Probabilistic Multivariate Early Warning Signals

Ville Laitinen(✉) and Leo Lahti

Department of Computing, University of Turku, Turku, Finland
{ville.laitinen,leo.lahti}@utu.fi

Abstract. A broad range of natural and social systems from human microbiome to financial markets can go through critical transitions, where the system suddenly collapses to another stable configuration. Anticipating such transition early and accurately can facilitate controlled system manipulation and mitigation of undesired outcomes. Generic data-driven indicators, such as autocorrelation and variance, have been shown to increase in the vicinity of an approaching tipping point, and statistical early warning signals have been reported across a range of systems. In practice, obtaining reliable predictions has proven to challenging, as the available methods deal with simplified one-dimensional representations of complex systems, and rely on the availability of large amounts of data. Here, we demonstrate that a probabilistic data aggregation strategy can provide new ways to improve early warning detection by more efficiently utilizing the available information from multivariate time series. In particular, we consider a probabilistic variant of a vector autoregression model as a novel early warning indicator and argue that it has certain advantages in model regularization, treatment of uncertainties, and parameter interpretation. We evaluate the performance against alternatives in a simulation benchmark and show improved sensitivity in warning signal detection in a common ecological model encompassing multiple interacting species.

Keywords: Early warning signals · Probabilistic programming · Complex systems

1 Introduction

The ability to anticipate and manage change plays a critical role in diverse domains from biomedicine to ecology, economics, or climate change [32]. Natural and social systems are complex arrangements of units and their interactions at various scales. Despite their complexity and size, such systems often exhibit remarkable stability where perturbations have only minor, and often temporary and reversible effects on the system. However, when the conditions are stretched far enough, a system may pass a critical threshold, a tipping point, leading to a rapid and potentially irreversible reorganization. These *critical transitions* can be observed across many different scenarios [23], including ecosystems [21,34], epidemics [27], and climate

I. Petre and A. Păun (Eds.): CMSB 2022, LNBI 13447, pp. 259–274, 2022.
https://doi.org/10.1007/978-3-031-15034-0_13

[24]. As large transitions may have far-reaching consequences, the ability to anticipate them can provide valuable tools to manage change.

Generic early warning signals (EWS) have been introduced to detect signs that could alarm us about approaching tipping points, bifurcation points in the parameter space of a complex system after which transition to an alternative regime become inevitable [9,33]. A major challenge is that the sequence of events leading to critical transitions can be subtle and gradual, with little or no apparent changes in the observable system state [16]. However, the underlying system dynamics may change in ways that can be observed and quantified. For instance, measurable aspects of system resilience tend to decrease as a tipping point is approaching [33]. This is associated with *critical slowing down*, which can be indicated by certain statistical properties such as autocorrelation and variance. Increasing lag-1 autocorrelation and variance are some of the most robust and widely utilized EWS [8,12,13]. Autocorrelation tends to increase with slowing down, because the systemic rate of change decreases, and as a result the states at over short time intervals tend to become more and more alike. Similarly, changes in variance are arising as the lowering mean-reversion rate allows the state variable to fluctuate more freely. An important property of statistical early warning indicators is their generality. They provide data-driven quantification that can be informative even when our understanding of the data generating processes is limited. This facilitates the detection of early warnings even when accurate mechanistic modeling is infeasible due to the complexity of the phenomena and limitations in data collection, and makes the generic EWS indicators applicable across a broad range of different systems [23,32].

Despite the recent advances, the generic EWS often rely on the availability of relatively long and dense time series and manual parameter tuning. The ability to utilize the available data more efficiently and detect EWS from more limited time series would be important in many application fields such as ecology and human biomedical studies where the sample sizes can be remarkably low due to ethical, financial, or other constraints. Quantification of uncertainty is another key aspect in EWS analysis. Data is always limited, and may come with uneven observation times, measurement noise, or possible biases. The ability to quantify and control uncertainty is particularly relevant with limited sample sizes. The probabilistic framework provides tools to incorporate uncertainty and prior information into the models [14], and could lead to a more sensitive EWS detection. We recently demonstrated this by introducing a univariate probabilistic method for EWS detection [22], showing improvements in automated model selection and increased sensitivity in EWS detection performance. Here, we extend this earlier work into a multivariate context.

The univariate representation provides convenient, intuitively appealing, and robust ways to summarize changes in complex systems. Yet, the reliance on one-dimensional summaries may neglect important information that can be used to enhance the EWS detection. An enhanced use of multivariate observations in EWS design can provide improved sensitivity especially in shorter multivariate time series, where the information content of the data is more limited.

The advantages of probabilistic methods in treating uncertainties, and the potential for improving EWS detection with data aggregation strategies motivated us to investigate the possibility of extending our earlier work on probabilistic EWS into the multivariate domain.

Hence, we are in this work designing and investigating a novel probabilistic EWS indicator for multivariate systems. More specifically, we formulated and implemented a probabilistic variant of the time-varying vector autoregressive-1 model. A non-probabilistic version of a similar model was recently studied in [18]. The probabilistic version allows alternative ways to pool information across the multivariate time series and deal with uncertainties in the modeling. Moreover, the proposed method supports automated parameter inference, circumventing the need to manually select model parameters, such as sliding window size, which have posed problems in many EWS methods [12]. Besides these theoretically appealing properties, simulations based on an ecological model that has been commonly used in the EWS context demonstrate possible benefits against the currently available alternatives.

The work is structured as follows. In *Methods* we describe the novel approach along with a set of previously studied EWS indicators. We then compare these indicators in a simulation study in *Results* and, finally, conclude in *Discussion* with suggestions for further extension.

2 Methods

In this section, we provide a short overview of the currently available, related methods based on a recent review [40]. We formulate the probabilistic variant, tvPVAR(1), and describe the simulation model that is used to generate data for the experiments.

