A dynamic probabilistic material flow modeling method

Nikolaus A. Bornhöft [a, b], Tian Yin Sun [a], Lorenz M. Hilty [a, b, *], Bernd Nowack [a]

[a] Empa, Swiss Federal Laboratories for Materials Science and Technology, Lerchenfeldstrasse 5, CH-9014 St. Gallen, Switzerland
[b] University of Zurich, Department of Informatics, Binzmühlestrasse 14, CH-8050 Zurich, Switzerland

ARTICLE INFO

ABSTRACT

Material flow modeling constitutes an important approach to predicting and understanding the flows of materials through the anthroposphere into the environment. The new "Dynamic Probabilistic Material Flow Analysis (DPMFA)" method, combining dynamic material flow modeling with probabilistic modeling, is presented in this paper. Material transfers that lead to particular environmental stocks are represented as systems of mass-balanced flows. The time-dynamic behavior of the system is calculated by adding up the flows over several consecutive periods, considering changes in the inflow to the system and intermediate delays in local stocks. Incomplete parameter knowledge is represented and propagated using Bayesian modeling. The method is implemented as a simulation framework in Python to support experts from different domains in the development of their application models. After the introduction of the method and its implementation, a case study is presented in which the framework is applied to predict the environmental concentrations of carbon nanotubes in Switzerland.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The quantification of the environmental concentration of an anthropogenic pollutant is a crucial step toward the determination of risks for humans and ecosystems emerging from the application of new materials. While direct, quantitative measurements are often not feasible, the representation of material flows that lead to those concentrations provides means for an indirect assessment. The knowledge about these flows is the starting point for multi-media environmental fate models, which regard systems as sets of clearly separated, distinguishable compartments and allow the investigation of material transfers between them (MacLeod et al., 2010). "Multimedia" in this context refers to the fact that multiple environmental media (air, surface water, groundwater, soil) are considered parts of the system under study.

In general, material flow modeling approaches are well suited to investigate a large range of anthropogenic pollutants. For the assessment of the arising environmental stocks, the relevant flow processes need to be investigated. Depending on the pollutant and the scope of the investigation, this may include the material pro-duction, the application and use in different products, subsequent waste handling processes, and flows between environmental me-dia. Different scopes of a study can introduce further aspects such as geographical distribution or a more detailed subdivision of (e.g. technical) processes.

Existing mass flow modeling approaches such as material flow analysis (MFA) (Baccini and Brunner, 1991) regard systems of stocks and flows using mass equations to derive dependent system di-mensions. They are supported by the software tool Stan (TU Vienna 2012) for general flow modeling purposes and the Umberto soft-ware (ifu Hamburg GmbH, 2014) for material flows in the domain of corporate environmental management. These programs (STAN and Umberto) also support uncertainty representation and propaga-tion, but are restricted to a set of given distribution functions. They also support a period-based time representation. However, the update of the system state is determined by an explicit definition of the flow model for every period and not based on an underlying set of rules (e.g., for the residence times in stocks).

In environmental modeling, however, often considerable un-certainties exist about the volume of a flow, the rates with which the total amount divides into partial flows, and the particular pathways they take. Available data sources may be based on imprecise, incomplete or even contradictory assumptions. The explicit representation of these uncertainties and their propagation through the model can lead to more meaningful simulation results, thus allowing more reliable predictions of the resulting environ-mental concentrations. Bayesian modeling provides a technique for

---

* Corresponding author. Empa, Swiss Federal Laboratories for Materials Science and Technology, Lerchenfeldstrasse 5, CH-9014 St. Gallen, Switzerland.
E-mail address: Lorenz.Hilty@empa.ch (L.M. Hilty).

representing and propagating incomplete system knowledge and translates uncertainty about the true value of a system variable to the model as a probability distribution for the model parameter in question. It represents the modelers' assumptions about the true value, which can vary both concerning the type and the parameters of the probability distribution. Based on the given distributions, the distributions of the dependent values are then inferred using Monte-Carlo (MC) simulation. Money et al. (2012) proposed a Bayesian network of several stages for forecasting environmental concentrations of nanoparticles.

The probabilistic material flow analysis (PMFA) approach was developed by Gottschalk and colleagues (Gottschalk et al., 2010). They built a flow model that includes a complete assessment of uncertainties in all model parameters. It applies Bayesian modeling to propagate incomplete knowledge about the absolute inflow to the system and the internal dependencies between the downstream flows. Over a large sample size, steady states of flows are calculated, each based on a sampled set of random values. From that the resulting absolute material flows are determined. PMFA has mainly been applied for assessing environmental flows of nanomaterials (Gottschalk et al., 2009, 2010; Gottschalk and Nowack, 2011; Sun et al., 2014).

The simulation of systems over significant periods enables the estimation of absolute stock volumes. This includes, in particular, systems with time-dependent inflows and residence times in stocks. To represent time-dependent residence times, dynamic models become necessary because the release of one period depends on the inflows of several previous periods and the delay characteristic of the stock. Such models partially include dynamic system behaviors, such as the scaling of a flow of a reference year to estimate annual flows for previous periods and add up those inflows to a stock to obtain absolute volumes (Gottschalk et al., 2009) or the calculation of flows over subsequent periods based on clocked releases defining rates from the absolute stock of a well-mixed reactor (Walser and Gottschalk, 2014). These models provide a probabilistic material flow representation and a limited representation of changes over time. However, time-dependent external material inflows and material release from stocks as functions with varying residence times and release rates are not included. Moreover, in the studies mentioned above, special-purpose models were developed for particular cases. These studies do not provide a general method of how to model systems of this type, nor do they provide a conceptual and operational framework to support the modeling and evaluation process.

Outside the field of probabilistic modeling, many material flow modeling methods are in use that provide means to represent dynamic system behavior over time. Müller et al. (2014) present a survey on a large range of these methods, focusing on the uncertainty handling of these methods. While a large share of the methods do not consider uncertainty at all (>50%), there are some that use sensitivity analysis (37%), Gaussian error propagation (6%) or parameter ranges (5%), but none supports full Bayesian uncertainty representation and propagation.

Dynamic Bayesian networks that are mainly used to learn and reproduce time-dependent system behavior (Daly et al., 2011) process uncertain knowledge in a time-dynamic model. However, this approach focuses on variances in state transitions and does not include flow-specific behavior.

To summarize, what is missing is a method for investigating the development of environmental stocks of a pollutant by building a model which satisfies the following requirements:

- It represents a system of mass balanced dependent flows,
- It considers changing material releases and intermediate delays in local stocks over a significant time horizon, and

- It provides means to represent and process incomplete parameter knowledge.

In (Bornhöft et al., 2013) we investigated several existing methods regarding their capabilities for meeting these requirements in more detail and revealed that no existing method fulfills these requirements. In the present article, we present a modeling approach that merges the advantages of the existing techniques of probabilistic material flow modeling with the existing approaches to dynamic material flow modeling. The combined method forms the basis for a software framework that supports the development, implementation, and simulation of dynamic probabilistic material flow models. We will describe how we implemented this framework as a software package using the Python language (Python Software Foundation, 2014) to support experts in building specific models in their field of application.

Finally, we will demonstrate the application of the framework using a realistic case study. This case includes the implementation of a model to investigate the system of flows of engineered Carbon Nanotubes (CNT) in Switzerland. Due to their toxic properties to humans and ecosystems, CNTs pose potential risks (Savolainen et al., 2010). Sun et al. (2014) presented a steady-state model to assess the inflows to different environmental compartments based on data for the year 2012. However, CNTs are very stable and accumulate in the environment over time. Moreover, they are usually applied in products with long lifetimes, which leads to significant material amounts bound in use stocks. A dynamic model is therefore needed to provide a more detailed and adequate system representation. Based on this example application, the new approach is discussed in more detail regarding general functionality and its opportunities and limitations.

