Outlier Detection for Polynomial Systems Using Semidefinite Relaxations

Steffen Borchers * **Rolf Findeisen** *

* *Institute for Automation Engineering, Otto-von-Guericke Universität Magdeburg, Germany*

Abstract: Outlier detection and analysis is a primary step in modelling towards obtaining unbiased estimates, model validation, and coherent analysis, because outliers may contain valuable information or lead to falsely rejecting hypotheses. In this work, we describe approaches for detecting outliers in measurements due to time-dependent and possibly non-homogeneously distributed measurement uncertainties within a set-membership setting. We propose a combinatorial outlier detection approach based on a rigorous invalidity criterion using semidefinite programming relaxations. To overcome combinatorial complexity issues, we furthermore propose a reachability-based approach to identify outlier candidates, which can be easily verified thereafter.

1. INTRODUCTION

Outlier analysis deals with the problem of detecting, and if appropriate removing, anomalous observations in data, and is a primary step towards obtaining estimates and coherent analysis [Ben-Gal, 2005]. As pointed out in Ben-Gal [2005] outliers may carry valuable information, although they may conversely lead to model misspecification or biased parameter estimates and are therefore important to identify prior to modelling and analysis [Williams et al., 2002, Liu et al., 2004].

Outliers often arise due to faults, changes in system behaviour, fraudulent behaviour, human error, instrument error or simply through natural deviations in populations [Hodge and Austin, 2004]. By now, many ways to detect and treat outliers depending on the specific applications exist. Even so an exact definition of an outlier often depends on hidden assumptions regarding the data structure and the applied detection method [Ben-Gal, 2005], some general definitions have been proposed. Hawkins [1980] defines an outlier as "an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism", and Barnett and Lewis [1994] stated that "an outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs". Furthermore, Johnson and Wichern [1992] defines an outlier as "an observation in a data set which appears to be inconsistent with the remainder of that set of data". For an in depth review, see e.g. Ben-Gal [2005], Chandola et al. [2009], for online detection see Liu et al. [2004].

Existing outlier detection methods can be classified according to whether an (error) model is utilized or not, i.e. parametric (supervised or semi-supervised classification) and non-parametric (non-supervised classification) respectively, see e.g. Hodge and Austin [2004] and Ben-Gal [2005] for a comprehensive survey. While non-parametric methods typically deal with large data sets and independent data, parametric outlier detection methods are commonly used for detecting outliers in time-series data (dependent data). Parametric outlier detection methods can be further classified into model-specific or model generic approaches [Ben-Gal et al., 2003]. While model-specific approaches rely on on a given (dynamical) model to perform the outlier analysis, e.g. cumulative sum or moving average filters, model-generic approaches provide the flexibility to estimate the underlying model parameters.

In this work, we consider outlier detection strategies for time-series data considering polynomial systems with possibly uncertain parameters. The methods are based on the set-based falsification, estimation, and analysis framework outlined in detail in Borchers et al. [2009], Rumschinski et al. [2010], and implemented within a MATLAB toolbox in Streif et al. [2012]. To this end, we first devise a combinatorial outlier detection approach to evaluate outlier hypotheses using infeasibility certificates. We discuss modifications of the approach to overcome the combinatorial complexity of the proposed approach. Furthermore, we consider a complementary outlier detection procedure based on reachability analysis.

2. PRELIMINARIES

Before we consider the outlier detection problem, we define the non-linear systems we consider, describe the available data, and summarize the results on model invalidation and estimation within the set-membership setting used.

Model and data uncertainty description

We consider polynomial discrete-time systems for $0 \leqslant k \leqslant N-1$ of the form

$$x_i[k+1] = f_i(x[k], p) \; i \in \{1, \ldots, n_x\}. \tag{1}$$

Remark 1. Note that it is possible to consider inputs, polynomial output equations, as well as implicit discrete-time systems, though for simplicity of presentation omitted here.

If the initial conditions and the model parameters are known precisely, such a model allows us to make predictions about the outcome of an experiment by numerical simulation.

