Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

# Bias Invariant Robust Estimation for Data Reconciliation of Nonlinear Dynamic Systems

Eduardo L. T. Conceição

CIEPQPF, Chemical Process Engineering and Forest Products Research Centre, University of Coimbra, Rua Sílvio Lima – Pólo II, 3030-790 COIMBRA, Portugal

**Correspondence**
Correspondence concerning this article should be addressed to E. L. T. Conceição at etc@eq.uc.pt

A few specific scenarios applied in examples of only five works suggest that the solely use of robust estimators for nonlinear dynamic data reconciliation is able to cope with biased measurements. We present a counterexample to that belief on a dynamic model of two CSTRs that gives rise to wrong transient behaviour. Motivated by this, we introduce and examine an invariant approach to measurement bias. It is based on the location invariance assured by a robust measure of scale used when detecting a sequence of consecutive differences between measured and reconciled values of the same sign. When applied to the counterexample, it can be seen that the procedure has the correct behaviour and shows good results.

**KEYWORDS:**
data reconciliation, nonlinear dynamic systems, measurement bias, outliers, robust estimation

## 1 | INTRODUCTION

It can be considered that the use of robust statistics [1,2] in data reconciliation problems is now well established for the straightforward treatment of outliers (a sporadic atypical measurement).[3] Recall that the presence of outliers precludes the exact knowledge of the data distribution, which is fundamental for classical statistical approaches to produce reliable estimates.

Within the framework of *robust data reconciliation* and *model-based* methods, only a very small number of works addressed so far the presence of a systematic error persistent in time, namely, measurement bias [4,5,6,7,8] or complete sensor failure [9,10,11] in the context of *nonlinear dynamic processes*. (Note, however, that Prata, Pinto, and Lima[4] consider three consecutive outliers of the same size for representing measurement bias.) Despite the lack of theoretical understanding, the empirical results in these papers suggest that the use of *univariate* robust estimators has the (desired) side effect of counteracting the detrimental effect of biases or complete failure. Having said that, notice that these estimators, being univariate, have been developed to exploit the

temporal redundancy arising from a sequence of observations, but not the spatial redundancy brought by the process model. Consequently, it should be emphasised that these results are not comprehensive enough to provide a definitive picture.

Two reasons motivate the above studies regarding the problem of instrument bias. The first is that robust estimators avoid the need to explicitly estimate the magnitude of the bias in order to compensate for its effects as is done in other methods, for instance, the method developed by Abu-el-zeet, Becerra, and Roberts [12] or the mixed integer non-linear approaches. [13] The second is that procedures for detection of bias or complete failure (apparently) are also not needed.

This last finding stands in contrast to Llanos, Sanchez, and Maronna [14], which more recently conducted a Monte Carlo simulation study using a steady state benchmark problem that assessed the simultaneous effect of biases and drifts on the performance of different types of robust M-estimators. Measurements were contaminated with either outliers or a mixture with equal amounts of biases and drifts. It was found that with respect to a baseline case without contamination, the mean square error (MSE) of the reconciled values significantly deteriorates when measurements are contaminated by the mixture of biases and drifts compared with almost no change when they are contaminated by outliers. From this, the authors concluded that the explicit detection of systematic errors combined with corrective actions should be used for robust data reconciliation when they are present. It seems reasonable to suppose that this is also valid for dynamic data reconciliation.

Also, our experience with minimising only a robust loss function under instrument bias for a particular case raises concerns regarding this type of approach.

Bias is characterised by having a constant magnitude meaning that the shape and precision of the true measurements are preserved in this abnormal measurement. It is therefore natural to choose a measure invariant under a shift in location in the formulation of the data reconciliation problem. For unbiased measurements, the formulation changes to a MSE-based one. Evidently, one should use robustified versions to cope with outliers. Another idea is to detect biases by a sequence of consecutive differences between measured and reconciled values with the same sign.

Taken together, these two ideas form the basis of our proposed method capable of simultaneously handle instrument biases and outliers, which is described in detail in the following section. Section 3 presents some aspects of the algorithm used for computing the reconciled values. We tested the method on the example problem detailed in Section 4. In Section 5 we give a counterexample to show that robust loss functions do not necessarily handle all cases of measurement bias. Then, the results of the experiments are presented in Section 6. Concluding remarks are given in Section 7.

## 2 | DESCRIPTION OF THE METHOD

It is assumed that process variables are directly measured. The presence of random noise or error, $\epsilon_{i,j}$, the possible contamination by outliers, $O_{i,j}$, and the potential occurrence of biases, $B_i$ in measurements can be represented by the following model

$$y_{i,j} = x_{i,j} + \epsilon_{i,j} + O_{i,j} + \left\{ B_i : i \in I \right\}, \tag{1}$$

where the index $i, j$ refers to the $i$th variable at sampling time $j$, $I$ is the set of biased variables, $y_{i,j}$ represents the process measurement, and $x_{i,j}$ stands for its true value. The measurement noise $\epsilon_{i,j}$ is typically modelled as zero-mean Gaussian with homoscedastic variance $\sigma^2_{\epsilon,i}$ and the assumption of independence among different variables.

Define a moving data window of fixed length $N$ advanced one time increment at a time. Then, the nonlinear programming formulation of the dynamic data reconciliation problem is

$$\hat{\boldsymbol{x}}_k = \arg\max_{\boldsymbol{x}_k} D\left\{ (y_{i,j} - x_{i,j}) : \underbrace{\overbrace{j = k-N+1, \dots, k}^{\text{moving window}}}_{k=N, N+1, N+2, \dots}, \; i = 1, \dots, n \right\} \tag{2}$$

subject to

$$\boldsymbol{f}\left( \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}, \boldsymbol{x}, t \right) = \boldsymbol{0} \tag{3a}$$

$$\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0} \tag{3b}$$

$$\boldsymbol{x}^{\mathrm{L}} \leq \boldsymbol{x} \leq \boldsymbol{x}^{\mathrm{U}} \tag{3c}$$

where $D$ plays a role analogous to a loss function and will be simply referred as loss function, $n$ is the number of measured variables, $k$ is the current time interval, $\boldsymbol{f}$ is the system model described as a set of differential algebraic equations, $\boldsymbol{g}$ is the set of inequality constraints, $\hat{\boldsymbol{x}}$ is the vector of reconciled values for measured and unmeasured variables, and $t$ denotes continuous time. Let us remark that in the above formulation we assume that the solution to model (3a) is to be obtained by a numerical solver as an inner loop of an optimisation algorithm.[15]

### 2.1 | Bias detection

For simplicity of treatment and as a first step towards understanding and elucidating some aspects of the behaviour of the proposed method, let us only consider the case of measurement bias of *infinite duration* (permanent bias). The test we consider is based on the existence of a sequence (of the same length as the moving time window) of residuals $r_{i,j} = y_{i,j} - \hat{x}_{i,j}$ with the same sign, attributed to the effect of bias.

