Bayesian Analysis (2013) **8**, Number 1, pp. 1–22

A Vine-copula Based Adaptive MCMC Sampler for Efficient Inference of Dynamical Systems

Daniel Schmidl *, Claudia Czado †, Sabine Hug ‡, and Fabian J. Theis §

Abstract. Statistical inference in high dimensional dynamical systems is often hindered by the unknown dependency structure of model parameters. In particular, the inference of parameterized differential equations (DEs) via Markov chain Monte Carlo (MCMC) samplers often suffers from high proposal rejection rates and is exacerbated by strong autocorrelation structures within the Markov chains leading to poor mixing properties. In this paper, we develop a novel vine-copula based adaptive MCMC approach for efficient parameter inference in dynamical systems with strong parameter interdependence. We exploit the concept of a vine-copula decomposition of distribution densities in order to generate problem-specific proposals for a hybrid independence/random walk Metropolis-Hastings (MH) sampler. The key advantage of this approach is that the corresponding MH proposals generate independent samples from the posterior distribution more efficiently than common competitors. All copula densities can be updated during the sampling procedure for fine-tuning. The performance of our method is assessed on two small-scale examples and finally evaluated on a delay DE model for the JAK2-STAT5 signaling pathway fitted to time-resolved western blot data. We compare our copula-based approach to an independence sampler, a second-order moment-based random walk MH algorithm, and an adaptive MH sampler.

Keywords: Parameter inference, Metropolis-Hastings algorithm, independence sampling, adaptive MCMC, vine, copula

1 Introduction

Dynamical systems are present in many scientific disciplines, including physics, engineering, bioinformatics, and many others. Such systems are frequently modeled by ordinary or delay differential equations and present modeling challenges due to scarce and noisy data for the typically large and complex models. Extensive research has been done on the inference of parameter values. As the data and models are generally very imprecise, determining single parameter values is often inadequate. In the last few years – especially in the fields of systems and computational biology – fully statistical Bayesian approaches were considered for parameter estimation in deterministic

*Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München and Institute of Mathematical Sciences, Technische Universität München, dschmidl@ma.tum.de

†Institute of Mathematical Sciences, Technische Universität München cczado@ma.tum.de

‡Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München and Institute of Mathematical Sciences, Technische Universität München sabine.hug@helmholtz-muenchen.de

§Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München and Institute of Mathematical Sciences, Technische Universität München
fabian.theis@helmholtz-muenchen.de

 DOI:10.1214/13-BA801

systems (Brown and Sethna (2003); Wilkinson (2006); Lawrence et al. (2010)). These Bayesian methods provide a way to combine the parameters of interest with the underlying data and *a priori* information, even when dealing with very complex models or partially unobserved quantities. As analytic computation of the posterior probability of model parameters quickly becomes intractable, numerical methods such as Markov chain Monte Carlo (MCMC) methods are typically employed (Brooks (1998); Gamerman and Lopes (2006)). Of the MCMC algorithms, one of the most successful and influential (Beichl and Sullivan (2000); Wilkinson (2007)) was developed by Metropolis and Hastings (Metropolis et al. (1953); Hastings (1970)).

Fine-tuning the Metropolis-Hastings (MH) algorithm for performing efficient inference is nevertheless a daunting task. The biggest hurdle in applying Bayesian methods arises due to their high computational costs (Wilkinson (2007)). Strong parameter dependencies and high dimensional sampling spaces limit MH algorithms to conservative parameter update schemes. In general, very many MCMC iterations may be required for vast traversals in the parameter space. Towards this end, a variety of algorithms have been developed to improve MCMC sampling efficiency (Girolami and Calderhead (2011); Liu (2008); Haario et al. (2001)). In this contribution we extend the MH algorithm by a novel problem-specific proposal function, which is based on the decomposition of the posterior distribution by means of a D-vine copula. Updating the proposal copula during the sampling process leads to an adaptive sampling scheme – an approach that has of late generated much interest in the MCMC community (Roberts and Rosenthal (2007); Holden et al. (2009); Rosenthal (2011)). Although copulas are well established modeling tools in fields like economics, finance, or geology, they have not yet been applied to tune the MH efficiency. Recently extensive research has been conducted regarding copulas and vine-copula decompositions (Kurowicka and Cooke (2006a); Aas et al. (2009); Kurowicka and Joe (2011)) for modeling systems with asymmetric tail dependencies – a characteristic also inherent to most posterior densities subject to MCMC sampling.

The paper is organized as follows. In the subsequent section we give a brief review of the random walk MH, the independence MH, and the Adaptive Metropolis (Haario et al. (2001)) algorithms. Section 3 reviews the concept of copulas and a D-vine decomposition of copulas. The main result is then contained in Section 4: we first present a basic version of our copula-based MH approach and subsequently extend it to an adaptive sampling scheme. Finally, we analyze the efficiency of our copula-MH approaches in Sections 5.1 - 5.3 by applying them to the problem of generating samples from a highly correlated normal distribution, a dynamic differential equation driven toy system, and a delay differential equation model of the JAK2-STAT5 signaling pathway. Throughout, we denote vectors/matrices by bold letters, while non-bold letters with subscript indices denote vector/matrix elements. Markov chains are displayed as sets, such as $\{\boldsymbol{\theta}^{(j)}\}_{j=0:K}$, where the superscript (j) denotes the j^{th} element. The notation $0:K$ abbreviates $0, 1, \ldots, K$. We also write $(\theta_{1:d})^\top$ for d-dimensional vectors $(\theta_1, \ldots, \theta_d)^\top$.

2 Monte Carlo methods in dynamical systems

Bayesian inference applied to dynamical systems is concerned with computing the probability distribution of the d-dimensional random parameter vector $\boldsymbol{\theta} = (\theta_{1:d})^\top \in \mathbb{R}^d$ of a parameterized system based on given observations $\boldsymbol{y} = (y_{1:n})^\top \in \mathbb{R}^n$. More precisely, in conjunction with a suitable error model, the data likelihood $p(\boldsymbol{y}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta})$ leads to computation of the posterior density $p(\boldsymbol{\theta}|\boldsymbol{y}) = \mathcal{L}(\boldsymbol{\theta})p(\boldsymbol{\theta})/p(\boldsymbol{y})$ (see Bayes and Price (1763)), where $p(\boldsymbol{y})$ is the normalizing capacity and $p(\boldsymbol{\theta})$ denotes the prior distribution. In dynamical systems parameters are often constrained to be non-negative. Sampling a Markov chain $\{\boldsymbol{\theta}^{(j)}\}_{j=0:K}$ from $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \mathcal{L}(\boldsymbol{\theta})p(\boldsymbol{\theta})$ can be done by the MH algorithm, which, starting at some initial $\boldsymbol{\theta}^{(0)}$, proposes for $j = 1 : K$ a sample $\tilde{\boldsymbol{\theta}}^{(j)}$ according to some transition density $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j-1)})$. The proposal $\tilde{\boldsymbol{\theta}}^{(j)}$ is then accepted with probability $\alpha(\tilde{\boldsymbol{\theta}}^{(j)}|\boldsymbol{\theta}^{(j-1)}) = \min\left\{\left(p(\tilde{\boldsymbol{\theta}}^{(j)}|\boldsymbol{y})q(\boldsymbol{\theta}^{(j-1)}|\tilde{\boldsymbol{\theta}}^{(j)})\right) / \left(p(\boldsymbol{\theta}^{(j-1)}|\boldsymbol{y})q(\tilde{\boldsymbol{\theta}}^{(j)}|\boldsymbol{\theta}^{(j-1)})\right), 1\right\}$ (see Robert and Casella (2004) for details). A very popular choice for the proposal density $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j-1)})$ is the d-dimensional normal distribution $\mathcal{N}_d(\boldsymbol{\theta}^{(j-1)}, \boldsymbol{\Sigma})$ with mean $\boldsymbol{\theta}^{(j-1)}$ and some fixed covariance matrix $\boldsymbol{\Sigma}$. We refer to this scheme as the *random walk Metropolis-Hastings (RWMH) algorithm*. If the proposal function is independent of all preceding samples the sampling scheme is called an *independence Metropolis-Hastings (IMH) algorithm*. The proposal function q is very crucial for the performance of the algorithm: an efficient MH algorithm yields high acceptance rates for the proposed samples with simultaneous low autocorrelation in the generated Markov chains, i.e. the chain is mixing well. This is especially hard to attain in high dimensions because small update step sizes result in high acceptance rates, but also in highly correlated Markov chain samples and vice versa. Towards this end, Haario et al. (2001) proposed an adaptive MH algorithm that gradually adjusts the RWMH proposal function during the sampling process: in this *Adaptive Metropolis algorithm (AM)* the RWMH proposal covariance matrix $\boldsymbol{\Sigma}$ for step (j) is defined by

$$\boldsymbol{\Sigma}^{(j)} = \begin{cases} \boldsymbol{\Sigma}^{(0)}, & j \leqslant j_0 \\ s_d \mathrm{cov}(\boldsymbol{\theta}^{(0)}, \ldots, \boldsymbol{\theta}^{(j-1)}) + s_d \varepsilon \boldsymbol{I}_d, & j > j_0, \end{cases} \tag{1}$$

for some initial covariance matrix $\boldsymbol{\Sigma}^{(0)} \in \mathbb{R}^{d\times d}$, the update initialization limit $j_0 > 0$, scaling constants $s_d > 0$ and $\varepsilon > 0$, and the d-dimensional identity matrix $\boldsymbol{I}_d \in \mathbb{R}^{d\times d}$.

