

METHODOLOGY ARTICLE

Open Access



Initial state perturbations as a validation method for data-driven fuzzy models of cellular networks

Lidija Magdevska^{1,2*} , Miha Mraz¹, Nikolaj Zimic¹ and Miha Moškon¹

Abstract

Background: Data-driven methods that automatically learn relations between attributes from given data are a popular tool for building mathematical models in computational biology. Since measurements are prone to errors, approaches dealing with uncertain data are especially suitable for this task. Fuzzy models are one such approach, but they contain a large amount of parameters and are thus susceptible to over-fitting. Validation methods that help detect over-fitting are therefore needed to eliminate inaccurate models.

Results: We propose a method to enlarge the validation datasets on which a fuzzy dynamic model of a cellular network can be tested. We apply our method to two data-driven dynamic models of the MAPK signalling pathway and two models of the mammalian circadian clock. We show that random initial state perturbations can drastically increase the mean error of predictions of an inaccurate computational model, while keeping errors of predictions of accurate models small.

Conclusions: With the improvement of validation methods, fuzzy models are becoming more accurate and are thus likely to gain new applications. This field of research is promising not only because fuzzy models can cope with uncertainty, but also because their run time is short compared to conventional modelling methods that are nowadays used in systems biology.

Keywords: Fuzzy logic, Model validation, Data-driven modelling, Dynamic modelling, MAPK signalling pathway, Circadian clock

Background

Computational models are depictions of reality that help us understand biological systems and direct experimental work in the field of systems biology [1]. A diverse range of methods for building models is available nowadays, with data-driven approaches playing an important role in cases where a large amount of experimental data exists and where prior knowledge of the system's structure is limited. A major advantage of these methods is that they can incorporate data directly without the need for expert knowledge to interpret the data, as their aim is to find correlations between data attributes [2, 3].

With experimental data, a certain level of measurement error appears [4]. A promising approach to dealing with this problem are Bayesian networks that allow the incorporation of qualitative data into the structure of the network, the likelihood function and the prior probability distribution of Bayes' rules [5], with a drawback that the prior probability distribution may sometimes not be available [6]. An alternative approach is fuzzy logic.

Fuzzy logic is an extension of traditional Boolean logic. The concept of a linguistic variable provides a means of approximate characterization of phenomena which are too complex or too ill-defined to be applicable in conventional quantitative terms [7]. To build a model, for each variable its term-set, the collection of linguistic (fuzzy) values, and a membership function are defined. Additionally, a set of fuzzy terms in the form of 'IF-THEN' rules is constructed, defining the relations between linguistic

*Correspondence: lm4828@student.uni-lj.si

¹ Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia

² Faculty of Mathematics and Physics, University of Ljubljana, Jadranska ulica 19, 1000 Ljubljana, Slovenia



variables [8]. Fuzzy models of cellular networks have been presented in [3, 6, 9–12].

Fuzzy models contain a large amount of parameters, hence they are susceptible to over-fitting. Additionally, it is possible that simulation results on small testing datasets fit the modelled system equally well for models with different sets of parameter values and topologies. This is especially likely in case of data-driven models as algorithms that build them do not account for the biological system's topology and may as such find a completely unsuitable solution. It is therefore important to expand the validation dataset in a way that helps us distinguish between accuracies of models with different topologies.

Computational models are typically validated on available experimental datasets and data that is collected from experiments that are performed after the establishment of the model. Models of signalling pathways often assume that the system's response only depends on the stimulus concentration [6, 13, 14], while they ignore the initial state of the system at the time of stimulation of the pathway. On the other hand protein concentrations are known to vary between cells and inside the same cell in different time points from 15 to 30% of their mean value [15]. This suggests that perturbations of protein initial concentrations could provide a successful method for fuzzy model validation.

First we apply our validation method to two fuzzy models of the classical cascade of the mitogen-activated protein kinase – MAPK. It is the most studied pathway from the MAPK signaling cascade family and coordinates many cellular activities in eukaryotic cells, such as gene expression, mitosis, metabolism, survival, apoptosis, and differentiation [16]. In cases where this signalling pathway is damaged, diseases such as cancer, Alzheimer's and Parkinson's disease may occur [17].

Later we apply the method to two fuzzy models of the mammalian circadian clock – CC, a timing system that forms rhythmic changes of processes in the body, with a period close to 24 h, allowing organisms to adapt to the cyclic changes in their habitats [18]. The disruption of this clock may cause a variety of pathologies, including cardiovascular and inflammatory diseases, cancer, and depression [19–22].

