Robust Model-Based Fault Detection Using Monte-Carlo Methods and Highest Density Regions

Felix Mardt[1] and Frank Thielecke[2]

[1,2] *Institute of Aircraft Systems Engineering – Hamburg University of Technology, Hamburg, 21129, Germany*
felix.mardt@tuhh.de
frank.thielecke@tuhh.de

ABSTRACT

One of the major problems of model-based fault detection is to account for model and measurement uncertainties in order to robustly detect occurring faults. This paper presents a method which utilizes Monte Carlo simulations to solve this problem for hybrid nonlinear models. By sampling the a-priori and statistically identified uncertainty distributions, corresponding residual values are obtained. The distributions of these residuals are analysed using highest density regions to obtain information about the probability of receiving the observed measurements given a fault-free model. In addition to the basic method, an extended method utilizing explicit fault models is presented. Both methods are implemented in form of an algorithm and, in order to provide a proof of concept, applied to the model of a cooling system for an unmanned aerial vehicle.

1. INTRODUCTION

With the increasing complexity of technical systems the task of fault detection and isolation (FDI) becomes more and more difficult. While the need for sophisticated safety critical real-time FDI is generally well covered due to government regulations and safety concerns, the maintenance related FDI has historically been less emphasized. This leads to increased maintenance costs during the system's life cycle caused by false alarms and missed detections.

In order to improve the maintenance related FDI, the Institute of Aircraft Systems Engineering at the Hamburg University of Technology is working on SPYDER, a **S**oftware **P**ackage for s**Y**stem **D**iagnosis engine**ER**ing. SPYDER utilizes available knowledge about the behaviour of a system in terms of physical models to design a diagnostic engine. The nonlinear, hybrid dynamic models are converted into convenient, overdetermined, steady state submodels which are employed for fault detection purposes.

The faults are detected by evaluating residuals derived from the submodel's equations. These residuals are ideally equal to zero in the fault-free case and different from zero in the case of a fault. In real world applications, however, these residuals usually differ from zero even in the fault-free case due to e.g. measurement errors and model uncertainties. This paper presents a method which allows the evaluation of whether an observed set of measurements could possibly stem from a fault-free system by utilizing Monte Carlo Methods with identified and a-priori uncertainty distributions.

The paper is structured as follows. Section 2 gives a short overview over previous work in this field of study. Section 3 systematically defines the problem and some basic definitions which are used to present the method in Section 4. After the theoretical introduction of the concept, Section 5 presents the implementation of the method in form of an algorithm. This algorithm is applied in Section 6 to a model of a cooling system for an unmanned aerial vehicle. The section discusses the obtained results and insights gained from the application. The paper closes with a conclusion and an outlook in Section 7.

2. LITERATURE REVIEW

There is an exhaustive literature available on the topic of robust fault detection, especially for model-based FDI. Generally, there are two approaches to handle uncertainties.

The first one, called active approach, relies on the design of residuals which are decoupled from the uncertainties. This is done by designing filters and observers, which decouple the residual from the uncertainties while preserving sensitivity to faults (Chen & Patton, 2012). This approach requires specific model characteristics and structures to be applicable and is far from the general approach envisaged here.

The second, called passive approach, accepts uncertainties in the residuals and handles them after the evaluation by applying thresholds. These thresholds can be static values or, in

more sophisticated methods, adaptive depending on the measurements or system's state. The simplest way to determine thresholds is to record non-faulty residual signals as either absolute values (Yu, Wang, Luo, & Huang, 2010), or as statistical measures (Staroswiecki & Comtet-Varga, 2001; Svärd, Nyberg, Frisk, & Krysander, 2011) and compare the residuals to the non faulty measures. This approach requires historical data and thus is not applicable to new systems or sensors. Additionally, it neglects that there commonly is expert knowledge available about the uncertainties of the system which can be used to base the thresholds on.

The set-membership approach (Ingimundarson, Bravo, Puig, Alamo, & Guerra, 2009) uses a-priori knowledge about the uncertainties of the measurements and parameters as bounded intervals in conjunction with a model of the system to calculate the set of possible residual values (direct test) or possible parameters (inverse test).

The inverse test relies on parameter estimation and system identification techniques. These methods generally require dedicated input signals in various system states to accurately determine the model's parameters. This usually involves a great effort and is infeasible or impossible in a maintenance context.

The direct test propagates uncertainties through the process model or some kind of approximation and thusly determines intervals of possible residuals of a fault-free model. This propagation can be done by e.g. simple (Fagarasan, Ploix, & Gentil, 2004) or complex (Armengol, Vehí, Sainz, Herrero, & Gelso, 2008) interval arithmetic. Other methods rely on numerical optimization techniques to find the residual's extrema. Most of these methods require the model to fulfil certain characteristics like linearity or continuity with respect to the uncertainties.

The direct and inverse tests are concepts also found in a dedicated research field called uncertainty quantification (UQ) (Sullivan, 2015). "UQ is the end-to-end study of the reliability of scientific interferences" (U.S. Deptartment of Energy, 2009) and thus the broader approach to handle uncertainties in interference in fields like economics, meteorology and general risk assessment. In UQ the direct test is often approached in a probabilistic way, specifying the output in a statistical sense rather than strict intervals as in the set-membership approach explained above. One of the simplest yet most powerful tools for UQ are Monte Carlo Simulations (MCS) (Rubinstein & Kroese, 2016). MCS utilizes samples of the input distributions of a model to sample the output distribution. Due to this black-box approach there are no restrictions posed upon the structure or characteristics of the model. This fact in conjunction with the additional information about the probability of output values is the reason MCS is chosen to increase the robustness of model-based FDI.

The concept of using MCS for FDI purposes is not new. E.g. particle filters or Sequential Monte Carlo methods can be used to estimate internal states and parameters of dynamic models to detect faulty states of a system (Li & Kadirkamanathan, 2004). As mentioned in Section 1 the aim of this paper is to detect faults through steady state residuals which means the updating nature of a filter is not needed in this context. In (Wang & Haves, 2014) MCS is used to generate diagnostic results for samples of possible measurements and to a limited extent parameter ranges. The approach diagnoses faults only if a majority of the diagnostic results do so and is therefore prone to missed detections. A similar approach is chosen here with the difference that only the detection problem is solved using a statistical hypothesis test. This should reduce the missed detection rate and allow the usual separation of detection and diagnosis.