2. Description of the method

We propose a new method that combines the advantages of the existing approaches to probabilistic and dynamic material flow modeling: dynamic probabilistic material flow analysis (DPMFA). It aims to close the gap in existing techniques for exposure assessment by providing means to model and simulate systems of complex, dependent material flows, consider the dynamic behavior of the system over time, and explicitly represent and propagate incomplete parameter knowledge. For that purpose, a set of components is provided as building blocks for the model. These components need to be instantiated and linked together to represent the investigated system, and to allow simulation and evaluation.

We first outline the main idea of the approach, describing the basic structure of the models, the simulation processes and how the elements of the previously introduced modeling methods are combined. The implementation of the framework as a software package in Python is described on that basis in a second step.

Each DPMFA model is an abstraction and idealization of an original system of flows in the technosphere and the ecosphere. The model is reduced to the parts and aspects that determine the behavior investigated. Following the scope of the simulation study, the system is first subdivided into a set of compartments. They constitute the static model structure and structure the system into spatially or logically separated units (e.g., as in Fig. 1). The actual breakdown depends on the objective and the scope of the study. All material inflows, transfers, accumulations, and releases refer to these compartments.

Simulation experiments need to be performed with the model to assess material stocks and flows over time. Based on the results of these experiments, conclusions about the processes of the original system are drawn. The general simulation mechanism for investigating the flows between the compartments is structured as
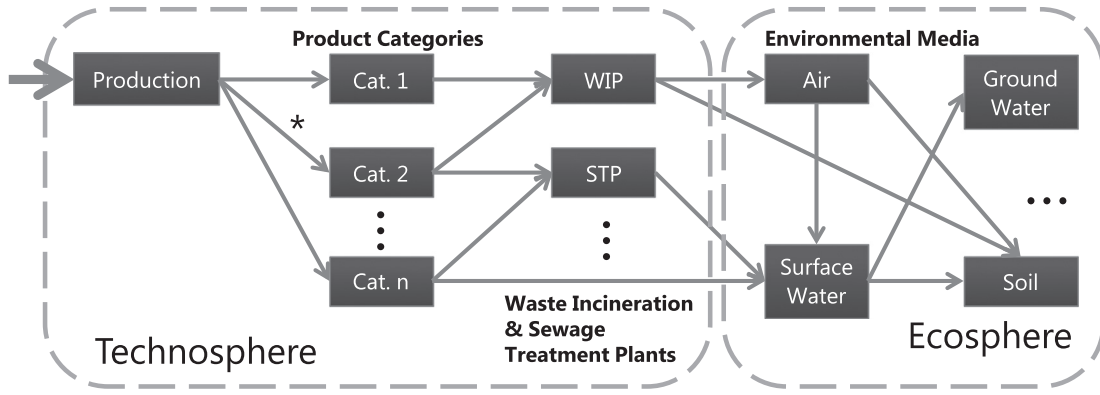
Fig. 1. Pathways of material flows of anthropogenic pollutants from the technosphere via different product categories, waste incineration and sewage treatment plants to the ecosphere. Specific system compartments and flow dependencies need to be implemented for each particular material and scope.

a 3-layer process (see Fig. 2). On the first — the Bayesian — layer, parameter uncertainty about the flow dependencies between the system compartments and the absolute annual inflow is represented by Bayesian probability distributions. These uncertainties are then propagated through the model for the entire simulation time using Monte-Carlo techniques.

The second layer refers to the time-dynamic model behavior. Time is represented as a sequence of successive periods (usually years). For each period within the time horizon of the simulation, the external inflows to the model, the material accumulation in stocks, and their local material releases are determined and added up.

To enable this, the third layer provides a mechanism that calculates the absolute material flows for a period based on absolute material releases and the flow matrix, taking all transfer dependencies into account.

2.1. Static structure

The static model structure consists of a set of persistent entities. They represent the local relations of the compartments and are assembled to derive the global system behavior. The basic model components are flow, stock, and sink compartments and external inflows.

- A flow compartment includes material inflows and relative outflows of a delimited spatial of logical system area.
- A stock compartment is a component with a temporary total or partial material accumulation and later re-release of the material. Stock compartments include local material in- and outflows and provide a delay function that determines material accumulations and releases.
- A sink compartment is a component with permanent material accumulations.
- An external inflow is a source that implies a time-dependent exogenous input to a stock or flow compartment (e.g., through production or import).

The dynamic model behavior emerges from the interplay of these static components over time.

2.2. Material flows

The calculation of absolute values for the material transfers is derived from existing material flow analysis approaches using a classical Leontief model (Leontief, 1986). It represents the material flows of one period as immediate and simultaneous. While exogenous inflows to the system are defined as absolute material inflow values to a compartment, endogenous flows from a compartment
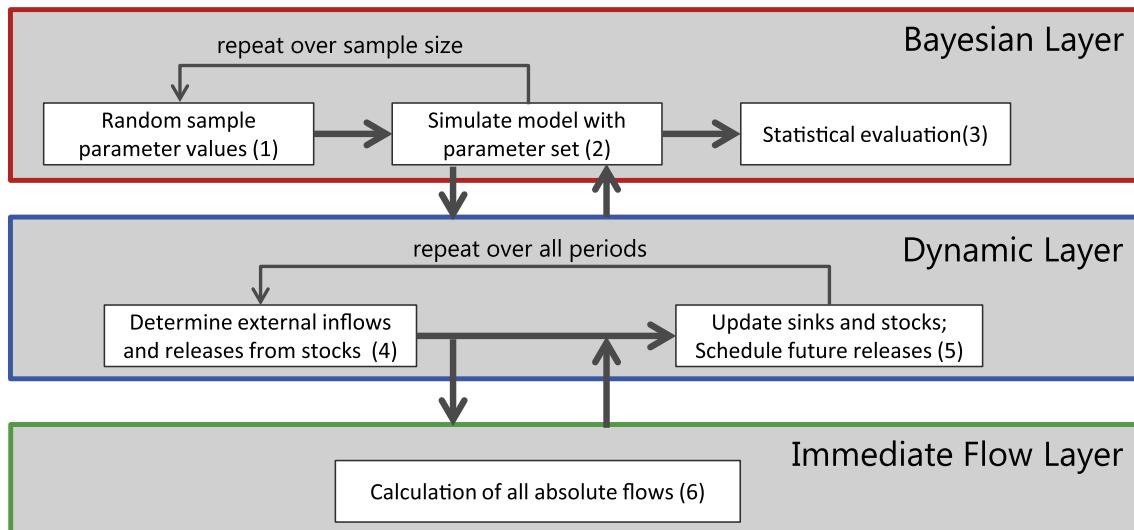

Fig. 2. Dynamic Probabilistic Material Flow Analysis —structure of the simulation process.

are defined by transfer coefficients (TC). The transfer coefficient TC_{js} defines the relative mass flow m from compartment j to s as a proportion of the sum of all inflows to compartment j (Eq. (1)).

$$TC_{js} = \frac{m_{js}}{\sum_r m_{rj}} \tag{1}$$

To determine the absolute flows of the model, all transfer coefficients are assembled to the flow matrix A (Eq. (2)).

$$A = \begin{array}{c|ccc|ccc}
 & C_1 & \dots & C_m & C_{m+1} & \dots & C_n \\
\hline
C_1 & 1 & \dots & -TC_{m\,1} & 0 & 0 & 0 \\
\vdots & \vdots & 1 & \vdots & 0 & 0 & 0 \\
C_m & -TC_{1\,m} & \dots & 1 & 0 & 0 & 0 \\
\hline
C_{m+1} & -TC_{1\,m+1} & \dots & -TC_{1\,m+1} & 1 & 0 & 0 \\
\vdots & \vdots & \dots & \vdots & 0 & 1 & 0 \\
C_n & -TC_{1\,n} & \dots & -TC_{m\,n} & 0 & 0 & 1
\end{array} \tag{2}$$

The flow rates from one compartment to another are read diagonally from top to left. The compartments C_1 to C_m represent immediate flow dependencies, compartments C_{m+1} to C_n sinks. The absolute material inflows to the system are expressed as an input vector I (Eq. (3)).

$$I = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ \cdot \\ \cdot \\ \cdot \\ g_n \end{pmatrix} \tag{3}$$

This vector comprises the sum of the current external inflows and the releases from the model stocks to all compartments C_1 to C_n as elements g_1 to g_n. Solving the System (Eq. (4)) for an unknown column vector X leads to a steady state of flows.

$$AX = I \tag{4}$$

The column vector X determines the inflows to the compartments with which the stocks are incremented. If the sum of each column of a flow compartment in the coefficient matrix is zero and the entire inflow is allocated to the sink columns as a non-zero
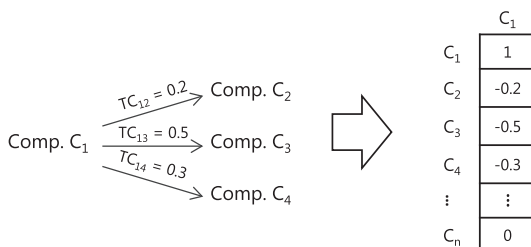
value, the system is mass-balanced. All material inflows are distributed to the sinks based on the relative local flow dependencies.