2.1 Autocorrelation Based EWS

Let us start by summarizing relevant methodology based a recent comparison between currently available indicators for detecting EWS in multivariate data [40]. In the present work we focus on autocorrelation-based indicators since these have shown robustness compared to the alternatives [13,40], and can be naturally extended into the probabilistic framework that we explore in this study. More specifically, the methods detecting changes in lag-1 autocorrelation are based on the standard autoregressive-1 process, AR(1), defined by the recursion.

$$X_{t+1} = \phi X_t + \sigma \epsilon_t \tag{1}$$

where X_t is the centered (zero-mean) state variable at time t, ϕ the autoregressive parameter, ϵ_t a zero-mean, unit variance Gaussian random variable scaled with σ. The main interest here lies in the autoregressive parameter ϕ, which directly measures the lag-1 autocorrelation of the system.

Many variants and extensions of the AR(1) process have been studied in univariate context [12], and applications in multivariate data are also possible.

For instance, *maximum autocorrelation* (ac/max) [10] is based on fitting the AR(1) model separately to each node, or feature, of the system and then selecting the one with the highest autocorrelation as a proxy for the entire system. Other options include *average autocorrelation* (ac/mean) across the features, or *degenerate fingerprinting* which measures autocorrelation along the first principal component of the multivariate data [17]. Min/Max autocorrelation factors analysis (MAF) [39] is another method based on dimension reduction that aims to identify the subspace with the highest autocorrelation in a multidimensional system. This algorithm generates a set of vectors (MAFs) to project the multidimensional data onto a subspace where autocorrelation is maximized. Eigenvalues of the MAF subspace quantify autocorrelation in the respective directions, with lower values indicating higher autocorrelation. In addition to *MAF eigenvalues* (eigen/MAF), we used autocorrelation (ac/MAF) and variance (var/MAF) projected onto the 1st MAF as indicators. For a more detailed description of the these methods, see [39]. We did omit some of the methods considered in [40], such as information dissipation length [29] and time [28], since they require larger amounts of data, and our interest lies mainly in practically motivated situations where the sample sizes are modest.

The EWS detection based on these previously suggested indicators was carried out following standard procedures [12]. We estimated the early warning indicators in sliding windows along the time series, resulting in a trajectory of the indicator, which is the autoregressive parameter ϕ in our case. Except where otherwise noted, we set the sliding window to 50% of length of the time series, which is a common default choice in the EWS literature. In order to remove the effect of non-stationary trends in the data that could lead to spurious conclusions [12], we used Gaussian detrending (R function *stats::smooth*) as a preprocessing step before quantifying the indicator. We used a bandwidth of 10% of the total time series length, except where otherwise noted, which we chose based on visual assessment; the bandwidth length was chosen so that it removes long-term mean level variations unrelated to the short term correlation structure while aiming to avoid overfitting to short-term variations.

We then measured the strength of each estimated EWS by computing Kendall's rank correlation τ between the estimated autocorrelation trajectory and time. The rank correlation receives values in $[-1, 1]$, with $\tau = 1$ indicating a monotonously increasing trajectory, and strong a EWS, while $\tau \approx 0$ implies a negative finding. The rank correlation is defined as $\tau = (N_{\text{concordant pairs}} - N_{\text{disconcordant pairs}})/N_{\text{all pairs}}$, where N refers to the number of elements in the subscript set, and a pair $(t_i, \phi_i), (t_j, \phi_j)$ is said to be concordant if $t_i \leq t_j$ implies $\phi_i \leq \phi_j$ and disconcordant otherwise.

For hypothesis testing on these EWS indicators, we utilized the so-called surrogate data analysis methods [12]. This technique generates an approximate sampling distribution for Kendall's τ, which is then compared to the actual point estimate. The sampling distribution represents results that would be recovered under the null hypothesis that the indicator trend has arisen simply by change. We generated a collection of time series from the simulation model presented

below in the Subsect. 2.3. We used constant parameters that produce data where the conditions remain constant and any estimated parameter trajectory is expected to have no correlation with time ($\tau = 0$). We generated 500 replicates of surrogate data for each experimental condition, and then estimated the EWS indicators and Kendall's rank correlations for these surrogate data sets, yielding approximate sampling distributions under the null hypothesis. P-values for a positive trend value were then computed as the proportion of the sampling distribution that exceeded (or were identical to) the actual point estimate.

2.2 The Probabilistic Time-Varying Vector Autoregressive-1 Process

We recently studied a probabilistic time-varying AR(1) process for detecting autocorrelation changes in univariate systems [22]. Here we investigate a multivariate extension of this model, the time-varying probabilistic vector AR(1) model, tvPVAR(1); for simplicity, we refer to this method as *ac/pooled* in the later comparisons. A non-probabilistic state space variant of this model was previously studied in [18]. Compared to the standard AR(1) process in Eq. 1, the time-varying model allows time-dependent variation in the model parameters.

The tvPVAR(1) model is defined as the recursion

$$\boldsymbol{X}_{t+1} = \Phi_t \cdot \boldsymbol{X}_t + \boldsymbol{\epsilon_t} \tag{2}$$

where $\boldsymbol{X}_t$ is the D-dimensional centered state vector at time t, Φ_t the autoregressive matrix, ϵ_t the multivariate Gaussian random variable with covariance matrix Σ. The degrees of freedom grow rapidly as a function of dimensionality, which makes parameter estimation challenging especially when the sample size is low compared to the dimensionality of the data. We are hence making certain simplifying assumptions. First, we assume that Σ is diagonal and constant over time. Second, we assume that $\Phi_t = \phi_t I$, where ϕ_t is a real number for all t and I is the identity matrix. The latter assumption amounts to parameter pooling (whence the name ac/pooled), which means that a single parameter (ϕ_t) represents several units. Intuitively, this provides a measure for the average systemic autocorrelation.