3. Problem Formulation

Consider a model M of a physical process P

$$P \sim M = \{e, x, y, \theta, f\}, \tag{1}$$

where

$$e(x, y, \theta, f) = \begin{bmatrix} e_1(x, y, \theta, f) \\ \dots \\ e_{n_e}(x, y, \theta, f) \end{bmatrix} \tag{2}$$

are potentially nonlinear, static equations e, unknown internal states $x \in \mathbb{R}^{n_x}$, known measurements $y \in \mathbb{R}^{n_y}$, model parameters $\theta \in \mathbb{R}^{n_\theta}$ and faults $f \in \mathbb{R}^{n_f}$. Note that the effect of a fault f_i on an equation e_j does not have to be modelled explicitly. A binary information that f_i has an effect on e_j suffices for now.

Now consider a subsystem $M^* : \{e^*, x^*, y^*, \theta^*, f^*\}$ of M where $\mathrm{card}(x^*) < \mathrm{card}(e^*)$ such that there are more equations in M^* than needed to calculate the unknowns x^*. This means there is at least one equation which contains redundant information and can potentially be used to test the subsystem for it's integrity. Those equations are called analytical redundancy relations (ARR). Thus, if $\mathrm{card}(e^*) - \mathrm{card}(x^*) = 1$ and the system of equations $\{e^* \setminus e_i\}$ can be algebraically solved for all x^*, e_i is an ARR. Constructing a residual r, which can be used to test the consistency between a measurement and the model, can be done by e.g. subtracting the left (LHS) from the right-hand side (RHS) of e_i

$$\begin{aligned} r(y^*, \theta^*) &= \mathrm{RHS}(e_i) - \mathrm{LHS}(e_i) \\ &= \begin{cases} = 0 & \text{if } y^* \text{ consistent with } M^* \\ \neq 0 & \text{otherwise} \end{cases} . \end{aligned} \tag{3}$$

Ideally an inconsistency and thus $r \neq 0$ only occurs if one of the faults f^* is present. In real world applications, however, the residual's values are almost always different from zero even in the fault-free case. This is mainly due to the following three effects:

1. Modelling error: neglected effects in the modelling process of M (note the $\sim$ in Eq. (1)).
2. Parameter uncertainties: exactly determining θ is practically impossible and the parameters might change over time.
3. Measurements error: y cannot be measured exactly.

Which means rather than constant values for y and θ a set of possible parameters S_θ and measurements S_y have to be considered. Thus, the residual also equate to a set

$$S_r = \{r(\tilde{y}, \tilde{\theta}) | \tilde{y} \in S_y, \tilde{\theta} \in S_\theta\}. \tag{4}$$

Assuming the modelling errors are negligible compared to the other two effects the test for consistency in Eq. (3) becomes

$$\begin{array}{ll} 0 \in S_r & \text{if } y^* \text{ consistent with } M^* \\ 0 \notin S_r & \text{otherwise} \end{array} . \tag{5}$$

Solving this problem for nonlinear, hybrid steady state residuals is the main objective of this paper.

Note that for fault diagnosis and isolation purposes usually a bank of residuals is used

$$R(y, \theta) = [r_1, ..., r_{n_r}],$$

where each residual r_i is sensitive to a different set of faults. In this paper the problem is explicitly separated into n_r subproblems, one for each residual. The alternative is to analyse the multidimensional problem $O \in R$, where O is the origin of the n_r-dimensional space. This formulation adds the benefit of merging information from each dimension (Adrot & Flaus, 2008). Interpreting the result of a multidimensional analysis is a much more complex task and not the focus of this work.

4. METHODOLOGY

The problem formulated above can be split into two subproblems:

1. calculating the set of possible residual values S_r and
2. determining whether $0 \in S_r$.

If the set S_r can be calculated exactly, the second step becomes trivial. As stated in Section 2 there is no method which is capable of this calculation for arbitrary models which is why a MCS based approach is chosen here. MCS uses samples of the input sets of a model to generate samples of the output set. For the problem described above this means generating a set

$$S_{\mathrm{r,MC}} = \{r(\tilde{y}_1, \tilde{\theta}_1), ..., r(\tilde{y}_{n_s}, \tilde{\theta}_{n_s}) | \tilde{y}_i \in S_y, \tilde{\theta}_i \in S_\theta\}$$

of n_s samples by evaluating r n_s times. Since $S_{\mathrm{r,MC}}$ consists of discrete samples of the otherwise piecewise-continuous S_r, the sub-problem 2 becomes more difficult. This comes with the benefit that the resulting distribution contains information not only about the range of possible values in S_r, but also about the probability of occurrence. This probability is only useful if the sampling of y and θ is based on reliable probability distributions. Determining these distributions poses an additional challenge.

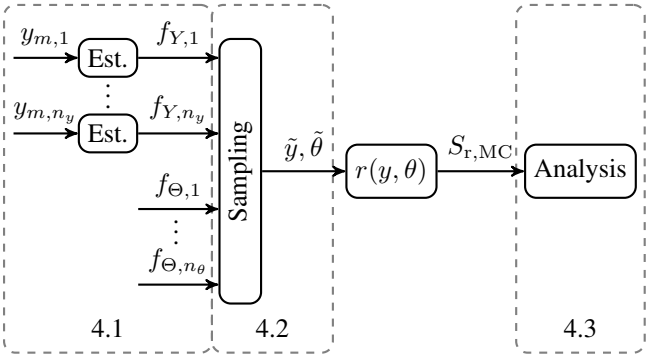

Figure 1. Basic method overview (including the respective sections)

Each of the following subsections deals with a different aspect of the presented method depicted in Figure 1. Section 4.1 discusses the modelling of uncertainty for the measurements and parameters, Section 4.2 covers the sampling method and Section 4.3 describes the method used to analyse the results and solve the second sub-problem stated above. An extension to the presented method in case of available explicit fault models is presented in Section 4.4.

4.1. Modelling Measurement and Parameter Uncertainty

As stated above, the modelling of the parameter and measurement uncertainty in terms of probability distributions is crucial for the usability of the result. Too wide distributions lead to missed detections of faults, while too narrow distributions can lead to false alarms. As depicted in Figure 1, the distributions are modelled in terms of probability density functions (PDFs) $f_{Y,i}$ and $f_{\theta,i}$. PDFs specify the relative probability of each possible value of a continuous random variable e.g. Y such that

$$\Pr[a \leq Y \leq b] = \int_a^b f_Y(y) dy$$

is the probability of a sample y falling into the range $[a, b]$. In the following, the process of determining the PDFs $f_{Y,i}$ and $f_{\Theta,i}$ for the measurements and the parameters respectively is discussed.

