A toolbox to find the best mechanistic model to predict the behavior of environmental systems

van Turnhout, André G.; Kleerebezem, Robbert; Heimovaara, Timo J.

DOI
10.1016/j.envsoft.2016.05.002

Publication date
2016

Document Version
Accepted author manuscript

Published in
Environmental Modelling & Software

Citation (APA)
van Turnhout, A. G., Kleerebezem, R., & Heimovaara, T. J. (2016). A toolbox to find the best mechanistic model to predict the behavior of environmental systems. *Environmental Modelling & Software*, *83*, 344-355. https://doi.org/10.1016/j.envsoft.2016.05.002

Elsevier Editorial System(tm) for Environmental Modelling & Software
Manuscript Draft

Manuscript Number: ENVSOFT-D-15-00385R3

Title: A toolbox to find the best mechanistic model to predict the behavior of environmental systems.

Article Type: Research Paper

Keywords: Bayesian inference; mechanistic biogeochemical modeling; integrative assessment of environmental systems; qualitative and quantitative model selection; optimal data description

Corresponding Author: Mr. André Gerard van Turnhout, Ir.

Corresponding Author's Institution: TU Delft

First Author: André Gerard van Turnhout, Ir.

Order of Authors: André Gerard van Turnhout, Ir.; Robbert Kleerebezem, Dr. Ir.; Timo J Heimovaara, Dr. Ir.

Abstract: Reliable prediction of the long-term behavior of environmental systems such as Municipal Solid Waste (MSW) landfills is challenging. While many driving forces influence this behavior, characterization of them is limited by measurement techniques. Therefore, a model structure for reliable prediction needs to optimally combine all measured information with suitable mechanistic information from literature. How to get such an optimal model structure? This study presents a toolbox to find and build the model structure that describes an environmental system as close as possible. The toolbox combines environmental frameworks to include all suitable mechanistic information; it fully couples kinetic and equilibrium reactions and contains multiple resources to obtain biogeochemical parameters.
Several possible optimal model structures are quickly built and evaluated with objective statistical performance criteria obtained via Bayesian inference. By applying the novel methodology, we select the best model structure for anaerobic digestion of MSW in full scale landfills.

Response to Reviewers: There were no comments from the reviewers.

Dear editor,

Thank you for your fast reply, we are very happy that our manuscript is considered for publication in Environmental Modelling & Software (EM&S). We also thank you for your additional comments which enabled us to improve the transparency and framing of the presented methodology and results. We double checked the grammar in the paper and changed the following according to your comments.

In line 55, we included two citations of papers from EM&S (Vrugt et al. (2016) and Kelly et al. (2013)) which express the need for a more objective integrated assessment of environmental systems. These are included to show that novel methods such as the one we present are necessary to decrease uncertainty in modeling of environmental systems. Our toolbox decreases model uncertainty by combining mechanistic information from different environmental fields and evaluating models based on qualitative and quantitative criteria.

In line 56, we included a reference to Bennett et al. (2013) who propose a generic evaluation procedure of model performance. Such an approach strengthens the credibility and relevance of modelling reported which is in line with the methodology we present. In lines 83-92, we indicate that our method supports all five evaluation steps proposed by Bennett et al. (2013). In our evaluation, however, the step of checking the data is minimal because the available measured data and information about its error was very limited. Nevertheless, we believe that the characteristics of our novel methodology are sufficiently demonstrated and that the best model is satisfactory given the scope of the paper.

In section 2.4, we added a subsection describing the qualitative performance criteria that are generated by the toolbox. In section 2.4.3, we justify our choice in quantitative criteria. We chose a set of criteria that enabled us to quantify aspects related to the (mechanistic) uncertainty of the models which is in line with the purpose of our modeling exercise. In addition, we indicate that other criteria suited for other modeling purposes are readily implemented.

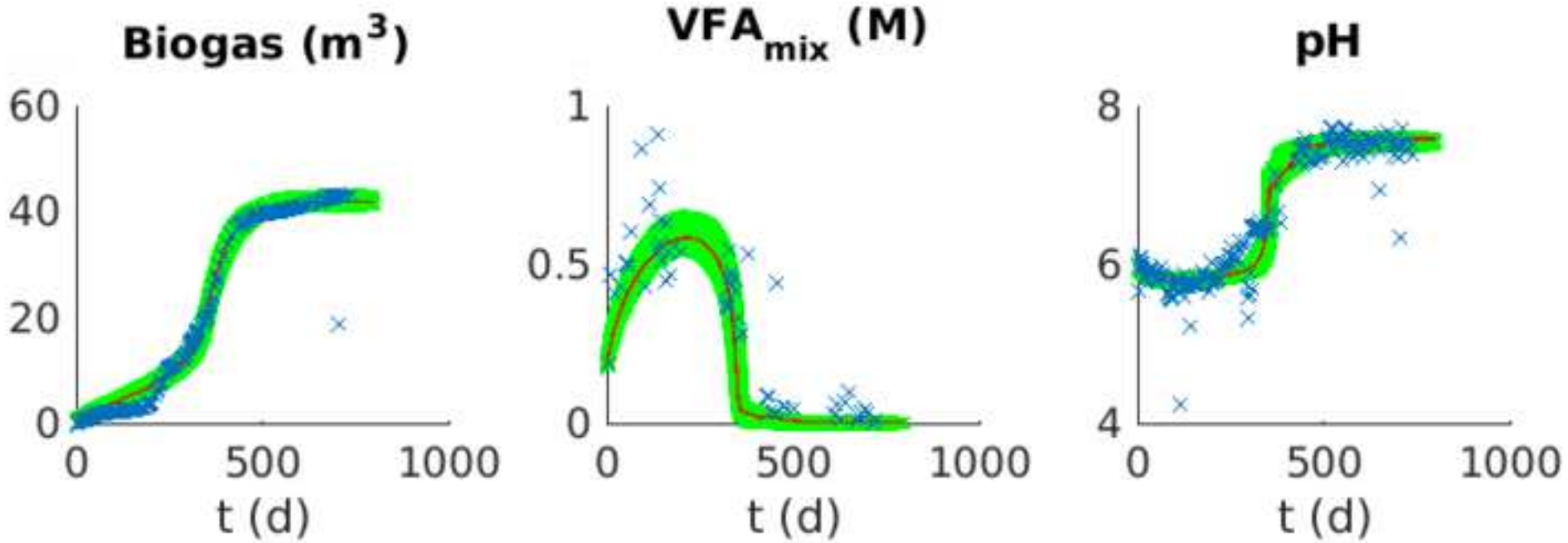
Results and discussion are only slightly modified because we believe that the changes made in the introduction and theory already improved their framing and justification significantly. Furthermore, we modified the conclusion about the best model into two steps. First we conclude which model is best for lysimeter scale and then we conclude which model is best for full scale.

We thank you for your time and hope our paper is now fit to contribute to Environmental Modelling & Software.

Sincerely yours,

André van Turnhout

Graphical Abstract



Highlights

- Finding the best mechanistic model to predict the behavior of environmental systems.
- The toolbox allows to quickly assess multiple biogeochemical reaction networks.
- Coupled kinetic & equilibrium processes and mechanistic parameter resources.
- Networks are evaluated with objective statistical criteria given measured data.
- The toolbox enabled to find the optimal network for anaerobic digestion of MSW.

A toolbox to find the best mechanistic model to predict the behavior of environmental systems.

Illustrated by modeling anaerobic degradation of Municipal Solid Waste

André G. van Turnhout[a,*], Robbert Kleerebezem[b], Timo J. Heimovaara[a]

[a] *Geoscience & Engineering Department, Faculty of CiTG, Delft University of Technology, Stevinweg 1, 2628CN, Delft, Netherlands*
[b] *Environmental Technology Group, Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC, Delft, Netherlands*

Abstract

Reliable prediction of the long-term behavior of environmental systems such as Municipal Solid Waste (MSW) landfills is challenging. While many driving forces influence this behavior, characterization of them is limited by measurement techniques. Therefore, a model structure for reliable prediction needs to optimally combine all measured information with suitable mechanistic information from literature. How to get such an optimal model structure? This study presents a toolbox to find and build the model structure that describes an environmental system as close as possible. The toolbox combines environmental frameworks to include all suitable mechanistic information; it fully couples kinetic and equilibrium reactions and contains multiple resources to obtain biogeochemical parameters. Several possible optimal model structures are quickly built and evaluated with objective statistical performance criteria obtained via Bayesian inference. By applying the novel methodology, we select the best model structure for anaerobic digestion of MSW in full scale landfills.

Keywords: Bayesian inference, mechanistic biogeochemical modeling, integrative assessment of environmental systems, qualitative and quantitative model selection, optimal data description

1. Introduction

It is challenging to make reliable predictions of the long-term behavior of environmental systems such as Municipal Solid Waste (MSW) landfills. In such systems, there are many driving forces influencing behavior, while current measurement techniques are not sufficient to characterize them. Therefore, for reliable predictions we require a mechanistic description with minimal uncertainty for extrapolation of measured data.

Reliable prediction of long-term emissions is needed for landfill management. Decisions about ending of landfill after care strongly depend on these predictions because after care can only be stopped

*Corresponding author, email address: a.g.vanturnhout@tudelft.nl

when emissions are below certain threshold values. Furthermore, reliable predictions also improve estimations of energy recovery from emissions such as methane. This energy is directly utilized in the facilities at the landfill.

So far, prediction of emissions by any modeling strategy is highly uncertain. In general, models have been developed according to two strategies. One strategy extrapolates measured emissions using empirical relations. Some of these models fit exponential equations to measured gas and leachate data [6, 32, 9, 13]. Others apply neural networks on emission data [26, 14]. Although the fit of these empirical models with measured data is very good, their extrapolations are poor because these models are not constrained by mechanistic principles. These models do not consider the impact of changes in environmental conditions on the emissions while multiple studies have shown that e.g. pH and mass transport limitation significantly influence the performance of waste water and solid waste treatment[34, 3, 1, 40, 42].

The other modeling strategy to predict emissions is mechanistic [29, 45, 20, 8, 7, 17]. These models vary a lot in complexity and the type of mechanistic information they include. Model concepts range from single point to three dimensional implementation and from a single type of framework such as biochemistry to frameworks coupling biochemistry, hydrology and settlement. Although these models do restrict their predictions with mechanistic principles and have given interesting insights, their prediction accuracy is often also very poor. Prediction with these models is poor because the uncertainty in mechanistic assumptions is very large, especially for very complex models. In addition it is difficult to reduce this uncertainty by model calibration because measured data is limited. The large uncertainty in mechanistic correctness is reflected by the wide spread of parameter values published in literature [22].