Many models have been built to analyse the dynamics of both systems. These models, however, use conventional computational biology methods [23–32] that have a long execution time and cannot deal with uncertain data.

Methods

Training, testing and validation datasets

Training, testing and validation sets for the MAPK signalling pathway were generated from the model presented in [23]. The model is based on ordinary differential

equations (ODEs) and was run in MATLAB for a time span of 30 min using the built-in ode45 function, with data being collected once per minute. Training and testing data were generated with constant initial conditions and variation of the epidermal growth factor – EGF (stimulus) concentration. All perturbations of the EGF concentration were inside the range that was experimentally tested in [23]. The validation set was generated by random perturbations of both initial conditions and EGF concentration. Training set of the mammalian CC was generated from the findings published in [32] following the recommendations of [33]. As test and validation datasets the raw data measured in liver under dark-dark conditions [32] were used.

Data-driven fuzzy models

In this article, two algorithms for building fuzzy models are used. Both algorithms use Zadeh-Mamdani fuzzy rules [34] that are of the form

$$\text{IF } x \text{ is } \tilde{A} \text{ THEN } y \text{ is } \tilde{B}, \tag{1}$$

where $(x \text{ is } \tilde{A})$ and $(y \text{ is } \tilde{B})$ are two fuzzy terms. The input variable x belongs to the fuzzy set \tilde{A} with the membership function value $\mu_{\tilde{A}}(x)$, and the output variable y belongs to the fuzzy set \tilde{B} with the membership function value $\mu_{\tilde{B}}(y)$. A general form of this rule that allows us to use an arbitrary number of input and output variables is

$$\begin{aligned} &\text{IF } x_1 \text{ is } \tilde{A}_1 \text{ AND } x_2 \text{ is } \tilde{A}_2 \text{ AND } \dots \text{ AND } x_{k_1} \text{ is } \tilde{A}_{k_1} \\ &\text{THEN } y_1 \text{ is } \tilde{B}_1 \text{ AND } y_2 \text{ is } \tilde{B}_2 \text{ AND } \dots \text{ AND } y_{k_2} \text{ is } \tilde{B}_{k_2}. \end{aligned} \tag{2}$$

For input and output variables we assume a Gaussian membership function that is defined with a mean value c and standard deviation σ , and is calculated from the expression

$$\mu_{\tilde{A}}(x) = e^{-\frac{(x-c)^2}{2\sigma^2}}. \tag{3}$$

For defuzzification of output variables, the center of gravity (COG) method [35] is used. The crisp value R' of a result of processing R that is described with a continuous membership function $\mu_{\tilde{R}}(y)$ equals

$$R' = \frac{\int_0^\infty y \mu_{\tilde{R}}(y) dy}{\int_0^\infty \mu_{\tilde{R}}(y) dy}. \tag{4}$$

Additionally, we assume that the next state of the system only depends on the previous state and the value of the stimulus.

Fuzzy c-means clustering algorithm (FCM)

The fuzzy c-means clustering algorithm (FCM) [36] is a basic fuzzy algorithm for clustering that searches for a

fuzzy partition $U = [u_{ik}]$ of data collection by minimising the generalised least squares functional

$$J_m(X, U, v) = \sum_{k=1}^N \sum_{i=1}^c u_{ik}^m d^2(x_k, v_i), \quad (5)$$

where $X = \{x_1, x_2, \dots, x_N\} \subset \mathbb{R}^n$ is a set of data, c the number of clusters in the set X ($2 \leq c < N$), $m \geq 1$ the degree of fuzzification to remove noise from data, d a distance function, U the fuzzy partition of set X , and $v = [v_i]$ the vector of cluster centres. The minimisation is run iteratively under the following conditions:

$$0 \leq u_{ik} \leq 1; 1 \leq i \leq c, 1 \leq k \leq N, \quad (6)$$

$$0 < \sum_{k=1}^N u_{ik} \leq n; 1 \leq i \leq c, \quad (7)$$

$$\sum_{i=1}^c u_{ik} = 1; 1 \leq k \leq N. \quad (8)$$

After each iteration, centres v_i and membership degrees u_{ik} are updated using the following procedure:

$$v_i = \frac{\sum_{k=1}^N u_{ik}^m x_k}{\sum_{k=1}^N u_{ik}^m}; 1 \leq i \leq c, \quad (9)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d(x_k, v_i)}{d(x_k, v_j)} \right)^{\frac{2}{m-1}}}; 1 \leq k \leq N, 1 \leq i \leq c. \quad (10)$$

For a fuzzy model with n input and m output variables, its learning with FCM uses $(n + m)$ -dimensional vectors as data, where each vector contains known values of input and expected values of output variables at given learning inputs. These data are then clustered in c groups with every group representing one fuzzy rule. Membership functions of fuzzy variables are determined from the groups' centres.