Modelling Measurement Uncertainty

Measurement errors, which are the reason for measurement uncertainties, are usually split up into a random and a systematic component. Since the latter part is systematic it can be compensated via calibration of the respective sensor. For the sake of simplicity, it is assumed that all sensors are exactly

calibrated such that only random errors occur. This means the measurement results y_m are randomly distributed around the real value y according to PDFs $[f_{Y,1}, ..., f_{Y,n_y}]$. Modelling $f_{Y,i}$, to obtain a sampling distribution for the MCS, can be done in different ways depending on the number of samples and the knowledge about the sensor's accuracy.

If only a single measurement is taken, no information about the variability of the measurement in form of data is available. In this case, the only way to model the PDF is knowledge about the sensor given by the manufacturer of that sensor or obtained from previous measurements. In most cases, the manufacturer specifies intervals for a sensor's accuracy. If no further information about the form of the PDF – i.e. in terms of a standard deviation $\sigma_{y,i}$ – is given, a uniform distribution

$$\mathbb{U}(y_{\text{min}}, y_{\text{max}})$$

with equal probabilities for every value in the interval can be used.

Since the aim of this method is to detect faults in steady state, usually a couple of measurements from the same state and thus the same real value are available. This means that statistics can be applied to narrow the uncertainty of the measurement.

For $n_m > 1$ measured values $y_{m,i}$, which are randomly distributed around a constant value y_i, the best estimate of this value is the arithmetic mean of the measurements

$$E[Y_i] = \bar{y}_{m,i} = \frac{1}{n_m} \sum_{j=1}^{n_m} y_{m,i,j}. \tag{6}$$

If the standard deviation of the sensor is known, the distribution of this estimate can be modelled as

$$f_{\bar{Y}_i} = \mathcal{N}\left(\bar{y}_{m,i}, \sigma_{\bar{y}_i} = \frac{\sigma_{y_i}}{\sqrt{n_m}}\right).$$

If no a-priori information about the distribution is available, the sample standard deviation of the mean

$$\hat{\sigma}_{\bar{y}_i} = \sqrt{\frac{\sum_{j=1}^{n_m} (y_{m,i,j} - \bar{y}_{m,i})^2}{n_m(n_m - 1)}} \tag{7}$$

can be used in conjunction with either the normal distribution

$$f_{\bar{Y}_i} = \mathcal{N}(\bar{y}_{m,i}, \hat{\sigma}_{\bar{y}_i}) \tag{8}$$

or, to account for small sample sizes, the scaled and shifted t-distribution

$$f_{\bar{Y}_i} = \mathcal{T}(\bar{y}_{m,i}, \hat{\sigma}_{\bar{y}_i}, n_m),$$

such that $(Y - \bar{y}_{m,i})/\hat{\sigma}_{\bar{y}_i}$ follows a standard t-distribution with $n_m - 1$ degrees of freedom. Since these are the approximate distributions of the best estimate of y_i, they can be used to sample possible values of y_i as long as the assumptions above hold. According to the central limit theorem the statistical approach applies for arbitrary noise distributions as long as they are centred around the real value.

Modelling of Parameter Uncertainty

The modelling of parameter uncertainties has to be based on a-priori knowledge about the parameters. Possible sources for this knowledge are:

- physical limitations,
- parameter identification,
- measurements,
- expert knowledge.

Physical limitations is the the most straight forward source. It is based on the fact that for some parameters a certain range of values is physically impossible. For example, the air density is known to be in a specific interval if an interval for the operating temperature of the system is given. On the one hand these intervals are reliable, but on the other they tend to introduce a rather high level of uncertainty. The process of tying a PDF to these intervals is not straight forward and must be chosen carefully by the designing expert. In case of no additional information about the probability of operating ranges, a uniform distribution over the possible interval should be used.

Parameter identification is the process of fitting model parameters using real world data. This is a sound source of knowledge but might be a time consuming task or infeasible due to unavailable data at the development stage. Note that parameters might change over time and once estimated parameters possibly become outdated somewhere in a system's life. Some parameter identification techniques come with error bands or uncertainty models for the estimated parameters. These values can be used directly to model the corresponding PDF.

Measurements are also a sound source of knowledge. If it is possible to measure a parameter, e.g. a geometric length, directly, the same procedures described for the modelling of measurement uncertainties above can be applied. Similar to identified parameters, the measurements can become obsolete and might have to be updated during the system's life.

Expert knowledge is a rather fuzzy source of knowledge but nevertheless valuable. The modelling expert has potentially worked on similar models in the past and gathered knowledge about possible parameter ranges, which can be utilized. Usually it is difficult for a human to comprehend complex distributions such as the normal or beta distribution. For this reason simple distributions like the triangle or uniform should be used to model expert knowledge.

Up until now, each PDF has been regarded as independent of each other. This holds for the sensor uncertainties, since it is assumed that the errors are essentially random and thus independent. This might not be the case for the parameters.

Consider for example the viscosity and density of air. Both of these quantities are dependent on air temperature and thus correlated. These correlations need to be addressed in the uncertainty modelling phase due to their potentially immense impact on the output distribution.

One statistical way to tackle the problem of correlating parameters is the usage of multivariate joint distributions e.g. multivariate Gaussian. Choosing the parameters for these models for physical relations is not straight forward. A more natural way is to model the correlations directly. Most of the correlated parameters are due to physical coupling as in the example above. By modelling this coupling directly and sampling the underlying inputs, the correlation is modelled implicitly. For the viscosity and density of air this means, specifying the range and distribution of possible operating temperatures and calculating the dependent parameters based on the sampled temperature. If the air temperature is measured and thus part of y, it can be used as an input directly. Note that this is essentially an extension of the underlying model and might introduce new parameter and model uncertainties. Similar to the regular process of physical modelling, the level of detail for the application has to be assessed carefully.

4.2. Sampling

Sampling from f_Y and f_Θ can be done using a pseudo random number generator. These generators mimic randomness by calculating deterministic sequences of numbers which pass specified tests for randomness. They are deterministic in a sense that given the same initial input - a so called seed - they produce the exact same sequence of random numbers. These generators are part of every statistical software suite and can be used to sample from arbitrary distributions.

There are different techniques to reduce the variance of the sampling process and consequently the number of samples needed. The most basic one is latin hypercube sampling, which splits the sample space into equally probable intervals and takes the same amount of samples from each of those intervals. This method preserves the shape of the PDF and simultaneously ensures a uniform coverage of the sample space. This results in the need of less samples for the same amount of coverage and thus reduces computational load. This approach is applied here.