The probabilistic formulation requires us to define the likelihood of the data, and the priors for the model parameters. Likelihood for the data $\boldsymbol{X_t}$, $t = 1, \ldots, T$ is given by

$$\mathcal{L}(\Phi_t, \Sigma | \boldsymbol{X}_t) = \prod_{t=1}^{T-1} \mathrm{MVN}(\boldsymbol{X}_{t+1} | \phi_t I \cdot \boldsymbol{X}_t, \Sigma), \tag{3}$$

where MVN refers to the multivariate normal distribution.

Regarding the prior distribution, we use a Gaussian process (GP) prior for ϕ_t. A Gaussian process $\mathcal{GP}(M, K)$ is formally defined as a collection of random variables where each finite set of these variables is multivariate normally distributed with mean M and covariance K [30]. GPs are a standard choice for Bayesian nonparametric regression, and incorporating them in the model forms a key difference compared to the autocorrelation metrics presented in Subsect. 2.1.

More specifically, GPs, as we utilize them, provide a means to regularize the posterior of ϕ_t towards areas of the parameter space that correspond to behaviour likely to be encountered in real systems, such as differentiability.

We utilize the Matèrn-3/2 covariance function that models the covariance between two random variables X_i and X_j as $k_{3/2}(\rho, \alpha) = \alpha^2\left(1 + \frac{\sqrt{3}r}{\rho}\right)\exp(-\frac{\sqrt{3}r}{\rho})$, where α^2 is the process variance, ρ the length scale and $r = |X_i - X_j|$ [30,37]. In general, the Matèrn class is larger set of covariance functions characterized by a parameter ν which determines the level of differentiability of the output functions. We set $\nu = 3/2$, which restricts the posterior of ϕ_t to continuous and differentiable functions. This is a reasonable condition that allows flexibility in the model while avoiding overfitting to occassional large deviations. The length scale parameter ρ controls the dependence over time, while the process variance α^2 controls the average distance from the mean M. We set $M = 0$ and $\alpha = 1$, which restrains a majority of the prior values between -1 and 1. This is a justified choice as autoregressive-1 models are stationary if and only if the autoregressive parameter is within this interval. Length scale ρ was set, unless otherwise noted, to the length of the time series. We used the Cholesky factored parameterization of GPs for posterior sampling [20]. This models the process as a latent vector η which is mapped to the output space as $\phi_t = L\eta$ where L is the lower triangular matrix with positive diagonal from the Cholesky decomposition $k_{3/2} = LL^T$.

The fitting procedure is illustrated in Fig. 1. Hypothesis testing was carried out by first computing Kendall's τ for each posterior sample for ϕ_t. This provides a posterior distribution for τ, and the mass of this distribution on the positive side of the real line reflects the posterior probability of an increasing autocorrelation in the time series. The "Bayesian P-value" [14] can then be computed as the proportion on the negative side, facilitating comparison with non-probabilistic methods that generate frequentist P-values. While the Bayesian posterior proportion and frequentist P-value follow non-analogous logics and interpretations, they are comparable as metrics for statistical evidence, and below we use them interchangeably for a lack of choice. We set the EWS detection level at $P = 0.1$ and compare the methods in terms of the standard true positive rate (TPR) and true negative rate (TNR).

We implemented the tvPVAR(1) model in the probabilistic programming language Stan [36], utilizing the R interface RStan, and used its No-U-Turn variant of the Hamiltonian Monte Carlo algorithm with 2 chains, both with 2000 iterations for a given fit to sample the posterior. Sampling convergence was assessed with the $\hat{R}$ statistic which remained below the recommended limit 1.1 [15]. In addition, we encountered no divergent transitions indicating that the algorithm had converged and produced reliable estimates.

2.3 Simulation Model

We evaluate the performance of the EWS detection methods based on simulations from a well-studied ecological model [25]. The model characterizes systems with competition and mutualism, such as plant-pollinator interactions.