To improve prediction accuracy, the challenge is to select a model structure that is constrained by mechanistic principles and has minimum uncertainty in assumptions. In order to find such a model, a more objective integrated assessment of environmental systems is needed [43, 15] in which model performance evaluation is generalized [4]. This type of assessment requires several prerequisites for developing models. First, mechanisms from different environmental fields should be available to be combined in order to include all suitable mechanistic information. Second, well established mechanistic parameters with relative low uncertainty should be readily obtainable from databases or derivation methods. Third, model performance should be able to be analyzed qualitatively and quantitatively. Statistical analysis of the remaining uncertainty in parameters in light of the measured data should allow us to quantify the uncertainty in model structure, calibrated parameter values and mechanistic correctness of the model. Finally, multiple reaction networks should be able to be quickly built and

the results quickly assessed in order to find the optimal model. Current modeling approaches do not fulfill all these prerequisites and therefore limit the possibility to find optimal models for prediction purposes.

The aim of this study is to develop a methodology that includes all the prerequisites previously mentioned in order to enable us to find an optimal mechanistic model for anaerobic degradation of MSW. The toolbox allows us to quickly build mechanistic biogeochemical reaction networks that may include kinetic reactions, equilibrium reactions and environmental inhibitions in a multi-phase system. Parameter values are obtained from an extensive geochemical database [21], a derivation method based on thermodynamic principles [16] and calibration based on measured data. Performances of these reaction networks are evaluated quantitatively with statistical criteria obtained by applying Bayesian inference with the DREAM algorithm [44, 18]. We illustrate our generic approach by finding an optimal mechanistic model for describing a data set measured on a series of landfill lysimeters, previously published by Valencia et al. [36, 37, 35, 38, 39], by evaluating four possible reaction networks. Our aim is to find a description which we can apply to full-scale anaerobic landfills.

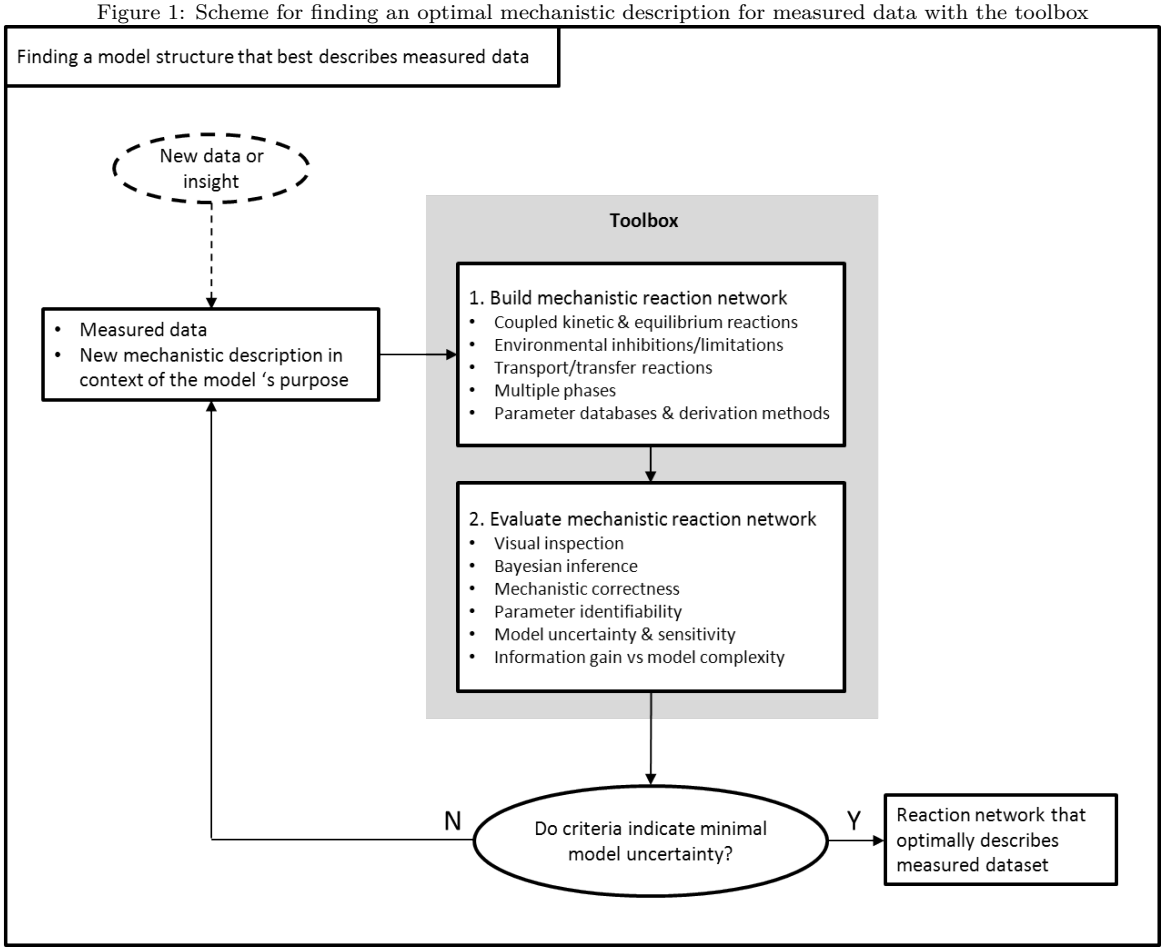
2. Theory

2.1. The cycle of finding an optimal model structure

The toolbox we present here is a novel combination of several approaches enabling us to find an optimal mechanistic model for describing measured data. This method follows the scheme depicted in figure 1 and allows a generic evaluation of model performance such as proposed by Bennett et al. (2013) [4]. We start by defining a first possible mechanistic reaction network which should provide a good fit to the measured data.

Subsequently, this reaction network is evaluated qualitatively by plotting the modeled data against the measured data and quantitatively by using a range of performance criteria which we obtain by applying Bayesian inference on the most uncertain parameters. These criteria allow us to judge aspects such as mechanistic correctness, parameter identifiability, model uncertainty, model sensitivity and model complexity. Based on the results and the purpose of the model, we decide if the model describes the measured data with sufficient (mechanistic) accuracy. If the accuracy is not sufficient, a next iteration is started where an update to the reaction network is evaluated until an optimal model is found. All prerequisites combined in the toolbox are discussed in detail below.

Figure 1: Scheme for finding an optimal mechanistic description for measured data with the toolbox



2.2. Defining the model structure with mechanistic information from different environmental frameworks

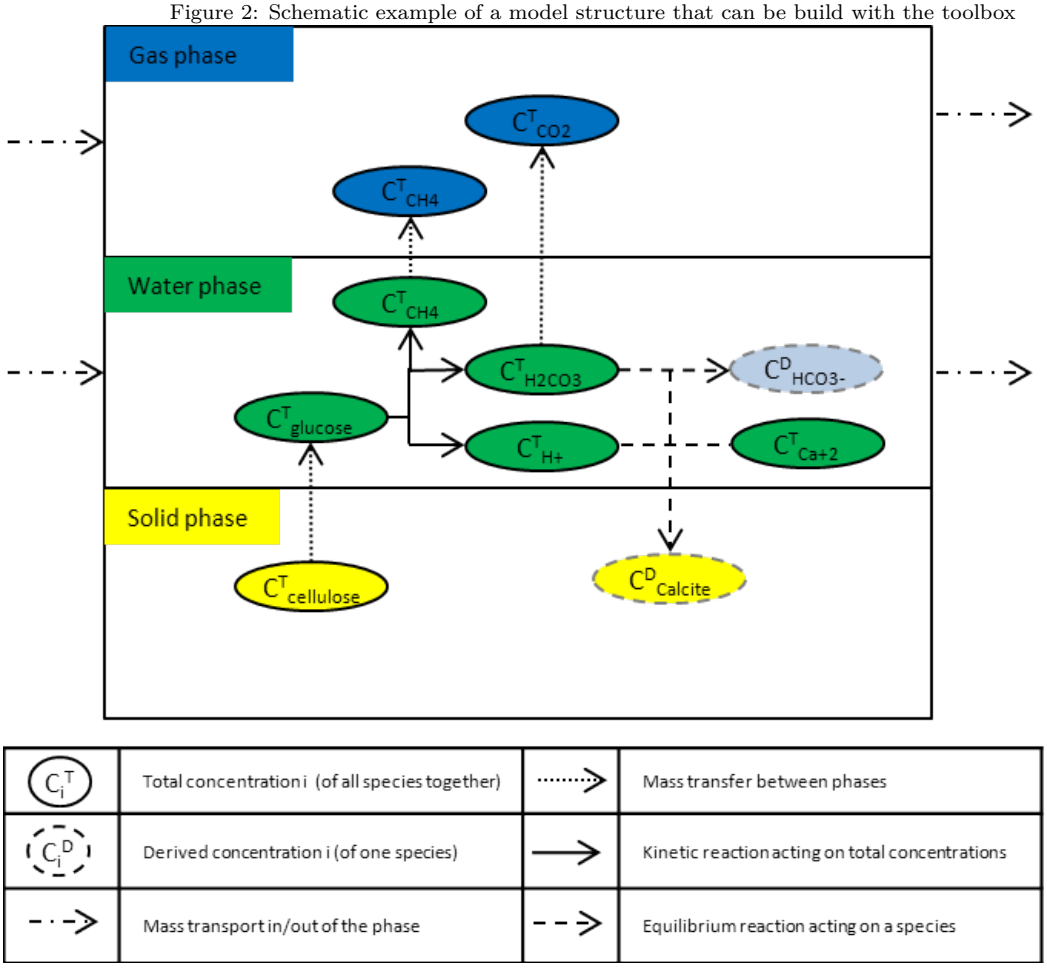
2.2.1. Multiphase, multicomponent and multiprocess environment

Model structures can be built that contain biogeochemical reaction networks within a multiphase environment with mass transport across the model domain and mass transfer between the phases. A schematic example of such a structure is given in figure 2. The model describes the fate of chemical compounds present in the solid, gas and liquid phase. In each phase we track the change of the total concentrations ($C^{\mathrm{T}}_{\mathrm{ref}}$) and derived concentrations (C^{D}) as a function of time. Total concentrations are calculated with the mass balances of building blocks of species e.g. H^+ and ${CO_3}^{-2}$ which are expressed in terms of one reference species. Using chemical equilibrium approaches, we can calculate the derived concentrations from the set of total concentrations (i.e. mass balances). Changes in total concentrations with time are caused by kinetic processes which can be either biochemical redox reactions, transfer of compounds between phases and transport in or out of the model domain. Biochemical

reactions are influenced by environmental inhibitions and limitations.

To support quick implementation of a model structure, it is completely defined within a single spreadsheet and solved with a generic matrix calculation method.

Figure 2: Schematic example of a model structure that can be build with the toolbox



2.2.2. Fully coupled rate dependent and equilibrium processes

Rate dependent processes and equilibrium processes are fully coupled within the model structure. This means that all derived concentrations are automatically recalculated when the total concentrations change because of biochemistry and transport or transfer processes. For example, the total concentration of H_2CO_3 in a standard water-carbonate system (i.e. $C^{\mathrm{T}}_{\mathrm{H_2CO_3}}$, $C^{\mathrm{T}}_{\mathrm{Ca^{+2}}}$ and $C^{\mathrm{T}}_{\mathrm{H^+}}$) changes because of gas production. The model updates the $C^{\mathrm{T}}_{\mathrm{H_2CO_3}}$ and as a result the pH and concentrations of CO_3^{-2}, HCO_3^{-}, H_2CO_3 and $CaCO_3$ are automatically updated. Full coupling allows accurate calculation of derived concentrations and their influence on the rate dependent processes in the system.