4.3. Analysis of MCS Output

As stated in the introduction of this section, the output set not only contains information about the range of possible values, but also about their probability of occurrence. This allows tying a level of confidence to Eq. (5) in order to potentially improve the missed detection rate and compare the consistency of different models. Thus, the detection problem becomes a

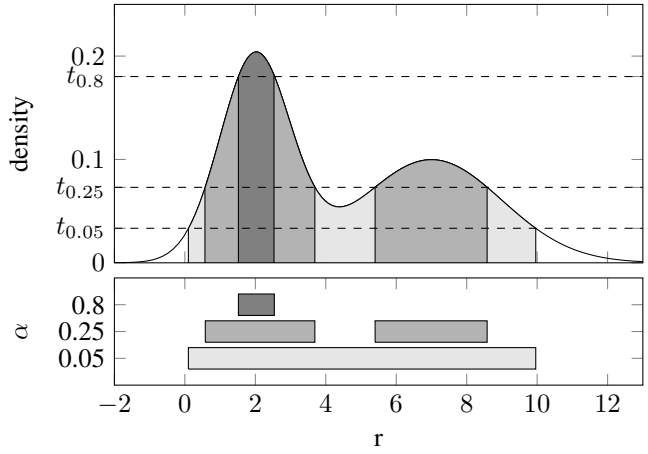

Figure 2. PDF and HDR for a multimodal distribution

statistical hypothesis test, where the null-hypothesis

$$H_0 : y^* \text{ consistent with } M^*$$

is rejected when $r = 0$ is sufficiently unlikely. Since the residuals are continuous, the probability of $r = 0$ is practically zero. This is a common phenomenon in the domain of hypothesis testing and is usually approached by using a probability of observing a result more unlikely than the current one. The calculation of this probability is straight forward when the observed outcome is normally distributed. Since the residuals are potentially nonlinear and discontinuous, their probability distribution is generally not normal and might be multi-modal and discontinuous as well.

To solve this problem, highest density regions (HDRs) are used (Hyndman, 1996). These regions are defined as the subset $S_{\mathrm{r},\alpha}$ of the sample space S_r such that

$$S_{\mathrm{r},\alpha} = \{r | f_R(r) \geq t_\alpha\}, \tag{9}$$

where f_R is the PDF of r and t_α is a constant probability density such that $\Pr(r \in S_{\mathrm{r},\alpha}) \geq 1 - \alpha$. This means $S_{\mathrm{r},\alpha}$ includes the most probable $100(1-\alpha)\,\%$ of values. Figure 2 shows the 20, 75 and $95\,\%$ HDRs for a given distribution. The $t_{0.8}$, $t_{0.25}$ and $t_{0.05}$ are marked as dashed lines and the corresponding HDRs as shaded areas below the PDF in the upper panel. The lower panel depicts the HDRs without overlays. Note that $S_{\mathrm{r},\alpha}$ can include multiple intervals for multi-modal distributions.

Using HDRs the hypothesis test can be stated as

$$H_0 : \begin{array}{ll} \text{accepted if} & 0 \in S_{\mathrm{r},\alpha_{\mathrm{cl}}} \\ \text{rejected if} & \text{else} \end{array} \tag{10}$$

where α_{cl} is chosen according to the required confidence level. Note that this approach is equal to the set membership problem for $\alpha_{\mathrm{cl}} = 0$.

In practice, the PDF f_R of the residual is not known and only the sample set $S_{\mathrm{r,MC}}$ is available. The most common technique to calculate an estimate of the PDF of a population given samples from that population, is the kernel density estimation (KDE). KDE approximates the PDF via a sum of kernel functions at each sample point multiplied by a smoothing factor called bandwidth. The bandwidth has a large impact on the resulting PDF and there is no general correct choice for it. For HDR estimation (Samworth, Wand, & others, 2010) proposed an optimal bandwidth selection algorithm which is applied here. This technique finds a bandwidth such that $(1-\alpha)100\%$ of the samples are inside the HDRs. This means that enough samples have to be generated to accurately capture the real distribution. The process of choosing the needed sample size n_s is described in Section 5.

4.4. Extension using Explicit Fault Models

The analysis method presented above uses arbitrary thresholds to improve the fault sensitivity by rejecting H_0 if $r = 0$ is sufficiently unlikely. This is based on the assumption that every combination of parameters and measurements which is less likely than the specified threshold is due to a fault. This dismissing of every unlikely residual can lead to reduced missed detection rates, but can also lead to higher false alarm rates depending on the fault modes and their behaviour. A solution to this is to include faulty behaviour models into the residuals and compare the probabilities of $r = 0$ of the faulty and the fault-free model.

Including information about faults into the model is done by modelling the effect of f on e in Eq. (2) explicitly such that f are input variables into the model, which are zero in the fault-free case. Thus, the diagnosable subsystem M^* and the residual r also depend on f^* as inputs. Note that the modelling of faults generally requires additional identification and validation of the models and might even be infeasible for complex faults. If the modelling can be done reliably, however, it provides an immense benefit to the model-based fault detection.

When also including the modelled faulty behaviour the null-hypothesis becomes

$$H_0 : y^* \text{ consistens with } M^* \text{ given } f^* = 0$$

and an alternate hypothesis can be formulated

$$H_1 : y^* \text{ consistens with } M^* \text{ given } f^* \geq t_f,$$

where t_f is an n_{f^*}-dimensional value specifying thresholds, above which the fault levels are considered unacceptable. It is assumed that the faults are monotonic in the sense that a higher value of f^* is more severe. Both hypotheses can be tested according to the basic method presented above. For the faulty-model the fault variables are sampled uniformly from a specified interval according to

$$\mathbb{U}(t_f, f^*_{\max}).$$

The resulting sets $S_{\mathrm{r},\mathrm{f}^*=0}$ and $S_{\mathrm{r},\mathrm{f}^*\geq\mathrm{t_f}}$ for the fault-free and faulty case respectively are then used to calculate the HDRs. This leads to the new hypothesis test

$$H_0 : \begin{array}{ll} \text{accepted if} & \exists\alpha_{\mathrm{cl}} : \quad 0 \in S_{r,f^*=0,\alpha_{\mathrm{cl}}} \wedge \\ & \qquad\qquad\ \ 0 \notin S_{r,f^*\geq t_f,\alpha_{\mathrm{cl}}} \\ \text{rejected if} & \text{else} \end{array} .$$

By comparing both confidence levels, a more sound guess can be made in accordance to the actual expected faulty behaviour.

5. IMPLEMENTATION

The previously described methods are concepts for the detection of discrepancies between a model and a series of measurements. The actual implementation of these methods raises additional points like: points in time of execution, necessary number of samples and which calculations have to be performed continuously. This section deals with these points and describes a way of implementing the presented methods.

