Chapter 3

Deterministic Sampling for Quantification of Modeling Uncertainty of Signals

Jan Peter Hessling

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/52193

1. Introduction

Statistical signal processing [1] traditionally focuses on extraction of information from noisy measurements. Typically, parameters or states are estimated by various filtering operations. Here, the quality of signal processing operations will be assessed by evaluating the statistical uncertainty of the result [2]. The processing could for instance simulate, correct, modulate, evaluate, or control the response of a physical system. Depending on the addressed task and the system, this can often be formulated in terms of a differential or difference *signal processing model equation* in time, with uncertain parameters and driven by an exciting input signal corrupted by noise. The quantity of primary interest may not be the output signal but can be extracted from it. If this uncertain dynamic model is linear-in-response it can be translated into a linear digital filter for highly efficient and standardized evaluation [3]. A statistical model of the parameters describing to which degree the dynamic model is known and accurate will be assumed given, instead of being the target of investigation as in system identification [4]. *Model uncertainty* (of parameters) is then *propagated* to *model-ing uncertainty* (of the result). The two are to be clearly distinguished – the former relate to the input while the latter relate to the output of the model.

Quantification of uncertainty of complex computations is an emerging topic, driven by the general need for quality assessment and rapid development of modern computers. Applications include e.g. various mechanical and electrical applications [5-7] using uncertain differential equations, and statistical signal processing. The so-called brute force Monte Carlo method [8-9] is the indisputable reference method to propagate model uncertainty. Its main disadvantage is its slow convergence, or requirement of using many samples of the model (large ensembles). Thus, it cannot be used for demanding complex models. The ensemble size is a key aspect which motivates deterministic sampling. Small ensembles are found by substituting the random generator with a customized deterministic sampling rule. Since any computerized random generator produces a pseudo-random rather than a truly random sequence, this is equivalent of modifying the random generator to be *accurate* for *small* ensembles of *definite* size, rather than being *asymptotically exact* (infinite ensembles). Correctness of very large ensembles is of theoretical but hardly practical interest for complex models, if the convergence to the asymptotic result is very slow.

2. Modeling uncertainty of signals

2.1. Problem definition

Suppose the (output) signal $y(x,\ t) \in \mathbb{R}$ of interest is generated from the (input) signal $x(t) \in \mathbb{R}$ passing through a dynamic system H, with parameters $a_k \in \mathbb{R}$, $b_k \in \mathbb{R}$,

$$\left[\sum_{k=0}^{u} a_k D^k\right] y = \left[\sum_{k=0}^{v} b_k D^k\right] x, \quad a_0 = 1. \tag{1}$$

The model is given in $n = u + v + 1$ uncertain parameters, which can be arranged in a column vector $q = \begin{pmatrix} b_0 & \cdots & b_v & a_1 & \cdots & a_u \end{pmatrix}^T$. For systems continuous-in-time (CT), $D = \partial_t$ is the differential operator in time while for systems discrete-in-time (DT), $D = \Delta^{-1}$ is the negative unit displacement operator, $\Delta^{-1} x_k = x_{k-1}$. There are several approximate methods to sample CT systems to DT systems, see [3] and references therein. The discretization techniques are beyond the scope of this presentation and DT systems will be assumed. If $u \geq 1$, there is feedback in the system which results in an impulse response $h(q,\ t)$ of infinite duration. For finite accuracy however, the duration is finite. The system is linear-in-response, $y(x = \alpha x_1 + \beta x_2,\ t) = \alpha y(x_1,\ t) + \beta y(x_2,\ t)$. Most importantly, the system is non-linear-in-parameters if $u \geq 1$. This is the typical situation addressed here.

Systems of the form in Eq. 1 may be directly realized as digital filters,$y(q,\ x,\ t) = h(q) * x(t)$, where $*$ denotes the filtering operation. The coefficients b_k and a_k are the numerator and denominator coefficients of the filter with impulse response $h(q)$, respectively. Its z-transform $H(q,\ z)$ is obtained with the substitution $\Delta \to z$. The parameterization can be changed to for instance gain$K$, poles p_k and zeros z_k, or poles p_k and residues r_k,

$$Y(z) = H(z)X(z): \quad H(q,z) = \frac{\sum_{k=0}^{v} b_k z^{-k}}{\sum_{k=0}^{u} a_k z^{-k}} = K \frac{\prod_k (z - z_k)/(1 - z_k)}{\prod_k (z - p_k)/(1 - p_k)} = \sum_k \frac{r_k}{(z - p_k)} \tag{2}$$

The parameterization should be carefully chosen as it affects the convergence rate of Taylor expansions (section 3.1) as well as the physical interpretation. The parameters and their statistics are preferably extracted from measurements using system identification techniques [4]. Note that complex-valued poles and zeros are conjugated in pairs [10].

