Journal of the Royal Statistical Society Series C: Applied Statistics, 2026, **75**, 1003–1025
https://doi.org/10.1093/jrsssc/qlag003
Advance access publication 2 March 2026
Original Article

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series C
Applied Statistics

Statistical monitoring with novel temporal edge network processes

Anna Malinovskaya[1], Rebecca Killick[2], Kathryn Leeming[3] and Philipp Otto[4]

[1]Institute of Cartography and Geoinformatics, Leibniz University Hannover, 30167 Hannover, Germany
[2]Department of Mathematics and Statistics, Lancaster University, LA1 4YW Lancaster, UK
[3]British Geological Survey, NG12 5GG Keyworth, UK
[4]School of Mathematics and Statistics, University of Glasgow, G12 8QQ Glasgow, UK

Address for correspondence: Philipp Otto, School of Mathematics and Statistics, University of Glasgow, G12 8QQ Glasgow, UK. Email: philipp.otto@glasgow.ac.uk

Abstract

Conventional modelling of networks evolving in time focuses on capturing variations in the network structure. However, the network might be static from the origin or experience only deterministic, regulated changes in its structure, providing either a physical infrastructure or a specified connection arrangement for some other processes. Thus, to detect the change in network use, we need to focus on the processes happening on the network. In this work, we present the concept of monitoring random temporal edge network processes that take place on the edges of a graph with a fixed structure. Our framework is based on the generalized network autoregressive statistical models with time-dependent exogenous variables (GNARX models) and cumulative sum control charts. To demonstrate its effective detection of various types of changes, we conduct a simulation study and monitor cross-border physical electricity flows in Europe.

Keywords: CUSUM control charts, GNARX model, network modelling, network monitoring

1 Introduction

Network analysis is an important statistical inference tool in various disciplines, including economics, social sciences, and chemistry (De Masi & Gallegati, 2007; Khrabrov & Cybenko, 2010; McDermott et al., 2021); it provides insights into the network structure and beyond. In the case of random network models, both nodes and edges may appear and disappear over time. However, processes also originate in network structures with limited or no temporal dynamics, which are analysed with static network models. In both cases, we need to inspect the data over time to understand whether a change has occurred. This procedure is also known as *network monitoring*. Hence, to distinguish the case where the network is considered to be a random observation from the case when the network structure is deterministic but the process on it is random, we utilize the terminology *random network monitoring* and *fixed network monitoring* introduced in Stevens et al. (2021b).

Usually, the modelling and monitoring of networks are focused on detecting changes and anomalies reflected in the geometric properties of a graph, falling under the type of *random network monitoring*. Illustrative studies include monitoring e-mail communication and the daily flights within a country for detecting unusual communication patterns (cf. Malinovskaya & Otto, 2021; Perry, 2020). However, for some networks, the development of connections or inclusion of new nodes either is no longer possible or would not be of considerable monitoring interest.

Received: March 8, 2024. **Revised:** December 8, 2025. **Accepted:** January 9, 2026

In this situation, *fixed network monitoring* becomes relevant, particularly for the dynamics taking place on such networks, often captured as network attributes or covariates.

In general, the integration of nodal or edge attributes for improving the modelling of the underlying network formation mechanism and, in turn, the network monitoring is not a new concept (Azarnoush et al., 2016; Shaghaghi & Saghaei, 2020). However, in existing research, the attributes are considered to regulate the presence or absence of an edge. In other words, the contextual information available on either vertices or edges is viewed as an extra dimension to the graph, helping to explain the likelihood of the observed links (Miller et al., 2013). In contrast, we consider the process happening on edges as the primary or only source of information about the network state.

It is helpful to distinguish random processes that evolve on a fixed network structure from analyses in which the network itself, or its dynamics, constitutes the primary object of inference. The first well-established perspective is the analysis of spatial point patterns on networks (cf. Baddeley et al., 2021), where the accurate location of the object on the physical network is of interest. The second area concerns the analysis of network flows (cf. Ahuja et al., 1993; Kolaczyk & Csárdi, 2020), where the physical constraints of a network structure and the flow itself play a vital role. In this work, however, we could think of the analysed process as being a flow or traffic with no associated constraints unless they are imposed explicitly.

We start with the description of temporal edge network (TEN) processes in Section 2 and then explain our monitoring framework in Section 3. Afterwards, we perform a simulation study by testing the proposed methodology under different anomalous scenarios in Section 4. To illustrate the importance and the idea of how to monitor the real-world TEN processes, we monitor cross-border physical electricity flows in Europe in Section 5. In the last section, we discuss further research directions and summarize the perspectives and limitations of the proposed approach.

2 TEN process

Consider a network $G = (V, E)$, where the elements of V represent the vertices (or nodes) and E – edges (or links). Furthermore, we assume a fixed structure of G over time described by an adjacency matrix $\mathbf{Y} := (Y_{ij})_{i,j=1,\ldots,|V|}$, with $|V|$ being the number of nodes. When two vertices $i, j \in V$ are connected by an edge $e \in E$, they are called adjacent. Usually, the entries of $\mathbf{Y}$ are binary, i.e. $Y_{ij} = 1$ if $(i, j) \in E,\ i \neq j$, and 0 otherwise. The graph can be directed or undirected; in the latter case, $\mathbf{Y}$ is symmetric. To each existing edge $e \in E$, we relate time series $\{x_{e,t}\}$, where $t = 1, \ldots, T$, being the attributed process of G. The complete representation $\mathbf{X} = (X_{e,t})_{e=1,\ldots,|E|,t=1,\ldots,T}$, where $|E|$ denotes the number of edges, together with the graph G forms a TEN process illustrated in Figure 1 (top) for two time stamps. The following section proposes an approach to modelling and monitoring TEN processes.

3 Network monitoring framework

Usually, statistical network monitoring is subdivided into two parts: (1) **network modelling** and (2) **process monitoring** (Stevens et al., 2021a). Speaking broadly about process monitoring, it is important to distinguish between data-driven and model-based methods. Data-driven methods regard change point/fault detection as a discrimination/classification problem, requiring a substantial amount of reliable, high-quality data for their development. The broadly applied techniques include principal component analysis, partial least squares, and machine learning algorithms such as neural networks (Tidriri et al., 2016). In contrast, the model-based techniques compare the results obtained by using a model to the actual observations, analysing so called ‘deviance residuals’ (Atoui & Cohen, 2021).

Considering the network monitoring specifically, it is common to introduce a network model to later perform surveillance over network parameters or some other characteristics (see examples in Flossdorf & Jentsch, 2021; Hosseini & Noorossana, 2018; Wilson et al., 2019). Using the standard definition of model-based network monitoring, where we base the monitoring on deviance residuals, the techniques arising from the time series field, which accounts for a network structure, become the point of attention. Though still being an emerging research area, the introduction of a network autoregressive (NAR) models by Knight et al. (2016) and Zhu et al. (2017) has facilitated

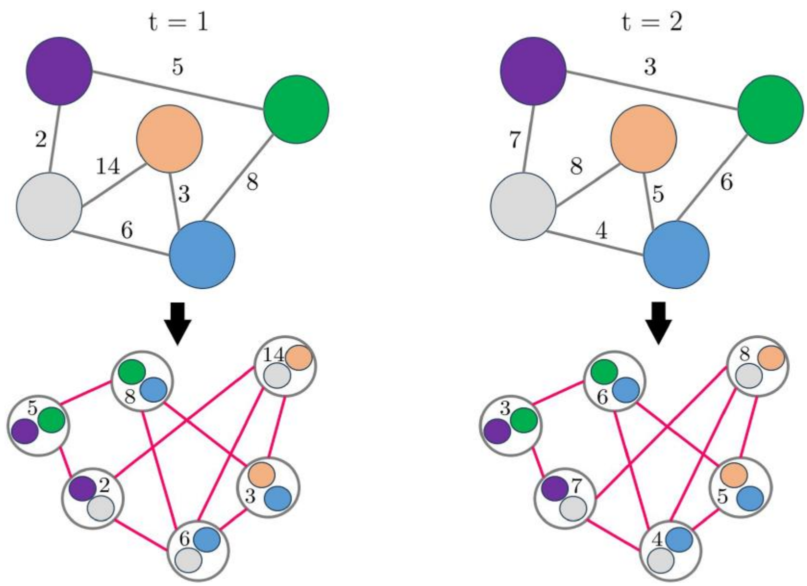

Figure 1. Visual illustration of a TEN process (top) for two time points with an adapted representation (below) described in Section 3.1.2.

several substantial contributions such as the development of generalized network autoregressive models (cf. Knight et al., 2020; Nason & Wei, 2022), presenting the foundation for our monitoring procedure described next. First, we define the model that can be suitable for TEN processes in Section 3.1.1, proceeding with the monitoring concept in Section 3.2.

3.1 Modelling of TEN processes

Before starting the discussion about network modelling, especially about the special case of modelling TEN processes, we provide some general references for readers with limited exposure to network science. To find out about real-world examples of systems that can vastly benefit from using the network analysis, particularly the temporal or dynamic network approach, we recommend looking at Blonder et al. (2012) and Holme and Saramäki (2012, 2023). If one is particularly interested in the theoretical foundation of network science, namely graph theory, Diestel (2024) is a valuable source. For gaining a broad but comprehensive overview of what network science is, especially which models exist, Newman (2018) is a suitable reference.

Regarding the variety of network models, one can distinguish between random network models and models for processes on (stochastic temporal) networks. The first group considers the network (or graph) structure as a random variable. Available models include Markov process models (cf. Snijders, 2005), (temporal) exponential random graph models (cf. Hanneke et al., 2010), and latent process-based models, e.g. (dynamic) stochastic block models (SBMs) (cf. Matias & Miele, 2017). In contrast, the latter group assumes a known network structure, which could be constant or dynamic over time, but the stochastic dynamics occur on the nodes or edges of the network, such as generalized network autoregressive (GNAR, cf. Knight et al., 2020) models, random walks on stochastic temporal networks (cf. Hoffmann et al., 2013), non-Poissonian processes on networks (cf. Hoffmann et al., 2012), or volatility models on networks (cf. Mattera & Otto, 2024). For our TEN process, we assume a static structure of a network over time. One could adapt the aforementioned random network models to this context, but they involve computationally intensive estimation that relies on numeric approximation. Hence, we consider alternative models designed for fixed network structures.

Multivariate time series modelling with an underlying network dependency structure is gaining popularity (cf. Knight et al., 2016). Leeming (2019) and Knight et al. (2020) present the GNAR framework, where a network is associated with multivariate time series and modelled as such.

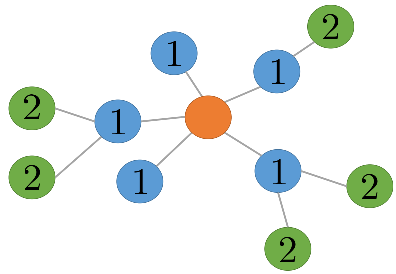

Figure 2. Stage 1 (blue) and Stage 2 (green) neighbourhood of the orange (central) node.

