Fuzzy Stochastic Petri Nets for Modeling Biological Systems with Uncertain Kinetic Parameters

Fei Liu[1]*, Monika Heiner[2], Ming Yang[1]*

1 Control and Simulation Center, Harbin Institute of Technology, Harbin, 150080 China, **2** Department of Computer Science, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, 03013 Germany

* liufei@hit.edu.cn (FL); myang@hit.edu.cn (MY)

Abstract

Stochastic Petri nets (SPNs) have been widely used to model randomness which is an inherent feature of biological systems. However, for many biological systems, some kinetic parameters may be uncertain due to incomplete, vague or missing kinetic data (often called fuzzy uncertainty), or naturally vary, e.g., between different individuals, experimental conditions, etc. (often called variability), which has prevented a wider application of SPNs that require accurate parameters. Considering the strength of fuzzy sets to deal with uncertain information, we apply a specific type of stochastic Petri nets, fuzzy stochastic Petri nets (FSPNs), to model and analyze biological systems with uncertain kinetic parameters. FSPNs combine SPNs and fuzzy sets, thereby taking into account both randomness and fuzziness of biological systems. For a biological system, SPNs model the randomness, while fuzzy sets model kinetic parameters with fuzzy uncertainty or variability by associating each parameter with a fuzzy number instead of a crisp real value. We introduce a simulation-based analysis method for FSPNs to explore the uncertainties of outputs resulting from the uncertainties associated with input parameters, which works equally well for bounded and unbounded models. We illustrate our approach using a yeast polarization model having an infinite state space, which shows the appropriateness of FSPNs in combination with simulation-based analysis for modeling and analyzing biological systems with uncertain information.

Introduction

To achieve a system-level understanding of biological systems, modeling and simulation play a crucial role as they permit to represent, explain, analyze and predict the system behavior from a holistic point of view. Randomness is an inherent property of many biological processes induced by gene regulatory networks, signal transduction networks or metabolic networks [1]. Randomness may arise from the external environment, intrinsic noise or low number of molecules, and may dramatically affect the behavior of these networks [2, 3]. To deal with randomness, stochastic modeling methods have been used, e.g., chemical master equations, stochastic differential equations [1], stochastic Pi-calculus [4], stochastic process algebra [5], or stochastic

Petri nets (SPNs) [6]. These approaches address the stochastic aspect of biological systems and thus describe their behavior more accurately than deterministic approaches like ordinary or partial differential equations. Among them, SPNs have recently become a promising tool. They have been widely used for modeling and analyzing stochastic biological systems.

The construction of an SPN model for a biological system basically involves two steps: building the model structure, which represents the system's underlying biochemical reactions and their interactions, and specifying the kinetic parameters of transitions, which determine the stochastic rates of biochemical reactions. Compared with the former step, the latter one is usually more challenging. On the one hand, the information available for some kinetic parameters may be either incomplete, vague or merely qualitative (i.e., no data); in these circumstances, these kinetic parameters cannot be measured or estimated accurately (we speak of fuzzy uncertainty). On the other hand, kinetic parameters may naturally vary, e.g., between different individuals, experimental conditions, etc. (often called variability). Such fuzzy uncertainty or variability might appropriately be addressed by fuzzy sets, which permit to turn vagueness of information into mathematical representations by means of fuzzy numbers and fuzzy operations on them. However, SPNs currently cannot cope with such biological systems where both randomness and fuzziness coexist.

Therefore, in this paper we will apply a specific type of Petri nets, fuzzy stochastic Petri nets (FSPNs), to model and analyze biological systems with uncertain kinetic information. FSPNs combine SPNs and fuzzy sets, thereby taking into account both randomness and fuzziness of biological systems. For a biological system, SPNs model the randomness, while fuzzy sets model the kinetic parameters with fuzzy uncertainty or variability by associating each kinetic parameter with a fuzzy number instead of a crisp real value. With fuzzy sets, uncertain kinetic parameters of biological systems can be characterized by experts based on their knowledge or belief, presumably supported by experimental data. This approach could be useful in many areas, such as for finding suitable, i.e. biological meaningful start values or value ranges for parameter estimation, or later for the evaluation of the results obtained by help of models, respectively. Moreover, if one wants to make progress even in the light of partially missing precise data, there is no way around to use expert belief (knowledge) to obtain a sophisticated guess which is in accordance with the modeled biological reality. FSPNs allow us to consider stochastic systems which—due to incomplete or vague knowledge—have uncertain parameter values, and to analyze the uncertainties of outputs (with output referring to a moment of the stochastic process) resulting from the uncertainties associated with the input parameters. Please note, in the whole paper we assume that "uncertainty" only relates to fuzziness and has nothing to do with stochasticity. Our efficient analysis approach builds on simulation, which works equally for bounded and unbounded models, i.e., for finite or infinite state spaces; with other words, the size of the state space does not matter.

We mainly consider the following scenario where our approach may be of help to cope with the challenges on hand. For a given biological system (phenomenon), there may exist different qualitative and quantitative experimental findings (data or models), each of which may describe complementary aspects of the system. The appropriate combination of all these findings to form a more detailed model is of great significance to gain a deeper understanding of the system [7, 8]. A simplistic approach might be to convert all of these qualitative and quantitative findings into qualitative descriptions and to build a model from a merely qualitative point of view. But it becomes more challenging the other way around. In this paper, we aim to use our approach for integrating qualitative experimental findings into a quantitative model and to explore the system to be studied from the quantitative point of view. This may be achieved by describing qualitative or semi-quantitative findings using fuzzy sets.

The main contributions of this paper are as follows.

- We apply a new modeling formalism, FSPNs, to cope simultaneously with both randomness and fuzziness of biological systems by combining SPNs and fuzzy sets. Thus, a biological system can be analyzed even if it has some parameters that are unknown or not precisely estimated.

- We present a simulation-based analysis method for FSPNs, and thus we can analyze the uncertainties of outputs caused by the uncertainties associated with the input parameters for both bounded or unbounded models independently of the size of the state space.

- Our approach permits to explore both the transient and steady state behavior, and can be easily varied to cope with different types of fuzzy numbers.

This paper is structured as follows. After a brief review of related work in Section 2, we describe fuzzy stochastic Petri nets in Section 3, followed by a simulation-based analysis method for FSPNs in Section 4. After that, we use an example of the yeast polarization model to illustrate our method in Section 5, followed by the discussions and conclusions in Section 6.

Related work

Stochastic Petri nets for modeling biological systems

Stochastic modeling is able to capture the randomness and fine grain behavior of biological systems which are not appropriately described by deterministic methods confining themselves to the averaged behavior. SPNs have been widely used for modeling stochastic biological systems since Gross et al. first applied stochastic Petri nets to molecular biology [9]. For example, Srivastava et al. [10] modeled the Escherichia coli stress circuit and performed stochastic kinetic analysis using SPNs. Marwan et al. [11] reconstructed the regulatory network controling the commitment and sporulation in *Physarum polycephalum* by use of hierarchically structured SPNs. Mura et al. [12] described the cell cycle in yeast with SPNs, and Lamprecht et al. [13] develop an SPN model for Ca^{2+} release sites consisting of a number of intracellular Ca^{2+} channels that exhibit stochastic Ca^{2+} excitability. Marwan et al. [14] investigated phosphate regulation in enteric bacteria by use of SPNs, while Castaldi et al. [15] gave an SPN model of the tissue factor induced coagulation cascade. Colored SPNs, an extension of SPNs by data structures as known from programming languages, have been used by Liu et al. [16] to discuss the modeling of membrane systems deploying an example of the virus infection, and by Pârvu et al. [17] to explore the stochasticity of bacterial colony growth in space with phase variable genes. See [6, 18, 19] and references therein for more applications. Like other quantitative modeling methods, SPNs rely on accurate kinetic data; in any case, the results shall be very carefully cross-validated.

Petri net based parameter estimation

Parameter estimation plays an important role in the construction of quantitative models. The aim is to tune the kinetic parameters to fit simulation results to in vivo/vitro experiment observations. A lot of work has been done to address this challenging problem, some of them within a Petri net framework. For example, Shaw et al. [20] proposed an automatic parameterization method for SPN models of biological networks using a distributed genetic algorithm, and illustrated its application using the stress response pathway in the bacterium E. coli. Koh et al. [21] gave a decompositional approach to parameter estimation in pathway models using hybrid functional Petri nets (HFPN). It exploits the structure of a large pathway model to decompose it into smaller components and then estimates the parameters independently. Donaldson and Gilbert [22] characterize the desired behavior of SPN models by properties expressed in a

Probabilistic Linear-time Temporal Logic with numerical constraints and combine Monte Carlo Model Checking with genetic algorithms to alter the parameters in the SPN model. Li et al. [23] utilized a data assimilation framework and model checking to estimate kinetic parameters of biological pathways, which were described by hybrid functional Petri net with extension (HFPNe).