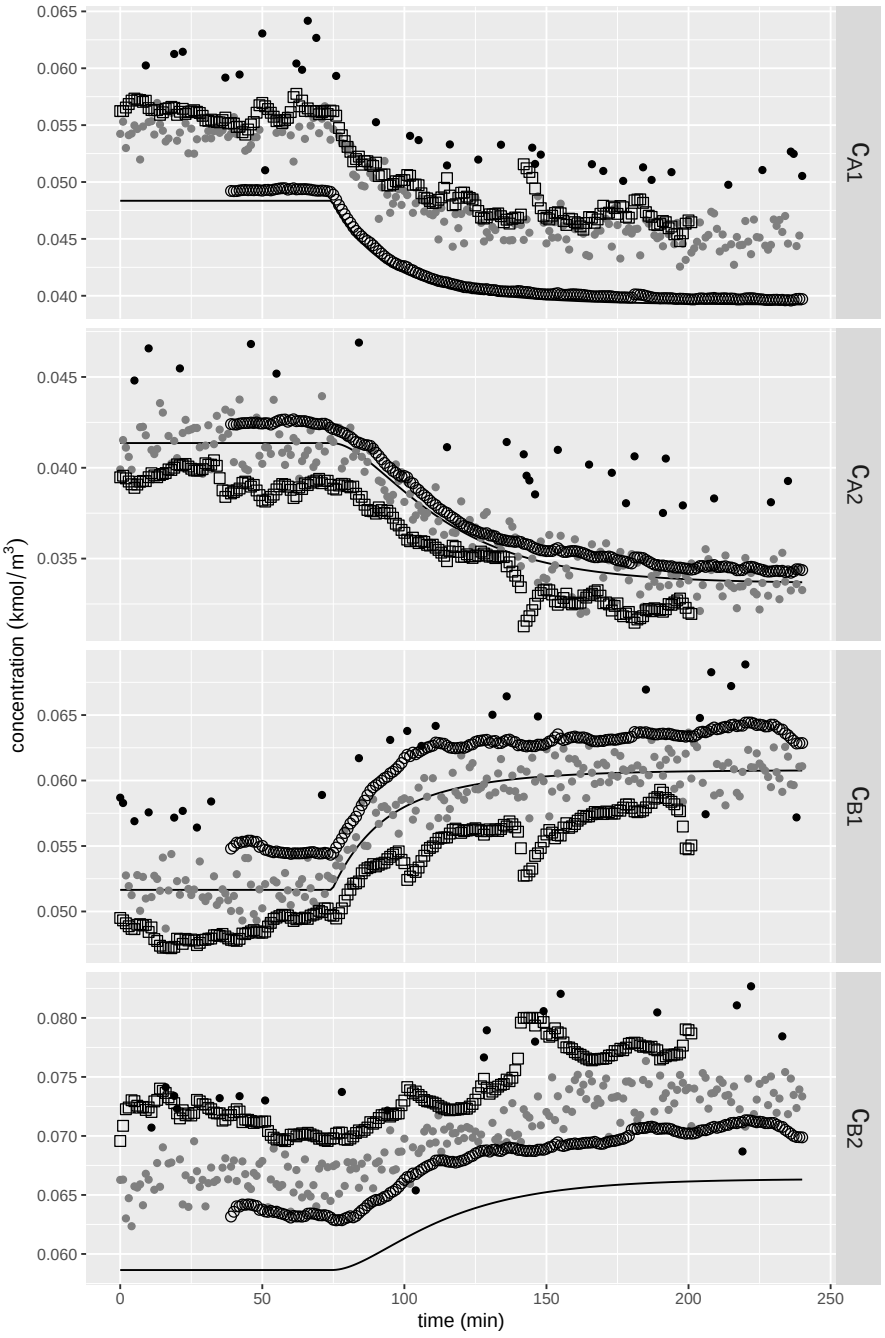


**FIGURE 1** Reconciled results solely using Hampel's redescending $\rho$-function according to equation (22) for window length 40. Filled grey and black circles depict the measured values and the measurement outliers, respectively; the full line refers to the simulated true values; the open squares denote the reconciled values of the variables *at the beginning* of the moving window, while open circles represent those *at the end* of the moving window.

Note that the presence of outliers affects the performance of the test. For example, a single outlier of magnitude greater and opposite sign than those of the bias will increase the probability of a type II error. To cope with this, we disregard the outliers

when evaluating the sign. An observation of variable $i$ at time period $j$ is classified as an outlier whenever

$$\text{outl}_{i,j} = \frac{|r_{i,j} - \mu(R_i)|}{\sigma(R_i)} > c, \quad i = 1, \dots, n, \tag{4}$$

with

$$R_i = \{r_{i,k-N+1}, \dots, r_{i,k}\}, \tag{5}$$

where $\mu$ and $\sigma$ are univariate *robust* location and dispersion or scale estimators, respectively, and the cutoff $c = 2.5$.[16]

Summing up, let $J_i = \{k-N+1 \leq j \leq k : \text{outl}_{i,j} \leq c\}$ be the set of indexes $j$ of inliers within the window. Formally we have

$$r_{i,j} \lesseqgtr 0, \ \forall j \in J_i \Rightarrow i \in I, \quad i = 1, \dots, n. \tag{6}$$

## 2.2 | Loss function

Because a univariate scale estimator is translation invariant, then, according to (1),

$$\sigma(R_i) = \sigma\{\epsilon_{i,j} + O_{i,j}) : j = k-N+1, \dots, k\} \tag{7}$$

for all $i \in I$. This makes it convenient to adopt a robust dispersion measure for the loss function for *biased* measurements, while for the remaining measured variables we instead use a root mean square error based loss function

$$\text{RMSE}(R_i) = \sqrt{\mu(R_i)^2 + \sigma(R_i)^2} \tag{8}$$

for all $i \notin I$.

There remains the problem of combining both of them, which may be addressed using the *desirability function* approach.[17] For each measured variable ($i = 1, \dots, n$), the desirability function $d(v_i)$ assigns a number between 0 (not acceptable) and 1 (completely desirable value) to the possible values of a statistic $v_i$ using

$$d(v_i) = \begin{cases} 1 & \text{if } v_i < v^L \\ \dfrac{v_i - v^U}{v^L - v^U} & \text{if } v^L \leq v_i \leq v^U \\ 0 & \text{if } v_i > t^U \end{cases} \tag{9}$$

with $v^L$ being the smallest possible value and $v^U$ a convenient upper bound. Notice that smaller values of $v_i$ are more desirable. Let $Y_i = \{y_{i,k-N+1}, \dots, y_{i,k}\}$ be the univariate sample for a measured variable $i$. Next we define

$$v_i = \begin{cases} \sigma(R_i)/|\mu(Y_i)| & \text{if } i \in I \\ \text{RMSE}(R_i)/|\mu(Y_i)| & \text{otherwise.} \end{cases} \tag{10}$$

The values of $v^L = 0$ and $v^U = 1$ are seen to work well here. Finally, the overall desirability $D$ is defined as the geometric mean of the individual desirability values

$$D = \left( \prod_{i=1}^{n} d(v_i) \right)^{1/n}, \tag{11}$$

which is precisely the loss function in (2).

We choose the robust univariate location and scale estimators described by Maronna and Zamar[18] because computational efficiency is of key importance. A brief description follows. Define the functions

$$W_c(x) = \begin{cases} \left(1 - \left(\frac{x}{c}\right)^2\right)^2 & \text{if } |x| \leq c \\ 0 & \text{if } |x| > c \end{cases} \tag{12a}$$

and

$$\rho_c(x) = \min(x^2, c^2), \tag{12b}$$

where $W_c$ is Tukey's redescending bisquare weight function and $c > 0$ is a tuning constant. Their estimators are defined as

$$\mu(R_i) = \frac{\sum_j w_j r_{i,j}}{\sum_j w_j}, \tag{13}$$

$$\sigma(R_i)^2 = \frac{\sigma_0^2}{N} \sum_j \rho_{c_2}\left(\frac{r_{i,j} - \mu(R_i)}{\sigma_0}\right), \tag{14}$$

with

$$\sigma_0 = \text{med}(|R_i - \text{med}(R_i)|)$$

and

$$w_j = W_{c_1}\left(\frac{r_{i,j} - \text{med}(R_i)}{\sigma_0}\right),$$

where med stands for median. The authors recommend the tuning parameters $c_1 = 4.5$ and $c_2 = 3$. To conclude, it should be pointed out that minimising the scale (14) (equivalent to maximizing the corresponding desirability) is similar in spirit to other robust estimators based on scale estimators such as those in the regression context (see for example Section 5.4 of Maronna, Martin, Yohai, and Salibián-Barrera[1]).

