Methods

Kathrin Reibelt*, Jörg Matthes and Veit Hagenmeyer

D4R: a new direct discrete dynamic data reconciliation method for the detection of cyber attacks

D4R: Eine neue Direkt Diskrete Dynamische Data Reconciliation – Methode für die Detektion von Cyberangriffen

https://doi.org/10.1515/auto-2025-0076
Received August 5, 2025; accepted November 25, 2025

Abstract: A novel hybrid method of data reconciliation and gross error detection, applicable for systems with a mixture of dynamic and static system constraints, is developed for the detection of cyber attacks. The requirements for the new application of data reconciliation and similar methods in cybersecurity differ from the requirements for the established use of data reconciliation in automation and control engineering. For the detection of cyber attacks aiming at physical damage the main focus is on significant gross error detection while for classical applications a robust optimization and smoothing of measurement data is the main concern. Therefore the new hybrid method of direct discrete dynamic data reconciliation, as well as similar methods of data reconciliation and Kalman filters with their referring methods of gross error detection are evaluated regarding their aptitude for attack detection in cybersecurity. All considered methods are compared regarding properties resulting from the specific optimization procedure and the detection. The new direct discrete dynamic data reconciliation is indeed shown to outperform the other methods regarding the detection of cyber attacks.

Keywords: gross error detection; cyber-physical systems; cybersecurity; false data injection; dynamic data reconciliation; Kalman filter

Zusammenfassung: Für die Detektion von Cyberangriffen, wurde eine neue, hybride Methode der Data Reconciliation und Großfehlerdetektion entwickelt, die auf Systeme mit einer Mischung aus dynamischen und statischen Systemgleichungen anwendbar ist. Die Anforderungen an die Data Reconciliation und ähnliche Methoden für den neuen Anwendungsfall in der Cybersecurity unterscheidet sich deutlich von den Anforderungen der klassischen Anwendungsfälle in Automatisierung und Regelungstechnik. Zur Detektion von Cyberangriffen, die auf physikalische Schäden abzielen, liegt der Fokus auf der Detektion relevanter Großfehler, während für die klassischen Anwendungsfälle hauptsächlich eine robuste Optimierung und Glättung der Messdaten benötigt wird. Daher werden die neue hybride direkte diskrete dynamische Data Reconciliation, ebenso wie ähnliche Methoden der Data Reconciliation und Kalmanfilter, mit den jeweils zugehörigen Großfehlerdetektionsmethoden geprüft, ob sie zur Angriffsdetektion in der Cybersecurity geeignet sind. Alle betrachteten Methoden werden hinsichtlich der Eigenschaften verglichen, die aus dem jeweiligen Optimierungsprozess und der Detektion folgen. Hierbei zeigt sich, dass die neue direkte diskrete dynamische Data Reconciliation den anderen Methoden hinsichtlich der Detektion von Cyberangriffen überlegen ist.

Schlagwörter: Großfehlerdetektion; Cyber-Physikalische Systeme; Cybersicherheit; False Data Injection; Dynamische Data Reconciliation; Kalman Filter

1 Introduction

Cyber-physical systems are not only threatened by common cyber-attacks, damaging or stealing data, but also by new cyber-attacks aiming at damages in the physical part of the system. Security measures in the systems ensure

***Corresponding author: Kathrin Reibelt**, KIT: Karlsruher Institut fur Technologie, IAI: Institute for Automation and Applied Informatics, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Baden-Württemberg, Germany, E-mail: kathrin.reibelt@kit.edu
Jörg Matthes and Veit Hagenmeyer, KIT: Karlsruher Institut fur Technologie, IAI: Institute for Automation and Applied Informatics, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Baden-Württemberg, Germany

compliance of the system states with restrictions that prevent damages. Therefore, cyber-attacks aiming at physical damages have to use false data injection to undermine those measures. These new cyber-attacks are the subject of the detection methods in this paper. For detection of false data injection in connected cyber-physical systems physical correlations can be considered. In addition, the new detection method utilizes prior knowledge regarding informational properties of the systems components like degree of exposition or commonalities as described in Reibelt [1]. The quality of this cyber-attack detection relies on the full chain of data preparation, anomaly detection and the new consideration of informational properties. Based on models, describing the physical correlations and constraints, there are several methods for improving measurement data and detection of gross errors, which can be understood as anomaly, indicating a cyber attack. Gross errors are errors greater than the measurement noise, classically considered to indicate a sensor fault, outage, or leakage but are also feasible to detect false data injection. The methods differ regarding their optimization and detection process. The comparison of different methods for evaluating the measurements regarding the physical information for cybersecurity shows significant differences in the influence of the procedures, in particular between methods of data reconciliation and Kalman filters. Thereby, confusion in the naming of methods can be found in the literature, as the traditional method of Dynamic Data Reconciliation [2] indeed is a kind of Kalman filter but not a real data reconciliation method.

The present paper extends a new direct discrete dynamic data reconciliation introduced in Reibelt [1] by including both dynamic and static system relations for applicability in common systems. This method is set-up with methods of gross error detection for the detection of cyber attacks. The new method and similar, preexisting methods are evaluated in terms of their usability for the first two steps of detecting false data injection, the data optimization and gross error detection. Some advantages of classical static data reconciliation regarding gross error detection are only transferred to dynamic systems by the new method of direct discrete dynamic data reconciliation introduced in Section 2.2. For enabling a reasonable selection for new applications in general, the present paper shows a detailed comparison of the effects of the different optimization processes especially with regard to gross error detection.

2 Methods

Methods of data reconciliation for dynamic systems are developed from traditional static data reconciliation. The new method (Section 2.2), as well as three preexisting ones, are based on the time-discrete, linearized representation of the system model. The introductions of these methods (Section 2.3) and an additional one using polynomial series expansion (Section 2.4) already focus on commonalities and differences. Traditional dynamic data reconciliation (Section 2.6) is described after the Kalman filter (Section 2.5) is introduced, as there is a close affinity.

2.1 Introduction to data reconciliation (DR)

For DR all relations and laws for measurements and input values are described in the system equation using the array of variables y and the system matrix $\boldsymbol{A}$:

$$\boldsymbol{A} \cdot y(t) = 0 \tag{1}$$

For optimizing the variables, the difference between the measured $y(t)$ and the reconciled values $\hat{y}(t)$ is minimized with respect to their variance and with the system equation as side condition [3]. This means, in DR the reconciled variable values $\hat{y}(t)$ strictly fulfill the system equation (1). Conventionally, the optimization is solved by Lagrange variation, but it is not limited to this method [4]. Gross Error Detection of DR utilizes either the measurement deviation (difference between measured/input values and reconciled values) or the residuals of the system equation (1) for detection and localization of the error.