If however the parameters and the initial conditions are unknown, they have to be identified from experimental data beforehand. This defines the estimation problem,

i.e. to convert observations into information about the process and hence to "infer the values of the variables (parameters or states) that characterize the process" [Tarantola, 2005]. In the following, observations/measurements of the process are denoted by

$$\tilde{x}_i[k],\ i \in \{1,\dots,n_x\},\ k \in \{0,1,\dots,N-1\}.$$

Characteristically, such measurements are always subjected to uncertainty, e.g. measurement noise or due to inherent variability of the considered process. In the following, we describe uncertainty of the measurement data by sets bounding its possible values. Such bounding sets can be simple intervals, e.g.

$$\tilde{x}_i[k] \in [\underline{x}_i[k], \overline{x}_i[k]] \doteq [x_i[k]]_{meas}.$$

More generally, polytopic bounding sets are used to describe uncertainties. For simplicity of presentation, we thus collect the uncertainty bounds for all measured compounds by (polytopic) sets

$$x[k] \in X_k \subseteq \mathbb{R}^{n_x},\ \ k \in \{0,1,\dots,N-1\},$$

and denote their collection by

$$D_{meas} = \{X_0, X_1, \cdots, X_{N-1}\}. \tag{2}$$

In addition to measurement data, knowledge about feasible values of the states or the parameters may be available. Such information is very important for testing and estimation, because often experimental data is sparse. For example, the system's states can be constrained by first principles, e.g. by considering conservation relations (mass, momentum, energy,...) or symmetry properties, see e.g. Ederer and Gilles [2007]. For shorthand, we denote the available a priori knowledge by

$$p \in P_0 \subseteq \mathbb{R}^{n_p},\ x[k] \in X_k \subseteq \mathbb{R}^{n_x}\ \forall k \in \{0,1,\dots,N-1\}$$

where P_0 and X are (polytopic) bounding sets of the parameters and states respectively. We assume given, for each state component x_i, an a priori bounding *interval* $[x_i]_{prior}$.

Remark 2. It is also possible to take so called qualitative constraints, capturing for example temporal logic conditions on the behaviour of the system. These constraints can be described by boolean variables as in [Rumschinski et al., 2012].

We furthermore require a measure of the distance of two sets; to this end, we consider

Definition 3. (Minimum distance). We define the minimum distance between two non-empty compact sets $A \in \mathbb{R}^n$, $B \in \mathbb{R}^n$ by the minimum distances between any two of their respective points, i.e.

$$d_{min}(A,B) = \min\{\|a-b\| : a \in A,\, b \in B\}.$$

Note that this distance is zero if the two sets overlap.

Invalidation and estimation

Our approach is motivated by a set-based invalidation and estimation scheme. We focus on a conceptual description of the invalidation and estimation approach. Further details, e.g. concerning the relaxation step, can be found in Borchers et al. [2009], Rumschinski et al. [2010], Streif et al. [2012].

The invalidation and estimation problems are approached by combining the model equations and the data within the following general problem

$$OP: \begin{cases} \min c(x,p) \text{ s.t. } \forall\, k \in \{0,\dots,N-1\}: \\ x_i[k+1] = f_i(x[k],p),\ i \in \{1,\dots,n_x\}, \\ x[k] \in X_k, \\ p \in P_0. \end{cases} \tag{3}$$

Hereby, $c(x,p)$ denotes a (polynomial) objective function, e.g. the weighted sum of least squares is given by

$$\sum_{\substack{1 \leqslant i \leqslant n_x \\ 0 \leqslant k \leqslant N-1}} \frac{1}{a_i}(x_i[k] - \tilde{x}_i[k])^2, \tag{4}$$

where a_i denotes the weighting factors, $\tilde{x}_i[k]$ the observations, and $x_i[k]$ the respective model prediction depending on the unknown parameters.

We denote the solution OP by c^*. Solutions of the above (non-convex) problem provide the desired results. In particular, if above OP has no solution for any choice of $c(x,p)$, then by construction, the model is *inconsistent* with the data. This way, model hypotheses can be falsified. To obtain an outer-bounding set of the parameters (or states) that describe the measurements, we have to determine all feasible solutions. To this end, consider the objective $c(x,p) = p_i$ of OP, and the respective solution c^*. This solution defines by construction a lower bound of the parameter, $\underline{p}_i \doteq c^* \leqslant p_i$. To obtain an upper bound, we consider $c(x,p) = -p_i$, and respectively $p_i \leqslant -c^* \doteq \overline{p}_i$. The interval $p_i \in [\underline{p}_i, \overline{p}_i]$ denotes the parameter uncertainty interval. Analogously, state uncertainty intervals can be obtained by solving OP.