Since the aim of this method is to detect faults with no immediate safety effect, the execution of the algorithms is not time critical. The algorithms do not even have to run in parallel to the monitored process and can be applied subsequently to recorded measurements of the system. To have an up-to-date information about the system's health status, however, a periodic execution of the algorithms during regular operation is advised. For mobile applications in e.g. planes or cars this could mean executing the algorithms on-board, which is why one of the implementation's aims is to keep the execution time at a required minimum.

The main part of the presented method is to generate samples and use these to calculate the HDRs. The implementation of these two steps is shown as a flow chart in Figure 3. The first step is to detect whether the system is in steady state. This is due to the fact that the considered residuals are based on steady state equations and thus invalid during transients. If the model is in steady state, the measured values can be used to estimate the statistical measures of the measurement uncertainty distributions. As soon as the model leaves a steady state, the collected statistics can be used to detect potential faults of the system. This has to be done only once at the end of every steady state period, since the parameter distributions are considered constant and the measurement distributions become narrower with every new measured value. This leads to a more precise detection result at every new time step and thus only the point in time containing the most information and least uncertainties is analysed. To ensure this only happens once, the $ssEval$-flag – a variable – is initially set to 0 and subsequently to 1 after the last steady state period has

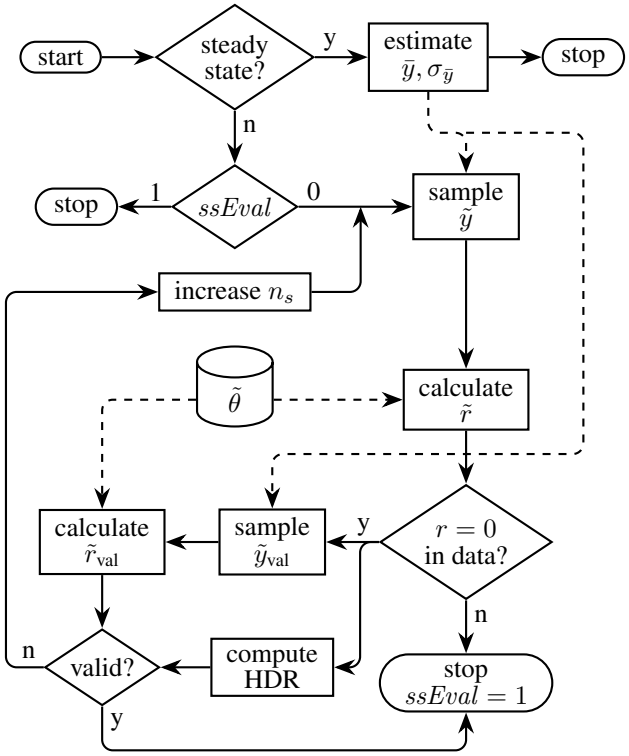

Figure 3. Implementation flowchart

been analysed.

The calculation of the HDRs itself starts the same way as described in Section 4 above. A latin hypercube sampling based on the estimated distribution parameters is used to create a specified number of samples. These samples together with parameter samples are then plugged into the residual equations. Since it is assumed that the distributions of parameters do not change during short term operation, they can be sampled once and kept in a database until changes to those distributions become necessary. This saves on-line computation time. Subsequently, the calculated residual samples $\tilde{r}$ are checked on whether their range contains $r = 0$. This is done by the simple test

$$0 \in [\min(\tilde{r}), \max(\tilde{r})],$$

since this is a necessary condition for the existence of an HDR which contains $r = 0$. If this test fails, the model is considered not consistent with the measurement and the algorithm is stopped.

After the test for a general possibility of $r = 0$, two tasks will be executed in parallel. The first one is the calculation of the HDRs. This is done for different levels $\alpha_{\text{cl},i}$ to get a discrete sampling of the probability space in contrast to just binary information as proposed in Eq. (10). The second parallel task is an additional sampling of values $\tilde{y}_{\text{val}}$ and corresponding residual values $\tilde{r}_{\text{val}}$. These additional samples are used as validation data to test if the HDRs are accurate for data which was not used to generate said HDRs. For this purpose a mean error

$$\frac{1}{n_\alpha} \sum_{i=1}^{n_\alpha} \left| \frac{\text{card}(\{r_{\text{val}} | r_{\text{val}} \in S_{r,\alpha_i}\})}{n_{\text{val}}} - (1 - \alpha_i) \right|$$

is introduced, which measures the average difference between the required and the actual fraction of validation data falling into each interval. By comparing this error to a fixed threshold, a test is introduced whether the calculated intervals are accurate. A failed test indicates that the data set used to calculate the HDRs does not contain enough information about the actual probability distribution of r. To compute more accurate HDRs the sample size n_s is increased and another iteration of the calculation of HDRs is initiated. This iterative procedure ensures that the number of samples is accurately chosen for each case.

The described algorithm can also be applied to the residual including faulty behaviour to get the corresponding HDRs. The only difference is the use of uniformly sampled fault inputs on an expected value range for calculation of HDRs and validation data.

The following and somewhat trivial interpretation of the HDRs is not depicted, since it heavily depends on the chosen method of subsequent fault diagnosis.

6. Method Application

The following section describes the application of the proposed methods. The system and model to which the methods are applied is presented in Section 6.1, Section 6.2 describes the process of uncertainty modelling and the results of the application are presented and discussed in Section 6.3.

6.1. System and Model Description

The model, to which the presented method is applied, is one of an air cooling system for an unmanned aerial vehicle (UAV) with vertical take-off and landing capabilities. The UAV itself is based on an existing Vehicle and the modelled aerodynamics and flight mechanics are validated using the real aircraft. The systems inside the UAV, namely the fuel cell and the cooling system, are not part of the real vehicle. They were chosen, designed and implemented virtually to test and develop different health monitoring algorithms in an ongoing national research project called Real Time Analysis Predictive Health Monitoring (RTAPHM).

The modelled cooling system uses ambient air to cool the fuel cell (FC) and motor power electronics (PE) on board of the UAV. A basic schematic of the main components is depicted in Figure 4.