3 Vine-copula decompositions

We now introduce the basic copula proposal function (for a thorough introduction on copulas and applications see Joe (1997) and Nelsen (2006)). A d-dimensional copula, C, is a multivariate distribution $C : [0,1]^d \longrightarrow [0,1]$ with uniform marginal distributions on $[0,1]$. According to Sklar's theorem (Sklar (1959)) there exists for any continuous multivariate cumulative distribution function (cdf) $F(\boldsymbol{x})$ with $\boldsymbol{x} = (x_{1:d})^\top$ and $\boldsymbol{u} = (u_{1:d})^\top$ a unique copula $C(\boldsymbol{u})$ such that for the marginal cdf's $F_i(x_i)$, $i = 1 : d$,

$$F(\boldsymbol{x}) = C(F_1(x_1), F_2(x_2), ..., F_d(x_d)). \tag{2}$$

A random vector $\boldsymbol{U} \sim C$ can hence be transformed to the random vector $\boldsymbol{X} \sim F$ via $x_i = F^{-1}(u_i)$ for $i = 1 : d$. We consider only absolutely continuous distributions $F(\boldsymbol{x})$ with joint density functions $f(\boldsymbol{x})$ and marginal density functions $f_i(x_i)$ for $i = 1 : d$. Then relationship (2) implies $f(\boldsymbol{x}) = c(F_1(x_1), \ldots, F_d(x_d)) \cdot f_1(x_1) \cdot \ldots \cdot f_d(x_d)$, where $c(\boldsymbol{u})$ is the density function corresponding to $C(\boldsymbol{u})$, i.e. $f(\boldsymbol{x})$ can be decomposed into the product of its marginals and the function $c(\boldsymbol{u})$. Here, $c(\boldsymbol{u})$ contains the full dependency structure of the random vector $\boldsymbol{X} \sim F$. For $d > 2$, there exist only very few copula distributions that allow for efficient sample generation. However, this class of multivariate copulas has recently been greatly extended by pair-copula based vine-copulas. Here, $c(\boldsymbol{u})$ is decomposed into the product of bivariate copulas as proposed by Joe (1996). Bedford and Cooke (2001), Bedford and Cooke (2002), and Kurowicka and Cooke (2006b) organized this *a priori* non-unique decomposition using a collection of linked trees, called vines. A special class of vines, the so-called D-vines, can be constructed as follows (Czado (2010)): The starting point is a recursive decomposition of a multivariate density into products of conditional densities. Let $\mathbf{X} := (X_{1:d})^\top$ be a set of variables with joint distribution function F and density function f. Consider the decomposition

$$f(\boldsymbol{x}) = [\prod_{t=2}^{n} f(x_t|x_{1:(t-1)})] \cdot f(x_1). \tag{3}$$

Here $F(\cdot|\cdot)$ and later $f(\cdot|\cdot)$ denote conditional cdf's and density functions, respectively. Using Sklar's theorem for dimension $d = 2$, we can express the conditional density of X_1 given $X_2 = x_2$ as

$$f(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1), \tag{4}$$

where $c_{12}(\cdot, \cdot)$ is an arbitrary bivariate copula density function. For distinct indices $i, j, i_1 : i_k$ with $i < j$, $i_1 < \cdots < i_k$ we set $c_{i,j|i_1:i_k} := c_{i,j|i_1:i_k}(F(x_i|x_{i_1:i_k}), F(x_j|x_{i_1:i_k}))$. Applying (4) to the conditional distribution of (X_1, X_t) given $X_{2:(t-1)}$ we can express $f(x_t|x_{1:(t-1)})$ recursively as $f(x_t|x_{1:(t-1)}) = [\prod_{s=1}^{t-2} c_{s,t|(s+1):(t-1)}] \cdot c_{(t-1),t} \cdot f_t(x_t)$. In combination with (3) and $s = i, t = i + j$ it follows that

$$f(\boldsymbol{x}) = [\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,(i+j)|(i+1):(i+j-1)}] \cdot [\prod_{k=1}^{d} f_k(x_k)]. \tag{5}$$

Note that the decomposition (5) of the joint density consists of pair-copula densities $c_{i,j|i_1:i_k}(\cdot, \cdot)$ evaluated at conditional distribution functions $F(x_i|x_{i_1:i_k})$ and $F(x_j|x_{i_1:i_k})$ for specified indices $i, j, i_1 : i_k$ and marginal densities f_k. This pair-copula decomposition was named a *D-vine distribution* by Bedford and Cooke. In general, the conditional pair-copula densities in (5) might depend on the conditioning values $x_{i_1:i_k}$. Here, however, we assume the restriction that $c_{i,j|i_1:i_k}(\cdot, \cdot)$ does not depend on $x_{i_1:i_k}$ for any i, j. This means that the decomposition (5) captures the dependency on the conditioning values solely through the arguments $F(x_i|x_{i_1:i_k})$ and $F(x_j|x_{i_1:i_k})$. Hobæk Haff et al. (2010) showed that this restriction is not severe. Aas et al. (2009) were the first to consider standard estimation methods for parameters of vine-copulas. These include stepwise and maximum likelihood estimation (MLE) – see Czado (2010) and Kurowicka and Joe

(2011) for current developments in this active area. Since the number of parameters grows quadratically in the dimension d, it is useful to consider a stepwise estimation approach, where we estimate the parameters from pair-copulas with no conditioning variables to the ones with $d-2$ conditioning variables. For the copula parameters with a single conditioning value, we transform the data with the appropriate conditional cdf's using the estimated parameters of the pair-copulas without conditioning to determine pseudo realizations needed in the pair-copulas with a single conditioning variable. We proceed as before until all parameters have been estimated. These so-called sequential estimates have been shown to be consistent and asymptotically normally distributed (Hobæk Haff (2013)). They are then used as starting values for numerically determining the maximum likelihood estimates. When several bivariate copula families for a pair-copula term are available, the family is chosen according to the Akaike information criterion (AIC). Brechmann (2010) showed that the AIC performs well with regard to several alternatives. Bayesian analyses of D-vines using MCMC are also available (Min and Czado (2010)). Additionally, model selection methods are implemented using indicator variables (Smith et al. (2010)) and reversible jump MCMC (Min and Czado (2011)). The R package *CDVine* of Brechmann and Schepsmeier (2013) applies the maximum likelihood/AIC approach described above.

4 Copula-based independence MH approach

We now want to introduce the basic version of our *hybrid copula-based independence/random walk Metropolis-Hastings approach (CIMH)*, which essentially constitutes an independence-sampling algorithm with a proposal function similar to the limiting distribution, resulting in high sampling efficiency. The copula proposal function is extended by two additional proposal functions, the first of which is a random walk density and the second a heavy-tailed independence density. Here, the latter is essential to safeguard convergence. Overall, we end up with a hybrid copula-based random walk transition density. The sampling scheme consists of four steps: (i) a prerun, (ii) a uniformization step of the prerun samples, (iii) a D-vine copula decomposition of the prerun samples, and (iv) the generation of a Markov chain by means of the hybrid copula-based independence/random walk sampler. We assume throughout that the sampling space $\Omega \subset \mathbb{R}^d$ is Borel measurable with measure μ.

4.1 The general sampling approach

(i) *Prerun:* Our goal is to construct a Markov chain $\{\boldsymbol{\theta}^{(j)}\}_{j=0:K}$ that has the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$ as its unique stationary and limiting distribution via an independence MH algorithm. For this, we first generate an initial Markov chain $\{\breve{\boldsymbol{\theta}}^{(j)}\}_{j=0:K'}$ for some $K' > 2$, the so-called *prerun samples*, using e.g. RWMH or any other sampling algorithm.

(ii) *Uniformization:* Based on $\{\breve{\boldsymbol{\theta}}^{(j)}\}_j$, we fit an $\boldsymbol{\eta}$-parameterized D-vine copula $c_{1:d}(\boldsymbol{u}|\boldsymbol{\eta})$ in step (iii). As seen in Section 3 copulas are defined on $[0,1]^d$. Hence, each prerun sample $\breve{\boldsymbol{\theta}}^{(j)}$ needs to be transformed to $[0,1]^d$. Depending on the shape of the histograms

of the d sample marginals $\breve{\boldsymbol{\theta}}_i := (\breve{\theta}_i^{(1)}, \ldots, \breve{\theta}_i^{(K')})^\top$ we fit for $i = 1:d$ $\boldsymbol{\gamma}_i$-parameterized continuous cdf's $G_i(\theta|\boldsymbol{\gamma}_i)$ to the respective sample marginal. Note that for all i the support of $G_i(\theta|\boldsymbol{\gamma}_i)$ needs to cover the respective dimension of the sample space. Each $\breve{\boldsymbol{\theta}}^{(j)}$ is then transformed to $\breve{\boldsymbol{u}}^{(j)} := (G_1(\breve{\theta}_1^{(j)}|\hat{\boldsymbol{\gamma}}_1), \ldots, G_d(\breve{\theta}_d^{(j)}|\hat{\boldsymbol{\gamma}}_d))^\top \in [0,1]^d$ based on the estimates $\hat{\boldsymbol{\gamma}}_i$ of $\boldsymbol{\gamma}_i$. In the following we refer to $\breve{\boldsymbol{u}}^{(j)}$ as *copula data*. Let us consider a simple example: say, for instance, $d = 2$ and the sample marginals of $\{\breve{\boldsymbol{\theta}}^{(j)}\}_j$ are normally distributed. Based on the estimated sample means $\hat{\mu}_1, \hat{\mu}_2$ and sample variances $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ of $\{\breve{\boldsymbol{\theta}}^{(j)}\}_j$ we transform $\breve{u}_1^{(j)} = \Phi\left((\breve{\theta}_1^{(j)} - \hat{\mu}_1)/\hat{\sigma}_1\right)$ and $\breve{u}_2^{(j)} = \Phi\left((\breve{\theta}_2^{(j)} - \hat{\mu}_2)/\hat{\sigma}_2\right)$, where $\Phi(\cdot)$ is the cdf of a standard normal random variable. Step (ii) does not change the dependency structure inherent to the prerun samples $\{\breve{\boldsymbol{\theta}}^{(j)}\}_j$, which is exclusively modeled by the D-vine copula (see Aas et al. (2009)). This implies that the estimated Kendall's τ's of $\{\breve{\boldsymbol{\theta}}^{(j)}\}_j$ are identical to the estimated Kendall's τ's of $\{\breve{\boldsymbol{u}}^{(j)}\}_j$.