Details on how to derive concentrations from mass balances according to equilibrium reactions are given in Bethke et al. (2008) [5].

2.2.3. Type of rate dependent processes

Biogeochemical reaction rates ($R^{\mathrm{K}}_{C_{\mathrm{i}}^{\mathrm{T}}}$) are implemented as

$$R^{\mathrm{K}}_{C_{\mathrm{i}}{}^{\mathrm{T}}} = \mu^{\max} \cdot C^{\mathrm{T}'} \cdot I \cdot s_{\mathrm{i}} \tag{1}$$

where $\mu^{\max}$ is the maximum rate, $C^{\mathrm{T}'}$ is the concentration that drives the reaction, I is the total inhibition factor ranging from 0 to 1 and s_{i} is the stoichiometry. This type of reaction supports Monod kinetics when the inhibition factor is (partly) determined by substrate limitation and the total concentration of bacteria is set as the driving concentration.

The rates of change in the liquid phase ($\mathrm{p_l}$) and the gas phase ($\mathrm{p_g}$) due to the presence of a mass transfer flux between phases are implemented as

$$F^{\mathrm{MF}}_{C_{\mathrm{i}}^{\mathrm{T,p_l}}} = k_{\mathrm{l}}a \cdot \left(C_{\mathrm{i}}^{\mathrm{p_l}'} - C_{\mathrm{i}}^{\mathrm{p_g}'} \cdot K_{\mathrm{H}} \cdot R \cdot T \right) \tag{2}$$

$$F^{\mathrm{MF}}_{C_{\mathrm{i}}^{\mathrm{T,p_g}}} = F^{\mathrm{MF}}_{C_{\mathrm{i}}^{\mathrm{T,p_l}}} \cdot \frac{V_{\mathrm{l}}}{V_{\mathrm{g}}} \tag{3}$$

where $k_{\mathrm{l}}a$ is the mass transfer constant, $C^{\mathrm{p_l}'}$ and $C^{\mathrm{p_g}'}$ are the driving concentrations (total or derived) in the water phase and the gas phase, K_{H} is the Henry coefficient, T is the temperature, R is the universal gas constant, V_{l} is the volume of the water phase and V_{g} is the volume of the gas phase.

The rate of change due to a mass transport flux ($F^{\mathrm{MT}}_{C_{\mathrm{i}}^{\mathrm{T}}}$) is implemented as

$$F^{\mathrm{MT}}_{C_{\mathrm{i}}^{\mathrm{T}}} = \frac{\phi \cdot C_{\mathrm{i}}^{\mathrm{T}}}{V} \tag{4}$$

where ϕ is the mass transport flow and V is the volume of the phase.

2.2.4. Type of equilibrium processes

Two types of equilibrium processes can be implemented in a model structure. The first type is a true equilibrium reaction relating concentrations of species according to mass action law. The second type is an equilibrium between the gas and the liquid phase. This equilibrium is implemented using mass transfer reactions (eq. 2) with high values for $k_{\mathrm{l}}a$.

2.2.5. Environmental inhibitions and limitations

Rate inhibiting and limiting reaction terms are implemented using four mechanisms of biochemical reactions taken from environmental studies (equations 5 to 8). These are substrate limitation (f^{SL}) [23], non-competitive inhibition (f^{NC}) [10], inhibition of sulphate reduction by sulfide (f^{SS}) and inhibition

of sulphate reduction by protonated Volatile Fatty Acids (VFA) (f^{SA}) [31]. In these equations, C_{inh} is the inhibiting concentration (either a total or derived concentration), K_{inh} is the inhibition or half saturation constant and l is a shape parameter. Each mechanism gives an inhibition factor f ranging from 0 to 1. Multiplication of all inhibition factors acting on a biochemical reaction gives its total inhibition factor I (eq. 1).

$$f^{\mathrm{SL}}_{(C_{\mathrm{inh}})} = \frac{C_{\mathrm{inh}}}{C_{\mathrm{inh}} + K_{\mathrm{inh}}} \tag{5}$$

$$f^{\mathrm{NC}}_{(C_{\mathrm{inh}})} = \left(\frac{K_{\mathrm{inh}}}{K_{\mathrm{inh}} + C_{\mathrm{inh}}}\right)^{l} \tag{6}$$

$$f^{\mathrm{SS}}_{(C_{\mathrm{inh}})} = \left(1 - \frac{C_{\mathrm{inh}}}{K_{\mathrm{inh}}}\right)^{l} \tag{7}$$

$$f^{\mathrm{SA}}_{(C_{\mathrm{inh}})} = \left(1 + \left(\frac{C_{\mathrm{inh}}}{K_{\mathrm{inh}}}\right)^{l}\right)^{-1} \tag{8}$$

2.2.6. Selecting mechanistic parameter values from databases and derivation methods

Parameter values or bandwidths are entered into the spreadsheet where the model structure is defined. The geochemical parameters for the equilibrium reactions are automatically retrieved from an extensive geochemical database which is part of the Orchestra chemical equilibrium model ([21]). A wide variety in validated geochemical equilibria can be selected. As an option, activity correction can be calculated with the Davies equation.

Stoichiometry and rate parameters of biochemical metabolic reactions can be derived from thermodynamic principles with an additional spreadsheet. This derivation method is adapted from Kleerebezem and van Loosdrecht, (2010) [16]. It allows to constrain the model with established mechanistic information.

As an example, we illustrate the method with the derivation of a metabolic reaction for methanogenesis. Metabolism is a combination of catabolism (releasing energy) and anabolism (biological growth) which are coupled using the yield factor,

$$Y_{\mathrm{xs}} = \frac{\mathrm{CmolX}}{\mathrm{CmolS}} \tag{9}$$

which specifies how much growth (X) occurs from substrate (S) given the amount of energy generated by the catabolic reaction. CmolX is the number of moles of C in the biomass and CmolS is the number of moles of C in the substrate. The redox half reactions for catabolism can be written as

$$1\mathrm{C_2H_3O_2^-} + 2\mathrm{H_2O} \rightleftharpoons 2\mathrm{CO_2} + 7\mathrm{H^+} + 8\mathrm{e^-} \tag{10}$$

$$1C_2H_3O_2^- + 9H^+ + 8e^- \rightleftharpoons 2CH_4 + 2H_2O \tag{11}$$

which combine to the following catabolic reaction:

$$1C_2H_3O_2^- + 1H^+ \rightleftharpoons 1CO_2 + 1CH_4. \tag{12}$$

Following the same principle, the anabolic reaction is as follows,

$$0.5C_2H_3O_2^- + 0.3H^+ + 0.2NH_4^+ \rightleftharpoons 1CH_{1.4}O_{0.4}N_{0.2} + 0.6H_2O \tag{13}$$

where we assume the generic molecular composition of bacteria ($CH_{1.4}O_{0.4}N_{0.2}$) to be the same as used by Henze et al. (1995) [12].

The growth yield is calculated with the method proposed by Kleerebezem and van Loosdrecht, (2011). This is a partially empirical method based on the Gibbs reaction energy of the catabolic and anabolic reaction [11]. In this case, $Y_{xs} = 0.04$ which leads to the following total metabolic reaction:

$$\begin{aligned} &12.9C_2H_3O_2^- + 12.7H^+ + 0.2NH_4^+ \rightleftharpoons \\ &1CH_{1.4}O_{0.4}N_{0.2} + 12.4CO_2 + 12.4CH_4 + 0.6H_2O. \end{aligned} \tag{14}$$

In addition to the stoichiometry, also the rate parameters for metabolic reactions can be estimated from thermodynamic principles [11].

Estimated parameter values can be corrected for temperature using several relations. Maximum rates, equilibrium constants and the solubility of Calcite are corrected with equations 15-17 respectively [41] [28]. In these equations, T_e is the temperature of the environment, T_r is the reference temperature of the parameter, $\triangle H^0$ is the standard enthalpy, k is 64 $\frac{kJ}{mol}$, a is -171.9065, b is 0.077993, c is 2893.319 and d is 71.595.

$$\ln\left(\mu_e^{max}\right) = \ln\left(\mu_r^{max}\right) + \ln\left(\frac{k}{R} \cdot \left(\frac{1}{T_r} - \frac{1}{T_e}\right)\right) \tag{15}$$

$$\ln\left(K_e\right) = \ln\left(K_r\right) + \ln\left(\frac{\triangle H^o}{R \cdot T_r^2} \cdot \left(T_e - T_r\right)\right) \tag{16}$$

$$\log_{10}\left(K_{eq(Calcite)}\right) = a + b \cdot T_e + \frac{c}{T_e} + d \cdot \log_{10}\left(T_e\right) \tag{17}$$

2.3. Solving the model structure with a generic matrix calculation method

The model structure is automatically assembled in to a system of ordinary differential equations (ODE) with a generic matrix calculation method adapted from Reichel et al. (2007) [29]. This system of ODEs is fully coupled with the equilibrium calculator ORCHESTRA [21]. The ODEs are solved in MATLAB and the equilibrium calculator is part of the JAVA memory space of MATLAB. This makes

the algorithm very efficient since both solvers use the same memory space while information about the total and derived concentrations are exchanged via a JAVA link.

