. Szolovits, and M. Feng, "A multivariate timeseries modeling approach to severity of illness assessment and forecasting in icu with sparse, heterogeneous clinical data." in *AAAI*, 2015, pp. 446–453.

[9] E. De Brouwer, J. Simm, A. Arany, and Y. Moreau, "Deep ensemble tensor factorization for longitudinal patient trajectories classification," *arXiv preprint arXiv:1811.10501*, 2018.

[10] J. Simm, A. Arany, P. Zakeri, T. Haber, J. K. Wegner, V. Chupakhin, H. Ceulemans, and Y. Moreau, "Macau: Scalable bayesian factorization with high-dimensional side information using mcmc," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2017, pp. 1–6.

[11] J. Futoma, S. Hariharan, and K. Heller, "Learning to detect sepsis with a multitask gaussian process rnn classifier," *arXiv preprint arXiv:1706.04152*, 2017.

[12] M. Moor, M. Horn, B. Rieck, D. Roqueiro, and K. Borgwardt, "Temporal convolutional networks and dynamic time warping can drastically improve the early prediction of sepsis," *arXiv preprint arXiv:1902.01659*, 2019.

[13] Y. Rubanova, R. T. Chen, and D. Duvenaud, "Latent odes for irregularly-sampled time series," *arXiv preprint arXiv:1907.03907*, 2019.

[14] D.-E. Gratie, B. Iancu, and I. Petre, "Ode analysis of biological systems," in *International School on Formal Methods for the Design of Computer, Communication and Software Systems*. Springer, 2013, pp. 29–62.

[15] S. M. Baker, K. Schallau, and B. H. Junker, "Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models," *Journal of integrative bioinformatics*, vol. 7, no. 3, pp. 254–262, 2010.

[16] C. G. Moles, P. Mendes, and J. R. Banga, "Parameter estimation in biochemical pathways: a comparison of global optimization methods," *Genome research*, vol. 13, no. 11, pp. 2467–2474, 2003.

[17] P. Mendes and D. Kell, "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation." *Bioinformatics (Oxford, England)*, vol. 14, no. 10, pp. 869–883, 1998.