



**Országos Tudományos Diákköri
Konferencia
2025**



**Model-centric Integration of Expert Knowledge by Importance
Sampling**

Készítette: Kenyeres Éva

Témavezető: Prof. Abonyi János

Pannon Egyetem, Mérnöki Kar, Folyamatmérnöki Intézeti Tanszék

A pályamunka lezárásának dátuma: 2023.04.14.

Contents

1	Introduction	2
2	The proposed method for integration of expert knowledge into the system model	6
2.1	Representation of uncertain expert knowledge	8
2.2	Parameter estimation by importance sampling (IS)	9
2.2.1	Importance sampling based on the expert knowledge	10
2.2.2	Resampling of the parameters	11
2.3	Application of the proposed method for the validation of experts knowledge	12
2.4	Weighting the initial expert judgments	13
3	Applications to a waste separation system	15
3.1	Introduction to the waste separation technology	16
3.2	Available expert knowledge	17
3.3	Parameter estimation by importance sampling	17
3.3.1	Formalization of the system model	17
3.3.2	Identifying the estimable parameters	19
3.4	Results of the parameter estimation	20
3.5	Evaluation of the expert judgments	24
3.6	Elimination of the incorrect expert knowledge	26
4	Conclusions	28

Abstract

This study presents a method for integrating and validating uncertain expert knowledge and simulation models.

The parameters of the models of complex systems are often unknown, and measurement data is not available. Experience-based knowledge of experts can substitute the missing information, which is always subjective. Sometimes they are even not able to give an exact value for the unknown variable, only an interval it has to be in. Therefore, probability distributions can appropriately represent the knowledge of experts. If the model structure is known, this information can be implemented in it.

In the present study, a method is introduced based on the Monte Carlo (MC) simulation and importance sampling (IS) techniques for handling uncertain variables, which is applicable for parameter estimation and investigating the consistency of expert knowledge. Expert judgments can also be evaluated according to the results.

The contributions of the work: 1) An expert knowledge-based parameter estimation technique is given; 2) MC and IS techniques are used together to handle uncertain expert assessments; 3) A method is given for validating expert knowledge.

The experiments mentioned above are executed in a case study of an operating waste separation system.

Results verify that assessments of different experts can be efficiently integrated into system models. Thereby, the behavior of the systems is comprehensively estimable and the consistency of expert judgments can be evaluated.

Keywords: Expert knowledge, Parameter estimation, Waste separation system, Importance sampling, Monte Carlo simulation

Chapter 1

Introduction

The behavior of complex systems is difficult to predict. On the one hand, a lot of variables can be defined in these systems which depend on each other. On the other hand, the number of connections between the variables is also high. This aggregation of variables and connections is difficult to see through without using a model of the system. Although a model never can describe reality entirely, the aim is to approximate the real system as far as possible. If a valid model is available, it can also be used for the development and optimization of the system.

Models often appear in a mathematical form, where the outputs ($\mathbf{y}$) of the system are functions of the inputs ($\mathbf{u}$) and the parameters (Θ):

$$\mathbf{y} = f(\mathbf{u}, \Theta) \quad (1.1)$$

However, model parameters and inputs are often not known or are uncertain, especially when the technology is in the design phase. For example, the composition of the feed, the effect of a new catalyst on the reaction rate in a reactor, the heat transfer coefficient of a new structural material or coating, or the separation efficiency of a distillation column. These parameters are often estimated using available measurement data of the model output. For example, the kinetic parameters of a biochemical reaction system can be estimated from measurements of concentration [1]. However, measurement data are not always available. In this case, another source of information is needed. For instance, the experience-based knowledge of experts can be used.

As expert knowledge is subjective, the assessments of more experts can be aggregated and represented by a probability distribution [2]. If the model structure of the complex system ($f()$) is known, expert knowledge can be integrated into. An example of expert knowledge-based parameter estimation of algae growth is described in [3].

Thereby, the unknown parameters or inputs of the system can be estimated, thus refining the model, and outputs can also be calculated. The uncertain variables of the

model needed to be estimated are gathered in vector $\mathbf{x}$:

$$\mathbf{x} = [\mathbf{u}^T, \boldsymbol{\Theta}^T, \mathbf{y}^T]^T \quad (1.2)$$

and one elements of $\mathbf{x}$ is denoted by x_i in the paper. In the introduced case study about waste separation, the inputs are known from measurements, thus only the parameters and outputs needed to be estimated, according to the experience-based knowledge of experts.

As in the present study, experts are often not able to provide exact information. In such cases, the uncertain information can be handled using a probabilistic approach, similarly to fuzzy models. Recently it has been shown that expert knowledge can be implemented in decision making using fuzzy logic illustrated by an example of choosing the most suitable solar panel system [4].

As for industrial systems, operators usually have some uncertain intuitions about the parameters and the outputs of the technology. Transforming this information into a mathematical form, the gained probability distributions make it possible to estimate the parameters and outputs of the system if the model structure is known. The aim is to bring the two imprecise information of the parameters and the outputs to a consensus as seen in Figure 1.1.

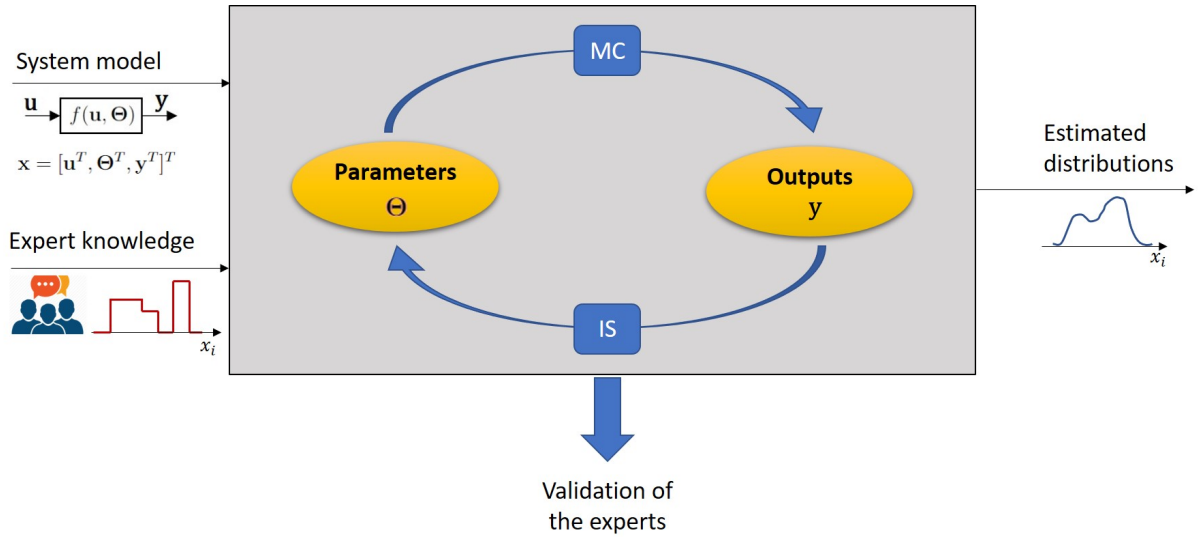


Figure 1.1: Parameter estimation and evaluation of expert knowledge based on the integration of expert-based information into the system model. The expert knowledge is available in the form of probability distributions of the parameters and outputs. Outputs are estimated by Monte Carlo (MC) simulation based on the system model and the sampled probability distributions of parameters, and parameters are deduced and refined according to the expert-based outputs by importance sampling (IS), iteratively. Expert knowledge can be also evaluated according to the results. $\mathbf{u}$, $\boldsymbol{\Theta}$ and $\mathbf{y}$ are represent the input, parameter and output vector of the system, respectively. The ones of them needed to be estimated gathered to vector $\mathbf{x}$, and its elements are marked by x_i .

The outputs can be evaluated by a Monte Carlo (MC) simulation from the expert-based probability distributions of the parameters and the model of the system, and results can be compared to those suggested by the experts. Thereby, the consistency of the expert knowledge about the parameters and outputs can be examined. According to the probability distributions of expert-based outputs, the parameters can be refined further by importance sampling (IS) technique. Executing these two steps iteratively, the imprecise informations of the parameters and outputs can be brought to a consensus, and the calculations converge to a result in the form of probability distributions.

MC simulations are often used to handle parameter uncertainty of a model. An example of this technique is presented in the case of a catchment model in [5]. Expert knowledge was also implemented to parameter estimation sometimes, e.g., in the case of a groundwater model [6], however, examples are rarely documented in the industrial field.

One contribution of this work is that the utilization of expert knowledge is introduced for an industrial system. A method is given how subjective, uncertain information about model parameters and outputs from more experts can serve as a basis for estimating these values. The introduced importance sampling based method applies to stationary systems. However, particle filter can be utilized similarly to the above-mentioned for dynamic ones.

As another novelty of this study a technique is provided that shows how results can be used to evaluate the reliability of expert knowledge. Comparing to the estimated probability distributions to the original ones provided by experts, the judgments of the experts can be characterized by goodness values thus evaluated. With this method, the incorrect judgments can be identified and even eliminated from the estimation by an outer iteration circle.

The above-mentioned experiments are executed and results are introduced through the case study of an operating Hungarian waste separation technology. This system satisfies the conditions above, namely, the parameters are not known, measurements are rarely available in this type of processes, and expert knowledge has a distribution because of subjectivity. Moreover, experts are not able to give their assessments about the parameters and outputs as exact values, only as intervals. With these features, the system is applicable to achieve the above-mentioned research goals.

The contributions of the study are listed below:

- An industrial example is shown how expert knowledge can be implemented to the modelling and analysis of the system.
- Monte Carlo (MC) method and importance sampling (IS) technique are integrated to handle uncertain expert knowledge.
- An expert knowledge based parameter estimation technique is given.

- The judgments of experts are evaluated and weighted according to the estimation results.
- Incorrect expert judgments are eliminated from the estimation by an outer iteration circle.