2.2.1. Flow compartments

In the model, the relative transfer dependencies are bound to flow compartments, which represent points in the system where material flows are gathered and split up. Several transfers can be bound to one flow compartment. Each transfer includes a target compartment and a transfer coefficient. The combination of all outgoing transfer coefficients from a compartment enables to ensure a mass-balanced system. Therefore, the outgoing transfers from each compartment need to sum up to 1 to create a global balance. To assemble the flow matrix (Fig. 3) the outgoing TCs from the flow compartments are transformed into the columns of the matrix.

2.3. Time-dynamic behavior

Time advancement is represented in the model as a series of subsequent periods T_0 to T_n of equal length. In each period, the model-wide material flows are determined and used to update the stocks and sinks:

- First, the external inflows and the material releases from stocks are determined (Fig. 2, Box 4).
- Second, the flows of the period are determined based on the inflows and releases (Fig. 2, Box 6) by assigning the respective material inflows to the input vector I(Eq. (3)) and by solving the flow matrix of the system (Eq. (4)).
- Finally, the stocks and sinks are incremented with their particular inflows from the solution vector X (Fig. 2, Box 5).

Once the model is simulated over the required time interval, the total material in a sink at the end of this interval can be predicted.

2.3.1. External inflows

Material inflows from an external source to a system compartment are defined as absolute material inputs for a particular system compartment and period. A time dynamic development of these inflows is either represented by a list defining an input volume for each period or as a function of time over all periods. The particular inflow of a period to a particular model compartment C_i is added to the inflow vector I at the element g_i.

2.3.2. Stock compartments

Stock compartments represent material flows through system areas, where at least a part of the material transfer is not immediate.

Therefore, the stock compartments include:

- A set of transfer coefficients that determine the proportions of the material leaving the compartment to particular subsequent compartments; this is analogous to the Flow Compartments (Fig. 3). However, due to residence times >0 of the material in stock, the periodic outflow to a stock compartment does not match its inflow. For a consistent definition of the relative proportions of the outgoing flows, the TCs are here defined as the relative ratio to the total outflow of a stock compartment.
- A release function $releaseFct(t)$ that defines relative times and proportions for the materials (re-) release based on the time of the material inflow t_0.

The release function defines the residence times and the rates with which materials that enter the stock compartment are

Fig. 3. Outgoing TCs from Compartment 1. The set of TCs corresponds to the respective column of the flow matrix (Eq. (2)).

released again. For the calculation, the immediate release in period 0 and those in later periods are treated in different ways. The portion immediately released is included to the flow matrix A. Therefore, the outgoing TCs from the stock are multiplied with the immediate release rate $releaseFkt(t_0)$ and added to the flow matrix as column, in just the same way as the TCs from the flow compartments. The portion of the material that is released with some delay is treated as described below.

To determine the dynamic development of the stored amounts in stock and the time-dependent material releases, a stock compartment includes the following elements:

- An *Inventory* displaying the current, absolute stocked amounts. To enable the evaluation of the stocked values, the inventory is modeled as a list, recording the stock for all periods.
- A *ReleaseList* that includes the scheduled material releases for the future periods

During the calculation of the flows of a period i, the following steps are performed in a stock compartment.

I. At the beginning of the period T_i
a) Transfer the stocked amount from the previous period to the current period as the initial value (Eq. (5)). (This step is omitted in the first period):

$$Inventory(T_i) = Inventory(T_{i-1}) \tag{5}$$

b) Determine the total release from the stock for the period (Eq. (6)) and reduce the inventory by that value (Eq. (7)):

$$currentRelease = ReleaseList(T_i) \tag{6}$$

$$Inventory(T_i) = Inventory(T_i) - currentRelease \tag{7}$$

c) For **each** outgoing TC_{js} from stock compartment j, include the portion of the current release to the inflow vector I (Eq. (8)):

$$I(g_s) = I(g_s) + currentRelease*TC_{js} \tag{8}$$

II. After calculation of the global flows (Fig. 2, Box 6) the stock compartment is updated with the respective material inflow from the solution vector X
a) Add the not-immediately released portion to the inventory (Eq. (9)):

$$Inventory(T_i) = Inventory(T_i) + X(s)*(1 - releaseFct(t_0)) \tag{9}$$

b) Schedule the material releases for the **future** periods (Eq. (10)):

$$\begin{aligned} &For\ each\ k\ from\ 1\ to\ m: \\ &\quad if\ i + k \leq n: \\ &\qquad ReleaseList(T_{i+k}) \\ &\qquad\quad = ReleaseList(T_{i+k}) + X(s) * releaseFct(k) \end{aligned} \tag{10}$$

2.4. Uncertainty representation and processing

In exposure assessment modeling, incomplete knowledge may concern the point in time, the location or the extent of a flow. This uncertainty is mainly epistemic, which means it relates to a general lack of knowledge about the true value of a system variable. Such uncertain variables are represented using (Bayesian) likelihood distributions, which include all plausible values and assign normalized probability densities. The dependent system variables (e.g. a stock at a particular time) are calculated using Monte-Carlo simulation, i.e., the model is repeatedly evaluated over a large sample size m. For each single run $i \in m$, all uncertain parameters are assigned a random number from the associated parameter distributions (Fig. 2, Box 1). With this parameter setting, the model is calculated over all periods as described above (Fig. 2, Box 2). As a result, the dependent model variables (e.g. stocks) are available as an $m \times n$ matrix. Based on that representation, statistical evaluations and visualizations can be performed (Fig. 2, Box 3).

The parameter distributions are either regarded as parametric distribution functions or as non-parametric distributions. Depending on the origin of the available data, there may by samples from direct observations, results of previous simulation steps, or probability distribution functions representing the assumed characteristics of the distribution. Since it is possible to sample random values from either variant for the Monte-Carlo simulation, both are suitable for representing uncertain knowledge about absolute inflows and transfer coefficients in the model.

The representation of uncertainty in transfer coefficients and external inflows has some important characteristics. For modeling TCs the mass balance of the system needs to be preserved. While in a deterministic mass balanced flow model the sum of the outgoing TCs from one flow compartment or stock have to sum up to 1, in the probabilistic case the marginal distributions for the model parameters have to be chosen in such a way that their *expected values* sum up to 1.

Moreover, in the simulation process, the dependent random values are adjusted after sampling to avoid combinations violating mass balance constraints. The modeler can chose to do so either by a normalization factor over all involved TCs or—in the case of transfer coefficients from underlying information of strongly differing reliability — by defining an order of priority to first adjust the parameter values based on the least reliable data.

The external inflow to a particular compartment over time can be represented either as a list of single probability distributions for each period or by one marginal distribution representing an uncertain base value and a deterministic growth function. The two variants imply different underlying assumptions. The use of a common base value for all periods emphasizes the inter-periodic dependencies while the absolute value is not exactly known. Expressed as a list of single inflows, the random samples for the periods are assumed to be independent. They implicitly show variant behavior and increasing degrees of freedom of the model with the number of simulated periods and thus a growth of the complexity of model behavior for longer time spans.