Classical DR is used for steady-state system equations, that sufficiently describe the behavior in a stationary operation point in chemical facilities where data reconciliation is traditionally used. Dynamic system equations are conventionally given in the form

$$\mathrm{d}x/\mathrm{d}t = \boldsymbol{A} \cdot x(t) + \boldsymbol{B} \cdot u(t) \tag{2}$$

where for most applications u is the array of the input variables with the input matrix $\boldsymbol{B}$ and $x(t)$ is the array of states/measurements. For DR the division of the variables in u and x is just considered as separation between variables appearing in derivative form and variables only appearing directly. Therefore both, $\boldsymbol{A}$ and $\boldsymbol{B}$ are regarded as system matrices. So the main task is to treat the derivative part $\mathrm{d}x/\mathrm{d}t$ in (2).

In classical DR several methods are considered for gross error detection. The measurement test uses the measurement deviation for detection and localization of gross errors [5]. The global test evaluates the sum of the equation residuals [6]. The hypotheses test compares actual residuals to expectation values referring to the possible gross errors [7], [8] and the nodal test identifies the variables included in the equations or contributing to nodes enabling a balance evaluation, that show residuals above the amount explicable by statistical errors only [9].

2.2 New method of direct discrete dynamic data reconciliation (D4R)

The dynamic system equations can be expressed in a time-discrete manner:

$$x(k+1) = \boldsymbol{A}_d \cdot x(k) + \boldsymbol{B}_d \cdot u(k) \tag{3}$$

in addition to the static system equations:

$$y(k) = \boldsymbol{C} \cdot x(k) + \boldsymbol{D} \cdot u(k) \tag{4}$$

Matrices $\boldsymbol{C}$ and $\boldsymbol{D}$ in (4) are identical to the ones in time continuous representation. There are different approaches for determining $\boldsymbol{A}_d$ and $\boldsymbol{B}_d$ in (3) from the time continuous matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ in (2), depending on the model used for signal reconstruction. For the following examples $\boldsymbol{A}_d = e^{\boldsymbol{A}\cdot Ts}$ and $\boldsymbol{B}_d = e^{\boldsymbol{A}\cdot Ts}\cdot\boldsymbol{B}$ is used.

In Reibelt et al. [10] a new method of Data Reconciliation handling dynamics using the discrete equation form (3) is introduced. The variable arrays $x(k)$ and $u(k)$ containing known (measured and input) values are expanded by several time steps, indicated by an index t. Then $x_t(k+1)$ can be expressed by a matrix multiplication with

$$\boldsymbol{Q} = \begin{bmatrix} 0\, 1_x\, 0 & \cdots \\ 0\, 0\;\, 1_x & \ddots \\ \vdots\, \vdots\;\; \ddots & \ddots \end{bmatrix} => \;\; x_t(k+1) = \boldsymbol{Q}\cdot x_t(k) \tag{5}$$

With (5) the dynamic system equations become

$$\left(\boldsymbol{A}_d - \boldsymbol{Q}\right)\cdot x_t(k) + \boldsymbol{B}_d \cdot u_t(k) = 0 \tag{6}$$

To return to the original form of the system equations in DR, with its known optimization solution, the variable arrays representing a piece of signal in the considered time span $x_t(k)$ and $u_t(k)$ are combined to one common variable array and the matrices of (6) are combined into one matrix:

$$\begin{bmatrix} \boldsymbol{A}_d & \boldsymbol{B}_d & -1_x & 0 & 0 & \cdots \\ 0 & 0 & \boldsymbol{A}_d & \boldsymbol{B}_d & -1_x & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \cdot \begin{bmatrix} x\,(k) \\ u\,(k) \\ x\,(k+1) \\ u\,(k+1) \\ \vdots \end{bmatrix} = 0 \tag{7}$$

$$\Rightarrow \qquad \boldsymbol{A}_{dyn} \cdot y_t(k) = 0 \tag{7}$$

For equation (7), the optimization result is already known from classical data reconciliation.

This form introduced in Reibelt et al. [10] considers dynamic relations only.

To consider static relations as they frequently appear it has to be extended.

If the system considered also includes static system relations conventionally matching the form

$$\begin{pmatrix} x(k) \\ u(k) \end{pmatrix} = \begin{bmatrix} \boldsymbol{C}\, \boldsymbol{D} \end{bmatrix} \cdot \begin{pmatrix} x(k) \\ u(k) \end{pmatrix} \tag{8}$$

The static relations (8) can be included by extending the system matrix respectively.

$$\begin{bmatrix} \boldsymbol{A}_d & \boldsymbol{B}_d & -1_x & 0 & 0 & \cdots \\ \boldsymbol{C} - \begin{bmatrix} 1_x \\ \boldsymbol{0}_u \end{bmatrix} & \boldsymbol{D} - \begin{bmatrix} \boldsymbol{0}_x \\ 1_u \end{bmatrix} & 0 & 0 & 0 & \cdots \\ 0 & 0 & \boldsymbol{A}_d & \boldsymbol{B}_d & -1_x & \ddots \\ 0 & 0 & \boldsymbol{C} - \begin{bmatrix} 1_x \\ \boldsymbol{0}_u \end{bmatrix} & \boldsymbol{D} - \begin{bmatrix} \boldsymbol{0}_x \\ 1_u \end{bmatrix} & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{9}$$

$$\cdot \begin{bmatrix} x\,(k) \\ u\,(k) \\ x\,(k+1) \\ u\,(k+1) \\ \vdots \end{bmatrix} = 0 \tag{9}$$

The form in equation (9) enables the application of classical DR optimization for the traces in the chosen time period. Within this new D4R approach, three methods are considered for gross error detection: The measurement test, the global test, and the hypotheses test [10].

2.3 Preexisting methods using discrete system equations

For system equations matching the pattern

$$x(k) = x(k-1) + \boldsymbol{B}_d \cdot u(k) \tag{10}$$

the idea of using a discrete system equation and creating a variable vector of $x_t(k)$ and $u_t(k)$ including two time steps is shown in Rolins et al. [11]. The treatment of several time steps is done by determining the reconciled value each time instant by averaging over the previous and the following step. This Two Step Averaging (TSA) results in values that do not fulfill the system equation completely. Applying Data Reconciliation to this system using several time steps is presented in Darouach et al. [12]. This solution can be considered equivalent to D4R, but is limited to the simplified case of a system matching the pattern

$$x(k+1) = x(k) + \boldsymbol{B}_d \cdot u(k+1) \tag{11}$$

with the identity matrix being the system matrix $\boldsymbol{A}_d$ like in TSA. Moreover, the considered input values $u(k+1)$ are the ones of the future time step.

