Structural inference using nonlinear dynamics

Chris Oates*and Sach Mukherjee†

March 22, 2012

Abstract

Network inference methods are widely used to study regulatory interplay in biological systems. Such methods are usually based on simple, often linear, approximations to underlying dynamics. We present a network inference methodology that is rooted in nonlinear biochemical kinetics. This is done by considering a dynamical system that depends on a reaction graph, summarizing all biochemical reactions and associated parameters. We assume that neither graph nor parameters are known; inference regarding the graph is carried out within a Bayesian framework, using an efficient Monte Carlo approach to integrate out parameters. In this way, we take account of model complexity as well as fit-to-data. Focusing on protein signaling networks, we show results on data simulated from a recent dynamical model of MAPK signaling. We find that the method is able to effectively recover regulatory relationships. Furthermore, the approach facilitates modeling of interventional data, since it respects the roles of individual variables.

Statistical network inference techniques are widely used in the analysis of multivariate biochemical data [5, 11, 21, 24, 31, 36]. The objective is to make inferences regarding a network N whose vertices are identified with biomolecular components (e.g. genes or proteins) and edges with regulatory interplay between those components. Network edges often have the causal interpretation that intervention on the parent influences the child.

Network inference methods are typically rooted in simple descriptions of biological dynamics, usually linear or discrete [4, 5, 11, 25, 31, 36, 39]. The statistical and computational tractability of such formulations facilitates inference over large spaces of networks. However, biochemical networks are structural summaries of dynamical systems that are generally nonlinear. There is a extensive literature on relevant chemical kinetics [3]; in the absence of structural uncertainty, nonlinear ordinary differential equations (ODEs) are widely used to model dynamics [3, 7, 18, 19, 32]. However, the case where structure is unknown or uncertain has received less attention and to date nonlinear dynamical models have not been integrated into network inference over large model spaces.

*University of Warwick, Coventry, United Kingdom
†Netherlands Cancer Institute, Amsterdam, The Netherlands

Here, we propose an approach that integrates nonlinear dynamical models into network inference. We do so by considering a dynamical system $\boldsymbol{f}_G$ that depends on a "reaction graph" G, summarizing all biochemical reactions in the system. A network N can be thought of as a coarse summary of the reaction graph G, such that the former is a function $N(G)$ of the latter (a detailed example appears below). Letting $\boldsymbol{X}(t)$ denote a state vector describing system configuration at time t, we have $\dot{\boldsymbol{X}}(t) = \boldsymbol{f}_G(\boldsymbol{X}(t), \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ collects together all unknown parameters, including the reaction rates. Given time-course data $\mathcal{D}$, we carry out inference within a Bayesian framework to obtain a posterior distribution over reaction graphs G,

$$p(G|\mathcal{D}) \propto p(G) \underbrace{\int p(\mathcal{D}|\boldsymbol{\theta}, G) p(\boldsymbol{\theta}|G) d\boldsymbol{\theta}}_{\text{marginal likelihood } p(\mathcal{D}|G)} \tag{1}$$

where the marginal likelihood $p(\mathcal{D}|G)$ captures how well the biochemical model G describes data $\mathcal{D}$, taking into account model complexity. Importantly, the likelihood $p(\mathcal{D}|\boldsymbol{\theta}, G)$ is now rooted in the dynamical model $\boldsymbol{f}_G$, which we describe below.

This approach, which we call "Network Inference using Chemical Kinetics" (NICK), has a number of advantages over inference using on simple (linear or discrete) statistical models. For instance basing inference on richer dynamical models may provide gains with respect to estimation of network structure. Further, since inference respects the mechanistic roles of individual variables, integration of interventional data is facilitated and scientific interpretability of results enhanced. The approach we propose is general and can be used in many settings where kinetic formulations may be automatically generated, including gene regulation and protein signaling. For definiteness in this paper we focus on protein signaling networks mediated by phosphorylation. Protein phosphorylation is central to diverse biological processes and plays a key role in disease states including cancer [35]. Protein phosphorylation dynamics have been widely studied [18]; here we employ a Michaelis-Menten kinetic formulation [19]. Moreover, recent advances in high-throughput proteomics have begun to permit multivariate, data-driven study of signaling [15, 9]; we show an application to reverse-phase protein array data below.

Network inference using linear statistical models may be cast as approximate inference based on linear ODEs [25], with a number of existing methods exploiting this connection [2, 13]. In the context of gene regulation, [1, 17] combined linear ODEs with Gaussian processes. On the other hand, [38] recently demonstrated that formal statistical model selection based on four hand-crafted ODE nonlinear models could be used to elucidate signaling mechanism in human embryonic kidney cells. This work unifies nonlinear model selection with network inference in the context of biological networks.