2.5. Implementation of the method

Based on the DPMFA method, a software framework was developed to support the design and use (i.e., the simulation) of dynamic probabilistic material flow models. It is designed as a Python (2014) package and utilizes the *SciPy* library (Jones et al., 2001) for statistical computation and in particular the *NumPy* package (van der Walt et al., 2011) for matrix representation and calculation.

The program package implements the principle of separation of model and experiment (Page and Kreutzer, 2005). At its core, it provides the infrastructure to perform simulation experiments using the Simulator class. This class is provided as a black-box component and is used unchanged by a modeler working with the package. The modeler implements the system-specific logic by
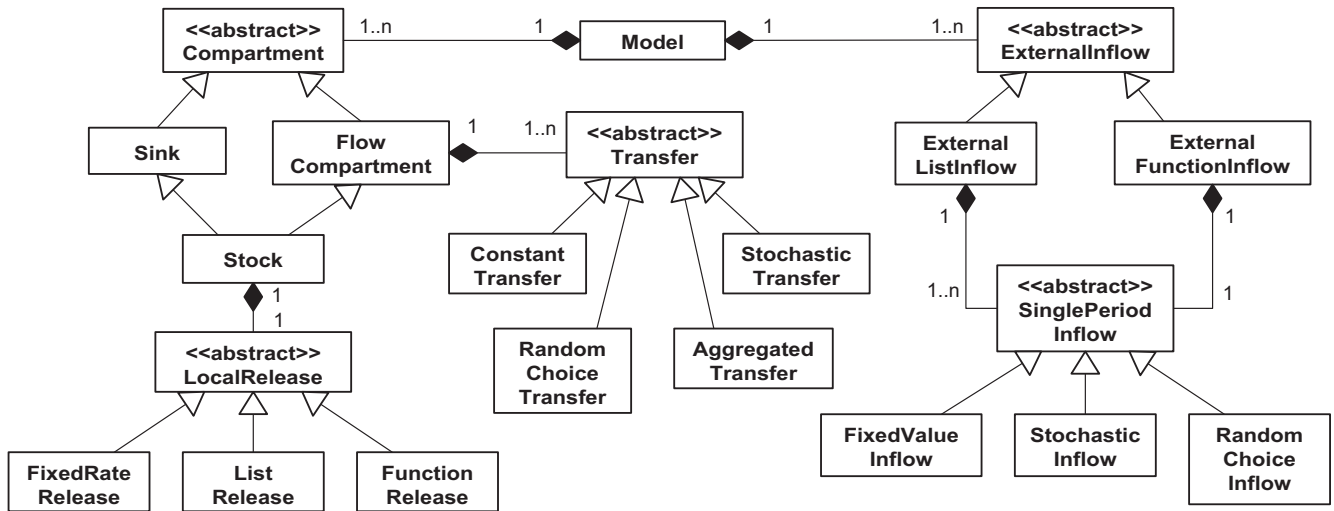
Fig. 4. UML diagram; composition of the DPMFA model structure.
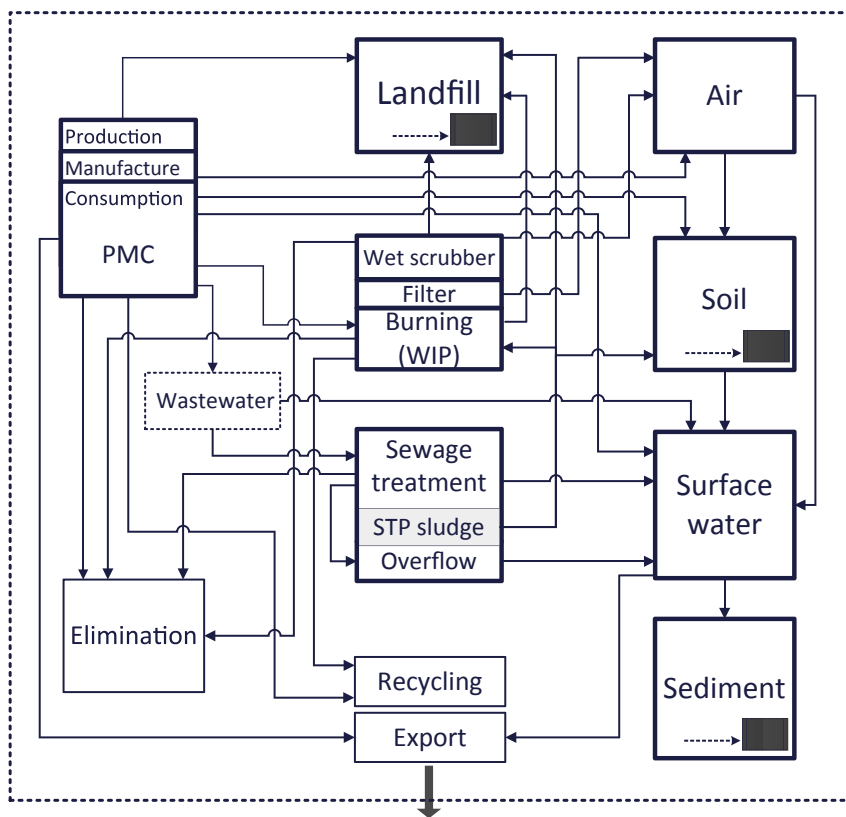

Fig. 5. Simplified pathways of CNTs to the environment. CNT production, distribution to different product categories and category-specific release are pooled in PMC. Technical waste and waste water treatment processes are pooled as well.

assembling predefined components. These are provided as white-box components that the modeler has to adapt to fit the particular behavior of the system under study.

2.5.1. Simulator

The Simulator performs experiments to generate and evaluate the Model behavior. As part of the simulation process — as described by our overall simulation algorithm above (Fig. 2) — the model parameters specified under uncertainty are assigned

random values from the underlying Bayesian probability distributions. Statistical evaluations of the observations over sufficiently large sample sizes approximate the distribution of the variables under the assumptions of the marginal distributions. For each of these parameter sets, the model is simulated over the total investigated time span.

In an iteration over all periods, the Simulator determines the external inflows to the system and the local inflows from the stocks. These flows are then distributed to the different model
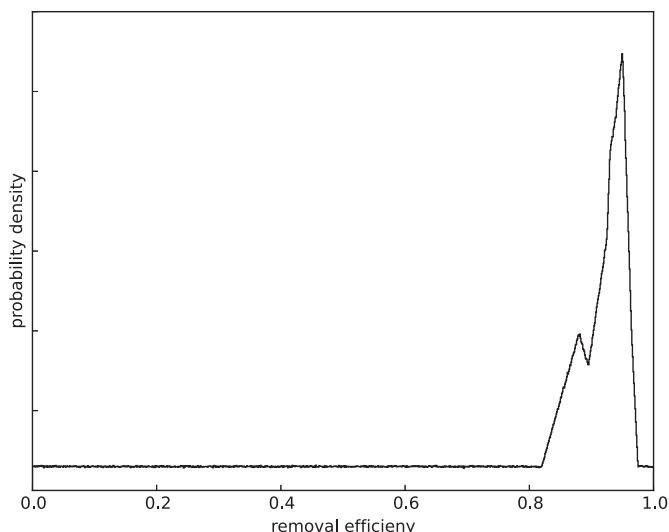
Fig. 6. Likelihood function of the CNT removal efficiency in sewage treatment plants (STP).

Fig. 7. CNT inflow to sediment compartment in the static model: Density function, mode and mean value of the sample.

compartments by solving the flow matrix of the model − which is assembled from the internal flow dependencies − with the current inflow vector. Based on the inflows, the model stocks and sinks are updated. During the experiment, the Simulator keeps track of the values of model variables (e.g., the amount of material in a stock).