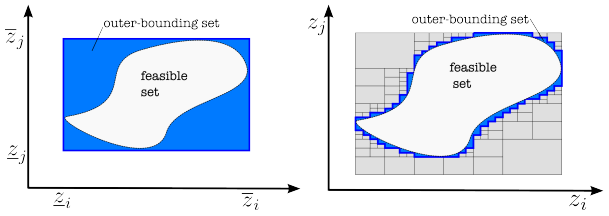
Remark 4. For optimization purposes, the sum of least squares (4) can be considered as objective function.

Due to nonlinearities of the model and data, the resulting optimization problem is frequently non-convex and ill-posed (see e.g. O'Sullivan [1986]). Thus, OP is in general a difficult optimization/feasibility problem. Showing that no solution exists, or finding the desired optimum, can thus be very challenging.

If the model equations and the constraints are polynomial functions, it is always possible to relax the non-convex OP into a semidefinite, and hence convex, problem (SDP). Advantageously, SDPs can be solved in polynomial time with arbitrary precision [Nesterov and Nemirovski, 1994], e.g. via primal-dual interior-point methods. Relaxation tightness is in general very difficult to assess, besides some particular problem classes where exact solutions are obtained by SDP relaxations Kim and Kojima [2003]. In general, the relaxation gap is difficult to assess, can however be handled by including strengthening constraints, by considering moments relaxation [Lasserre, 2001], or by partitioning strategies. Here, we utilize an iteration and a bisection algorithm, which is sufficient for this purpose while being computational tractable. To improve the performance, we furthermore consider relaxation to a linear optimization problem (LP) following the relaxation hierarchy proposed by Kojima [2002].

The most important relation of the relaxed problem SDP and the original problem OP is that any solution of OP is also a solution of SDP. This implies that no feasible solutions are missed. Furthermore, if both problems are feasible, then the optimum of SDP is a lower bound for the global minimum of OP. Infeasibility and optimality *certificates* are subsequently obtained via the dual problem of SDP, see e.g. Boyd and Vandenberghe [2004].

In particular, so derived infeasibility certificates provide a rigorous falsification criterion and an efficient test of incon-



(a) interval bounding (b) partitioning

Fig. 1. Illustration of the set-membership estimation techniques. z_i denotes the variable of interest, e.g. a parameter p_i or a state $x_i[k]$.

sistency, utilized for outlier detection later on, summarized as follows:

Result 1. (Invalidation). If the dual problem SDP is unbounded, then OP has no consistent solution. Hence, the model hypothesis is inconsistent with the data and rejected.

Dual feasible solutions can instead be used for estimation, in particular to outer-bound the uncertainty intervals of the parameters and the states:

Result 2. (Estimation). The parameter uncertainty interval $[p_i]_{post} = [\underline{p}_i, \overline{p}_i]$ (state uncertainty interval $[x_i[k]]_{post} = [\underline{x}_i[k], \overline{x}_i[k]]$ respectively) is obtained from two feasible solutions of the dual SDP.

For an illustration of the estimation techniques interval bounding and partitioning, see Fig. 1.

Remark 5. The set-based framework does allow simultaneously estimating the unknown model parameters and states.

With these preparations, we can now approach the outlier detection problem.

3. OUTLIER ANALYSIS

We consider the model as given in (1), and the available a priori data P_0, X and D_{meas}. We start from the observation that the model and the data are inconsistent according to Result 1.

Instead immediately rejecting the model, we consider the possibility that the data contains *some* outliers, see for an illustration Fig. 2.

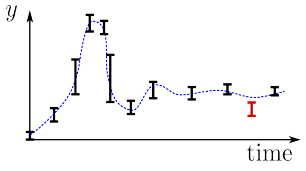


Fig. 2. Illustration non-homogeneous data uncertainties and outliers in measurements.