The remainder of the paper is organized as follows. First, we introduce the model and associated statistical formulation. Second, we discuss inference. Third, we show empirical results, comparing NICK to several existing network

inference approaches. We carry out *in silico* assessment using a recent mechanistic model of MAPK signaling [38], under a range of regimes including interventional data. Finally, we discuss our findings, including directions for further work and limitations of the proposed approach.

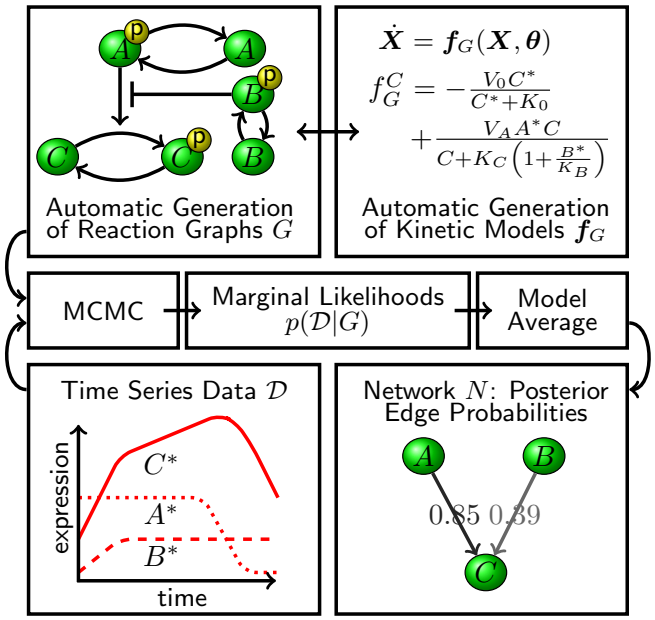

Figure 1: Network inference using chemical kinetics. Reaction graph G summarizes interplay that is described quantitatively by a kinetic model $\boldsymbol{f}_G$. Candidate graphs G are scored against observed time course data $\mathcal{D}$ in a Bayesian framework, using the marginal likelihood $p(\mathcal{D}|G)$. Evidence in favor of graph edges is captured by the the posterior probability of the edge, obtained by averaging over the space of graphs.

1 Methods

In this Section we describe the proposed methodology in the specific context of protein phosphorylation networks. We begin by defining reaction graphs G for this setting and considering relevant chemical kinetics. We then explain how statistical network inference rooted in these kinetics may be carried out.

1.1 Reaction graphs for protein phosphorylation

We consider p proteins $X_i \in \mathcal{V} = \{X_1 \dots X_p\}$. Each X_i can be phosphorylated to X_i^* (* denotes the phosphorylated form); the set of phosphorylated proteins

is $\mathcal{V}^*$. Phosphorylation of X_i is catalyzed by enzymes $E \in \mathcal{E}_i$; the subscript indicates that each protein may have a specific set of enzymes (enzymes catalyzing phosphorylation are known as kinases, we use both terms interchangeably). We consider the case in which the kinases themselves are phosphorylated proteins, i.e. the sets $\mathcal{E}_i$ are subsets of $\mathcal{V}^*$ (if phosphorylation of X_i is not driven by an enzyme in $\mathcal{V}^*$, we set $\mathcal{E}_i = \emptyset$). For simplicity we do not consider other post-translational modifications such as ubiquitinylation, nor spatial effects such as translocation. The ability of enzyme $E \in \mathcal{E}_i$ to catalyze phosphorylation of X_i may be inhibited by proteins $I \in \mathcal{I}_{i,E} \subset \mathcal{V}^*$; the double subscript indicates that inhibition is specific to both target X_i and enzyme E (see below). The reaction graph G provides a visual representation of the sets $\mathcal{E}_i$ and $\mathcal{I}_{i,E}$; Fig. 1 contains an illustrative example. A coarser network $N(G)$ is formed by drawing a directed graph with exactly p vertices and edges (i, j) indicating that X_i^* is either an enzyme catalyzing phosphorylation of X_j, or an inhibitor of such an enzyme. That is, $(i, j) \in N \iff i \in \mathcal{E}_j \vee \exists E \cdot i \in \mathcal{I}_{j,E}$. For the example shown in Fig. 1, the network N is the directed graph $A \to C \leftarrow B$.

1.2 Phosphorylation kinetics

The reaction graph G can be decomposed into local graphs G_i describing enzymes and their inhibitors for protein X_i. For simplicity of exposition, in what follows we consider phosphorylation of one protein X_i and inference concerning G_i. Thus, X_i plays the role of the substrate; following conventional notation in enzyme kinetics, we refer to X_i using the symbol S (the corresponding set of kinases is $\mathcal{E}_S$ and for kinase E the set of inhibitors $\mathcal{I}_{S,E}$). Square brackets $[\cdot]$ denote concentration.