(iii) *Copula decomposition:* Based on $\{\breve{\boldsymbol{u}}^{(j)}\}_j$ reorder the dimensions $1:d$ by means of a permutation function $\iota : \{1:d\} \longrightarrow \{1:d\}, i \mapsto \iota(i)$ such that for $i = 1:(d-1)$ the pair $(u_{\iota(i)}, u_{\iota(i)+1})$ exhibits the highest pairwise absolute Kendall's τ. Hence, each prerun sample $\breve{\boldsymbol{u}}^{(j)} = (\breve{u}_{1:d}^{(j)})^\top$ is transformed to $\tilde{\boldsymbol{u}}^{(j)} := (\breve{u}_{\iota(1)}^{(j)}, \ldots, \breve{u}_{\iota(d)}^{(j)})^\top$. While this step is not essential, it is known from parameter estimation in copulas that when unconditioned pair-copulas cover the strongest pairwise dependencies the estimation error can be reduced. Using $\{\tilde{\boldsymbol{u}}^{(j)}\}_j$ we fit an $\boldsymbol{\eta}$-parameterized D-vine copula density

$$c_{1:d}(\boldsymbol{u}|\boldsymbol{\eta}) = \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{j,j+i|j+1:j+i-1}(F(u_j|u_{j+1:j+i-1}, \boldsymbol{\eta}), F(u_{j+i}|u_{j+1:j+i-1}, \boldsymbol{\eta})|\boldsymbol{\eta}), \quad (6)$$

where $F(u_\ell|\boldsymbol{u}_\mathcal{D}, \boldsymbol{\eta})$ is the $\boldsymbol{\eta}$-parameterized conditional cdf of U_ℓ given $\boldsymbol{U}_\mathcal{D} = \boldsymbol{u}_\mathcal{D}$ and $\boldsymbol{u}_\mathcal{D}$ is a set of $[0,1]$-valued variables. Here, the order of the variables in (6) corresponds to the reordering by ι. In our notation the parameter $\boldsymbol{\eta} = \{\boldsymbol{\eta}_{i,j+i|(j+1):(j+i-1)}\}$ for $j = 1:(d-1)$ and $i = 1:(d-j)$ contains the copula parameters and types. The D-vine copula can be fitted using the maximum likelihood/AIC approach of Section 3. Note that since the number of copula-parameters grows quadratically in the dimension d, as mentioned above, the algorithm scales by $\mathcal{O}(d^2)$.

(iv) *Generation of the Markov chain:* The copula proposal function is defined as follows: for generating d-dimensional copula proposals $\tilde{\boldsymbol{\theta}} \in \Omega$, we sample $\tilde{\boldsymbol{u}} \sim c_{1:d}(\boldsymbol{u}|\hat{\boldsymbol{\eta}})$ from the estimated copula $c_{1:d}(\boldsymbol{u}|\hat{\boldsymbol{\eta}})$. The sample $\tilde{\boldsymbol{u}}$ is then transformed by $\tilde{\theta}_i := G^{-1}_{\iota^{-1}(i)}(\tilde{u}_{\iota^{-1}(i)}|\hat{\boldsymbol{\gamma}}_{\iota^{-1}(i)})$ to yield $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_{1:d})^\top$. In the setting of the example above ι is the identity function. The corresponding samples $\tilde{\boldsymbol{\theta}}$ on $\mathbb{R}^2$ are then for $i = 1, 2$ given by $\tilde{\theta}_i = G^{-1}_{\iota^{-1}(i)}(\tilde{u}_{\iota^{-1}(i)}|\hat{\mu}_{\iota^{-1}(i)}, \hat{\sigma}^2_{\iota^{-1}(i)}) = \Phi^{-1}(\tilde{u}_i)\hat{\sigma}_i + \hat{\mu}_i$. Thus, all copula proposals $\tilde{\boldsymbol{\theta}}$ are generated according to the joint density function

$$q_1(\boldsymbol{\theta}|\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}) := c_{1:d}(G_1(\theta_1|\hat{\boldsymbol{\gamma}}_1), \ldots, G_d(\theta_d|\hat{\boldsymbol{\gamma}}_d)|\hat{\boldsymbol{\eta}}) \cdot \prod_{i=1}^{d} g_i(\theta_i|\hat{\boldsymbol{\gamma}}_i), \quad (7)$$

where $g_i(\theta|\hat{\boldsymbol{\gamma}}_i)$ are the density functions corresponding to $G_i(\theta|\hat{\boldsymbol{\gamma}}_i)$. Now, let $q_2(\boldsymbol{\theta}|\boldsymbol{\theta}')$ be a random Metropolis-Hastings transition density of choice and $q_3(\boldsymbol{\theta})$ a (compared

to the posterior density $p(\boldsymbol{\theta}|\boldsymbol{y})$) heavy-tailed independence transition density. For fixed constants $r_1 \in [0,1)$ and $r_2 \in [0,1)$ with $r_1 + r_2 < 1$ we define the *copula-based hybrid independence/random walk transition density for CIMH* via the density function

$$q^{cop}(\boldsymbol{\theta}|\boldsymbol{\theta}', \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}) := r_1 q_1(\boldsymbol{\theta}|\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}) + r_2 q_2(\boldsymbol{\theta}|\boldsymbol{\theta}') + (1 - r_1 - r_2) q_3(\boldsymbol{\theta}). \tag{8}$$

With respect to readability we simply write $q^{cop}(\boldsymbol{\theta}|\boldsymbol{\theta}')$ instead of $q^{cop}(\boldsymbol{\theta}|\boldsymbol{\theta}', \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}})$. For proving convergence of the proposed sampling scheme we make use of the strong Doeblin condition, which requires that there exist an integer $s > 0$ and a constant $a_s \in (0,1]$ such that $(q^{cop})^s(\boldsymbol{\theta}, \boldsymbol{\theta}') \geqslant a_s p(\boldsymbol{\theta}|\boldsymbol{y})$ for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Omega$. Here, $(q^{cop})^s$ is the s-step transition density for $\boldsymbol{\theta}$ implicitly defined for the density after j iterations, P^j, by $P^{i+s}(\boldsymbol{\theta}) = \int_\Omega (q^{cop})^s(\boldsymbol{\theta}'|\boldsymbol{\theta}) P^i(\boldsymbol{\theta}') \, \mathrm{d}\mu(\boldsymbol{\theta}')$ with $(q^{cop})^1 = q^{cop}$ (compare Holden (2000)). The transition density q_3 guarantees that the proposal distribution q^{cop} has uniformly heavier tails than the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$. It follows that the strong Doeblin condition holds (see Holden et al. (2009)). As the MH acceptance probability satisfies the detailed balance condition, Holden (2000) showed that the strong Doeblin condition implies convergence of the Markov chain. The constants r_1 and r_2 are generally chosen such that $r_1 + r_2$ is close to one in order to "waste" as few samples as possible.

4.2 Adaptive copula-sampling scheme

Short preruns might cause insufficient sampling from the posterior's marginals' tails in order to fit an efficient proposal copula. To avoid setting $r_1 + r_2 \ll 1$ and thus generating ineffective proposals, we propose an extension of the basic CIMH approach by sequentially updating the copula functions based on preceding Markov chain samples. This changes the proposal function during the sampling process and leads to an adaptive MCMC scheme: for integers $R, S > 0$ we set the *copula update-probability* for the j^{th} MCMC step, $P(j)$, to

$$P(j) = \begin{cases} 1, & \text{if } j \bmod R = 0 \text{ and } j < R \cdot S, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

That is, the estimated copula parameters $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\eta}}$ become dependent on the proposal step j, resulting in a step-dependent proposal function $q^{cop}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}, \hat{\boldsymbol{\gamma}}^{(j)}, \hat{\boldsymbol{\eta}}^{(j)})$, where $\hat{\boldsymbol{\gamma}}^{(j)}$ and $\hat{\boldsymbol{\eta}}^{(j)}$ are updated based on the concatenated prerun samples and the samples generated up to step j according to the copula update-probability. By construction the support of each $q^{cop}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}, \hat{\boldsymbol{\gamma}}^{(j)}, \hat{\boldsymbol{\eta}}^{(j)})$ covers Ω. It follows that the associated transition kernel of the Metropolis-Hastings algorithm is ergodic. The copula proposal function is updated at most S times. Proposition 3 in Roberts and Rosenthal (2007) then ensures that this finite adaption scheme preserves ergodicity and therefore yields a valid adaptive MCMC sampler. We refer to this *hybrid adaptive copula-update independence/random walk MH approach* as *ACIMH*. The pseudo code for ACIMH is shown in Algorithm 1.

Algorithm 1: The ACIMH algorithm

(i) **Input**: RWMH prerun samples $\{\breve{\boldsymbol{\theta}}^{(j)}\}_{j=0:K'}$ with $\breve{\boldsymbol{\theta}}^{(j)} = (\breve{\theta}_1^{(j)}, \ldots, \breve{\theta}_d^{(j)})^\top$, update and sampling parameters R, S, r_1, and r_2, chain length K, starting value $\boldsymbol{\theta}_0$, and transition densities q_2 and q_3.
Output: Markov chain $\{\boldsymbol{\theta}^{(j)}\}_{j=0:K}$.