In contrast, our approach in this paper differs from parameter estimation in the following way. Parameter estimation means to tune parameters and find crisp values to fit the simulation results to in vivo/vitro experiment observations, i.e., removing parameter uncertainties, while we want to derive the uncertainties of outputs which are caused by uncertain input parameters, i.e., keeping parameter uncertainties.

Fuzzy logic for modeling biological systems

Fuzzy logic is able to appropriately capture systems with only qualitative or incomplete quantitative information by describing the dynamic behavior of a system with a set of fuzzy rules. It has been applied for the modeling of a number of biological systems, see e.g., [24]. Fuzzy Petri nets, combining fuzzy logic and (qualitative) Petri nets, facilitate the use of fuzzy logic by help of the graphical representation coming with Petri nets. They have occasionally been used for biological modeling, see, e.g., [25, 26]. The advantage of fuzzy logic or fuzzy Petri nets is that they can be deployed to model biological systems if the kinetic data are incomplete or even unavailable; the disadvantage is that they are usually used for constructing qualitative models, rather than quantitative ones.

To overcome this, researchers started to combine quantitative Petri nets with fuzzy logic. For example, Valette et al. [27] introduced a class of fuzzy time Petri nets by associating transitions with fuzzy enabling durations. Ding et al. [28] gave a class of fuzzy timed Petri nets in which a fuzzy number is associated with a transition. Tüysüz et al. [29] combined stochastic Petri nets with fuzzy sets and applied it to the modeling of a flexible manufacturing cell with fuzzy parameters. But to the best of our knowledge, there are so far no applications of these fuzzy quantitative Petri nets to the modeling of biological systems.

In this paper, we combine SPNs and fuzzy logic in order to quantitatively model biological systems with uncertain kinetic parameters. We will use a similar way to associate the rate of a stochastic transition with a fuzzy number as in [29]. However, in [29], they analytically determine the steady state distribution of an SPN model by generating the corresponding continuous time Markov chain (CTMC). As this approach requires a finite state space, they only can deal with bounded Petri nets. But, there are many biological models that are unbounded. Unbounded SPNs have an infinite state space, thus the CTMC cannot be exhaustively constructed. Consequently, all numerical approaches requiring a closed representation of the CTMC do not work. Even if the model is bounded, the size of the state space often exceeds the computational resources required for numerical analyses; see [30] for details.

Therefore, in this paper we will adopt a simulation approach to tackle both bounded and unbounded models in order to overcome the limitations of the method developed in [29]. Moreover, our approach immediately reveals both the transient and steady state behavior, and can be easily varied to cope with different types of fuzzy numbers.

Fuzzy stochastic Petri nets

Stochastic Petri nets

Petri nets [31, 32] (see Fig 1 for an example) are weighted, directed, bipartite multigraphs, which consist of places, transitions and arcs. In biological models, places may represent chemical species or any kind of compounds, e.g., genes, mRNA, proteins or protein complexes, while
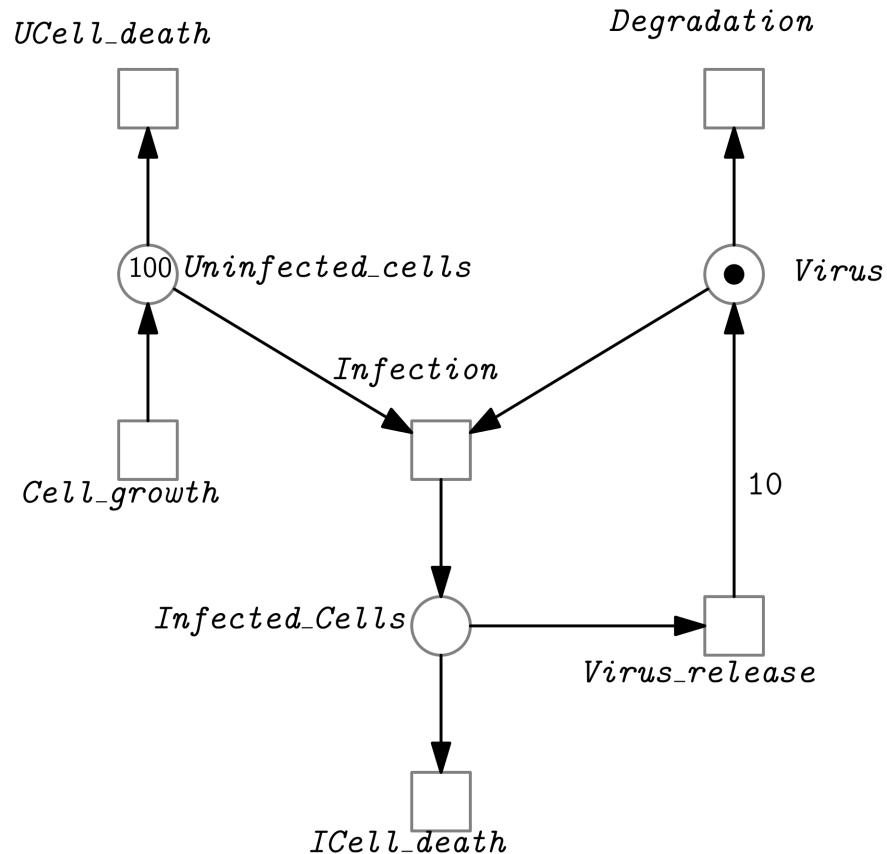
Fig 1. A Petri net model for the infection of healthy cells by a virus [34]. Cells grow or die. The virus may enter a healthy cell (UCell) and infect it (ICell). Then the virus starts the replication of itself and more viruses are released (note the arc weight of 10). Besides, infected cells may die and viruses may degrade.

transitions represent any kind of chemical reactions, e.g. association, disassociation, translation, translocation or transcription [33]. For a chemical reaction, its precursors correspond to the preplaces of a transition, while products correspond to its postplaces. The preplaces (postplaces) of a transition are also called the input (output) places of the transition. The arcs connect places with transitions, or transitions with places. The weight of an arc indicates its multiplicity, reflecting e.g., stoichiometries of a chemical reaction. The weight 1 is usually not explicitly labeled. A place may contain an arbitrary (natural) number of tokens, represented as black dots if the number of tokens is smaller, or a natural number if it is larger. A specific distribution of tokens over all places of a Petri net is called a marking, which represents a state of the Petri net. The formal definition of a (qualitative) Petri net is given in the following.

A Petri net PN is a five-tuple $N = <P, T, F, f, M_0>$, where

- P is a finite, non-empty set of places.

- T is a finite, non-empty set of transitions.

- $F \subseteq (P \times T) \cup (T \times P)$ is a finite set of directed arcs.

- $f : F \rightarrow \mathbb{N}^+$ is a function that assigns a positive integer to each arc $a \in F$.

- $M_0 : P \rightarrow \mathbb{N}_0$ gives the initial marking, which assigns a non-negative integer to each place $p \in P$.

A transition is called enabled if each of its preplaces contains tokens whose number is greater than or equal to the weight of the corresponding arcs. If a transition is enabled, it may fire. The firing of a transition moves tokens from its preplaces to its postplaces according to the weights of the corresponding arcs, updating the current marking to a new reachable one. The repeated firing of transitions establishes the behavior of a net. The set of markings reachable from the initial marking form the state space of the net. These reachable markings and transitions between them constitute the reachability graph of the net.

Stochastic Petri nets are a popular extension of Petri nets [6, 35], and more specifically can be considered as an extension of Time Petri nets (TPN) [36]. In TPN, deterministic firing delays are associated to transitions. Once a transition is enabled, it may fire after the elapsing of the associated deterministic delay time, provided it did not loose the concession to fire while waiting. In SPN, these firing delays are considered as random variables with a negative exponential probability distribution function (PDF). For this purpose, a stochastic hazard function $h(t, \theta)$ (or firing rate) is assigned to each transition $t \in T$, whereby

$$h(t, \theta) : \mathbb{N}_0^{|{}^\bullet t|} \to \mathbb{R}^+. \tag{1}$$

θ is a rate constant (also called kinetic parameter) which serves as parameter for the PDF, $\mathbb{R}^+$ is the set of all non-negative real numbers, and ${}^\bullet t$ denotes the preplaces of t. A classical type of hazard functions are the Gillespie propensities: $h(t, \theta) = \theta \cdot \#\{\text{reactant combinations}\}$. The stochastic hazard function $h(t, \theta)$ relates to a random variable Y_t, defined by the following exponential probability distribution (written as $Y_t \sim Exp(h(t, \theta))$):

$$F_{Y_t}(\tau) = 1 - e^{-h(t,\theta)\cdot\tau}, \tau \geq 0 . \tag{2}$$

The semantics of a stochastic Petri net is defined by a continuous time Markov chain (CTMC), a class of stochastic processes, which is constructed from the reachability graph of the underlying qualitative Petri net by labeling the arcs between the states (markings) with the corresponding transition rates, i.e., all states are still reachable; for more details see [6].