In the case of a cellular network model the input variables are concentrations of chemical species, while the output variables are the changes in concentrations of chemical species in two consecutive measurements. The change of concentration of the stimulus is ignored, as we assume that it is constant throughout the whole simulation time span. Since the training and testing datasets contain absolute concentration values, the learning method determines the changes, while the final model computes absolute values from input values and fuzzy model outputs.

This learning method is performed using the MATLAB function `genfis3`. Since its results are non-deterministic, the method is run 10 times and the model with the smallest error on the training set is selected for further observations.

Multi-attribute fuzzy time series method

Fuzzy time series is a prediction model that allows modelling dynamic processes in which linguistic values are observed. The model assumes that an observation in a time point is the result of observations from the past [37]. One of the procedures to build a fuzzy time series is the multi-attribute fuzzy time series method [38], later denoted as MAFTS. It consists of four steps:

- 1 The clustering of time series $S(t)$ into c clusters using FCM to identify patterns,
- 2 The ranking of each cluster and fuzzification of time series $S(t)$ to a fuzzy time series $F(t)$,
- 3 The determination of fuzzy rules,
- 4 The prediction of new data and defuzzification of results.

Data used for clustering is a set of concentrations of chemical species. The data of each chemical species is clustered separately to determine membership functions of the corresponding variable. Mean values of the Gaussian membership functions are determined as cluster centres obtained by FCM, while standard deviations are set to a constant percentage (3.5% in case of the MAPK signalling pathway and 0.8% in case of the CC) of the length of the interval on which a fuzzy variable is defined, in order to reduce the number of parameters that have to be learnt. Since membership functions for each protein are determined separately, linguistic names can be given to linguistic values. Each fuzzy variable gets either 3 or 5 fuzzy values denoted *low*, *medium*, and *high* (with 5 fuzzy values also *very low*, and *very high*), so that their mean values correspond to the linguistic meaning of the linguistic values. The number of fuzzy values per variable was set as in [6, 10], but could be extended in case of inaccuracy of the built model or reduced in case of over-fitting. The domain of a fuzzy variable is defined as a closed interval from 0 to the maximum value achieved by the variable on the training data.

Data points are fuzzified so that the fuzzy value with the maximal membership function value is chosen for each fuzzy variable. For each pair of consecutive data points, one fuzzy rule is determined. Fuzzy values of the fuzzy variables at the earlier time point are included in the IF part of the rule, and the fuzzy values at the later time point in the THEN part of the rule. Input and output variables of the fuzzy model are hence concentrations of chemical species. The stimulus concentration is not predicted as we assume that it is constant through the whole simulation time span.

The MATLAB function `fc` is used to cluster protein concentrations. Since its results are non-deterministic and it sometimes returns results of numeric type NaN, learning is repeated until a valid numeric result for cluster centres is obtained.

Model evaluation metric

Model accuracy is evaluated using a mean absolute error (MAE)

$$\text{MAE} = \frac{\sum_{i=1}^n \text{abs}(\epsilon_i)}{n}, \quad (11)$$

and a root mean square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n \epsilon_i^2}{n}}, \quad (12)$$

where n denotes the number of test instances and ϵ_i the prediction error of the i -th test instance [39]. The prediction error is measured as the average normalized difference between the true values and the predicted values of a component (variable) within a test instance. Each component was normalized by the maximal value of its domain.

Results and discussion

In order to gather validation data for dynamic models, experimental data needs to be sampled in a series of time-points after perturbations of experimental conditions. An appropriate design of time-series experiments is difficult and may contain redundant information leading to the inefficient use of experimental resources [40]. An alternative approach for model validation is therefore a comparison with existing models that allows us to sample validation data of arbitrary size. This is especially useful when accurate models exist, but are too slow to be effectively incorporated in experimental work.