An approach by Yin et al. [13] considers the difference of the values of two neighboring time steps in the variable arrays. The system matrices are adapted accordingly. This results in a static system description with the known solution. Although the new set of values in this approach of Difference Variables (DV) fulfills its system equation, for those variables appearing only in derivative form, only the differences are reconciled. This solution in Yin et al. [13] only considers one instant in time and the previous one for creating the differences. Only reconciling the differences can lead to a slow deflection of the reconciled values from the measurement. For gross error detection and localization, the nodal test is applied, followed by a measurement test of the suspected variables. Residuals are evaluated for the pre-selection of suspected variables and the measurement deviation is used for the final determination of the faulty variable.

2.4 Polynomial series expansion (PSE)

One approach from Bagajewicz et al. [14] and Bennouna et al. [15] is accessing the derivative using the polynomial series expansion of the variable values:

$$x = \sum_{k=0}^{s} \alpha_k \cdot t^k \qquad u = \sum_{k=0}^{s} \beta_k \cdot t^k \tag{12}$$

With (12) the derivative can be directly expressed as:

$$\mathrm{d}x/\mathrm{d}t = \sum_{k=0}^{s} \alpha_k \cdot k \cdot t^{k-1} = \sum_{k=0}^{s-1} \alpha_{k+1} \cdot (k+1) \cdot t^k \tag{13}$$

Inserting (12) and (13) the system equation becomes

$$\sum_{k=0}^{s-1} \alpha_{k+1} \cdot (k+1) \cdot t^k = \boldsymbol{A} \cdot \left(\sum_{k=0}^{s} \alpha_k \cdot t^k \right) + \boldsymbol{B} \cdot \left(\sum_{k=0}^{s} \beta_k \cdot t^k \right) \tag{14}$$

which leads to the side condition $(k+1) \cdot \alpha_{k+1} = \boldsymbol{A} \cdot \alpha_k + \boldsymbol{B} \cdot \beta_k$. The polynomial series expansion can be expressed using a time matrix $\boldsymbol{T}$

$$T = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 1 & t_1 & t_1^2 & \cdots \\ 1 & t_2 & t_2^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$
$$\Rightarrow \quad x\left(t_1 \ldots t_2\right) = \boldsymbol{T} \cdot \alpha \qquad u\left(t_1 \ldots t_2\right) = \boldsymbol{T} \cdot \beta \tag{15}$$

This shows that the optimization of the polynomially approximated variables x and u can be expressed by an optimization of α and β that can be calculated. For gross error detection and gross error localization, then the measurement deviation is evaluated [15].

2.5 Kalman filter (KF)

In the KFs [16], the variables are divided in input values u, internal parameters x and the observable variables or system states y. Model inaccuracy and noise are contained in variable z. The system is discretely described by equations:

$$\begin{aligned} x(k+1) &= \boldsymbol{A}_d \cdot x(k) + \boldsymbol{B}_d \cdot u(k) + \boldsymbol{G}_d \cdot z(k) \\ \text{and} \quad y(k) &= \boldsymbol{C} \cdot x(k) + \boldsymbol{D} \cdot u(k) \end{aligned} \tag{16}$$

The optimization is divided into two steps. The first step is the prediction of the internal parameters $\hat{x}(k+1)$ based on the current input values $u(k)$ and the estimated parameters $\tilde{x}(k)$ of the previous time step using the system model

$$\hat{x}(k+1) = \boldsymbol{A}_d \cdot \tilde{x}(k) + \boldsymbol{B}_d \cdot u(k) \tag{17}$$

The second step takes place in the next time step and optimizes the predicted value from (17) using the measured value. The correction summand consists of the difference between predicted and measured value weighted by a Kalman gain $\boldsymbol{K}(k)$:

$$\tilde{x}(k) = \hat{x}(k) + \boldsymbol{K}(k) \cdot \left(y(k) - \boldsymbol{C} \cdot \hat{x}(k) - \boldsymbol{D} \cdot u(k) \right) \tag{18}$$

In parallel the covariance matrix $\hat{\boldsymbol{P}}(k)$ of the error of estimation $\hat{\varepsilon}(k) = x(k) - \hat{x}(k)$ is predicted in the first step

$$\hat{\boldsymbol{P}}(k+1) = \boldsymbol{A}_d \cdot \tilde{\boldsymbol{P}}(k) \cdot \boldsymbol{A}_d^T + \boldsymbol{G}_d \cdot \boldsymbol{Q}(k) \cdot \boldsymbol{G}_d^T \tag{19}$$

and optimized in the second step

$$\tilde{\boldsymbol{P}}(k) = (1 - \boldsymbol{K}(k) \cdot \boldsymbol{C}) \cdot \hat{\boldsymbol{P}}(k) \tag{20}$$

The optimization includes the variance $\boldsymbol{Q}(k)$ of model inaccuracy and noise $z(k)$. As the index k already implies, the Kalman gain $\boldsymbol{K}(k)$ is also adapted every time step through

$$\boldsymbol{K}(k) = \hat{\boldsymbol{P}}(k) \cdot \boldsymbol{C}^T \cdot \left(\boldsymbol{C} \cdot \hat{\boldsymbol{P}}(k) \cdot \boldsymbol{C}^T + \boldsymbol{R}(k) \right)^{-1} \tag{21}$$

with respect to the predicted covariance matrix of the error of estimation and the variance $\boldsymbol{R}(k)$ of the measurement noise $v(k)$.

For gross error detection based on Kalman Filter, the postulated error is included as an internal parameter in array $x(k)$ and its amplitude is estimated in every time step [16]. So for false data injection, where the measured or input values affected are unknown, this leads to a limitation of detectable manipulations or a huge effort for calculating several ‘Kalman Filters’ considering all possible fault hypotheses.