The problem to be addressed is the statistical evaluation of any function $g(y(t)=h(q,\ t)*x(t))$, given statistical models of q and x. It will here consist of evaluating its time-dependent mean $\langle g(y)\rangle$ and standard deviation $\sqrt{\langle [g(y)-\langle g(y)\rangle]^2\rangle}$. Without loss of generality, the analysis will be made for $g(y)=y$. Digital filtering will be utilized for evaluating samples of the model, i.e. filtering with definite sets of q and signals x.

2.2. Nomenclature

Statistical expectations of any signal, model or function $g(q)$ over finite discrete E as well as continuous ensembles or probability distributions (no subscript) are defined as,

$$\begin{aligned} \langle g\rangle_E &= \frac{1}{m}\sum_{k=1}^{m} g\left(\hat{q}^{(k)}\right), \\ \langle g\rangle &= \int_Q g(q) f_q(q) dq. \end{aligned} \tag{3}$$

Samples of q are labeled $\hat{q}$, with their components organized in columns. Sample indices will be given as superscripts in parenthesis, eg. $\hat{q}^{(k)}$ is a column vector denoting the $k-$th sample of parameter q. Variations from the mean are written as $\delta q_{(E)} \equiv q-\langle q\rangle_{(E)}$.

Only uniform (UNI) and normal distributions (NRM) will be utilized. Either the mean and standard deviation, or the interval in brackets will be given in parenthesis, e.g. $q\sim\mathrm{UNI}\left(0.5, 1/2\sqrt{3}\right)=\mathrm{UNI}([0,\ 1])$. Statistical moments $M_i^{(k)}=\sqrt[k]{\langle(\delta q_i)^k\rangle}$ carry the information contained in the marginalized probability density functions (pdf) $f_i(\delta q_i)\equiv\int_Q f_q(\delta q) d q_1 \cdots d q_{i-1} d q_{i+1} \cdots d q_n$, where Q denotes the sample space. While $M_i^{(2)}$ describes the width of $f_i(\delta q_i)$, $M_i^{(3)}$ is related but different to its *skewness* [11]. Further, the shape is reflected in $M_i^{(4)}$, similarly to the *curtosis* [11]. Since UNI(0,1) and NRM(0,1) are normalized and symmetric $f_i(\delta q_i)=f_i(-\delta q_i)$, $M_i^{(2)}=1$ and $M_i^{(3)}=0$. Their differences are first reflected in their fourth moment, $M_i^{(4)}=1/4\sqrt{5},\ 1/2\sqrt{3}\approx(0.11,\ 0.29)$ for UNI(0,1) and NRM(0,1), respectively. The maximum variation of the parameter q_i is expressed by the range $M_i^{(\infty)}\equiv\lim_{k\to\infty}|M_i^{(k)}|=\max(|\delta q_i|)$. Dependencies are expressed in mixed moments $\left\langle(\delta q_{i1})^{k_1}(\delta q_{i2})^{k_2}\cdots\right\rangle_{(E)}$. The discussion will be limited to correlations described by the covariance matrix $\mathrm{cov}(q)=\langle\delta q\delta q^T\rangle$, where the vector multiplication is an outer product.

Matrix size will be indicated with subscripts, e.g. $V_{n\times m}$ is a matrix of n rows and m columns with elements V_{jk}, $j=1, \ldots n$ and $k=1, \ldots m$. The identity matrix will be denoted$I$, while matrices with equal elements (i) will have their size attached, $(i_{n\times n})_{jk}=i$. For a matrix (vector) D, diag(D) is a vector (diagonal matrix) with components (diagonal elements) equal to the diagonal elements (components) of D. The trace of a matrix is denoted Tr.

A method will be stated intrusive if manipulations of the model are required. For the targeted highly complex models, it will be assumed that the computational cost for their evaluation dominates all other calculations. The efficiency ρ of any method will accordingly be defined by the least required number of evaluations of the original model.