The GNAR models time series that occur at nodes of the network, e.g. growth rates of gross domestic product. There are several model extensions, such as the GNAR-edge model that considers time-varying edge weights (see Mantziou et al., 2023) and the GNAR model that incorporates time-dependent exogenous variables (GNARX) (see Nason & Wei, 2022), which is relevant for our monitoring framework. After the introduction of the original GNARX model in Section 3.1.1, we present its extension from nodal time series to time series on network edges in Section 3.1.2.

3.1.1 GNARX model

In this section, consider $\{x_{i,t}\}$ as a time series related to each node i. The GNARX model $(p, \mathbf{s}, \mathbf{q})$ with H exogenous regressors $\{z_{h,i,t} : h = 1, \ldots, H,\ i \in V,\ t = 1, \ldots, T\}$ and the autoregressive order p, where $(p, \mathbf{s}, \mathbf{q}) \in \mathbb{N} \times \mathbb{N}_0^p \times \mathbb{N}_0^H$ holding for all vertices $i \in V$, is specified as

$$\begin{aligned} x_{i,t} = \sum_{l=1}^{p} \left(\alpha_{i,l} x_{i,t-l} + \sum_{r=1}^{s_l} \beta_{l,r} \sum_{j \in \mathcal{N}^{(r)}(i)} \omega_{i,j} x_{j,t-l} \right) \\ + \sum_{h=1}^{H} \sum_{q=0}^{q_h} \gamma_{h,q} z_{h,i,t-q} + \epsilon_{i,t}. \end{aligned} \tag{1}$$

The order p also determines the maximum order of neighbour time lags, i.e. $\mathbf{s} = (s_1, \ldots, s_p)$ where s_l is the maximum stage of neighbour dependence for time lag l. For example, $s_1 = 2$ means that nodes depend on their first- and second-stage neighbours in G in the first time lag. Similarly, the maximum time lag of the hth exogenous regressor $\{z_{h,i,t}\}$ is defined as q_h, and collectively, these are $\mathbf{q} = (q_1, \ldots, q_H)$. In case $q_h = 0$, the current value of the exogenous variable is considered at time point t. For the empirical case study, we fix $p = 1$, $\mathbf{s} = (1)$, i.e. $s_1 = 1$, and $\mathbf{q} = (0, 0)$ for all models.

The noise is denoted by $\epsilon_{i,t}$ and is assumed to be independent and identically distributed at each vertex i with mean zero and variance σ_i^2. The parameters $\alpha_{i,l},\ \beta_{l,r},\ \gamma_{h,q} \in \mathbb{R}$ define autoregressive influence, neighbouring influence, and external influence from regressors, respectively. It is possible to estimate a global-α model, where $\alpha_{i,l} = \alpha_l$, assuming the same autoregressive process for all nodes i.

The set $\mathcal{N}^{(r)}(i)$ denotes the rth stage neighbourhood set of node $i \in G$. For instance, the Stage 1 neighbours of a node $i \in V$ are the adjacent nodes $j \in V$, connected by an edge. Furthermore, the Stage 2 neighbourhood set of a node $I \in V$ is the Stage 1 neighbours of the adjacent nodes $j \in V$ as can be seen in Figure 2. Moreover, the weights $\omega \in [0, 1]$ are associated with every pair of nodes that, in our case, depend on the size of the neighbour set as explained in Knight et al. (2020).

To obtain the estimates of the parameters $\alpha_{i,l}, \beta_{l,r}, \gamma_{h,q}$, we fit the model using the least squares method with the design matrix that contains a column for each parameter we attempt to estimate. These estimates can be used for computing the $\hat{x}_{i,t}$ value from $\{x_{i,t-1}\}$. The deviance residuals are then determined as

$$u_{i,t} = x_{i,t} - \hat{x}_{i,t}, \tag{2}$$

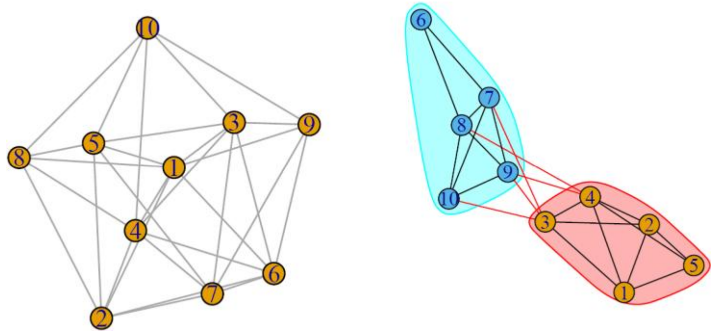

Figure 3. Representation of two different connectivity types: A graph structure generated by an Erdős–Rényi model (left) vs. by an SBM (right).

which can be utilized in our monitoring framework. However, the GNARX model is defined for time series related to nodes. Thus, we need to extend the representation of TEN processes to enable the correct application of the GNARX model when monitoring.

3.1.2 Extension to TEN processes

There are different ways of thinking about why we require an alternative representation. First, our main focus is on discovering changes between and within the streams, i.e. a process captured on the edges. Thus, we indirectly consider it to be a network that is not of single nodes anymore but of pairs of nodes. Second, the influence from node to node and from one stream to its neighbouring streams portrays two distinct standpoints. In the case of the dependency between two traffics, more than two nodes, as well as other underlying mechanisms that determine the exchange and its strength, are involved. Thus, we need a different adjacency matrix to describe this novel dependency structure.

To distinguishing between the structures G and $\mathbf{Y}$, and the novel representation of the TEN process as G', we introduce $\mathbf{Y}'$ as the new adjacency matrix. Now, as displayed in Figure 1 (below), the edges $e \in E$ from the original graph G become vertices $\iota \in V'$ with the number of nodes being $|V'| = |E|$ so that the time series $\{x_{e,t}\}$ placed on edges $e \in E$ are now placed on the nodes, becoming $\{x_{\iota,t}\}$. The change to $\{x_{\iota,t}\}$ indicates that we start regarding Equation 1 as one determined for edges in a form of nodes rather than the original nodes. Hence, $\omega_{i,j}$ and $\{x_{\iota,t}\}$ have no relation to each other apart from being defined in exactly the same way for different types of nodes. Consequently, we would also have new edges $\xi \in E'$.

Considering the connectivity structure captured by the adjacency matrix, without any expert knowledge, the natural choice for creating $\mathbf{Y}'$ would be to use a model for generating random graphs. A potential candidate is the Erdős–Rényi model (cf. Erdős & Rényi, 1959), where all graphs on a fixed vertex set with a fixed number of edges are equally likely. Another possibility is to consider sampling from a SBM, which produces graphs with a community structure (cf. Holland et al., 1983). For example, links within a community may be more probable than between communities. Both structure types are showcased in Figure 3.

If a TEN process is well understood, the application of the SBM, where the flows should be first subdivided into clusters, is a good choice. Alternatively, it is possible to run a clustering algorithm. If the cluster structure is not suitable or no specific knowledge about the TEN process is available, the structure of G' can be sampled by applying the Erdős–Rényi model. Both of those options are tested in Section 4.

However, there is an additional approach to choosing a new connectivity structure by referring to other types of graphs from graph theory. We use a deterministic approach to construct $\mathbf{Y}'$ in this case. One relevant representation technique is the line (also known as edge-to-vertex dual) graph $L(G)$ of G. Here, $L(G)$ is constructed on E, where $e \in E$ are adjacent as nodes $\iota \in V'$ if and only if they share a common node as edges in G (Diestel, 2017). In other words, each edge $\iota = (i, j)$ in the original graph G is treated as a node in the line graph $\mathbf{Y}'$, and two edges ι and $\upsilon = (k, l)$ are adjacent in $\mathbf{Y}'$ if they share a common node, i.e. $\{i, j\} \cap \{k, l\} \neq \emptyset$. The weights $\omega'_{\iota,\upsilon}$ are defined accordingly, for instance, as uniform row-normalized values $\omega'_{\iota,\upsilon} = 1/|\mathcal{N}(\iota)|$ if $\upsilon \in \mathcal{N}(\iota)$, or as normalized functions of the original edge weights to preserve network structure. This connectivity structure is adapted to the original TEN process shown in Figure 1 (bottom plot) and is also applied in the empirical study in Section 5. Such formation can suit real-world applications well as it offers a more logical structure in case of limited knowledge about existing communities or insufficiency of a connectivity structure generated from a random graph model.

After redefining the representation of TEN processes, we can proceed with their modelling by applying the GNARX model and monitoring them using a residual-based control chart introduced in the next section.

3.2 Monitoring of TEN processes

Recall that, for network monitoring, we need a model and a monitoring procedure. For monitoring, it is vital to determine the framework in which we are aiming to perform it. One possibility is to monitor the estimates of the model parameters, although this is computationally intensive (cf. Malinovskaya & Otto, 2021). A more efficient alternative is to perform monitoring based on deviance residuals. For instance, Miller et al. (2013) compute graph residuals as the difference between the observed graph and its expected value. In modelling the TEN with the GNARX model [see Equation (2)], it is uncomplicated to calculate the residuals as the deviations of the current observation from the GNARX prediction. We expect the deviance residuals to increase as soon as a change occurs, indicating a lack of fit. According to Alexopoulos et al. (2004), who offers a detailed introduction to statistical process monitoring (SPM) and especially to the forecast-based monitoring methods, the prediction errors are uncorrelated when the model or prediction is accurate. Thus, we can apply traditional SPM techniques such as control charts (cf. Montgomery, 2012) to these deviance residuals. A technique that is particularly suitable for testing whether the current residual-based statistic is anomalous is a control chart. Psarakis et al. (2014) and Ali et al. (2016) provide a comprehensive overview of this methodology.

Usually, control charts are both straightforward to implement and interpret. Grundy et al. (2026) propose a framework for monitoring forecast models, showing that changes in complex data are reflected in the deviance residuals. They utilize the Cumulative Sum (CUSUM) control chart, and we extend their approach to monitoring deviance residuals of the GNARX model for TENs, as shown subsequently. A comprehensive overview of residual control charts and their comparison to other types of control charts is provided in Knoth and Schmid (2004) and Jensen et al. (2006).

Applying control charts in an online manner, i.e. performing a sequential change point detection at each time point t, we test whether the process remains in control. Hence, the specific hypothesis, as well as the definition of a change point, are clarified in Section 3.2.1 before coming to the description of the implemented control chart in Section 3.2.2.