2.6 Dynamic data reconciliation (DDR)

The dynamic data reconciliation according to [2] expects model errors δ_t ($\hat{y}_t = x_t + \delta_t$) in addition to measurement errors ε_t ($y_t = x_t + \varepsilon_t$) just like Kalman Filters. It starts with a model based prediction $\hat{y}_t$, where the procedure is not described in detail. These predicted values $\hat{y}_t$ are corrected by adding the weighted measurement deviation:

$$\hat{x}_t = \hat{y}_t + \boldsymbol{K}\left(y_t - \hat{y}_t\right) \tag{22}$$

Thus, DDR (22) is a predictor-corrector-algorithm just like in the KF. In contrast to the KF correction, the gain is not adapted every time step, but only depends on the variance matrices of the measurement errors $\boldsymbol{V}$ and the model errors $\boldsymbol{R}$ that are supposed to be known [2]:

$$\boldsymbol{K} = \left(\boldsymbol{V}^{-1} + \boldsymbol{R}^{-1}\right)^{-1} \cdot \boldsymbol{V}^{-1} = \left(\boldsymbol{1} + \boldsymbol{V} \cdot \boldsymbol{R}^{-1}\right)^{-1} \; [2] \tag{23}$$

The resulting values $\hat{x}_t$ do not fulfill the model completely and can be considered as estimation. Thus the DDR presented by Bai is not actually a DR method. As errors are only reduced with respect to the variance, the measurement deviation still contains information for the detection of gross errors but it is distorted. For gross error detection, a nodal test is applied to the residuals. For every residual a test is conducted, if the respective residual is explicable by statistical errors only. If residual values are suspicious above a predefined significance, variables appearing in all suspicious residuals but nowhere else are tested using the measurement deviation [17].

3 Comparison of methods

The different methods are compared regarding the optimization process and its effects. In particular, the effects on the gross error detection are evaluated. A brief summary is given in Table 1.

To illustrate the effects described in the following comparison, two basic example systems were used. All of the methods discussed except of DV (i.e. D4R, TSA, PSE, KF, and DDR) are treating real dynamic equations. The dynamic system equations used for the following demonstration example are (24) and (25):

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \underbrace{\begin{bmatrix} -1/1 & 0 & 0 & 0 \\ 0.25/2 & -1/2 & 0 & 0 \\ 0.75/3 & 0 & -1/3 & 0 \\ 0 & 1/4 & 1/4 & -1/4 \end{bmatrix}}_{\boldsymbol{A}} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \underbrace{\begin{bmatrix} 1/1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{B}} \cdot u_1 \tag{24}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\boldsymbol{C}} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{D}} \cdot u_1 \tag{25}$$

Input signal u_1 is a sinewave, all initial values are zero. Discretization was done by the c2d Matlab function using the default zero-order hold. For DV indeed a static system model is used reconciling the differences of the variables between the time steps. The system equation used for the following example of DV is (26)

$$\begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix} \cdot \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \Delta x_4 \\ \Delta x_5 \\ \Delta x_6 \end{pmatrix} = 0 \tag{26}$$

where the x_2-trace as well as its share of *0.25* of x_1 was adopted from the dynamic example to create a comparable plot. Normal distributed errors were added to all variables. In all cases x_2 was plotted.

3.1 What are the application cases of the methods and the consequences for their capabilities?

The methods belonging directly to Data Reconciliation, PSE, D4R, DV, and TSA are used for monitoring systems with noisy measurements. Their purpose is the identification of the most probable 'true value' of measurements and input values. PSE, D4R, and DV also contain gross error detection methods for detection of leakages, sensor outages, and calibration issues.

KFs, as well as DDR, focus on the estimation of the system state to provide sustainable information for robust and stable controlling of the process which requires explicit smoothing of traces. Gross error detection in KFs and DDR is used for the purpose of correcting the knowledge of the system state. It is not meant to inform about the error or take measures against its root cause. Especially in KFs, gross error detection is limited to a few selected measurement variables within in the system that are expected to show sensor drifts.

Table 1: Comparison of the data improvement methods by different criteria.

	Data reconciliation				
	Discrete				
	D4R/DV	**TSA**	**PSE**	**DDR**	**KF**
Application	None/flow process monitoring and gross error detection	Monitoring	Wind turbine monitoring and gross error detection	Chemical process distillation process control	Various dynamic systems process control
Considered elements	System states: measurements and input values			Measurements only	
Trusted elements	Model			Input values	
Smoothing depending on	Implicit smoothing within considered time period/none	Averaging of values from neighboring time steps	Order of series expansion without respect to system dynamics	Measurement deviation weighted by variances	Measurement deviation and scattering of previous values
Optimization/estimation/ prediction	Optimization	Optimization and averaging	Optimization	Prediction and correction to estimate	
Optimization condition	Minimizing square variable deviation divided by variance side condition: system equation not described			No actual optimization	Minimizing covariance of estimation error
Dynamic	Non-causal (except DV), no time shift			Causal, time delay	
Conclusion	Strong/weak reduction of statistical faults, values fit model	Weak reduction of statistical faults, smoothing by averaging, values don't fit model	Strong smoothing prior to reconciliation values fit model	Reduction of statistical faults, smoothing after use of model, values don't fit model	Strong reduction of statistical faults, smoothing after use of model, values don't fit model
Gross error detection	Based on residuals or deviation of measured and input values from system model constraints	Not defined, not sensible for gross error detection	Deviation of measured and input values from series expanded and recon. Values (general bias error model in series expansion)	Measurement deviation or modeled as time constant bias co-calculated by reconciliation	Expected gross errors are introduced in the model as variable and estimated every time step
Conclusion	Well appropriate for gross error detection	Limited use because model fulfillment is destroyed by averaging	Limited use as data is debased prior to reconciliation	Well appropriate if input variables are reliable	Appropriate only if gross errors are to be expected in a small number of variables

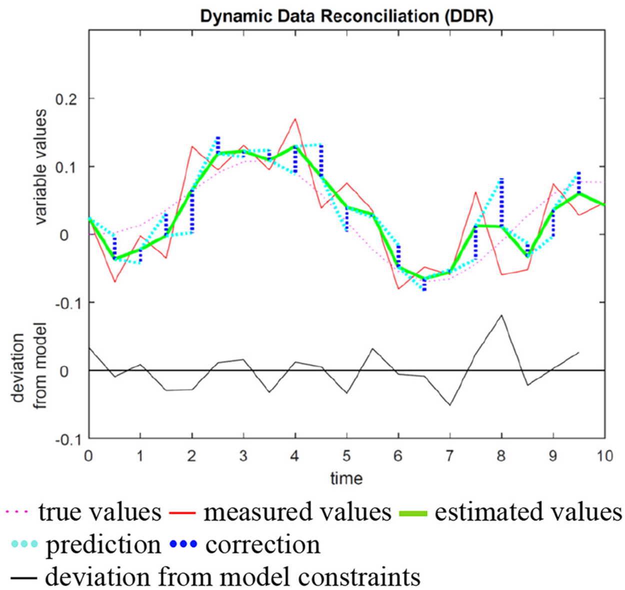

Figure 1: For DDR the next value is predicted and then corrected towards the actual measurement, just like for KF. The correction provides some trace smoothing. The model constraints are not fulfilled, resulting in a deviation shown in the lower plot.