All of these values are logged in the form of a matrix over all samples and periodic values for later statistical evaluation. To facilitate an aggregated evaluation, categories can be assigned to the model compartments. After a simulation experiment is executed, the Simulator provides several functions for a category-based evaluation, e.g., to provide total material inflow or outflow or the total material stocked.

2.5.2. Model

The model builder implements a specific simulation model by customizing and combining basic model components:

- Model Compartments representing system entities, which all material flows, accumulations, and releases are related to,
- Transfers defining the internal, relative flow dependencies,
- LocalReleases defining the residence times of materials from Stocks and the release rates, and
- ExternalInflows representing exogenous inputs to the system.

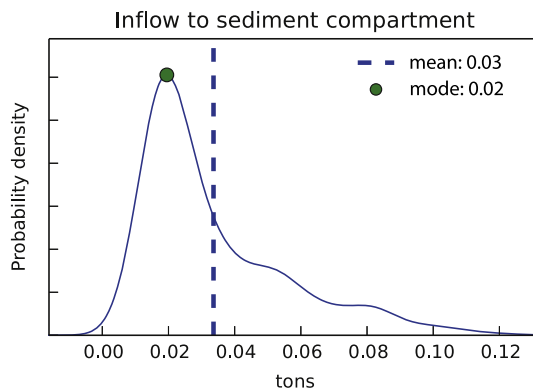An overview of the model structure is shown in Fig. 4 as a class

diagram. The diagram illustrates the model composition and the hierarchy of the included component types. The Compartments are specified by subclasses. FlowCompartments are branches of a flow within one period; Sinks mark the material accumulation at an endpoint of a flow process, and Stocks represent material flows that are delayed for a particular period of time and later transferred further.

Different Transfer types are used to model flow dependencies as relative transfer coefficients to particular subsequent target Compartments. ConstTransfers define deterministic values as transfer coefficients. StochasticTransfer, RandomChoiceTransfer, and AggregatedTransfer use probability distributions to represent incomplete knowledge about the true values of transfer coefficients. Random values are sampled for those Transfers during the simulation process. StochasticTransfers are parameterized with probability distribution functions and respective parameter lists. RandomChoiceTransfers hold lists of values to draw randomly from. AggregatedTransfers allow weighted combinations of the previously stated Transfers.

All transfers are bound to sources, which can either be Flow-Compartments or Stocks. To ensure the mass balance of the system, the local transfer coefficients for the relative outflows from such a source have to sum up to 1. This adjustment step is performed after the random values are sampled from the underlying probability distributions. The modeler can either chose to apply a normalization of the corresponding transfers or to define a prioritization to adjust the random numbers from the least credible underlying data. Combinations of both approaches are feasible as well.

Stocks represent delayed flow processes. The model builder defines their particular release times and rates as LocalRelease

Table 1

Simulation results − model sinks in tons of CNT/year: Comparison of the mean values of the inflows to the model sinks for 2012. The left column shows the results of the deterministic model in R, using the expectation values of the parameter distributions from Sun et al., 2014. The middle column shows the results of the same deterministic model implemented using the new package. The right column shows the simulation outcome of the probabilistic version of the model implemented with the new package (mean values).

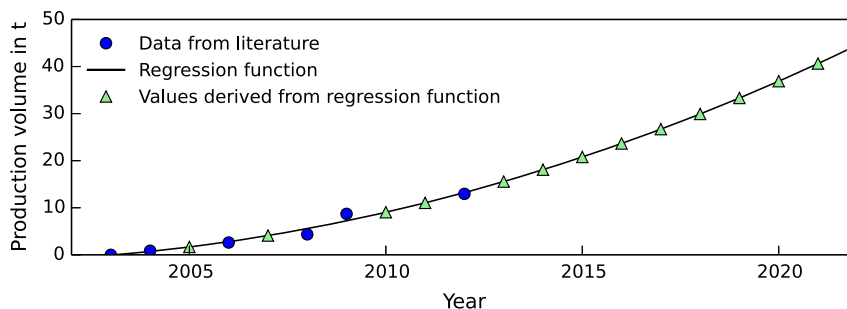|	Deterministic model based on (Sun et al., 2014), implemented in R	Deterministic model, implemented using the new simulation package	Probabilistic model, implemented using the new simulation package (mean values)
Elimination	7.83	7.82	7.82
Landfill	0.96	0.96	0.97
Soil	0.14	0.15	0.15
Sediment	0.03	0.03	0.03
Cement Plant	0.01	0.01	0.01
Recycling	3.10	3.10	3.08
Export	0.87	0.86	0.87
Sum	12.94	12.94	12.93

Fig. 8. Annual production volumes in tons/y; the value for 2012 is taken from Sun et al. (2014), previous years from the survey by Piccinno et al. (2012). Future and missing values were estimated using a quadratic regression function.

strategies. The target compartments and the relative transfer co-efficients are defined as Transfer objects the same way as for FlowCompartments. To implement LocalReleases, their subclasses need to be implemented. FixedRateRelease defines constant rates for all following periods, ListRelease an explicit list of all future release rates, and FunctionRelease gives a mathematical function for the particular rates and periods.

ExternalInflows are implemented as ExternalListInflow to define explicit inflow amounts for each period or as External-FunctionInflow with a (growth) function on a base value. To define the base value or the individual values for the list, the model builder has to define SinglePeriodInflows. These can be either deterministic FixedValueInflows or a probability distribution function, namely StochasticInflow or RandomChoiceInflow from a given sample.

3. Example application of the method

The capabilities of the DPMFA method and the corresponding Python package are illustrated by applying them to a case study of practical relevance. Here, we modeled the flows of carbon nano-tubes (CNT) in Switzerland to predict current and future material stocks in the technosphere and the environment. CNTs appear to be a useful and challenging example application because of their stability and toxicological properties as well as a lack of analytical methods for a direct measurement of environmental concentra-tions (Wick et al., 2011). CNT technology is relatively new, and there is a strong increase in current and expected production volumes. Moreover, a large proportion of the produced material is used in long-lasting applications such as polymer composites, which leads to the development of significant use stocks.

The CNT flows were previously modeled using MFA (Mueller and Nowack, 2008) and PMFA (Gottschalk et al., 2009; Sun et al., 2014). The investigated flows include the production of the CNTs, their application in different product categories, their release dur-ing the life cycles of the products to technical and environmental system compartments, and the subsequent environmental fate, namely their final accumulation as a pollutant.

The model was simulated on a standard laptop[1] with an Intel i5-4200U CPU @1.6 GHz processor and 8 Gb memory.

3.1. The static case

The basic structure of the model, as shown in Fig. 5, is derived from a steady-state model that we developed earlier to predict CNT flows in Switzerland (Sun et al., 2014). This model includes 31 compartments and sinks and 80 transfers, where all TCs are

modeled using parameter distributions. Fig. 6 exemplarily shows the sewage treatment efficiency as one of those distributions. This distribution determines the proportion of CNTs from waste water that are bound to Sewage Treatment Plant (STP) sludge. The dis-tribution is the result of combining several sources of uncertain evidence. It supports a range of values between 0 and 100% with a high likelihood of between 82% and 97%.

In Sun et al. (2014) the model was originally implemented as a special-purpose application using the R programming language. From that work, we adopted the subdivision of the system into particular compartments and the probability distribution functions that define the transfer coefficients of the flow dependencies be-tween the compartments. This static model will now be re-built and extended to a dynamic model to demonstrate our new approach.

We first re-implemented the static model using our approach to cross-check the consistency between the two approaches for the static case. To facilitate the cross-check, we created a deterministic version of the model by replacing the parameter distributions with their expectation values and then implemented the deterministic version both in R (as the original model of Sun et al., 2014) and in Python using the new package. With that, it was possible to compare the basic functionality of the flow calculations of the two implementations.

Then we re-implemented the stochastic version of the original model of Sun et al. (2014) using the new Python package as well. The purpose was to check the influence of the randomness of the underlying probability distributions on the simulation results. For the stochastic version, we used the same probability distributions as Sun et al. (2014) did. We simulated 50,000 runs which was considered a sufficient sample size. (See the Discussion and Outlook section for a discussion of sample sizes.)