Thus, we consider the following definition of an outlier:

Definition 6. (Outlier). An outlier denotes a measurement which is not appropriately described by the uncertainty description and the model, i.e. the real measurement is not covered.

The number of outliers in the measurement data is denoted by n_o, which is of course typically unknown.

With these preparations, we can now turn on the outlier detection and analysis problem.

3.1 Combinatorial outlier detection

An approach to detect possible outliers in the data set D_{meas} with the total number of measurements $M = n_x N$ consists in formulating outlier hypotheses, i.e. to select possible outlier candidates and discard them, and subsequently to perform an inconsistency test utilizing Result 1. Technically, when we say a measurement $\tilde{x}_i[k]$ for some $i \in \{1, \ldots, n_x\}$ and $k \in \{0, 1, \ldots, N-1\}$ is *discarded*, we set $[x_i[k]]_{meas} \doteq [x_i]_{prior}$. We denote this by $D_{meas} \setminus x_i[k]$.

To this end, we consider first the simple case where a single outlier is suspected. Then, the proposed outlier detection strategy consists in *discarding* the measurements $x_i[k]$ (starting e.g. with $i = 1$ and $k = 0$) one by one, and perform the inconsistency test Result 1 with the model and $D_{meas} \setminus x_i[k]$.

Proposition 7. (Single outlier case). A single outlier is detected by at most M infeasibility certificates according to Result 1.

The respective outlier, $x_i[k]$, can be further analysed, e.g. the distance of the outlier $[x_i[k]]_{meas}$ from the reachable state $[x_i[k]]_{post}$ (refer Result 2) can be estimated.

Remark 3.1. Note that, even in this most simple case, there exists the possibility that there exists no unique solution to the outlier problem, i.e. there might exist alternative, possibly exclusive, outliers. This is independent from the proposed method, rather an intrinsic issue of the (in this case ill-posed) outlier detection problem.

Multiple outliers The proposed outlier detection approach extends to the multiple outlier case. To this end, we *discard* n_o measurements from the data set analogously to the single outlier case, and perform the consistency test according to Result 1. We have:

Proposition 8. (Combinatorial outlier detection). To detect n_o outliers within M measurements, at most

$$\binom{M}{n_o} = \frac{M!}{n_o!(M-n_o)!} \tag{5}$$

infeasibility certificates according to Result 1 are required.

Proof. The number of possible arrangements of n_o outliers in M measurements is equivalent to the classical combinatorial problem, i.e. n_o-combinations (zero elements) in a sequence of otherwise one elements of length M. Hence, the number of possible combinations is the given by the binomial coefficient (5). □

Note that in practice the number of outliers is unknown. Therefore, a strategy in this case consists in increasing successively the number of suspected outliers starting from a single one. This ensures to obtain a minimal number of outliers, with the advantage that masking and swamping are avoided, although uniqueness can not be guaranteed in general, refer Remark 3.1.

The disadvantage of this approach is that the number of combinations and therefore the number of required evaluations according to Result 1 increases with the number of suspected outliers n_o; particularly if M is large, the complexity of the proposed approach is approximately $(M)^{n_o}$, i.e. exponential in the number of suspected outliers. Hence, the combinatorial outlier detection approach is in general not suited for large data sets (with possibly many outliers), given that no particular outlier hypotheses can be formulated and a combinatorial search has to be

considered. This is because the outliers have to be detected instantaneously (all at once).

For larger data sets, a sequential approach may therefore be more advantageous, e.g. to remove the most extreme outliers first. To do so, a possibility consists in 'relaxing' the uncertainty description, i.e. by introducing an additional pessimism. By choosing the pessimism appropriately, the number of outliers is decreased, which in turn facilitates the combinatorial detection approach.

To formalize the relaxation of the uncertainty description, we consider a tolerance given as follows.

Definition 9. (Tolerance). The tolerance $\epsilon \geqslant 0$ admits the following properties:

- If $\epsilon = 0$, then $D^{\epsilon}_{meas} = D_{meas}$.
- For any $0 < \epsilon$, we have $D_{meas} \subset D^{\epsilon}_{meas}$.
- For any $\epsilon_2 < \epsilon_1$, we have $D^{\epsilon_2}_{meas} \subset D^{\epsilon_1}_{meas}$.