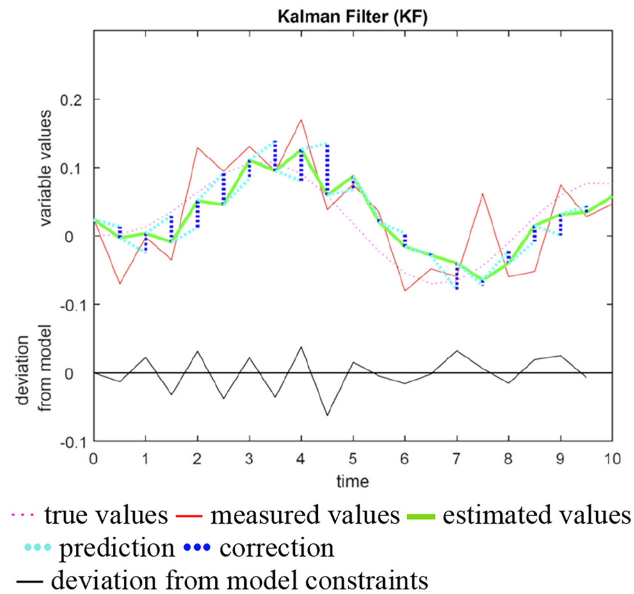

Figure 2: For KF, the next value is predicted and then corrected towards the actual measurement. The correction provides some trace smoothing. The model constraints are not fulfilled, resulting in a deviation shown in the lower plot.

3.2 Whats the conditions and proceedings of smoothing by the methods?

Smoothing in the time dimension helps to reduce statistical faults, but can also mask actual deviations. For both, KF and DDR, estimated values are smoothed by a correction after predicting the value using the model. In both cases the correction depends on the difference between the measured and the predicted values weighted by a correction factor (Figures 1 and 2). For DDR the correction factor depends on the variances of the measurements and the model faults and therefore is time invariant. In contrast in KFs, the Kalman gain is adapted every time step minimizing the covariance of the estimation error. So the smoothing and adaption depends on the previous values.

In both KF and DDR, the estimated values can still violate the systems equations since they result from a weighted mean of predictions and possibly model-violating measurements. Actual faults can be masked. KF and DDR need to smooth the traces, but smoothing is not limited to them. In TSA a smoothing effect is achieved by averaging the solutions of the neighboring time steps (Figure 3). This only provides a small reduction of statistical faults. In PSE the traces are smoothed by the series expansion, prior to the data reconciliation. The extent of the smoothing strongly depends on the order of the series expansion, which must suit the shape of the considered traces (Figure 4). This smoothing by serial expansion masks a lot of artefacts. In the D4R discrete treatment, data reconciliation returns a solution fulfilling the dynamic system equations, a continuously differentiable trace, which is a kind of smooth trace within the considered time span, but does not result from smoothing (Figure 5). So actual faults are not masked. The reduction

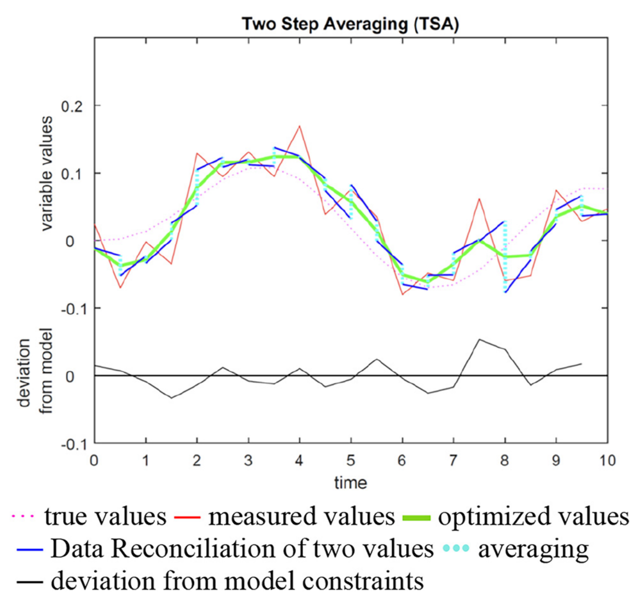

Figure 3: For TSA data reconciliation is applied for all neighboring values. This gives two reconciled values for every time instant. To get the optimized value these two values are averaged, which provides some smoothing. The optimized values deviate from model constraints (lower plot).

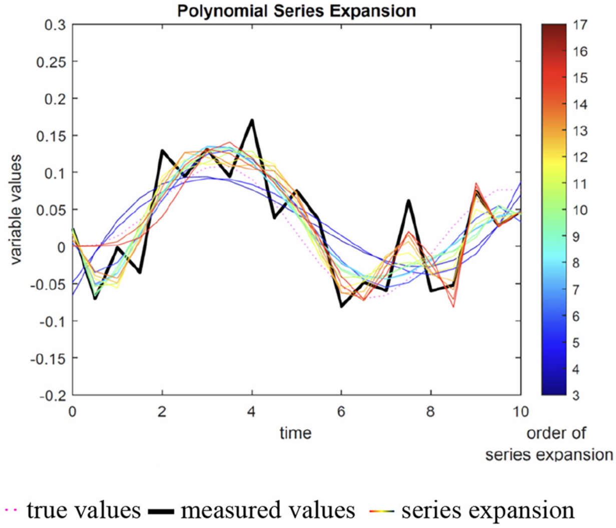

Figure 4: This plot illustrates the difficulties with series expansion prior to data reconciliation. Whether the traces are described properly, strongly depends on the order of series expansion. The smoothing also depends on the order of series expansion and does not respect the model constraints.