The set of ODEs for each phase can be written as

$$\frac{\mathrm{d}\boldsymbol{C}^{\mathrm{T,water}}}{\mathrm{d}t} - \boldsymbol{F}^{\mathrm{T,water}} = \boldsymbol{R}^{\mathrm{K}} \tag{18}$$

$$\frac{\mathrm{d}\boldsymbol{C}^{\mathrm{T,gas}}}{\mathrm{d}t} - \boldsymbol{F}^{\mathrm{T},gas} = 0 \tag{19}$$

where $\boldsymbol{F}^{\mathrm{T}}$ are the total fluxes into the respective phase and $\boldsymbol{R}^{\mathrm{K}}$ are the total biochemical rates. The total concentrations in the solid phase are included in the ODEs of the water phase and are expressed per volume of water phase. The total fluxes $\boldsymbol{F}^{\mathrm{T},water}$, $\boldsymbol{F}^{\mathrm{T},gas}$ and the total biochemical rates $\boldsymbol{R}^{\mathrm{K}}$ are calculated by vector summation over the number of fluxes or reactions j per process,

$$\boldsymbol{R}^{\mathrm{K}} = \boldsymbol{\mu}^{\max} \circ \boldsymbol{C}^{\mathrm{T}^{'}} \circ \boldsymbol{I} \cdot \boldsymbol{S} \tag{20}$$

$$\boldsymbol{F}^{\mathrm{T,water}} = \sum_{j} \boldsymbol{F}^{\mathrm{MT,water}} - \sum_{j} \boldsymbol{F}^{\mathrm{MF,water}} \tag{21}$$

$$\boldsymbol{F}^{\mathrm{T,gas}} = \sum_{j} \boldsymbol{F}^{\mathrm{MF,gas}} - \sum_{j} \boldsymbol{F}^{\mathrm{MT,gas}} - \boldsymbol{x} \cdot \sum \left(\sum_{j} \boldsymbol{F}^{\mathrm{MF,gas}} - \sum_{j} \boldsymbol{F}^{\mathrm{MT,gas}} \right) \tag{22}$$

where $\boldsymbol{\mu}^{\max}$, $\boldsymbol{C}^{\mathrm{T}^{'}}$ and $\boldsymbol{I}$ have size $1 \times j$, $\circ$ is the symbol for element-wise multiplication, $\boldsymbol{S}$ are the stoichiometry with size $j \times n_{\mathrm{l}}$, $\boldsymbol{F}^{\mathrm{MF,water}}$ and $\boldsymbol{F}^{\mathrm{MT,water}}$ have size $j \times n_{\mathrm{l}}$, $\boldsymbol{F}^{\mathrm{MF,gas}}$ and $\boldsymbol{F}^{\mathrm{MT,gas}}$ have size $j \times n_{\mathrm{g}}$ and $\boldsymbol{x}$ are the molar fraction of the total concentrations in the gas phase. The last term in $\boldsymbol{F}^{\mathrm{T,gas}}$ constrains the gas phase with a constant volume and a constant total pressure.

2.4. Evaluating model performance

2.4.1. Qualitative evaluation

Model performance is evaluated visually by automatically generated plots combining the model results with the measured data and plots of all modeled states. Because of the strong human capacity for pattern detection, these plots may indicate extreme, under- or non-modeled behavior. In addition, residuals are evaluated with QQ plots and auto-correlation plots.

2.4.2. Applying Bayesian inference

In order to find a model structure which provides the best fit to the data, we also evaluate the outcome of the simulation with several objective criteria. All these criteria are evaluated using the results of Bayesian inference applied to the set of most uncertain parameters ($\boldsymbol{\theta}$). The outcome of Bayesian inference,

$$p\left(\boldsymbol{\theta}|\hat{\mathbf{y}}\right) \propto p\left(\boldsymbol{\theta}\right) \cdot L\left(\boldsymbol{\theta}|\hat{\mathbf{y}}\right), \tag{23}$$

is the joint posterior probability distribution ($p(\boldsymbol{\theta}|\hat{\mathbf{y}})$) of the set of parameters given the measured data. This distribution reflects the uncertainty in the different parameters. The posterior distribution is calculated from the prior distribution of the parameters ($p(\boldsymbol{\theta})$) and the likelihood of the parameters in light of the measured data ($L(\boldsymbol{\theta}|\hat{\mathbf{y}})$). The posterior distribution is obtained using an adapted version of the DREAM$_{(\mathrm{ZS})}$ algorithm [44, 18] where algorithmic settings and parameters are set to the recommended and default values.

The toolbox uses a Gaussian objective function, which includes the set of standard deviations of total error ($\boldsymbol{\sigma}$) for each subset of data, to evaluate the likelihood,

$$\ln\left(L\left(\boldsymbol{\theta}|\hat{\mathbf{y}}\right)\right) = -\frac{n}{2}\cdot\ln\left(2\pi\right) - \sum\ln\left(\boldsymbol{\sigma}\right) - \frac{1}{2}\cdot\sum\left(\frac{\hat{\mathbf{y}}-\mathbf{y}}{\boldsymbol{\sigma}}\right)^2 \tag{24}$$

where $\mathbf{y}$ is the modeled data and $\boldsymbol{\sigma} = [\boldsymbol{\sigma}^{\mathrm{subset1}}_{(m\times 1)}; \boldsymbol{\sigma}^{\mathrm{subset2}}_{(m\times 1)}; ...]$ with m being the size of a subset of data and all entries in $\boldsymbol{\sigma}_{(m\times 1)}$ have the same value σ^{subset} for that subset. Because it is often difficult to estimate the measurement error, and it is more or less impossible to estimate the model error, we chose to expand the set of uncertain parameters $\boldsymbol{\theta}$ with the standard deviations σ^{subset} for each dataset.

Prior distributions of uncertain model parameters are assumed to be uniform with an initial search range that is 100× wider to ensure global convergence. Prior distributions of standard deviations are also uniform ranging from $\frac{1}{5}\cdot\sigma^{\mathrm{subset}}$ to $5\cdot\sigma^{\mathrm{subset}}$. Ranges in the prior distributions can be different for other problems.

2.4.3. Quantitative criteria

The result from the Bayesian inference allows us to quantitatively compare the performance of different model structures. Since our modeling objective is to find the model structure with minimal (mechanistic) uncertainty, we implemented criteria that quantify aspects related to this. For other modeling objectives, other criteria/metrics [4, 43] may be more suited which are readily implemented.

One criterion is the difference in best fit or lowest total error (measurement error & model error) between model structures. In addition, we can also assess the probability distribution of total errors which is related to the combined probability distributions in calibrated parameters. The latter directly reflects the uncertainty of the complete model structure in light of the measured data. Even more, the uncertainty of total error per subset of data can be compared between models with the marginal probability distributions of the standard deviations.

A second criterion is the uncertainty in calibrated parameters for different model structures. Parameter uncertainty is quantified by the width of its marginal posterior distribution. A wider distribution indicates more uncertainty under the condition of the same prior distribution. For a quantitative

comparison, the gain of information from marginal prior to marginal posterior for parameter i can be quantified with the Kullback-Leibler divergence (D_{KL}). This metric,

$$D_{\mathrm{KL}} = \int \left(p\left(\theta_i|\hat{\boldsymbol{y}}\right) \cdot \ln\left(\frac{p\left(\theta_i|\hat{\boldsymbol{y}}\right)}{p\left(\theta_i\right)}\right) \cdot \mathrm{d}\theta_i, \tag{25}$$

is a non-symmetric measure of the dissimilarity between two probability distributions. Higher values of D_{KL} indicate a larger information gain and therefore less uncertainty.

A third criterion is the presence of correlations between parameters in the model structure which can be seen in the marginal posterior distributions. These correlations point to possibilities for optimizing model structure by adding restrictions or lumping of correlated parameters.

A fourth criterion is the mechanistic completeness of model structure. Mechanistic completeness is evaluated by comparing the calibrated parameter bandwidths (i.e. 5%-95% quantiles) with 'ideal' parameter values measured under non-limiting conditions. A close match between calibrated parameters and 'ideal' values indicates a mechanistically complete model structure. Parameters that strongly deviate from 'ideal' values indicate missing mechanistic processes.

A final criterion is the amount of information a model structure provides given its complexity. More complex models with a large number of parameters can be compared with simpler ones with the marginalized likelihood (L^{m}) which is the probability of the measured data given the model structure, not assuming any particular model parameters. It can be approximated with the harmonic mean of likelihoods [24],

$$L^{\mathrm{m}} = \left(\frac{1}{N}\sum_{i=1}^{N} L\left(\boldsymbol{\theta}|\hat{\boldsymbol{y}}\right)^{-1}\right)^{-1} \tag{26}$$

where N is the number of likelihoods. The model structure with the highest marginal likelihood has the best balance between information content and model complexity.

3. Results

3.1. The final four evaluated biogeochemical reaction networks

The dataset we used for testing our toolbox was obtained from Roberto Valencia [39, 36]. We present the final four model structures that were the outcome of our search for an optimal model with the methodology provided by the toolbox. These four reaction networks are schematically depicted in Figure 3. In this figure, the most simple network 1 is presented with black lines. Our hypothesis for this network is that four kinetic reactions control the measured emissions. The first one is hydrolysis lumped with acidogenesis. It converts the available biodegradable Solid Organic Matter (SOM) into a mix of Volatile Fatty Acids ($\mathrm{VFA_{mix}}$) with hydrolysis as rate limiting step. The second one is

methanogenesis which converts the VFA_{mix} into methane and carbon dioxide. The final two reactions are decay of the two types of bacteria involved in lumped hydrolysis and methanogenesis with rates which are 5% of the maximum growth rates [1]. Furthermore, only methanogenesis is influenced by substrate limitation of VFA_{mix}. The stoichiometry of the kinetic reactions is listed in table 1. We included the most common geochemical equilibrium reactions assuming a readily available excess of Calcite to represent the high alkalinity of MSW. In addition, we assume that the exchange between concentrations in the gas and liquid phases is instantaneous.

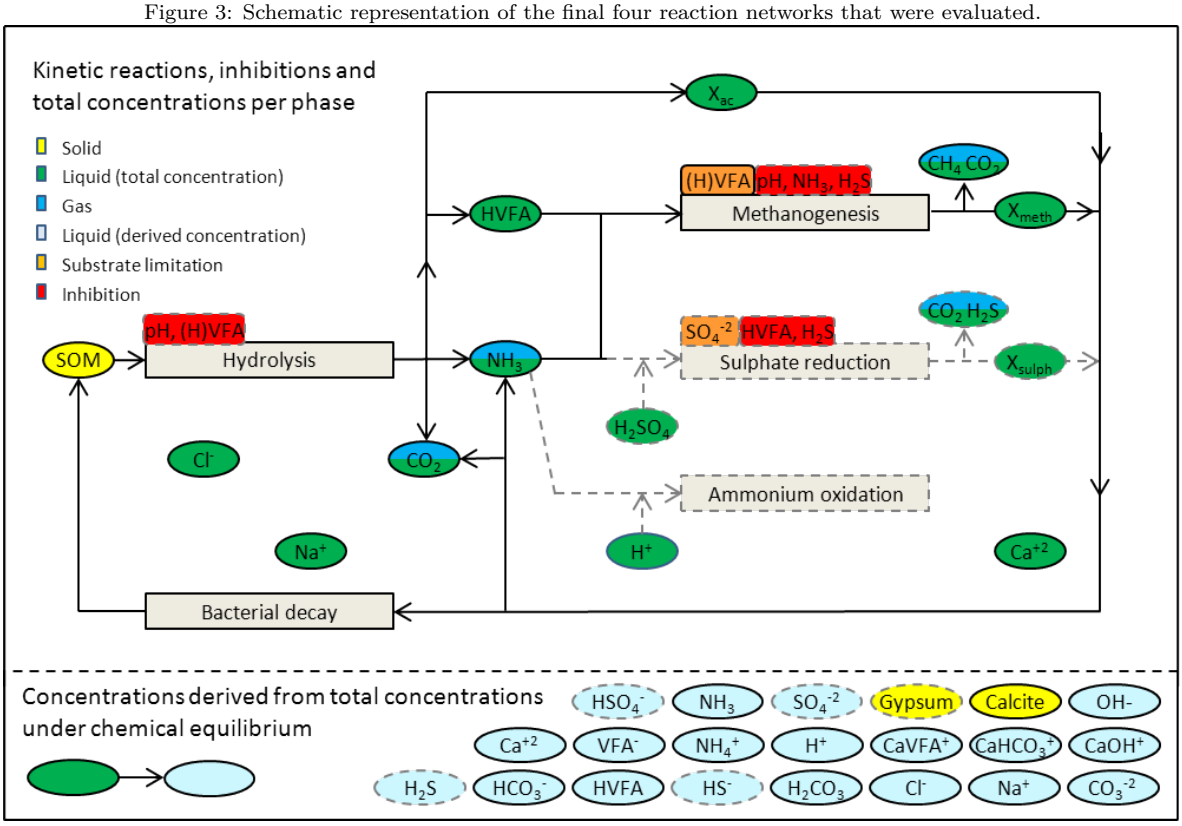
According to the experimental design, gas vents to the atmosphere and no other transport flows are included. Because leachate was recirculated during the experiment, we simplified the problem to ideally mixed batch conditions. Model parameters which we considered to be well known are taken from literature. These parameter values are listed in table 2 together with the initial experimental conditions. The model parameters which we considered to be unknown or uncertain are listed in table 3 together with the prior ranges found in literature. These unknown parameters in the model's reaction network are obtained via Bayesian inference which fits the model outputs to the measurements.