Table 1 shows the material inflows to the model sinks as simulation results; in columns 1 and 2 for the deterministic ver-sions of the model in R and using the new Python package, respectively, and in column 3 for the probabilistic version.

The agreement between the simulation results was high. Small discrepancies between the two deterministic implementations can be explained by small numerical errors caused by differences in the underlying algorithms, i.e., for solving the flow matrix, or in number representation. But all in all, the two implementations can be seen as almost equivalent. Differences between the determin-istic and the probabilistic model can be explained by the stochastic error, introduced by the randomness of the probabilistic model, which is small due to the large sample size.

In previous works by Gottschalk et al. (2009, 2010; Gottschalk and Nowack, 2011) and Sun et al. (2014), we focused on the mode value to represent a sample by its most probable single value. Here we mainly use the mean value of the sample. This has some
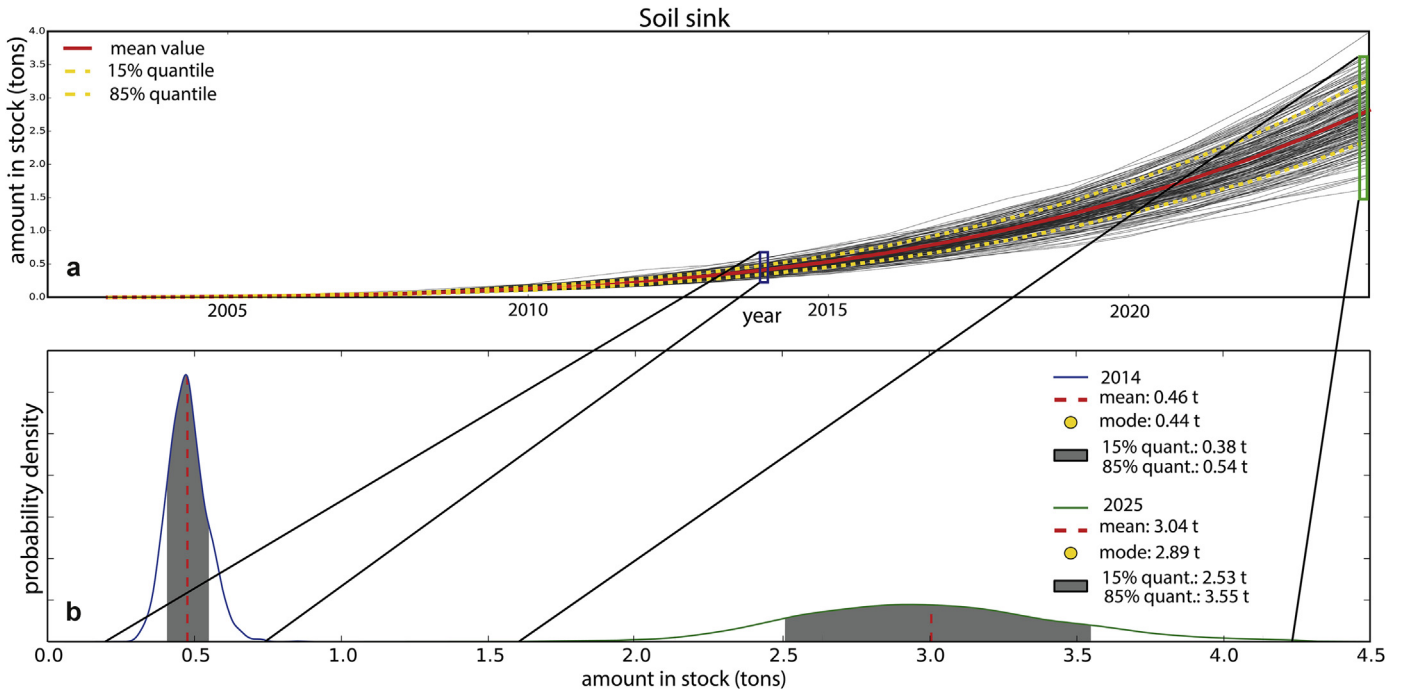
---

[1] HP EliteBook 840 G1.

Fig. 9. a,b: Amount of CNT in soil over time; each gray curve represents a random set of parameter values. For the years 2014 and 2025 the sample is projected to a density function.

Table 2
Mean values (in tons) of the samples of CNTs bound in the technosphere in different product categories, predicted values for 2012 and 2014, and predictions for 2025 under the assumption of growing production volumes or of an immediate production stop in 2015.

	2012	2014	2025	
			Growing prod.	Stopped prod.
Polymer composites	30.60	46.47	172.72	5.09
Consumer electronics	4.00	6.08	22.86	0.96
Automotive	1.87	2.95	12.84	1.78
Sum	36.47	55.50	208.42	7.83

advantages because the mean values show a system of balanced flows. Also, mean values are more robust, especially on small and scattered samples. The computation of a "real" mode value can be performed only for a discrete set of different values. For continuous variables, the maximum of a density function of the sample, such as the Gaussian kernel-density estimator (Scott, 1992), are often used instead. Depending on the used estimator and its parameters, different maximum values are chosen. However, both the mean value and the mode value represent only a single aspect of a

Table 3
Mean material amounts in sinks in tons, predicted values for 2012 and 2014 and predictions for 2025 under the assumption of growing production volumes or an immediate production stop in 2015.

	2012	2014	2025	
			Growing prod.	Stopped prod.
Elimination	5.14	10.74	109.49	61.11
Landfill	1.47	2.59	18.19	7.63
Soil	0.27	0.46	3.04	1.20
Sediment	0.13	0.22	1.11	0.29
Cement Plant	0.05	0.07	0.37	0.09
Recycling	9.51	15.57	84.64	24.79
Export	2.34	3.86	21.75	6.80
Sum	18.91	33.51	238.59	101.91

probability sample (Fig. 7). For more comprehensive insights, the sample itself or at least several dimensions of it have to be considered.

3.2. The dynamic case

We extended the static model to a dynamic one by applying historical production volumes as model inflows for previous periods and projections for future periods. This extension demonstrates the advantage of the DPMFA package. It enables the assessment of the absolute material amount in a stock from the sum of the preceding material flows.

The modeled time span begins in 2003 to cover the significant time period in which CNT have been applied on the industrial scale. The annual production volumes are derived from Sun et al. (2014) and Piccinno et al. (2012). Missing values for past and future periods are estimated using a quadratic regression function (Fig. 8). To represent uncertainty about the true production volumes, a standard deviation (SD) is assumed that complies with the relative SD in the sample of the system input from the Sun data. This is implemented as ExternalListInflow of single StochasticInflows using normal distributions with a respective parameterization.