Initialize $s \leftarrow 0$ and set $\boldsymbol{\theta}^{(0)} \leftarrow \boldsymbol{\theta}_0$.
Infer dimension permutation function ι.
for $j \leftarrow 0$ **to** K **do**
  if j *mod* $R = 0$ *and* $j < R \cdot S$ **then**
    Update $s \leftarrow s + 1$.
(ii) **for** $i \leftarrow 1$ **to** d **do**
      Fit $\hat{\boldsymbol{\gamma}}_i^{(s)}$ of parameterized cdf $G_i(\cdot|\boldsymbol{\gamma}_i)$ based on $\{\breve{\theta}_i^{(k)}\}_{k=0:(K'+j)}$.
      for $k \leftarrow 0$ **to** $K' + j$ **do**
        Set $\breve{u}_i^{(k)} \leftarrow G_i(\breve{\theta}_i^{(k)}|\hat{\boldsymbol{\gamma}}_i^{(s)})$.
(iii) Fit $\hat{\boldsymbol{\eta}}^{(s)}$ of D-vine copula $c_{1:d}(u_1, \ldots, u_d|\boldsymbol{\eta})$ on $\{(\breve{u}_{\iota(1)}^{(k)}, \ldots, \breve{u}_{\iota(d)}^{(k)})^\top\}_{k=0:(K'+j)}$.
(iv) **if** $j > 0$ **then**
    According to r_2, $1 - r_1 - r_2$, and r_1 sample $\tilde{\boldsymbol{\theta}} \sim q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j-1)})$, $\tilde{\boldsymbol{\theta}} \sim q_3(\boldsymbol{\theta})$, or $(\tilde{u}_1, \ldots, \tilde{u}_d)^\top \sim c_{1:d}(u_1, \ldots, u_d|\hat{\boldsymbol{\eta}}^{(s)})$, respectively. In the latter case
    for $i \leftarrow 1$ **to** d **do**
      transform $\tilde{\theta}_i \leftarrow G_{\iota(i)}^{-1}(\tilde{u}_{\iota^{-1}(i)}|\hat{\boldsymbol{\gamma}}_{\iota(i)}^{(s)})$
    and define $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_d)^\top$. Set

$$\boldsymbol{\theta}^{(j)} \leftarrow \begin{cases} \tilde{\boldsymbol{\theta}} & \text{with prob. } \alpha^{cop}(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(j-1)}) = \min\left\{\frac{p(\tilde{\boldsymbol{\theta}}|\boldsymbol{y})q^{cop}(\boldsymbol{\theta}^{(j-1)}|\tilde{\boldsymbol{\theta}})}{p(\boldsymbol{\theta}^{(j-1)}|\boldsymbol{y})q^{cop}(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(j-1)})}, 1\right\}, \\ \boldsymbol{\theta}^{(j-1)} & \text{with prob. } 1 - \alpha^{cop}(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(j-1)}). \end{cases}$$

  Set $\breve{\boldsymbol{\theta}}^{(K'+j+1)} \leftarrow \boldsymbol{\theta}^{(j)}$.

5 Performance and results

For benchmarking CIMH and ACIMH the algorithms were tested on three examples. First, we draw samples from a strongly correlated bivariate normal distribution. This system serves as a simple proof-of-concept of an analytically tractable system. Subsequently, we turn to dynamical systems defined by differential equations (DEs). More precisely, Example 2 examines the performance for ordinary non-linear parameter dependencies and parameter distributions with non-symmetric tail dependencies. Finally, we apply our sampling approaches to a delay differential equation (DDE) model of the JAK2-STAT5 signaling pathway as published by Swameye et al. (2003). Here, a sophisticated proposal generation is crucial as there exists no closed form solution of the DDE system, calling for a computationally very expensive numerical solution for every evaluation of the likelihood. Moreover, the seven parameters involved show high interdependency, which additionally complicates the inference.

We evaluated the following performance indices: (I_1) the quotient of acceptance rate

and estimated inefficiency factor (INEFF) and (I_2) the estimated effective sample size (ESS) per second. Here, ESS $= \max_{i=1:d} \left\{ (K+1) / \left(1 + 2 \sum_{\tau=1}^{K_c^i} (1 - \frac{\tau}{K+1}) \hat{\rho}_i(\tau) \right) \right\}$ for the MCMC sampling length K and the estimated autocorrelation functions $\hat{\rho}_i(\tau) = \frac{1}{\hat{\sigma}^2(K+1-\tau)} \sum_{j=\tau}^{K} (\theta_i^{(j)} - \hat{\mu}_i)(\theta_i^{(j-\tau)} - \hat{\mu}_i)$ of lag τ and dimension i of a Markov chain $\{\boldsymbol{\theta}^{(j)}\}_{j=0:K} = \{(\theta_i^{(j)})_{i=1:d}\}_{j=0:K}$ (Hoffman and Gelman (2011), Appendix A). The estimates $\hat{\mu}_i$ and $\hat{\sigma}^2$ of the mean and variance of dimension i are computed from a separate very long (1,000,000 sample) MCMC run in order to avoid underestimation of the autocorrelation function. Since $\hat{\rho}_i(\tau)$ becomes noisy for large τ, the sum is truncated at $K_c^i = \text{argmin}\{\hat{\rho}_i(\tau)\}$, such that $\hat{\rho}_i(\tau) < 0.05$. The ESS assesses the mixing property of the Markov chain (He et al. (2007)). Conversely, the INEFF is simply given by INEFF$=(K+1)/$ESS. It is equal to the average number of Markov chain states that two samples need to be separated in order to be considered independent. For our applications (I_1) was motivated by the trade-off between high acceptance rates and low INEFFs for Markov chains with small proposal variances, and the rejection of a large fraction of proposed moves for the MH algorithm when variance is too high (see Roberts et al. (1997); Liu (2008); Girolami and Calderhead (2011)). Clearly $(I_1) \in [0, 1]$, with higher values being superior. As all algorithms were implemented in MATLAB (R2012a) using the same underlying MH code, (I_2) is a well justified measure for the mixing speed of an algorithm. Time here denotes the CPU-time on a single core of a 24 core AMD Opteron 6234 (2.4 GHz) machine.

The performances of CIMH and ACIMH were compared to (a) an RWMH algorithm, (b) an IMH algorithm, and (c) the AM algorithm using (I_1) and (I_2). We applied a joint parameter update scheme in each proposal function. In order to set up the proposal functions of RWMH, IMH, CIMH, and ACIMH, we ran another random walk MH algorithm to generate so-called *prerun samples*. The latter used an $\mathcal{N}(\boldsymbol{\theta}^c, \boldsymbol{\Sigma})$ proposal function, where $\boldsymbol{\theta}^c$ denotes the current Markov chain sample and $\boldsymbol{\Sigma}$ is a fixed covariance matrix defined as follows: we determined the maximum a posteriori estimates for all d parameters using a simulated annealing algorithm (Kirkpatrick et al. (1983)); denoting these estimates by s_i the i^{th} diagonal element of $\boldsymbol{\Sigma}$ was set to $k_p \cdot s_i$, where k_p was adjusted in each example to yield an acceptance rate of approximately 23% as suggested in Roberts et al. (1997) – our exact limits were set to 10% and 36%. Two major issues of the prerun sampler are (P_1) a rather strong autocorrelation between subsequent MCMC iterations and (P_2) its failure to incorporate any information about the limiting distribution when proposing new samples. To address (P_1) we set up the IMH whose proposals are generated independently of the current Markov chain state: as for the copula-based algorithms we fitted one-dimensional parameterized cdf's $G_i(\theta|\boldsymbol{\gamma}_i)$ to each of the d empirical marginal parameter distributions sampled in the prerun. In fact, these were identical for IMH, CIMH, and ACIMH. The IMH proposals $\tilde{\theta}_i^{(j)}$ were jointly generated by sampling $d \cdot (K+1)$ independent samples $u_i^{(j)} \sim \mathcal{U}[0,1]$ $(i = 1:d,$ $j = 0:K)$, which are subsequently transformed to $\tilde{\theta}_i^{(j)} = G_i^{-1}(u_i^{(j)}|\hat{\boldsymbol{\gamma}}_i)$. In other words, IMH generates proposals assuming an independent parameter structure. Rather than directly reducing the autocorrelation in the Markov chain, RWMH exploits the expected covariance matrix $\hat{\boldsymbol{C}}$ of the prerun and addresses (P_2): the RWMH proposal function is

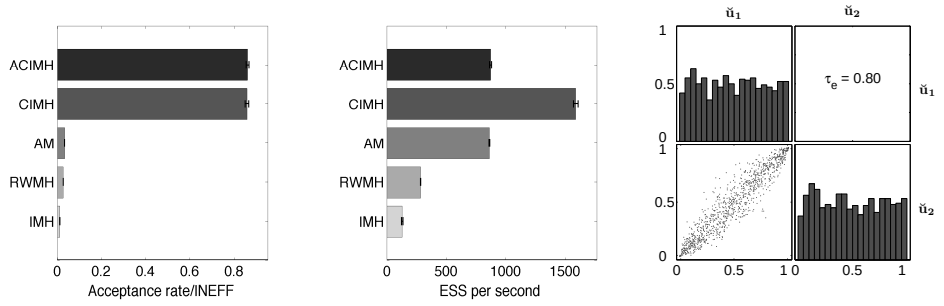
given by $\mathcal{N}(\boldsymbol{\theta}^c, k_{RW} \cdot \hat{\boldsymbol{C}})$. Again, k_{RW} was chosen to yield an approximate acceptance rate of 23%. In order to save the sampling time of the prerun AM is addressing (P_2) by gradually updating the proposal covariance matrix during the sampling process. For our applications we set $\boldsymbol{\Sigma}^{(0)} = \boldsymbol{I}_d$, $\varepsilon = 1 \cdot 10^{-7}$, and s_d to yield an approximate acceptance rate of 23% (determined by an additional independent MCMC run). The update initialization limit j_0 was set to coincide with the first accepted Markov chain sample different from $\boldsymbol{\theta}^{(0)}$. While IMH and RWMH can in some sense be seen as antagonistic approaches w.r.t. (P_1) and (P_2), CIMH and ACIMH address both issues at once. For performance assessment CIMH and ACIMH were applied as introduced in Section 4. Throughout, the Metropolis-Hastings transition density q_2 was taken to be identical with the one of RWMH, reusing the tuning parameter k_{RW}; q_3 and the proposal probabilities r_1 and r_2 were adjusted individually (see Sections 5.1 - 5.3). For thorough performance evaluation, the first two examples were each run 100 times for 50,000 MCMC iterations, the last one 10 times for 50,000 MCMC iterations. The same set-up was applied for prerun sampling. In all examples the copula update parameters for ACIMH were set to $R = 10{,}000$ and $S = 4$. While the copulas were fitted on 1,000 equally spaced prerun samples in the first two examples, we used 3,000 samples for the JAK2-STAT5 inference, owing to the complexity of the system. The time for the prerun was added to the sampling times of IMH, RWMH, CIMH, and ACIMH. For copula fitting and sample generation the *CDVine* R-package (Brechmann and Schepsmeier (2013)) was used.