3.2.1 Hypothesis formulation

As we base our monitoring part on the deviance residuals $u_{\iota,t}$ obtained by finding the difference between the actual process and the predictor at each flow [Equation (2)], obtaining an accurate model before monitoring starts is vital. Our objective is then to test

$$H_{0,t}: \text{The observed TEN process coincides with the fitted GNARX model}$$

against the alternative

$$H_{1,t}: \text{The observed TEN process does not coincide with the fitted GNARX model}$$

for each t. Now, the question arises of which variable to use for the test statistic that could provide information about when H_0 is violated. The change point τ is defined as

$$\mathbf{x}_t \sim \begin{cases} F(\boldsymbol{\mu}, \Sigma), & \text{if } t < \tau \\ F_\tau(\boldsymbol{\mu}_\tau, \Sigma_\tau), & \text{if } t \geq \tau \end{cases}$$

where $\mathbf{x}_t = (x_{1,t}, \ldots, x_{|V'|,t})'$, and $\boldsymbol{\mu}$ and Σ define the mean vector and variance–covariance matrix of the network flow distribution F, respectively. However, the change in the mean and/or variance of the raw data would lead to the respective changes in the deviance residuals (Grundy et al., 2026). Hence, by constructing a test statistic based on $\mathbf{u}_t = (u_{1,t}, \ldots, u_{|V'|,t})'$, we can accordingly determine the time point τ when the change has occurred. Following, we introduce the change point detection framework specified by Grundy et al. (2026) and discuss the respective test statistic for applying residual-based CUSUM control charts.

3.2.2 *Residual-based CUSUM control chart*

Before describing the technicalities of the considered control chart, it is worth beginning with an explanation of its key elements and framework for application in practice. Generally, a control chart, a graphical tool for detecting unusual variations of the process, consists of three components: (1) a central line (CL) for plotting the average of the process, (2) an upper control line for the upper control limit, and (3) a lower control line for the lower control limit. When the process considerably deviates from its expected state, the control statistic starts crossing the limits, indicating a substantial change beyond random variation.

Usually, there are two phases involved in the implementation of control charts. Phase I corresponds to the exploration and calibration period, where we estimate the target process and calibrate the control chart, e.g. computing control limits. It is based on the assumption that the process during this phase is in control, i.e. stable, predictable, and repeatable (Vining, 2009). In Phase II, the actual application of the control chart begins, which could be viewed as a sequential implementation of the hypothesis test introduced in Section 3.2.1. The more precise the estimation of the parameters in Phase I, the more reliable the performance of the control chart in Phase II. The desired operation of control charts during Phase II is a quick detection of the change point in a process when it experiences an out-of-control state, i.e. unusual process variation is present. At the same time, we strive for a long-running scheme without false alarms, i.e. the process remains in control but the scheme signals a change. This translates to a balance between small control limits for fast detection and large control limits for small false alarm rates.

In real-world applications, it is usually unknown what properties, e.g. mean or variance, of a process may change. Thus, an important criterion for selecting a suitable control chart is its ability to simultaneously track many types of changes. Another point is the decision about what exactly to monitor—the data itself or process related to it. As discussed by Grundy et al. (2026), the monitoring of the data directly might deteriorate in the case of temporal dependency, complex trends, or seasonality effects within its structure. However, monitoring deviance residuals of the process omits these issues and can reflect changes in either the process mean or the process variance.

The proposed control charting procedure employs parallel univariate control charts to detect *local changes* in individual flows. In the second step, these detected local changes are aggregated to identify significant *global changes*. As under the real settings, it is usually unknown whether the change occurs in a mean and/or variance of the process. Thus, if the task is to perform joint monitoring, there are two possible directions: (1) design monitoring using two change point models or (2) apply a single detector that is capable of detecting changes in both location and scale simultaneously. Application of control charts that belong to the latter group yields multiple benefits: first, the Type I error can be controlled more effectively. Second, in case central assumptions about the data distribution are violated, there are possibilities to use intuitive nonparametric approaches such as data-depth based control charts (e.g. R. Y. Liu, 1995; Y. Liu et al., 2019) or other ranking-based techniques (cf. Ross et al., 2011). Generally speaking, a joint control chart consolidates the process, simplifying analysis and interpretation. For monitoring TEN processes, we utilize a more general type of Page's CUSUM detector (cf. Page, 1954), which is based on the (centred) squared data and can detect a combination of the mean and/or variance change in the original data. For monitoring deviance residuals $u_{\iota,t}$ obtained for each flow ι, Grundy et al. (2026) adapt Page's CUSUM test statistic (cf. Page, 1954) to the centred squared deviance residuals as follows:

$$Q_\iota(m, k) = \sum_{t=m+1}^{m+k} (u_{\iota,t} - \hat{b}_\iota)^2 - \frac{k}{m} \sum_{t=1}^{m} (u_{\iota,t} - \hat{b}_\iota)^2, \tag{3}$$

where m corresponds to the length of Phase I, k is the current time point in Phase II, and $\hat{b}_\iota$ is the mean estimate of the deviance residuals computed from Phase I. From that, we compute the control chart statistic for flow ι as

$$D_\iota(m, k) = \max_{0 \leq a \leq k} |Q_\iota(m, k) - Q_\iota(m, a)|, \tag{4}$$

and the corresponding upper control limit is given by

$$UCL_\iota = \hat{\sigma}_\iota \zeta_\alpha g(m, k, \nu), \tag{5}$$

where ζ_α is the critical value of the distribution with the significance level α and $\hat{\sigma}_\iota$ is the estimate of the standard deviation of the centred squared deviance residuals belonging to the flow ι. The estimate $\hat{\sigma}_\iota$ may also be obtained from the long-run variance, which accounts for remaining temporal autocorrelation in the residual sequence (including the possibility of negative or alternating correlations arising from general over-adjustment of temporal structure), as applied in Section 5. Methodologically, one may either standardize the statistic $D_\iota(m, k)$ directly by dividing by $\hat{\sigma}_\iota$ or equivalently incorporate $\hat{\sigma}_\iota$ into the upper control limit UCL_ι.

The part $g(m, k, \nu)$ defines the weight function with the tuning parameter ν. Both are described in detail in Horváth et al. (2004) and Grundy et al. (2026). In this work, we consider $\nu = 0$. The derivation of the applied UCL_ι can be found in the online supplementary material. We consider the current time point to be $k = \tau$, i.e. the control chart detects a *local change* point if $D_\iota(m, k) \geq UCL_\iota$. Notably, no lower control limit is defined as the control statistic obtains non-negative values only and, therefore, has its natural lower control limit at zero.

To detect *globally relevant changes*, the changes detected in individual nodes or edges can be aggregated in various ways to provide a comprehensive assessment of the system. Firstly, a straightforward approach is to identify global changes based on a percentage of nodes or edges exhibiting local changes by controlling family-wise error rates (FWERs), as the deviance residuals are assumed to be independent and identically distributed under the null hypothesis. In other words, if the process coincides with the target process, all dependencies are captured in the fitted GNARX model, leaving the deviance residuals independent with the same in-control mean and variance. Secondly, domain experts can determine a relevant threshold percentage of nodes or edges, leveraging their knowledge of the system's context and structure, e.g. the system could only be critically affected if a certain percentage share of flows signalled a change. This is different to the first case, in which a significant change in at least one of the flows triggers a global alarm. Thirdly, local signals can be combined in a weighted fashion to account for the importance of specific edge flows. For instance, central or essential nodes could be assigned higher weights due to their pivotal roles in the network, or higher weights could be assigned to signals that appear in close proximity (i.e. clustering effects). Conversely, the weights could also be determined based on the entropy or dispersion of local signals, emphasizing nodes or edges that deviate from expected patterns. The choice of the aggregation method ultimately depends on the specific application. For our energy-flow monitoring procedure in Section 5, we employ the second type and consider the percentage of flows with signals, denoted as *change intensity*, to aggregate the univariate control charts.

3.2.3 Monitoring scope and performance function

The considered CUSUM chart belongs to the univariate control charts, meaning that the scheme monitors only one process variable. In our case, we simultaneously implement $n = |V'|$ control charts to monitor the TEN process completely. Consequently, it is also possible to perform local monitoring and control the behaviour of only one essential flow from a complete TEN process.

In this work, however, we are interested in proposing a monitoring procedure suitable for tracking the complete TEN process. While the univariate charts may signal at different time points, we additionally want to trigger a *global alarm* when a relevant percentage of flows have experienced a change. Thus, as a performance metric, we introduce a cumulative intensity function of a change that presents a cumulative sum of flows that triggered a change at a time point t,

$$I_{\mathbf{X}}(T) = \sum_{k=1}^{T} \frac{\sum_{\iota=1}^{n} \mathbf{1}_{[UCL_{\iota},\infty)}[D_{\iota}(m, k)]}{n}. \tag{6}$$

If in one flow ι, a change was detected, the monitoring of this flow stops, resulting in $\max(I) = 1$. Thinking about how a signal is produced by computing $I_{\mathbf{X}}(T)$, we need to define a suitable threshold W. The definition of a relevant share to trigger a global alarm requires some expert knowledge of the specific process. As described in Section 3.2.2, instead of the equally weighted percentage of univariate signals, a relevant global change could alternatively involve considering central/essential nodes or clustered signals via weighted averages of (univariate) signals.

In the subsequent section, we illustrate a simulation study that handles different cases of changes and compares monitoring results with respect to the selected model for constructing the adjacency matrix $\mathbf{Y}'$.

4 Simulation study

The reason to conduct a simulation study is twofold: first, we aim to quantify the detection speed of changes in the effects captured by the model parameters $\alpha_{i,l}, \beta_{l,r}$ or $\gamma_{h,q}$. Second, as mentioned in Section 2, there are different possibilities to construct the adjacency matrix $\mathbf{Y}'$ as introduced in Section 3.1.2. Here, we examine two distinct random graph models that cover both the availability and absence of expert knowledge about suitable sampling models, namely the Erdős–Rényi model and the SBM, both described in Section 3.1.2.

In the next part, we explain the design of the simulation study, present the results, and compare them to the generated graph structure.

4.1 Experimental setting

To perform a simulation study, we have to decide on the design of a TEN process, the initial parameters, and the test cases. To keep the study simple for reproducing it but comprehensive for designing different anomaly types and testing two distinct $\mathbf{Y}'$ structures, we create a TEN process with ten flows and medium connectivity. That means, for sampling $\mathbf{Y}'$ from the Erdős–Rényi model, we set the number of nodes to be ten and the number of connections to be 30, considering that everything else is random. For sampling from an SBM, we separate the flows into two clusters, c_1 and c_2, of equal size. Additionally, we need to define the probabilities that the flows are connected within a cluster or between the clusters.

Regarding the settings for sampling from a GNARX model, we design a process with a global autoregressive effect of order $p = 1$. Also, we consider the neighbourhood effect of order $\mathbf{s} = (1)$ together with two exogenous regressors with time lags $\mathbf{q} = (0, 0)$. To be precise, we generate separate data for estimating the GNARX model (600 observations) before designing Phase I with 200 observations and Phase II with 100 observations, subdivided into 50 in-control and 50 out-of-control samples. The Phase I length was fixed at $m = 200$ to balance two competing requirements: On the one hand, sufficiently many observations are needed for stable parameter estimation; on the other hand, longer windows make the assumption of a homogeneous, change-free Phase I segment more restrictive in practice. The detailed description of each simulation step is presented in Algorithm 1. In total, we conduct 500 iterations for each test case, presenting the outcomes subsequently.