As for the roadmap of the work, in Section 2, the methodological background of the research is introduced, containing some ideas about the representation possibilities of uncertain expert knowledge in Section 2.1, describing the proposed parameter estimation technique based on IS in Section 2.2, introducing the application of the results for the validation of expert judgments in 2.3, and providing a method for the elimination of incorrect expert assessments in Section 2.4. The experimental results are provided in Section 3. The investigated waste separation system is introduced in Section 3.1, the available expert knowledge about the system in Section 3.2, the experiments and results of parameter estimation is presented in detail in Section 3.3 and 3.4, the evaluation of the expert judgments in Section 3.5, and the investigations on the outer iteration circle in Section 3.6. Finally, in Section 4, the main conclusions, limitations of the method and future research directions are revealed.

Chapter 2

The proposed method for integration of expert knowledge into the system model

In this section, the methodological background of the work is introduced. The initial, uncertain expert knowledge about the parameters and outputs is transformed into probability distributions. MC simulation is used to handle this knowledge, and IS to estimate the most likely probability distributions of the parameters and outputs based on the available information. To evaluate the experts, the distributions created from the intervals provided by them are weighted, comparing to the estimated distributions, thus goodness values are generated. Figure 2.1 outlines the method which elements are introduced in detail in this section.

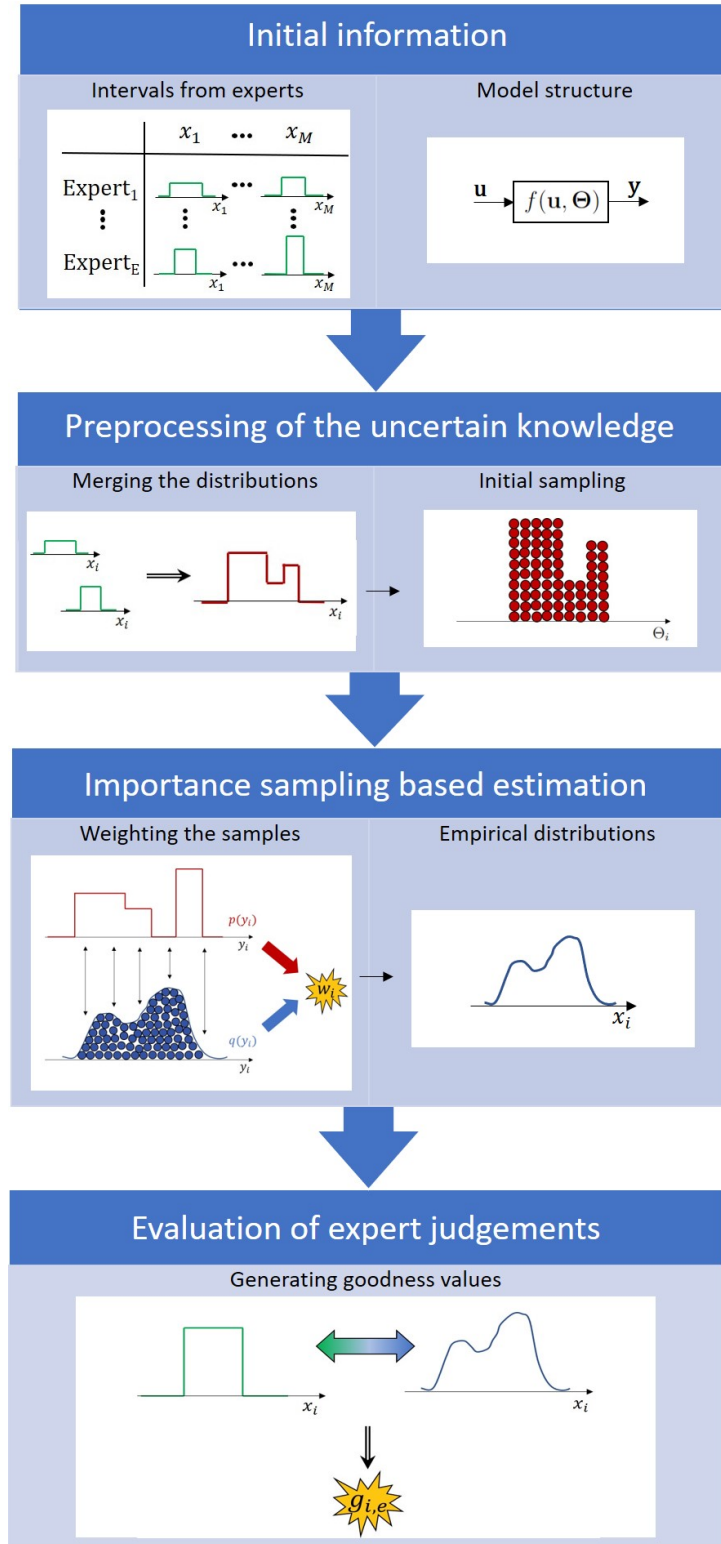


Figure 2.1: **The steps of the procedure used in the study to integrate expert knowledge into the system model.** The initial information from the experts is represented by uniform distributions, which are aggregated for each estimable x_i variable. Sample elements are drawn from the aggregated distributions of the parameters, and are weighted (w_i) according to those of the outputs (y_i). Thereby, empirical distributions are generated, and compared to the original expert-based distributions, thus creating goodness values ($g_{i,e}$) of expert judgments.

2.1 Representation of uncertain expert knowledge

Expert knowledge is a set of subjective assessments of several professionals who created their judgments independently from each other. If they can give their judgment about the unknown variable as an exact value, the uncertainty of the assessment can be modeled well with normal distribution. However, sometimes experts are not able to give an exact estimation for the unknown values, but they provide an interval by minimum $x_{i,min}$ and maximum $x_{i,max}$ bounds, e.g. in the case of the e th expert:

$$x_i \in [x_{i,min}^e, x_{i,max}^e] \quad (2.1)$$

This information is usually handled in the form of uniform distribution if there is no reason to assume that the experts consider some elements of the given interval more likely than others:

$$p_e(x_i) = \begin{cases} \frac{1}{x_{i,max}^e - x_{i,min}^e} & \text{for } x_{i,min}^e \leq x_i \leq x_{i,max}^e \\ 0 & \text{for } x_i < x_{i,min}^e \text{ or } x_i > x_{i,max}^e \end{cases} \quad (2.2)$$

$$(2.3)$$

where $p_e(x_i)$ represents the opinion of the e th expert about x_i .

This representation of expert knowledge was used, for example, for estimating the numerical meaning of linguistic variables [7].

If the assessments of more expert are available, the individual distributions should be aggregated for every x_i as seen in the second row of Figure 2.1 by forming the mean of them. Thereby, every expert judgments were handled with the same weights:

$$p(x_i) = \frac{\sum_{e=1}^E p_e(x_i)}{E} \quad (2.4)$$

where E denotes to the number of experts.

If some variables of the system are available as a probability distribution instead of an exact number, a sample has to be generated that represents them. Monte Carlo simulation is a good solution for the problem. According to the technique, samples are drawn from the probability distributions of the uncertain inputs ($p(u_i)$) and parameters ($p(\Theta_i)$):

$$p(x_i) \approx \{x_i^j\}_{j=1}^N \quad (2.5)$$

where x_i^j is the j th sample element from $p(x_i)$, and N is the number of the samples.

Thereby, the original continuous probability distribution is represented by a discrete sample as seen in Figure 2.2.

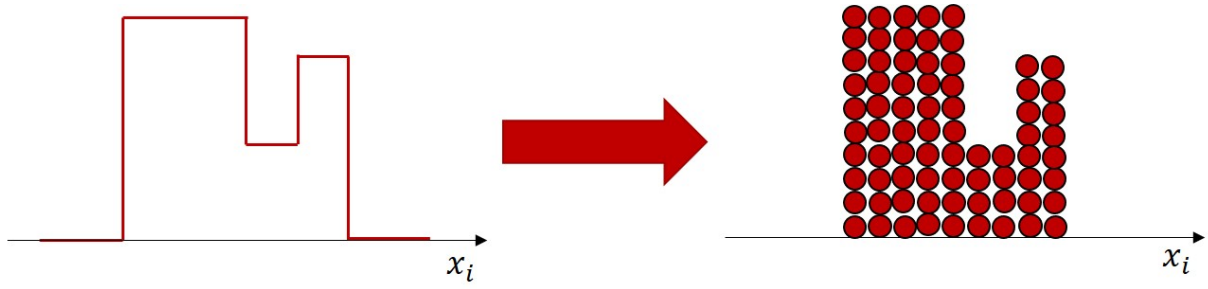


Figure 2.2: **The outline of MC simulation.** Samples are drawn from the continuous distribution to represent it. The variable that the distribution belong to is denoted by x_i .

Then, the output of the model can be calculated for every sample element according to Equation 1.1. If the sample size is large enough, the resulting probability distributions of the outputs will be consequently the same for every run of the simulation.

Monte Carlo (MC) simulation is widely used to handle input or parameter uncertainty. For example, the performance of a solar collector was predicted under uncertain operating conditions [8], the cooling load prediction of an HVAC system refined besides uncertain inputs [9], or the product yield of a polymerization process with several possible reaction paths estimated besides uncertain kinetic parameters [10].

In this study, the input of the system is available from measurements, however, the uncertain information about the parameters and outputs provided by the experts in the form of data ranges is treated by the technique described above. The judgments of experts were handled as uniform distributions, and as the assessments of more experts were available, the distributions were aggregated. Then, the distributions of the parameters were sampled and a Monte Carlo simulation was executed to calculate the estimated outputs, and the aggregated distributions of outputs are used to validate the results that will be described in the next section.

2.2 Parameter estimation by importance sampling (IS)

If the assessments about the parameters and the outputs were gained from two independent groups of experts, these pieces of information do not harmonize with each other. In the following, a technique is introduced to how these elements of knowledge can be made consistent.