The air enters the system through a ram air intake, which uses

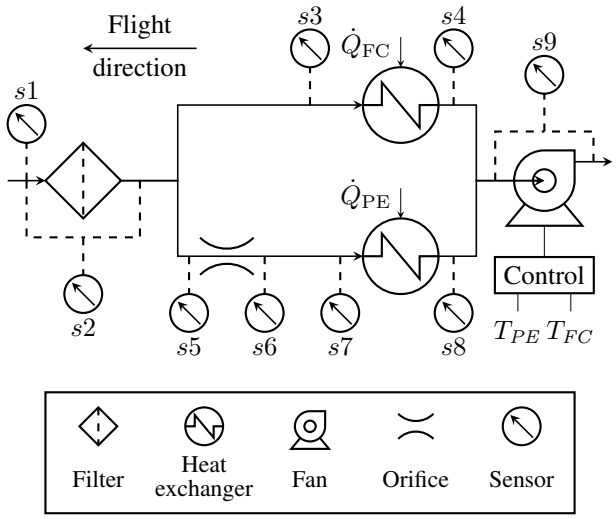

Figure 4. Cooling system schematic

dynamic pressure in flight to increase the static pressure inside the duct. Driven by the resulting difference in pressure the air flows through a filter, which protects downstream components from unwanted solid particles, before diverging into two parallel sections. The first section cools the FC through a heat exchanger. The second section also incorporates a similar but smaller heat exchanger for the PE. Due to the fact that the heat flow emitted by the PE is much less than the flow emitted by the FC the PE needs less cooling air. To set the flow split accordingly an orifice is part of the PE section. After cooling both the PE and FC the air flows through a fan which pulls air through the system when there is no dynamic pressure provided by a forward movement of the UAV – i.e. in ground or hover/take-off operation. The schematic only shows the main components of the system and not the shape and or length of the connecting ducts.

The fan control is based on the component temperatures T_{FC} and T_{PE}. The purpose of all other sensors si depicted in Figure 4 is the detection and isolation of occurring faults in the components.

The model of the described system is implemented in Matlab Simscape based on dynamic pneumatic and thermal components. Some of the equations are highly nonlinear and contain multiple modes of operations. The sensor models include a relative and absolute accuracy as well as white noise, which is added onto the real measurement, to incorporate measurement uncertainties. In addition to the nominal behaviour each of the depicted components incorporates one fault mode with maintenance relevance. The implemented fault modes are:

- clogging of the filter,
- clogging of the orifice,
- fouling of both heat exchangers,
- reduced fan efficiency.

Table 1. Residual overview

Comp.	r_i	n_θ	y				
Filter	r_{filter}	2	$\dot{m}_{s1}$,	dp_{s2}			
Orifice	r_{orifice}	2	T_{s1},	$\dot{m}_{s1}$,	p_{s5},	p_{s6}	
Fan	r_{fan}	4	$\dot{m}_{s1}$,	dp_{s9},	Ctl_{fan},	$P_{\mathrm{el,fan}}$	
FC	r_{FE}	9	T_{s1},	$\dot{m}_{s3}$,	p_{s3},	p_{s4}	
PE	r_{PE}	9	T_{s1},	$\dot{m}_{s1}$,	$\dot{m}_{s3}$,	p_{s7},	p_{s8}

Each of these faults is modelled as a continuous change of behaviour with a normalized input controlling its severity.

Fault Detection

The detection of faults of the above described system is done by using one residual for each of the faulty components. These residuals were generated based on the model equations in steady state. Table 1 lists all of the residuals including their name, the specific component, the number of uncertain parameters as well as the input signals. The input signals are: the pressures p, the differential pressures dp, temperatures T and mass flows $\dot{m}$ with their respective measurement point given as an index. Ctl_{fan} and $P_{\mathrm{el,fan}}$ are the control signal and the electrical power drawn by the fan respectively.

6.2. Uncertainty Modelling

The uncertainty distributions for the parameters have been chosen according to Section 4.1. Since the system only exists virtually, a parameter identification was not possible. Thus, the boundaries of the parameters are based on physical limitations and probable measurement uncertainties where applicable. The remaining distributions were estimated heuristically and the boundaries of the parameters converted to distributions by using uniform distributions.

The sensor uncertainties are assumed to be completely unknown and estimated online according to Eq. 6 and 7. This is feasible due to the relatively high sample rate which is also the reason for using the normal distribution in Eq. 8 instead of the T-distribution to model the measurement uncertainties.

The faulty residual distributions are generated by sampling the fault inputs uniformly from the interval $[0.25, 1]$.

6.3. Simulation results

In order to test the methods, a 15 minute flight mission is simulated. The height (H) profile for this scenario is shown in the upper panel of Figure 5 and comprises a preflight, flight, idle and a battery recharge phase. To test the fault detection capabilities, two faults in terms of a reduced fan efficiency and a partially clogged filter ($f_{\mathrm{fan}} = f_{\mathrm{filter}} = 0.3$) are injected. All other components are simulated fault-freely.

The residuals during each of the steady state periods for three components are depicted in the subsequent panels of Figure

5. There are a total of five steady state periods:

- one in the preflight phase,
- two in the flight phase, where the second and shorter one occurs during the vertical descent,
- one in the idle phase,
- two in the battery recharge phase, where the interruption is due to a fan ramp up at $655\,\mathrm{s}$.

During flight the fan residual is not evaluated, since the fan is switched off. Each residual has been evaluated with different y and θ values. The dashed lines show the residuals evaluated with the ideal sensor and parameter values without any uncertainty. Looking at these values it is obvious that the fan and the filter are degraded since the signal is clearly different from zero while the fuel cell is consistent with the nominal model. In practice, however, measurement noise and parameter uncertainties are present. These effects are shown by the gray lines. Especially the measurement noise impedes the interpretation of the residuals. But even when eliminating the measurement noise, by using the average of each sensor signal for this specific steady state, shown by the solid black lines, all three residual values are different from zero. The parameters θ_{est} were taken as the most probable value of each parameter's distribution.

To test whether the observed deviation from zero in the residuals could be due to parameter and measurement uncertainties, the proposed methods are applied. In Figure 6 the detailed analyses of the second steady state period at the end of the battery recharge phase is depicted. Each panel shows the estimated probability densities for one of the residuals for the fault-free and faulty case. The densities were calculated using the bandwidth algorithms by Samworth and Wand. The HDRs which include $r = 0$ are given in terms of an interval for the corresponding α value in the title of each panel. If $r = 0$ is not part of the data, only the area between the minimum and maximum of r is shaded in the specific color. In this case, $\alpha = 1$ is given, since no HDRs were computed.