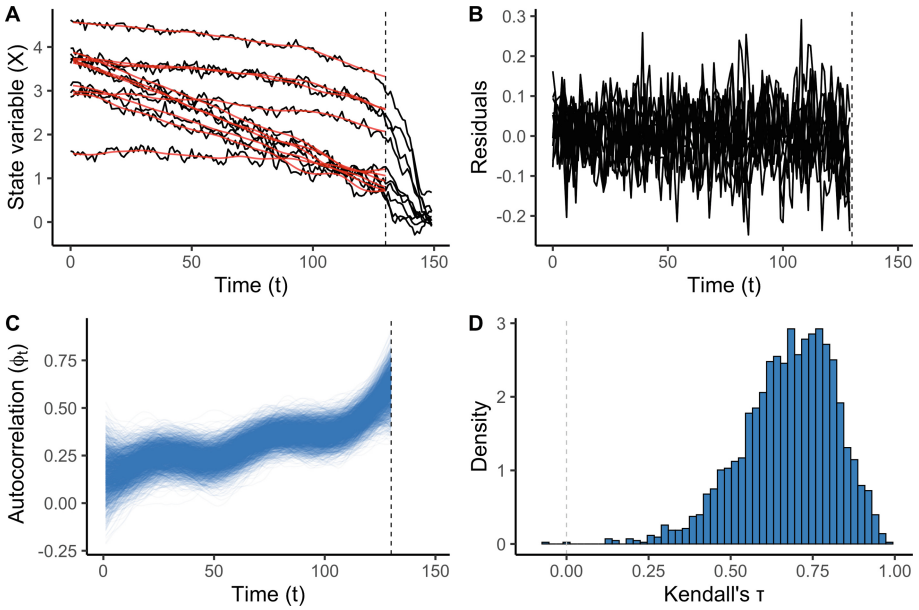

Fig. 1. Early warning signal detection with the proposed ac/pooled indicator (tvP-VAR(1) model) in simulated data. **A** The observed system state $X_i, i = 1, \ldots, N$ as a function of time (black) with estimated time series trends (red). The detrending is based on Gaussian kernel smoother applied separately on the individual features and removes mean-level variations unrelated autocorrelation. The system gradually approaches a tipping point before a system-wide collapse occurs starting at the vertical dashed line ($T = 134$). **B** Residuals from the detrending process are used to look for the EWS. **C** An increasing trend can be observed in the posterior samples of the autocorrelation parameter ϕ_t from tvPVAR(1). **D** Posterior of the Kendall's rank correlation for ϕ_t. More than 99.9% of the posterior mass is above 0 indicating strong evidence for an increasing average autocorrelation and an EWS before the observed state transition.

The deterministic part of the model consists of logistic growth which is stimulated by intergroup mutualism and limited by competition within the same group. The model is defined by the stochastic differential equation

$$dX_i^{(P)} = X_i^{(P)} \left(r_i^{(P)} + \frac{\Sigma_{k=1}^{S_P} \gamma_{ik}^{(P)} X_i^{(A)}}{1 + h^{(P)} \Sigma_{k=1}^{S_P} \gamma_{ik}^{(P)} X_i^{(A)}} - \Sigma_{l=1}^{S_A} c_{il}^{(P)} X_l^{(P)} \right) dt + \sigma_j^{(P)} dW$$

$$dX_j^{(A)} = X_j^{(A)} \left(r_j^{(A)} + \frac{\Sigma_{k=1}^{S_A} \gamma_{jk}^{(A)} X_j^{(P)}}{1 + h^{(A)} \Sigma_{k=1}^{S_A} \gamma_{jk}^{(A)} X_j^{(P)}} - \Sigma_{l=1}^{S_P} c_{jl}^{(A)} X_l^{(A)} \right) dt + \sigma_j^{(A)} dW$$

where X_i represents the abundance of species i, r the growth rate and h the half saturation constant, which was set to 0.5 for all species. The matrices γ_{ik} and c_{ij} represent the intergroup mutualism and interspecies competition, respectively.

The last term is the stochastic part of the system, a Wiener process with variance σ_i^2. The superscripts P and A refer to pollinators and plants, respectively.

The system can be pushed towards a critical transition by gradually decreasing the growth rate of the pollinator species [25], which could result from increasingly harsh environmental conditions, for instance. We randomly sampled initial pollinator growth rates $r^{(P)}$ from $\mathcal{N}(0, 0.1^2)$, and decreased them linearly to -1.5 during the simulation time, except in the cases where a group of pollinators were left undisturbed. In the latter case the growth rate was kept constant over the simulation. We randomly sampled rest of the parameters from the following distributions: $r_i^{(A)} \sim \mathcal{N}(-0.1, 0.05^2)$; $\gamma_{ij}^{(A)}, \gamma_{ij}^{(B)} \sim \text{Unif}(0.6, 1)$ for the off-diagonal elements and $\gamma_{ij}^{(A)} = \gamma_{ij}^{(B)} = 1$ when $i = j$; $c_{ij}^{(A)}, c_{ij}^{(B)} \sim \text{Unif}(0, 0.1)$ for the off-diagonal elements and $c_{ij}^{(A)} = c_{ij}^{(B)} = 0.3$ when $i = j$. We set the initial abundances to 2.5 for all species and then simulated the dynamics for 20 time points with constant conditions, during which the system settled into a stable state, and then discarded this settling period before EWS analysis. The stochastic noise σ was set to 0.1 in all cases. The chosen parameter sampling distributions and constants were based on previous studies [25,40].

We used the Euler-Maruyama discretization with time-step $\Delta t = 0.01$ to simulate the model and discarded all but every 100th observation, giving integer valued time points. For each replicate we required all species to be present after the settling period and defined a species to be present if its abundance is larger than 0.05. If this condition was not met, we repeated the parameter sampling process until a viable community emerged. We defined extinction to occur when any of the community species fell below 0.05. We only used the part preceding extinction in the EWS detection.

To assess the performance and robustness of the various EWS indicators in different settings, we generated data sets with varying data characteristics. For the first part of the experiments we simulated a community with $D = 10$ species in total, varied the number of perturbed pollinators from 1 to 5 and simulated $T = 150$ time points per replicate with no observation error. In the second part we simulated data with $D = 10$, $T = 150$ and included random Gaussian observation error, with standard deviations 0, 0.05, 0.1 and 0.2. In the third part we used time series lengths of $T = 50, 150, 250$, with $D = 10$ and no observation error. In the final part we varied the total number of features $D = 4, 10, 20$, with $T = 150$ and no observation error. In order to assess the indicators' specificity, we also generated data with corresponding data characteristics but where the conditions were kept constant, and no extinction took place. In each distinct set of data characteristics 50 replicates were generated, and in every simulation half of the community species were plants and half were pollinators.

3 Results

3.1 Simulation Benchmark

In this section we compare the performance of the alternative EWS indicators presented in *Methods* in ecological simulations. We limit our presentation here to

the five top-performing autocorrelation-based methods, based on average TPR over all of our experiments. This filtering process excluded the degenerate fingerprinting and MAF eigenvalue indicators, and retained the ac/pooled, ac/mean, ac/max, ac/MAF and var/MAF.

We studied the effect of four different data characteristics on the detection performance: the number of perturbed features when the full dimensionality is kept constant (at 10), Gaussian additive observations error with four levels of standard deviations, time series length, and the total dimensionality of the system. Figure 2 provides a graphical presentation of the results.