According to the estimated outputs in Section 2.1, the model of the system can be validated by involving some measurement data and comparing them to the estimated values. However, if measurement data are not available, asking for some information from the experts is again a favorable possibility [11].

Similarly to the case of parameters, experts sometimes can provide only intervals of

the outputs, that can be modeled by uniform distributions. Thereby, an aggregated distribution of them provides the basis of model validation, too.

Using these pieces of information, the parameters of the model can be estimated by an iterative technique: the sample elements of the output got by the MC simulation are weighted by IS, thus new, empirical probability distributions of the parameters are generated, and new samples are drawn from them. Then, the output is calculated again for every sample element. These steps are repeated until the results converge. An outline of the procedure can be seen in Figure 2.3.

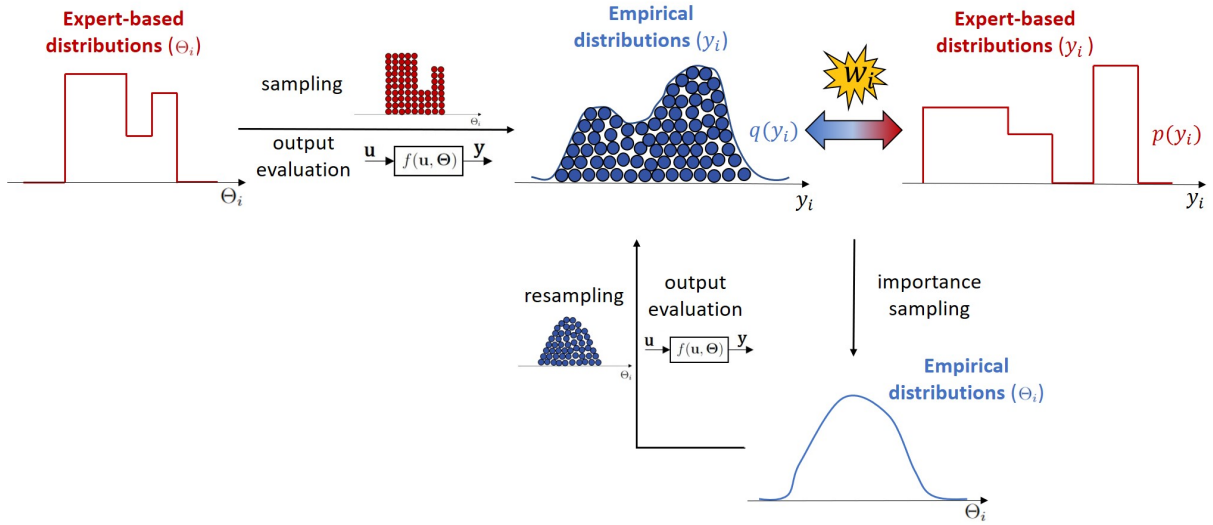


Figure 2.3: **The schematic figure of the parameter estimation technique using expert knowledge.** It is assumed that the input vector ($\mathbf{u}$) of the system is known. Θ and $\mathbf{y}$ denote the parameter and output vector of the system with the elements of Θ_i and y_i , respectively. w_i refers to the weights generated by importance sampling. The expert-based distributions are represented by red color and the empirical distributions by blue.

2.2.1 Importance sampling based on the expert knowledge

The importance sampling (IS) technique is suitable to give weights to the sample elements by comparing the resulting distribution of outputs with the expert-based ones in every iteration step. Thereby, new empirical distributions of the parameters can be generated.

As performed in Section 2.1, the estimated outputs are available as sampled densities from the MC simulation. The empirical probability density functions (pdf) fitted to these will serve as so-called importance functions or proposal distributions, denoted to $q(y_i)$ in Figure 2.3 where y_i is an element of the output vector ($\mathbf{y}$). The expert-based distributions used for the comparison are denoted to $p(y_i)$ as seen in Figure 2.3.

Interoperability between $p(y_i)$ and $q(y_i)$ can be described as:

$$E_p[y_i] = \int y_i p(y_i) dy_i = \int y_i \frac{p(y_i)}{q(y_i)} q(y_i) dy_i = E_q[y_i \cdot \frac{p(y_i)}{q(y_i)}]. \quad (2.6)$$

where $E_p[\cdot]$ and $E_q[\cdot]$ denotes the expected values of the argument that is distributed according to $p(\cdot)$ and $q(\cdot)$, respectively.

Therefore, regarding discrete distributions:

$$E_p[y_i] \approx \frac{1}{N} \sum_{j=1}^N y_{i,p}^j = \frac{1}{N} \sum_{j=1}^N y_{i,q}^j \cdot \frac{p(y_{i,q}^j)}{q(y_{i,q}^j)} = \frac{1}{N} \sum_{j=1}^N y_{i,q}^j w_i^j \quad (2.7)$$

where $y_{i,p}^j$ and $y_{i,q}^j$ are the j th element sampled from $p(y_i)$ and $q(y_i)$, respectively, w_i^j denotes to the so-called importance weight of $y_{i,q}^j$, and N to the sample size.

Equations 2.6 and 2.7 clarify that sample elements can be weighted by calculating w_i^j .

The original sample elements have equal weights. They get new, normalized weights according to the $p(y_i)$ distribution provided by the experts as

$$w_i^j = \frac{p(y_i^j)}{q(y_i^j)}, \quad (2.8)$$

where y_i^j is the i th output belonging to the j th sample element, and w_i^j denotes to the weight of this element according to y_i^j .

The method is suitable to apply for weighting the sample elements to accomplish the parameter estimation.

2.2.2 Resampling of the parameters

As the sample elements were weighted by importance sampling, new probability distributions of the parameters come up. New sample elements are drawn from these empirical distributions.

The new elements are drawn from the old ones with the probability of their weights. Firstly, N random numbers (s_n) are generated from the uniform distribution on $(0, 1]$. Then, for every s_n , the m th sample element is chosen, if

$$\sum_{j=1}^{m-1} w_i^j < s_n \leq \sum_{j=1}^m w_i^j \quad (2.9)$$

where w_i^j represents the weights of the elements [12].

Figure 2.4 introduces the scheme of resampling.

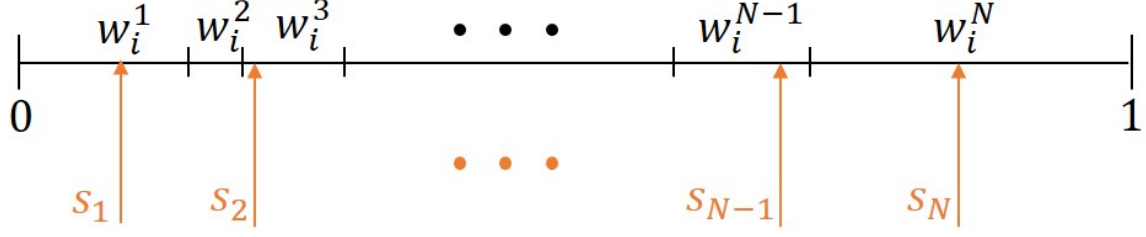


Figure 2.4: **The schematic figure of resampling.** The generated $s_j \in (0, 1]$ random numbers are represented by orange arrows and the weights of the sample elements are denoted by w_i^j ($j = 1, \dots, N$).

The new samples are handled with equal weights. In the present work, more parameters are needed to estimate, thus all of them are resampled and assigned to the sample elements independently. Then, the outputs are calculated again for each sample element, and the iteration seen in Figure 2.3 is executed until the empirical distributions converge.

2.3 Application of the proposed method for the validation of experts knowledge

According to the results, the expert knowledge can be evaluated. The estimated empirical distributions ($q(x_i)$) of the parameters and outputs can be compared to the original ones from the experts ($p_e(x_i)$).

One of the most simple technique to compare two probability distributions is the measure of Jaccard index, which will be used as goodness values of expert judgments. As the original method is able to represent the similarity of two datasets, here a generalization of the technique is needed to apply.

Firstly, the two density functions desired to compare should be discretized by a Δx_i value and the density functions evaluated at that points. Then, the Jaccard index of the two resulting datasets can be calculated [13]. Thereby, the goodness values of the expert judgments:

$$g_{i,e} = J(\mathbf{p}_e, \mathbf{q}) = \frac{\sum_{k=1}^K \min(p_{e,k}, q_k)}{\sum_{k=1}^K \max(p_{e,k}, q_k)} \quad (2.10)$$

where $g_{i,e}$ marks the goodness value belonging to the judgment of the e th expert about x_i , $\mathbf{p}_e$ and $\mathbf{q}$ represent the two datasets gained from $p_e(x_i)$ and $q(x_i)$, K refers to the size and $p_{e,k}$ and q_k denote to the elements of them, respectively.

The scheme of the method can seen in Figure 2.5.

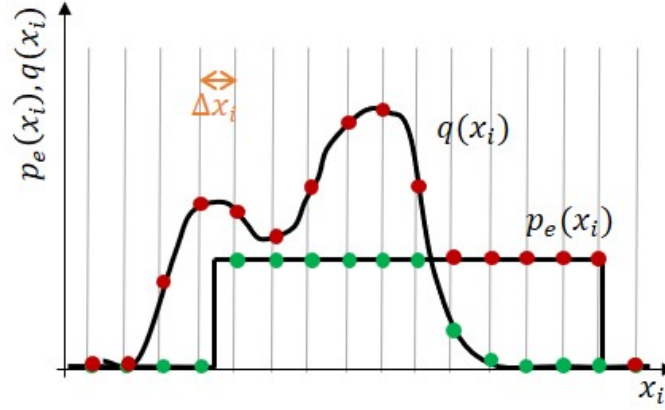


Figure 2.5: **Comparison of two probability distribution by calculating the Jaccard index.** The $p_e(x_i)$ and $q(x_i)$ distributions are discretized by Δx_i and probability density functions are evaluated at that points. The elements two datasets are compared pairwise: the maximum and minimum values are denoted to red and green color, respectively. Then, Equation 2.10 is used to calculated Jaccard index.