Fuzzy model of the MAPK signalling pathway

We generated two data-driven fuzzy models of the MAPK signalling pathway from the same training dataset. The first model was generated using FCM with 20 clusters and the second model with MAFTS with 5 fuzzy values per variable. Both models simulate the dynamics of the MAPK signalling pathway by iterative runs of the inference system. Given an initial condition and EGF concentration models returns a time series of 30 consecutive states of the system.

We are searching for a model that describes the dynamics of a signalling pathway. In contrast to some prediction models, where, given a state, the model has to produce an accurate prediction of the next state (i.e. the state in the next time point), later called *next state prediction*, we attempt to find a model that given an initial condition and a stimulus concentration, predicts an accurate series of consecutive states. We call the later a *whole time series prediction*.

MAE and RMSE were hence calculated on two testing sets and two validation sets. One of the sets used the predictions of the next state from a given state, while the other predicted a series of states from a given initial state.

The errors of the generated fuzzy models were of similar size for the testing sets that included the results of a whole time series, while the next state prediction was better using the model generated with FCM (Table 1). At this stage of validation, we could thus assume that the model generated with FCM is either more accurate than the model generated with MAFTS or that they are both approximately as accurate.

We then generated validation data with initial state perturbations to validate our assumption. Validation data were generated with two distinct approaches. In the first case only the initial state was randomly selected so that it belonged to the domain on which the models are defined, while the EGF concentration was randomly taken from the set of EGF concentrations that occur in training data. In the second case both the initial state and stimulus concentration were randomly selected from the domain. MAE and RMSE were measured as before.

We found out that in both cases errors of the model generated with FCM increased notably compared to the testing data (Tables 2 and 3), while the errors of the model generated with MAFTS increased only slightly. The main reason for the increase of the whole series prediction error of the model generated with FCM is that the model estimates the difference in concentration and not the concentration itself, allowing the concentration prediction to increase above the maximum value of the domain. Once the input variables of the FCM model are outside the domain, the results are unlikely to be in the domain, leading to large errors. Such errors are likely to occur whenever replacing ODE models with fuzzy models with an aim to speed them up.

Our results show that the model generated with MAFTS is much more accurate than the model generated with FCM, although we were unable to form this conclusion from the testing datasets generated by exclusively EGF concentration perturbations. These findings suggest that perturbations of initial conditions can simplify the process of model validation as even a small dataset can sometimes eliminate an inaccurate fuzzy model.

Table 1 Test sets errors

	FCM model	MAFTS model
MAE (next state)	0.07	0.14
MAE (whole series)	0.76	0.24
RMSE (next state)	0.02	0.10
RMSE (whole series)	0.47	0.15

MAE and RMSE measured on models generated with FCM and MAFTS with respect to the testing sets where either the next state or a whole time series is predicted

Table 2 Errors on validation sets with initial state perturbations

	FCM model	MAFTS model
MAE (next state)	$0.20 * 10^3$	0.15
MAE (whole series)	$1.41 * 10^3$	0.24
RMSE (next state)	$3.28 * 10^3$	0.22
RMSE (whole series)	$8.67 * 10^3$	0.31

MAE and RMSE measured on models generated with FCM and MAFTS with respect to the validation sets with initial state perturbations where either the next state or a whole time series is predicted

Fuzzy models of the mammalian circadian clock

The observations of the models of the MAPK signalling pathway might suggest that sensitivity to perturbations is a feature of FCM models. For this reason we generated two data-driven fuzzy models of the mammalian circadian clock from the same training dataset using MAFTS. In the first case we used 3 fuzzy values per variable, and in the second case we used 5 fuzzy values per variable. Both models again simulate the dynamics of the network by iterative runs of the inference system.

Korenčić et al. [32] suggests that the effect of transcription factors on gene expression at a given time point can be modelled as an effect of gene expression levels at earlier time points. This delay corresponds to the time needed for post-transcriptional modifications and differs between genes. In order to integrate this approach to MAFTS, the previous state was defined as a set of gene expression levels before delay time points. The initial condition in this case is therefore a series of four states, as the largest delay observed in [32] corresponds to four hours. In each model a series of 24 states corresponds to the 24 h day cycle. As with the previous case study we attempt to find a model that, given an initial condition, predicts an accurate series of consecutive states, however, in this case it is more important that the system keeps oscillating than to obtain low MAE or RMSE. Without any initial state perturbations both models produced oscillations with a 24 h period.