2.3. Fundamentals of non-linear propagation of uncertainty

Linearity in parameters (LP) is to be distinguished from linearity in response (LR),

$$\begin{array}{llll} LR: & y(q,a_1x_1+a_2x_2,t) & = & a_1y(q,x_1,t)+a_2y(q,x_2,t), \quad \forall x_1,x_2 \\ LP: & y(q_1+q_2,x,t) & = & y(q_1,x,t)+C(x,t)^T(q_2-q_1) \quad \forall q_1,q_2 \end{array}, \tag{4}$$

for some vector $C_{n\times 1}$. Different concepts of linearity are used, $y(q_1+q_2, x, t)\neq y(q_1, x, t)+y(q_2, x, t)$ for LP models. Strictly speaking, LP denotes models that are affine, i.e. written as linear combinations of their parameters. Most constructed systems are designed to be as close to LR as possible while most models are not LP. There is hence no contradiction in non-linear (LP) propagation of uncertainty with linear (LR) digital filters, as here.

For non-linear propagation of uncertainty, the asymmetry of the resulting pdf is central. It can be expressed as a lack of commutation of non-linear propagation and statistical evaluation of a center value ($\cdot_C$), as measured with the *scent* [12],

$$\zeta \equiv y_C(q)-y(q_C). \tag{5}$$

The method for evaluating the center is left unspecified, as there are several alternatives. The most common choice is to use the mean, $\cdot_C=\langle\cdot\rangle$. The lowest order approximation of the scent can then be obtained by calculating the expectation of a Taylor expansion (section 3.1), $\zeta=\mathrm{Tr}[\mathrm{cov}(q)\cdot H(y)]/2$, where $H(y)_{jk}=\partial^2 y/\partial q_j\partial q_k$ is the Hessian matrix signal of y, evaluated at $\langle q\rangle$. The scent is related to the skewness $\gamma=\langle\delta y^3\rangle/\langle\delta y^2\rangle^{3/2}$. The *additional* asymmetry caused by the non-linearity of the model is measured with the scent but differently. The scent addresses how parametric uncertainties are propagated and not how the result is distributed, e.g. $\zeta=0$ for all LP models for which γ may attain any value. A finite scent thus implies the model is not LP, but not the reverse. The scent should not be confused with bias. Bias is a property of an estimator, while scent is a property of a model. For every model, such as the

REF (section 6.1), many different estimators of $y_C(q)$ can be used, e.g. the different ensembles in section 5.6, see result in Fig. 5 (left). Consequently, an unbiased estimator of $y_C(q)$ correctly accounts for rather than ignores its finite scent, or deviation from $y(q_C)$.

The scent is important since y_C and not $y(q_c)$ is the main result utilized in applications. The corresponding difference [13] in the standard deviation $M_y^{(2)}$ from its linearized approximation $\sqrt{\nabla y^T \mathrm{cov}(q) \nabla y}$, with $(\nabla y)_{jk} = \partial_j y(t_k)$, affects the confidence in the result. Its accuracy is usually less critical. An accurate evaluation of the scent is perhaps the strongest feature of the unscented Kalman filter, which provides the foundation for the presented approach as well as the origin of the term ‘scent’.

3. Conventional methods

A brief resume of the most traditional related methods of uncertainty propagation, applicable to signal processing models, is here given together with their pros and cons. Advanced intrusive methods like e.g. polynomial chaos expansions [14-15] not directly related to the proposed method are omitted.

3.1. Taylor expansions

The indisputable default methods of uncertainty propagation are based on Taylor expansions. These methods are intrusive if the differentiations are made analytically. Convergent series require regular differentiable models and numerical or analytical complexity make them error prone. Their applicability is therefore limited for complex models.