For reaction network 2, we increased the complexity by adding the main environmental inhibitions that are known to influence the kinetic reactions in wastewater treatment. These inhibitions are non-competitive inhibition of hydrolysis by pH and total concentration of VFA_{mix} and non-competitive inhibition of methanogenesis by pH and ammonia.

In reaction network 3, we further increased the complexity by adding a kinetic ammonium oxidation reaction. We implemented this reaction to find an explanation for the decreasing measured ammonium concentration in time which normally does not occur under anaerobic conditions.

Finally, network 4 extends the reaction network further in order to include sulphate and sulfide. These compounds are usually present in anaerobic environments and are known to inhibit methanogenesis and oxidize available biodegradable SOM. Adding mechanistic information about these processes may therefore optimize model performance. The following is included: 1) a kinetic sulphate reduction reaction, 2) corresponding equilibrium reactions, 3) inhibition of methanogenesis by H_2S and 4) inhibition of sulphate reduction by H_2S and protonated VFA_{mix}.

Figure 3: Schematic representation of the final four reaction networks that were evaluated.



The most basic reaction network 1 is presented with the black lines. Complexity is increased stepwise from network 1-4 which is indicated by the dashed gray lines. Network 2 includes inhibition of hydrolysis by pH and $C^{T}_{VFA_{mix}}$ and inhibition of methanogenesis by pH and ammonia. In network 3, an ammonium oxidation reaction is added. Network 4 is extended with sulphate reduction for which total concentrations, derived concentrations and inhibitions are included accordingly. X represents bacterial biomass. Compounds that are in equilibrium between the gas phase and the liquid phase have two colors.

Table 1: Stoichiometry for the total concentrations in the biochemical reactions

	$SOM^{3)}$	$VFA_{mix}{}^{3)}$	H_2CO_3	NH_3	H_2O	CH_4	$SO_4{}^{-2}$	H_2S	X_{acid}	X_{meth}	X_{sulph}	H^+
Hydrolysis[1)]	-1	0.3	0.01	0.024	-0.088	-	-	-	0.18	-	-	-
Methanogenesis	-	-6.67	7.23	-0.2	-8.76	9.37	-	-	-	1	-	-
Decay[2)]	0.92	-	0.08	0.14	-0.42	-	-	-	(-1)	(-1)	(-1)	-
Ammonium oxidation	-	-	-	-1	-	-	-	-	-	-	-	-1
Sulphate reduction		-10.6	27.1	-0.2	-27.1		-15.3	15.3			1	-30.5

The gray cells highlight the reaction driving concentrations. 1) In this reaction, the growth yield for acidogenesis on glucose is used: $Y_{XS} = 1.09$. 2) Each bacterial biomass (X) decays individually. 3) The elemental compositions of SOM is $CH_{1.79}O_{0.63}N_{0.06}$ and VFA_{mix} is $C_{2.64}H_{5.28}O_2$ which were derived from the data measured in the lysimeter experiments.

Table 2: Model parameters and initial conditions

			K_{inh}		l		Initial conditions			
μ^{max}_{aci}	0.12	[1]	$f^{NC}_{(hyd,C^D_{H^+})}$	1×10^{-5}	2	[34]	$C^{T1)}_{SOM}$	5.44	$C^T_{Cl^-}$	0.1
μ^{max}_{decay}	$0.05 \cdot \mu^{max}_{growth}$	[1]	$f^{NC}_{(hyd,C^T_{VFA})}$	2.34×10^{-2}	1	[34]	$C^T_{SO_4{}^{-2}}$	0.16	V_g	250
K_{H,CH_4}	1.5×10^{-3}	[2]	$f^{NC}_{(meth,C^D_{H^+})}$	5×10^{-7}	2	[34]	$C^T_{NH_3}$	0.065	p_{tot}	1
K_{H,NH_3}	71.4	[2]	$f^{NC}_{(meth,C^D_{NH_3})}$	1.5×10^{-3}	1	[34]	$C^T_{H_2CO_3}$	1.13	T	303
K_{H,H_2S}	0.11	[2]	$f^{NC}_{(meth,C^D_{H_2S})}$	4.29×10^{-2}	1	[27]	$C^T_{Ca^{+2}}$	1.16	pH	6.15
K_{H,H_2O}	2.3×10^3	[2]	$f^{SS}_{(sulph,C^D_{H_2S})}$	1.61×10^{-2}	0.401	[31]	$C^T_{Na^+}$	0.2	V_l	325
K_{H,CO_2}	0.03	[2]	$f^{SA}_{(sulph,C^D_{HVFA})}$	9×10^{-4}	1.08	[31]				

1) Initial concentration of SOM is taken from the total amount of carbon produced as biogas in the experiment. Units of maximum rates, Henry constants, inhibition constants, concentrations, temperatures, pressures and volumes are respectively d^{-1}, $\frac{mol}{L \cdot atm}$, $\frac{mol}{L}$, $\frac{mol}{L}$, atm, K and L. All parameter values are corrected for 303K except $K_{inh,C^D_{H_2S}}$ (298 K).

3.2. Outcome of the Bayesian inference applied on the four networks

The result of the Bayesian inference allows us to evaluate the quality of performance of our four model structures. We start with the most basic evaluation, visual inspection of fit. Figure 4 presents the modeled data with the highest likelihood (in red) and the related uncertainty bandwidths (in green) together with the measured data (in blue) for each network. Visual inspection clearly shows that some networks perform better than others. The fit for the cumulative biogas is better in networks $2-4$ and the fit for ammonium concentration is better in networks $3-4$. In addition, the uncertainty is lower in networks $2-4$ indicated by slightly narrower green bandwidths.

Visual judgment, however, reveals very little about the mechanistic correctness and parameter uncertainty in the models. The D_{KL} values in table 3 indicate the uncertainty of the calibrated parameters relative to the different model structures. Higher values mean lower uncertainty. The uncertainty in the calibration of $K_{\mathrm{inh},C_{\mathrm{VFA}}^{\mathrm{T}}}^{\mathrm{SL(meth)}}$, for example, is much lower for network $2-4$ than network 1. Preferably, we select a model with a low uncertainty in all calibrated parameters. The uncertainty in the total error for each subset of data is shown per network by the D_{KL} values of the standard deviations. This shows for instance that the uncertainty for the subset of ammonium is much lower in networks $3-4$ than networks $1-2$.

The posterior bandwidths ($5\%-95\%$ quantiles) in table 3 reveal how close the calibrated parameters are to 'ideal' values measured under non-limiting conditions. Parameters that are close are indicated in bold. Ideally for mechanistic completeness, all calibrated parameters should be close to 'ideal' values. Network $2-4$ are therefore more mechanistic correct than network 1 because the calibrated bandwidths of $\mu_{\mathrm{hyd}}^{\mathrm{max}}$ and $\mu_{\mathrm{meth}}^{\mathrm{max}}$ are closer to 'ideal' values.

Most parameters in the networks are not correlated except for the maximum rate and the half saturation constant of methanogenesis in the less complex networks. These correlations are presented in Figure 5. Interestingly, the correlation becomes weaker as the complexity and mechanistic information content of the model structures increases.

The log-normal marginal likelihoods-values in figure 4 show the performance of the networks with respect to the balance of information content (model complexity) versus model uncertainty. They show that although network 4 is most complex, the added information improves the maximum likelihood without increasing model uncertainty too much.

When evaluating model uncertainty based on statistics, it is important to check if the statistical assumptions applied are valid. This means that the probability of the residuals should be normally distributed because we apply a Gaussian likelihood. We check this with Q-Q plots and normalized autocorrelation functions (ACF) for the residuals of the best fit results of network 4 which are presented

in figure 6. The Q-Q plots show residual quantiles (in blue) vs quantiles from the theoretical normal distribution. When the residual quantiles are normally distributed they should coincide with the theoretical red line. Apparently, not all residuals are normally distributed. Therefore, some error is present in the statistical evidence we apply to evaluate our models. The same violation of normal distribution is indicated by the autocorrelation of the residuals. However, for the type of evaluations we apply this statistical error has minimal impact. If one is interested in a more statistical correct evaluation, we suggest to use the general likelihood function presented by Schoups and Vrugt et al. (2010)[33].