The uppermost panel in Figure 6 shows the possible residual values for the fan. It can be seen that with the given parameter range a value of $r = 0$ is impossible and none of the residual samples came close to zero. Thus, the fault can be detected using only information about the nominal behaviour. When also using the faulty model, this assumption is confirmed since $r = 0$ lies in between the most probable 85 and $90\,\%$ of all data. The low probability of observing a, in this case, correct result is remarkable and needs to be explained. The reason for this is the same reason for why all of the shown faulty residual distributions in Figure 6 are wide spread. The faults are continuous inputs into these models and are assumed to be uniformly distributed according to $\mathbb{U}(t_f, f_{\text{max}})$. If these intervals cover a wide range of operational modes, the residuals are widespread as seen here. This has two downsides. The first one is slower convergence of

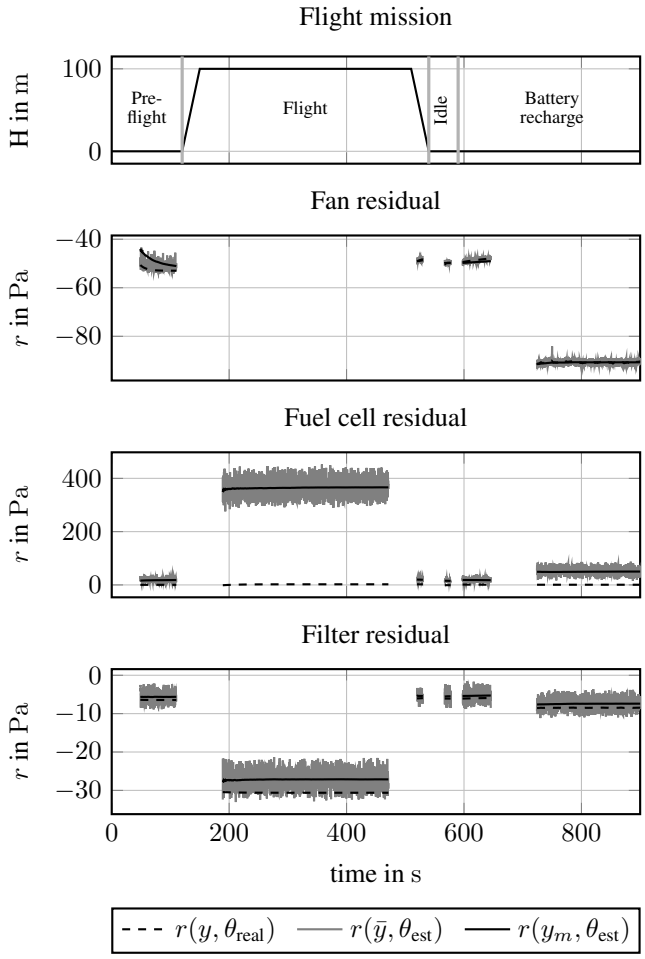

Figure 5. Flight mission height profile and residuals for the fan, fuel cell and filter for a basic flight mission. The legend shown only applies to the last three panels.

the HDR algorithm since small deviations in samples have large impact on the intervals. The second one is that the relative probabilities get lower compared to the nominal model or one with a fixed fault parameters. This is the effect observed here and can lead to misleading results when comparing the probabilities of the nominal and the faulty residuals. This effect can be mitigated by splitting the fault model into multiple smaller intervals of fault values. When splitting the fault sample interval for the fan residual from the initial $[0.25, 1]$ into $[0.25, 0.625]$ and $[0.625, 1]$, the result becomes $0.25 < \alpha_{r_f} \leq 0.3$ for the first interval, while the latter does not contain $r = 0$. As expected, further splitting increases the possibility of $r = 0$ for the interval in which the real value lies even more. Consequently, it can be beneficial to divide the faulty residuals when facing continuous fault inputs for the cost of more computing effort. This needs to be considered when using the extended method.

The second panel in Figure 6 shows the residuals for the filter. The overall result is similar to the ones obtained for the

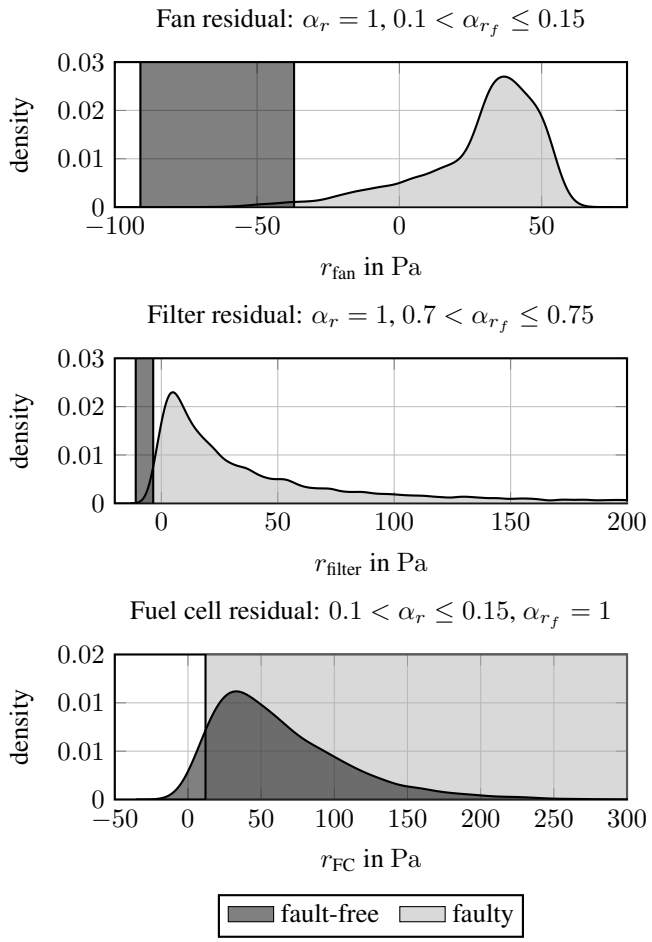

Figure 6. Probability densities of the faulty and fault-free residuals for the steady state interval ending at 900 s.

fan. The probability of consistency between the model and the observed measurements is basically zero and the faulty model matches the data well.

For the fuel cell residuals depicted in the third and last panel the opposite is the case. Only the nominal model is consistent with the measured values. However, the probability of observing these values is relatively small with only 85 to 90 % of the most probable residual values containing $r = 0$. This is a case where the extended method is useful since it provides the additional information that the observations are not consistent with a faulty model. Thus, even with a small probability H_0 can be accepted.

Receiving improbable results for the nominal model without any information about the faulty behaviour of the system should lead to further investigations of the uncertainty assumptions. In this case, the steady state interval is relatively long and thus the measurement uncertainties only have a minor impact on the resulting distribution. When revisiting the parameter distributions and e.g. halving their width, the interval in which α_r for the fuel cell residual lies is more than doubled to $[0.25, 0.3]$. This shows the immense impact of the a-priori distributions on the result.