CNTs applied in some products have a considerable residence time. This constitutes material stocks with releases after a delay period. Polymer composites, consumer electronics, and automotive have been identified as product categories forming significant intermediate stocks of CNTs (Sun et al., 2014). The delay period of consumer electronics is approximated by a list of relative circulation times of computer notebooks (Stiftung Entsorgung Schweiz et al., 2014) as ListRelease. The mean circulation time in the automotive industry is modeled as a normal distribution with a mean of 11.9 years (Kraftfahrt-Bundesamt, 2013) and a standard deviation of 5 years. For polymer composites, a mean delay of 7 years is assumed and approximated by a normal distribution with a mean of 7 and an SD of 3 years. The material releases from both stocks are modeled using a FunctionRelease.
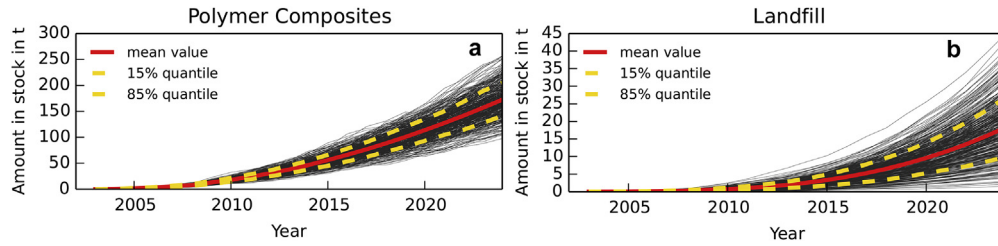
Fig. 10. a,b: Growth Scenario − CNTs bound in products containing polymer composites as stock of the technosphere (a) and in landfills (b) over time.
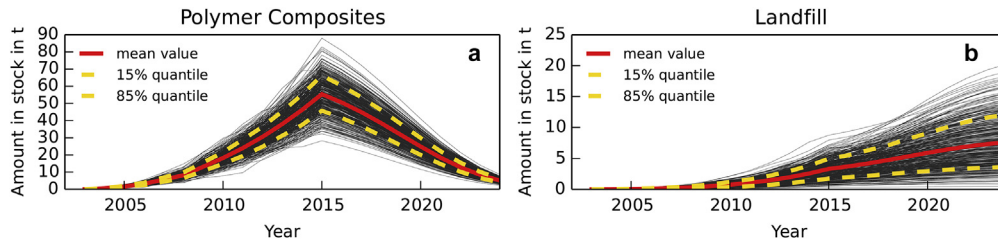

Fig. 11. a,b: Production stop in 2015 scenario − CNTs bound in products containing polymer composites as stock of the technosphere (a) and in landfills (b) over time.

3.3. Simulation results

The dynamic model was investigated for the period from 2003 to 2025 to predict its material stocks and flows over time. The environmental concentrations of CNTs in soil were determined for the years 2014 and 2025 as examples. Afterward, a second scenario was simulated to investigate the assumption of an immediate production stop of CNTs from 2015 on. Both scenarios were run over a sample size of 50,000 simulation runs. The computation of each took approximately 8:30 min. In the first scenario, growing production volumes (Fig. 8) were assumed.

The change of the amount of CNTs in the soil compartment over time is shown in Fig. 9a. Each individual curve represents the progress of the material amount in the compartment for one random set of parameter values from the underlying probability distributions, so areas of a high density of curves indicate values with a high likelihood. In the diagram, the number of curves was limited to 500 to increase the clarity of the representation. However, the mean values and quantiles stated still refer to the full sample. For the years 2014 and 2025, each of the samples of CNTs accumulated in the soil compartment were projected to a density distribution, from which mean and mode values, as well as quantiles, were derived (Fig. 9b). Based on the mean values and the significant mass of natural and urban soil of 6.25E+12 kg in Switzerland (Sun et al., 2014), the predicted environmental concentration in soil is 74 ng/kg for 2014 and 486 ng/kg for 2025.

Besides the growth of the amount of material stocked (and with it the environmental concentration), the uncertainty about the true values increases over time as well. While for 2014 the range between the 15% and the 85% quantile is approximately 0.16 tons, for 2025 it is 1.02 tons. The distribution of the CNTs among the different stocks for the years 2012 and 2014 is presented as mean values of the respective samples in Tables 2 and 3 (columns 1 and 2).

Table 2 shows the in-use stocks of CNTs for the years 2012 and 2014 and for both scenarios in 2025. Table 3 shows the accumulated amounts for the model sinks of the technosphere and environmental media. Currently, a large part of the material is still bound in products (in-use stock) − 36.47t(2012), 55.50t (2014) − while only 18.91t (2012) and 33.51t (2014) have been further transferred. This means that in 2012 a share of 65.85% (62.35% in 2014) of the mass

that entered the system has not yet been released to the environment. The material that is released from the product categories leaves the system to a large extent via export (3.86t) and recycling (15.57t). Waste incineration and sewage treatment eliminate 10.74t, and subsequently, 2.59t are bound in landfills. The release to the environment has resulted in an amount of 0.46t in soils and 0.22t in sediments (2014) so far. The progress of the stocked material in "polymer composites" as a compartment of the technosphere and in "landfill" as a model sink are pictured in Fig. 10a and Fig. 10b, respectively.

The second scenario investigates the system under the assumption of an immediate production stop from the year 2015 on. This leads to a peak of CNTs bound in the technosphere and a subsequent steady release (Fig. 11a,b).

The simulation results of the projected "growth"-scenario show a strong increase of both the amount of CNTs bound in polymer composites products and in landfill over time. The development of the material amounts in landfill is delayed relative to the material stock in polymer composites and shows a significant increase in the years from 2020 on. In the "production stop" scenario, the amount of CNTs bound in polymer composites slowly runs out, leaving only 5.09t in 2025. The total amount in landfill stabilizes at an amount of 7.63 t at the end of the time considered, and the predicted soil concentration is 192 ng/kg. Both scenarios show relatively little uncertainty about the product stocks. In contrast, the spread between the 15% and 85% quantiles of the landfill stock is approximately the same as the mean value. Outliers even reach roughly three times the mean amount.

4. Discussion and outlook

Dynamic probabilistic material flow modeling (DPMFA) as a new approach to material flow modeling provides a method for indirectly assessing material accumulations in stocks − both in the technosphere and in the environment − considering a variety of dependent partial flows and epistemic uncertainties. The simulation package to support the modeling process also provides components to represent local system behavior and a simulation environment to investigate dependent variables such as stocks at a particular time.

The suitability of the method and that of the Python package

supporting it for modeling and simulating these systems were illustrated through their application to predicting stocks of engineered CNTs in the environment. This is an exemplary case and the new method is applicable virtually to all MFA and dynamic MFA modeling cases, e.g., the ones reviewed by Müller et al. (2014), if and when the modelers want to consider the uncertainties for all relevant model parameters.

The DPMFA method enables the assessment of environmental concentrations, exposure to humans and ecosystems, and emerging risks. Moreover, the implementation of the example model showed that in the case of CNTs, delayed material transfers and the existence of intermediate stocks in the technosphere have a large impact on estimated current and future environmental concentrations. Whereas it was possible before to perform such simulations with traditional dynamic material flow models, it was so far not possible to fully include the uncertainties of the model parameters. Considering the intermediate stocks enables a closer investigation of the actual material amounts released to the environment and the prospective future releases. Within the scope of exposure assessment modeling, the new DPMFA method represents a significant step forward compared to established MFA methods because it allows consideration of a large range of different types of uncertainty for all relevant model parameters. The modeler can choose freely whether to use distributions, functions, or discrete data to describe the uncertainty of all parameters, thus making full use of the available data while representing the varying quantities and qualities of uncertainty as adequate as possible.

The time representation as a series of subsequent periods of equal length is an abstraction from the continuous nature of the flows in the real system. There are two good reasons for this abstraction. First, it enables efficient computation. Second, it corresponds to the way most data is available — as time series, namely as periodic (e.g., annual) values. Given that a continuous model would introduce assumptions (by implicit interpolation) that are often not warranted by data, this would induce a potential discretization error that would be rather inherent to the data than explicitly introduced during the modeling process.

The implementation of our approach as a Python package was chosen because it leads to several advantages. As a package on language level, it provides great flexibility for representing specific system characteristics, e.g., by implementing particular distribution functions for specific behaviors. The modeler is supported with virtually any parametric or non-parametric distribution function. As a tradeoff, programming skills are required. However, as Python is a language that is easy and convenient to learn, this disadvantage remains limited. At the same time it allows the modeler to embed the model into a larger project and to utilize the functionality of further associated libraries, e.g., for the preparation and management of large amounts of data with *pandas* (McKinney, 2014) or for plotting and evaluating simulation output with *matplotlib* (Hunter et al., 2007).

To ensure the computability also of larger models, the method accepts some limitations. The package does not support the representation of uncertainty about the time of a particular release from stock. However, material amounts in environmental stocks depend primarily on the total inflow to the system and the proportion transferred to the compartment. Especially for longer observation periods, the exact duration of a delay process has comparatively little impact on the total amount stored. Accordingly, uncertainty about these processes has only little influence and is therefore considered less relevant. Moreover, the transfer coefficients describing the relations between flows are considered stable over the investigated time (simulation length). Under this assumption, the model complexity mainly depends on the number of included model compartments and flow dependencies.