Exemplary, a tolerance can be modelled by an additional absolute or relative error according to Equation (8) and (9).

By considering a tolerance, the sequential outlier detection procedure is devised as follows:

(1) Introduce a positive tolerance ϵ according to Def. 9 (D^{ϵ}_{meas}).
(2) Perform the combinatorial outlier detection approach and remove the outliers.
(3) Decrease the tolerance and repeat the procedure until no tolerance is required.

The challenge hereby clearly consists in choosing an appropriate tolerance. Note that, in general, the tolerance can be considered as a 'weighting' of uncertainty; if no additional knowledge is available, it may be reasonable to consider a tolerance which acts equally onto all measurements, e.g. an additional absolute or relative error. Conversely, knowledge might be available that certain outputs (or specific measurements) might be more prone to errors than others; then, the tolerance can of course be adapted accordingly.

Regression Example

As an illustrative though simple example, we consider a static regression problem given the six (uncertain) measurements pairs (x, y) with (0,0.8±0.25), (1,2±0.25), (2,2.2±0.25), (3,4.1±0.25), (4,4.85±0.25), (5,5.2±0.25).

We first examine a linear regression model given by

$$y = a_1 x + a_0,$$

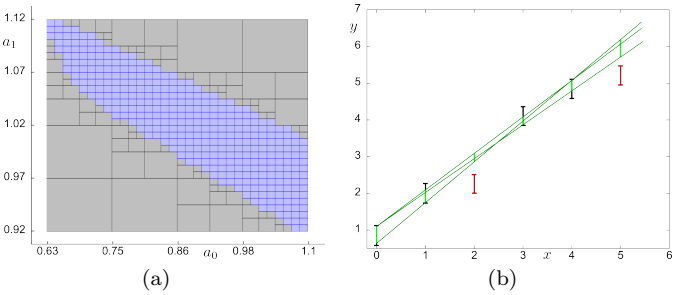
where $a_1 \in \mathbb{R}$ and $a_0 \in \mathbb{R}$ are the unknown parameters.

Regarding the linear regression model, we found at least two outliers according to Prop. 8, i.e. at $x = 2$ and $x = 5$. Regarding the remaining measurements, the parameters are estimated according to Result 2 as depicted in Fig. 3(a). The reachable states are indicated in green color. The respective solution sets are shown in Fig. 3(b).

Additionally, we consider a quadratic regression model given by

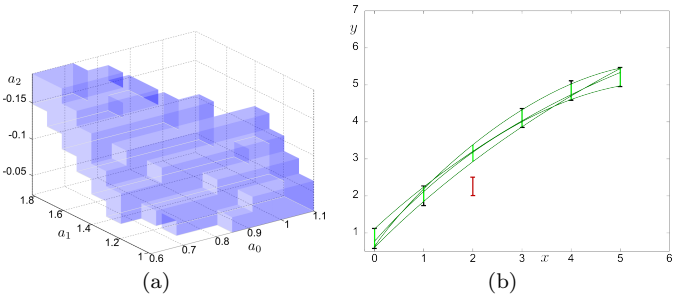
$$y = a_2 x^2 + a_1 x + a_0,$$

where $a_2 \in \mathbb{R}$, $a_1 \in \mathbb{R}$, $a_0 \in \mathbb{R}$ are the unknown parameters. For this regression model, we found one outlier at $x = 2$. With the remaining measurements, the parameters were estimated according to Result 2, shown in Fig.4(a). The respective solution sets are depicted in Fig.4(b).



(a) (b)

Fig. 3. Combinatorial outlier detection for the linear regression model. Left: Consistent parameters after removal of the outliers. Right: Detected outliers and reachable sets.



(a) (b)

Fig. 4. Combinatorial outlier detection for the quadratic regression model. Left: Consistent parameters after removal of the outliers. Right: Detected outliers and reachable sets.

3.2 Reachability-based outlier detection

Complementary to the combinatorial detection approach, reachability analysis can be considered for detecting outliers. We start from the observation that the model (1) and the data are inconsistent according to Result 1 due to outliers.

The reachability-based outlier detection approach is performed as follows.