Perturbations of initial conditions were up to 1% of their value, which is less than the differences between measure-

Table 3 Errors on validation sets with initial state and stimulus concentration perturbations

	FCM model	MAFTS model
MAE (next state)	$0.29 * 10^3$	0.16
MAE (whole series)	$2.02 * 10^3$	0.25
RMSE (next state)	$4.35 * 10^3$	0.23
RMSE (whole series)	$11.5 * 10^3$	0.31

MAE and RMSE measured on models generated with FCM and MAFTS with respect to the validation sets with initial state and stimulus concentration perturbations where either the next state or a whole time series is predicted

ments in different mice at the same time point in [32], meaning that they should not affect the dynamics of the system. As Fig. 1 shows the model with 5 fuzzy values per variable keeps oscillating, while the model with only 3 fuzzy values stops oscillating after 10 h of simulation.

While in this case the inaccuracy is not a consequence of over-fitting, we show that initial state perturbations can also help as a testing method to determine the minimal number of fuzzy values needed to accurately describe the dynamics of a cellular network.

Discussion

The size of available datasets limits many validation methods not only due to the complexity of the experimental work, but also due to the long runtime of simulations of large ODE and partial differential equations (PDE) models that are still the most popular approach for the depiction of signalling pathways and gene regulatory networks. This also holds true for the reference ODE model used in this study, but we were still able to generate a validation dataset of sufficient size to disprove the fuzzy model generated with FCM.

This limitation should, however, not prevent one from using the proposed method, as simulations of fuzzy models are much faster than the corresponding ODE reference models and several fuzzy models can be validated using the same validation datasets. Additionally, our method can be extended to cases where appropriate experimental data or any type of an accurate quantitative model of the observed biological system is available.

Conclusions

Validation of computational models of biological systems is often problematic, as only small experimental datasets are available for comparison. In this paper we provided a description of an approach that helps in eliminating inaccurate fuzzy data-driven models through initial state perturbations of a dynamic system. We demonstrated the method's applicability by comparing two data-driven fuzzy models of the MAPK signalling cascade and two data-driven fuzzy models of the mammalian CC, where we successfully detected an over-fitted model. With the improvement of validation methods fuzzy models are not only becoming more accurate, but are also becoming a more promising alternative to conventional modelling methods as they can cope with uncertain data and can predict outputs quickly. The presented method can be also extended to the validation of fuzzy dynamic models of a diverse spectrum of biological systems, providing an opportunity for new applications of fuzzy logic to systems biology. The latter can gain importance through data-driven models built directly from experimental data or as a way to speed up existing models that are accurate but too slow for frequent usage.