Applying the technique for comparing the resulting empirical probability densities and the original expert-based ones, the expert assessments can be characterized by the calculated goodness values ($g_{i,e}$), thus evaluated.

2.4 Weighting the initial expert judgments

The goodness values resulted in Section 2.3 are also applicable to give weights to the initial uniform distributions from the experts ($p_e(x_i)$) and generate an outer iteration circle to refine the estimation further. The weights can be created by the normalization of the goodness values for each x_i :

$$\tilde{g}_{i,e} = \frac{g_{i,e}}{\sum_{e=1}^E g_{i,e}} \quad (2.11)$$

where $\tilde{g}_{i,e}$ refers to the normalized goodness value of the judgment of the e th expert about x_i .

Using these weights, new aggregated distributions can be created from the initial ones as:

$$p(x_i) = \sum_{e=1}^E \tilde{g}_{i,e} p_e(x_i) \quad (2.12)$$

Thereby, feedback is gained from the estimated distributions as seen in Figure 2.6.

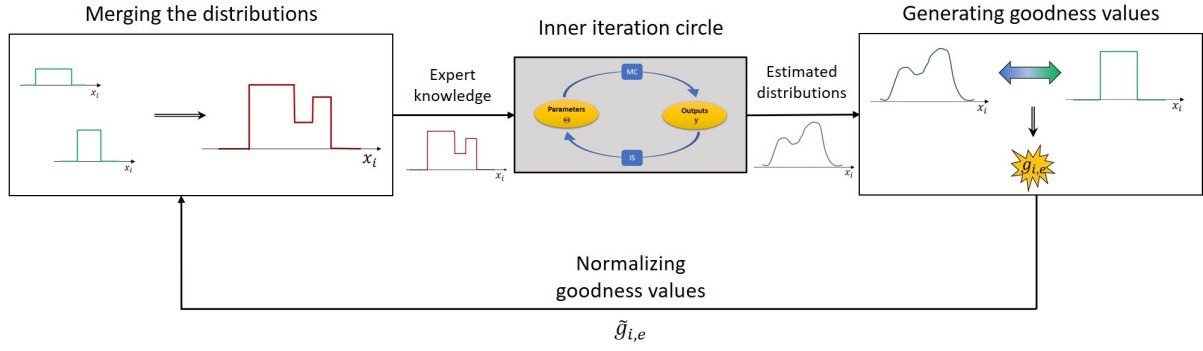


Figure 2.6: **Outline of the outer iteration circle.** The initial distributions from experts are merged again in every iteration according to the normalized goodness values ($\tilde{g}_{i,e}$) gained by comparing these distributions to the estimated ones resulted in the inner circle.

Weighting the initial distributions according to their correctness gives the possibility to eliminate the incorrect expert judgments, thus refining the estimation. The convergence in the outer circle is also investigated in this study.

Chapter 3

Applications to a waste separation system

Nowadays there is a great emphasis on waste management in our society. As the population and consumption grow, the amount of municipal waste is also becoming higher and higher which has to be handled in some way. The main area to focus on is waste recycling.

To be able to recycle waste, it should be efficiently separated by different types of materials at first, and contaminants have to be removed. For this aim, many types of waste industrial equipment can be used based on the composition of the waste, thus building the waste sorting technology. As the quality of municipal waste is not constant because of the changing material regulations, technologies also have to be adapted to these changes. This flexibility requires knowing the model of the system and its parameters as far as possible.

Due to the importance of the issue, my investigations were conducted on a real, operating waste separation system in Hungary. This gives the chance to gather information from experts who work with the technology for a long time. The system, the handling of expert knowledge and the output estimation of the system were described thoroughly in [14]. In the present study, the parameter estimation and the evaluation of the expert knowledge is investigated.

Thereinafter, the technology is described shortly in Section 3.1. Section 3.2 contains a short description of the available expert-based data, Section 3.3 and 3.4 the principles and the results of parameter estimation, respectively. Section 3.5 introduces how the results can be used for the evaluation of expert knowledge. Section 3.6 presents the results applying an outer iteration circle to eliminate the incorrect expert knowledge from the estimation.

3.1 Introduction to the waste separation technology

The technological scheme of the waste processing plant was chosen to examine can be seen in Figure 3.1 [14]. It operates with seven units. The pre-shredder has one input and one output, as it changes only the size distribution of the waste. Thereby, the compositions of its input and output are the same. All the other units can be considered as flow splitters: they have one input and two outputs in all cases. Magnetic and eddy-current separators remove iron and aluminium from the waste, respectively. Drum screen extracts non-recyclable components (RDF: refuse-derived fuel) of the waste, those go for burning to produce electricity. The air separator splits the rest of the waste based on its shape to 2D and 3D components. Finally, optical separators working in the near-infrared (NIR) spectra sort the waste by material quality.

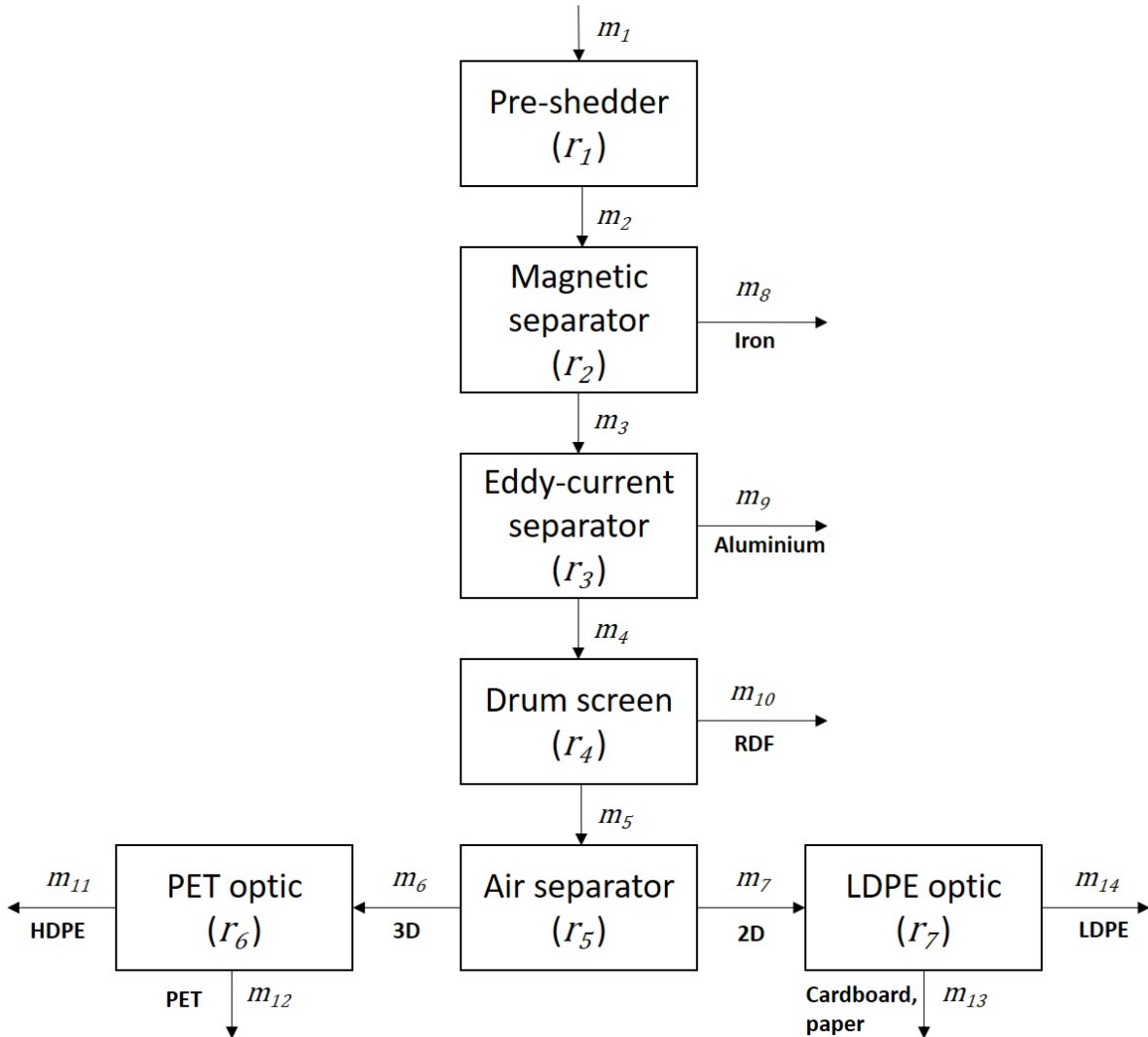


Figure 3.1: **Block diagram of the waste processing technology.** In the figure, m_i ($i = 1, \dots, 14$) represents the technological mass flows, and r_j ($j = 1, \dots, 7$) represents the set of the separation efficiencies of the j th unit on components. The titles of the arrows indicate in which component the related flow is rich. RDF: refuse-derived fuel.

3.2 Available expert knowledge

Expert knowledge was surveyed by asking them about the separation efficiencies and some yields of the system. As introduced in [14], they provided their assessment about all of the r_j ($j = 1, \dots, 7$) values for all components, and the yields of the PET, LDPE, iron and aluminium components for the outflows they accumulate. Knowing the feed of the system from measurements, the information about the yields depends only the leaving component mass flows.

The separation efficiencies and the yields were asked from two independent groups of experts with two and three members, respectively [14]. As was revealed in Section 2.1, they were not able to provide exact values, but intervals which were handled as aggregated uniform distributions.

3.3 Parameter estimation by importance sampling

MC simulation has already been used for the chosen waste separation technology to represent the expert-based distributions of separation efficiencies by discrete sample, and estimate the outflows. It was shown that if the sample size is large enough, the resulting distributions of the outflows are the same for all run of the simulation. These distributions were also compared with expert-based intervals about the outflows where it was available, thus validating the model [14].