Algorithm 1 Simulation steps for one iteration for the test case $\beta + 0.3$

1. **Setting:** Choose the lengths of the burn-in period (300), model fitting period (600), Phase I (200) and Phase II (100), subdivided into in-control (50) and out-of-control (50) parts. Set the data-generating parameters of the GNARX model: $\alpha_{\iota,l} = \alpha = 0.2$, $\beta_{l,r} = \beta = 0.3$ with $p = 1$ and $\mathbf{s} = (\mathbf{1})$, $\gamma_{h,q} = \gamma_1 = 2$ and $\gamma_2 = 3$ with $\mathbf{q} = (\mathbf{0}, \mathbf{0})$. Select the change: $\beta_\tau = \beta + 0.3$.
2. **TEN Generation:**
(a) Generate the adjacency matrix $\mathbf{Y}' := (Y'_{\iota v})_{\iota, v=1,\ldots,|V'|}$ by sampling from either the Erdős–Rényi model with $|V'| = 10$ and $|E'| = 30$ or the SBM with $P_{c_1,c_2} = P_{c_2,c_1} = 0.2$ and $P_{c_1,c_1} = P_{c_2,c_2} = 0.8$.
(b) For each $t = 1, \ldots, 1200$ generate the GNARX process:
   (i) Sample error values $\epsilon_{\iota,t}$, where $\epsilon_{\iota,t} \sim \mathcal{N}(0, 1)$.
   (ii) Sample covariate values $z_{1,\iota,t} \sim \mathcal{N}(0, 1)$ and $z_{2,\iota,t} \sim \mathcal{N}(0, 1)$.
   (iii) For $\iota = 1$ to $\iota = |V'|$.
      1. If $t \leq 1150$, generate the data with α, β, γ_1 and γ_2 parameters,

$$x_{\iota,t} = \alpha x_{\iota,t-1} + \beta \sum_{v \in N^{(1)}(\iota)} \omega_{\iota,v} x_{v,t-1} + \gamma_1 z_{1,\iota,t} + \gamma_2 z_{2,\iota,t} + \epsilon_{\iota,t}.$$

      2. If $t \geq 1151$, generate the data with $\alpha, \beta_\tau, \gamma_1$ and γ_2 parameters having an effect either on all flows or only on flows from c_1, adjusting the equation in (1).
3. **TEN modelling:**
(a) Fit the GNARX model with $p = r = 1$ using the first 600 generated observations after discarding the burn-in period $t = 1, \ldots, 300$.
(b) Save the estimated parameters $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}_1$, and $\tilde{\gamma}_2$.

4. **TEN Monitoring:**
(a) Phase I: Estimate the target process by using $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}_1$, and $\tilde{\gamma}_2$ parameters to compute deviance residuals $u_{\iota,t}$ and calibrate the CUSUM control charts described in Section 3.2.2.
(b) Phase II: Run monitoring using the calibrated CUSUM control charts and compute the cumulative intensity function provided in Equation~\eqref{eq:intensity}. Save the time stamp t when a change is detected.

4.2 Results and comparison of connectivity structures

As explained in Section 3.2, we aim to detect changes in the whole TEN process, i.e. the anomalous network states. Therefore, we average the cumulative sum of alarms, i.e. compute the cumulative change intensity given in Equation (6) for each iteration and then determine its mean values for each time stamp in Phase II. With that, we summarize the performance of the control chart in the simulation study. It is worth noting that we do not define a threshold W here, focusing on the general detection capability of the approach as well as fluctuation in the performance. It is important to note that as soon as the control chart signals for a particular flow, the monitoring of that flow stops. This means that no multiple change points for the same flow are possible.

For each parameter $\tilde{\alpha}, \tilde{\beta}$, and $\tilde{\gamma}_1$, we have three test cases of changes whose values gradually increase. Comparing the different types of sampling the adjacency matrix $\mathbf{Y}'$, we exchange the initially estimated parameters with anomalous ones either in all flows or only in flows that are part of c_1.

In Figures 4, 5, and 6, we display the monitoring results of Phase II (PII). The blue area defines the in-control and the red the out-of-control part. First, in all plots, we notice the same pattern: The curve that shows the simulation with $\mathbf{Y}'$ sampled from the SBM stays below another curve that corresponds to the setting with an adjacency matrix generated from the Erdős–Rényi model. That corresponds to the way the change in the effects was implemented: while in the case of the SBM structure, only one community (half of the flows) was affected by anomalous parameter sizes, the simulation that involves the Erdős–Rényi structure experiences the change in the complete network, i.e. all flows were affected. Thus, the cumulative intensity change reaches a value of only 0.5, which is the normalized size of a community with five flows, which presents anomalous behaviour

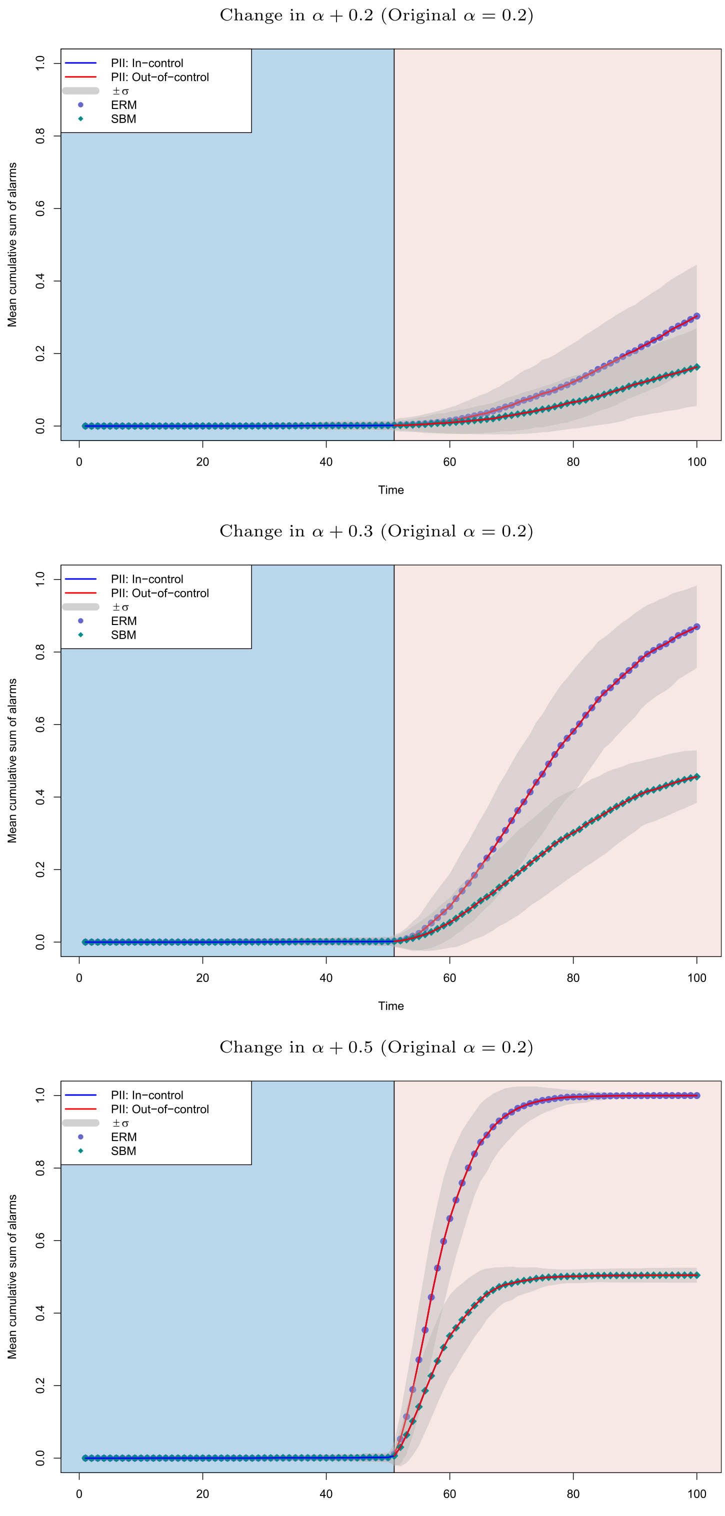

Figure 4. Simulation study: Visualisation of the simulation results over 500 iterations with differences in α. ERM corresponds to the Erdős–Rényi model. The variation between simulation iterations is quantified by plotting the standard deviation ($\pm\sigma$) of the average run length (ARL) values across runs.

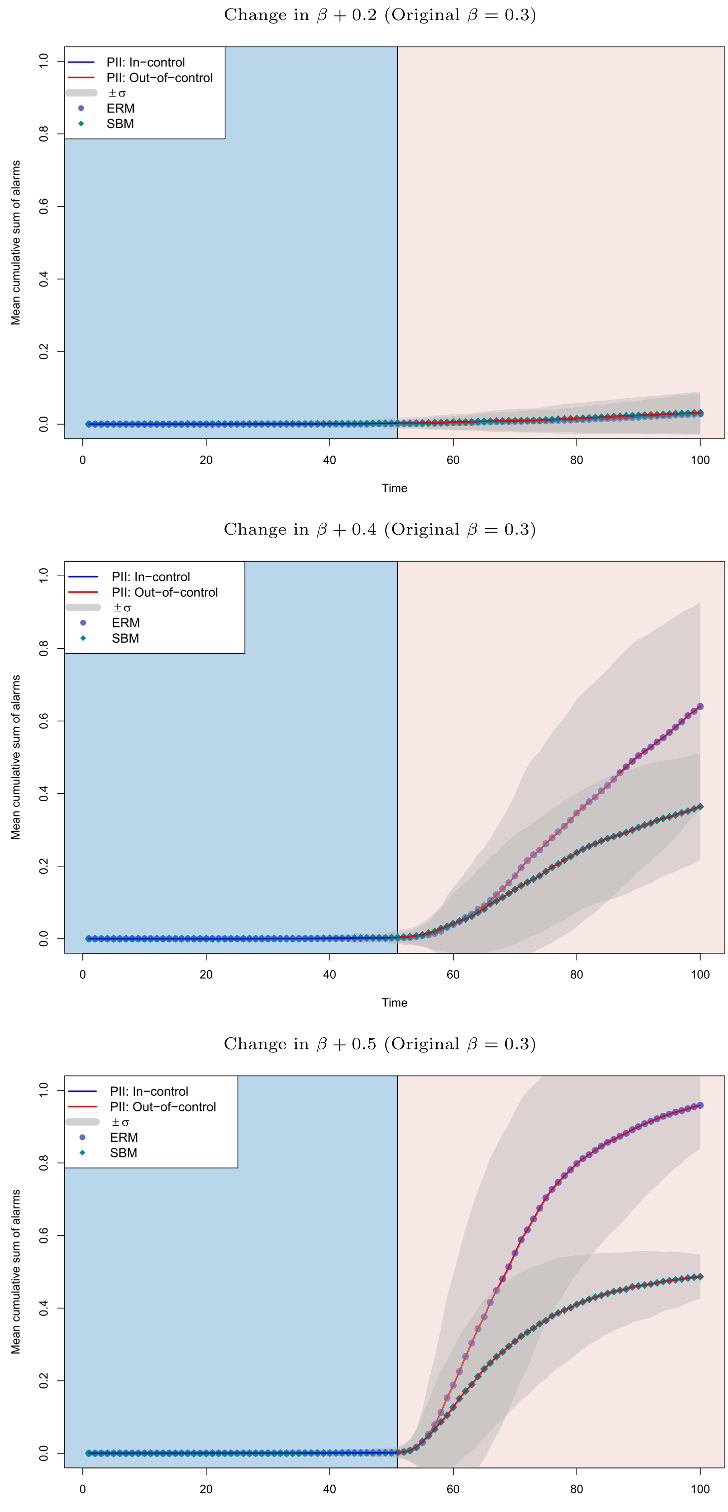

Figure 5. Simulation study: Visualisation of the simulation results over 500 iterations with differences in β. ERM corresponds to the Erdős–Rényi model. The variation between simulation iterations is quantified by plotting the standard deviation ($\pm\sigma$) of the average run length (ARL) values across runs.