5.1 Sampling from a strongly correlated 2-dim. normal distribution

In our first example we draw samples from a strongly correlated bivariate normal distribution $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respective mean and covariance matrix

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & 0.95 \cdot \sqrt{3} \\ 0.95 \cdot \sqrt{3} & 3 \end{pmatrix}.$$

Here, $\rho = 0.95$. We chose this example as it is illustrative as well as analytically tractable. Canonically, the cdf's of $\mathcal{N}(0,1)$ and $\mathcal{N}(0,3)$ were used to transform the prerun samples to $[0,1]^2$. The independence proposal density q_3 was taken to be a bivariate Student-t distribution with location parameter $(0,1)^\top$ and identity scale matrix. Furthermore, we set $r_1 = 0.99$ and $r_2 = 0$. All samplers were started at the origin and approximated the two-dimensional normal distribution with negligible errors. In fact, the residual differences between the estimated and true parameter values of $\mu_1, \mu_2, \sigma_1, \sigma_2$, and ρ were on average less than $8.2 \cdot 10^{-2}$ throughout all approaches during the 100 runs. Although the sampling times for IMH $(16.9 \pm 1 \cdot 10^{-2}$ sec.) were on average slightly faster than the ones of RWMH $(21.3 \pm 5 \cdot 10^{-3}$ sec.), RWMH turned out to be superior w.r.t. (I_1) and (I_2) as can be seen from Figure 1(a) and 1(b). Here, missing information about the correlation structure in the proposal function is clearly hampering the performance of IMH. While CIMH and ACIMH outperformed all other algorithms on (I_1), ACIMH and AM attained only comparable (I_2)-values. Due to the simplicity of the example and the prerun- $(9.6 \pm 2 \cdot 10^{-3}$ sec.) and copula-fitting time



(a) (b) (c)

Figure 1: Results for the two-dimensional normal distribution. Figure (a): Quotient of acceptance rate and INEFF, (I_1). Figure (b): ESS per second, (I_2). Error bars show the estimated standard errors based on 100 runs. Figure (c): Marginal copula data (c.f. Section 4.1 (ii)) used to fit the CIMH and ACIMH copula of the first run. The diagonal displays the histograms of the MCMC sample marginals and τ_e the corresponding empirical Kendall's τ.

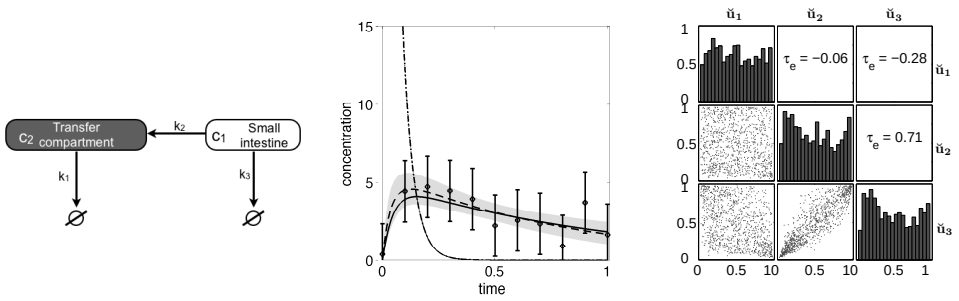
$(31.2 \pm 0.3$ sec.) saved by AM the latter result is not astonishing. Copula refitting did not decrease the INEFF (INEFF CIMH: $1.1 \pm 1 \cdot 10^{-2}$; INEFF ACIMH: $1.1 \pm 1 \cdot 10^{-2}$). Hence, saving the additional time needed for copula refitting, CIMH turned out to be the most efficient algorithm w.r.t. (I_2), outperforming all other algorithms by far. We have to point out that (I_1) is very close to one for CIMH and ACIMH. This means that in almost every MCMC iteration an independent sample was generated. At first sight this might almost seem too good of a result, but clearly, due to the simplicity of the problem, the copula was fitted almost perfectly (c.f. Figure 1(c)) leading to an independent proposal function $q^{cop}(\boldsymbol{\theta}|\boldsymbol{\theta}') = q^{cop}(\boldsymbol{\theta})$ that was very close to the true sampling distribution $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This pushed the MH acceptance probability close to one. (I_1) can hence be seen as a combined goodness-of-fit index for the fitted marginal cdf's and vine-copula decomposition. The copula families for ACIMH did not change in any of the 100 runs, meaning that the dependency structure was already well covered by the preruns. The bivariate copula $c_{1,2}(u_1, u_2|\boldsymbol{\eta})$ of the first of the 100 runs was fitted to be Gaussian with an estimated parameter value of $\hat{\boldsymbol{\eta}} = 0.953$. This is very close to the actual correlation value of $\rho = 0.95$. The corresponding Kendall's τ for the copula parameters was estimated to be $\hat{\tau}_m = 0.805$, which coincides with the Kendall's τ estimated for the prerun (c.f. Figure 1(c)). All other runs showed similar outcomes (results not shown). A nice connection between RWMH, AM and the copula-based algorithms is given by the fact that all four were using a Gaussian copula for proposal generation. However, as RWMH and AM were applying it only for locally proposing new samples, the (I_1) indices were low compared to the ones of CIMH and ACIMH.

5.2 Inference of a small compartmental model

We will now consider posterior inference in dynamical systems. We focus on the issue of parameter inference in DEs, a topic very prominent e.g. in the field of computational systems biology. Here, the DE parameters represent rate constants controlling the turn-over of biochemical substances. Despite the arrival of new, high-throughput measurement techniques, compared to model complexity most systems in this field suffer from very low observation numbers and noisy measurements. The current example is motivated by a model for the biokinetic behavior of zirconium (Zr) in the human body. Compartmentalizing major organs, Li et al. (2011) analyzed the circulation of Zr after ingestion. The paper compares transfer rates of two competing models w.r.t. sensitivity and predictability in order to establish a new model for radiation risk analysis. Both models are structurally identical as far as the interaction of "small intestine" and the "transfer compartment" is concerned, which is what our toy model is based on: after ingestion Zr passes through "small intestine". Subsequently it is either excreted directly or via the "transfer compartment" as depicted in (Figure 2(a)). Since taking accurate measurements of Zr in the "small intestine" compartment is technically not possible, we chose to generate data for the "transfer compartment" only. The differential equations underlying the data are

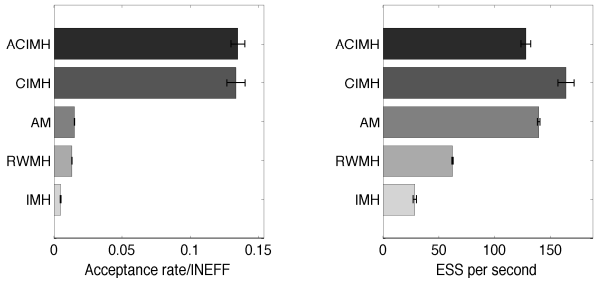
$$\frac{\mathrm{d}c_1(t)}{\mathrm{d}t} = -k_2 c_1(t) - k_3 c_1(t) \qquad \text{and} \qquad \frac{\mathrm{d}c_2(t)}{\mathrm{d}t} = k_2 c_1(t) - k_1 c_2(t), \tag{10}$$

making our model in parts similar to the ones proposed in Li et al. (2011). The dependency of $c_i(t)$ on k_1, k_2, and k_3 is omitted for readability. We generated our data for $k_1 = 1$, $k_2 = 1$, and $k_3 = 20$ at the time points $t_i = 0, 0.1, 0.2, \ldots 1.0$ as $y_i = c_2(t_i) + \varepsilon_i$ with $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1^2)$ for $i = 1 : 11$. Here, $c_1(0) = 100$ and $c_2(0) = 0$ is assumed to be known and the prior distributions were set to $k_1, k_2 \sim \mathcal{N}_{[0,1000]}(1, 1^2)$, and $k_3 \sim \mathcal{N}_{[0,1000]}(20, 20^2)$, where $\mathcal{N}_{[a,b]}(\mu, \sigma^2)$ denotes the $[a, b]$-truncated normal distribution. We started all algorithms at the true k_1-, k_2-, and k_3-values which made a burn-in phase unnecessary. The MCMC sample-based solutions to equation (10) nicely approximated the data. Figure 2(b) depicts the data as well as the posterior median solution with corresponding 95% credible interval for the first ACIMH run, i.e. at time point t equation (10) was solved numerically for all ACIMH MCMC samples; subsequently the point-wise median over all solutions – called the *posterior median solution* – and its 95% credible interval were computed. Note that neither the posterior median solution, nor the credible interval boundaries need to solve equation (10). As independence proposal density q_3 we chose a uniform distribution on $[0, 1000]^3$ and set $r_1 = 0.99$ and $r_2 = 0$. There is an interesting dependency structure between the parameters k_1, k_2, and k_3 inherent to the system. Recalling that the dependency structure between the k_i's and the $\breve{u}_i$'s is identical, k_2 and k_3 show strong positive non-symmetric lower tail dependency, while k_1 is almost independent of k_2 and k_3 (Figure 2(c)). This strong lower tail dependence was covered by a rotated 180° BB6 copula $c_{2,3}$ with estimated parameters $\hat{\eta}_{2,3} = (1.22, 3.08)$ and corresponding estimated Kendall's $\hat{\tau} = 0.71$ (see Brechmann and Schepsmeier (2013) for a definition of this copula type). The estimated Kendall's τ's of $\hat{c}_{2,3}$ nicely coincided with the estimated Kendall's τ's of the copula-samples (Figure 2(c)) which indicates a good parameter dependency coverage by $\hat{c}_{2,3}$.