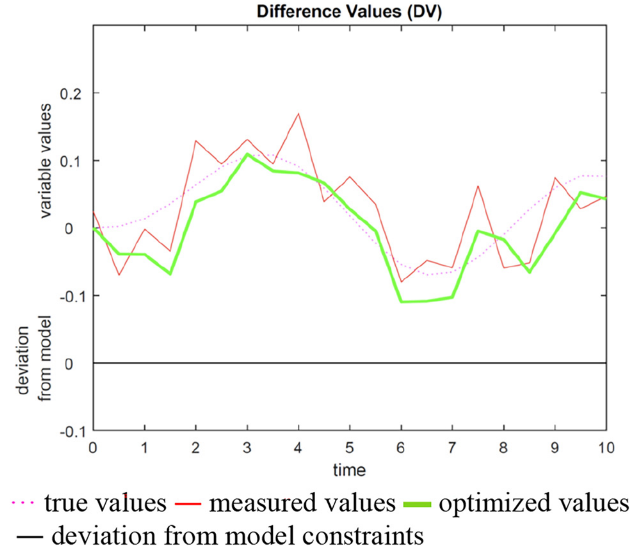

Figure 6: DV requires the fulfillment of the model constraints, thus the deviation is 0. As only the differences between two time steps are reconciled, there is no smoothing and the reconciled traces are slightly deflected.

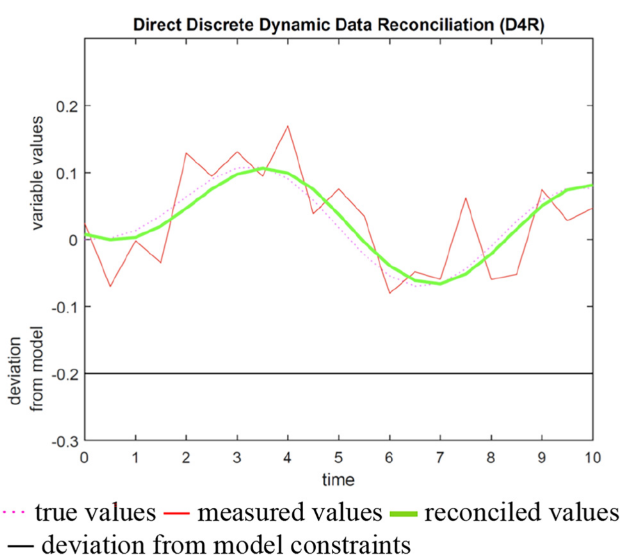

Figure 5: D4R consequently requires the fulfillment of the model constraints, thus the deviation is 0. This also leads to smoothing and strongly reduces statistical errors.

of statistical faults depends on the length of the considered time span. A long time span leads to small statistical faults, but requires a long data history and causes high calculation effort and time. Therefore the length of the considered time span is limited by the response time required for defense measures against cyber attacks. DV does not provide any smoothing (Figure 6).

3.3 How is data improved by the different methods?

In PSE, D4R, DV, and TSA an optimization is undertaken by minimizing the square difference between measurement/input values and reconciled values weighted by their variance with the system model as a side condition. These reconciled values fulfill the system equations. In TSA, as a subsequent step, the two neighboring values are averaged to the final values, which do not fulfill the system equations anymore. In contrast in PSE, D4R, and DV the final reconciled values do fulfill the system equations. Although the reconciled values in D4R fulfill the system equations, they do not follow the true values exactly (Figure 5). The improvement of the correction factor in KFs is an optimization minimizing the estimation error. But for KFs and DDR no actual optimization of the measurement values is done. The estimate of the previous time step is trusted and used to predict the current values, and the current measurements are only included for correction of the predicted values limited by the correction factor. KF and DDR are causal, only previous values have influence on the estimate of future values, but future values are not used for improvement of previous values. Together with the damping effect of the correction factor, this also leads to a slight time delay of the estimated values in case of fast state changes (Figure 7). In DV, the optimization only includes the current and the previous point in time. So DV is causal, too. In contrast, PSE,

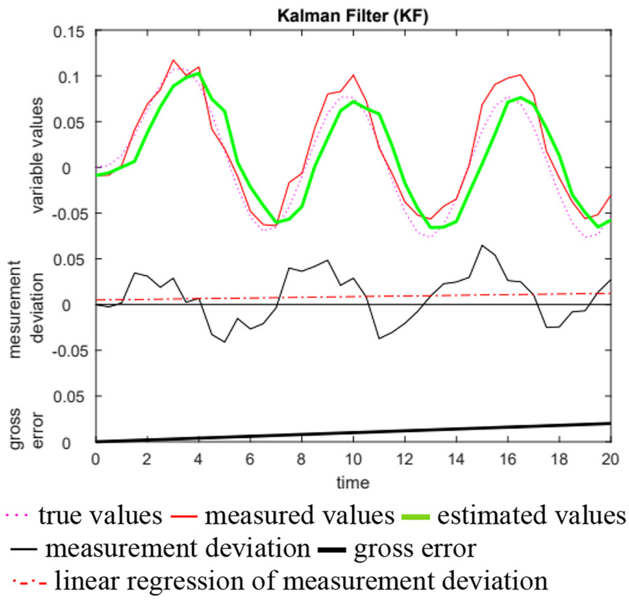

Figure 7: In KFs the adaptive correction can follow a slowly increasing gross error and therefore impede the detection as the measurement deviation keeps oscillating around 0. For dynamic behavior KFs always have a slight time delay.

D4R, and TSA are non-causal because there is no dedicated direction of influence for optimization.

3.4 How are gross errors detected in the various methods?

PSE, D4R, DV, and TSA trust in the model and consider the system states consisting of the measurements and input values as affected by statistical errors and potentially by gross errors. In contrast, KF and DDR trust the input values and consider the models as faulty and only the measurements as affected by statistical errors. Potential gross errors in KF are modeled and estimated for selected measurement variables only. In DDR gross errors are potentially expected in all measurement variables, while input values are trusted. For gross error detection in DDR, two methods are proposed. Either the measurement deviations are evaluated, which is the deviation from the model constraints in the last step, weighted by the correction factor. Or the gross error is modeled as time constant bias of the measurements that is co-calculated during the reconciliation process. This means it represents the average bias of the measurement errors in the considered time interval. In KFs gross errors are calculated by modeling them as additional variables and estimating them in every time step. This method is limited to a few selected measurement variables that are suspected to be contaminated by a gross error, usually a sensor drift. For all other measurements, a gross error cannot be detected but might appear as pretended gross errors in related variables due to smearing effects. For 'KFs' the deviation between measured and estimated values is minimized and does not provide reliable information about gross error, so a measurement test cannot be applied. Slowly and steadily increasing gross errors are compensated by the correction gain and therefore not detectable (Figure 7). In contrast in D4R increasing gross errors are represented by increasing measurement deviations (Figure 8). In addition to the measurement deviation, the residuals of the system equations are used for gross error detection in D4R. Both contain reliable information about gross errors as the data improvement in D4R provides significant reduction of statistical faults without distorting the data. In PSE gross error detection is also based on evaluating measurement deviations. The detection performance strongly depends on the extent the data is properly described by the series expansion. For DV first a so-called nodal test is applied to the residuals. If values in the residual array are suspicious, variables appearing in all equations of these residuals, are tested using the measurement deviation. The performance is similar to gross error detection in steady-state data reconciliation. For TSA no gross error detection is proposed. As the reduction of statistical errors is only based on two instants in time, only a limited improvement is achievable. In addition, the averaging adds a deviation from the system model (Figure 3). So deviations between measurements or input values and the optimized values by TSA are distorted and therefore difficult to evaluate.