We carry out network inference rooted in a dynamical system $\boldsymbol{f}_G$ based on Michaelis-Menten phosphorylation kinetics [18, 19, 32]. The rate of phosphorylation due to kinase E is given by $V_E[E][S]/([S] + K_E)$, which acknowledges variation of kinase concentration $[E]$ and permits kinase-specific response profiles, parameterized by K_E, with rate constant V_E. We entertain competitive inhibition, where substrate S and inhibitor I compete for the same binding site on the enzyme E ($EI \rightleftharpoons E \rightleftharpoons ES \to E{+}S^*$). When multiple inhibitors (I^A, I^B) are present, they are assumed to act exclusively, competing for the same binding site on the enzyme ($EI^A \rightleftharpoons E \rightleftharpoons EI^B$), corresponding mathematically to a rescaling of the Michaelis-Menten parameter $K_E \mapsto K_E \left(1 + \sum_{I \in \mathcal{I}_{S,E}} \frac{[I]}{K_I}\right)$. We do not model phosphatase specificity; in particular, dephosphorylation is assumed to occur at a rate $V_0[S^*]/([S^*] + K_0)$, depending on a Michaelis-Menten parameter K_0 and taking a maximal value V_0. Combining these assumptions produces a kinetic model for phosphorylation of substrate S, given by

$$f_{G_S}(\boldsymbol{X}_S, \boldsymbol{\theta}_S) = -\frac{V_0[S^*]}{[S^*] + K_0} + \sum_{E \in \mathcal{E}_S} \frac{V_E[E][S]}{[S] + K_E \left(1 + \sum_{I \in \mathcal{I}_{S,E}} \frac{[I]}{K_I}\right)} \tag{2}$$

where state vector $\boldsymbol{X}_S$ collects together concentrations of the product (S^*), kinases (E) and inhibitors (I), parameter vector $\boldsymbol{\theta}_S$ contains the maximum rates

($\boldsymbol{V}$) and Michaelis-Menten constants ($\boldsymbol{K}$), and the (local) graph G_S specifies the sets $\mathcal{E}_S$ and $\mathcal{I}_{S,E}$. Substrate concentrations $[S]$ are implicitly conditioned upon throughout, we do not model transcriptional dynamics here. To simplify notation we suppress the subscript S below and use $\boldsymbol{X}$ to refer to the complete and reduced state vector interchangeably, depending on context; similarly for $\boldsymbol{\theta}$. The complete dynamical system $\boldsymbol{f}_G$ is given by taking, for each species $S \in \mathcal{V}$, a model of the form Eqn. (2). In this way we are effectively able to automate the generation of biochemical hypotheses.

1.3 Statistical formulation

Data $\mathcal{D}$ comprises observations $Y_i(t_j), Y_i^*(t_j)$ proportional to the concentrations of unphosphorylated and phosphorylated forms respectively of species i at discrete times t_j, $0 \leq j \leq n$. In all examples, data are preprocessed using a three point moving average, then normalized to ensure unit mean expression of each protein. Observables are related to dynamics via an Euler approximation $Z_i(t_j) = (Y_i(t_j) - Y_i(t_{j-1}))/(t_j - t_{j-1})$, itself normalized to have unit standard deviation. Below we describe inference regarding the local reaction graph G_S for a single species S; iterating over $S \in \mathcal{V}$ permits inference concerning the complete reaction graph G. The kinetic model $\boldsymbol{f}_G$ (Eqn. (2)) is formulated as a statistical model by constructing a design matrix $\boldsymbol{D}(\boldsymbol{K})$ containing (unknown) Michaelis-Menten parameters $\boldsymbol{K}$,

$$\boldsymbol{D} = \Bigg[-\frac{Y_S^*}{Y_S^* + K_0}, \ldots, \underbrace{\frac{Y_E^* Y_S}{Y_S + K_E\left(1 + \sum_{I \in \mathcal{I}_{S,E}} \frac{Y_I^*}{K_I}\right)}}_{E \in \mathcal{E}_S}, \ldots \Bigg], \tag{3}$$

and then interpreting Eqn. (2) as

$$\boldsymbol{Z} \quad = \quad \boldsymbol{D}(\boldsymbol{K})\boldsymbol{V} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}) \tag{4}$$

where $\boldsymbol{Z} = [Z_S(t_1), \ldots, Z_S(t_n)]^T$, $\mathcal{N}$ denotes a Normal distribution, σ is a noise parameter, $\boldsymbol{I}$ the identity matrix and as above $\boldsymbol{V}$ the vector of maximum rates.