Formally, the stochastic process underlying a stochastic Petri net can be described as a collection of random variables indexed by means of time indices τ from a time set T', given a probability space (Ω, Σ, Pr), where Ω is a sample space, Σ is a σ-algebra of subsets of Ω and Pr is a probability measure, that is,

$$\{X_\tau : \tau \in T'\} \tag{3}$$

where each X_τ (where X is the marking of the stochastic Petri net) is a random variable on Ω. See e.g., [37] for more details.

For example, if we assign the rate functions given in Table 1 to the model in Fig 1, then we obtain an SPN model; see Fig 2 for a stochastic simulation plot. The model assumes infinite cell

Table 1. Rate functions for the virus infection SPN model [34]. Here, the place names like Uninfected_cells refer to the markings of these places.

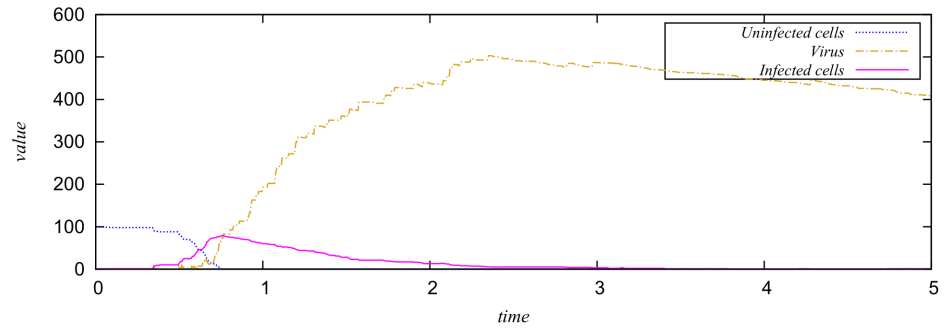| transition t	rate function $h(t, \theta)$	kinetic parameter θ
Cell_growth	θ_1	1.0
UCell_death	$\theta_2 *Uninfected\_cells$	0.1
Infection	$\theta_3 *Uninfected\_cells *Virus$	1.0
Virus_release	$\theta_4 *Infected\_Cells$	1.0
ICell_death	$\theta_5 *Infected\_cells$	0.5
Degradation	$\theta_6 *Virus$	0.1

doi:10.1371/journal.pone.0149674.t001

Fig 2. A stochastic simulation plot for the virus infection SPN model shown in Fig 1 with the rate functions given in Table 1.

growth (its corresponding transition does not have a precondition). Thus, the corresponding CTMC is infinite. The averaged amount of uninfected cells depends on the ratio of the firing rates of the pre- and posttransitions.

Fuzzy sets

Fuzzy theory was introduced by Zadeh [38] to deal with uncertainty associated with imprecision or vagueness rather than with randomness. A fuzzy set $\tilde{\xi}$ on a universal set $\mathbb{X}$ is defined by its membership function

$$\mu_{\tilde{\xi}} : \mathbb{X} \to [0, 1], \tag{4}$$

which assigns to each element $x \in \mathbb{X}$ a real value $\mu_{\tilde{\xi}}(x)$ in $[0, 1]$. The support of a fuzzy set $\tilde{\xi}$ is the crisp subset of all elements in $\mathbb{X}$ with non-zero membership degree, i.e.,

$$supp(\tilde{\xi}) = \{x | \mu_{\tilde{\xi}}(x) > 0, x \in \mathbb{X}\}. \tag{5}$$

The core of a fuzzy set $\tilde{\xi}$ is the crisp subset of all elements in $\mathbb{X}$ with membership degree equal to 1, i.e.,

$$core(\tilde{\xi}) = \{x | \mu_{\tilde{\xi}}(x) = 1, x \in \mathbb{X}\}. \tag{6}$$

The α-cut of a fuzzy set $\tilde{\xi}$ for a level $\alpha \in [0, 1]$ is the crisp subset of $\mathbb{X}$, containing all the elements whose membership degree is greater than or equal to a given α, i.e.,

$$\tilde{\xi}_{\alpha} = \{x | \mu_{\tilde{\xi}}(x) \geq \alpha, x \in \mathbb{X}, \alpha \in [0, 1]\}. \tag{7}$$

A fuzzy number is a special convex (that is, each of its α-cuts is a convex set) normalized fuzzy set defined on the real set $\mathbb{R}$. Among different fuzzy numbers, triangular and trapezoidal shapes are widely used. For example, a triangular fuzzy number, denoted by $\tilde{\xi} = (a, b, c), a \leq b \leq c$, is defined as follows (see Fig 3):

$$\mu_{\tilde{\xi}}(x) = \begin{cases} 0 & \text{if } x \leq a, \\ \dfrac{x - a}{b - a} & \text{if } a \leq x \leq b, \\ \dfrac{c - x}{c - b} & \text{if } b \leq x \leq c, \\ 0 & \text{if } x \geq c, \end{cases} \tag{8}$$
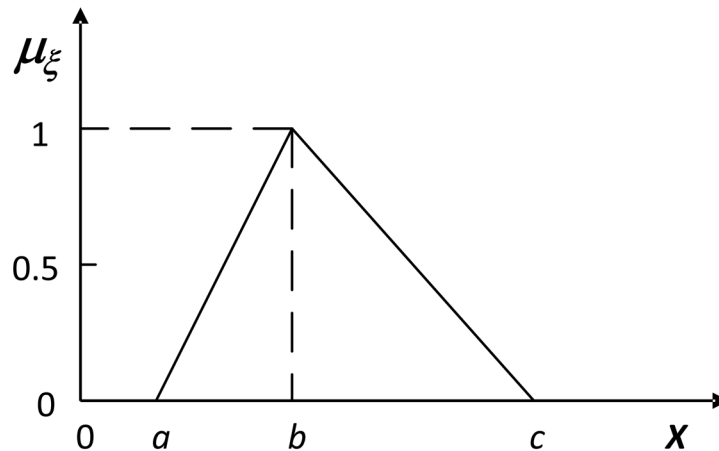
Fig 3. A triangular fuzzy number.

and its α-cut is simply $\tilde{\xi}_\alpha = [a + \alpha(b - a), c - \alpha(c - b)]$, for any $\alpha \in [0, 1]$. We then denote by Γ the set of fuzzy numbers whose lower bound is greater than 0.

Fuzzy stochastic Petri nets

Fuzzy stochastic Petri nets are an extension of SPNs by associating each transition $t \in T$ with a stochastic hazard function $h(t, \theta)$, where

$$h(t, \theta) : \mathbb{N}_0^{|{}^\bullet t|} \to \mathbb{R}^+ \cup \Gamma, \tag{9}$$

and the kinetic parameter θ is described by either a fuzzy number in Γ or a real value in $\mathbb{R}^+$. Please note that in our definition, we allow θ to be a real value if it is measured accurately, or a fuzzy number if it is uncertain or varies. When each fuzzy number θ in $h(t, \theta)$ is defuzzified or sampled to a crisp value, the FSPN turns into an SPN.

Formally, a fuzzy stochastic Petri net is a six-tuple $N = <P, T, F, f, v, M_0>$, where

- $<P, T, F, f, M_0>$ is a Petri net.

- $v: T \to H$ is a function that assigns a stochastic hazard function $h(t, \theta)$ to each transition $t \in T$, whereby $H := \bigcup_{t \in T}\{h(t, \theta)\}$ is the set of all stochastic hazard functions, $v(t) = h(t, \theta)$ for all transitions $t \in T$, and $h(t, \theta)$ is defined by Eq 9.

For example, if we assign the rate functions given in Table 2 to the model in Fig 1, then we obtain an FSPN model.

In contrast with the probability measure and random variables in the probability theory which allow us to deal with stochasticity, the possibility measure and fuzzy variables in the possibility theory [39] are deployed to describe the fuzzy uncertainty [40]. Suppose $(\Theta, P(\Theta), Pos)$ is a possibility space, where Θ is the sample space, $P(\Theta)$ is the power set of Θ, and Pos is the possibility measure on Θ. A mapping from Θ to the real set $\mathbb{R}$ defines a fuzzy variable. For example, a fuzzy number is a fuzzy variable. A random fuzzy variable ξ is defined as a mapping from the possibility space $(\Theta, P(\Theta), Pos)$ to a set of random variables [41]. For example, suppose $\xi \sim Exp(\lambda)$, where λ is a fuzzy variable; then ξ is a random fuzzy variable. For each $\theta \in \Theta$, if the expected value $E[\xi(\theta)]$ is finite, then $E[\xi(\theta)]$ is also a fuzzy variable in $(\Theta, P(\Theta), Pos)$.