In general, the required computational effort to simulate a DPMFA model can be a limiting factor regarding model complexity, simulation length, time granularity, and desired precision of the simulation outcome. The used sample size of 50.000 illustrates a realistic, rather large sample, which leads to results that are stable between different simulation experiments. The computation of the model did not pose particular difficulties. In Gottschalk et al. (2010) the model stability of a PMFA model is discussed based on the match of significant numbers of the model output with the deterministic counterpart of the model as well as in between two simulation experiments of the same sample size. To estimate the required sample size for a particular precision of the results general estimations for Bayesian computation can be applied (e.g. Carlin and Louis, 2000).

For the given scope of the method — the assessment of environmental stocks and flows under substantial uncertainties — the simulation package was shown to be suitable. Considering a much higher degree of detail either of the system representation or the time resolution might be desirable in some cases. However, a particular degree of detail of the model only makes sense if it is not considerably exceeded by the existing uncertainties.

As probabilistic — Bayesian — prediction models, our models represent incomplete knowledge about the true value of a parameter as probability distributions. To ensure to comprise the true parameter value, also wrong, but plausible values are included. Instead of a validation of the model in terms of confirming or rejecting it, it can be improved by proving or rejecting some of the assumptions made, which reduces the incorporated uncertainty.

Future work could provide additional components that are adapted to particular application domains. Moreover, the modeling process could be enhanced by higher-level modeling constructs, in particular for hierarchical modeling and graphic model representation. Also, while the actual processing of uncertain knowledge about material flows is clear, support for the modeler in the formulation of a probability distribution based on heterogeneous, diverse, and incomplete knowledge about a system variable could be improved.

Software availability

The simulation framework is available as a software package via PyPI, the Python Package Index at: https://pypi.python.org/pypi/dpmfa-simulator.

Acknowledgment

References

Baccini, P., Brunner, P.H., 1991. Metabolism of the Anthroposphere. Springer, Berlin, Germany. http://dx.doi.org/10.1007/978-3-662-02693-9.

Bornhöft, N.A., Nowack, B., Hilty, L.M., 2013. Material flow modelling for environmental exposure assessment — a critical review of four approaches using the comparative implementation of an idealized example. In: Göbel, J., Wohlgemuth, V. (Eds.), 27th International Conference on Informatics for Environmental Protection. B. Page, a. G. Fleischer. Shaker, Hamburg, Germany, pp. 379—388.

Carlin, B.P., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis. Chapman & Hall/CRC, Boca Raton.

Daly, R., Shen, Q., Aitken, S., 2011. Learning Bayesian networks: approaches and issues. Knowl. Eng. Rev. 26 (2), 99—157. http://dx.doi.org/10.1017/S0269888910000251.

Gottschalk, F., Nowack, B., 2011. Release of engineered nanomaterials to the environment. J. Environ. Monit. 13, 1145–1155. http://dx.doi.org/10.1039/C0EM00547A.

Gottschalk, F., Scholz, R.W., Nowack, B., 2010. Probabilistic material flow modeling for assessing the environmental exposure to compounds: methodology and an application to engineered nano-TiO2 particles. Environ. Model. Softw. 25, 320–332. http://dx.doi.org/10.1016/j.envsoft.2009.08.011.

Gottschalk, F., Sonderer, T., Scholz, R.W., Nowack, B., 2009. Modeled environmental concentrations of engineered nanomaterials (TiO$_2$, ZnO, Ag, CNT, fullerenes) for different regions. Environ. Sci. Technol. 43, 9216–9222. http://dx.doi.org/10.1021/es9015553.

Hunter, J.D., others, 2007. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9 (3), 90–95. http://dx.doi.org/10.1109/MCSE.2007.55.

ifu Hamburg GmbH, 2014. "Umberto" V. 5.6 from. www.umberto.de. accessed: 05.09.14.

Jones, E., Oliphant, T., Peterson, P., others, 2001. SciPy: Open Source Scientific Tools for Python. URL http://www.scipy.org/.

Kraftfahrt-Bundesamt, 2013. Pressemitteilung Nr. 30/2003 - Alter Fahrzeugbestand.

Leontief, W., 1986. Input-output Economics. Oxford University Press.

MacLeod, M., Scheringer, M., McKone, T.E., Hungerbuhler, K., 2010. The state of multimedia mass-balance modeling in environmental science and decision-making. Environ. Sci. Technol. 44 (22), 8360–8364. http://dx.doi.org/10.1021/es103297w.

McKinney, W., 2014. "pandas: a python Data Analysis library." V. 0.15.1 from. http://pandas.pydata.org/. accessed: 01.12.14.

Money, E.S., Reckhow, K.H., Wiesner, M.R., 2012. The use of Bayesian networks for nanoparticle risk forecasting: model formulation and baseline evaluation. Sci. Total Environ. 426 (0), 436–445. http://dx.doi.org/10.1016/j.scitotenv.2012.03.064.

Mueller, N.C., Nowack, B., 2008. Exposure modeling of engineered nanoparticles in the environment. Environ. Sci. Technol. 42, 4447–4453. http://dx.doi.org/10.1021/es7029637.

Müller, E., Hilty, L.M., Widmer, R., Schluep, M., Faulstich, M., 2014. Modeling metal stocks and flows: a review of dynamic material flow analysis methods. Environ. Sci. Technol. 48 (4), 2102–2113. http://dx.doi.org/10.1021/es403506a.

Page, B., Kreutzer, W., 2005. The Java Simulation Handbook : Simulating Discrete Event Systems with UML and Java. Shaker, Aachen, Germany.

Piccinno, F., Gottschalk, F., Seeger, S., Nowack, B., 2012. Industrial production quantities and uses of ten engineered nanomaterials in Europe and the world. J. Nanoparticle Res. 14 (9), 1–11. http://dx.doi.org/10.1007/s11051-012-1109-9.

Python Software Foundation, 2014. Python Language Reference, Vol. 2.7 from. http://www.python.org. accessed: 2014-08-07.

Savolainen, K., Alenius, H., Norppa, H., Pylkkänen, L., Tuomi, T., Kasper, G., 2010. Risk assessment of engineered nanomaterials and nanotechnologies—a review. Toxicology 269 (2–3), 92–104. http://dx.doi.org/10.1016/j.tox.2010.01.013.

Scott, D.W., 1992. Multivariate Density Estimation : Theory, Practice, and Visualization. Wiley, New York. http://dx.doi.org/10.1002/9780470316849.

Stiftung entsorgung schweiz, 2014. Swico, and stiftung Licht recycling schweiz (2014). Fachbericht.

Sun, T.Y., Gottschalk, F., Hungerbühler, K., Nowack, B., 2014. Comprehensive probabilistic modelling of environmental emissions of engineered nanomaterials. Environ. Pollut. 185 (0), 69–76. http://dx.doi.org/10.1016/j.envpol.2013.10.004.

TU Vienna, 2012. Institute for Water Quality, Resource and Waste Management. In: STAN, SubSTance Flow ANalysis, Vol. 2.5 from. http://www.stan2web.net/. accessed: 05.09.14.

van der Walt, S., Colbert, S.C., Varoquaux, G., 2011. The NumPy array: a structure for efficient numerical computation. Comput. Sci. Eng. 13 (2), 22–30. http://dx.doi.org/10.1109/MCSE.2011.37.

Walser, T., Gottschalk, F., 2014. Stochastic fate analysis of engineered nanoparticles in incineration plants. J. Clean. Prod. 80, 241–251. http://dx.doi.org/10.3390/ijerph120505581.

Wick, P., Clift, M.J.D., Rösslein, M., Rothen-Rutishauser, B., 2011. A brief summary of carbon nanotubes science and technology: a health and safety perspective. ChemSusChem 4 (7), 905–911. http://dx.doi.org/10.1002/cssc.201100161.