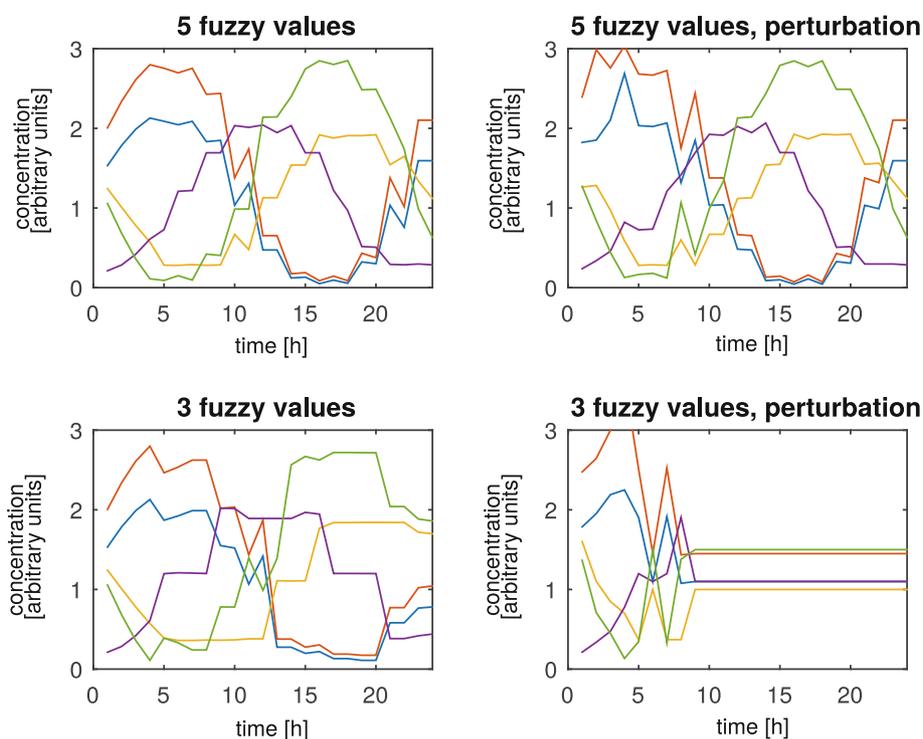


Fig. 1 Comparison of fuzzy models of the circadian clock. Simulation results of both fuzzy models. After initial state perturbations the model with 5 fuzzy values per variable keeps oscillating, while the model with only 3 fuzzy values stops. Without initial state perturbations both models showed oscillations with a period of approximately 24 h

Abbreviations

CC: Circadian clock; EGF: Epidermal growth factor; FCM: Fuzzy c-means clustering algorithm; MAE: Mean absolute error; MAFTS: Multi-attribute fuzzy time series method; MAPK: Mitogen-activated protein kinase; ODE: Ordinary differential equations; RMSE: Root mean square error

Funding

The research was partially supported by the scientific-research programme Pervasive Computing (P2-0359) financed by the Slovenian Research Agency in the years from 2013 to 2023, by the basic research project CholesterOR in metabolic liver diseases (J1-9176) financed by the Slovenian Research Agency in the years from 2018 to 2021, and a scholarship of the City of Ljubljana. Neither funding body played any role in the design of the study, nor collection, analysis, and interpretation of data, nor in writing the manuscript.

Availability of data and materials

All code is available for download at: <https://github.com/magdevska/fuzzy-model-validation>.

Authors' contributions

LM designed the method, performed the experiments, and wrote the manuscript. LM and MMo devised the study. MMo supervised the study. MMo, MMr and NZ provided critical feedback and helped shape the research, analysis and manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 26 February 2018 Accepted: 10 September 2018

Published online: 21 September 2018

References

- Patterson EA, Whelan MP. A framework to establish credibility of computational models in biology. *Prog Biophys Mol Biol*. 2017;129:13–19.
- Janes KA, Lauffenburger DA. A biological approach to computational models of proteomic networks. *Curr Opin Chem Biol*. 2006;10(1):73–80.
- Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, Lauffenburger DA. Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. *PLoS Comput Biol*. 2009;5(4):1000340.
- Tahera K, Ibrahim RN, Lochert PB. A fuzzy logic approach for dealing with qualitative quality characteristics of a process. *Expert Syst Appl*. 2008;34(4):2630–8.
- Lucas PJ. Bayesian network modelling through qualitative patterns. *Artif Intell*. 2005;163(2):233–63.
- Huang Z, Hahn J. Fuzzy modeling of signal transduction networks. *Chem Eng Sci*. 2009;64(9):2044–56.
- Gaweda AE, Zurada JM. Data-driven linguistic modeling using relational fuzzy rules. *IEEE Trans Fuzzy Syst*. 2003;11(1):121–34.
- Virant J. *Design Considerations of Time in Fuzzy Systems*, vol. 35. Dordrecht: Springer; 2000.
- Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA. Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Computational Biol*. 2011;7(3):1001099.
- Bordon J, Moškon M, Zimic N, Mraz M. Fuzzy logic as a computational tool for quantitative modelling of biological systems with uncertain