Further, for a fuzzy stochastic Petri net, each X_τ in Eq 3 can be considered as a random fuzzy variable. Accordingly, a moment of X_τ like $E[X_\tau]$ can be considered as a fuzzy variable

Table 2. Rate functions for the virus infection FSPN model. Here, the place names like Uninfected_cells refer to the markings of these places.

transition t	rate function $h(t, \theta)$	kinetic parameter θ
Cell_growth	θ_1	(0.9, 1, 1.4)
UCell_death	$\theta_2 * Uninfected\_cells$	(0.05, 0.1, 0.2)
Infection	$\theta_3 * Uninfected\_cells * Virus$	(0.9, 1, 1.2)
Virus_release	$\theta_4 * Infected\_Cells$	(0.9, 1, 1.2)
ICell_death	$\theta_5 * Infected\_cells$	(0.4, 0.5, 0.6)
Degradation	$\theta_6 * Virus$	(0.05, 0.1, 0.15)

doi:10.1371/journal.pone.0149674.t002

(fuzzy moment). Therefore, the semantics of a fuzzy stochastic Petri net can be regarded as a fuzzy stochastic process (that is a mapping from the possibility space to the set of random stochastic processes) entirely determined by its "fuzzy probability distribution" (of the family of random fuzzy variables), which can be described by the set of fuzzy moments like $E[X_\tau]$.

We can further define an observation function or performance measure $f(X_\tau)$ based on X_τ, and thus a moment of $f(X_\tau)$ like $E[f(X_\tau)]$ is also a fuzzy variable, with which we can analyze a fuzzy stochastic Petri net. Therefore, for a fuzzy stochastic Petri net, according to the membership functions of fuzzy input parameters, we obtain the membership functions of outputs, as each output can be considered as a fuzzy variable.

FSPN provides an alternative approach to immediately study those kinds of systems where precise values for some parameters cannot be obtained for the time being. Thus, instead of idle waiting for complete (crisp) parameters, the fuzzy set approach permits to make progress building on expert belief. As to how to build a FSPN model, parameter estimation should always be the first choice to possibly obtain the precise values of most parameters, and fuzzy values are assigned by professionals to those parameters and situations only where experimental data are not available yet. Besides, to determine the values of uncertain or unknown kinetic parameters, we could adopt the following scheme. The professional experts of the field may use their special insight knowledge of the system to be studied or the available parameter estimation results, to specify the pessimistic value, the most possible value and the optimistic value for an uncertain or unknown parameter, which then form a fuzzy number for this parameter. Experts in this context are usually those most familiar with the modeled system or pathway under consideration.

Simulation and analysis

SPN models that are not bounded induce an infinite state space, and thus an infinite CTMC, which prevents the use of any analysis technique building on a finite CTMC representation. Simulation-based behavior exploration is the only possible approach, if the state space is too huge to be analyzed numerically or even infinite. In this case, the size of the state space one can reach depends on the number of simulations that are done and of course on the length of the simulation traces. In this paper, we will overcome this problem by a simulation-based analysis method. The general idea of this method is as follows.

According to Zadeh's extension principle [38], a fuzzy variable (number) can be represented as a union of its α-cuts. Therefore, if we decompose all fuzzy parameters into its α-cuts, and then run stochastic simulations at each α level, we obtain the α-cut for each output of interest. We then compose all the α-cuts and obtain the membership function for each output, which reflects the effect of the uncertainties of the input parameters. The whole procedure is given in Algorithm 1, which consists of the following three steps.

Algorithm 1 Simulation algorithm for an FSPN model.
Input: An FSPN model
Output: Membership functions of the performance measures of interest
1: **for** each α level α_j, $j = 1, 2, \ldots, J$ **do**
2:   **for** each fuzzy number $\tilde{\theta}_i$, $i = 1, 2, \ldots, I$ **do**
3:     Obtain its α-cut, represented as $(\tilde{\theta}_i)_{\alpha_j} = [L_i^j, U_i^j]$;
4:     Discretize each α-cut $[L_i^j, U_i^j]$ and obtain crisp values for each
parameter;
5:   **end for**
6:   **for** each combination of values for all parameters **do**
7:     Run stochastic simulations for M replications;
8:     **for** each performance measure Y_m, $m = 1, 2, \ldots, M$ **do**
9:      Compute the sample mean $\overline{Y}_m$;
10:      Compose all the α-cuts of Y_m to obtain its membership function;
11:     **end for**
12:   **end for**
13: **end for**

(1). For each fuzzy number $\tilde{\theta}_i$, $i = 1, 2, \ldots, I$, decompose its membership function into some
of its α-cuts using the same step size, denoted, e.g., by α_j, $j = 1, 2, \ldots, J$, where J is the number of the considered α levels; see Fig 4 for a demo. The step size of the α levels should be
determined with great care according to the given problem. A smaller step may achieve a
precise membership function for a given output, but may also cause a significant increase
in computational costs. The α-cut for each fuzzy number $\tilde{\theta}_i$ can be represented as
$(\tilde{\theta}_i)_{\alpha_j} = [L_i^j, U_i^j]$. After that we discretize each α-cut $[L_i^j, U_i^j]$ using different kinds of sampling methods, such as Latin Hypercube sampling (LHS), and obtain a set of crisp values
for each parameter. We denote by K the number of the crisp values after discretization.
To obtain crisp values (points), a simple division of the whole α-cut into a number of
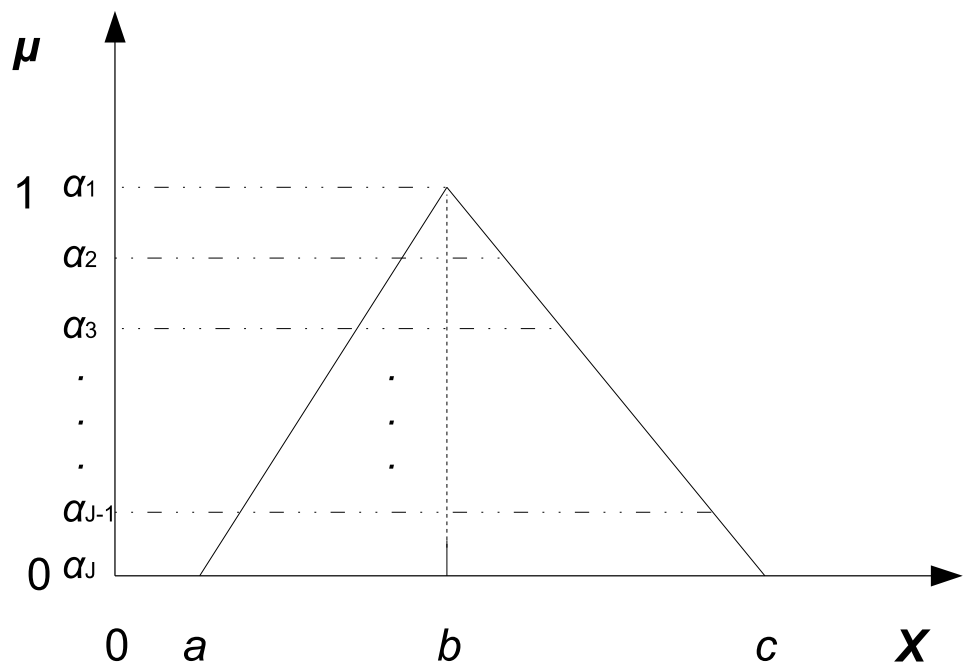smaller equally sized intervals is usually enough. At each α level α_j, we obtain all

Fig 4. Decomposition of a fuzzy number into its α-cuts.

combinations of values for all parameters. Please note that this approach can be applied regardless of the form of fuzzy numbers. But, for sake of simplicity, we have chosen in the following a rather simple version of fuzzy numbers, which can effortlessly be replaced by any more sophisticated fuzzy number scheme.

(2). For each combination, we apply Gillespie's stochastic simulation algorithm (SSA) [42] to produce simulation results, and then analyze some key performance measures (or outputs), denoted by Y_m, $m = 1, 2, \ldots, M$, where M is the number of performance measures of interest. For example, a performance measure could be the average number of molecules of a species at some time point.

For a transient simulation (i.e., a simulation which aims to study the transient behavior), we usually run a number of independent and identically distributed (i.i.d.) replications to compute the sample mean as the point estimator of the expectation for a performance measure. For example,

$$\overline{Y}_m = \frac{1}{N} \sum_{n=1}^{N} (Y_m)_n, \tag{10}$$

where $(Y_m)_n$ is the output data at simulation replication n, and N is the number of replications. Further, we can compute a confidence interval for the expectation; see, e.g., [43] for more details about transient simulation analysis. A crucial point here is to determine the required amount of simulation replications (or samples) in order to achieve a reliable estimation of the measure of interest. For this, we can use the idea of the confidence interval as described in [44]. The confidence interval relates to a confidence level, which usually has values of 90%, 95%, or 99%. For example, choosing the confidence level of 95% and the accuracy of 10^{-2}, we need to perform—according to Eq 4 in [44]—38,000 stochastic simulation replications. As this is usually computationally intensive due to the numerous numerical simulations, we have to decide on an appropriate number of simulation runs by choosing an acceptable confidence level and accuracy according to e.g., [44]. That is, sometimes, we have to choose a small confidence level (e.g., 90%) and low accuracy (e.g., 10^{-1}) in order to decrease the number of simulation runs.