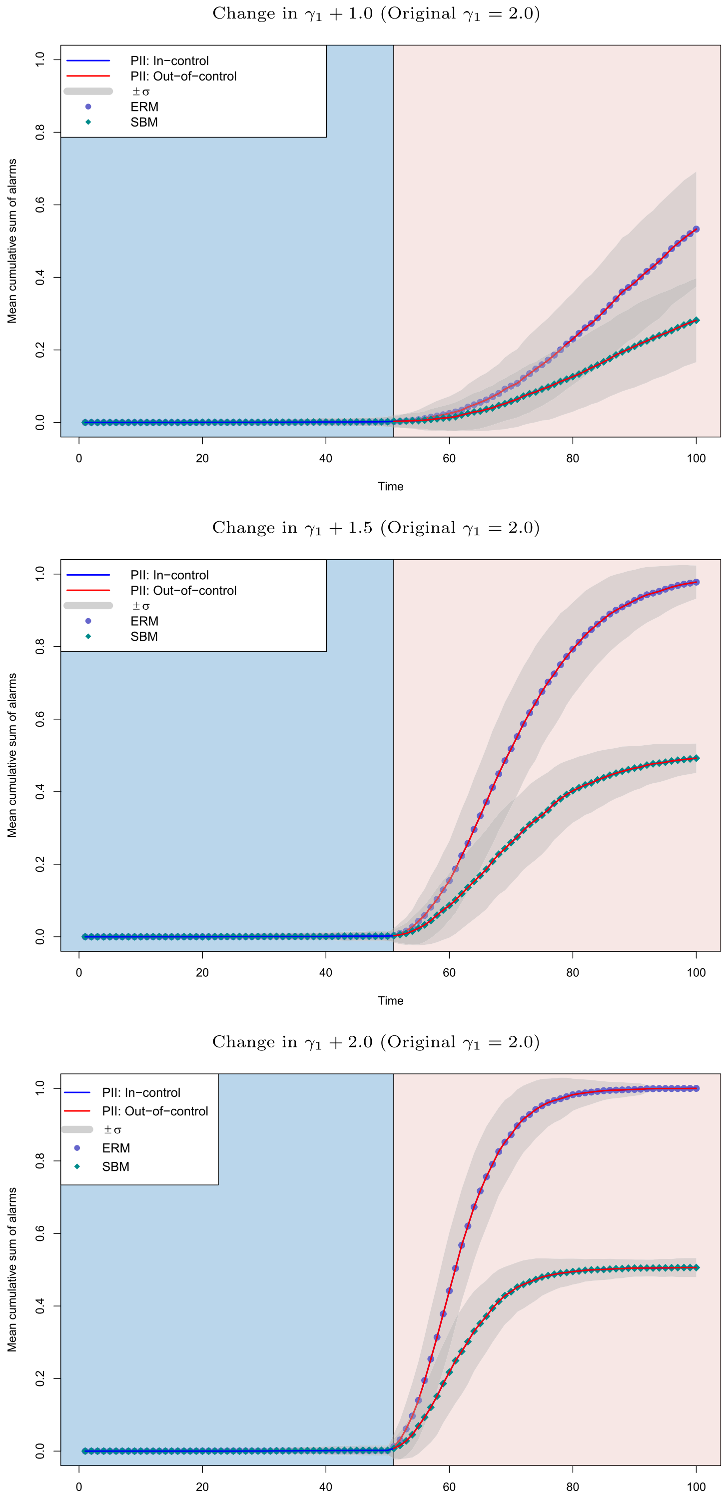

Figure 6. Simulation study: Visualisation of the simulation results over 500 iterations with differences in γ_1. ERM corresponds to the Erdős–Rényi model. The variation between simulation iterations is quantified by plotting the standard deviation ($\pm\sigma$) of the average run length (ARL) values across runs.

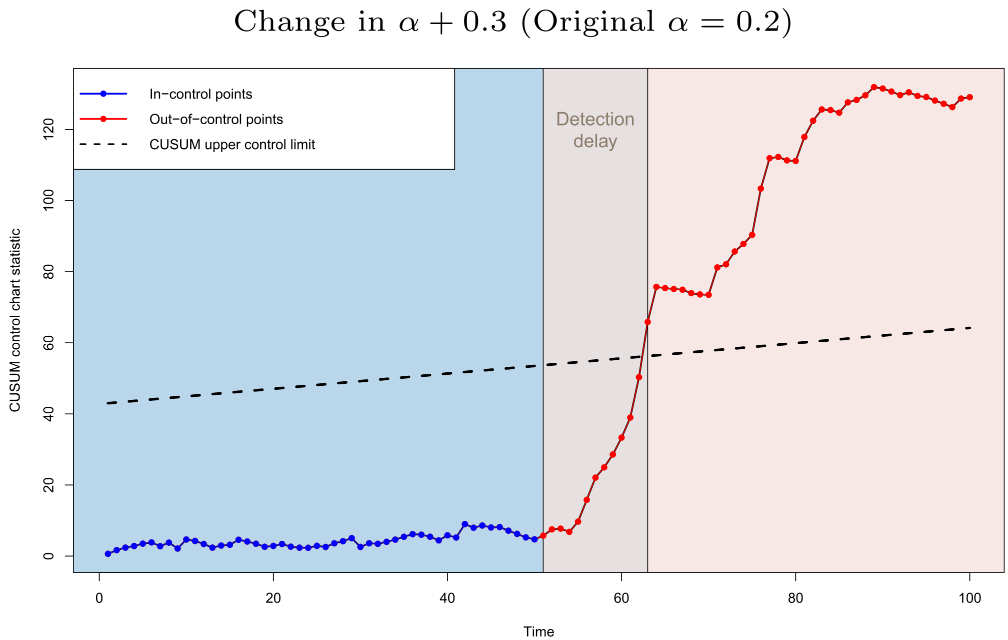

Figure 7. Simulation study: example of monitoring one specific flow from a TEN process.

when using the SBM. Second, we recognize that the most difficult type of anomaly for the proposed approach is the change in the neighbourhood effects captured by the parameter β. This can be seen from the wider uncertainty span and longer run length to detect the insertion of a change in β. Exact values of average run lengths (ARLs) for different threshold values W can be found in the online supplementary material.

Overall, the simulation study reveals that the monitoring approach is highly effective, meaning that no signals occur when no actual change has been introduced, but is only efficient in detecting the time stamp quickly after the change if we are aware of the underlying structure of G' or aim to detect medium or large process innovations.

If one specific flow reflects the state of the whole system or is particularly relevant to the flawless functionality of a network, we could focus solely on its monitoring. For that, we would apply the CUSUM control chart and directly determine the change point. Visually, a possible outcome could look as displayed in Figure 7.

In the following section, as a real-world illustration, we present the monitoring of cross-border physical electricity flows across Europe.

5 Monitoring of cross-border physical electricity flows

Cross-country analysis of electricity trade, especially in the context of renewable energy, is a relevant field of study that contributes to evaluating transmission infrastructure policies (cf. Abrell & Rausch, 2016). It is worth noting that the cross-border physical flow network indicates the actual flow of electricity (corresponding to the laws of physics) and could differ from the scheduled commercial exchanges, which reflect the economic relations between the market parties and, therefore, be of particular technical importance, allowing for the flawless electricity trade across Europe. To the best of our knowledge, no studies are related to the change point detection in European cross-border physical flows (CBPFs). We illustrate the monitoring of these processes below.

5.1 Data description

The European Network of Transmission System Operators for Electricity (ENTSO-E) provides data about the cross-border physical flow and scheduled commercial electricity exchanges of

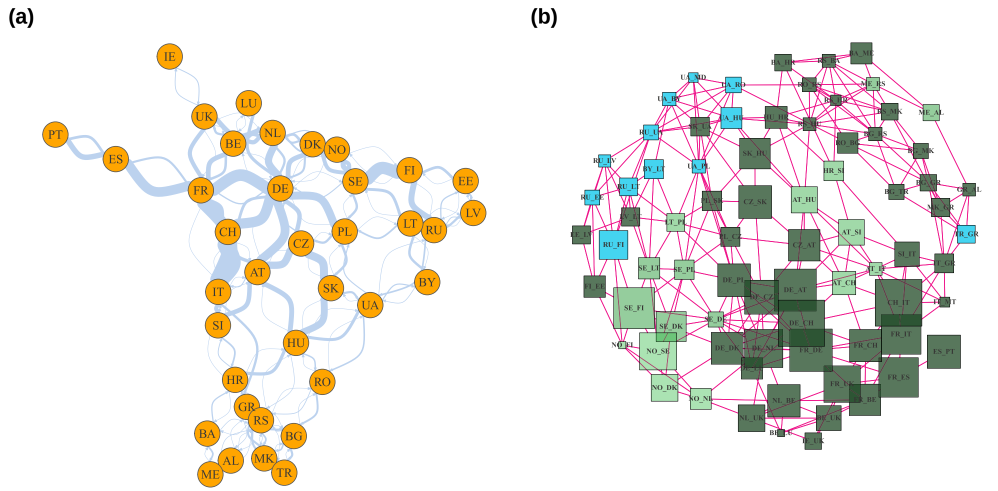

Figure 8. Conventional representation of the cumulated cross-border physical electricity flow in the year 2019 (a) and new representation (b) following the description in Section 3.1.2. (a) The edge thickness implicates the strength of the flow. (b) The size of the nodes indicates the strength of the electricity exchange and the colour its proportion of electricity generated with renewable energy sources (light green corresponds to the higher proportion, and blue corresponds to the missing information).

more than 40 countries.[1] It offers hourly data for all years from 2014. However, after careful investigation of the data quality, we decided to concentrate on the period from 1 January 2018 to 27 November 2022, and use a weekly aggregation of the electricity values given in megawatts. Considering the data quality, on the weekly level, there were no missing data for any flows.