Using importance sampling (IS), a feedback is gained: sample elements can be weighted based on the available distributions about the outflows, and separation efficiencies resampled according to these weights. Thereby, the parameter estimation of the system can be executed by the iterative circle introduced before in Figure 1.1. In the present case, separation efficiencies serve as parameters (Θ), and the yields form the outputs ($\mathbf{y}$) of the system.

For applying the method, firstly, the structure of the system model is defined in Section 3.3.1. Then, the estimable parameters are identified in Section 3.3.2.

3.3.1 Formalization of the system model

The model of the introduced waste separation system has one input ($u = m_1$), which mass and composition are known from measurements. The component mass flows going in and out of the units are used as state variables. The parameters of the model are the separation efficiencies ($\Theta = [r_1, r_2, \dots, r_7]^T$) which are specified by components. They always give the ratio of the outflow marked with a smaller number and the inflow of the unit for a chosen component.

The state-space model of the system in steady-state can be set up for component c as

$$0 = \mathbf{m}^c(\mathbf{A}^c - \mathbf{I}) + \mathbf{B}^c u^c \quad (3.1)$$

where $\mathbf{m}^c$ denotes to a 14-element vector that contains the component mass flows of component c . $\mathbf{A}^c$ refers to the state transition matrix of component c with a size 14×14 , which non-zero elements represent the connections between the flows. $\mathbf{I}$ refers to the unit matrix with the same size as $\mathbf{A}^c$. $\mathbf{B}^c$ denotes to the input vector $([1, 0, \dots, 0]^T)$, and $u^c = m_1^c$ to the input of the system in the case of component c .

Matrix $\mathbf{A}^c$ defines the connections between the flows containing the separation efficiencies belonging to component c . The structure of it is the following.

$$\mathbf{A}^c = \begin{bmatrix} 0 & r_1^c & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & r_2^c & 0 & 0 & 0 & 0 & 1 - r_2^c & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r_3^c & 0 & 0 & 0 & 0 & 1 - r_3^c & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & r_4^c & 0 & 0 & 0 & 0 & 1 - r_4^c & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & r_5^c & 1 - r_5^c & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & r_6^c & 1 - r_6^c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & r_7^c & 1 - r_7^c & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

As for the outputs, they can be defined as:

$$\mathbf{y}^c = \mathbf{C}^c \mathbf{m}^c \quad (3.2)$$

where the output vector is marked by $\mathbf{C}^c$, and the output of the system by $\mathbf{y}^c$ for component c .

Vector $\mathbf{C}^c$ selects the outputs of the system from $\mathbf{m}^c$ states and transform them to yields:

$$\mathbf{C}^c = [0, 0, 0, 0, 0, 0, 0, \frac{1}{u^c}, \frac{1}{u^c}, \frac{1}{u^c}, \frac{1}{u^c}, \frac{1}{u^c}, \frac{1}{u^c}, \frac{1}{u^c}] \quad (3.3)$$

It can be seen that the formalized model of the waste separation system is congruent with the generalized one described by Equation 1.1.

3.3.2 Identifying the estimable parameters

As seen from the formalization in Section 3.3.1, the component mass flows of the system can be calculated independently for each component from the relevant separation efficiencies. Therefore, the r_i^c ($i = 1, \dots, 7$) parameters have an effect only on the mass flows of the c component, and on the others not. As expert-based information about the output is available only to four outflows out of the seven, and only to the component in which that flow is rich, separation efficiencies belonging to the rest of the components are not estimable based on the provided data. The reduced $\mathbf{y}$ vector can be defined as:

$$\mathbf{y} = [\frac{m_{12}^{PET}}{u^{PET}}, \frac{m_{14}^{LDPE}}{u^{LDPE}}, \frac{m_8^{Iron}}{u^{Iron}}, \frac{m_9^{Aluminium}}{u^{Aluminium}}]^T \quad (3.4)$$

In addition to the above-mentioned, only the parameters of those units affect the outputs, which are before the output location in the row of units as seen in Figure 3.2.

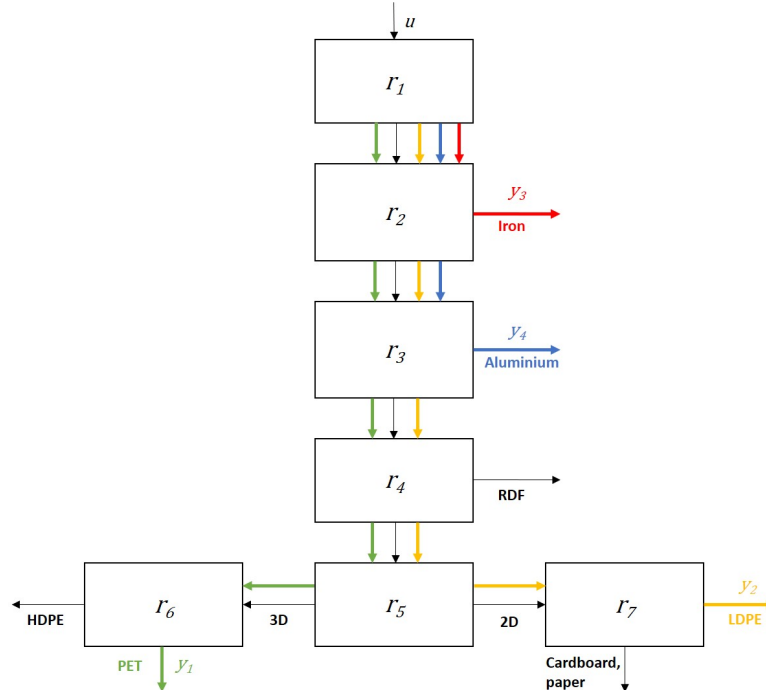


Figure 3.2: **Effects of the parameters on the outputs.** The four outputs about which expert-based information was obtained are referred to y_i ($i = 1, \dots, 4$). Green, yellow, red and blue colors mark which units have an effect on the different outputs. u denotes to the input, and r_i ($i = 1, \dots, 7$) represents the parameters of the i th unit.

Considering the above-mentioned issues, the parameters that can be estimated are summarized in Table 3.1. The table does not contain the pre-shedder unit as it has only one output, thus its separation efficiencies are surely 100% for all components.

Table 3.1: **Summary of the estimable parameters.** The components of which expert-based distribution about the output is available are represented by red color. The gray cells represent the estimable parameters, and the white cells the non-estimable ones. Columns specify the unit and rows refer to the component to which the estimable or non-estimable separation efficiency belongs.

	Magnetic separator (r_2^c)	Eddy-current separator (r_3^c)	Drum screen (r_4^c)	Air separator (r_5^c)	PET optic (r_6^c)	LDPE optic (r_7^c)
PET	✓	✓	✓	✓	✓	
LDPE	✓	✓	✓	✓		✓
HDPE						
Hazardous						
Iron	✓					
Aluminium	✓	✓				
RDF						
Cardboard						
Paper						

3.4 Results of the parameter estimation

The parameter estimation of the system was executed by the iterative circle represented in Figure 2.3 for the parameters marked in Table 3.1. The simulation stop criteria was defined by the two-sample Kolmogorov-Smirnov test besides 20% confidence interval. As will be seen later, this criteria is strict enough to stop the iteration. $M = 3000$ sample size was used which is large enough to not affect the result [14].

During the simulation, all the estimable r_i^c parameters introduced in Table 3.1 were resampled according to the importance weights generated based on the relevant output simultaneously. The simulation stopped when distributions of all the parameters and outputs converged. For that, 19 iterations were needed besides 20% confidence interval. Figure 3.3 and 3.4 shows the convergence of outputs, and Figure 3.5 that of the $r_3^{Aluminium}$ parameter.

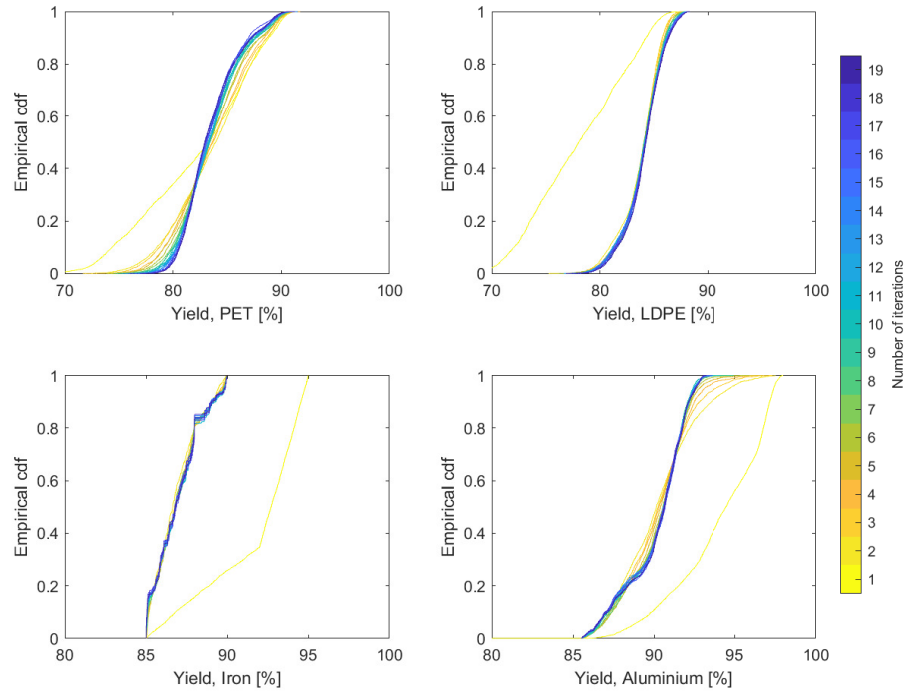


Figure 3.3: **Empirical cumulative density functions (cdf) of the outputs along the iteration steps.** The lines turn blue from yellow as the iteration is progressing.