For a steady state simulation (i.e., a simulation which aims to study the steady state behavior), the number of samples required to estimate a performance measure can be obtained using a single long simulation run by viewing the estimator as the long-run average level. But we have to ensure that the simulation reaches a stationary state and the effect of the initial values has been overcome. In particular, given the simulation observations $(Y_m)_i$, $i = 1, 2, \ldots, L$, where L is the length of a simulation run, we could use

$$\overline{Y}_m^l = \frac{1}{L - l} \sum_{i=l+1}^{L} (Y_m)_i, \tag{11}$$

to estimate the steady state mean of a performance measure. Here we adopt the most commonly used method given in [45] for eliminating the bias caused by the initial conditions, that is, identifying an index l from which the warmup period finishes and truncating the observations before the index l. See, e.g., [43] for more details about steady state simulation.

(3). The sample mean $\overline{Y}_m$ for each combination consists of the α-cut of a measure Y_m at level α_j. In this way, we obtain all the α-cuts of a measure Y_m.

According to Zadeh's extension principle [38], saying that a fuzzy variable can be represented as a union of its α-cuts, we can compose all the α-cuts of a measure Y_m to obtain its
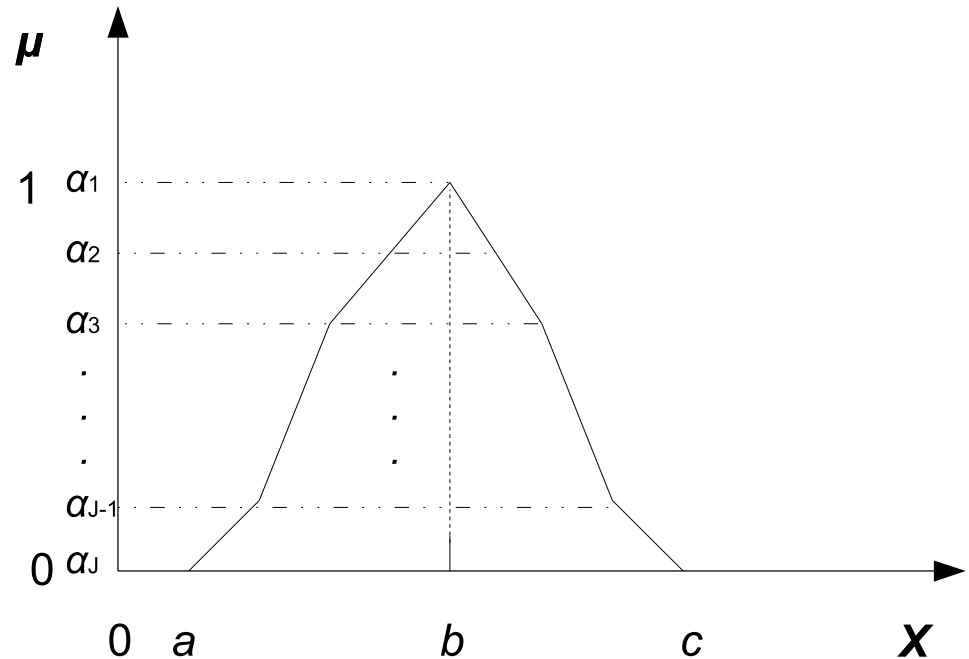
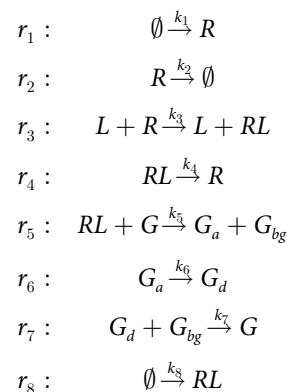**Fig 5. Composition of the *α*-cuts to obtain a fuzzy number.**

membership function, which reflects the effect of the uncertainties of parameters on the uncertainties of the measure; see Fig 5 for a demo.

Reproducibility. We use an extended version of Snoopy [46], a unifying, platform-independent and easily extensible Petri net tool, for modeling our FSPNs. Then, we deploy Matlab [47] for performing fuzzy and stochastic simulations by implementing the algorithm described above and for drawing all the plots given in this paper. The interested reader can download the extended version of Snoopy (Windows OS only), the Matlab code, and all the models used in the paper from http://dx.doi.org/10.6084/m9.figshare.1451230, which then will permit to reproduce our simulation results.

Example

We now illustrate our approach using the yeast polarization model describing the pheromone-induced G-protein cycle in *Saccharomyces cerevisiae* [48, 49], which consists of the following eight reactions.
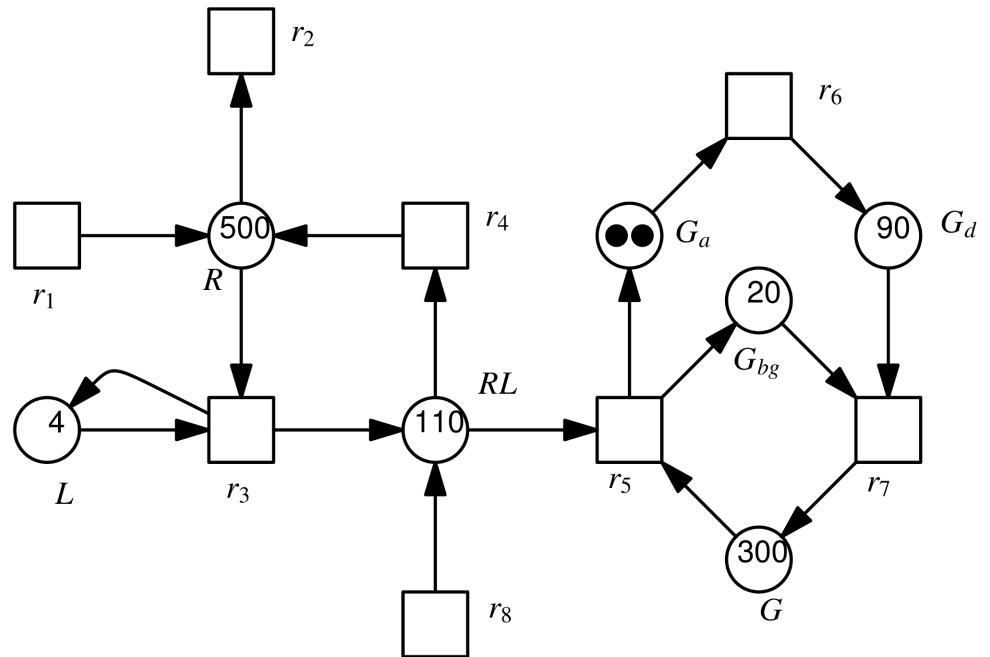
$$r_1 : \qquad \emptyset \xrightarrow{k_1} R$$

$$r_2 : \qquad R \xrightarrow{k_2} \emptyset$$

$$r_3 : \quad L + R \xrightarrow{k_3} L + RL$$

$$r_4 : \qquad RL \xrightarrow{k_4} R$$

$$r_5 : \quad RL + G \xrightarrow{k_5} G_a + G_{bg}$$

$$r_6 : \qquad G_a \xrightarrow{k_6} G_d$$

$$r_7 : \quad G_d + G_{bg} \xrightarrow{k_7} G$$

$$r_8 : \qquad \emptyset \xrightarrow{k_8} RL$$

Fig 6. An FSPN model for yeast polarization based on the eight reactions given above.

In this model, R, L, and RL represent the pheromone receptors, ligands, and receptor-ligand complexes, respectively. G represents the G-protein, and G_a, G_d and G_{bg} its three separate units. The ligands L bind with the receptors R to form complexes RL, which activate the G-proteins G to separate its two subunits G_a and G_{bg}. G_a, acting as an autophosphotase, can be dephosphorylated to G_d, which then rebinds with G_{bg} to go back to G. This completes the whole pheromone-induced G-protein cycle.

We create an FSPN model of the G-protein cycle based on the reactions given above, which is shown in Fig 6. Each species is modeled as a place and each reaction as a transition, and thus the FSPN model has 7 places and 8 transitions. That is, we can write the place set as $P = (L, R, RL, G, G_{bg}, G_a, G_d)$, and the transition set as $T = (r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8)$. Besides, we denote by $k = (k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8)$ the kinetic parameter set.

We start our analysis by determining some structural properties of the FSPN model using the Petri net analysis tool Charlie [50] (see [6] for the technical notions used in the following brief discussion). It turns out that this model is unbounded due to the two input transitions r_1, r_8. So its infinite CTMC cannot be constructed, but we can use our simulation-based method to explore its behavior. We further compute the invariants of the Petri nets, and obtain three minimal semi-positive place invariant, (L), $(G + G_{bg})$, and $(G + G_a + G_d)$, leaving R and RL as the places where the number of tokens is structurally unbounded. As L establishes a place invariant on its own, the token number on L will never change and remains during the execution always the same as at the initial marking. Thus, we can neglect it during the following analysis. Computing the transition invariants reveals that all reactions may contribute to the steady state behavior (technically speaking, the Petri net is covered with transition invariants). Finally, the so-called Siphon-Trap Property (STP) holds, which ensures for the given net structure (Extended Simple) that the Petri net is live, meaning all reactions will occur forever.