The graph in Figure 8a represents the aggregated quantity of electricity transmitted in the year 2019 so that the wideness of edges reflects the total amount of power passed from one to another country. The higher this value is, the wider the edge becomes. For most of the country pairs, we can notice the existence of parallel edges, i.e. one flow f_1 coming from France (FR) to Spain (ES) and another flow f_2 back. If we directly apply a new representation as introduced in Section 3.1.2, we would obtain a network that would be considerably bigger than the original graph. Hence, we need to aggregate both flows f_1 and f_2 and take one vertex to represent a country pair. There are different possibilities for doing so. For that, we create three metrics and discuss them subsequently.

5.2 Phase I modelling

To avoid a considerable expansion in the new representation of the CBPFs because of parallel flows, we consider three aggregation strategies for directional electricity flows f_1 and f_2 (in units of megawatts per week) and fit separate GNARX models for each aggregated representation.

The first statistic, $\mathscr{M}_1 = \log(f_1 + f_2 + 1)$, is a Box–Cox transformation with parameters $\lambda_1 = 0$ and $\lambda_2 = 1$ (Box & Cox, 1964), applied to the total bidirectional flow. It reflects the overall intensity of electricity exchange between two countries, irrespective of direction. This transformation stabilizes variance and reduces the influence of large flows while preserving monotonicity.

The second statistic, $\mathscr{M}_2 = \log(f_1 + 1) - \log(f_2 + 1)$, measures the directional imbalance in log-scale and indicates the prevailing direction of net transfer. Positive values imply dominance of exports from the country associated with f_1, while negative values indicate dominance of f_2. This metric is well-suited to detect persistent asymmetries in the flow direction, which may reflect structural dependencies or policy-driven exchange patterns.

[1] Central collection and publication of electricity generation, transportation and consumption data and information for the pan-European market. ENTSO-E Transparency platform (2023). https://transparency.entsoe.eu.

Table 1. Phase I modelling for different types of aggregation statistics.

	Model $\mathscr{M}_1$	Model $\mathscr{M}_2$	Model $\mathscr{M}_3$
$\tilde{\alpha}$	0.896***	0.904***	0.910***
	(0.005)	(0.005)	(0.005)
$\tilde{\beta}$	0.094***	0.019$^{\cdot}$	0.020$^{\cdot}$
	(0.006)	(0.011)	(0.011)
$\tilde{\gamma}_1$	0.003$^{\cdot}$	−0.034***	−0.003***
	(0.002)	(0.006)	(0.001)
$\tilde{\gamma}_2$	0.005$^{\cdot}$	0.032***	0.003***
	(0.002)	(0.006)	(0.001)
RMSE (dimensionless)	0.89	2.75	0.34
$\bar{\mathscr{M}}$	10.84	0.14	0.03
$\tilde{\sigma}_{\mathscr{M}}$	2.00	6.81	0.84

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p \le 0.1$

The third statistic, $\mathscr{M}_3 = \frac{f_1 - f_2}{f_1 + f_2}$, captures relative imbalance in a scale-free, normalized form bounded in $[-1, 1]$. It represents a symmetric flow skewness index that is insensitive to absolute magnitude. This makes it useful for identifying shifts in directional preference in edges with relatively low overall flow.

Each of the three metrics captures different structural aspects: $\mathscr{M}_1$ highlights the variations in total volume (amplitude-driven changes), $\mathscr{M}_2$ focuses on net directional dominance (signed imbalance), and $\mathscr{M}_3$ isolates proportional asymmetry (scale-invariant imbalance). In this sense, $\mathscr{M}_1$ is more sensitive to high-volume events, whereas $\mathscr{M}_3$ emphasizes subtle directional shifts. All metrics $\mathscr{M}_i$ are computed on logarithmically or ratio-transformed flow variables that are dimensionless, since the transformations involve ratios of power flows to a fixed reference value $(1\text{MW} \cdot \text{week}^{-1})$. Consequently, the summary quantities $\bar{M}$, σ_M, and RMSE in Table 1 are also dimensionless and indicate relative, rather than absolute, deviations in flow behaviour.

Figure 8b illustrates a new representation of CBPFs, where the new adjacency structure is constructed using the rules of a line graph described in Section 3.1.2.

It is vital to highlight that the presented statistics measure different features; hence, there is no universal approach. The first statistic measures total flow of energy, the second one measures the energy flow between countries, and the third one presents the relative measure of asymmetry. There are options to choose from these statistics in terms of what is expected to happen. Another point of view on that is how much information can be preserved by combining both flows. In case the flows are rather proportional, e.g. the strength of electricity flow from Germany to France is about the same as from France to Germany, the first statistic $\mathscr{M}_1$ is suitable. In case the flows strongly differ, it is more insightful to use either the second $\mathscr{M}_2$ statistic or the third statistic $\mathscr{M}_3$.

As no exceptional events are known in the years 2018 and 2019 that could considerably disturb the CBPF network from general geopolitical developments reported in news, we decide to choose these 2 years as Phase I. Any fluctuations within this window are regarded as part of the normal process dynamics and are therefore absorbed into the Phase I variance estimates. Testing different model settings, the final set-up involves a global autoregressive parameter $\tilde{\alpha}$ with $p = 1$, neighbourhood effect $\tilde{\beta}$ with $\mathbf{s} = (1)$, and two covariates with $\mathbf{q} = (0, 0)$ lags that correspond to the weekly aggregated Box–Cox transformed total amount of electricity generated by renewable energy sources only (in contrast to the general electricity flows f_1 and f_2 that have aggregated sources of energy). These two covariates represent renewable energy generation for each country involved in the flow with the effects $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ and were estimated at the global level. This means, for each row in the design matrix, we take the weekly values, apply transformation to each covariate and place it into the design matrix without temporal lag effects, after that we estimate the overall effect on the flows. These data are obtained from the Actual Generation per Production Type dataset, available through ENTSO-E Transparency Platform. For country pairs with missing renewable energy data, values were set to zero. These were predominantly non-EU countries with different disclosure requirements. Alternative sources lacked the required temporal granularity so that the covariates' effect was not estimated for these country pairs.

Originally, we have considered two energy generation types, namely fossil-based and renewable. The fossil type included fossil brown coal/lignite, fossil coal-derived gas, fossil gas, fossil hard coal, fossil oil, fossil oil shale, and fossil peat. The renewable group included biomass, geothermal, hydropumped storage, hydro run-of-river and Poundage, hydro water reservoir, solar, wind offshore, wind onshore, marine, and other renewable. Based on the statistical information criteria like the Bayesian information criterion, we used only the renewable energy data in the final model as covariates.

Table 1 presents the results of estimating the parameters for each choice of the aggregation statistic $\mathscr{M}$ using 7828 observations $[76 \times (52 \times 2 - p)]$, where 76 defines the number of bilateral exchanges (unilateral in case there were no parallel edge existing). As we can observe, the GNARX model fits the data in Phase I well and confirms the relevance of accounting for the network structure as well as external effects when modelling TEN processes. Residual temporal dependence was accounted for via a long-run variance estimator for the centred squared residuals, ensuring appropriate calibration of the control limits. Using the displayed coefficients, we proceed with the monitoring of Phase II, which consists of the years 2020–2022.

It is worth noting that although our simulation and empirical studies do not control for seasonality or annual cycles because we could not validate clearly their presence after data transformation, the correct monitoring relies on the incorporation of seasonality when needed. Otherwise, the variance will increase in case of uncounted seasonality, leading quickly to false alarms. For a full industrial-level application, a formal Phase I stability analysis, such as iterative checks of homogeneity or the removal of atypical periods, would be essential to ensure reliable parameter estimation. Such an analysis was not undertaken here, but would be a natural extension of the empirical study.

5.3 Phase II monitoring

Figures 9, 10, and 11 display the monitoring results during Phase II. Out of 76 country pairs, between 27 and 39 pairs have had a change point according to the implemented CUSUM control charts based on either $\mathscr{M}_1$, $\mathscr{M}_2$, or $\mathscr{M}_3$ statistics. We highlight two periods during Phase II: the period related to severe situations due to the COVID-19 pandemic and the period starting in February 2022 where several crises are expected or happened due to the Russian–Ukrainian war. Considering the global threshold W, we define it to be $W = 0.2$, i.e. when in at least 16 bilateral exchanges, a change point is detected. To make it more interpretable, we note that requiring 16 of 76 bilateral exchanges to signal implies that between 7 and 32 countries are involved, depending on the network structure. The minimum occurs when edges cluster among a small group of countries, while the maximum reflects disjoint exchanges. This mapping offers a tangible understanding of how widespread a structural change must be to trigger a global alarm. The derivation of the threshold UCL_ι used in univariate CUSUM control charts can be found in the online supplementary material.

Comparing the three different aggregation methods, we notice similarities in detecting changes during the pandemic. For instance, according to Kuik et al. (2022), the mild weather conditions and low electricity demand during the coronavirus pandemic lead to the fact that the electricity demand could often be met from renewables alone. This could be the main cause of many signals produced by the control charts during this period.

In addition to pandemic-related shifts, we observe signals that align with known grid events. Notably, the CUSUM alarms in January 2021 correspond to the major grid disturbance reported by ENTSO-E: the split[2] of the Continental Europe synchronous area on 8 January, caused by a substation failure in Croatia. Related signals appear on January 10 (days 4 January–10 January 2021) in AT–HU ($\mathscr{M}_1$), FR–BE ($\mathscr{M}_2$), and FR–IT ($\mathscr{M}_3$), reflecting immediate redispatch and rerouted flows and emergency support from France to Italy.

[2] https://www.entsoe.eu/news/2021/07/15/final-report-on-the-separation-of-the-continental-europe-power-system-on-8-january-2021/.