1.4 Bayesian inference

In the Bayesian setting prior distributions are needed to complete the model specification. We employ a truncated Normal prior $p(\boldsymbol{K}|G_S) = \mathcal{N}_T(\boldsymbol{K}; \boldsymbol{1}, \nu \boldsymbol{I})$ for Michaelis-Menten parameters where $\boldsymbol{1} = (1, \ldots, 1)$ and variance hyperparameter $\nu = 1/2$. Truncation ensures positivity of parameters; additional information on truncated Normals is provided in the Appendix. We use a Jeffreys prior $p(\sigma|G_S) \propto 1/\sigma$ over the noise parameter and a truncated g-prior $p(\boldsymbol{V}|\boldsymbol{K}, \sigma, G_S) = \mathcal{N}_T(\boldsymbol{V}; \boldsymbol{1}, n\sigma^2(\boldsymbol{D}'\boldsymbol{D})^{-1})$ for maximum rates $\boldsymbol{V}$ [40]. This formulation is partially conjugate, with the conditional density $p(\boldsymbol{V}, \sigma|\boldsymbol{K}, G_S, \mathcal{D})$ given in closed form as

$$p(\boldsymbol{V}, \sigma|\boldsymbol{K}, G_S, \mathcal{D}) \quad = \quad \mathcal{N}_T(\boldsymbol{V}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathcal{IG}(\sigma; a, b), \tag{5}$$

where $\boldsymbol{\mu} = \mathbf{1}/(n+1) + n/(n+1) \times (\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{Z}$, $\boldsymbol{\Sigma} = \sigma^2 n/(n+1) \times (\boldsymbol{D}'\boldsymbol{D})^{-1}$, $a = (n-1)/2$, $b = (1/2)(\mathbf{1}'\boldsymbol{D}'\boldsymbol{D}\mathbf{1}/n + \boldsymbol{Z}'\boldsymbol{Z} - n/(n+1) \times \boldsymbol{Z}'\boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{Z})$ and $\mathcal{IG}(\bullet; a, b)$ is an inverse gamma density with shape and scale parameters a, b respectively. In all experiments we used a non-informative network prior $p(G)$ (see Appendix); informative network priors are also available [23].

1.5 Marginal likelihood

Partial conjugacy of the above formulation permits an efficient Metropolis-within-Gibbs Markov chain Monte Carlo (MCMC) sampling scheme for the parameter posterior distribution $p(\boldsymbol{\theta}_S|G_S, \mathcal{D})$. The conditional $p(\boldsymbol{V}, \sigma|\boldsymbol{K}, G_S, \mathcal{D})$ is given in closed form as in Eqn. 5 above, while a Metropolis-Hastings acceptance step allows sampling from the remaining conditional $p(\boldsymbol{K}|\boldsymbol{V}, \sigma, G_S, \mathcal{D})$. To estimate marginal likelihoods from the output of this Metropolis-within-Gibbs sampler we employ a scheme due to [8], whereby the "basic marginal likelihood identity"

$$p(\mathcal{D}|G_S) = \frac{p(\mathcal{D}|\boldsymbol{\theta}_S, G_S)p(\boldsymbol{\theta}_S|G_S)}{p(\boldsymbol{\theta}_S|\mathcal{D}, G_S)} \tag{6}$$

is evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_S^*$ using a Monte Carlo estimate of the posterior ordinate $p(\boldsymbol{\theta}_S^*|\mathcal{D}, G_S)$. The point $\boldsymbol{\theta}_S^*$ is usually taken to be the posterior mode, however we found that in this application the posterior mean provided superior estimator precision, since in practice the mode is difficult to obtain.

1.6 Intervention

Interventions play an important role in experiments aimed at uncovering causal relationships between biological components. In such experiments, data are obtained under treatments that externally influence network nodes. Inhibitors of protein phosphorylation are now widely available; such inhibitors typically bind to the kinase domain of their target, preventing enzymatic activity by the target. We consider such inhibitors in biological experiments below. Within our framework, these interventions are modeled in a natural way by setting to zero terms in the design matrix $\boldsymbol{D}$ corresponding to the inhibited kinase E in the treated samples. This removes the influence of E from the model. We discuss causal interpretations in Discussion below.

1.7 Network inference

Evidence for influence (either kinase or kinase inhibiting activity) of protein i on protein j is summarized by the posterior probability of a directed edge (i, j). This is obtained by averaging over reaction graphs,

$$p((i,j) \in N|\mathcal{D}) = \frac{\sum_{G:(i,j)\in N(G)} p(\mathcal{D}|G)p(G)}{\sum_G p(\mathcal{D}|G)p(G)}. \tag{7}$$

Following work in structural inference for graphical models [11, 12] we bound graph in-degree degree to render Eqn. (7) tractable. In particular, we bound $|\mathcal{E}_S| \leq c_1$ and $|\mathcal{I}_{S,E}| \leq c_2$, with c_1, c_2 set to small values for computational tractability. Alternatively MCMC could be employed over the space of reaction graphs [14].