(1) Introduce a positive tolerance ϵ according to Def. 9 (D^{ϵ}_{meas}).
(2) Estimate the reachable sets $[x_i[k]]_{post}$, for all $i \in \{1, \ldots, n_x\}$ and $k \in \{0, 1, \ldots, N-1\}$ regarding the data D^{ϵ}_{meas} (Result 2).
(3) Compare the reachable sets $[x_i[k]]_{post}$ and the measurements $[x_i[k]]_{meas}$ considering the minimum distance according Def. 3. Each measurement with $d_{\min}([x_i[k]]_{meas}, [x_i[k]]_{post}) > 0$ is an outlier, and can be removed safely from the data.
(4) Decrease the tolerance ϵ and repeat the procedure until $\epsilon = 0$ or a desired threshold is achieved.

The proposed approach avoids combinatorial evaluation of all possible outlier hypotheses, and numerical complexity does not depend on the number of outliers. Though, due to the introduced additional tolerance, results may be overly pessimistic, and finding outliers can not be guaranteed this way (in contrast to the combinatorial detection strategy). An advantage of this approach is however that the model parameters can be adapted and refined successively without requiring that all outliers are already removed from the data. We provide next an example from systems biology for the reachability-based outlier detection.

Cell Growth Example We consider a mechanistic description of a cell growth process in batch, including the uptake of glucose (x_2) and glutamine (x_3), the release of lactate (x_4) and ammonia (x_1), as well as the dynamics of dead (x_5) and viable cells (x_6), following [Bailey and Ollis, 1986] and references therein:

$$\begin{aligned}
\dot{x}_1 &= \tfrac{\mu_{max}}{Y'_{\mathrm{X/Amn}}} \frac{x_3}{x_3 + K_{Gln}} x_6 + K_{\mathrm{deg}} \mathrm{x}_3 \\
\dot{x}_2 &= -\tfrac{\mu_{max}}{Y'_{\mathrm{X/Glc}}} \frac{x_2}{x_2 + K_{Glc}} x_6 \\
\dot{x}_3 &= -\tfrac{\mu_{max}}{Y'_{\mathrm{X/Gln}}} \frac{x_3}{x_3 + K_{Gln}} x_6 - K_{\mathrm{deg}} x_3 \\
\dot{x}_4 &= \tfrac{\mu_{max}}{Y'_{\mathrm{X/Lac}}} \frac{x_2}{x_2 + K_{Glc}} x_6 \\
\dot{x}_5 &= K_D x_6 - K_{\mathrm{lys}} x_5 \\
\dot{x}_6 &= (\mu - K_D) x_6,
\end{aligned} \tag{6}$$

where

$$\mu = \mu_{\max} \frac{x_2}{x_2 + K_{Glc}}. \tag{7}$$

For this cell growth process, a batch experiment has been performed within a bioreactor, and the measurements for all six states were recorded, for details, see Borchers et al. [2013]. A summary of the measurement errors obtained by assay validation is shown in Tab. 1.

From statistical analysis, lower and upper bounding values of the uncertainty are derived. In case variances are homogeneously distributed, the standard deviation σ_i is considered to derive the 1-sigma confidence intervals of the measurement uncertainties: with

$$\begin{aligned}
\underline{x}_i[k] &= \tilde{x}_i[k] - \sigma_i, \\
\overline{x}_i[k] &= \tilde{x}_i[k] + \sigma_i.
\end{aligned} \tag{8}$$

In case a relative standard deviation of the method r_i is used to describe the measurement uncertainty, we have

$$\begin{aligned}
\underline{x}_i[k] &= \tilde{x}_i[k](1 - r_i/100), \\
\overline{x}_i[k] &= \tilde{x}_i[k](1 + r_i/100).
\end{aligned} \tag{9}$$

Besides the values of the parameters K_{deg} and K_{lys}, which are known from previous experiments, the parameters of the model (6) are unknown.

By employing an initial invalidity test according to Result 1, we found that the model and the data were inconsistent. We suggested that this is due to outliers in the measurement data.