Figure 4: Measured data and best fit model results for the four reaction networks

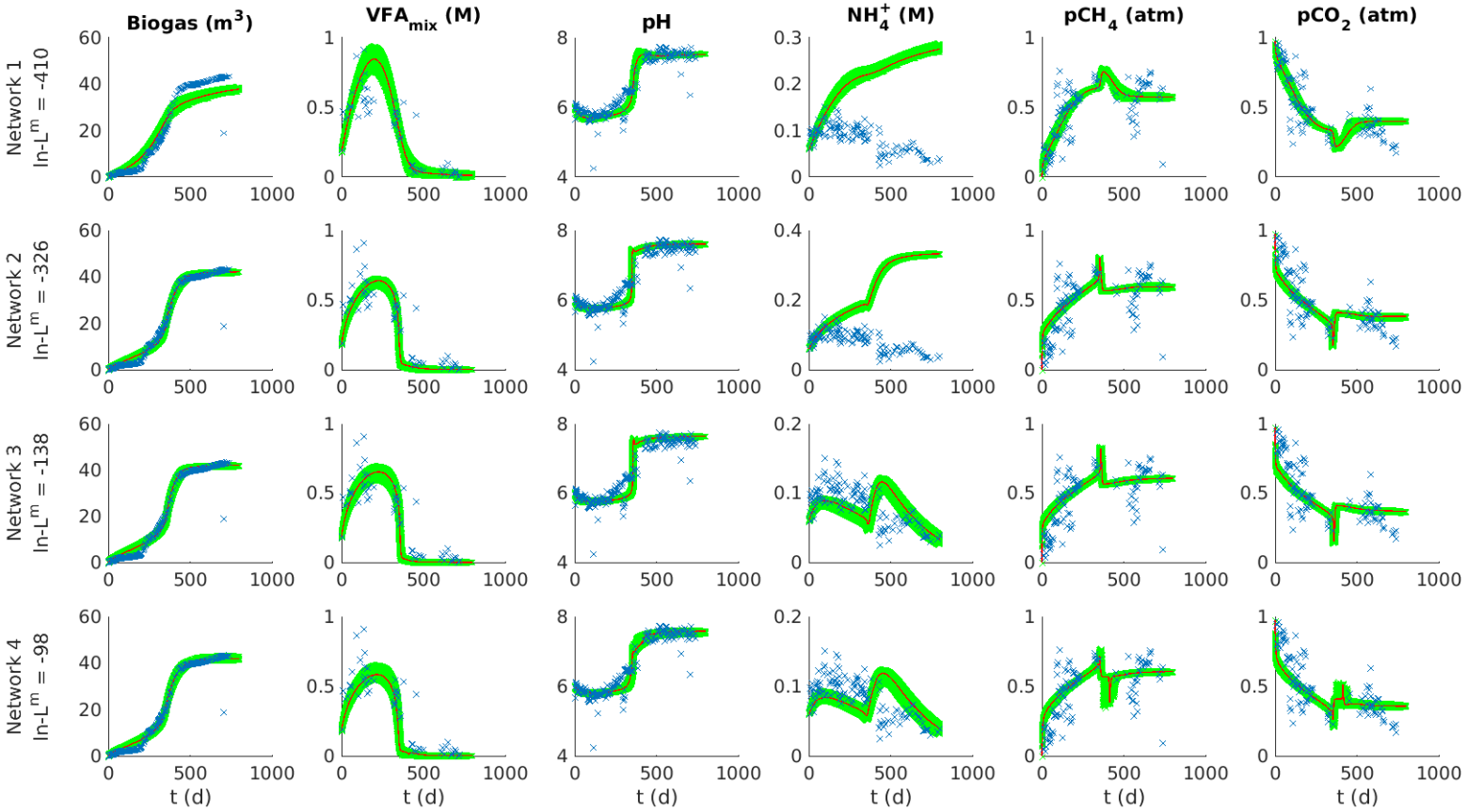


Table 3: The prior ranges, posterior ranges ($5\% - 95\%$ quantiles) and information gain (D_{KL}) for the inferred parameter per network

	$\mu_{\mathrm{hyd}}^{\mathrm{max}}$ (d^{-1})		$\mu_{\mathrm{meth}}^{\mathrm{max}}$ (d^{-1})		$K_{\mathrm{inh},C_{\mathrm{VFA}}^{\mathrm{T}}}^{\mathrm{SL(meth)}}$ (mM)		$C_{\mathrm{X_{meth}}}^{\mathrm{T(ini)}}$ (mM)	
	quantiles	D_{KL}	quantiles	D_{KL}	quantiles	D_{KL}	quantiles	D_{KL}
Prior	0.09-0.26 [41]		0.04-0.47 [22]		0.03-420 [22]		0.27-19 [25]	
Network 1	0.004-0.0045	10.6	0.013-0.029	7.72	330-1318	2.12	**8.6-13.5**	5.70
Network 2	0.052-0.060	7.84	**0.127-0.199**	6.24	**60-152**	4.54	**8.8-12.6**	5.97
Network 3	0.056-0.069	7.30	**0.095-0.167**	6.23	**95.5-184**	4.54	**12.9-25.3**	4.83
Network 4	0.050-0.061	7.55	**0.095-0.145**	6.63	**152-259**	4.24	**14.5-25.8**	4.93

	$\mu_{\mathrm{NH_4}}^{\mathrm{max}}$ (d^{-1})		$\mu_{\mathrm{sulph}}^{\mathrm{max}}$ (d^{-1})		$K_{\mathrm{inh},C_{\mathrm{SO_4}^{-2}}^{\mathrm{T}}}^{\mathrm{SL(sulph)}}$ (mM)		$C_{\mathrm{X_{sulph}}}^{T(ini)}$ (mM)	
	quantiles	D_{KL}	quantiles	D_{KL}	quantiles	D_{KL}	quantiles	D_{KL}
Prior	0.001-1		0.024-2.4 [31]		0.178-0.26 [30]		0.27-19 [25]	
Network 1								
Network 2								
Network 3	0.0044-0.0052	6.94						
Network 4	0.0043-0.0052	6.84	16.47-210	0.26	952-25085	1.06	0.095-120	2.95

	σ_{biogas} (m^3)		σ_{VFA} (mM)		σ_{pH}		$\sigma_{\mathrm{NH_4}^+}$ (mM)	
	quantiles	D_{KL}	quantiles	D_{KL}	quantiles	D_{KL}	quantiles	D_{KL}
Prior	3.41-85.2		0.05-1.24		0.16-3.95		0.005-0.136	
Network 1	4.19-5.14	8.86	0.13-0.19	7.47	0.24-0.28	8.82	0.12-0.15	5.93
Network 2	2.12-2.52	9.76	0.12-0.17	7.62	0.27-0.32	8.86	0.14-0.17	5.76
Network 3	2.19-2.62	8.68	0.12-0.17	7.65	0.27-0.32	8.18	0.029-0.036	7.24
Network 4	2.06-2.49	9.64	0.12-0.17	7.58	0.23-0.26	9.11	0.029-0.038	7.06

	$\sigma_{\mathrm{pCO_2}}$ (atm)		$\sigma_{\mathrm{pCH_4}}$ (atm)	
	quantiles	D_{KL}	quantiles	D_{KL}
Prior	0.04-1.01		0.04-1.05	
Network 1	0.12-0.14	7.99	0.11-0.13	8.10
Network 2	0.13-0.16	7.89	0.13-0.16	7.91
Network 3	0.13-0.17	7.87	0.13-0.16	7.82
Network 4	0.13-0.17	7.82	0.13-0.16	7.79

Quantiles that are comparable with 'ideal' values measured under non-limiting conditions are presented in bold.

Figure 5: Correlations between parameters per network

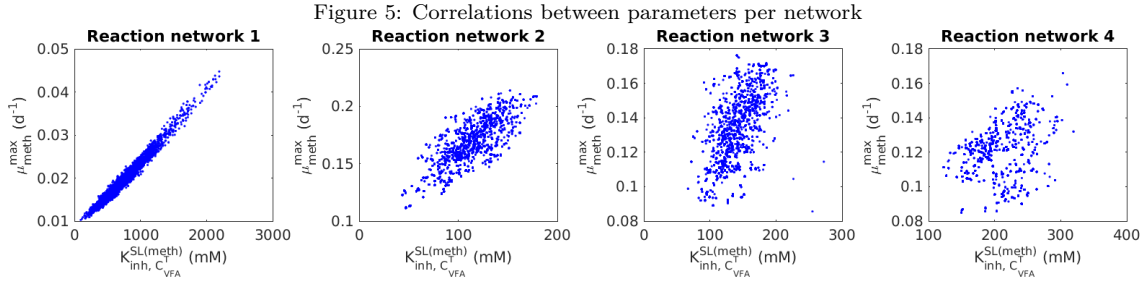
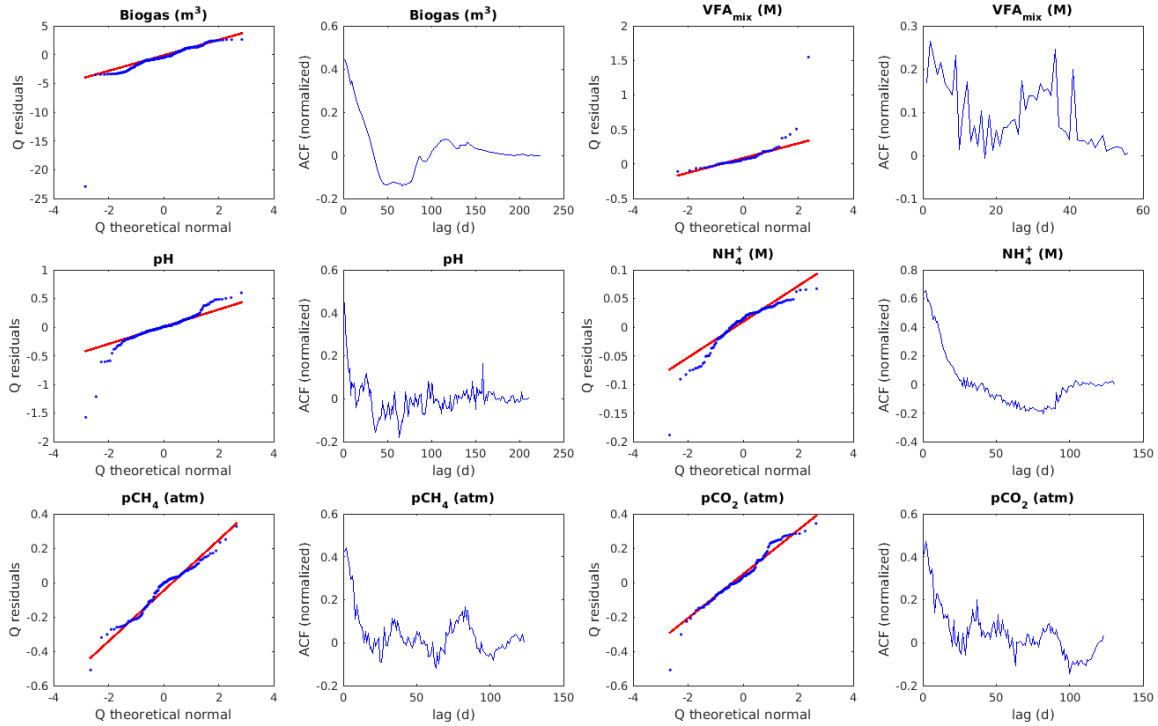


Figure 6: Q-Q plots and autocorrelation functions (ACF) per sub dataset of the best fit residuals of reaction network 4



4. Discussion & Conclusions

The aim of the methods combined in the toolbox is to choose an optimal network with which we can describe anaerobic digestion of MSW. We believe this novel approach helps to reduce the ambiguity in model structure and calibrated parameter values reported in literature for a given environmental system. It strengthens (qualitative) judgment of model performance by providing quantitative criteria for evaluating (mechanistic) model uncertainty and it allows integrated environmental assessment. As a result, we obtain model structures which allow us to make predictions with a higher certainty because we have included more (objective) information for selecting the reaction network.