We set the initial marking to $M_0(P) = (4, 500, 110, 300, 20, 2, 90)$ and use the parameter values $k = (0.38, 0.04, 0.082, 0.12, 0.021, 0.1, 0.005, 13.21)$, both taken from [51]. In this case, the
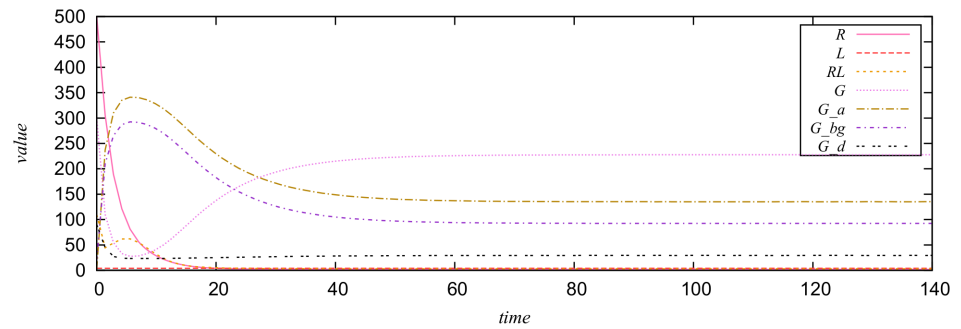
Fig 7. A stochastic plot of the yeast polarization model.

model is an SPN model, and we run stochastic simulation for 38,000 replications to obtain an average simulation trace, which is given in Fig 7. The plot shows that each species reaches a steady state in this example. But note that the convergence of expected values does not necessarily mean that the stochastic process reaches a steady state.

We then explore how the uncertainties of kinetic parameters affect the outputs using the simulation-based method introduced above. For each fuzzy parameter, we consider 11 α levels, i.e., 0.0, 0.1, . . ., 1.0, and equally discretize each α-cut into 11 points. For each combination of parameter values, we run a long simulation by setting the simulation end time to 38,000 and compute the steady state mean for each performance measure of interest using Eq 11. The reason for choosing this simulation end time is as follows. For transient simulation, 38,000 simulation runs offer 38,000 samples to compute the sample mean according to Eq 10, which results in the confidence level of about 95% and the accuracy of 10^{-2} according to Eq 4 in [44]. If we set the simulation end time to 38,000 and sample a trace every single simulation time unit, we also obtain 38,000 samples to compute a sample mean, which results in a similar confidence level and accuracy as when doing 38,000 runs for transient simulation.

First, we consider only one fuzzy parameter k_6 = (0.05, 0.1, 0.15) and explore how it affects the outputs of the model. The membership function of the steady state mean for each species is computed and illustrated in Fig 8, which clearly shows the uncertainty distribution of the steady state mean for each species. For example, the steady state mean for G can be at least about 109, while at most about 259 (at α level 0.0), and the most possible value is about 226 (at α level 1.0). For other species, we can obtain similar results. If we shorten the uncertainty interval of the parameter k_6, e.g., k_6 = (0.08, 0.1, 0.12), the range of the steady state mean for G will be shortened too, which is illustrated in Fig 9.

We then consider two fuzzy parameters, k_6 = (0.08, 0.1, 0.12) and k_8 = (10, 13.21, 15), and study how they affect the outputs of the model. The membership function of the steady state mean for each species is computed and illustrated in Fig 10. Of course, we can consider more fuzzy parameters, and we will obtain similar results as given above.

Now, let's go back to the scenario discussed in the Introduction to see how to take advantage of our approach. If there exist qualitative as well as quantitative experimental finding for a biological phenomenon, each of which may describe different aspects of the system, we could either use stochastic Petri nets to describe the quantitative findings, with the kinetic parameters to be determined by parameter estimation methods, or use qualitative Petri nets to represent the qualitative findings, e.g., the biochemical reactions of the system with unknown parameters. If we assign fuzzy values to these unknown parameters with the help of biologists providing the required insights, we could combine the stochastic and qualitative Petri net models to form a more detailed model and then study it from a quantitative point of view. On the other
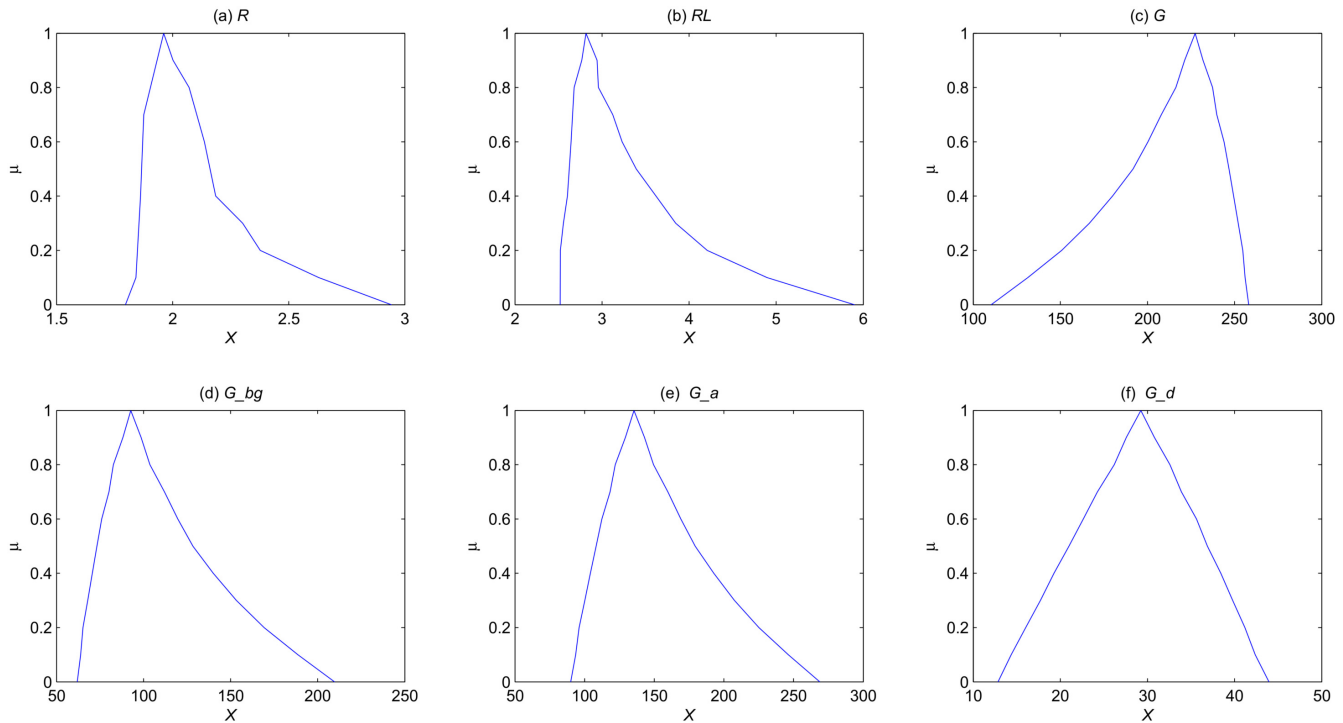
Fig 8. The membership function of the steady state mean for each species of the yeast polarization model in the setting of one fuzzy parameter k_6 = (0.05, 0.1, 0.15).