To this end, we first performed a single outlier hypothesis test according to Prop. 7, with negative result. We concluded that multiple outliers were present. Next, we introduced an additional tolerance (ϵ=10 %, relative error) onto all measurements, denoted by D^{ϵ}_{meas}. Using the relaxed measurement data, we then performed a combined parameter estimation and reachability analysis. The results are shown in Table 2 and Fig. 5 respectively. By direct comparison of the so obtained reachable states with the measurement data, three outliers were detected. The result was last validated performing an invalidity test after removing the three outliers.

The example nicely illustrates the utility of the proposed approach for biotechnological applications.

4. CONCLUSIONS

We considered the detection of outliers considering nonlinear systems and time-dependent data (possibly inhomogeneous). Using an invalidation approach within a set-based setting, we presented a combinatorial approach. The approach is suitable only if the number of outliers is small. For larger data sets with possibly several outliers, we proposed extensions of combinatorial setting considering sequential detection and a detection strategy based on reachability analysis. Both approaches provide conclusive results, under the assumptions made on the measurement uncertainties.

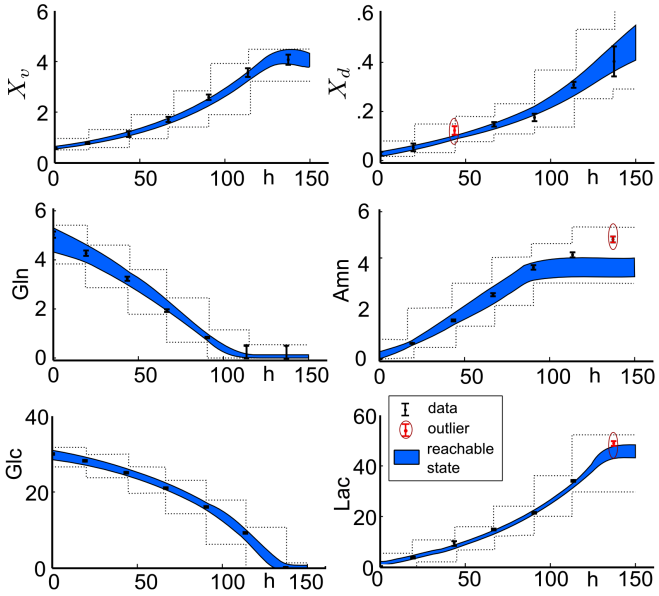


Fig. 5. Reachability-based outlier detection.

By estimating the reachable states regarding a more conservative uncertainty description, the most extreme outliers can thus be detected and removed first while the model parameters can be refined simultaneously without removal of all outliers. Subsequently, the pessimism can be reduced and the procedure repeated until all outliers are detected. The results have to be validated thereafter using a single inconsistency test.

An important consequence of the model-generic setting is that the detection of outliers inherently depends on the model used for detection.While outliers may be present with respect to one model (hypothesis), other outliers or none at all may result considering another model. Deriving respective reliability criteria to compare hypotheses under outliers is an ongoing research direction. A further step is to devise a moving horizon based *on-line* outlier detection for control and supervision of processes.

APPENDIX

Table 1. Statistical analysis of the measurement errors by validation assay.

	x_1 [mM]	x_2 [mM]	x_3^* [mM]	x_4 [mM]	x_5^* [$10^6 \frac{cells}{ml}$]	x_6^*
SD	0.03	0.39	(0.08)	0.30	(0.02)	(0.02)
% SD	(2.1%)	(1.9%)	5.9%	(1.7%)	6.2%	6.2

*non-homogeneous variance. SD: standard deviation (σ_i). % SD: relative standard deviation (r_i).

REFERENCES

J.E. Bailey and D.F. Ollis. *Biochemical engineering fundamentals*. McGraw-Hill Education, 1986.