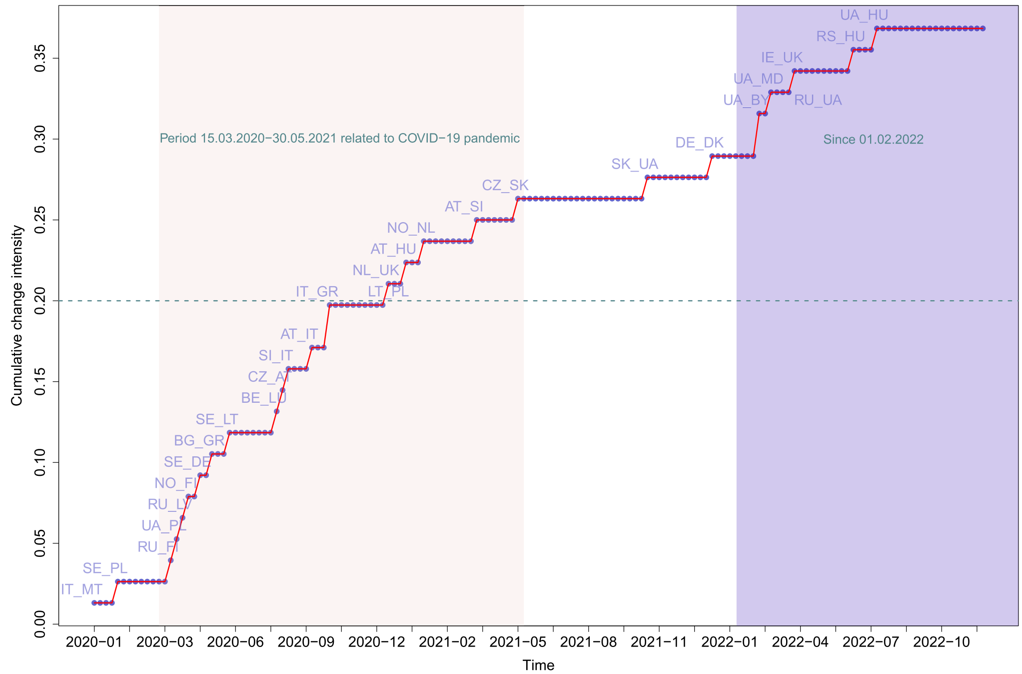

Figure 9. Results of monitoring $\mathscr{M}_1$, obtaining in total 28 change points.

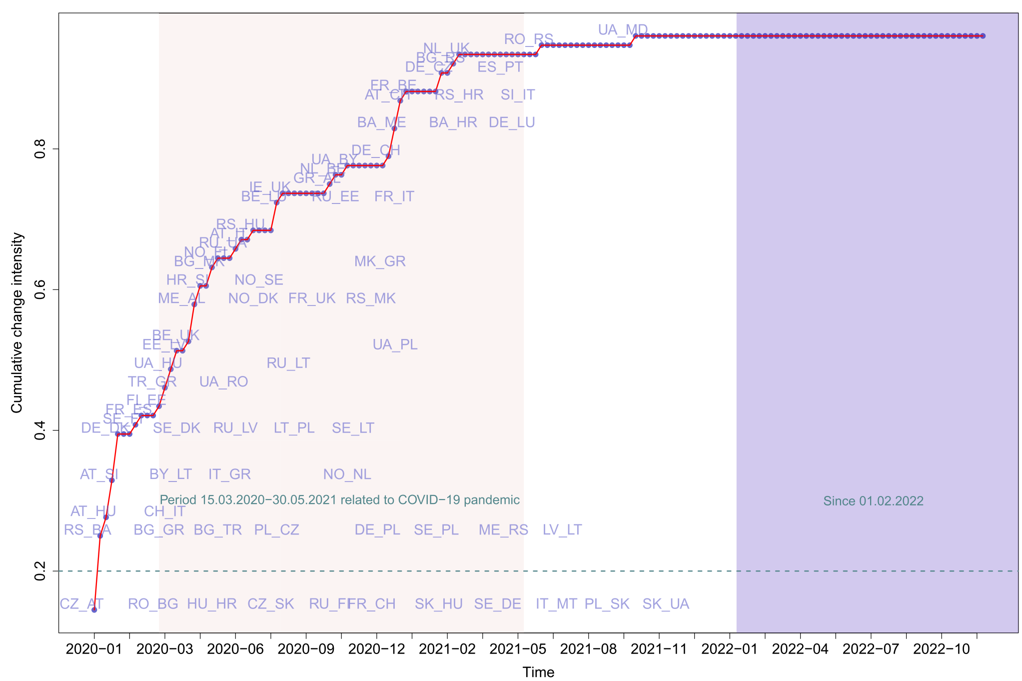

Figure 10. Results of monitoring $\mathscr{M}_2$, obtaining in total 73 change points.

Then, the first detection of $\mathscr{M}_1$ on 6 March 2022 (days 28 February–6 March) coincides with the beginning of the war between Russia and Ukraine. Surprising could be the fact that another country pair experiences a change in the same week, namely Belarus and Ukraine. Later, another

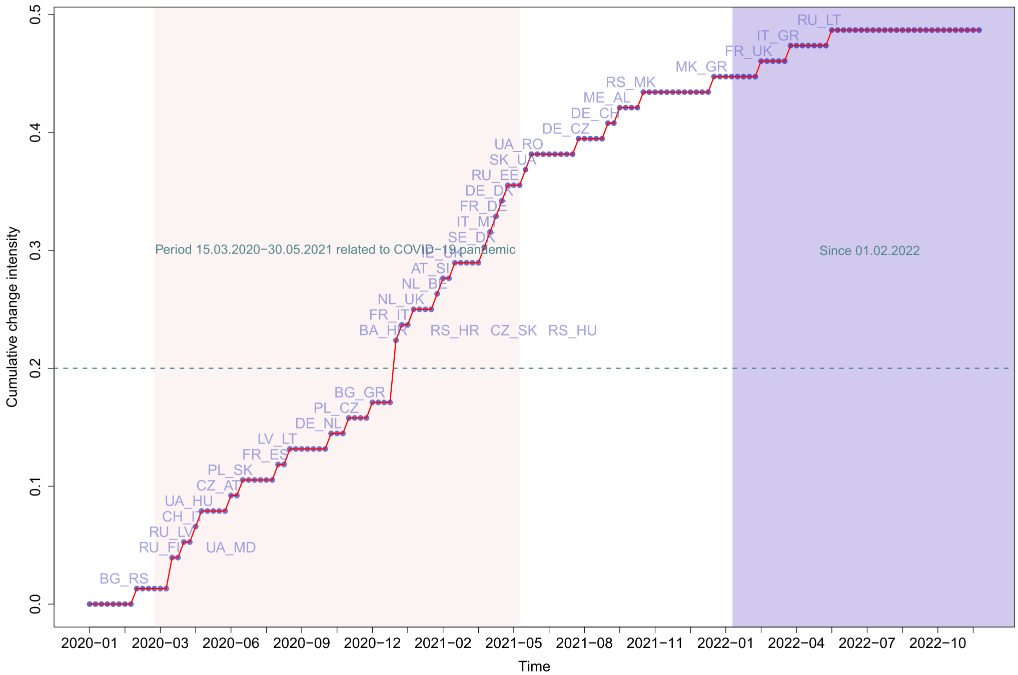

Figure 11. Results of monitoring $\mathscr{M}_3$, obtaining in total 37 change points.

change detection occurs at the end of May–beginning of June in the pair Russia and Lithuania ($\mathscr{M}_3$).[3]

Overall, we can notice that without expert knowledge of the CBPF network, it is challenging to identify a particular reason for the change point; however, we also can see a reliable performance of the proposed framework in detecting major events such as the pandemic and the energy crisis related to the begin of the Russian–Ukrainian war. The caveat to be aware of is the interpretation of the chosen aggregation statistic $\mathscr{M}_1$, $\mathscr{M}_2$, and $\mathscr{M}_3$ as in some of the cases, the signalled anomalies considerably differ.

6 Conclusion

The network with a given structure but a random process on its edges, which we define in this work as a TEN, can be of particular interest for guaranteeing the safety of the infrastructure and foreseeing possible accidents. This work presents a monitoring framework to detect anomalies in TENs by combining the GNARX model and the CUSUM control chart based on residuals.

The question of why monitoring networks with a fixed structure needs special treatment might still exist. For example, why not try to 'randomize' the connections? The explanation is straightforward: It could be possible to select a flow threshold for creating dynamics in the graph by deleting or adding the links; however, the anomalies detected in this case are similar to those that are the focus of the random network monitoring. In our case, we are rather interested in anomalies occurring in the edge process, e.g. changes in the flow's strength or some other temporal deviations. Hence, we introduce a new way of representing TEN processes where edges become the main focus of both modelling and monitoring parts.

The proposed change point detection framework is applied only to the TENs observed at discrete times. It is an open research field on how and when to extend the monitoring to continuous times. Equally important is to research when different adjacency matrices for different time points should be introduced and whether it will benefit the overall monitoring procedure. From the theoretical perspective, some future research should be performed around the comparison of the

[3] https://enmin.lrv.lt/en/news/no-more-russian-oil-gas-and-electricity-imports-in-lithuania-from-sunday/.

properties, which are held when creating the new adjacency matrix using Erdős–Rényi and SBM generators, i.e. random approaches, vs. a line graph that represents a deterministic way to create $\mathbf{Y}'$. Also, to monitor a TEN process at once, we have introduced a cumulative change intensity function. However, an appealing way to perform such monitoring would be a suitable multivariate control chart or another statistical monitoring tool that allows for medium-sized multivariate processes—given that one accounts for the complexity of interpreting the results in case of an alarm.

While this study focuses on CUSUM charts due to their established use in detecting mean and variance changes, alternative charting methods (e.g. EWMA charts) represent a natural direction for future comparative work. Moreover, future research should address the formal definition of the threshold W based on ARL criteria, accounting for the cross-network and temporal correlations that significantly influence threshold determination.

As an empirical illustration, we have monitored a network of bilateral electricity flows across Europe. Our framework could potentially be applied to monitor traffic flows on roads or rail systems to optimize transportation infrastructure and within the computer communication domain, aiding cybersecurity. In an environmental context, TENs may be useful in evaluating water transfers between areas or companies, contributing to sustainable water resource management.

While the present study demonstrates the feasibility of GNARX-edge modelling and monitoring for moderate-size networks and weekly data, extensions to high-frequency settings or larger networks may require additional methodological developments. In particular, sparse estimation techniques or scalable approximations could be explored to reduce computational complexity in applications with $|E| \gg 100$ or sub-daily temporal resolution. Investigating such extensions is an important direction for future research.

This work is devoted to networks with static structures that fit well with the considered process on edges: cross-border physical electricity flows. However, thinking of potential processes with a changing underlying structure over time, it is important to determine whether the monitoring framework would experience substantial changes in terms of re-estimating the parameters in Phase I or challenges in performing multivariate monitoring when some nodes or edges disappear with time. Moreover, the effectiveness of the proposed monitoring procedure strongly relies on the assumption that the GNARX model is suitable for the considered data. Thus, if the data cannot be well represented by the GNARX model, e.g. count data, an alternative modelling approach or an extension to the existing model is required.

Acknowledgments

The authors thank the Editor and the Associate Editor for managing the review process, and the four anonymous reviewers involved at different stages of review for their detailed and constructive comments. KL publishes with the permission of the Executive Director of the British Geological Survey, part of UK Research and Innovation (UKRI).

Conflicts of interest: None declared.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under Project Number 412992257.