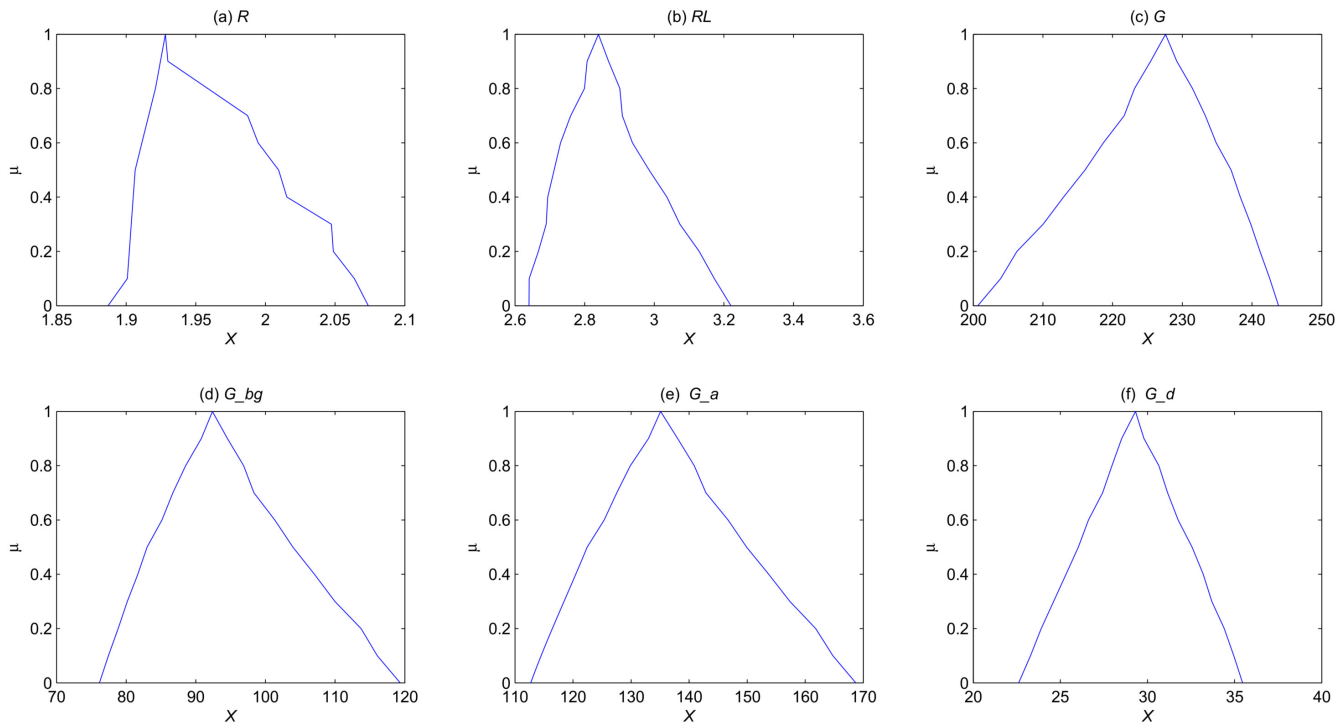doi:10.1371/journal.pone.0149674.g008

Fig 9. The membership function of the steady mean for each species of the yeast polarization model in the setting of one fuzzy parameter k_6 = (0.08, 0.1, 0.12).
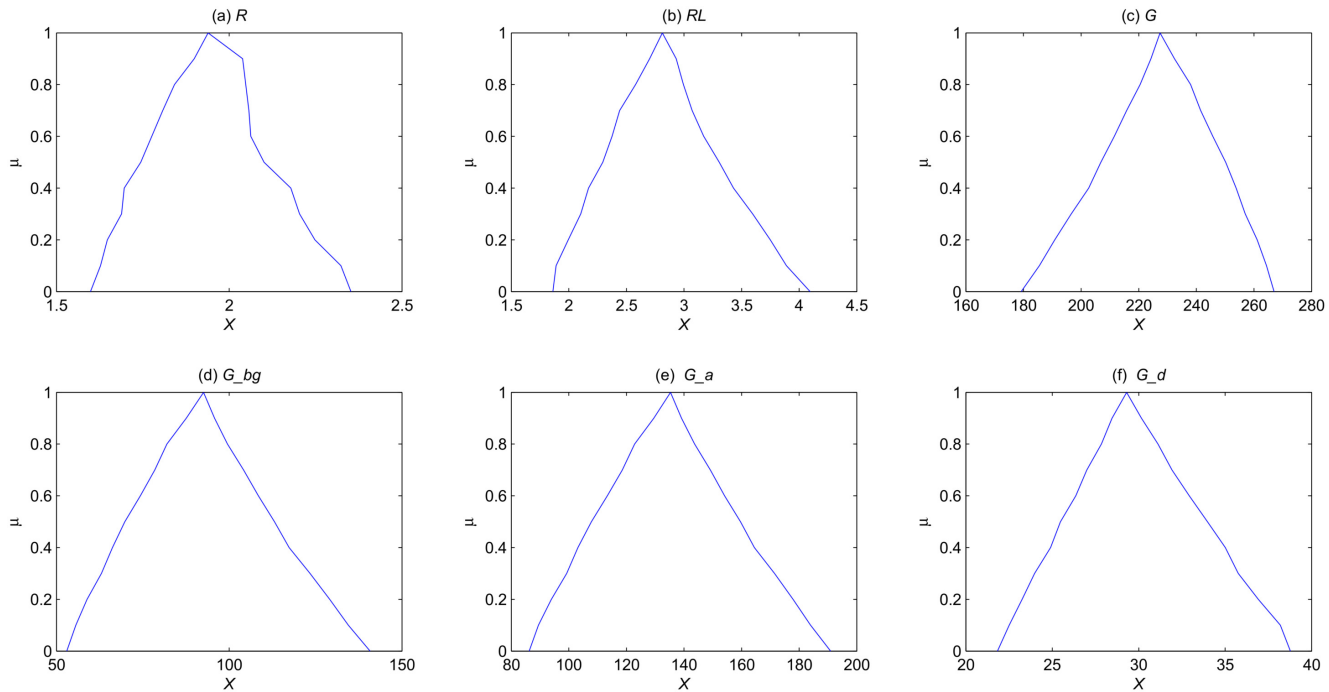
doi:10.1371/journal.pone.0149674.g009

Fig 10. The membership function of the steady state state mean for each species of the yeast polarization model in the setting of two fuzzy parameters k_6 = (0.08, 0.1, 0.12) and k_8 = (10, 13.21, 15).

doi:10.1371/journal.pone.0149674.g010

hand, if some parameters naturally vary, e.g., between different individuals, and we want to see the effect of these variabilities on the outputs, we could also use fuzzy sets to describe the variabilities of parameters and use our approach for its study.

Discussions and Conclusion

Taking into account the fact that in biological systems some kinetic parameters may be uncertain due to incomplete, vague or missing kinetic data, or naturally vary, e.g., between different individuals, experimental conditions, etc., we apply FSPNs by combining the strength of SPNs to model stochastic systems with the strength of fuzzy sets to deal with uncertain information.

We introduced a simulation-based analysis method and its implementation to simulate and analyze our FSPNs. Simulation-based analyzis techniques enjoy a couple of advantages. Most importantly, the size of the state space does not influence the efficiency of the approach; they work equally well for finite, but very large or even infinite state spaces. They also allow us to explore both transient and steady state behavior, are able to cope with non-markovian stochastic models, and can easily be adjusted to deal with various types of fuzzy numbers (beyond triangular-shaped ones). FSPNs combined with simulation provide us with a powerful tool to conveniently model and analyze biological systems with uncertain kinetic parameters.

We demonstrated the application of FSPNs and our analysis method using the yeast polarization model as case study. In summary, our approach is useful to integrate qualitative experimental findings into a quantitative model and to explore the system under study from the quantitative point of view. FSPNs provide a good means to consider parameter uncertainties in a model and to efficiently analyze how uncertain parameters affect the outputs of a model.

The way to deal with uncertain kinetic parameters and the simulation-based analysis method given in the paper equally apply to all quantitative modeling paradigms, i.e., also to

continuous and hybrid Petri nets [46]. In a next step, we will consider more case studies which could take advantage by combining quantitative Petri nets with fuzzy sets.

Acknowledgments

Author Contributions

Conceived and designed the experiments: FL MH. Performed the experiments: FL. Analyzed the data: FL MH MY. Contributed reagents/materials/analysis tools: FL MH MY. Wrote the paper: FL MH MY.

References

1. Meng TC, Somani S, Dhar P. Modeling and simulation of biological systems with stochasticity. In Silico Biol. 2004; 4(3):293–309. PMID: 15724281

2. Meister A, Du C, Li Y, Wong W. Modeling stochastic noise in gene regulatory systems. Quantitative Biology. 2014;p. 1–29.

3. Kepler TB, Elston TC. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophysics Journal. 2001; 81:3116–3136. doi: 10.1016/S0006-3495(01)75949-8

4. Phillips A, Cardelli L. Efficient, Correct Simulation of Biological Processes in the Stochastic Pi-calculus. In: Proc. of the 5th International Conference on Computational Methods in Systems Biology. LNCS 4695. Springer; 2007. p. 184–199.

5. Yang X, Han R, Guo Y, Bradley J, Cox B, Dickinson R, et al. Modelling and performance analysis of clinical pathways using the stochastic process algebra PEPA. BMC Bioinformatics. 2011;(13:):1–17.

6. Heiner M, Gilbert D, Donaldson R. Petri Nets for Systems and Synthetic Biology. In: Proc. of the 8th international conference on Formal methods for computational systems biology. LNCS 5016. Springer; 2008. p. 215–264.

7. Blätke MA, Heiner M, Marwan W. Predicting Phenotype from Genotype through Automatically Composed Petri Nets. In: Computational Methods in Systems Biology. LNCS 7605. Springer; 2012. p. 87–106.