(a) (b) (c)

Figure 2: (a) Schematic representation of the small compartmental model. The concentration of the shaded "transfer compartment" is measured at eleven time points $t_i = 0, 0.1, \ldots, 1.0$. The rates k_1 and k_3 lead to unobserved downstream compartments and are therefore considered as degradation rates. (b) Depicted are the true underlying concentration $c_2(t)$ of the "transfer compartment" (solid line), the posterior median solution (dashed line) as well as its 95% credible interval (shaded area) of the first ACIMH run. The dots depict noisy data y_i including the 95% confidence intervals of the normal error model placed on the observed data points (vertical lines). The unobserved concentration $c_1(t)$ of the "small intestine" compartment is shown as a dashed-dotted line. (c) Copula data (c.f. Section 4.1 (ii)) of the first run used to fit the CIMH copula. For uniformization of k_1, k_2, and k_3 the cdf's of $\mathcal{N}(1.25, 0.43^2)$, $\mathcal{N}(1.54, 0.57^2)$, and $\mathcal{N}(27.37, 11.10^2)$ were applied. The diagonal displays the histograms of the MCMC sample marginals and τ_e the respective empirical Kendall's τ.



(a) (b)

Figure 3: Results for the compartmental model. Figure (a): Quotient of acceptance rate and INEFF. Figure (b): ESS per second. Error bars show the estimated standard errors based on 100 runs.

As in our first example, compared to AM, CIMH, and ACIMH, RWMH and IMH

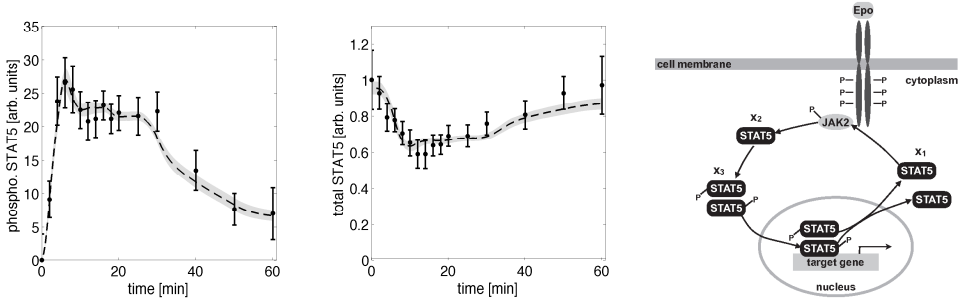
Sampler	IMH	RWMH	AM	CIMH	ACIMH
$M[k_1\|\boldsymbol{y}]$	1.17	1.16	1.18	1.19	1.18
$CI[k_1\|\boldsymbol{y}]$	(0.61;1.97)	(0.61;1.98)	(0.61;1.98)	(0.61;1.97)	(0.61;1.97)
$M[k_2\|\boldsymbol{y}]$	1.26	1.33	1.35	1.29	1.27
$CI[k_2\|\boldsymbol{y}]$	(0.74;2.56)	(0.73;2.57)	(0.73;2.56)	(0.73;2.57)	(0.73;2.58)
$M[k_3\|\boldsymbol{y}]$	23.84	22.14	21.70	23.99	22.62
$CI[k_3\|\boldsymbol{y}]$	(11.76;46.91)	(11.61;47.05)	(11.63;46.91)	(11.66;47.10)	(11.65;47.19)

Table 1: Small compartmental model. Estimated marginal posterior modes $M[\cdot|\boldsymbol{y}]$ (MAP estimates) and 90% posterior quantile-based credible intervals $CI[\cdot|\boldsymbol{y}]$ for k_1, k_2, and k_3 for the concatenated data of 100 runs.

ran into problems generating adequate proposals for the this time non-linear parameter dependency (compare Figures 3(a) and 3(b)). Although on average three times faster than ACIMH, AM performed equally well w.r.t. (I_2). Moreover, although taking on average more than 1.3 times as long as any non copula-based sampler, CIMH nicely detected the parameter dependency structure and yielded the best results. The average ESS increased slightly when updating the copulas (ESS CIMH: 1.0126 ± 440; ESS ACIMH: 10284 ± 360). This means that the copula structure is recursively adjusted to better fit the true underlying dependency structure of k_1, k_2, and k_3. Again, compared to CIMH the additional time for refitting the copula lowered the efficiency of ACIMH w.r.t. (I_2). For the inference of the marginal maximum a posteriori (MAP) estimates, we applied a kernel density estimator to the respective sampled Markov chains. The posterior mean and mode estimates including 90% credible intervals are given in Table 1. All predicted modes slightly overestimated the true values $k_1 = k_2 = 1$ and $k_3 = 20$.

5.3 Inference of the JAK2-STAT5 signaling pathway

We now apply our sampling schemes to the JAK2-STAT5 signaling pathway. Cellular signaling pathways are processing and transmitting intercellular signals in order to control processes such as cell cycle or immunoresponses (Cooper and Hausman (1997)). The mechanism is generally based on a number of phosphorylation and dephosphorylation steps within a complex protein interaction network. A graphical representation of the JAK2-STAT5 pathway is shown in Figure 4(c). It plays a key role in gene regulation (Subramaniam et al. (2001)) and is scientifically of major interest as malfunctioning results in diseases like leukemia or bronchial asthma (Igaz et al. (2001)).



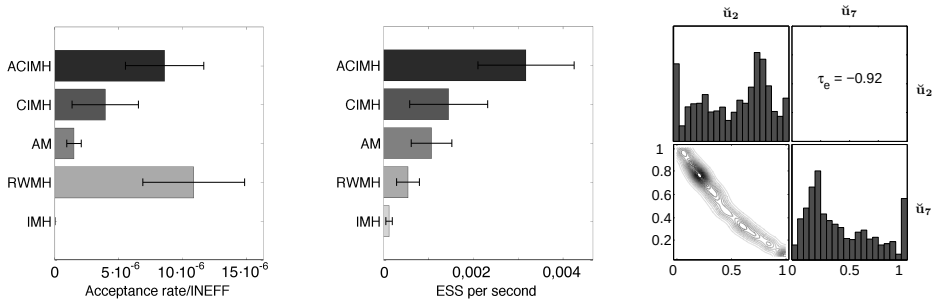
(a) (b) (c)

Figure 4: (a) Time courses for the numerical solution of phosphorylated STAT5 in the cytoplasm ($y_1(t)$). Depicted are the posterior median solution (dashed line) and the corresponding 95% credible interval (shaded area) of the first ACIMH run. The dots represent given measurements y_i including 95% confidence intervals (vertical lines). (b) Similarly to (a), the results for the numerical solution of total STAT5 in the cytoplasm ($y_2(t)$). (c) Graphical representation of the JAK2-STAT5 pathway: Erythropoietin (Epo) binds to the transmembrane receptor. Monomeric STAT5 (x_1) is tyrosine phosphorylated (x_2) by the activated JAK2/receptor complex in the cytoplasm. After dimerizing the phosphorylated JAK5-homodimer (x_3) enters the nucleus and binds to the promoter target gene region. It is then dephosphorylated and released back into the cytoplasm.

Our analysis is based on the data and DDE model of Swameye et al. (2003):

$$\begin{aligned}
\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} &= -k_1 x_1(t) Epo(t) + 2k_4 x_3(t+\tau) \\
\frac{\mathrm{d}x_2(t)}{\mathrm{d}t} &= -k_2 x_2^2(t) + k_1 x_1(t) Epo(t) \\
\frac{\mathrm{d}x_3(t)}{\mathrm{d}t} &= -k_3 x_3(t) + \frac{1}{2} k_2 x_2^2(t) \\
\frac{\mathrm{d}x_4(t)}{\mathrm{d}t} &= -k_4 x_3(t+\tau) + k_3 x_3(t),
\end{aligned} \tag{11}$$

with $x_1(0) = 1$ and $x_2(0) = x_2(0) = x_4(0) = 0$, where $Epo(t)$ denotes the time-dependent Epo stimulation function, τ the time-lag between STAT5 entering the nucleus and dephosphorylated cytoplasmic release, and $x_4(t)$ the concentration of STAT5 in the nucleus. Due to the law of mass conservation, we need to claim $k_3 \geqslant k_4$. The data we used for inference was provided by J.Timmer at http://webber.physik.uni-freiburg.de/~jeti/PNAS_Swameye_Data/. It contains (including 95% confidence intervals) the amount of phosphorylated STAT5, $y_1^\varepsilon(t_i) = k_5(x_2(t_i)+2x_3(t_i)+\varepsilon_1(t_i))$, and the total concentration of cytoplasmic STAT5, $y_2^\varepsilon(t_i) = k_6(x_1(t_i)+x_2(t_i)+2x_3(t_i)+\varepsilon_2(t_i))$, at 16 time points $t_1, \ldots, t_{16}$ (in minutes) in the interval $[0, 60]$. Here, k_5 and k_6 are introduced since all measurements are relative. The errors $\varepsilon_j(t_i)$ are measurement errors