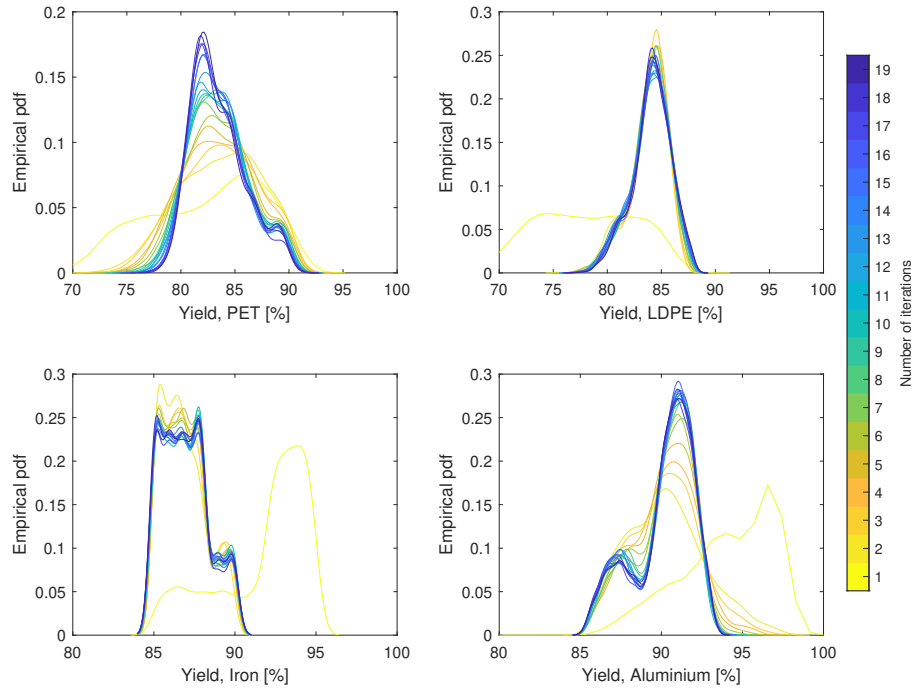


Figure 3.4: **Empirical probability density functions (pdf) of the outputs along the iteration steps.** The lines turn blue from yellow as the iteration is progressing.

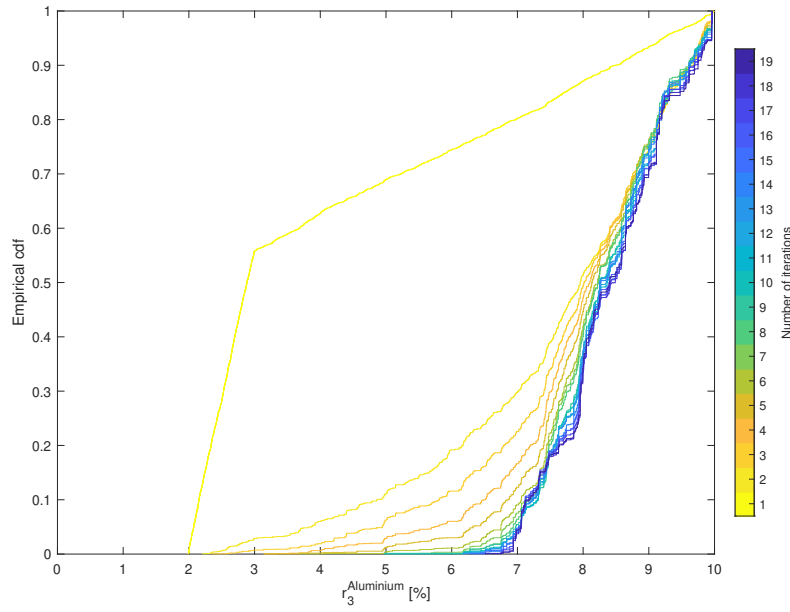


Figure 3.5: **Empirical cumulative density functions (cdf) of the $r_3^{Aluminium}$ parameter along the iteration steps.** The lines turn blue from yellow as the iteration is progressing.

As seen in Figure 3.3, 3.4 and 3.5, the convergence is satisfactory, thus no more iteration is needed. The convergence of the other parameters provides similar results.

The estimated outputs are shown with the intervals provided by experts in Figure 3.6.

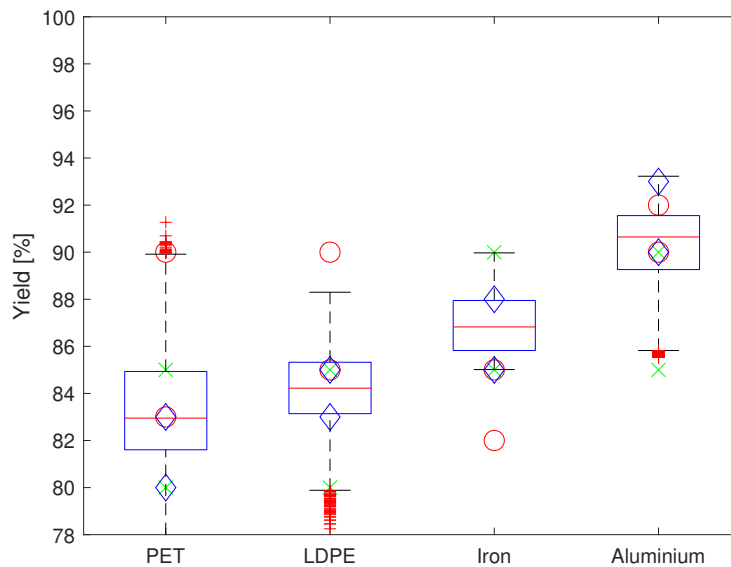


Figure 3.6: **The estimated and the expert-based outputs.** Formers are represented by boxes with a red line in the median, and latters by markers (red o, green x and blue diamond) at the edges of the provided intervals by the three experts.

Comparing these results with that of [14], it can be concluded that standard deviations are much smaller in case of iron and aluminium output, and medians are more likely within the limits of experts. Thereby, the output estimation is more precise with the improved method used in this study. This result induces that the estimated parameter distributions must also be more accurate than that provided by the experts.

As for parameter estimation, results belonging to component LDPE can be seen in Figure 3.7.

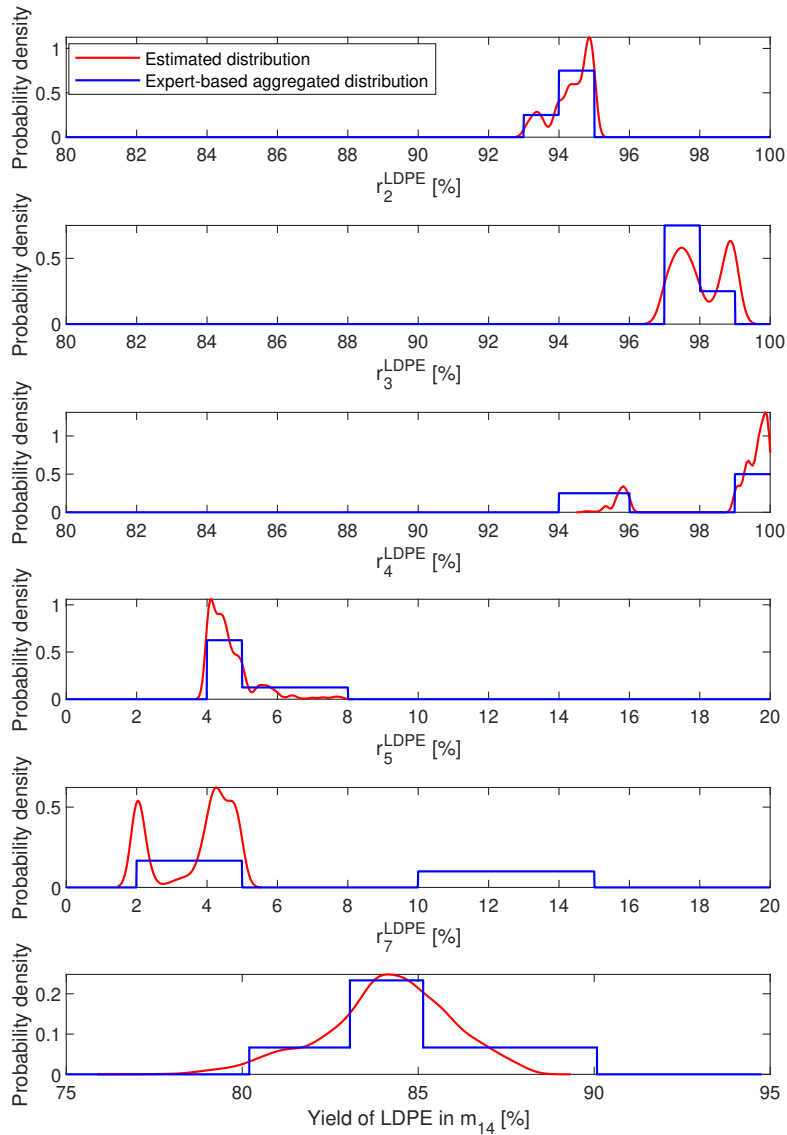


Figure 3.7: Estimated and expert-based distributions of the estimable parameters and output belonging to LDPE.

The estimation results can be compared to the expert knowledge based on Figure 3.7. As seen, estimated distributions fall into the intervals covered by expert-based information in all the cases. Otherwise, the estimated output is nearly normal, because it depends on the product of five uncertain parameters, so the differences equalize.

As also seen in Figure 3.7, in some cases, experts gave disjunct intervals. It can be easily assumed that in these cases one of them is wrong. The estimated distributions usually cover only the preferable one of these intervals, and the other is neglected. Therefore, if the estimated distributions are used for sampling the parameters, some wrong information are eliminated, thus more exact output estimation can be provided as was declared above.

The individual expert judgments can also be evaluated according to the estimation results. In Section 3.5, the method is introduced in detail.