8. Blätke M, Dittrich A, Rohr C, Heiner M, Schaper F, Marwan W. JAK/STAT signalling—an executable model assembled from molecule-centred modules demonstrating a module-oriented database concept for systems and synthetic biology. Molecular BioSystem. 2013; 9(6):1290–1307. doi: 10.1039/c3mb25593j

9. Goss PJE, Peccoud J. Quantitative Modeling of Stochastic Systems in Molecular Biology by Using Stochastic Petri Nets. The Proceedings of the National Academy of Sciences USA. 1998; 95(12):6750–6755. doi: 10.1073/pnas.95.12.6750

10. Srivastava R, Peterson MS, Bentley WE. Stochastic kinetic analysis of the Escherichia coli stress circuit using $\sigma(32)$-targeted antisense. Biotechnology and Bioengineering. 2001; 75(1):120–129. doi: 10.1002/bit.1171 PMID: 11536134

11. Marwan W, Sujatha A, Starostzik C. Reconstructing the regulatory network controling commitment and sporulation in Physarum polycephalum based on hierarchical Petri net modeling and simulation. J Theor Biol. 2005; 236:349–365. doi: 10.1016/j.jtbi.2005.03.018 PMID: 15904935

12. Mura I, Csikász-Nagy A. Stochastic Petri Net extension of a yeast cell cycle model. Journal of Theoretical Biology. 2008; 254(4):850–860. doi: 10.1016/j.jtbi.2008.07.019 PMID: 18703074

13. Lamprecht R, Smith GD, Kemper P. Stochastic Petri net models of Ca^{2+} signaling complexes and their analysis. Natural Computing. 2011; 10:1045–1075. doi: 10.1007/s11047-009-9143-y

14. Marwan W, Rohr C, Heiner M. 21. In: Helden J, Toussaint A, Thieffry D, editors. Petri nets in Snoopy: A unifying framework for the graphical display, computational modelling, and simulation of bacterial regulatory networks. vol. 804 of Methods in Molecular Biology. Humana Press; 2012. p. 409–437.

15. Castaldi D, Maccagnola D, Mari D, Archetti F. Stochastic Simulation of the Coagulation Cascade: A Petri Net Based Approach. In: Euro-Par 2012 Workshops. LNCS 7640. Springer; 2013. p. 248–262.

16. Liu F, Heiner M. Multiscale modelling of coupled Ca^{2+} channels using coloured stochastic Petri nets. IET Systems Biology. 2013 August; 7(4):106–113. doi: 10.1049/iet-syb.2012.0017 PMID: 23919953

17. Pârvu O, Gilbert D, Heiner M, Liu F, Saunders N, Shaw S. Spatial-temporal modelling and analysis of bacterial colonies with phase variable genes. ACM Transactions on Modeling and Computer Simulation (TOMACS). 2015 May; 25(2):25p.

18. Liu F, Heiner M. Modeling membrane systems using colored stochastic Petri nets. Nat Computing. 2013; 12(4):617–629. doi: 10.1007/s11047-013-9367-8

19. Blätke M, Heiner M, Marwan W. 7. In: Robeva R, editor. BioModel Engineering with Petri Nets. Elsevier Inc.; 2015. p. 141–193.

20. Shaw O, Steggles J, Wipat A. Automatic Parameterisation of Stochastic Petri Net Models of Biological Networks. Electronic Notes in Theoretical Computer Science. 2006; 151:111–129. doi: 10.1016/j.entcs.2006.03.015

21. Koh G, Teong HFC, Clément MV, Hsu D, Thiagarajan PS. A decompositional approach to parameter estimation in pathway modeling: a case study of the Akt and MAPK pathways and their crosstalk. Electronic Notes in Theoretical Computer Science. 2006; 22(14):271–280.

22. Donaldson R, Gilbert D. A model checking approach to the parameter estimation of biochemical pathways. In: Computational Methods in Systems Biology. LNCS 5307. Springer; 2008. p. 269–287.

23. Li C, Nagasaki M, Koch CH, Miyano S. Online model checking approach based parameter estimation to a neuronal fate decision simulation model in Caenorhabditis elegans with hybrid functional Petri net with extension. Mol BioSyst. 2011;(5:):1576–1592. doi: 10.1039/c0mb00253d PMID: 21373654

24. Gintrowski A. Modeling Gene Networks using Fuzzy Logic. In: Sixth Doctoral Workshop on Math. and Eng. Methods in Computer Science. Dagstuhl Publishing; 2010. p. 32–39.

25. Hamed RI, Ahson S, Parveen R. A new approach for modelling gene regulatory networks using fuzzy Petri nets. Journal of Integrative bioinformatics. 2010; 7(1):113.

26. Bordon J, Moskon M, Mraz M. Semi-quantitative modelling of biological systems with extended Fuzzy Petri nets. In: BioPPN 2013, a satellite event of PETRI NETS 2013. vol. 988. CEUR; 2013.

27. Valette R, Cardoso J, Dubois D. Monitoring manufacturing systems by means of Petri nets with imprecise markings. In: IEEE international symposium on intelligent control. IEEE; 1989. p. 233–238.

28. Ding Z, Bunke H, Schneider M, Kandel A. Fuzzy timed Petri net definitions, properties, and applications. Mathematical and Computer Modelling. 2005; 41:345–360. doi: 10.1016/j.mcm.2003.02.015

29. Tüysüz F, Kahraman C. Modeling a flexible manufacturing cell using stochastic Petri nets with fuzzy parameters. Expert Systems with Applications. 2010; 37(5):3910–3920. doi: 10.1016/j.eswa.2009.11.026

30. Heiner M, Rohr C, Schwarick M, Streif S. A Comparative Study of Stochastic Analysis Techniques. In: Proc. 8th International Conference on Computational Methods in Systems Biology (CMSB 2010). ACM digital library; 2010. p. 96–106.

31. Petri CA. Kommunikation mit Automaten. Bonn: Institut für Instrumentelle Mathematik, Schriften des IIM Nr. 2; 1962.

32. Murata T. Petri nets: Properties, analysis and applications. Proceedings of the IEEE. 1989; 77(4):541–580. doi: 10.1109/5.24143

33. Tareen S, Ahmad J. Modelling and Analysis of the Feeding Regimen Induced Entrainment of Hepatocyte Circadian Oscillators Using Petri Nets. PLoS one. 2015; 10(3):e0117519. doi: 10.1371/journal.pone.0117519 PMID: 25789928

34. Blätke M, Heiner M, Marwan W. Tutorial—Petri Nets in Systems Biology. Otto von Guericke University Magdeburg, Magdeburg Centre for Systems Biology; 2011.

35. Marsan MA, Balbo G, Conte G, Donatelli S, Franceschinis G. Modelling with Generalized Stochastic Petri Nets. Wiley Series in Parallel Computing. John Wiley and Sons; 1994.

36. David R, Alla H. Discrete, Continuous, and Hybrid Petri Nets. Springer Verlag, Berlin; 2010.

37. Marsan MA, Balbo G, Conte G, Donatelli S, Franceschinis G. Modelling with Generalized Stochastic Petri Nets. John Wiley and Sons; 1995.

38. Zadeh LA. Fuzzy Sets. Inform and Control. 1965;(8:):338–353. doi: 10.1016/S0019-9958(65)90241-X

39. Wang S, Watada J. Fuzzy Stochastic Optimization. Springer New York; 2012.

40. Dubois D, Prade H. Possibility Theory. Plenum Press, New York; 1988.

41. Liu B. Theory and Practice of Uncertain Programming. Physica-Verlag, Heidelberg; 2002.

42. Gillespie DT. Exact Stochastic Simulation of Coupled Chemical Reactions. Journal of Physical Chemistry. 1977; 81(25):2340–2361. doi: 10.1021/j100540a008

43. Alexopoulos C. STATISTICAL ANALYSIS OF SIMULATION OUTPUT: STATE OF THE ART. In: Proceedings of the 2007 Winter Simulation Conference. IEEE; 2007. p. 150–161.

44. Sandmann W, Maier C. On the Statistical Accuracy of Stochastic Simulation Algorithms Implemented in Dizzy. In: Proc. of the 5th International Workshop on Computational Systems Biology; 2008. p. 153–157.

45. Law AM. Simulation Modeling and Analysis, Fourth Edition. Tucson, AZ: McGraw-Hill Companies, Inc; 2007.

46. Heiner M, Herajy M, Liu F, Rohr C, Schwarick M. Snoopy—a unifying Petri net tool. In: Proc. PETRI NETS 2012. LNCS 7347. Springer; 2012. p. 398–407.

47. Matlab. Matlab—The Language of Technical Computing; 2013. http://www.mathworks.com.

48. Daigle BJ Jr, Roh MK, Gillespie DT, Petzold LR. Automated estimation of rare event probabilities in bio-chemical systems. THE JOURNAL OF CHEMICAL PHYSICS. 2011; 134:044110–1–13. doi: 10.1063/1.3522769

49. Drawert B, Lawson MJ, Petzold L, Khammash M. The diffusive finite state projection algorithm for efficient simulation of the stochastic reaction-diffusion master equation. THE JOURNAL OF CHEMICAL PHYSICS. 2010; 132:074101–1–12. doi: 10.1063/1.3310809 PMID: 20170209

50. Heiner M, Schwarick M, Wegener J. Charlie—an extensible Petri net analysis tool. In: Devillers R, Valmari A, editors. Proc. PETRI NETS 2015. vol. 9115 of LNCS. Springer; 2015. p. 200–211.

51. Daigle BJ Jr, Roh MK, Petzold LR, Niemi J. Accelerated maximum likelihood parameter estimation for stochastic biochemical systems. BMC Bioinformatics. 2012; 13(68):1–18.