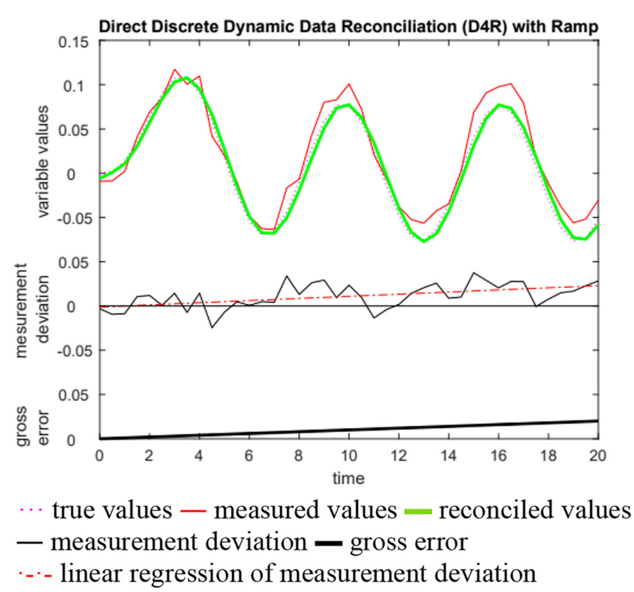

Figure 8: D4R request strict model fulfillment and the slowly, but systematically increasing measurement deviations enables the detection of a slowly increasing gross error.

4 Discussion

The differences observed affect the aptitude and applicability for systems and purposes. The restrictions and advantages of the considered methods is discussed in the following section.

KF perfectly fulfills its common use for improving noisy measurement data. It is very tolerant towards imprecise models due to the continuously tuned correction term. With this correction also faults in input variables are treated implicitly. But as the correction depends on the previous values, the results are hardly traceable or explicable. For gross error detection, KF is only feasible if integrity can be guaranteed for most of the measurement variables.

DDR requires knowledge of the variance of the model fault for determining the correction factor and considers the model error as normally distributed. The correction treats a certain amount of model faults and to some extent compensates faults in input variables. Although the dependency of the previous values is smaller compared to the KF as the correction gain is constant, it still is included in the prediction of values and therefore the improved values are not completely explicable. Gross error detection is only applicable for gross errors in measurement variables. Gross errors in input variables are not detectable so input values have to be reliable. This might not apply to all kind of actors and for manipulation detection, manipulation of input values has to be considered.

For PSE value traces have to be well followed by series expansion of a certain, known order. For system equations containing higher than first-order derivatives, PSE is very useful. The traces are smoothed by the series expansion and therefore are distorted to some extent before data reconciliation. So for gross error detection evaluating the difference between raw data and series expanded and reconciled data, the origin of the measurement deviation is not only gross errors.

TSA only considers two points in time for reconciliation, so the ability to reduce statistical errors is limited. The averaging of the values calculated for two neighboring transitions provides additional smoothing but leads to values that do not fulfill the system model. The remaining statistical errors and the distortion by averaging limit the capability of TSA for gross error detection.

In DV there are also only two points in time considered for reconciliation, so the traces are not smoothed. Optimized values do fulfill the system equation. Reduction of statistical errors is only based on redundancy in the system evaluated at one instant, values from a second time-step are only used for creating the difference. For gross error detection, the deviation between measured or input values and the reconciled values do provide information about the error. Compared to the similar method D4R considering several time steps, the gross error detection in DV is more disturbed by statistical errors.

D4R does not provide explicit smoothing, but the data reconciliation enforces compliance with the system model. The resulting traces are therefore smooth depending on the system equation. By considering several time steps statistical errors are not only reduced using the system redundancies but also by considering several time steps. As the reconciled values fulfill the system model, the measurement (and input value) deviation as well as the equation residuals provide information about gross errors. The reduction of statistical errors further improves the gross error detection.

D4R in the presented form can be applied for all linearizeable systems. It seems likely to be applicable for nonlinear systems, too. This has to be investigated in future work.

From the methods compared, D4R is the most applicable method for gross error detection and thus for the detection of cyber attacks.

5 Conclusions

A new method of direct discrete dynamic data reconciliation treating both, static and dynamic system equations has been presented together with a brief summary of similar preexisting methods for comparison.

All methods have been evaluated and compared for different aspects of their procedure, preconditions and results focusing on the aptitude for the detection of cyber attacks.

The established methods ‘Kalman Filter’ and ‘Dynamic Data Reconciliation’ use a predictor-corrector-form and can only detect faults in measurement variables, the ‘Kalman Filters’ can even detect gross errors only in a few selected measurement variables, where gross errors are previously modeled by the ‘Kalman Filters’ designer. However, the variables affected by cyber attacks are hardly predictable and not restricted to measurement variables. In particular, also targeting input values is reasonable for attackers, as they are linked to actuators. The estimated values, used as reference for gross error detection, are not completely predictable by calculation as they depend on the full history of values. As another shortcoming of the ‘Kalman Filter’ the Kalman gain compensates slowly increasing gross errors and make them undetectable. Furthermore, some of the traditional methods distort the value prior to or after the data reconciliation by polynomial series expansion or averaging (two step averaging), but not so the new direct discrete dynamic data reconciliation method. In these cases, it is not possible to distinguish to what extend the reconciliation procedure and to what extend potential gross errors contribute to any residuals, which hampers their application to the detection of cyber attacks.

The consideration of several points in time, as presented in the new direct discrete dynamic data reconciliation method, is shown to improve the reduction of statistical errors and therefore the detectability of gross errors.

The new method of direct discrete dynamic data reconciliation thus outperforms the existing methods regarding the aspects considered for comparison, as it can detect errors in all variables, and the smoothing is done based on the system dynamics which leads to a significant reduction of statistical errors, and the optimization leads to model compliant values. These reduced statistical errors and the model compliant values provide significant information for error detection. Direct discrete dynamic data reconciliation therefore is superior to the established methods and ready to be applied for the detection of cyber attacks in cyber-physical systems. In future work the applicability oft he new method has to be validated for real world applications.