2 Results

Data were generated from a computational model of the MAPK signaling pathway (Fig. 2(a)) due to [38], specified by a system of 25 ODEs of Michaelis-Menten type (Eqn. (2)). This archetypal protein signaling system provides an ideal test case, since the underlying network structure is known exactly and the dynamical model used for simulation has been validated against experimental data. Moreover the model permits *in silico* intervention mimicking the effect of experimental inhibitors and agonists. In general, interventions may affect nodes that are observed, or nodes that are not observed (but that may influence those that are). Accordingly, we considered three distinct regimes: (i) no intervention, only global stimulation ("Global"), (ii) intervention outside the observed nodes ("Outset") and (iii) intervention on the observed nodes ("Inset"). Specific simulation conditions for these regimes were: (i) EGF stimulation only; (ii) EGF stimulation plus combinatorial PDE3 inhibition (cilostamide), EPAC activation (EPAC agonist) and PKA activation (PKA agonist); (iii) EGF stimulation plus direct inhibition (of kinase activity) of Ras, Rap1 and B-Raf in turn. For each experimental condition, measurements were obtained at 15 evenly spaced time points in the dynamic range following EGF stimulation. We generated trajectories for active forms of Ras, and Rap1, and phosphorylated forms of Raf-1, B-Raf, MEK and ERK.

We compared our approach to three established network inference methods: (i) ARACNE, based on mutual information [20]; (ii) dynamic Bayesian networks using Bayesian model averaging [16] ("DBN"); and (iii) a semiparametric approach using Gaussian processes [1] ("GP"). DBN and GP are based on linear models. For NICK we took $c_1 = 2$, $c_2 = 0$. Parents were inferred for {Raf-1, B-Raf, MEK and ERK} from the set {Ras, Rap1, Raf-1, B-Raf, MEK}, in each of the three experiments, using each of the four methods. Disregarding self-edges made a total of $(2^4)^4 = 16,384$ possible networks.

To assess ability to infer regulatory relationships, we first carried out a simulation free of measurement noise. Results are shown in Fig. 2(b); we find that NICK is better able to recover key features of the data-generating graph than the other approaches. Next, we added Gaussian noise of magnitude 0.1 onto log-concentrations, in line with the typical signal-to-noise ratio of current protein array technologies [15]. To systematically assess estimation of network structure we computed the area under the receiver operating characteristic curve (AUR; this represents the probability that a randomly chosen edge in the data-generating network, i.e. a true edge, is given a higher score than a randomly chosen false edge; higher scores indicate better performance). Fig. 2(c) shows

mean AURs for all four approaches, for the three intervention cases ("Global", "Outset" and "Inset"). Mean AUR scores for NICK are consistently high in all simulation regimes, demonstrating robustness of the chemical kinetic approach. Under "Inset" intervention NICK delivers superior performance.

3 Discussion

We introduced an approach by which to automatically integrate detailed mechanistic models into network inference. Empirical results demonstrated that the approach is capable of recovering network structure from time-course data. Whilst we restricted our investigation to protein signaling, the approach we propose is general and can be readily applied to any setting where automatic generation of kinetic equations is possible. In particular, extension to gene regulation is straightforward and indeed a Michaelis-Menten formulation could also be used in that setting.

Network inference aims to estimate regulatory networks from biochemical data. Edges in these networks have a causal interpretation in that intervention on the parent is expected to influence the child; however, care is needed in drawing causal conclusions. At present we can experimentally access only a small fraction of relevant chemical species, therefore it is likely that many estimated edges will not be causal, but rather confounded by unobserved variables. Therefore, validation by interventional experiments is crucial to establish causal relationships; the output of network inference should be regarded as suggestions for specific interventional experiments. However, even influences validated by intervention may be indirect, via unmeasured, intermediate species that do not confound the causal effect; to establish physically direct regulation requires further biochemical and structural work.

Causal inference in graphical models remains an ongoing area of statistical research [10, 26]. The approach used here can be formulated as a graphical model; our treatment of "Inset" interventions can be thought of as changing the structure of the underlying graphical model, as in [26]. Our approach differs from linear models, including conventional continuous (static or dynamic) Bayesian networks, since the underlying non-linear models are not structurally symmetric; this may aid in causal inference. We note that we considered only the case in which the target of inhibitors is known. For inhibitors whose targets are unknown, the framework we propose could in principle be adapted to ask whether any of the observed proteins are likely targets, complementing via a non-linear model the work of [2].

There are two main statistical problems associated with Bayesian inference for nonlinear dynamical systems: (i) inference of model parameters, and (ii) computation of marginal likelihoods for model selection. The first problem has been tackled from many directions, including Approximate Bayesian Computation [33], MCMC [37], and Particle Filtering [28]. The second problem is a comparatively under-developed area of statistical research, with candidate approaches evaluated in [6, 34]. In particular thermodynamic integration is

recommended by [34, 38] for marginal likelihood estimation, however such techniques remain computationally intensive. Alternatively, variational techniques may used to facilitate Bayesian model selection over dynamical systems; this approximate approach has recently been exploited by the neuroscience community [27]. Here, we employed a scheme based on MCMC output due to [8], formulating a partially conjugate statistical model to permit an efficient Metropolis-within-Gibbs sampling scheme.