## 2.3 | Equivalent biases and regularization

Now we turn our attention to an unavoidable problem associated with measurement bias: besides the true set of bias locations and values, several other alternative sets (of the same or different cardinality as the true set) may also lead to the same value of (2) (see Chapter 6 of Bagajewicz[19] and the earlier paper by Bagajewicz and Jiang[20]). The point is that the problem is ill-posed because it is impossible to distinguish one set from another without some additional *a priori* information.

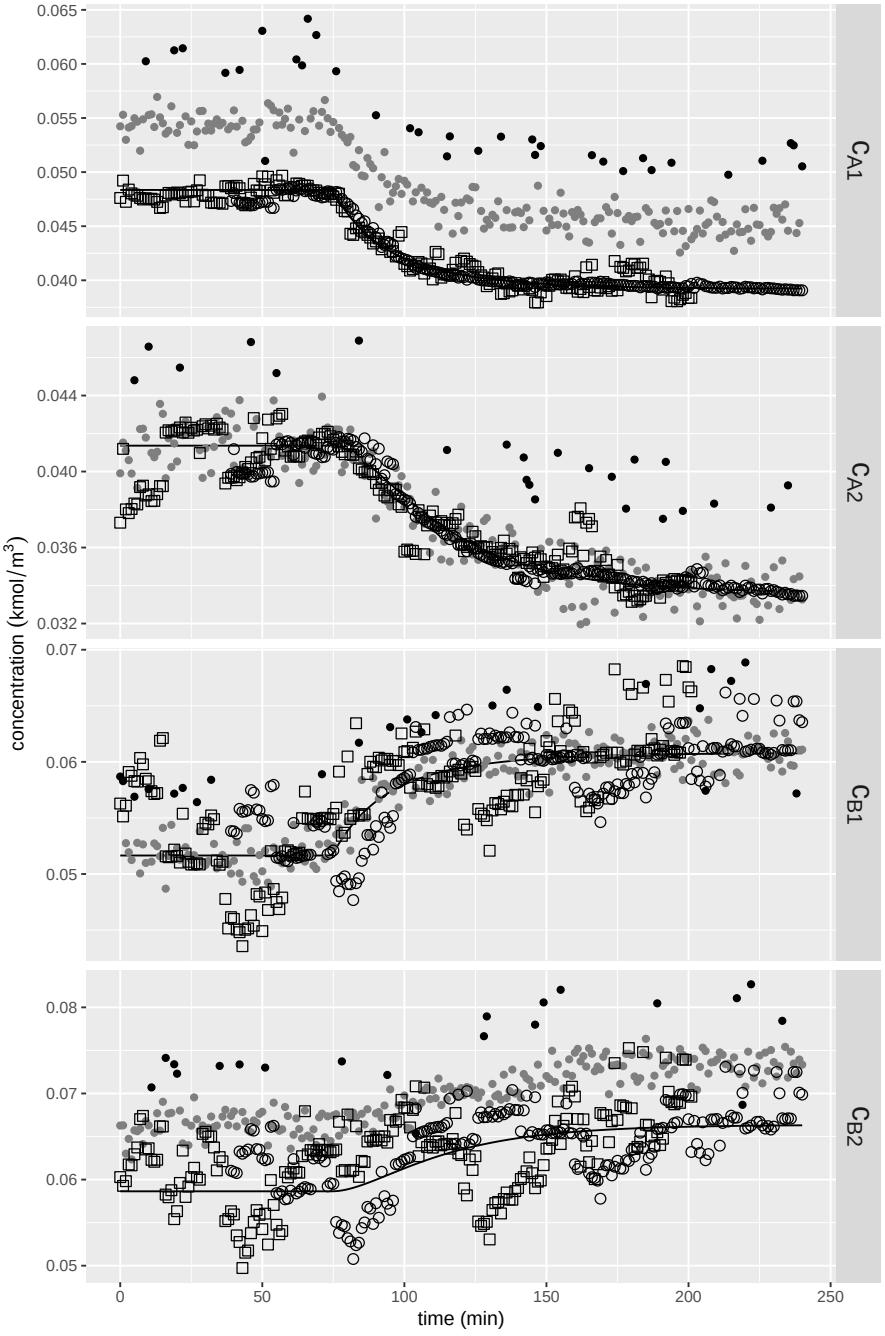


**FIGURE 2** Reconciled results using the new method without regularisation for window length 40. Filled grey and black circles depict the measured values and the measurement outliers, respectively; the full line refers to the simulated true values; the open squares denote the reconciled values of the variables *at the beginning* of the moving window, while open circles represent those *at the end* of the moving window.

To tackle this problem,[21] showed that the use of temporal redundancy accomplishes a better identification of the right set of biases. Another way to mitigate this issue is to impose an upper bound on the number of biases $\delta$, regularizing the optimisation