(a) (b) (c)

Figure 5: Results for the JAK2-STAT5 model. Figure (a): Quotient of acceptance rate and INEFF. Figure (b): ESS per second. Error bars show the estimated standard errors based on 10 runs. (c) Density plot of the $(\breve{u}_2, \breve{u}_7)$ copula data-pair corresponding to (k_2, k_6) of the first ACIMH run. Darker contour lines depict higher, lighter ones lower density values.

included in the data, which are assumed to be $\mathcal{N}(0, \sigma^2_{i,j})$ distributed, where $\sigma^2_{i,j}$ was estimated from various experiments. All seven parameters $\boldsymbol{\theta} = (k_1, k_2, k_3, k_4, \tau, k_5, k_6)^\top$ are time-independent. Again, for readability the dependence of the solutions $x_i(t)$ to (11) on $\boldsymbol{\theta}$ is omitted. A picture of the data can be seen in Figure 4(a) and 4(b). Similarly to Swameye et al. (2003) we reparameterized the DDE system in order to resolve structural parameter identifiability issues. For more details on the reparameterization see the supporting text to Swameye et al. (2003). A discussion on the structural parameter identifiability issues of the particular system can be found in Timmer et al. (2004) and Raue et al. (2009). Due to the lack of knowledge we chose the independent prior distributions $k_1, k_2, k_3, k_4, \tau, k_5, k_6 \overset{i.i.d.}{\sim} \mathcal{U}[0, 50]$. The lower limit 0 was canonically introduced by the non-negativity constraint for reaction rates. Since there is no analytical solution to (11), we applied MATLAB's dde23 solver to numerically derive the solutions $x_i(t)$ for $i = 1 : 4$. As dde23 is quite time consuming, generating good proposals is essential for efficient sampling from the highly dependent seven-dimensional parameter distribution.

We started the inference by choosing the independence proposal density q_3 to be uniform on $[0, 50]^7$ and setting $r_1 = 0.7$ and $r_2 = 0.25$. All algorithms were initialized on the outcome of a simulated annealing run, making the correction for a burn-in phase unnecessary. A look at the copula data revealed that fitting standard pair-copulas to the data is rather involved: the density plot of the $(\breve{u}_2, \breve{u}_7)$-pair of the first run (c.f. Figure 5(c)), for instance, has a non-standard bent ridge shape with a very dense region at high $\breve{u}_7$ and low $\breve{u}_2$ values. Here, $(\breve{u}_2, \breve{u}_7)$ corresponds to (k_2, k_6). This fitting issue results in rather low acceptance rates for CIMH ($1.5\% \pm 0.2\%$) and ACIMH ($1.7\% \pm 0.1\%$). Moreover, the estimated Kendall's τ's of the pairs (k_2, k_5), (k_2, k_6), (k_3, k_4), and (k_5, k_6), computed to very high values of $\hat{\tau}_{2,5} = -0.87$, $\hat{\tau}_{2,6} = -0.92$, $\hat{\tau}_{3,4} = 0.97$, and $\hat{\tau}_{5,6} = 0.89$,

Sampler	IMH	RWMH	AM	CIMH	ACIMH
$M[k_1\|\boldsymbol{y}]$	0.03	0.03	0.03	0.03	0.03
$CI[k_1\|\boldsymbol{y}]$	(0.02;0.03)	(0.03;0.04)	(0.02;0.04)	(0.03;0.04)	(0.03;0.04)
$M[k_2\|\boldsymbol{y}]$	1.88	1.03	2.44	1.24	1.06
$CI[k_2\|\boldsymbol{y}]$	(1.84;2.48)	(0.95;2.96)	(1.43;4.39)	(1.14; 4.40)	(0.78; 4.74)
$M[k_3\|\boldsymbol{y}]$	0.14	0.16	0.17	0.14	0.24
$CI[k_3\|\boldsymbol{y}]$	(0.13;0.17)	(0.13;0.38)	(0.12;0.25)	(0.13;0.24)	(0.14;0.33)
$M[k_4\|\boldsymbol{y}]$	0.14	0.16	0.17	0.14	0.24
$CI[k_4\|\boldsymbol{y}]$	(0.13;0.17)	(0.13;0.38)	(0.12;0.25)	(0.13;0.24)	(0.14;0.32)
$M[\tau\|\boldsymbol{y}]$	4.80	3.74	3.85	3.68	3.95
$CI[\tau\|\boldsymbol{y}]$	(3.64;5.61)	(2.63;4.94)	(2.85;5.41)	(2.89;4.91)	(2.70;4.54)
$M[k_5\|\boldsymbol{y}]$	37.45	35.62	36.65	36.66	35.43
$CI[k_5\|\boldsymbol{y}]$	(33.77;37.60)	(33.50;39.13)	(33.74;38.93)	(33.51;38.45)	(34.10;39.14)
$M[k_6\|\boldsymbol{y}]$	0.93	0.94	0.96	0.95	0.94
$CI[k_6\|\boldsymbol{y}]$	(0.91;0.99)	(0.90;0.98)	(0.91;0.99)	(0.91;0.98)	(0.91;0.99)

Table 2: JAK-STAT5 pathway model. Estimated marginal posterior modes $M[\cdot|\boldsymbol{y}]$ (MAP estimates) and 90% posterior quantile-based credible intervals $CI[\cdot|\boldsymbol{y}]$ for the parameters $k_1, k_2, k_3, k_4, \tau, k_5$ and k_6 for the concatenated data of 10 runs.

respectively. This strong parameter-dependency structure hampers MCMC inference severely and led to very low (I_1)-values for all algorithms (Figure 5(a)). Nevertheless, CIMH and ACIMH outperformed IMH, RWMH, and AM w.r.t. (I_2) (Figure 5(b)). Especially ACIMH exceeded all non-copula based sampling schemes by an average of more than 3-fold w.r.t. (I_2). The prerun samples were transformed to $[0,1]^7$ using fitted normal densities for the margins of k_1, k_2, k_3, k_4, τ, and k_5 and a fitted lognormal density for the margin of k_6. Owing to the complexity of the system, we used 3,000 samples to fit all copulas involved. By sequential adjustment of the proposal function during the sampling process ACIMH could increase (I_1) and (I_2) compared to CIMH. The average number of pair-copula family updates in every ACIMH run was 47%, i.e. almost every second pair-copula was fit to have different copula types compared to the fit before. Table 2 shows the marginal posterior modes (MAP estimates) and 90% posterior quantile-based credible intervals for the concatenated data of all 10 runs. The estimates of the time τ a STAT5 molecule remains in the nucleus is ≈ 4 minutes. This means that the cytoplasmic release turns out to be a bit faster than the value of ≈ 6.4 minutes computed by Swameye et al. (2003). Nevertheless $\tau \approx 4$ minutes is contained in their confidence interval of $(3.8; 6.9)$ minutes. Although all other results coincide well, longer MCMC chains are necessary to obtain more reliable estimates. Overall, the system represents a challenging example for MH algorithms and is thus a good performance benchmark.

5.4 Sensitivity to the choice of copula and prerun transformation

Two very crucial factors for the performance of CIMH and ACIMH are the goodness-of-fit of (i) the pair-copula decomposition and (ii) the cdf's for prerun transformation. We

already inferred the effect of applying oversimplified copula decompositions: as the Independence copula is defined by $C : [0,1]^d \longrightarrow [0,1]$, $(u_1, \ldots, u_d) \mapsto \prod_{i=1}^d u_i$, the IMH essentially constitutes a CIMH algorithm with an Independence-copula based proposal function. The JAK2-STAT5 example showed that we can run into serious problems in fitting an appropriate decomposition. More involved techniques such as fitting mixtures of pair-copulas as well as non- or semi-parametric copula density estimation and sample generation (Hu (2006)) are therefore needed in future applications. To assess the misspecification effect, we re-sampled the example of Section 5.2 using fitted exponential distribution functions for CIMH. The ESS per second dropped by a factor of 2.9, while the index (I_1) even decreased by a factor of 10.1. In order to improve the misspecification issue while avoiding the need for manual definition of the marginal distribution types we propose to apply a Gaussian mixture model (Singh and Dattatreya (2006)) for prerun uniformization in future applications. Here, the marginal prerun samples $\{\breve{\theta}_i^{(k)}\}_{k=0:K'}$ of dimension i could be used to define the marginal pdf $g_i(x)$ for uniformization along dimension i as $g_i(x) = \sum_{k=0}^{K'} \frac{1}{K'+1} \varphi_{(\breve{\theta}_i^{(k)}, \sigma^2)}(x)$ for some user-defined bandwidth σ, where $\varphi_{(\mu,\sigma^2)}(x)$ denotes the univariate pdf to $\mathcal{N}(\mu, \sigma^2)$. The corresponding cdf is then for the error function erf given by $G_i(x) = \frac{1}{2} + \frac{1}{2} \sum_{k=0}^{K'} \frac{1}{K'+1} \mathrm{erf}\left((x - \breve{\theta}_i^{(k)})/\sigma\sqrt{2}\right)$, while $G^{-1}_{\iota(i)}(u)$ might be obtained by inverse interpolation. This approach would lower the need for exploratory data analysis and human interaction considerably. The uniformization results are moreover expected to be more stable. However, the efficiency of these Gaussian mixture models requires more thorough analysis. As mentioned above, (I_1) can here be taken as a goodness-of-fit index.