Since inference in our approach decomposes over proteins i and for a given protein, over local models G_i, the computations are "naïvely" parallelizable. In principle, this should in the future allow our procedure to be scaled up to networks orders of magnitude larger than the systems considered here. Equally, MCMC approaches that sample from the graph space could be used in this setting [11, 14]. Nevertheless, the computational burden of our approach is higher than for corresponding linear formulations. In contrast, linear (or discrete) models are crude approximations to underlying dynamics, but using conjugate Bayesian or penalized likelihood approaches permit network inference that scales well to relatively high-dimensional systems. In this sense, the approach proposed here and linear network inference methods are complementary.

Although we mainly focused on the problem of structural inference, our approach also yields estimates of kinetic parameters. Importantly, in contrast to approaches that assume the reaction graph is known *a priori*, our approach permits inference concerning kinetic parameters even when the graph itself is unknown or uncertain, as is relevant in perturbed biological states such as cancers. Moreover, the use of chemical kinetic models may enhance predictive ability. For example the dynamic behavior of phosphoprotein concentrations obtained under our methodology are, unlike existing linear approaches, physically plausible (i.e. smooth, bounded and non-negative). It may therefore be possible to extend the approach to quantitatively predict dynamic cellular response to an external stimulus, such as a drug treatment, even when the underlying networks are subject to uncertainty.

4 Acknowledgments

The authors wish to acknowledge correspondence with Gareth Roberts, Mark Girolami and Vladislav Vyshemirsky. This work was supported by EPSRC EP/E501311/1 (CO & SM) and NCI U54 CA 112970 (SM).

5 Appendix

5.1 Truncated normal densities

We used truncated normal distributions as priors for kinetics parameters, as described in Main Text. Here, we describe truncated normal distributions and how we sampled from them.

A random variable $Y \in \mathbb{R}^p$ has a truncated multivariate normal distribution with mean μ and covariance Σ, denoted $Y \sim N_T(\mu, \Sigma)$, if Y has probability density function

$$p_Y(y) \propto \exp\left(-\frac{(y-\mu)^T\Sigma^{-1}(y-\mu)}{2}\right)\mathbb{I}(y \geq 0). \tag{8}$$

where $\mathbb{I}$ here is the indicator function, such that $\mathbb{I}(y \geq 0) = 1$ if $y \geq 0$, otherwise $\mathbb{I}(y \geq 0) = 0$. (The notation $y \geq 0$ is taken to mean that $y_i \geq 0$ for all $i = 1, \ldots, p$.) The density p_Y is related to the standard normal probability density ϕ via $p_Y(y) = C^{-1}\phi(y)\mathbb{I}(y \geq 0)$, so evaluation of p_Y requires

$$C = \int_{y \geq 0} \phi(y; \mu, \Sigma)dy = \int_{z \leq 0} \phi(z; -\mu, \Sigma)dz := \Phi(0; -\mu, \Sigma), \tag{9}$$

where Φ is the normal cumulative distribution function.

In general, sampling efficiently from truncated multivariate normal distributions is challenging. For example a rejection sampler based on an unconditioned normal density becomes inefficient when the measure of the target density's support is small. One approach is to construct a Gibbs sampler based on Eqn. 8 (see [29, 30]) but this is considerable effort for obtaining random samples for our purposes. However if the target distribution is non-degenerate (i.e. Σ is positive definite) then there exists a bijective mapping onto a product of standard truncated normal densities, which we exploit for sampling. Specifically, if $Y \sim N(\mu, \Sigma)$ then we can write $Y = \mu + AZ$ where $Z \sim N(0, I)$ where I is the identity matrix and A arises from the Cholesky decomposition $\Sigma = AA^T$. Positive definiteness ensures that the Cholesky decomposition exists and is unique. Moreover A is invertible, being lower triangular with strictly positive diagonal entries. Since $Y \geq 0$ if and only if $Z \geq -A^{-1}\mu$, we have the basis for efficient sampling (Algorithm 1). In the case that the target distribution approximates a point mass (this arises from conditioning on a rare event in the tails of a normal distribution), the algorithm uses numerical regularization.

5.2 Network prior

In the Bayesian setting, we must specify a prior probability distribution over reaction graphs G. Here, we used a non-informative network prior; uniform over the number of kinases, and for a given kinase, uniform over the number of kinase inhibitors:

$$p(G) = \prod_{i=1}^{p} \binom{p}{|\mathcal{E}_i|}^{-1} \prod_{E \in \mathcal{E}_i} \binom{p}{|\mathcal{I}_{i,E}|}^{-1} \tag{10}$$

References

[1] Äijö T, Lähdesmäki H (2010) Learning gene regulatory networks from gene expression measurements using non-parametric molecular kinetics. *Bioinf.* 25:2937-2944.