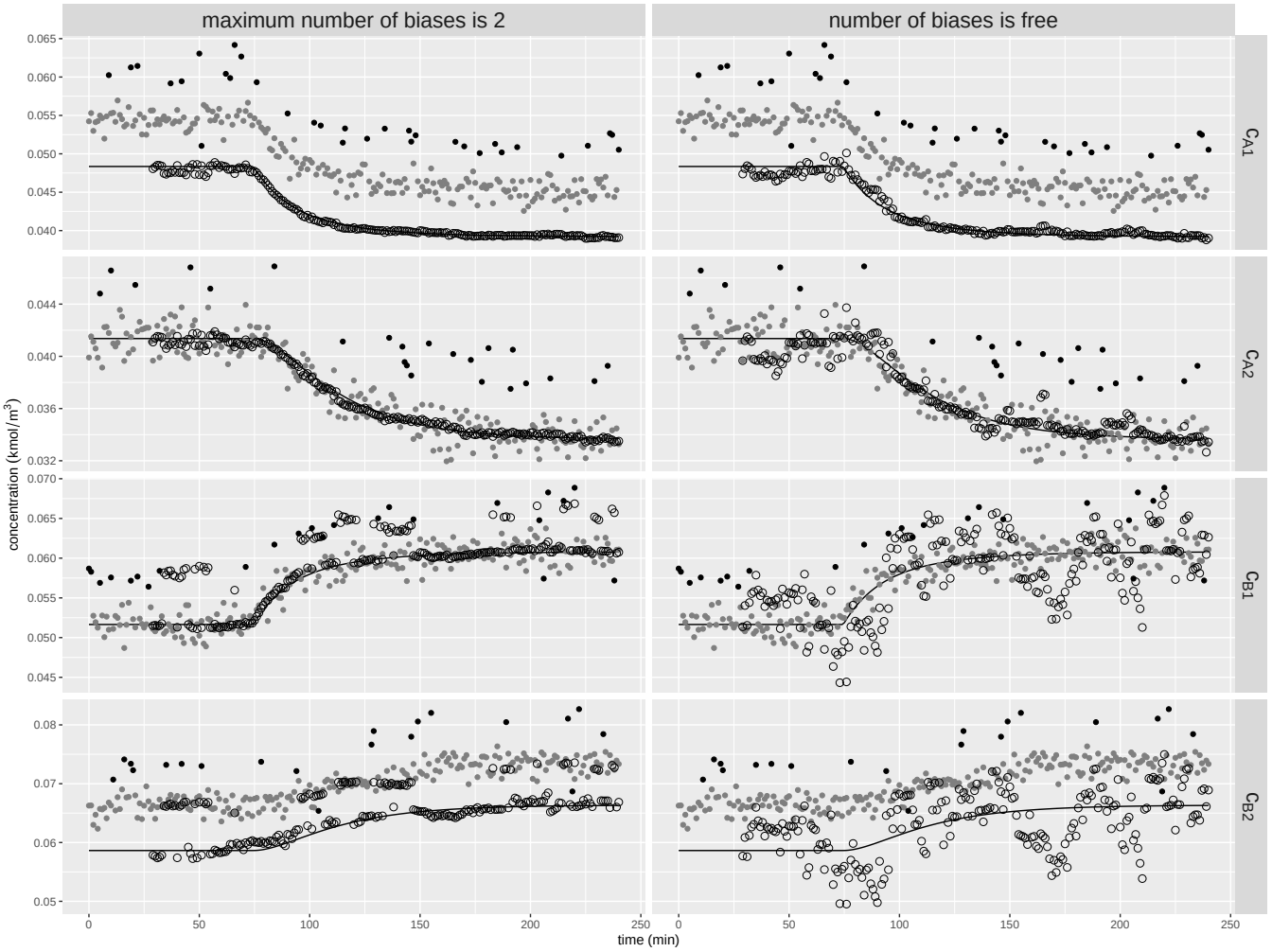


**FIGURE 3** Reconciled results using the new method with (left panels) and without (right panels) regularisation for window length 30. Filled grey and black circles depict the measured values and the measurement outliers, respectively; the full line refers to the simulated true values; the open circles denote the reconciled values of the variables *at the end* of the moving window.

problem in the following form[22]

$$\hat{\boldsymbol{x}}_k = \arg\max_{\boldsymbol{x}_k} D \tag{15a}$$

subject to

$$\#I \leq \delta \tag{15b}$$

where # denotes the cardinal of a set.

## 3 | SOLUTION TECHNIQUE

We start by assuming the case when there are no inequality constraints (3b). Given the complexity of the proposed loss function owing to the logical structure of (10), the nonconvexity of (12a) and (12b), and points $\{-c, c\}$ of non-differentiability in (12b), the estimates $\hat{\boldsymbol{x}}$ are obtained by means of a global optimisation heuristic. (The same type of problem occurs in the desirability function, but this has been overcome in the modified form by Castillo, Montgomery, and McCarville.[23]) More specifically, we adopt with some minor modifications a variant[24] of the original Differential Evolution (DE) algorithm,[25] which we will now succinctly describe.

The DE algorithm evolves a population of $N_\mathrm{p}$ potential solutions $\boldsymbol{P} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{N_\mathrm{p}}\}$ consisting of vectors of real numbers with $p$ elements each $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})$, which are created through a random initialisation.

In each iteration $t$, for each $\boldsymbol{x}_i \in \boldsymbol{P}$ a mutant vector $\boldsymbol{m}_i$ is generated from the individuals in $\boldsymbol{P}$ by means of the DE/current-to-best/1 mutation strategy as

$$\boldsymbol{m}_i = \underbrace{\boldsymbol{x}_i + F(\boldsymbol{x}_\mathrm{best} - \boldsymbol{x}_i)}_{\boldsymbol{b}_i} + F(\boldsymbol{x}_{r_1} - \boldsymbol{x}_{r_2}), \tag{16}$$

with

$$r_1, r_2 \in \{1, 2, \dots, N_\mathrm{p}\}, \quad r_2 \neq r_1 \neq i,$$

where $r_1$ and $r_2$ are randomly chosen population indices, $\boldsymbol{x}_\mathrm{best}$ is the solution with the best objective function value in the current population $\boldsymbol{P}$, and $F$ is a control parameter called the *scale factor*.

Then, each element of the mutant vector $\boldsymbol{m}_i$ is randomly exchanged with the corresponding element of the parent $\boldsymbol{x}_i$ and the trial vector $\boldsymbol{o}_i$ is generated

$$o_{i,j} = m_{i,j} \quad \text{for} \quad j = j_\mathrm{rand}, \tag{17a}$$

and

$$o_{i,j} = \begin{cases} m_{i,j} & \text{if } U(0,1) \leq C_\mathrm{r} \\ x_{i,j} & \text{otherwise} \end{cases} \qquad \forall j \in \{1, 2, \dots, p\} \setminus j_\mathrm{rand}, \tag{17b}$$

where $j_\mathrm{rand}$ is a discrete uniform random variable over $\{1, 2, \dots, p\}$, $U(0, 1)$ is a uniform random number within the range $[0, 1]$, and the so-called *crossover rate* control parameter $C_\mathrm{r}$ is obtained for each individual $i = 1, 2, \dots N_\mathrm{p}$, and for every iteration $t$ as a realisation of a normally distributed variable $N(0.5, 0.15^2)$.[26] Condition (17a) avoids the trivial case where $\boldsymbol{o}_i = \boldsymbol{x}_i$. This is known as binomial (bin) crossover.

The resulting trial vector $\boldsymbol{o}_i$ is compared one-to-one with its respective parent $\boldsymbol{x}_i$, keeping the better of them for the next iteration

$$\boldsymbol{x}_i^{t+1} = \begin{cases} \boldsymbol{o}_i^t & \text{if } D(\boldsymbol{o}_i^t) > D(\boldsymbol{x}_i^t) \\ \boldsymbol{x}_i^t & \text{otherwise.} \end{cases} \tag{18}$$

Selection is implemented by immediately replacing a losing parent vector $\boldsymbol{x}_i^t$ in $\boldsymbol{P}$.[27]

Lee, Han, and Chang[24] added what can be seen as a simple local search to the scale factor $F$ during the generation of every new trial vector $\boldsymbol{o}_i$. The algorithm, given a value $F_\mathrm{b}$ of the scale factor as a starting point, is thus

1. Set $F \leftarrow F_\mathrm{b}$.
2. For $k = 0, 1, 2, \ldots$ :
   (a) Compute $\boldsymbol{m}_{i,k}$ in (16) and then $\boldsymbol{o}_{i,k}$ in (17a)-(17b).
   (b) Update $F \leftarrow F\xi$.
3. Stop when $D(\boldsymbol{o}_{i,k+1}) \leq D(\boldsymbol{o}_{i,k})$.
4. Put $\boldsymbol{o}_i^t \leftarrow \boldsymbol{o}_{i,k}$ in (18).

Our experiments showed that the choice $F_\mathrm{b} = 0.2$ performed well. Instead of using a fixed value for the step factor $\xi$, a uniformly selected random real number from $[1.3, 1.7]$ is generated for each individual $i = 1, \ldots N_\mathrm{p}$, and for every iteration $t$. Our experiments showed that the choice $F_\mathrm{b} = 0.2$ performed well.