- kinetic data. *IEEE/ACM Trans Comput Biol Bioinform.* 2015;12(5):1199–205.
11. Woolf PJ, Wang Y. A fuzzy logic approach to analyzing gene expression data. *Physiol Genomics.* 2000;3(1):9–15.
 12. Resson H, Wang D, Varghese RS, Reynolds R. Fuzzy logic-based gene regulatory network. In: *The 12th IEEE International Conference on Fuzzy Systems, 2003. FUZZ'03.* Piscataway: IEEE; 2003. p. 1210–5.
 13. Apgar JF, Toettcher JE, Endy D, White FM, Tidor B. Stimulus design for model selection and validation in cell signaling. *PLoS Comput Biol.* 2008;4(2):30.
 14. Puchrová T. *Modelling and experimental validation of signalling pathways with relevance to homologous mammalian systems.* Pilsen: University of West Bohemia; 2015.
 15. Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, Rosenfeld N, Danon T, Perzov N, Alon U. Variability and memory of protein levels in human cells. *Nature.* 2006;444(7119):643–6.
 16. Roux PP, Blenis J. ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiol Mol Biol Rev.* 2004;68:320–44.
 17. Kim EK, Choi E-J. Pathological roles of MAPK signaling pathways in human diseases. *Biochim Biophys Acta.* 2010;1802:396–405.
 18. Reppert SM, Weaver DR. Molecular analysis of mammalian circadian rhythms. *Annu Rev Physiol.* 2001;63(1):647–76.
 19. Oishi K, Ohkura N, Amagai N, Ishida N. Involvement of circadian clock gene clock in diabetes-induced circadian augmentation of plasminogen activator inhibitor-1 (pai-1) expression in the mouse heart. *FEBS Lett.* 2005;579(17):3555–9.
 20. Cao Q, Gery S, Dashti A, Yin D, Zhou Y, Gu J, Koeffler HP. A role for the clock gene per1 in prostate cancer. *Cancer Res.* 2009;69(19):7619–25.
 21. McCarthy MJ, Welsh DK. Cellular circadian clocks in mood disorders. *J Biol Rhythm.* 2012;27(5):339–52.
 22. Labrecque N, Cermakian N. Circadian clocks in the immune system. *J Biol Rhythm.* 2015;30(4):277–90.
 23. Kochańczyk M, Kocieniewski P, Kozłowska E, Jaruszewicz-Błońska J, Sparta B, Pargett M, Albeck JG, Hlavacek WS, Lipniacki T. Relaxation oscillations and hierarchy of feedbacks in MAPK signaling. *Sci Rep.* 2017;7:38244.
 24. Levchenko A, Bruck J, Sternberg PW. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc Natl Acad Sci.* 2000;97(11):5818–23.
 25. Kamioka Y, Yasuda S, Fujita Y, Aoki K, Matsuda M. Multiple decisive phosphorylation sites for the negative feedback regulation of SOS1 via ERK. *J Biol Chem.* 2010;285:33540–8.
 26. Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol.* 2002;20(4):370–5.
 27. Bhalla U. S. Signaling in small subcellular volumes. I. Stochastic and diffusion effects on individual pathways. *Biophys J.* 2004;87(2):733–44.
 28. Yamada S, Taketomi T, Yoshimura A. Model analysis of difference between EGF pathway and FGF pathway. *Biochem Biophys Res Commun.* 2004;314(4):1113–20.
 29. Leloup J-C, Goldbeter A. Toward a detailed computational model for the mammalian circadian clock. *Proc Natl Acad Sci.* 2003;100(12):7051–6.
 30. Forger DB, Peskin CS. A detailed predictive model of the mammalian circadian clock. *Proc Natl Acad Sci.* 2003;100(25):14806–11.
 31. Mirsky HP, Liu AC, Welsh DK, Kay SA, Doyle FJ. A model of the cell-autonomous mammalian circadian clock. *Proc Natl Acad Sci.* 2009;106(27):11107–12.
 32. Korenčič A, Bordyugov G, Lehmann R, Rozman D, Herzog H, et al. Timing of circadian genes in mammalian tissues. *Sci Rep.* 2014;4:5782.
 33. Hughes ME, Abruzzi KC, Allada R, Anafi R, Arpat AB, Asher G, Baldi P, De Bekker C, Bell-Pedersen D, Blau J, et al. Guidelines for genome-scale analysis of biological rhythms. *J Biol Rhythm.* 2017;32(5):380–93.
 34. Mamdani EH, Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man-Machine Stud.* 1975;7(1):1–13.
 35. Zimmermann H-J. *Fuzzy Set Theory and Its Applications.* New York: Springer; 2001.
 36. Bezdek JC, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. *Comput Geosci.* 1984;10(2-3):191–203.
 37. Song Q, Chissom BS. Fuzzy time series and its models. *Fuzzy Sets Syst.* 1993;54(3):269–77.
 38. Cheng C-H, Cheng G-W, Wang J-W. Multi-attribute fuzzy time series method based on fuzzy clustering. *Expert Syst Appl.* 2008;34(2):1235–42.
 39. Sammut C, Webb GI. *Encyclopedia of Machine Learning.* New York: Springer; 2011.
 40. Hecker M, Lambeck S, Toepfer S, Van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models—a review. *Biosystems.* 2009;96(1):86–103.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