The true positive rate (TPR) in EWS detection increased with the number of affected features (the number of different pollinator species). No clear differences between the alternative indicators were observed when only 1–2 features were affected. However, when a larger fraction of the system, or a higher number of features was affected, a clear distinction between the methods emerges and ac/pooled achieves superior classification accuracy.

Regarding observation error, the performance for all indicators decreased at the error level 0.1 or higher, compared to the noise-free case. At the lower error levels we observed mixed results, with increasing accuracy in some cases. Random variation may explain these differences (ANOVA $F = 0.33$, $P = 0.86$ between error levels 0 and 0.05).

Increasing time series length led to a better accuracy. At $T = 50$ ac/pooled achieved a TPR of approximately 0.5, whereas the other models performed only slightly above the theoretical level for random guess, 0.1. We observed a large further improvement in EWS detection with the longer $T = 150$ set. Difference between 150 and 250 time points did not amount to a large improvement in TPR. This would either imply that the accuracy began to saturate, or that substantially larger amounts of samples are needed for further improvements in the TPR. We have omitted the analysis of longer time series in the present work because the longer time series are increasingly slow to model, and because our analysis is primary motivated by the practically important set of biomedical and ecological applications where the availability of longitudinal observations is limited to a few dozen time points.

By varying the total number of features in the data we observed, perhaps surprisingly, that EWS detection accuracy was best at the lowest-dimensional system $D = 4$. The dimensionalities of $D = 10$ and $D = 20$ had reduced, and approximately similar performance. The better performance at the lowest dimension level might be explained by a lower level of mutualistic links, which in turn causes transients to be short lived. This could be detected visually from the time series, as the lower dimension cases experienced a more sudden collapse compared to the higher dimensional cases that collapsed in a more gradual fashion.

In summary, our proposed indicator ac/pooled achieved the best performance (average TPR over all experiments 0.71), compared to ac/mean (0.51), var/MAF (0.4), ac/MAF (0.34), ac/max (0.31). One-way analysis of variance (ANOVA) indicated statistically significant differences between these methods ($F = 10.2$,

$P = 1.4 \cdot 10^{-6}$). We also looked at all of the aforementioned aspects in data with constant conditions and no EWS signal. Here, we found no meaningful differences between any of the methods in TNR (ANOVA $F = 2.0$, $P = 0.097$ over all experiments), which was close to the theoretical value expected to be get with a random guess, 0.9.

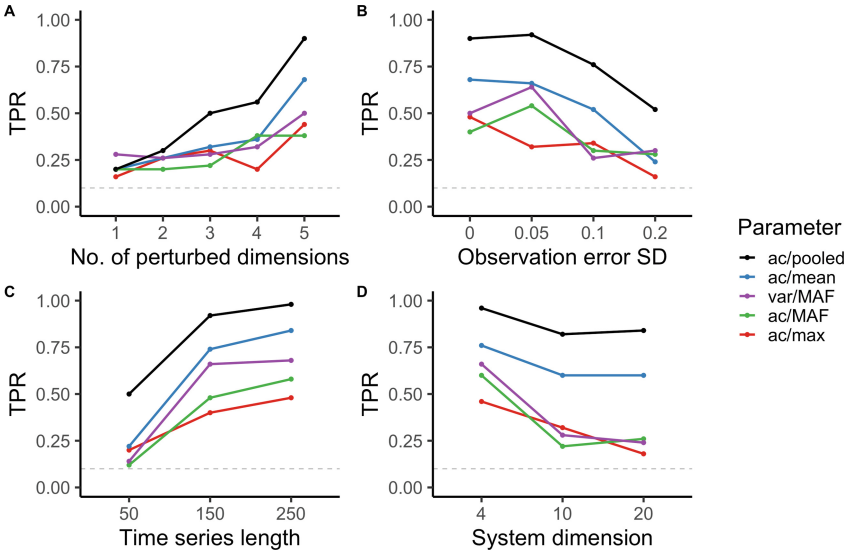

Fig. 2. EWS detection accuracy measured in true positive rate (TPR) depends on the data characteristics. **A** TPR increases as a function of perturbed dimensions. All replicates had 10 dimensions in total. **B** Increasing Gaussian additive observation error is associated with decreasing TPR. **C** Increasing time series lengths increase classification accuracy. The analyzed data sets were slightly shorter than indicated as only the part prior to the transition was used in analysis, and the exact number varied by simulation (49, 141 and 234 on average, respectively). **D** The total dimensionality of the system also influences the TPR, when half of the features are perturbed. In all panels the horizontal dashed line marks the theoretical level of a random classification, 0.1. At each x-axis value in each panel the TPR is based on 50 replicates of the simulation with randomly selected parameter and initial values.

3.2 Sensitivity Analysis

The detected EWS signal depends on chosen hyperparameters, and the values can cause spurious false positives or false negatives. Here, we investigated the effect of data detrending bandwidth and sliding window length, or Gaussian process length scale prior for the probabilistic model, on the EWS detection in representative time series.

We noticed that for ac/pooled and ac/mean the choice of these parameters did not notably influence the EWS detectability, and ac/pooled in fact produced

posterior evidence exceeding the EWS detection limit in all cases. At the lowest levels of the detrending bandwidth the P-values for both methods decreased, see Fig. 3. For the other methods the experiment showed that an EWS was correctly identified only in a small set of hyperparameter combinations, indicating remarkable sensitivity to critical modeling choices.

In time series with constant conditions and no expected EWS the results were more uniform: all methods correctly identified a true negative with practically all hyperparameter combinations (results not shown).