Bound constraints (3c) are enforced explicitly by modifying only those coordinate values which violate bounds:[28]

$$b_{i,j} = \begin{cases} b_{i,j} & \text{if } x_j^\mathrm{L} \leq b_{i,j} \leq x_j^\mathrm{U} \\ (x_j^\mathrm{L} + x_{i,j})/2 & \text{if } b_{i,j} < x_j^\mathrm{L} \\ (x_{i,j} + x_j^\mathrm{U})/2 & \text{if } b_{i,j} > x_j^\mathrm{U} \end{cases} \tag{19a}$$

and

$$o_{i,j} = \begin{cases} o_{i,j} & \text{if } x_j^\mathrm{L} \leq o_{i,j} \leq x_j^\mathrm{U} \\ (x_j^\mathrm{L} + b_{i,j})/2 & \text{if } o_{i,j} < x_j^\mathrm{L} \\ (b_{i,j} + x_j^\mathrm{U})/2 & \text{if } o_{i,j} > x_j^\mathrm{U} \end{cases} \tag{19b}$$

for all $j = 1, 2, \ldots, p$.

We now tackle the problem of handling the regularization constraint $\#\boldsymbol{I}(\boldsymbol{x}) \leq \delta$. This is done by applying the *superiority of feasible solutions* approach proposed by Deb,[29] which modifies the selection rule (18). More precisely, the trial vector $\boldsymbol{o}_i^t$ will be selected if

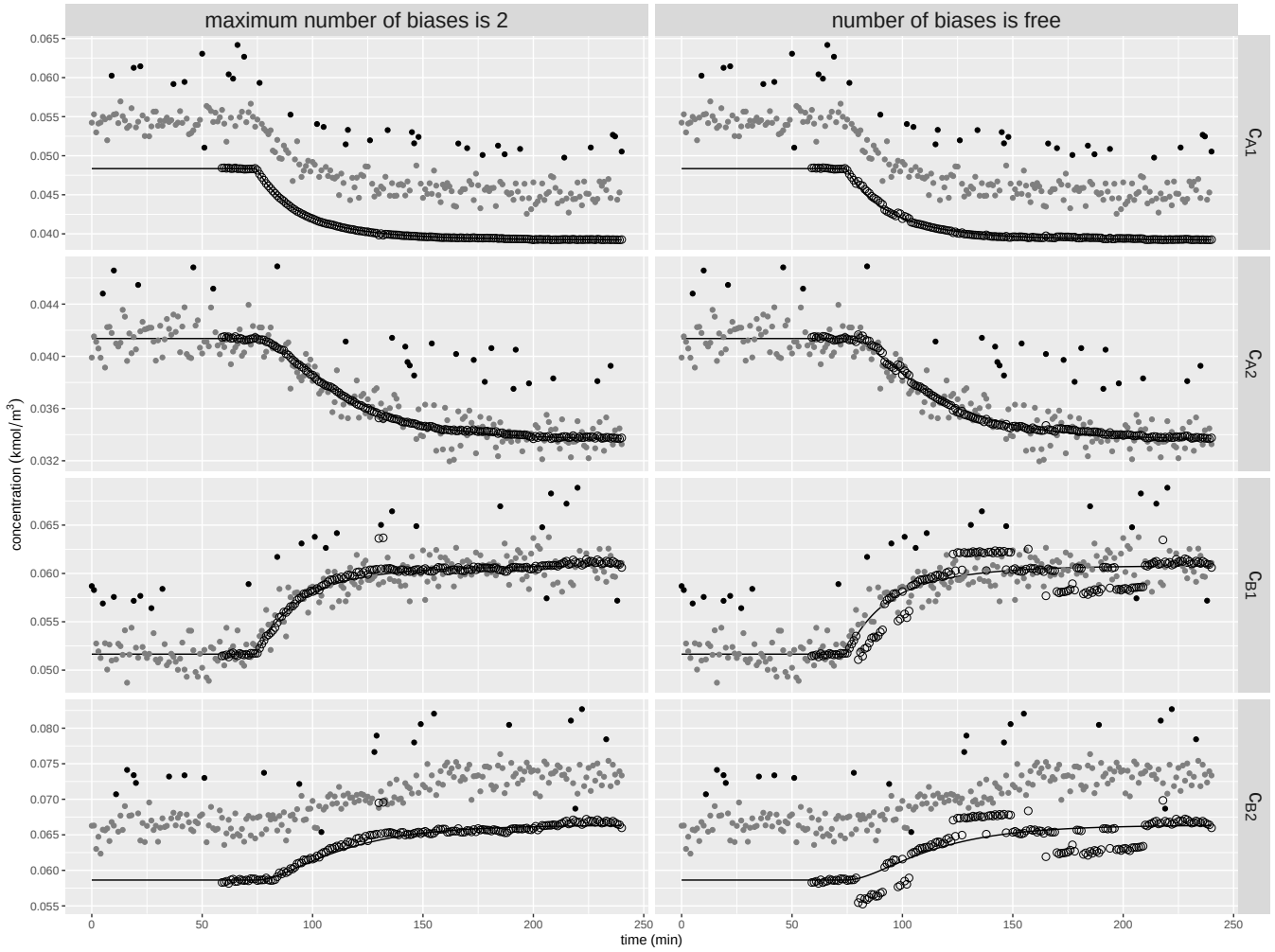


**FIGURE 4** Reconciled results using the new method with (left panels) and without (right panels) regularisation for window length 60. Filled grey and black circles depict the measured values and the measurement outliers, respectively; the full line refers to the simulated true values; the open circles denote the reconciled values of the variables *at the end* of the moving window.

- given two feasible solutions $\boldsymbol{o}_i^t$ and $\boldsymbol{x}_i^t$, it provides the better value for $D$

$$\#I(\boldsymbol{o}_i^t) \leq \delta \wedge \#I(\boldsymbol{x}_i^t) \leq \delta \wedge D(\boldsymbol{o}_i^t) > D(\boldsymbol{x}_i^t), \quad \text{or}$$

- it is a feasible solution and $\boldsymbol{x}_i^t$ is infeasible

$$\#I(\boldsymbol{o}_i^t) \leq \delta \wedge \#I(\boldsymbol{x}_i^t) > \delta, \quad \text{or}$$

- both $\boldsymbol{o}_i^t$ and $\boldsymbol{x}_i^t$ are infeasible, but it presents a lower or equal constraint violation

$$\#I(\boldsymbol{o}_i^t) > \delta \wedge \#I(\boldsymbol{x}_i^t) > \delta \wedge \#I(\boldsymbol{o}_i^t) \leq \#I(\boldsymbol{x}_i^t).$$

## 4 | ILLUSTRATIVE EXAMPLE

To illustrate the use of the proposed method, we apply it to the case of an reversible exothermic autocatalytic reaction between species A and B

$$\mathrm{A} + \mathrm{B} \underset{k^{\mathrm{r}}}{\overset{k^{\mathrm{f}}}{\rightleftharpoons}} 2\,\mathrm{B} \tag{20}$$

that takes place in two isothermal continuous stirred tank reactors arranged in series where 50 % of the product stream is recycled back into the first reactor as described by Abu-el-zeet, Becerra, and Roberts.[12]