Table 2. Results as lower and upper bounds for α_r and α_{r_f} at the end of each steady state phase

Time in s		110	471	530	578	646	900
Fan	α_r	1	n/a	1	1	1	1
	α_{r_f}	0.1 0.15	n/a	0.1 0.15	0.15 0.20	0.1 0.15	0.1 0.15
Filter	α_r	1	1	1	1	1	1
	α_{r_f}	0.7 0.75	0.7 0.75	0.7 0.75	0.7 0.75	0.7 0.75	0.7 0.75
FC	α_r	0.05 0.1	0.05 0.1	0.0 0.05	0.1 0.15	0.0 0.05	0.1 0.15
	α_{r_f}	1	1	1	1	1	1

The analysis results for all steady state phases are listed in Table 2. While the results for the other points in time are mostly consistent with the ones shown above, there are minor differences for the residual of the fuel cell. After ruling out convergence issues as well as the sensor uncertainties, it was found that this is due to the fact that the same parameter uncertainties can act differently on the residual distribution depending on the point of operation. Therefore, several operating points should always be considered before drawing conclusions.

As stated above, the measurement uncertainties have only a minor impact on the shown residual distributions. This holds even for the short steady state periods where less measurements are available. Thus, when considering refining the distributions the parameters should always be the starting point.

Convergence and Computation Time

Analysing the faulty and nominal residuals for all three shown residuals in the six steady state phases takes about 45 s on a desktop PC[1]. This includes an average sample size of $4.2e4$ for each MCS. Since this is a relatively short time an application on on-board hardware seems possible. For residuals including more sensor signals and equations this time will rise and the point needs to be revisited.

The creation of the parameter database takes less than half a second and the database itself takes up 160 MB of memory. This is due to the fact that the parameter samples are generated beforehand for each sample size that might occur. This means, permanently storing $1e6$ parameter values even though only $4.2e4$ are used on average. Considering the discrepancy between computing time and memory usage, it might be beneficial to create the parameter samples as

[1] Intel i5 at 4.10 GHz

needed, depending on the availability of memory and computing power.

During the analyses no convergence issues were encountered. The rate at which the solutions converged was slow and scaled roughly with $\sqrt{n_s}$ as reported in literature. It is advised to choose an initial sampling size which is large (about 1000) enough to avoid triggering one of the stopping criteria by chance.

7. Conclusion

This paper presents a fault detection method based on the Monte Carlo simulation of analytical redundancy relations. The inputs to the simulations are a-priori parameter and statistical measurement distributions of the respective uncertainties. This approach allows the handling of nonlinear, hybrid models and provides probability distributions for the residuals. The resulting distributions allow for a formulation of the fault detection problem in terms of hypothesis testing. The analysis of the non-normal distributions is done by computing highest density regions and examining the relative probability of a fault-free model. In an extended version of the method explicit fault models are used to improve the detection capabilities. Both methods are implemented in terms of an algorithm which computes uncertainty distributions from measurements and applies the described methods. The presented algorithm is applied to a model of a cooling system of an unmanned aerial vehicle. The application shows a basic proof of concept, the impact of the input distributions as well as an improvement to the extended method.

Since this paper did show a proof of concept by applying the methods to a model, the focus in future studies should be the application to real world systems in order to further analyse critical issues such as correlating input distributions and accuracy of the model itself.

Acknowledgment

This work was funded by the German Federal Ministry of Economic Affairs and Energy (BMWi) within the RTAPHM project (contract code: 20X1736M) in the national LuFo V-3 program. Their support is greatly appreciated.

References

Adrot, O., & Flaus, J.-M. (2008). Fault detection based on uncertain models with bounded parameters and bounded parameter variations. *IFAC Proceedings Volumes*, *41*(2), 7338–7343. (Publisher: Elsevier)

Armengol, J., Vehí, J., Sainz, M. n., Herrero, P., & Gelso, E. R. (2008). Squaltrack: A tool for robust fault detection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *39*(2), 475–488. (Publisher: IEEE)

Chen, J., & Patton, R. J. (2012). *Robust model-based fault diagnosis for dynamic systems* (Vol. 3). Springer Science & Business Media.

Fagarasan, I., Ploix, S., & Gentil, S. (2004). Causal fault detection and isolation based on a set-membership approach. *Automatica*, *40*(12), 2099–2110.

Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician*, *50*(2), 120–126. (Publisher: Taylor & Francis Group)

Ingimundarson, A., Bravo, J. M., Puig, V., Alamo, T., & Guerra, P. (2009). Robust fault detection using zonotope-based set-membership consistency test. *International journal of adaptive control and signal processing*, *23*(4), 311–330. (Publisher: Wiley Online Library)

Li, P., & Kadirkamanathan, V. (2004). Fault detection and isolation in non-linear stochastic systems—a combined adaptive Monte Carlo filtering and likelihood ratio approach. *International Journal of Control*, *77*(12), 1101–1114. (Publisher: Taylor & Francis)

Rubinstein, R. Y., & Kroese, D. P. (2016). *Simulation and the Monte Carlo method* (Vol. 10). John Wiley & Sons.

Samworth, R., Wand, M., & others. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. *The Annals of Statistics*, *38*(3), 1767–1792. (Publisher: Institute of Mathematical Statistics)

Staroswiecki, M., & Comtet-Varga, G. (2001). Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems. *Automatica*, *37*(5), 687–699. (Publisher: Elsevier)

Sullivan, T. J. (2015). *Introduction to uncertainty quantification* (Vol. 63). Springer.

Svärd, C., Nyberg, M., Frisk, E., & Krysander, M. (2011). A Data-Driven and Probabilistic Approach to Residual Evaluation for Fault Diagnosis. *Proceedings of the IEEE Conference on Decision and Control*, 95–102. doi: 10.1109/CDC.2011.6160714

U.S. Deptartment of Energy. (2009, October). *Scientific Grand Challenges for National Security* (Tech. Rep.).

Wang, L., & Haves, P. (2014). Monte Carlo analysis of the effect of uncertainties on model-based HVAC fault detection and diagnostics. *HVAC&R Research*, *20*(6), 616–627. (Publisher: Taylor & Francis)

Yu, M., Wang, D., Luo, M., & Huang, L. (2010). Prognosis of hybrid systems with multiple incipient faults: Augmented global analytical redundancy relations approach. *IEEE Transactions on systems, man, and cybernetics-part A: systems and humans*, *41*(3), 540–551. (Publisher: IEEE)