4.1. Performance of network 1

The most basic reaction network 1 reasonably reproduced a significant portion of the measured data (figure 4) with low uncertainty in total error (narrow green bandwidths). The model captures the values and trends of measured VFA_{mix}, pH, pCH_4 and pCO_2 in time. This good fit was achieved by selecting a set of reactions with the correct stoichiometry and parameter values from well established anaerobic reaction networks from literature, databases and derivation methods. The toolbox optimizes the implementation of such networks.

The model based on reaction network 1, however, does not provide a nice fit to all subsets of data. Using the results from the Bayesian inference, we modified the model structure in order to improve the results. Figure 5 shows that $K_{\mathrm{inh},C_{\mathrm{VFA}}^{\mathrm{T}}}^{\mathrm{SL(meth)}}$ is strongly correlated with $\mu_{\mathrm{meth}}^{\mathrm{max}}$. This suggests that both parameters can be lumped together or that another process is being compensated for by the correlation of these two parameters. The results also show that the optimized values of probability distributions of the maximum rates are low compared to 'ideal' values. We concluded that limitations or inhibitions are missing in the reaction network which reduce prediction accuracy and cause the poor fit for cumulative biogas and ammonium.

4.2. Performance of network 2

Network 2 is extended with reaction terms which include the main environmental inhibitions acting on hydrolysis and methanogenesis. Assessing the results for this network indicate that this extension significantly improves the model fit of the cumulative biogas. In addition the calibrated maximum rates are in closer agreement with 'ideal' values. We conclude that the mechanistic description is more correct, while all parameters are still identifiable within sharp bandwidths (although some with slightly broader probability distributions than in network 1).

Adding the inhibition terms weakened the correlation between $K_{\mathrm{inh},C_{\mathrm{VFA}}^{\mathrm{T}}}^{\mathrm{SL(meth)}}$ and $\mu_{\mathrm{meth}}^{\mathrm{max}}$. The value of $K_{\mathrm{inh},C_{\mathrm{VFA}}^{\mathrm{T}}}^{\mathrm{SL(meth)}}$ is still high compared to 'ideal' values which indicates that this parameter compensates for a missing mass transport limitation in the model.

The difference in calibrated values between network 1 and network 2 nicely illustrates how easily ambiguity is created in published parameter values and outcomes of models [22]. By assuming a slightly different model structure, different sets of calibrated values will be reported which in turn may be used in yet again different model structures leading to wrong predictions. The only way to reduce this ambiguity is to report calibrated parameters together with the reaction networks and model structures from which they have been defined. Ideally the reactions should be mechanistically correct and the resulting parameters should be identifiable (i.e. possessing a narrow probability distribution without being correlated to other parameters). For this, statistical evaluation of model outcome and

parameter distributions is necessary because basic evaluation by eye does not give the information needed to judge mechanistic correctness and identifiability.

In the case of reaction network 2, the evaluation indicates a structure that is nearly correct. However, the concentration of ammonium is still overestimated by the model. This is to be expected since hydrolysis releases ammonium and the anaerobic reaction network 2 does not include an ammonium removal process. The experimental results clearly indicate that such a process occurs in the lysimeter.

4.3. Performance of network 3

In network 3, a relatively simple ammonium oxidation process is added to the reaction network. As a result, the model fit of the ammonium concentration improved drastically, while the rest of the model performance stayed similar to that of network 2. The modeled pH and the calibrated parameter bandwidths are not significantly influenced. The results indicate that in this lysimeter experiment an ammonium oxidation rate of 0.0048 d^{-1} is likely. The ammonium oxidation in the experiment was most likely caused by intrusion of oxygen when leachate was recirculated.

4.4. Performance of network 4

Although the results of network 3 are very good, we chose to do one more iteration. The waste used in the lysimeter most probably contained a source of sulphate (gypsum from building waste). This implies that sulphate reduction can occur which may explain the delay in onset of biogas production observed in the measured data. Although neither sulphate or sulphide were measured in this experiment, the results indicate an improved fit because both the errors in the fit of cumulative biogas and pH decreased. The final parameter distributions of the other parameters remained the same as found for network 3. The calibrated parameter values for sulphate reduction have a wide range and are high compared to 'ideal' values so we have to be careful in judging the mechanistic correctness of this approach. The improved datafit can be explained by the mechanism of sulphate reduction but it just as well can be attributed to the additional degrees of freedom added by the extra reaction. Measurements of sulphate and sulphide species are necessary to make a conclusion about how to improve the reaction network.

4.5. Selecting the optimal model structure for anaerobic digestion of MSW

4.5.1. For lysimeter scale

Based on the evaluation of network performances, we choose reaction network 3 as the optimal model structure for anaerobic digestion of MSW on a lysimeter scale. All parameters are either based on thermodynamic principles, taken from well established geochemical databases or calibrated with bandwidths which are in agreement with 'ideal' values. This strong mechanistic character of the

model together with its low uncertainty in calibrated parameters increases confidence in (long-term) prediction accuracy compared to other model structures.

For the value of $K_{\mathrm{inh},C^{\mathrm{T}}_{\mathrm{VFA}}}^{\mathrm{SL(meth)}}$, we suggest to either keep it as a parameter to be calibrated as a 'material property' for the system under investigation or to further extend the model with a transport limitation mechanism such as diffusion with an extra transport parameter which then needs to be calibrated.

Network 4 is not preferred over network 3 because the mechanism of sulphate reduction can not be adequately calibrated with this dataset. The improved likelihood is not guaranteed when applying this reaction network to other experiments and conditions. Calibration of network 4 with an extended measured dataset including sulphate or sulfide measurements would be very interesting.

Another interesting option would be to include the measurement error in the likelihood function. This was not possible for this dataset since the measurement error was unknown. Nevertheless, when included a distinction between model and measurement error could be made. A fully correct model would by necessity include all measured data within the green uncertainty bandwidths in figure 4.

4.5.2. For full scale

For full scale application, we select network 2 as the optimal model structure. Network 2 is preferred over network 3 because the ammonium oxidation in network 3 is modeled with a non-mechanistic first order decay function based on the observations only. In this experiment, air intrusion during recirculation of leachate is causing this degradation. This is probably not representative for large scale anaerobic digestion of waste in landfills.

5. Acknowledgments

This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs. We thank Roberto Valencia for providing us with the original data from his lysimeter experiments. We also thank Jasper Vrugt for providing us with a very efficient DREAM$_{\mathrm{ZS}}$ module which is included in the toolbox. Furthermore, we like to thank Jasper Vrugt and the anonymous reviewers for their comments which helped us to greatly improve the paper.

6. Supporting Information

The toolbox is implemented in MATLAB[19] and the algorithm is available at DOI: 10.4121/uuid:aefbaeb9-85be-4fa6-a29e-a9cb71d0fb3a. Here, also a detailed manual is provided on how to use the toolbox. The toolbox can be used without the requirement of special MATLAB toolboxes. Ordinary differential equations are solved with ode15s. For the Bayesian inference, an efficient submodule of DREAM$_{ZS}$ is

used and provided by Jasper Vrugt. A single integration of a network takes approximately 2.5 seconds and a Bayesian inference run with 250.000 iterations takes approximately 1 week. Inference run time can be substantially decreased ($\frac{1}{3}$) when Markov chains are run in parallel.

[1] Angelidaki, I., Ellegaard, L., Ahring, B.K., 1999. A comprehensive model of anaerobic bioconversion of complex substrates to biogas. Biotechnology and bioengineering 63, 363–72. URL: http://www.ncbi.nlm.nih.gov/pubmed/10099616.

[2] Atkins, P., de Paula, J., 2011. Physical chemistry for the life sciences. Freeman, W.H. URL: http://books.google.nl/books/about/Physical_Chemistry_for_the_Life_Sciences.html?id=WPwA3E0XsOQC&

[3] Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S.V., Pavlostathis, S.G., Rozzi, A., Sanders, W.T.M., Siegrist, H., Vavilin, V.A., 2002. The IWA Anaerobic Digestion Model No 1 (ADM1). Water science and technology : a journal of the International Association on Water Pollution Research 45, 65–73. URL: http://www.ncbi.nlm.nih.gov/pubmed/12188579.

[4] Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environmental Modelling \& Software 40, 1–20. URL: http://dx.doi.org/10.1016/j.envsoft.2012.09.011, doi:10.1016/j.envsoft.2012.09.011.

[5] Bethke, C., 2008. Geochemical and biogeochemical reaction modeling. Cambridge University Press, Cambridge U.K. URL: http://www.langtoninfo.com/web_content/9780521875547_frontmatter.pdf.

[6] Fellner, J., Döberl, G., Allgaier, G., Brunner, P.H., 2009. Comparing field investigations with laboratory models to predict landfill leachate emissions. Waste management (New York, N.Y.) 29, 1844–51. URL: http://www.ncbi.nlm.nih.gov/pubmed/19171473, doi:10.1016/j.wasman.2008.12.022.

[7] Garcia de Cortazar, A.L., Monzon, I.T., 2007. MODUELO 2: A new version of an integrated simulation model for municipal solid waste landfills. Environmental Modelling \& Software 22, 59–72. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364815205002215, doi:10.1016/j.envsoft.2005.11.003.

[8] Gawande, N.A., Reinhart, D.R., Gour-Tsyh, Y., 2010. Modeling microbiological and chemical processes in municipal solid waste bioreactor, Part II:

Application of numerical model BIOKEMOD-3P. Waste management (New York, N.Y.) 30, 211–8. URL: http://www.ncbi.nlm.nih.gov/pubmed/19815404 http://www.ncbi.nlm.nih.gov/pubmed/19819123, doi:10.1016/j.wasman.2009.09.011.

[9] Gönüllü, M., 1994. Analytical modelling of organic contaminants in leachate. Waste management \& research URL: http://wmr.sagepub.com/content/12/2/141.short http://wmr.sagepub.com/content/12/4/339.short.

[10] Haldane, J., 1930. Enzymes. Journal of the Society of Chemical Industry 49, 919–920.

[11] Heijnen, J.J., Kleerebezem, R., 1999. Bioenergetics of microbial growth. Encyclopedia of Bioprocess Technology , 594–616URL: http://onlinelibrary.wiley.com/doi/10.1002/0471250589.ebt026/full.

[12] Henze, M., Harremoes, P., Jansen, J., Arvin, E., 1995. Wastewater Treatment - Biological and Chemical Processes.

[13] Kamalan, H., Sabour, M., Shariatmadari, N., 2011. A review on available landfill gas models. Journal of Environmental Science and Technology 4, 79–92. URL: http://agris.fao.org/agris-search/search/display.do?f=2012/DJ/DJ201229164103.xml;DJ2012029189.

[14] Karaca, F., Özkaya, B., 2006. NN-LEAP: A neural network-based model for controlling leachate flow-rate in a municipal solid waste landfill site. Environmental Modelling & Software 21, 1190–1197. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364815205001155, doi:10.1016/j.envsoft.2005.06.006.