The model is given by the following differential equations

$$\frac{\mathrm{d}c_{A1}}{\mathrm{d}t} = \frac{0.5}{\tau_1}(c_{A0} + c_{A2}) - \frac{c_{A1}}{\tau_1} - (k_1^{\mathrm{f}} c_{A1} c_{B1} - k_1^{\mathrm{r}} c_{B1}^2) \tag{21a}$$

$$\frac{\mathrm{d}c_{B1}}{\mathrm{d}t} = \frac{0.5}{\tau_1} c_{B2} - \frac{c_{B1}}{\tau_1} + (k_1^{\mathrm{f}} c_{A1} c_{B1} - k_1^{\mathrm{r}} c_{B1}^2) \tag{21b}$$

$$\frac{\mathrm{d}c_{A2}}{\mathrm{d}t} = \frac{c_{A1}}{\tau_2} - \frac{c_{A2}}{\tau_2} - (k_2^{\mathrm{f}} c_{A2} c_{B2} - k_2^{\mathrm{r}} c_{B2}^2) \tag{21c}$$

$$\frac{\mathrm{d}c_{B2}}{\mathrm{d}t} = \frac{c_{B1}}{\tau_2} - \frac{c_{B2}}{\tau_2} + (k_2^{\mathrm{f}} c_{A2} c_{B2} - k_2^{\mathrm{r}} c_{B2}^2) \tag{21d}$$

with

$$k = A \exp\left(-\frac{E_{\mathrm{a}}}{RT}\right),$$

where the subscripts 0, 1, and 2 refer to the feed stream, exit stream from the first reactor, and product stream, respectively, the superscripts f and r refer to the forward and reverse reaction, respectively, $c$ denotes concentration ($\mathrm{kmol\,m^{-3}}$), $\tau$ represents residence time, $k$ and $A$ are the rate constant and pre-exponential factor ($\mathrm{m^3\,kmol^{-1}\,s^{-1}}$), $E_{\mathrm{a}}$ is the activation energy, $R$ is the gas constant, and $T$ denotes temperature (K).

For all the simulation cases the following values were used

$$\tau_1 = 30\,\mathrm{min} \qquad \tau_2 = 25\,\mathrm{min}$$

$$E_{\mathrm{a}}^{\mathrm{f}}/R = 17\,786\,\mathrm{K} \qquad E_{\mathrm{a}}^{\mathrm{r}}/R = 23\,523\,\mathrm{K}$$

$$A^{\mathrm{f}} = 9.73 \times 10^{22}\,\frac{\mathrm{m^3}}{\mathrm{kmol\,s}} \qquad A^{\mathrm{r}} = 3.1 \times 10^{30}\,\frac{\mathrm{m^3}}{\mathrm{kmol\,s}}$$

$$T_1 = 307\,\mathrm{K} \qquad T_2 = 302\,\mathrm{K}$$

$$c_{A0} = 0.1\,\mathrm{kmol\,m^{-3}}.$$

Starting from steady-state operation, transients are introduced by a step change in the first reactor temperature $T_1$ from 307 K to 310 K at $t = 75$ min.

## 4.1 | Implementation and simulation settings

All the simulations reported here have been done with the free software `R`.[30] The model differential equations are integrated numerically using the `LSODA`[31] solver available through the package `deSolve`.[32] The package `robustbase`[33] contains `scaleTau2`, which is an implementation for the univariate estimators (13) and (14).

The population size used in DE was $N_\mathrm{p} = 40$, and the algorithm was stopped after 400 iterations.

We experimented with the moving window lengths $N = 30\,\mathrm{min}$, $40\,\mathrm{min}$, $50\,\mathrm{min}$ and $60\,\mathrm{min}$.

Measurement noise $\epsilon_{i,j}$ is simulated by sampling from a Gaussian distribution using the package `mvtnorm`[34,35] with standard deviation $\sigma_{\epsilon,i}$ 2.5 % of the nominal value of the outlet concentration at steady state

$$c_{A1} = 0.048\,35\,\mathrm{kmol\,m^{-3}} \qquad c_{A2} = 0.041\,36\,\mathrm{kmol\,m^{-3}}$$

$$c_{B1} = 0.051\,65\,\mathrm{kmol\,m^{-3}} \qquad c_{B2} = 0.058\,64\,\mathrm{kmol\,m^{-3}}.$$

The magnitude and sign of the biases is fixed at $B_i = K^\mathrm{b}\sigma_{\epsilon,i}$ with the constant $K^\mathrm{b} = 5$ for both $c_{A1}$ and $c_{B2}$. One-sided outliers are generated for all exit concentrations with probability 0.1 and magnitude and sign $O_{i,j} = K^\mathrm{o}\sigma_{\epsilon,i}$ with the constant $K^\mathrm{o}$ being 5 (outliers on the same side as bias with respect to the true values) or $-7$ (outliers and bias lie on opposite sides of the true values). Also, we should not forget that the Gaussian distribution may generate outliers by itself. We will consider any point satisfying $|(y_{i,j} - x_{i,j})|/\sigma_{\epsilon,i} \geq 2.5$ as an outlier.

Lastly, the search space is defined by the interval vector $(0.03, 0.03, 0.04, 0.04) \leq (c_{A1}, c_{A2}, c_{B1}, c_{B2})/\mathrm{kmol\,m^{-3}} \leq (0.06, 0.06, 0.08, 0.08)$ for $K^\mathrm{o} = 5$, with the lower bound of $c_{A2}$ changing to $0.025\,\mathrm{kmol\,m^{-3}}$ when using $K^\mathrm{o} = -7$.

# 5 | A COUNTEREXAMPLE

We give a specific counter-example for the redescending estimator of Hampel with tuning parameters $a = 1$, $b = 2$, $c = 6$ as in Nicholson, López-Negrete, and Biegler[5]

$$\rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } 0 \leq |x| \leq a \\ a|x| - \dfrac{a^2}{2} & \text{if } a < |x| \leq b \\ ab - \dfrac{a^2}{2} + \dfrac{a(c-b)}{2}\left[1 - \left(\dfrac{c-|x|}{c-b}\right)^2\right] & \text{if } b < |x| \leq c \\ ab - \dfrac{a^2}{2} + \dfrac{a(c-b)}{2} & \text{if } |x| > c \end{cases}$$

showing the inability of robust loss functions to cope with measurement bias. (According to these authors, tuning of the estimator parameters is not critical to obtain good performance.)