Data availability

All data used in this study are open source. The complete data set, R code, derived results, and reproducibility material are provided in the online supplementary material. Specifically, the online supplementary material includes the results of the simulation study, the derivation of the upper control limit for the CUSUM control charts used in the empirical analysis, and example data. The files required to run the simulation study are contained in the folder `SimulationStudy_Section4`, while the files required to reproduce the empirical study, including example data, are provided in the folder `MonitoringCBPF_Section5`. Further details are documented in the `README.txt` file included with the online supplementary material.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series C*.

References

Abrell J., & Rausch S. (2016). Cross-country electricity trade, renewable energy and European transmission infrastructure policy. *Journal of Environmental Economics and Management*, *79*(1), 87–113. https://doi.org/10.1016/j.jeem.2016.04.001

Ahuja R. K., Magnanti T. L., & Orlin J. B. (1993). *Network flows: Theory, algorithms, and applications*. Prentice Hall.

Alexopoulos C., Goldsman D., Tsui K.-L., & Jiang W. (2004). SPC monitoring and variance estimation. *Frontiers in Statistical Quality Control*, *7*, 194–209. https://doi.org/10.1007/978-3-7908-2674-6_13

Ali S., Pievatolo A., & Göb R. (2016). An overview of control charts for high-quality processes. *Quality and Reliability Engineering International*, *32*(7), 2171–2189. https://doi.org/10.1002/qre.1957

Atoui M. A., & Cohen A. (2021). Coupling data-driven and model-based methods to improve fault diagnosis. *Computers in Industry*, *128*(4), 103401. https://doi.org/10.1016/j.compind.2021.103401

Azarnoush B., Paynabar K., Bekki J., & Runger G. (2016). Monitoring temporal homogeneity in attributed network streams. *Journal of Quality Technology*, *48*(1), 28–43. https://doi.org/10.1080/00224065.2016.11918149

Baddeley A., Nair G., Rakshit S., McSwiggan G., & Davies T. M. (2021). Analysing point patterns on networks—a review. *Spatial Statistics*, *42*(4), 100435. https://doi.org/10.1016/j.spasta.2020.100435

Blonder B., Wey T. W., Dornhaus A., James R., & Sih A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, *3*(6), 958–972. https://doi.org/10.1111/j.2041-210X.2012.00236.x

Box G. E., & Cox D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

De Masi G., & Gallegati M. (2007). Debt-credit economic networks of banks and firms: the Italian case. In *Econophysics of markets and business networks: Proceedings of the econophys-Kolkata III* (pp. 159–171). Springer. https://doi.org/10.1007/978-88-470-0665-2_12

Diestel R. (2017). *Graph theory*. Springer.

Diestel R. (2024). *Graph theory*. Springer (print edition); Reinhard Diestel (eBooks).

Erdős P., & Rényi A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen*, *6*(290–297), 18. https://snap.stanford.edu/class/cs224w-readings/erdos59random.pdf

Flossdorf J., & Jentsch C. (2021). Change detection in dynamic networks using network characteristics. *IEEE Transactions on Signal and Information Processing over Networks*, *7*, 451–464. https://doi.org/10.1109/TSIPN.2021.3094900

Grundy T., Killick R., & Svetunkov I. (2026). Online detection of forecast model inadequacies using forecast errors. *Journal of Time Series Analysis*. https://doi.org/10.1111/jtsa.12843

Hanneke S., Fu W., & Xing E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, *4*, 585–605. https://doi.org/10.1214/09-EJS548

Hoffmann T., Porter M. A., & Lambiotte R. (2012). Generalized master equations for non-Poisson dynamics on networks. *Physical Review: E, Statistical, Nonlinear, and Soft Matter Physics*, *86*(4), 046102. https://doi.org/10.1103/PhysRevE.86.046102

Hoffmann T., Porter M. A., & Lambiotte R. (2013). Random walks on stochastic temporal networks. In *Temporal networks* (pp. 295–313). Springer.

Holland P. W., Laskey K. B., & Leinhardt S. (1983). Stochastic block models: First steps. *Social Networks*, *5*(2), 109–137. https://doi.org/10.1016/0378-8733(83)90021-7

Holme P., & Saramäki J. (2012). Temporal networks. *Physics Reports*, *519*(3), 97–125. https://doi.org/10.1016/j.physrep.2012.03.001

Holme P., & Saramäki J. (2023). A map of approaches to temporal networks. In *Temporal network theory* (pp. 1–24). Springer. https://doi.org/10.1007/978-3-031-30399-9_1

Horváth L., Hušková M., Kokoszka P., & Steinebach J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, *126*(1), 225–251. https://doi.org/10.1016/j.jspi.2003.07.014

Hosseini S. S., & Noorossana R. (2018). Performance evaluation of EWMA and CUSUM control charts to detect anomalies in social networks using average and standard deviation of degree measures. *Quality and Reliability Engineering International*, *34*(4), 477–500. https://doi.org/10.1002/qre.2267

Jensen W. A., Jones-Farmer L. A., Champ C. W., & Woodall W. H. (2006). Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology*, *38*(4), 349–364. https://doi.org/10.1080/00224065.2006.11918623

Khrabrov A., & Cybenko G. (2010). Discovering influence in communication networks using dynamic graph analysis. In *2010 IEEE Second International Conference on Social Computing* (pp. 288–294). IEEE. https://doi.org/10.1109/SocialCom.2010.48

Knight M., Leeming K., Nason G., & Nunes M. (2020). Generalized network autoregressive processes and the GNAR package. *Journal of Statistical Software*, *96*(5), 1–36. https://doi.org/10.18637/jss.v096.i05

Knight M. I., Nunes M., & Nason G. (2016). 'Modelling, detrending and decorrelation of network time series', arXiv, arXiv:1603.03221, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1603.03221

Knoth S., & Schmid W. (2004). Control charts for time series: A review. *Frontiers in Statistical Quality Control*, *7*(3), 210–236.

Kolaczyk E. D., & Csárdi G. (2020). *Analysis of network flow data*. Springer International Publishing. pp. 169–186.

Kuik F., Adolfsen J. F., Meyler A., & Lis E. (2022). Energy price developments in and out of the COVID-19 pandemic – from commodity prices to consumer prices. *Economic Bulletin Articles, European Central Bank*, *4*.

Leeming K. (2019). *New Methods in Time Series Analysis: Univariate Testing and Network Autoregression Modelling* [PhD thesis], University of Bristol.

Liu R. Y. (1995). Control charts for multivariate processes. *Journal of the American Statistical Association*, *90*(432), 1380–1387. https://doi.org/10.1080/01621459.1995.10476643

Liu Y., Liu L., Yan Y., Feng H., & Ding S. (2019). Analyzing dynamic change in social network based on distribution-free multivariate process control method. *Computers, Materials & Continua*, *60*(3), 1123–1139. https://doi.org/10.32604/cmc.2019.05619

Malinovskaya A., & Otto P. (2021). Online network monitoring. *Statistical Methods & Applications*, *30*(5), 1337–1364. https://doi.org/10.1007/s10260-021-00589-z

Mantziou A., Cucuringu M., Meirinhos V., & Reinert G. (2023). The GNAR-edge model: A network autoregressive model for networks with time-varying edge weights. *Journal of Complex Networks, 11*(6). https://doi.org/10.1093/comnet/cnad039

Matias C., & Miele V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *79*(4), 1119–1141. https://doi.org/10.1111/rssb.12200

Mattera R., & Otto P. (2024). Network log-ARCH models for forecasting stock market volatility. *International Journal of Forecasting*, *40*(4), 1539–1555. https://doi.org/10.1016/j.ijforecast.2024.01.002

McDermott M. J., Dwaraknath S. S., & Persson K. A. (2021). A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis. *Nature Communications*, *12*(1), 3097. https://doi.org/10.1038/s41467-021-23339-x

Miller B. A., Arcolano N., & Bliss N. T. (2013). Efficient anomaly detection in dynamic, attributed graphs: Emerging phenomena and big data. In *2013 IEEE International Conference on Intelligence and Security Informatics* (pp. 179–184). IEEE.

Montgomery D. C. (2012). *Statistical quality control*. Wiley Global Education.

Nason G. P., & Wei J. L. (2022). Quantifying the economic response to COVID-19 mitigations and death rates via forecasting purchasing managers' indices using generalised network autoregressive models with exogenous variables. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(4), 1778–1792. https://doi.org/10.1111/rssa.12875

Newman M. (2018). *Networks*. Oxford University Press.

Page E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1–2), 100–115. https://doi.org/10.1093/biomet/41.1-2.100

Perry M. B. (2020). An EWMA control chart for categorical processes with applications to social network monitoring. *Journal of Quality Technology*, *52*(2), 182–197. https://doi.org/10.1080/00224065.2019.1571343

Psarakis S., Vyniou A. K., & Castagliola P. (2014). Some recent developments on the effects of parameter estimation on control charts. *Quality and Reliability Engineering International*, *30*(8), 1113–1129. https://doi.org/10.1002/qre.1556

Ross G. J., Tasoulis D. K., & Adams N. M. (2011). Nonparametric monitoring of data streams for changes in location and scale. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, *53*(4), 379–389. https://doi.org/10.1198/TECH.2011.10069

Shaghaghi M., & Saghaei A. (2020). PCA likelihood ratio test approach for attributed social networks monitoring. *Communications in Statistics: Theory and Methods*, *49*(12), 2869–2886. https://doi.org/10.1080/03610926.2018.1491599

Snijders T. A. (2005). Models for longitudinal network data. *Models and Methods in Social Network Analysis*, *1*, 215–247.

Stevens N. T., Wilson J. D., Driscoll A. R., McCulloh I., Michailidis G., Paris C., Parker P., Paynabar K., Perry M. B., Reisi-Gahrooei M., & Sparks R. (2021a). Research in network monitoring: Connections with SPM and new directions. *Quality Engineering*, *33*(4), 736–748. https://doi.org/10.1080/08982112.2021.1974035

Stevens N. T., Wilson J. D., Driscoll A. R., McCulloh I., Michailidis G., Paris C., Paynabar K., Perry M. B., Reisi-Gahrooei M., Sengupta S., & Sparks R. (2021b). Foundations of network monitoring: Definitions and applications. *Quality Engineering*, *33*(4), 719–730. https://doi.org/10.1080/08982112.2021.1974033

Tidriri K., Chatti N., Verron S., & Tiplica T. (2016). Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. *Annual Reviews in Control*, *42*(8), 63–81. https://doi.org/10.1016/j.arcontrol.2016.09.008

Vining G. (2009). Technical advice: Phase I and phase II control charts. *Quality Engineering*, *21*(4), 478–479. https://doi.org/10.1080/08982110903185736

Wilson J. D., Stevens N. T., & Woodall W. H. (2019). Modeling and detecting change in temporal networks via the degree corrected stochastic block model. *Quality and Reliability Engineering International*, *35*(5), 1363–1378. https://doi.org/10.1002/qre.2520

Zhu X., Pan R., Li G., Liu Y., & Wang H. (2017). Network vector autoregression. *Annals of Statistics*, *45*(3), 1096–1123. https://doi.org/10.1214/16-AOS1476