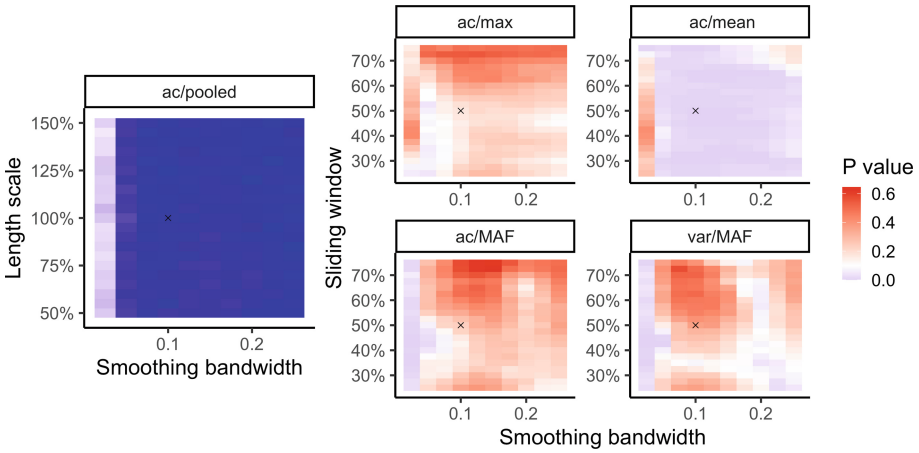

Fig. 3. Statistical evidence of the detected EWS depends on the hyperparameter combinations. The heatmap colors denote model evidence (posterior evidence and P-value for ac/pooled and all other models, respectively) for an increasing systemic autocorrelation at the hyperparameter values. The y-axis values represent proportion of the time series length. The black cross in the panels denotes the hyperparameter values we used in other experiments, Gaussian smoothing bandwidth of 0.1, and 100% and 50% of the times series length for length scale and sliding window, respectively.

4 Discussion

Early warnings have become an active area of research in the study of complex systems. Extensions of univariate indicators into the multivariate context provides opportunities to improve the sensitivity and accuracy of early warnings, as summarized in a recent review [40]. These earlier methods quantify (auto)correlation and variance in multivariate systems by optimized or averaged aggregate features and univariate projections. Related approaches have been also proposed based on neural networks [5], network analysis [26], and epidemic models [27]. These techniques typically rely on relatively large sample sizes and manual parameter adjustment, and lack explicit generative models for the data. These shortcomings form bottlenecks for practical application and interpretation. There is room for developing alternatives that are applicable to

longitudinal data sets with limited sample sizes, provide explicit quantification of uncertainties, and avoid the need for manual parameter adjustment.

Our work is motivated by biomedical and ecological applications, where the number of available time points is typically low even in the best case scenarios. Therefore we have limited our experiments to relatively low sample sizes of up to 250 time points. It is noteworthy that the typical EWS methods in the literature generally rely on several hundreds or even thousands of time points, while these sample sizes are inaccessible in many applications. Hence, our experiments also provide a useful comparison for the alternative methods in the low sample size scenarios. Furthermore, the importance of analysing uncertainties are emphasized with smaller sample sizes.

This work is an attempt to address the above-mentioned shortcomings by constructing a new class of early warning indicators for multivariate time series. We have constructed a probabilistic variant of the time-varying vector autoregressive-1 model ("ac/pooled") that can detect early warnings of critical slowing down and resilience loss by aggregating evidence through latent variables in a multivariate model. Whereas a similar pooling strategy could be considered for variance and other common EWS indicators, we have exclusively focused on autocorrelation-based methods in the current work because these outperformed variance-based indicators in our initial experiments and have shown robustness also in other recent benchmarking studies [13,40].

One of the advantages in using the probabilistic framework is that it makes the analysis of statistical certainty more straightforward. A posterior distribution for the test statistic, Kendall's τ can directly be computed from the model parameter ϕ_t. In contrast, in the non-probabilistic setting one needs to resort to indirect and time-consuming surrogate data analysis methods to generate an approximate sampling distribution for the test statistic. This can be relatively simple when a data generating model is known as in our simulation experiments. However, the data generating processes are often unknown in practice and need to be approximated in order to generate surrogate models. In univariate setting, for instance, the ARMA(p, q) model has been used to identify the optimal ARMA model parameters, in order to generate surrogates from this model [12]. In multivariate context the corresponding model is VARMA(p, q), but fitting this model is slow and potentially unreliable with higher dimension. Hence, the ability to directly estimate uncertainty in model parameters without such extra steps is beneficial. Another key advantage of the probabilistic framework is the ability to use prior distributions to regularize model fitting, and to incorporate available knowledge. This can be particularly useful when sample sizes are limited. By utilizing Gaussian process (GP) priors on ϕ_t we could restrict its posterior to differentiable functions, and by GP hyperparameter selection we could emphasize longer term trends in the target variable which are of most interest in EWS context. On the other hand the Matérn-3/2 covariance structure remains sufficiently flexible to detect relatively sudden changes as well [30,37].

Benchmarking experiments showed systematic and robust improvements of the new model compared to available alternatives. As expected, the detection

accuracy was in general better with larger perturbations. In all experiments, our proposed probabilistic multivariate indicator (ac/pooled) was systematically more sensitive than the other alternative autocorrelation based models and shorter time series were sufficient to achieve similar levels sensitivity than with alternative indicators. The highest TPR was achieved in systems with the smallest dimensionality. We speculate that such behavior could occur for instance when transients are more short-living with less species and mutualistic links that stabilize the system. All models performed equally well in terms of the true negative rate, close to their theoretical true negative rate corresponding to a random guess. No significant trends favouring any particular model in this regard were observed. Finally, our proposed method (ac/pooled) also outperformed the other methods in hyperparameter sensitivity analysis as it correctly detected the true EWS in a representative time series at all hyperparameter combinations.