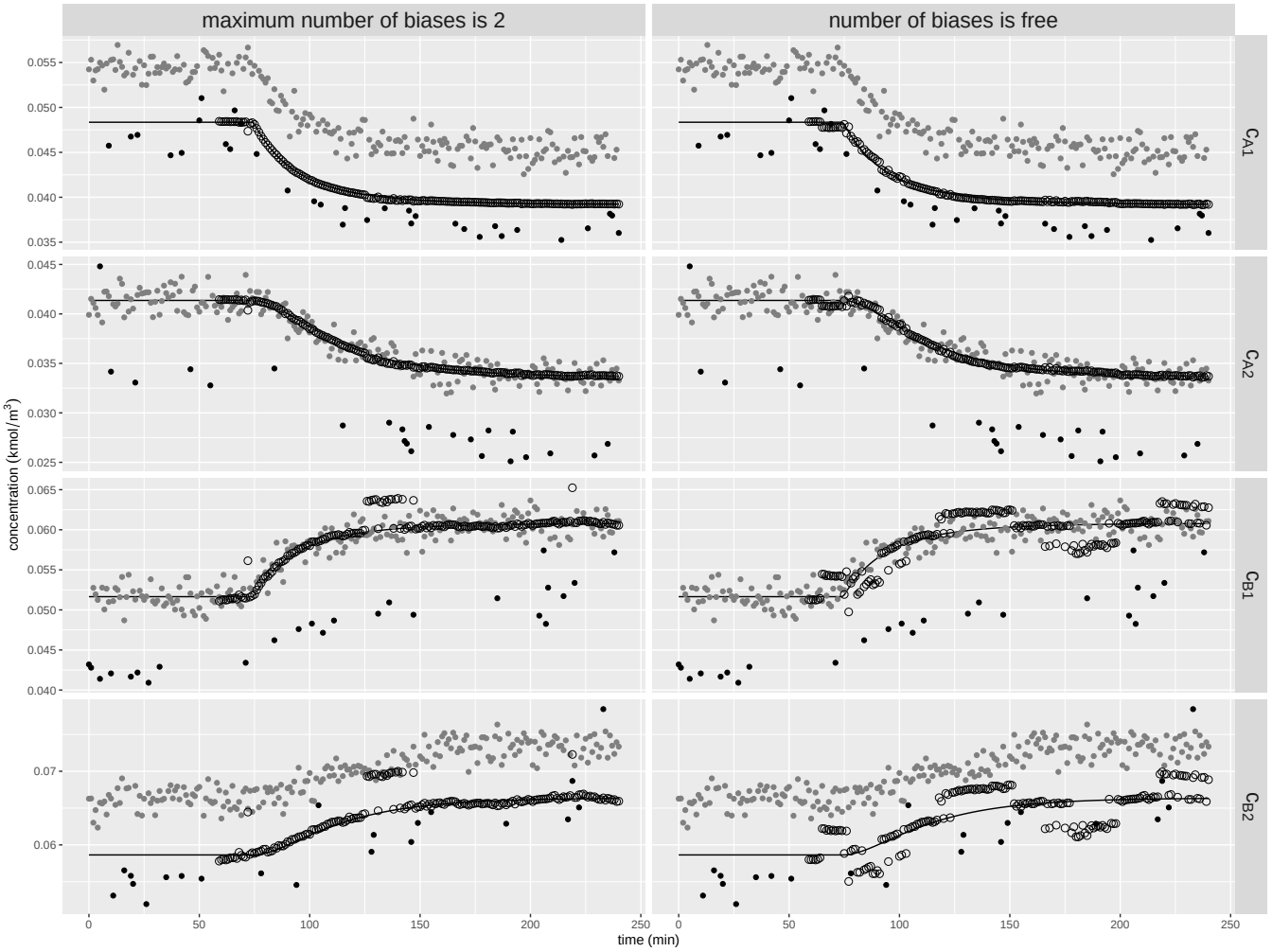


**FIGURE 5** Reconciled results using the new method with (left panels) and without (right panels) regularisation for window length 60 and in the case of bias and outliers located on opposite sides of the true values. Filled grey and black circles depict the measured values and the measurement outliers, respectively; the full line refers to the simulated true values; the open circles denote the reconciled values of the variables *at the end* of the moving window.

In this case, equation (2) becomes

$$\hat{\boldsymbol{x}}_k = \arg\min_{\boldsymbol{x}_k} \sum_{j=k-N+1}^{k} \sum_{i=1}^{n} \rho\left(\frac{y_{i,j} - x_{i,j}}{\sigma_{\epsilon,i}}\right), \tag{22}$$

and we fixed the denominator $\sigma_{\epsilon,i}$ at its true value to isolate the effect of the $\rho$-function.

The simulation results are shown in Figure 1 for a horizon size of $N = 40$ when outliers and biases have the same (positive) sign. For other sizes, namely 30, 50 and 60, we get rather similar results. Two behaviours can be observed: one for species A and one for species B. For species A, the reconciled values *at the end* of the moving window are quite close to the true values meaning that the bias in $c_{A1}$ was correctly identified. That is not the case with the reconciled values *at the beginning* of the moving window, which are shifted vertically with respect to those at the end. For species B, the reconciled values follow the lower and upper boundaries of the main bulk of the measurements, ignoring completely the existence of bias in $c_{B2}$.

The preceding results cast considerable doubt on the suitability of (22) for coping with measurement bias.

## 6 | RESULTS OF THE NUMERICAL EXPERIMENTS

First let us look at the behaviour of the new method in the least favourable condition *without regularisation* (Figure 2) when applied to the same case used in the counterexample. Compared to it, the distinction between the reconciled values at the beginning of the moving window and those at the end now disappears. Moreover, the bias in $c_{B2}$ is now detected. This is exactly what should be expected on general grounds. Finally, we observe a large difference in precision results between species A (high) and B (low). Indeed, when regularisation is not used, some of the reconciled values for $c_{B1}$ are so far away from the bulk of the data as to cause an incorrect classification as a biased point.

To examine the effect of regularisation, the true value of $\delta = 2$ was chosen for the upper bound on the number of biased measurements which is shown in the left panels of Figure 3, while leaving $\delta$ unbounded in the right panels for a window length of $N = 30$. Two important results are worth noticing when using regularisation. First, when the "right" $\delta$ is given, it leads to solutions of high quality, especially for species B (which exhibits high precision, contrary to the case without regularisation). Second, it reveals that there are two equivalent sets of biased measurements: $\{c_{A1}, c_{B2}\}$ and $\{c_{A1}, c_{B1}\}$. (It is important to recall that the criterion for the detection of measurement bias by definition ignores outliers.) The optimisation algorithm cycles through the two in a seemingly random manner, but note that $\{c_{A1}, c_{B2}\}$ is preponderant over $\{c_{A1}, c_{B1}\}$.

Larger window sizes $N$ improve monotonically the performance of the proposed approach, as can be seen in Figure 4 for $N = 60$. All but two points will correctly identify the measurement bias in $c_{B2}$ with regularisation (left panels), whereas the precision of the reconciled measurements for species B is considerably improved when not using regularisation (right panels).

Finally, we consider an edge case in which the biased measurements and outliers are located on opposite sides of the simulated true values. It is seen from Figure 5 ($N = 60$) that the proposed method is also able to cope with this type of problem.

## 7 | CONCLUDING REMARKS

Herein we solely focused on an initial characterisation of the behaviour of this method to assess if further development is of interest. One important and attractive feature is the fact that it does not require *a priori* knowledge on the variance of the measurement error.

Our approach was tested using two challenging scenarios with small biases and small outliers when compared to other works, with very encouraging preliminary results. This leads us to believe that it provides a sound basis for dynamic data reconciliation subject to measurement bias.

Besides the choice of the horizon size $N$ (long horizons are desirable), we would argue that there are three kinds of challenges to accomplish a practical implementation as follows: (1) one needs a method for selecting the exact number of biased measured

variables (regularization parameter),[22,36] (2) further, the ability to simultaneously locate multiple optima (multimodal optimisation) such as the one offered by evolutionary algorithms,[37] (3) and finally, to figure out how to integrate process-specific knowledge into a hybrid model.[38] Of course, the appropriate choice should be balanced by taking into account some prespecified computational budget.

## Literature Cited


1. Maronna RA, Martin RD, Yohai VJ, Salibián-Barrera M. *Robust Statistics: Theory and Methods (with R)*. Hoboken, NJ: Wiley. second ed. 2019.