3.5 Evaluation of the expert judgments

This section aims to give an evaluation method for the provided expert judgments by comparing the original expert-based distributions before aggregation to the estimated ones.

An example of graphical comparison of two distributions can be seen in the Q-Q plot of r_2^{LDPE} in the case of Expert 2. in Figure 3.8.

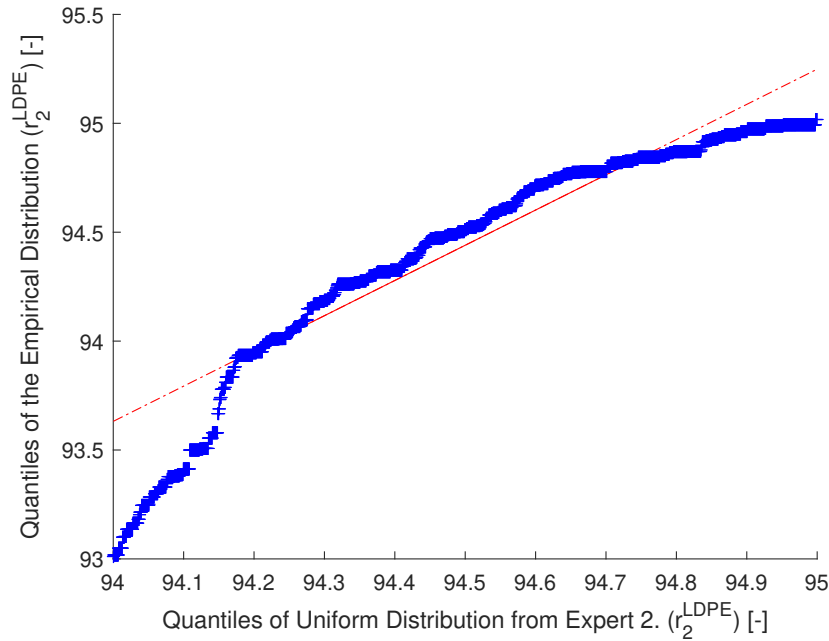


Figure 3.8: **Q-Q plot for comparing the uniform distribution provided by Expert 2.** and the estimated empirical distributions of r_2^{LDPE} . If the two distribution were the same, the plot would appear linear.

As seen in Figure 3.8, the plot is nearly linear, so it is supposed that the two distributions are close to each other in the case of r_2^{LDPE} . To give a more precise measure, Jaccard index is calculated which serves as the goodness values of the expert judgments.

Using the resulting empirical distributions in Section 3.4, the goodness values ($g_{i,e}$) of the expert judgments can be calculated according to the method introduced in Section 2.3. This measure represents how similar the distribution given by an expert is to the estimated one.

The outputs were assessed by three experts, and the parameters by two who were independent from the formers. The uniform distributions created from the intervals they were provided were compared to the estimated distributions. The calculated goodness values for the outputs and the parameters can be seen in Figure 3.9 and 3.10, respectively.

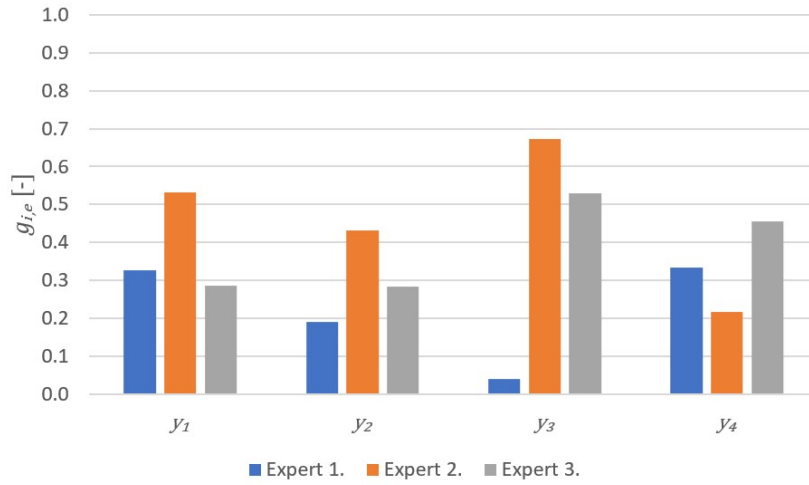


Figure 3.9: Goodness values belonging to the outputs.

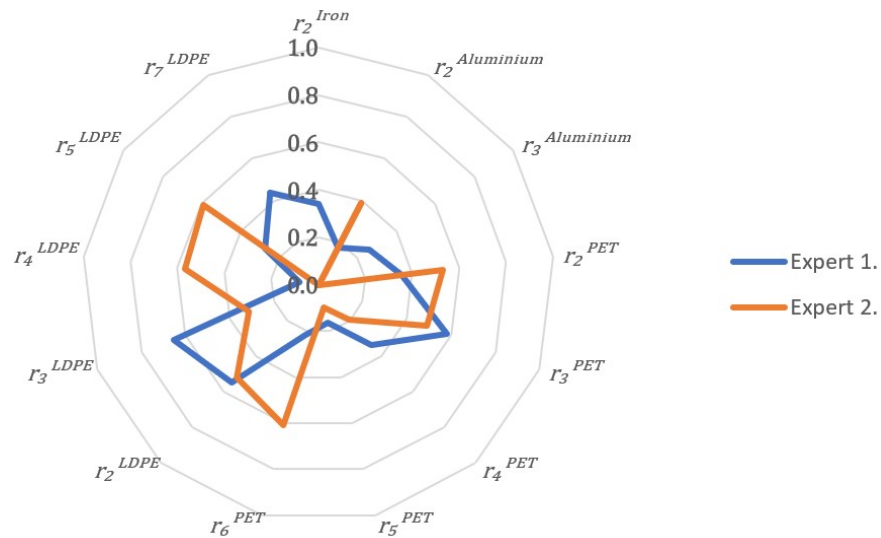


Figure 3.10: Goodness values belonging to the estimable parameters. The labels on the axes marks which parameter the goodness value belongs to.

Goodness values fall into the $[0, 1]$ interval showing the similarity of the two distributions. If they are the same, the goodness value belonging to them is 1, and if they are completely different, they get 0. As seen in Figure 3.9 and 3.10, there are a few values around 0 which suggests incorrect expert judgments would be preferable to eliminate from the estimation.

3.6 Elimination of the incorrect expert knowledge

As was declared in 3.5, there are some low-weight expert judgments in the case of the parameters and also the system outputs. To eliminate them from the estimation, an outer iteration circle is suggested to generate by the weighting of the initial distributions according to the technique introduced in detail in Section 2.4.

At the beginning of the procedure, during the aggregation of the probability distributions from experts, the judgments of each expert were considered with equal weights. However, as seen before, some assessments seem to be closer to the real values than others. By normalizing the goodness values for each variable, weights can be given to the initial uniform distributions from the experts, and new aggregated distributions can be created as the input of the estimation in the inner circle. Thereby, an outer iteration circle can be generated as was introduced in Figure 2.6.

New aggregated distributions were created in the case of the parameters and outputs, too. The convergence of the outputs of the inner circle in the case of the system outputs (**y**) are introduced in Figure 3.11 for the first 20 iteration steps in the outer circle.

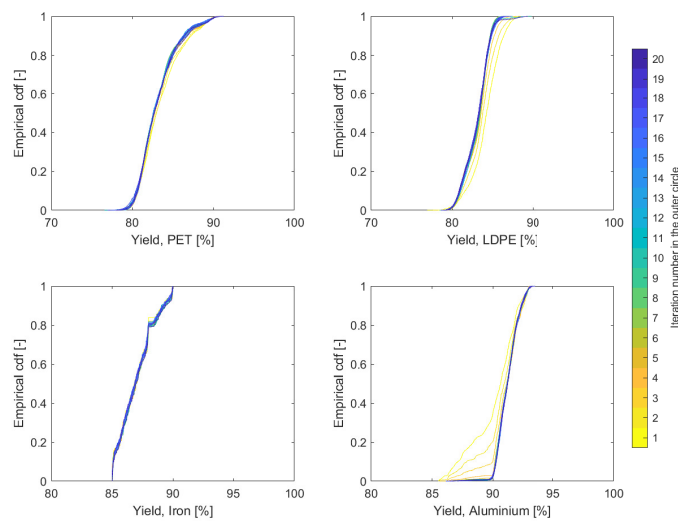


Figure 3.11: **Convergence of the empirical distributions of the four outputs (**y**) estimated in the inner circle.** The lines turn blue from yellow as the iteration is progressing in the outer iteration circle.

As seen in Figure 3.11, system outputs converge, thus the estimation is refined. The convergence of the normalized goodness values belonging to the outputs and the number of iterations in the inner circle along the outer iterations can be seen in Figure 3.12.

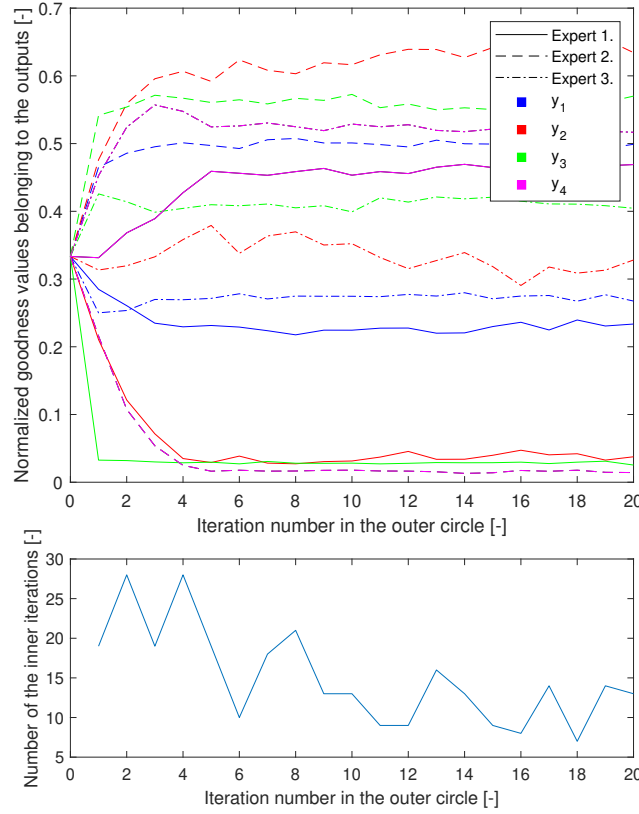


Figure 3.12: **Normalized goodness values ($\tilde{g}_{i,e}$) belonging to the system outputs (y) and the number of inner iterations along the outer iterations.** The three experts are marked with different line styles and the four outputs with different colors.