Algorithm 1 Efficient sampling from the (non-degenerate) truncated multivariate normal $Y \sim N_T(\mu, \Sigma)$, with numerical regularization. Here U is the uniform distribution, p is the dimension of Y and ϵ is taken to be machine precision.

$A \leftarrow \text{Cholesky}(\Sigma)$
$b \leftarrow -A^{-1}\mu$
for $i = 1$ **to** p **do**
  $u \sim U[\Phi(b_i), 1]$
  if $u > 1 - \epsilon$ **then**
    $z_i \leftarrow b_i$
  else
    $z_i \leftarrow \Phi^{-1}(u)$
  end if
end for
$y \leftarrow \mu + Az$

[2] Bansal M, Della Gatta G, di Bernardo D (2006) Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinf.* 22(7):815-822.

[3] Beard D, Qian H (2008) Chemical biophysics: quantitative analysis of cellular systems. *Cambridge University Press.*

[4] Bender et al. (2010) Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data. *Bioinf.* 26(ECCB 2010):i596-i602.

[5] Bonneau R (2008) Learning biological networks: from modules to dynamics. *Nat. Chem. Bio.* 4:658-664.

[6] Bos C (2002) A Comparison of Marginal Likelihood Computation Methods. *Compstat: Proceedings in Computational Statistics : 15th Symposium Held in Berlin, eds. Wolfgang Härdle, Bernd Rönz*, 111-116.

[7] Chen *et al.* (2009) Inputoutput behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol. Sys. Bio.* 5:239.

[8] Chib S, Jeliazkov I (2001) Marginal Likelihood from the Metropolis-Hastings Output. *J. Am. Stat. Ass.* 96(453):270-281.

[9] Ciaccio M, Wagner J, Chuu C, Lauffenburger D, Jones R (2010) Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nat. Meth.* 7(2):148-155.

[10] Eaton D, Murphy K (2007) Exact Bayesian structure learning from uncertain interventions. *In Proceedings of 11th Conference on Artificial Intelligence and Statistics (AISTATS-07).*

[11] Ellis B, Wong W (2008) Learning causal Bayesian network structures from experimental data. *J. Am. Stat. Ass.* 103(482):778-789.

[12] Friedman, N., Koller, D. (2003) Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. *Mach. Learn.* 50(1-2):95-125.

[13] Gardner T, di Bernardo D, Lorenz D, Collins J (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301(5629):102-105.

[14] Green P (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4):711-732.

[15] Hennessy *et al.* (2010) A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Nonmicrodissected Human Breast Cancer. *Clin. Proteom.* 6:129151.

[16] Hill S *et al.* (2011) Bayesian Inference of Signaling Network Topology in a Single Cancer, in preparation.

[17] Honkela *et al.* (2010) Model-based method for transcription factor target identification with limited data. *Proc. Nat. Acad. Sci.* 107(17):7793-7798.

[18] Kholodenko B (2006) Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Bio.* 7(3):165-176.

[19] Leskovac V (2003) Comprehensive enzyme kinetics. Springer.

[20] Margolin *et al.* (2006) ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinf.* 7(Suppl.1):S7.

[21] Markowetz F, Spang R. (2007) Inferring cellular networks - A review. *BMC Bioinf.* 8(Suppl. 6):S5.

[22] Memmott R, Dennis P (2009) Akt-dependent and -independent mechanisms of mTOR regulation in cancer. *Cellular Sig.* 21:656-664.

[23] Mukherjee S, Speed T (2008) Network inference using informative priors. *Proc. Nat. Acad. Sci.* 105(38):14313-14318.

[24] Nachman I, Regev A, Friedman N (2004) Inferring quantitative models of regulatory networks from expression data. *Bioinf.* 20(Suppl. 1):i248-i256.

[25] Oates C, Mukherjee S (2011) Network Inference and Biological Dynamics, to appear in *Ann. App. Stat.*

[26] Pearl J (2009) Causal inference in statistics: An overview. *Statistics Surveys* 3:96-146.

[27] Penny *et al.* (2010) Comparing Families of Dynamic Causal Models. *PLoS Comp. Bio.* 6(3):e1000709.

[28] Quach M, Brunel N, D'Alché-Buc F (2007) Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference. *Bioinf.* 23(23):3209-3216.

[29] Rodriguez-Yam, G., Davis, R.A., Scharf, L.L. (2002) A Bayesian model and Gibbs sampler for hyperspectral imaging, *Sensor Array and Multichannel Signal Processing Workshop Proceedings*, p105-9.