6 Conclusions

We have introduced a vine-copula based hybrid independence/random walk MH sampling scheme and tested its performance on two toy examples as well as a model of the JAK2-STAT5 signaling pathway. The basic algorithm was extended by a sequential copula updating scheme leading to an adaptive MCMC approach. Both algorithms were evaluated on the basis of the quotient of acceptance rate and inefficiency factor and the effective sampling size per second. As competing samplers a random walk MH, an independence sampler, as well as an adaptive Metropolis algorithm were chosen. Our copula-based approach generally covered the dependency structure of the posterior very well and outperformed all other sampling schemes in every example. It turned out that the basic hybrid algorithm performs best on simple systems as it does not lose time on extra copula updates. However, in very complex situations, such as the inference in the JAK2-STAT5 pathway, copula updates were needed to fine-tune the proposal distribution and thereby improve the performance. We primarily focused on parameter inference in dynamical systems. However, the field of application is not limited to this scenario. Both algorithms can be applied to any MCMC inference problem. They are expected to work well on highly dependent posterior distributions, but also very efficiently in simple systems. Further research is needed to improve the algorithms for sampling from highly complex posterior distributions. The JAK2-STAT5 pathway indicated that

non-standard copula and marginal distributions might be needed to guarantee efficient sampling. Gaussian mixture models for prerun sample transformation as well as non-parametric pair-copula distributions would be a first step to improve performance. In addition to an automated choice of univariate margins, first sequential selection methods of the vine structure, its pair copula families and corresponding parameters have been suggested in Dißmann et al. (2013) and implemented in the R package *VineCopula* (Schepsmeier et al. (2012)). These automated methods are able to handle parameter dimensions between 20 and 30, but their performance is subject to future research.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). "Pair-copula constructions of multiple dependence." *Insurance, Mathematics and Economics*, 44: 182–198. 2, 4, 6

Bayes, M. and Price, M. (1763). "An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS." *Philosophical Transactions (1683-1775)*, 370–418. 3

Bedford, T. and Cooke, R. (2001). "Probability density decomposition for conditionally dependent random variables modeled by vines." *Annals of Mathematics and Artificial Intelligence*, 32: 245–268. 4

— (2002). "Vines - a new graphical model for dependent random variables." *Annals of Statistics*, 30(4): 1031–1068. 4

Beichl, I. and Sullivan, F. (2000). "The Metropolis algorithm." *Computing in Science & Engineering*, 2(1): 65–69. 2

Brechmann, E. (2010). "Truncated and simplified regular vines and their applications." Diploma thesis, Technische Universität München. 5

Brechmann, E. and Schepsmeier, U. (2013). "CDVine: Modeling Dependence with C- and D-Vine Copulas in R." *Journal of Statistical Software*, 52(3): 1–27. 5, 10, 12

Brooks, S. (1998). "Markov chain Monte Carlo method and its application." *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1): 69–100. 2

Brown, K. and Sethna, J. (2003). "Statistical mechanical approaches to models with many poorly known parameters." *Physical Review E*, 68(2): 021904–1–021904–9. 2

Cooper, G. and Hausman, R. (1997). *The cell: a molecular approach*. Washington DC: ASM Press. 14

Czado, C. (2010). "Pair-copula constructions." In Jaworki, P., Durante, F., Härdle, W., and Rychlik, T. (eds.), *Copula Theory and its Applications*, 93–110. Dortrecht (NL): Springer. 4

Dißmann, J., Brechmann, E., Czado, C., and Kurowicka, D. (2013). "Selecting and estimating regular vine copulae and application to financial returns." *Computational Statistics & Data Analysis*, 59: 52–69. 19

Gamerman, D. and Lopes, H. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*, volume 68. Boca Raton: Chapman & Hall/CRC. 2

Girolami, M. and Calderhead, B. (2011). "Riemann manifold Langevin and Hamiltonian Monte Carlo methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2): 123–214. 2, 9

Haario, H., Saksman, E., and Tamminen, J. (2001). "An adaptive Metropolis algorithm." *Bernoulli Journal*, 7(2): 223–242. 2, 3

Hastings, W. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57(1): 97–109. 2

He, Y., Hodges, J., and Carlin, B. (2007). "Re-considering the variance parameterization in multiple precision models." *Bayesian Analysis*, 2(3): 529–556. 9

Hobæk Haff, I. (2013). "Parameter estimation for pair-copula constructions." *Bernoulli Journal*, in Press. 5

Hobæk Haff, I., Aas, K., and Frigessi, A. (2010). "On the simplified pair-copula construction–Simply useful or too simplistic?" *Journal of Multivariate Analysis*, 101(5): 1296–1310. 4

Hoffman, M. and Gelman, A. (2011). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *arXiv preprint arXiv:1111.4246*. 9

Holden, L. (2000). "Convergence of Markov chains in the relative supremum norm." *Journal of Applied Probability*, 37(4): 1074–1083. 7

Holden, L., Hauge, R., and Holden, M. (2009). "Adaptive independent Metropolis–Hastings." *The Annals of Applied Probability*, 19(1): 395–413. 2, 7

Hu, L. (2006). "Dependence patterns across financial markets: a mixed copula approach." *Applied Financial Economics*, 16(10): 717–729. 18

Igaz, P., Toth, S., and Falus, A. (2001). "Biological and clinical significance of the JAK-STAT pathway; lessons from knockout mice." *Inflammation Research*, 50(9): 435–441. 14

Joe, H. (1996). "Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters." In L. Rüschendorf and B. Schweizer and M. D. Taylor (ed.), *Distributions with Fixed Marginals and Related Topics*, volume 28 of *IMS Lecture Notes – Monograph Series*. 4

— (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall. 3

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). "Optimization by simulated annealing." *Science*, 220(4598): 671–680. 9

Kurowicka, D. and Cooke, R. (2006a). "Completion problem with partial correlation vines." *Linear Algebra and its Applications*, 418: 188–200. 2

— (2006b). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Chichester: Wiley. 4

Kurowicka, D. and Joe, H. (2011). *Dependence Modeling - Handbook on Vine Copulae*. Singapore: World Scientific Publishing Co. 2, 4

Lawrence, N., Girolami, M., and Rattray, M. (2010). *Learning and Inference in Computational Systems Biology*. Cambridge: The MIT Press. 2

Li, W., Greiter, M., Oeh, U., and Hoeschen, C. (2011). "Reliability of a new biokinetic model of zirconium in internal dosimetry. Part II. Parameter sensitivity analysis." *Health Physics*, 101(6): 677–692. 12

Liu, J. (2008). *Monte Carlo Strategies in Scientific Computing*. New York: Springer. 2, 9

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., et al. (1953). "Equation of state calculations by fast computing machines." *The Journal of Chemical Physics*, 21(6): 1087–1092. 2

Min, A. and Czado, C. (2010). "Bayesian inference for multivariate copulas using pair-copula constructions." *Journal of Financial Econometrics*, 8(4): 511–546. 5

— (2011). "Bayesian model selection for multivariate copulas using pair-copula constructions." *Canadian Journal of Statistics*, 39: 239–258. 5

Nelsen, R. (2006). *An Introduction to Copulas*. New York: Springer. 3

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood." *Bioinformatics*, 25(15): 1923–1929. 16

Robert, C. and Casella, G. (2004). *Monte Carlo statistical methods*. New York: Springer. 3

Roberts, G., Gelman, A., and Gilks, W. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithms." *The Annals of Applied Probability*, 7(1): 110–120. 9

Roberts, G. and Rosenthal, J. (2007). "Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms." *Journal of Applied Probability*, 44(2): 458–475. 2, 7

Rosenthal, J. (2011). "Optimal proposal distributions and adaptive MCMC." *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Press, Boca Raton*, 93–112. 2

Schepsmeier, U., Stoeber, J., and Brechmann, E. (2012). *VineCopula: Statistical inference of vine copulas*. Technische Universität München, Munich, Germany. URL http://cran.r-project.org/web/packages/VineCopula/index.html 19

Singh, L. and Dattatreya, G. (2006). "Gaussian Mixture Parameter Estimation for Cognitive Radio and Network Surveillance Applications." *WSEAS Transactions on Communications*, 5(3): 423–428. 18

Sklar, A. (1959). "Fonctions dé repartition á n dimensions et leurs marges." *Publications de l'Institut de Statistique de l'Université de Paris*, 8: 229–231. 3

Smith, M., Min, A., Almeida, C., and Czado, C. (2010). "Modeling longitudinal data using a pair-copula construction decomposition of serial dependence." *Journal of the American Statistical Association*, 105: 1467–1479. 5

Subramaniam, P., Torres, B., and Johnson, H. (2001). "So many ligands, so few transcription factors: a new paradigm for signaling through the STAT transcription factors." *Cytokine*, 15(4): 175–187. 14

Swameye, I., Müller, T., Timmer, J., Sandra, O., and Klingmüller, U. (2003). "Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling." *Proceedings of the National Academy of Sciences of the United States of America*, 100(3): 1028–1033. 8, 15, 16, 17

Timmer, J., Müller, T., Swameye, I., Sandra, O., and Klingmüller, U. (2004). "Modeling the nonlinear dynamics of cellular signal transduction." *International Journal of Bifurcation and Chaos*, 14(6): 2069–2079. 16

Wilkinson, D. (2006). *Stochastic Modelling for Systems Biology*. Boca Raton: Chapman & Hall/CRC. 2

— (2007). "Bayesian methods in bioinformatics and computational systems biology." *Briefings in Bioinformatics*, 8(2): 109–116. 2

Acknowledgments

This work was supported by the Federal Ministry of Education and Research (BMBF) in its MedSys initiative (project SysMBO), and the European Union within the ERC grant LatentCauses. The authors wish to thank Eike Brechmann and Ulf Schepsmeier for their support and adjustments regarding the CDVine R-package. We are very grateful to Christiane Fuchs for helpful discussions and Justin Feigelman and Kary Myers for carefully proof reading the manuscript. Furthermore, we would like to thank Professor Herbie Lee as well as the anonymous reviewer and associate editor for improving the manuscript considerably. A special thank you to Professor Marina Vannucci for the opportunity to present this paper as an invited discussion article.