As seen in Figure 3.12, all the lines start from 0.33, as the expert judgments were considered with equal weights at the first iteration. However, after a few iterations, all the lines fluctuate around specific values. Some of them are near zero, which means the incorrect judgments are largely eliminated from the estimation. It is also seen that the number of iterations in the inner circle decrease after the weights are set. This suggests that the aggregated distributions of the parameters and outputs forming the input of the inner circle become more consistent thus less iteration is needed for the convergence in the inner circle.

As seen from the results, the outer iteration circle helped to refine the estimation by decreasing the weights of the less correct expert judgments and increasing the better ones.

Chapter 4

Conclusions

This work introduced a method for the integration of expert knowledge into system modeling.

As shown in this study, lack of exact information about the parameters (or inputs), experience-based knowledge of experts can be applied well for estimating them along with the outputs of the technology.

As expert knowledge is subjective, it can be represented by probability distributions. In the present work, Monte Carlo (MC) simulation was used to calculate the system outputs besides uncertain parameters. Furthermore, importance sampling (IS) was applied to involve expert knowledge about the outputs by comparing them to the results, thus deducing the parameters and generating an iterative circle.

According to the results, the correctness of the expert judgments was also evaluated using the Jaccard index as a measure of similarity. Thereby, weighting the expert judgments, new aggregated distributions were created from the original expert assessments, and the estimation was refined further by an outer iteration circle.

The method was applied on an operating Hungarian waste separation system, where separation efficiencies by components served as model parameters, and yields as the outputs. Expert-based information was available about all the parameters and four outputs by two independent groups of experts. They could give only intervals instead of exact values, which were represented by uniform distributions.

In the present study, a method was provided applying IS to execute the parameter estimation based on the available expert knowledge, and investigating the consistency of the expert-based information about the parameters and the outputs. The estimated distributions of the yields by the IS-based technique were more precise than that applying only MC.

It could also be noticed that there were some disjunct expert opinions, which implies some of them is wrong. For investigating this issue, a technique was also provided for

the validation of expert knowledge according to the estimation results. The incorrectness of some expert judgments also appeared calculating the Jaccard index which gave values near zero in these cases. To eliminate these assessments, goodness values of the judgments were normalized, thus weighted, and new aggregated initial distributions were calculated for the parameters and outputs. It was introduced that estimated distributions can be further refined by creating an outer iteration circle, where weights of the expert judgments converge.

The main limitations of the proposed methodology are the following. First of all, a sufficient number of experts are needed, who are qualified at the required level and experienced enough. The advantage of the method is that incorrect expert judgments can be eliminated, however, only in the case if most of the judgments are correct, otherwise, the estimation will be wrong. Another limitations is, that the model structure of the system have to be known, otherwise, the proposed method cannot be applied. Last but not least, the experiments in the present study were executed on a stationary model, but not yet tested on a dynamic case.

In the future, the applicability of the methodology can be investigated further. The present technique is applicable in systems where models and expert knowledge are available. However, some modifications may be needed. In the case of dynamic models, particle filter is applicable instead of IS. Expert knowledge can also be available in a different form, e.g. other types of information, which is needed to be represented correctly. The topic of integrating expert knowledge and measurement data simultaneously is also a possible research direction, that helps us to get the most complete information about the system.

List of symbols

The following symbols are used in this manuscript:

$\mathbf{y}$	a vector gathers the outputs
$\mathbf{u}$	a vector gathers the inputs
Θ	parameter vector
$f()$	model structure of the complex system
$\mathbf{x}$	a vector gathers the estimable variables
M	the size of $\mathbf{x}$
Θ_i	an element of Θ
y_i	an element of $\mathbf{y}$
x_i	an element of $\mathbf{x}$
$x_{i,min}^e$	the lower bound of the interval provided by the e th expert as a judgment about x_i
$x_{i,max}^e$	the upper bound of the interval provided by the e th expert as a judgment about x_i
$p_e(x_i)$	the uniform distribution that the interval provided by the e th expert about x_i was transformed to
$p(x_i)$	the expert-based aggregated distribution of x_i
E	number of the experts
x_i^j	the j th sample element from $p(x_i)$
N	sample size
$q(x_i)$	the empirical probability distribution of x_i
$E_p[]$	expected value of the argument distributed according to $p()$
$E_q[]$	expected value of the argument distributed according to $q()$
$y_{i,p}^j$	the j th sample element drawn from $p(y_i)$
$y_{i,q}^j$	the j th sample element drawn from $q(y_i)$
w_i^j	the weight of the j th sample element according to the y_i output
s_n	the random numbers generated during resampling
$\mathbf{Pe}$	the dataset gained by the discretization of $p_e(x_i)$

$\mathbf{q}$	the dataset gained by the discretization of $q(x_i)$
$p_{e,k}$	an element of $\mathbf{p_e}$
q_k	an element of $\mathbf{q}$
K	the size of $\mathbf{p_e}$ and $\mathbf{q}$
$J()$	Jaccard-index
$g_{i,e}$	the goodness value belonging to the judgment of the eth expert about x_i
$\tilde{g}_{i,e}$	the normalized goodness value of the eth expert judgment about x_i
m_i	the component mass flows of the waste separation system
r_i	the separation efficiencies
$\mathbf{m}^c$	a 14-element vector contains the component mass flows of component c
$\mathbf{A}^c$	the state transition matrix of component c
$\mathbf{I}$	the unit matrix
$\mathbf{B}^c$	the input vector in case of component c
u^c	the input of the system in case of component c
$\mathbf{y}^c$	the outputs of the system in case of component c
$\mathbf{C}^c$	the output vector

Acknowledgments

I am grateful for my supervisor, János Abonyi, who gave me directions in the research, answered my every little question, and helped me with his valuable suggestions to make the essay better and better.

Bibliography

- [1] Walter Stortelder. Parameter estimation in dynamic systems. *Physica D-nonlinear Phenomena - PHYSICA D*, 42, 01 1998.
- [2] Robert T. Clemen and Robert L. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203, Apr 1999.
- [3] Oliver Krol, Nicolas Weiss, Thomas Bernard, and Felix Sawo. Including expert knowledge in finite element models by means of fuzzy based parameter estimation. 2009.
- [4] Muhammad Akram, Ghous Ali, José Carlos R. Alcantud, and Aneesha Riaz. Group decision-making with fermatean fuzzy soft expert knowledge. *Artificial Intelligence Review*, 55(7):5349–5389, Oct 2022.
- [5] George Kuczera and Eric Parent. Monte carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm. *Journal of Hydrology*, 211(1):69–85, 1998.
- [6] Mohammad Mahdi Rajabi and Behzad Ataie-Ashtiani. Efficient fuzzy bayesian inference algorithms for incorporating expert knowledge in parameter estimation. *Journal of Hydrology*, 536:255–272, 2016.
- [7] Simona Pedde, Kasper Kok, Janina Onigkeit, Calum Brown, Ian Holman, and Paula A. Harrison. Bridging uncertainty concepts across narratives and simulations in environmental scenarios. *Regional Environmental Change*, 19(3):655–666, Mar 2019.
- [8] Sajna Parimita Panigrahi, Sarat Kumar Maharana, Thejaraju Rajashekaraiah, Ravichandran Gopalashetty, Mohsen Sharifpur, Mohammad Hossein Ahmadi, C. Ahamed Saleel, and Mohamed Abbas. Flat unglazed transpired solar collector: Performance probability prediction approach using monte carlo simulation technique. *Energies*, 15(23), 2022.

- [9] Chengliang Fan, Yundan Liao, Guang Zhou, Xiaoqing Zhou, and Yunfei Ding. Improving cooling load prediction reliability for hvac system using monte-carlo simulation to deal with uncertainties in input variables. *Energy and Buildings*, 226:110372, 2020.
- [10] Onur Dogu, Andreas Eschenbacher, Robin John Varghese, Maarten Dobbelaere, Dagmar R. D’hooge, Paul H.M. Van Steenberge, and Kevin M. Van Geem. Bayesian tuned kinetic monte carlo modeling of polystyrene pyrolysis: Unraveling the pathways to its monomer, dimers, and trimers formation. *Chemical Engineering Journal*, 455:140708, 2023.
- [11] Edith Calixto-Pérez, Jesús Alarcón-Guerrero, Gabriel Ramos-Fernández, Pedro Américo D. Dias, Ariadna Rangel-Negrín, Monica Améndola-Pimenta, Cristina Domingo, Víctor Arroyo-Rodríguez, Gilberto Pozo-Montuy, Braulio Pinacho-Guendulain, Tania Urquiza-Haas, Patricia Koleff, and Enrique Martínez-Meyer. Integrating expert knowledge and ecological niche models to estimate mexican primates’ distribution. *Primates*, 59(5):451–467, Sep 2018.
- [12] J. Rawlings and D.Q. Mayne. *Model Predictive Control: Theory and Design*. 2009.
- [13] Luciano da F. Costa. Further generalizations of the jaccard index. *ArXiv*, abs/2110.09619, 2021.
- [14] Ágnes Bárkányi, Attila Egedy, Attila Sarkady, Róbert Kurdi, and János Abonyi. Expert-based modular simulator for municipal waste processing technology design. *Sustainability*, 14(24), 2022.