The current work provides a proof-of-concept study on the potential of probabilistic multivariate early warning signals based on a single well-characterized ecological model that has been used also in other EWS studies [11,25,40]. Additional simulation models, and variations in data resolution, interaction structures, multiplicative noise, or large dimensionality, will help to assess the broader utility of the approach in practical scenarios [2,8].

Naturally, the ultimate aim is to apply the presented method on real data. While a large proportion of EWS literature has been motivated by questions in ecology, the methods are mostly agnostic to the application. In many fields, the bottleneck that has prevented wide adoption of EWS methods has been the scarcity of time series data with sufficient sample size. In biomedicine, however, developments in measurement technologies are making it increasingly feasible to collect comprehensive data sets. For instance, the human gut microbiome provides a potential target system where alternative stable states have been associated with health outcomes, such as in *C. difficile*-induced dysbiosis [4]. Other potential applications include deterioration of complex diseases [6] and predicting epidemic outbreaks [35].

The proposed method relies on a simple diagonal structure for the transition matrix with tied parameters. This aggregates information across the system and reduces the deterministic part of the dynamics into a single variable. The downside is that this will neglect interactions and does not inform us about the specific parts of the system that are affected and under risk. Future extensions could hence benefit from allowing off-diagonal terms in the transition matrix with a suitable regularization or sparsity inducing priors in order to increase model flexibility and capture important additional aspects of covariance within the system. A similar but more restricted approach would be to enhance automated feature selection by allowing the diagonal elements to vary independently, analogously to the maximal lag-1 autocorrelation in the non-probabilistic context. This could be regularized for instance with a composite GP prior, consisting of a common process, as in our model, in addition to separate GPs for the distinct elements which would be used to characterize additional, individual trajectories.

Furthermore, while quantifying uncertainties in parameter inference, our current method lacks an explicit model for observation error; adding this would allow the analysis of alternative error structures and potentially expand the scope of the method to different types of systems. Whereas a time-varying vector autoregressive-1 state space model has been studied in EWS context as a potential solution [18], this was only tested on 2-dimensional simulated data and convergence issues could arise in higher dimensions, and parameter estimation with state space models can be challenging even in the simplest cases [3]. Further extensions could consider variance, network structure, and other aspects of the system. For instance, principal component analysis has been used to detect changes in maximal variance and to identify features that are potentially most vulnerable [7,10]. Probabilistic PCA [38] could add sensitivity to these analyses by explicitly distinguishing between measurement error and random variations in the data. Incorporating other aspects of dynamics, such as estimated exit times [1] or memory properties [19], could provide further means to enhance EWS, and combining the analysis of longitudinal time series and aspects of multivariate survival analysis (see e.g. [31]) as prior information, could provide interesting avenues for future research.

The detection of early warning signals for critical transitions is a highly topical yet challenging task. Given the limitations typically encountered in applied scenarios, it is paramount that the available information can be utilized in the most optimal way, and uncertainties communicated effectively. Our proof-of-concept study presents a step towards this direction, providing an example on how probabilistic multivariate analysis can support the development of more sensitive, robust, and intuitive alternatives for the currently available early warnings signals in complex dynamical systems.

Acknowledgements. This work has been supported by Academy of Finland (decisions 295741, 330887) and by Turku university graduate school (UTUGS). The authors wish to acknowledge CSC - IT Center for Science, Finland, for computational resources. The authors declare no conflict of interest.

Code Availability. R source code for the experiments is available at 10.5281/zenodo.6472720.

References

1. Arani, B.M.S., Carpenter, S.R., Lahti, L., van Nes, E.H., Scheffer, M.: Exit time as a measure of ecological resilience. Science **372**(6547), eaay4895 (2021). https://doi.org/10.1126/science.aay4895
2. Arkilanian, A.A., Clements, C.F., Ozgul, A., Baruah, G.: Effect of time series length and resolution on abundance- and trait-based early warning signals of population declines. Ecology **101**(7), e03040 (2020). https://doi.org/10.1002/ecy.3040
3. Auger-Méthé, M., et al.: State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. Sci. Rep. **6**(1), 26677 (2016). https://doi.org/10.1038/srep26677