[15] Kelly, R.A., Jakeman, A.J., Barreteau, O., Borsuk, M.E., ElSawah, S., Hamilton, S.H., Henriksen, H.J., Kuikka, S., Maier, H.R., Rizzoli, A.E., van Delden, H., Voinov, A.A., 2013. Selecting among five common modelling approaches for integrated environmental assessment and management. Environmental Modelling \& Software 47, 159–181. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-84879530766&partnerID=40&md5=adecc1fbe25545cf5 doi:10.1016/j.envsoft.2013.05.005.

[16] Kleerebezem, R., Van Loosdrecht, M.C.M., 2010. A Generalized Method for Thermodynamic State Analysis of Environmental Systems. Critical Reviews in Environmental Science and Technology 40, 1–54. URL: http://www.tandfonline.com/doi/abs/10.1080/10643380802000974, doi:10.1080/10643380802000974.

[17] Kouzeli-Katsiri, A., Bosdogianni, A., Christoulas, D., 1999. Prediction of leachate quality from sanitary landfills. Journal of Environmental Engineering 125, 950–958. URL: http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9372(1999)125:10(950).

[18] Laloy, E., Vrugt, J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing. Water Resources Research 48, W01526. URL: http://www.agu.org/pubs/crossref/2012/2011WR010608.shtml, doi:10.1029/2011WR010608.

[19] MATLAB Release 2015a, . The MathWorks, Inc.

[20] McDougall, J., 2007. A hydro-bio-mechanical model for settlement and other behaviour in landfilled waste. Computers and Geotechnics 34, 297–320. URL: http://www.sciencedirect.com/science/article/pii/S0266352X07000158 http://linkinghub.elsevier.com/retrieve/pii/S0266352X07000158, doi:10.1016/j.compgeo.2007.02.004.

[21] Meeussen, J.C.L., 2003. ORCHESTRA: an object-oriented framework for implementing chemical equilibrium models. Environmental science & technology 37, 1175–82. URL: http://www.ncbi.nlm.nih.gov/pubmed/12680672.

[22] Meima, J.A., Naranjo, N.M., Haarstrick, A., 2008. Sensitivity analysis and literature review of parameters controlling local biodegradation processes in municipal solid waste landfills. Waste management (New York, N.Y.) 28, 904–18. URL: http://www.ncbi.nlm.nih.gov/pubmed/17499984, doi:10.1016/j.wasman.2007.02.032.

[23] Monod, J., 1949. The growth of bacterial cultures. Annual Reviews in Microbiology URL: http://www.annualreviews.org/doi/pdf/10.1146/annurev.mi.03.100149.002103.

[24] Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. Journal of the Royal Statistical Society 56, 3–48. URL: http://www.jstor.org/stable/2346025.

[25] Nopharatana, A., Pullammanappallil, P.C., Clarke, W.P., 2007. Kinetics and dynamic modelling of batch anaerobic digestion of municipal solid waste in a stirred reactor. Waste management (New York, N.Y.) 27, 595–603. URL: http://www.ncbi.nlm.nih.gov/pubmed/16797956, doi:10.1016/j.wasman.2006.04.010.

[26] Ozkaya, B., Demir, A., Bilgili, M., 2007. Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. Environmental Modelling \& Software 22, 815–822. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364815206000776, doi:10.1016/j.envsoft.2006.03.004.

[27] Paula Jr, D.R., Foresti, E., 2009. Sulfide toxicity kinetics of a uasb reactor. Brazilian Journal of Chemical Engineering 26, 669–675. URL: http://www.scielo.br/scielo.php?pid=S0104-66322009000400005&script=sci_arttext.

[28] Plummer, L.N., Busenburg, E., 1982. The solubilities of calcite, aragonite and vaterite in CO2-H2O solutions between 0 and 90 C, and an evaluation of the aqueous model for the system CaCO3-CO2-H2O. Geochimica et cosmochimica acta 46.

[29] Reichel, T., Ivanova, L.K., Beaven, R.P., Haarstrick, A., 2007. Modeling Decomposition of MSW in a Consolidating Anaerobic Reactor. Environmental Engineering Science 24, 1072–1083. URL: http://www.liebertonline.com/doi/abs/10.1089/ees.2006.0230, doi:10.1089/ees.2006.0230.

[30] Roychoudhury, A.N., Van Cappellen, P., Kostka, J.E., Viollier, E., 2003. Kinetics of microbially mediated reactions: dissimilatory sulfate reduction in saltmarsh sediments (Sapelo Island, Georgia, USA). Estuarine, Coastal and Shelf Science 56, 1001–1010. URL: http://linkinghub.elsevier.com/retrieve/pii/S0272771402003256, doi:10.1016/S0272-7714(02)00325-6.

[31] Rzeczycka, M., Blaszczyk, M., 2005. Growth and activity of sulphate-reducing bacteria in media containing phosphogypsum and different sources of carbon. Polish Journal of environmental studies 14, 891–895. URL: http://6csnfn.pjoes.com/pdf/14.6/891-895.pdf.

[32] Scharff, H., van Zomeren, A., van der Sloot, H.A., 2011. Landfill sustainability and aftercare completion criteria. Waste management & research : the journal of the International Solid Wastes and Public Cleansing Association, ISWA 29, 30–40. URL: http://www.ncbi.nlm.nih.gov/pubmed/20921059, doi:10.1177/0734242X10384310.

[33] Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research 46, n/a–n/a. URL: http://doi.wiley.com/10.1029/2009WR008933, doi:10.1029/2009WR008933.

[34] Siegrist, H., Vogt, D., Garcia-Heras, J.L., Gujer, W., 2002. Mathematical model for meso- and thermophilic anaerobic sewage sludge digestion. Environmental science \& technology 36, 1113–23. URL: http://www.ncbi.nlm.nih.gov/pubmed/11917999.

[35] Valencia, R., den Hamer, D., Komboi, J., Lubberding, H.J., Gijzen, H.J., 2009a. Alternative treatment for septic tank sludge: Co-digestion with municipal solid waste in bioreactor landfill simulators. Journal of Environmental Management 90, 940–945. URL: http://dx.doi.org/10.1016/j.jenvman.2008.02.007, doi:10.1016/j.jenvman.2008.02.007.

[36] Valencia, R., van der Zon, W., Woelders, H., Lubberding, H.J., Gijzen, H.J., 2009b. Achieving "Final Storage Quality" of municipal solid waste in pilot scale bioreactor landfills. Waste management (New York, N.Y.) 29, 78–85. URL: http://www.ncbi.nlm.nih.gov/pubmed/18406126, doi:10.1016/j.wasman.2008.02.008.

[37] Valencia, R., van der Zon, W., Woelders, H., Lubberding, H.J., Gijzen, H.J., 2009c. The effect of hydraulic conditions on waste stabilisation in bioreactor landfill simulators. Bioresource Technology 100, 1754–1761. URL: http://dx.doi.org/10.1016/j.biortech.2008.09.055, doi:10.1016/j.biortech.2008.09.055.

[38] Valencia, R., Zon, W.V.D., Woelders, H., Lubberding, H.J., Gijzen, H.J., 2011. Anammox: an option for ammonium removal in bioreactor landfills. Waste management (New York, N.Y.) 31, 2287–93. URL: http://www.ncbi.nlm.nih.gov/pubmed/21795036, doi:10.1016/j.wasman.2011.06.012.

[39] Valencia Vazquez, R., 2008. ENHANCED STABILISATION OF MUNICIPAL SOLID WASTE IN BIOREACTOR LANDFILLS. Ph.D. thesis. Wageningen University and UNESCO IHE Institute for Water Education, Delft.

[40] Vavilin, V.A., Rytov, S.V., Lokshina, L.Y., Pavlostathis, S.G., Barlaz, M.A., 2003. Distributed model of solid waste anaerobic digestion: effects of leachate recirculation and pH adjustment. Biotechnology and bioengineering 81, 66–73. URL: http://www.ncbi.nlm.nih.gov/pubmed/12432582, doi:10.1002/bit.10450.

[41] Veeken, A.H., Hamelers, B., 1999. Effect of temperature on hydrolysis rates of selected biowaste components. Bioresource technology 69, 249–254. URL: http://www.sciencedirect.com/science/article/pii/S0960852498001886.

[42] Veeken, A.H., Hamelers, B., 2000. Effect of substrate-seed mixing and leachate recirculation on solid state digestion of biowaste. Water science and technology : a journal of the International Association on Water Pollution Research 41, 255–62. URL: http://www.ncbi.nlm.nih.gov/pubmed/11382000.

[43] Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. Environmental Modelling & Software 75, 273–316. URL: http://www.sciencedirect.com/science/article/pii/S1364815215300396, doi:10.1016/j.envsoft.2015.08.013.

[44] Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research 39. URL: http://www.agu.org/pubs/crossref/2003/2002WR001642.shtml, doi:10.1029/2002WR001642.

[45] White, J., Robinson, J., Ren, Q., 2004. Modelling the biochemical degradation of solid waste in landfills. Waste management (New York, N.Y.) 24, 227–40. URL: http://www.ncbi.nlm.nih.gov/pubmed/15016412, doi:10.1016/j.wasman.2003.11.009.

Dear editor,

Thank you for your fast reply, we are very happy that our manuscript is considered for publication in Environmental Modelling & Software (EM&S). We also thank you for your additional comments which enabled us to improve the transparency and framing of the presented methodology and results. We double checked the grammar in the paper and changed the following according to your comments.

In line 55, we included two citations of papers from EM&S (Vrugt et al. (2016) and Kelly et al. (2013)) which express the need for a more objective integrated assessment of environmental systems. These are included to show that novel methods such as the one we present are necessary to decrease uncertainty in modeling of environmental systems. Our toolbox decreases model uncertainty by combining mechanistic information from different environmental fields and evaluating models based on qualitative and quantitative criteria.

In line 56, we included a reference to Bennett et al. (2013) who propose a generic evaluation procedure of model performance. Such an approach strengthens the credibility and relevance of modelling reported which is in line with the methodology we present. In lines 83-92, we indicate that our method supports all five evaluation steps proposed by Bennett et al. (2013). In our evaluation, however, the step of checking the data is minimal because the available measured data and information about its error was very limited. Nevertheless, we believe that the characteristics of our novel methodology are sufficiently demonstrated and that the best model is satisfactory given the scope of the paper.

In section 2.4, we added a subsection describing the qualitative performance criteria that are generated by the toolbox. In section 2.4.3, we justify our choice in quantitative criteria. We chose a set of criteria that enabled us to quantify aspects related to the (mechanistic) uncertainty of the models which is in line with the purpose of our modeling exercise. In addition, we indicate that other criteria suited for other modeling purposes are readily implemented.

Results and discussion are only slightly modified because we believe that the changes made in the introduction and theory already improved their framing and justification significantly. Furthermore, we modified the conclusion about the best model into two steps. First we conclude which model is best for lysimeter scale and then we conclude which model is best for full scale.

We thank you for your time and hope our paper is now fit to contribute to Environmental Modelling & Software.

Sincerely yours,

André van Turnhout

LaTeX Source Files

Click here to download LaTeX Source Files: TEX_file.tar.gz