2. Medina MA, Ronchetti E. Robust statistics: a selective overview and new directions. *Wiley Interdisciplinary Reviews: Computational Statistics* 2015; 7(6): 372-393. doi: 10.1002/wics.1363

3. Llanos CE, Sánchez MC, Maronna RA. Robust estimators for data reconciliation. *Industrial & Engineering Chemistry Research* 2015; 54(18): 5096-5105. doi: 10.1021/ie504735a

4. Prata DM, Pinto JC, Lima EL. Comparative analysis of robust estimators on nonlinear dynamic data reconciliation. In: Braunschweig B, Joulia X., eds. *18th European Symposium on Computer Aided Process Engineering*. 25 of *Computer Aided Chemical Engineering*. Elsevier B.V.; 2008; Amsterdam: 501-506

5. Nicholson B, López-Negrete R, Biegler LT. On-line state estimation of nonlinear dynamic systems with gross errors. *Computers & Chemical Engineering* 2014; 70: 149-159. doi: 10.1016/j.compchemeng.2013.11.018

6. Zhang Z, Chen J. Correntropy based data reconciliation and gross error detection for nonlinear dynamic processes. *Computers & Chemical Engineering* 2015; 75: 120-134. doi: 10.1016/j.compchemeng.2015.01.005

7. Valluru J, Patwardhan SC, Biegler LT. Development of robust extended Kalman filter and moving window estimator for simultaneous state and parameter/disturbance estimation. *Journal of Process Control* 2018; 69: 158-178. doi: 10.1016/j.jprocont.2018.05.008

8. Rangegowda PH, Patwardhan SC, Biegler LT, Mukhopadhyay S. Simultaneous state and parameter estimation using robust receding-horizon nonlinear Kalman filter. *IFAC-PapersOnLine* 2019; 52(1): 10-15. doi: 10.1016/j.ifacol.2019.06.030

9. Albuquerque JS, Biegler LT. Data reconciliation and gross-error detection for dynamic systems. *AIChE Journal* 1996; 42(10): 2841-2856. doi: 10.1002/aic.690421014

10. Arora N, Biegler LT. Redescending estimators for data reconciliation and parameter estimation. *Computers & Chemical Engineering* 2001; 25(11-12): 1585-1599. doi: 10.1016/S0098-1354(01)00721-9

11. Prata DM, Schwaab M, Lima EL, Pinto JC. Simultaneous robust data reconciliation and gross error detection through particle swarm optimization for an industrial polypropylene reactor. *Chemical Engineering Science* 2010; 65(17): 4943-4954. doi: 10.1016/j.ces.2010.05.017

12. Abu-el-zeet ZH, Becerra VM, Roberts PD. Combined bias and outlier identification in dynamic data reconciliation. *Computers & Chemical Engineering* 2002; 26(6): 921-935. doi: 10.1016/S0098-1354(02)00018-2

13. Soderstrom TA, Himmelblau DM, Edgar TF. A mixed integer optimization approach for simultaneous data reconciliation and identification of measurement bias. *Control Engineering Practice* 2001; 9(8): 869-876. doi: 10.1016/S0967-0661(01)00056-9

14. Llanos CE, Sánchez MC, Maronna RA. Classification of systematic measurement errors within the framework of robust data reconciliation. *Industrial & Engineering Chemistry Research* 2017; 56(34): 9617-9628. doi: 10.1021/acs.iecr.7b00726

15. Biegler LT. Nonlinear programming strategies for dynamic chemical process optimization. *Theoretical Foundations of Chemical Engineering* 2014; 48(5): 541-554. doi: 10.1134/S0040579514050157

16. Rousseeuw PJ, Hubert M. Anomaly detection by robust statistics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2018; 8(2): e1236. doi: 10.1002/widm.1236

17. Derringer G, Suich R. Simultaneous optimization of several response variables. *Journal of Quality Technology* 1980; 12(4): 214-219. doi: 10.1080/00224065.1980.11980968

18. Maronna RA, Zamar RH. Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics* 2002; 44(4): 307-317. doi: 10.1198/004017002188618509

19. Bagajewicz MJ. *Smart Process Plants: Software and Hardware Solutions for Accurate Data and Profitable Operations*. New York: McGraw-Hill . 2010.

20. Bagajewicz MJ, Jiang Q. Gross error modeling and detection in plant linear dynamic reconciliation. *Computers & Chemical Engineering* 1998; 22(12): 1789-1809. doi: 10.1016/S0098-1354(98)00248-8

21. Llanos CE, Sánchez MC, Maronna RA. Robust estimation of nonredundant measurements and equivalent sets of observations. *Industrial & Engineering Chemistry Research* 2019; 58(42): 19551-19561. doi: 10.1021/acs.iecr.9b03040

22. Aster RC, Martin BB, Thuerber CH. *Parameter Estimation and Inverse Problems*. Elsevier. third ed. 2019.

23. Del Castillo H, Montgomery DC, McCarville DR. Modified desirability functions for multiple response optimization. *Journal of Quality Technology* 1996; 28(3): 337-345. doi: 10.1080/00224065.1996.11979684

24. Lee MH, Han C, Chang KS. Dynamic optimization of a continuous polymer reactor using a modified differential evolution algorithm. *Industrial & Engineering Chemistry Research* 1999; 38(12): 4825-4831. doi: 10.1021/ie980373x

25. Storn R, Price K. Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 1997; 11(4): 341-359. doi: 10.1023/A:1008202821328

26. Salman A, Engelbrecht AP, Omran MGH. Empirical analysis of self-adaptive differential evolution. *European Journal of Operational Research* 2007; 183(2): 785-804. doi: 10.1016/j.ejor.2006.10.020

27. Babu BV, Angira R. Modified differential evolution (MDE) for optimization of non-linear chemical processes. *Computers & Chemical Engineering* 2006; 30(6-7): 989-1002. doi: 10.1016/j.compchemeng.2005.12.020

28. Biedrzycki R, Arabas J, Jagodziński D. Bound constraints handling in Differential Evolution: An experimental study. *Swarm and Evolutionary Computation* 2019; 50: 100453. doi: 10.1016/j.swevo.2018.10.004

29. Deb K. An efficient constraint handling method for genetic algorithms. *Computer Methods in Applied Mechanics and Engineering* 2000; 186(2-4): 311-338. doi: 10.1016/S0045-7825(99)00389-8

30. R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2020.

31. Hindmarsh AC. ODEPACK, a systematized collection of ODE solvers. In: Stepleman RS., ed. *Scientific Computing*. 1 of *IMACS Transactions on Scientific Computation*. North-Holland; 1983; Amsterdam: 55-64.

32. Soetaert K, Petzoldt T, Setzer RW. Solving differential equations in R: Package deSolve. *Journal of Statistical Software* 2010; 33(9): 1-25. doi: 10.18637/jss.v033.i09

33. Maechler M, Rousseeuw P, Croux C, et al. *robustbase: Basic Robust Statistics*. 2019. R package version 0.93-5.

34. Genz A, Bretz F, Miwa T, et al. *mvtnorm: Multivariate Normal and t Distributions*. 2020. R package version 1.0-12.

35. Genz A, Bretz F. *Computation of Multivariate Normal and t Probabilities*. 195 of *Lecture Notes in Statistics*. Berlin: Springer-Verlag . 2009.

36. Vogel CR. *Computational Methods for Inverse Problems*. Philadelphia, PA: SIAM . 2002.

37. Das S, Maity S, Qu BY, Suganthan PN. Real-parameter evolutionary multimodal optimization — A survey of the state-of-the-art. *Swarm and Evolutionary Computation* 2011; 1(2): 71-88. doi: 10.1016/j.swevo.2011.05.005

38. Venkatasubramanian V. The promise of artificial intelligence in chemical engineering: Is it here, finally?. *AIChE Journal* 2019; 65(2): 466-478. doi: 10.1002/aic.16489