4. Belizário, J.E., Faintuch, J.: Microbiome and gut dysbiosis. In: Silvestre, R., Torrado, E. (eds.) Metabolic Interaction in Infection. ES, vol. 109, pp. 459–476. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74932-7_13
5. Bury, T.M., et al.: Deep learning for early warning signals of tipping points. Proc. Natl. Acad. Sci. **118**(39), e2106140118 (2021). https://doi.org/10.1073/pnas.2106140118
6. Chen, L., Liu, R., Liu, Z.P., Li, M., Aihara, K.: Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci. Rep. **2**, 342 (2012). https://doi.org/10.1038/srep00342
7. Chen, S., O'Dea, E., Drake, J., Epureanu, B.: Eigenvalues of the covariance matrix as early warning signals for critical transitions in ecological systems. Sci. Rep. **9**, 2572 (2019). https://doi.org/10.1038/s41598-019-38961-5
8. Clements, C.F., Drake, J.M., Griffiths, J.I., Ozgul, A.: Factors influencing the detectability of early warning signals of population collapse. Am. Nat. **186**(1), 50–58 (2015). https://doi.org/10.1086/681573
9. Clements, C.F., Ozgul, A.: Indicators of transitions in biological systems. Ecol. Lett. **21**(6), 905–919 (2018). https://doi.org/10.1111/ele.12948
10. Dakos, V.: Identifying best-indicator species for abrupt transitions in multispecies communities. Ecol. Ind. **94**, 494–502 (2018). https://doi.org/10.1016/j.ecolind.2017.10.024
11. Dakos, V., Bascompte, J.: Critical slowing down as early warning for the onset of collapse in mutualistic communities. Proc. Natl. Acad. Sci. **111**(49), 17546–17551 (2014). https://doi.org/10.1073/pnas.1406326111
12. Dakos, V., et al.: Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. PLoS ONE **7**(7), 1–20 (2012). https://doi.org/10.1371/journal.pone.0041010
13. Dakos, V., van Nes, E.H., D'Odorico, P., Scheffer, M.: Robustness of variance and autocorrelation as indicators of critical slowing down. Ecology **93**(2), 264–271 (2012). https://doi.org/10.1890/11-0889.1
14. Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D.: Bayesian Data Analysis, 3rd edn. Chapman and Hall/CRC (2013). https://doi.org/10.1201/b16018
15. Gelman, A., Rubin, D.B.: Inference from iterative simulation using multiple sequences. Stat. Sci. **7**(4), 457–472 (1992). https://doi.org/10.1214/ss/1177011136
16. Hastings, A., Wysham, D.B.: Regime shifts in ecological systems can occur with no warning. Ecol. Lett. **13**(4), 464–472 (2010). https://doi.org/10.1111/j.1461-0248.2010.01439.x
17. Held, H., Kleinen, T.: Detection of climate system bifurcations by degenerate fingerprinting. Geophys. Res. Lett. **312**, L23207 (2004). https://doi.org/10.1029/2004GL020972
18. Ives, A.R., Dakos, V.: Detecting dynamical changes in nonlinear time series using locally linear state-space models. Ecosphere **3**(6), 58 (2012). https://doi.org/10.1890/ES11-00347.1
19. Khalighi, M., Sommeria-Klein, G., Faust, K., Gonze, D., Lahti, L.: Quantifying the impact of ecological memory on the dynamics of interacting communities. PLOS Comput. Biol. (2022)
20. Kuss, M., Rasmussen, C.E.: Assessing approximate inference for binary Gaussian process classification. J. Mach. Learn. Res. **6**(Oct), 1679–1704 (2005)
21. Lahti, L., Salojärvi, J., Salonen, A., Scheffer, M., de Vos, W.M.: Tipping elements in the human intestinal ecosystem. Nat. Commun. **5**, 4344 (2014). https://doi.org/10.1038/ncomms5344

22. Laitinen, V., Dakos, V., Lahti, L.: Probabilistic early warning signals. Ecol. Evol. **11**(20), 14101–14114 (2021). https://doi.org/10.1002/ece3.8123
23. Lenton, T.M.: Tipping positive change. Philos. Trans. R. Soc. B: Biol. Sci. **375**(1794), 20190123 (2020). https://doi.org/10.1098/rstb.2019.0123
24. Lenton, T.M., et al.: Tipping elements in the earth's climate system. Proc. Natl. Acad. Sci. **105**(6), 1786–1793 (2008). https://doi.org/10.1073/pnas.0705414105
25. Lever, J.J., Nes, E., Scheffer, M., Bascompte, J.: The sudden collapse of pollinator communities. Ecol. Lett. **17**, 350–359 (2014). https://doi.org/10.1111/ele.12236
26. Liu, R., Chen, P., Aihara, K., Chen, L.: Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. Sci. Rep. **5**(1), 17501 (2015). https://doi.org/10.1038/srep17501
27. Proverbio, D., Kemp, F., Magni, S., Gonçalves, J.: Performance of early warning signals for disease re-emergence: a case study on Covid-19 data. PLoS Comput. Biol. **18**(3), 1–22 (2022). https://doi.org/10.1371/journal.pcbi.1009958
28. Quax, R., Apolloni, A., Sloot, P.: The diminishing role of hubs in dynamical processes on complex networks. J. R. Soc. Interface **10**, 20130568 (2013)
29. Quax, R., Kandhai, D., Sloot, P.M.A.: Information dissipation as an early-warning signal for the Lehman brothers collapse in financial time series. Sci. Rep. **3**(1), 1898 (2013). https://doi.org/10.1038/srep01898
30. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. The MIT Press (2006). https://doi.org/10.7551/mitpress/3206.001.0001
31. Salosensaari, A., et al.: Taxonomic signatures of cause-specific mortality risk in human gut microbiome. Nat. Commun. **12**, 2671 (2021). https://doi.org/10.1038/s41467-021-22962-y
32. Scheffer, M.: Critical Transitions in Nature and Society. Princeton University Press, New Jersey (2009). https://doi.org/10.1515/9781400833276
33. Scheffer, M., et al.: Early-warning signals for critical transitions. Nature **461**(7260), 53–59 (2009)
34. Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B.: Catastrophic shifts in ecosystems. Nature **413**(6856), 591–596 (2001). https://doi.org/10.1038/35098000
35. Southall, E., Brett, T.S., Tildesley, M.J., Dyson, L.: Early warning signals of infectious disease transitions: a review. J. R. Soc. Interface **18**(182), 20210555 (2021). https://doi.org/10.1098/rsif.2021.0555
36. Stan Development Team: RStan: the R interface to Stan (2020). R package version 2.21.2
37. Stein, M.L.: Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics, Springer, New York (1999). https://doi.org/10.1007/978-1-4612-1494-6
38. Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) **61**(3), 611–622 (1999). https://doi.org/10.1111/1467-9868.00196
39. Weinans, E., et al.: Finding the direction of lowest resilience in multivariate complex systems. J. R. Soc. Interface **16**, 20190629 (2019). https://doi.org/10.1098/rsif.2019.0629
40. Weinans, E., Quax, R., van Nes, E.H., van de Leemput, I.A.: Evaluating the performance of multivariate indicators of resilience loss. Sci. Rep. **11**(1), 9148 (2021). https://doi.org/10.1038/s41598-021-87839-y