Table 2. Summary of parameters corresponding to the bioreactor experiment.

par.	unit	references	bioreactor $[\underline{p}_i, \overline{p}_i]$	bioreactor *opt**
$\mu_{\max}$	$1/h$	2e-2–1.3e-1	[1.54,1.91]e-2	1.90e-2
$Y'_{X/Glc}$	$\frac{10^9}{mmol}$	6e-2–1.7	[0.93,1.93]e-1	1.44e-1
$Y'_{X/Gln}$	$\frac{10^9}{mmol}$	3e-2–1.6	[3.31,6.23]e-1	4.89e-1
$Y'_{X/Lac}$	$\frac{10^9}{mmol}$	7e-2–2.5e-1	[6.20,8.28]e-2	8.22e-2
$Y'_{X/Amn}$	$\frac{10^9}{mmol}$	5.0e-1–2.0	[3.98,6.03]e-1	5.48e-1
K_d	$1/h$	2.8e-4–3e-1	[1.66,3.45]e-3	2.66e-3
K_{Glc}	mM	1.5e-1–1.0	[0.89,2.43]	1.21
K_{Gln}	mM	6e-2–8.0e-1	[0.01,1.35]	0.55
K_{deg}	$1/h$	1.5e-3[1]	–	
K_{lys}	$1/h$	1.0e-2[1]	–	

$\underline{p}_i$ and $\overline{p}_i$ denote the limits of the 1-sigma parameter confidence interval. [1]unpublished data. *optimal parameter values regarding the sum of least squares (4).

V. Barnett and T. Lewis. *Outliers in Statistical Data. John Wiley, 1994.* John Wiley, 1994.

I. Ben-Gal. *Outlier detection, In: Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, volume 1. Kluwer Academic Publishers, 2005.

I. Ben-Gal, G. Morag, and A. Shmilovici. Context-based statistical process control. *Technometrics*, 45(4):293–311, 2003.

S. Borchers, P. Rumschinski, S. Bosio, R. Weismantel, and R. Findeisen. A set-based framework for coherent model invalidation and parameter estimation of discrete time nonlinear systems. In *Proc. 48th Conf. on Dec. and Cont. (CDC 2009), Shanghai, China*, pages 6786–92, 2009.

S. Borchers, S. Freund, A. Rath, S. Streif, U. Reichl, and R. Findeisen. Identification of growth phases and influencing factors in fermentation with AGE1.HN cells using set-based methods. *PLoS One*, 2013.

S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, UK, 2004.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.

M. Ederer and E.D. Gilles. Thermodynamically feasible kinetic models of reaction networks. *Biophysical journal*, 92(6):1846–1857, 2007.

D. Hawkins. *Identification of outliers.* Chapman and Hall, London, 1980.

V. Hodge and J. Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.

R.A. Johnson and D.W. Wichern. *Applied multivariate statistical analysis*, volume 258. Prentice Hall Upper Saddle River, 1992.

S. Kim and M. Kojima. Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations. *Comp. Opt. and Appl.*, 26(2):143–154, 2003.

M. Kojima. A general framework for convex relaxation of polynomial optimization problems over cones. *J. Op. Res. Soc. Japan*, 46:125–144, 2002.

J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11:796–817, 2001.

H. Liu, S. Shah, and W. Jiang. On-line outlier detection and data cleaning. *Computers & chemical engineering*, 28(9):1635–1647, 2004.

Y. Nesterov and A. Nemirovski. *Interior-point polynomial algorithms in convex Programming*, volume 13. SIAM J. Appl. Math., 1994.

F. O'Sullivan. A statistical perspective on ill-posed inverse problems. *Stat Sci*, pages 502–518, 1986.

P. Rumschinski, S. Borchers, S. Bosio, R. Weismantel, and R. Findeisen. Set-based dynamical parameter estimation and model invalidation for biochemical reaction networks. *BMC Sys. Biol.*, 4:69, 2010.

P. Rumschinski, S. Streif, and R. Findeisen. Combining qualitative information and semi-quantitative data for guaranteed invalidation of biochemical network models. *International Journal of Robust and Nonlinear Control*, 22(10):1157–1173, 2012.

S. Streif, A. Savchenko, P. Rumschinski, S. Borchers, and R. Findeisen. ADMIT matlab toolbox for guaranteed invalidation and estimation of biochemical reaction networks. *Bioinformatics*, 28(9):1290–91, 2012.

A. Tarantola. *Inverse problem theory and methods for model parameter estimation.* Society for Industrial Mathematics, 2005.

G. Williams, R. Baxter, H. He, S. Hawkins, and L. Gu. A comparative study of RNN for outlier detection in data mining. In *Data Mining, 2002. ICDM 2002. Proceedings. 2002 IEEE International Conference on*, pages 709–712. IEEE, 2002.