Acknowledgments: We acknowledge the financial support of the Helmholtz Association of German Research Centres (HGF) within the framework of the Program-Oriented Funding POF IV in the program Energy Systems Design (ESD).
Research ethics: Not applicable.
Informed consent: Not applicable.
Author contributions: Reibelt: Analysis and initial draft. Matthes: supervision, corrections and extension. Hagenmeyer: corrections, extension, acquisition of funds.
Use of Large Language Models, AI and Machine Learning Tools: Not applicable.
Conflict of interest: The authors state no conflict of interest.
Research funding: Helmholtz Association of German Research Centres (HGF) Program-Oriented Funding POF IV, program: Energy Systems Design (ESD).
Data availability: The data that support the findings of this study are available from the corresponding author, K. Reibelt, upon reasonable request.

References

[1] K. Reibelt, “Ein neues Verfahren zur Detektion von Cyber-Angriffen auf zukünftige Energiesysteme unter Nutzung physikalischer Modellansätze,” Doctoral thesis, Karlsruhe Institute of Technology, 2025. Available at: https://www.publikationen.bibliothek.kit.edu/1000183407.
[2] S. Bai, J. Thibault, and D. D. McLean, “Dynamic data reconciliation: Alternative to Kalman filter,” *J. Process Control*, vol. 16, no. 5, pp. 485—498, 2006.
[3] S. Narasimhan and C. Jordache, *Data Reconciliation and Gross Error Detection*, Houston, Texas, Gulf Professional Publishing, 1999.
[4] M. J. Leibman, T. F. Edgar, and L. S. Lasdon, “Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques,” *Comput. Chem. Eng.*, vol. 16, nos. 10—11, pp. 963—986, 1992.
[5] A. C. Tamhane, “A note on the use of residuals for detecting an outlier in linear regression,” *Biometrika*, vol. 69, no. 2, pp. 488—489, 2010.
[6] G. Almasy and T. Sztano, “Checking and correction of measurements on the basis of linear system model,” *Probl. Control Inf. Theory*, vol. 4, pp. 57—69, 1975.
[7] J. Rosenberg, R. S. H. Mah, and C. Iordache, “Evaluation of schemes for detecting and identifying gross errors in process data,” *Ind. Eng. Chem. Res.*, vol. 26, no. 3, pp. 555—564, 1987.
[8] S. Narasimhan and R. S. H. Mah, “Generalized likelihood ratio method for gross error identification,” *AIChE J.*, vol. 33, no. 9, pp. 1514—1521, 1987.
[9] R. S. H. Mah, G. M. Stanley, and D. M. Downing, “Reconciliation and rectification of process flow and inventory data,” *Ind. Eng. Chem. Process Des. Dev.*, vol. 15, no. 1, pp. 175—183, 1976.
[10] K. Reibelt, H. B. Keller, V. Hagenmeyer, and J. Matthes, “Dynamic model based detection of cyberattacks in industrial facilities,” in *Computer Aided Chemical Engineering 50, 31st European Symposium on Computer Aided Process Engineering: ESCAPE-31*, M. Türkay, Ed., 2021.
[11] D. K. Rolins and S. Devanathan, “Unbiased estimation in dynamic data reconciliation,” *AIChE J.*, vol. 39, no. 8, pp. 1330—1334, 1993.
[12] M. Darouach and M. Zasadzinski, “Data reconciliation in generalized linear dynamic systems,” *AIQChE J.*, vol. 37, no. 2, pp. 193—201, 1991.
[13] W. Yin, R. Gang, and W. Shuqing, “Linear dynamic data reconciliation: Refinery application,” *IFAC Proc.*, vol. 34, no. 25, pp. 561—566, 2001.
[14] M. Bagajewicz and Q. Jiang, “An integral approach to plant linear dynamic reconciliation,” *AIChE J.*, vol. 43, no. 10, pp. 2546—2558, 1997.
[15] O. Bennouna, N. Heraud, M. Rodriguez, and H. Camblong, “Data reconciliation & gross error detection applied to wind power,” *Proc. Inst. Mech. Eng., Part I: J. Syst. Control Eng.*, vol. 221, no. 3, pp. 497—506, 2007.
[16] R. Marchthaler and S. Dingler, *Kalman-Filter — Einführung in die Zustandsschätzung und ihre Anwendung für eingebettete Systeme*, Wiesbaden, Germany, Springer Vieweg, 2017.
[17] S. Bai, J. Thibault, and D. D. McLean, “Simultaneous measurement bias correction and dynamic data reconciliation,” *Can. J. Chem. Eng.*, vol. 85, no. 1, pp. 111—117, 2007.

Bionotes

Kathrin Reibelt
KIT: Karlsruher Institut fur Technologie, IAI: Institute for Automation and Applied Informatics, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Baden-Württemberg, Germany
kathrin.reibelt@kit.edu

Kathrin Reibelt received a diploma degree in physics from the Karlsruhe Institute of Technology (KIT) at the Institute for Data Processing and Electronics (IPE) in 2014. After gaining experience with informational engineering of industrial sensors in a private company, she returned to KIT for her doctorate. In summer 2025 she received the Dr.-Ing. degree from the faculty of informatics at KIT. The thesis was about including physical information to the detection of cyberattacks resulting in the new method of direct discrete Data Reconciliation published in this paper.

Jörg Matthes
KIT: Karlsruher Institut fur Technologie, IAI: Institute for Automation and Applied Informatics, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Baden-Württemberg, Germany

Jörg Matthes received the Dipl.-Ing. degree in mechatronics from the Technical University of Freiberg in 1999 and the Dr.-Ing. degree from the Technical University of Karlsruhe in 2004. Since 2000, he is a research staff member at the Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology (KIT). Since 2020 he is apl. Professor at KIT. His research interests are in control and flexibilization of energy intensive processes.

Veit Hagenmeyer
KIT: Karlsruher Institut fur Technologie, IAI: Institute for Automation and Applied Informatics, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Baden-Württemberg, Germany

Veit Hagenmeyer is Professor of Energy Informatics with the Faculty of Informatics, and the Director of the Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany. His research interests include modeling, optimization and control of sector-integrated energy systems, machine learning based forecasting of uncertain demand and production in energy systems mainly driven by renewables, and integrated cyber-security of such systems.