[30] Rodriguez-Yam, G., Davis, R.A., Scharf, L.L. (2004) Efficient Gibbs Sampling of Truncated Multivariate Normal with Application to Constrained Linear Regression, unpublished manuscript.

[31] Sachs K, Perez O, Pe'er D, Lauffenburger D, Nolan G (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 308:523529.

[32] Steijaert M, Van Den Brink J, Liekens A, Hilbers P, Ten Eikelder H (2010) Computing the Stochastic Dynamics of Phosphorylation Networks. *J. Comp. Bio.* 17(2):189-199.

[33] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. Roy. Soc. Inter.* 6(31):187-202.

[34] Vyshemirsky V, Girolami M (2008) Bayesian ranking of biochemical system models. *Bioinf.* 24(6):833-839.

[35] Weinberg R (2007) The Biology of Cancer. Garland Science.

[36] Werhli A, Grzegorczyk M, Husmeier D (2006) Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and Bayesian networks. *Bioinf.* 22(20):2523-2531.

[37] Wilkinson D (2010) Parameter inference for stochastic kinetic models of bacterial gene regulation: a Bayesian approach to systems biology. *Bayesian Statistics* 9:679-706.

[38] Xu *et al.* (2010) Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci. Sig.* 3(113):ra20.

[39] Yu J, Smith V, Wang P, Hartemink A, Jarvis E (2004) Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinf.* 20(18):3594-3603.

[40] Zellner A (1986) On Assessing Prior Distributions and Bayesian Regression Analysis With g-Prior Distributions. *Bayesian Inference and Decision Techniques - Essays in Honor of Bruno de Finetti, eds. P. K. Goel and A. Zellner*, 233-24.

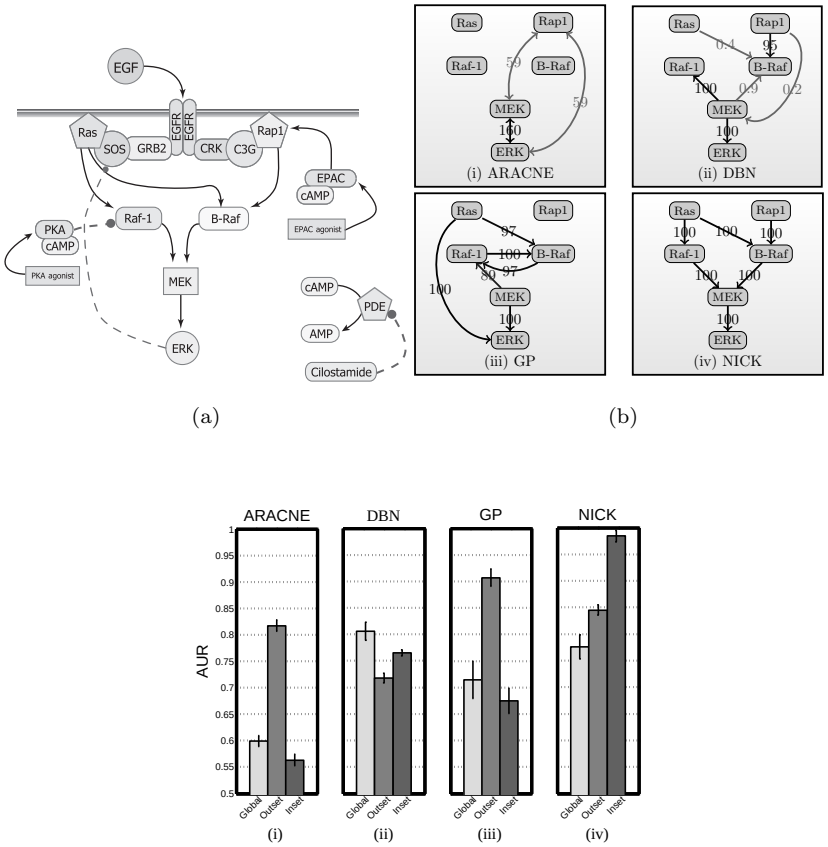

Figure 2: Simulation study. (a) Model of the ERK signaling pathway due to [38]. (b) Networks estimated from time course data, generated from model (a) under inhibition of network nodes, without measurement noise. (c) Mean area under the receiver operating characteristic curve (AUR) for each method, with Gaussian measurement noise. See text for details. [Network inference methods: (i) ARACNE, an information-theoretic approach, (ii) dynamic Bayesian network ("DBN"), (iii) Gaussian process-based network inference ("GP") and (iv) Network inference using chemical kinetics ("NICK"); Simulation regimes: global perturbation ("Global"), intervention on unobserved variables ("Outset"), intervention on observed variables ("Inset"); Panel (a) reproduced with permission from [38]; in Panel (b) edges annotated with weights indicating (100×) mutual information for ARACNE and posterior probability for other methods, 6 highest scoring edges shown in each case; in (c) mean AUR shown, errors bars indicate SEM.]