The transfer function $H(q, z)$ of the digital filter can be expanded in a Taylor series,

$$\begin{aligned}\delta H(q,z) = H(q,z) - H(\langle q \rangle, z) = \sum_{k=1}^{+\infty} \frac{1}{k!}\left(\delta q^T \nabla_q\right)^k H(q,z) = \delta q^T \nabla_q H + \frac{1}{2}\sum_{k,l}^{n} \delta q_k \delta q_l \frac{\partial^2 H}{\partial q_l \partial q_k} + \ldots \\ = \delta q^T E^{(1)}(\langle q \rangle, z) + \mathrm{Tr}\left\{\left[\delta q \delta q^T\right] \cdot E^{(2)}(\langle q \rangle, z)\right\} + \ldots\end{aligned} \tag{6}$$

This defines n sensitivity systems (column vector) $E_k^{(1)}(\langle q \rangle, z)$, $n(n+1)/2$ unique quadratic variation systems (matrix) $E^{(2)}(\langle q \rangle, z)$, and so on. These variation systems differ (intrusive) from $H(q, z)$ but may nevertheless be realized as digital filters [3,7,10], just as $H(q, z)$. The corresponding variation of $y(q, x, t) = h(q, t) * x(t)$ is given by,

$$\delta y(q,x,t) = \delta q^T \left[e^{(1)}(\langle q \rangle, t) * x(t) \right] + \frac{1}{2} \mathrm{Tr}\left\{ \left[\delta q \delta q^T \right] \cdot \left[e^{(2)}(\langle q \rangle, t) * x(t) \right] \right\} + \ldots, \tag{7}$$

where $e^{(k)}(\langle q\rangle, t)$ are the impulse responses of the systems $E^{(k)}(\langle q\rangle, z)$. Utilizing digital filters with impulse responses $e^{(k)}(\langle q\rangle, t)$, the differentiations are conveniently done *once*, and not repeatedly for every signal $x(t)$. The linearity in parameters of the model can easily be studied for many different input signals $x(t)$, by evaluating $e^{(k)}(\langle q\rangle, t) * x(t)$. Due to the large number of variation systems, higher order perturbation analyses rapidly become intractable though. The established method is limited to linearization (LIN) [16] $\left(e^{(1)}\right)$. It will always incorrectly yield vanishing scent, $\zeta=0$. A first order estimate of ζ is instead given by the expectation of second term in Eq. 7, $\zeta \approx \mathrm{Tr}[\mathrm{cov}(q) H(\langle q\rangle, t)]/2$, where the matrix of Hessian signals $H(\langle q\rangle, t)=e^{(2)}(\langle q\rangle, t) * x(t)$ is obtained with repeated digital filtering.

3.2. Brute force Monte Carlo

Monte Carlo (MC) methods [8-9], or *random sampling* of uncertain models was originally introduced and phrased 'statistical sampling' by Enrico Fermi already in the 1930's [17]. The MC methods *realize* uncertain signal processing models in finite *ensembles*. Every ensemble consists of a possible set of well-defined model systems, all (usually) having the same structure but slightly different parameter values. In the original so-called brute force Monte Carlo method, each set of parameters is assigned to the output of random generators with appropriate statistics. The convergence to the assigned statistics is very slow [5] but it is asymptotically exact and the required number of samples is essentially independent of the number of parameters. Hence it does not suffer from the curse-of-dimensionality of many other methods. The outstanding simplicity in application is likely the cause of its popularity, just as the slow convergence or low efficiency is the main reason for its failures.

In MC, arbitrary distributions and dependencies are usually obtained by means of transformations of samples of elementary distributions. Independent samples $\hat{q}^{(k)}$ of any probability density function (pdf) $\phi(x)$ can be constructed with the inverse transform method [9]. It consists of a calculation of the inverse of its cumulative distribution function (cdf) $\Phi(y)$ and generation of a uniformly distributed random sequence $\hat{z}^{(k)}$,

$$\hat{q}^{(k)} = \Phi^{-1}\left(\hat{z}^{(k)}\right), \quad \Phi(y) = \int_{-\infty}^{y} \phi(x)dx, \quad z \sim \mathrm{UNI}(0,1), \quad k = 1,2,\ldots m. \tag{8}$$

Covariance may be included with an appropriate transformation of samples of *canonical* parameters $\tilde{q}$: $q=U^T S\tilde{q}$ with $\mathrm{cov}(\tilde{q})=I$,

$$\mathrm{cov}(q) = \left\langle \delta q \delta q^T \right\rangle = \left\langle U^T S\delta\tilde{q}\left(U^T S\delta\tilde{q}\right)^T \right\rangle = U^T S\left\langle \delta\tilde{q}\delta\tilde{q}^T \right\rangle SU = U^T S^2 U, \quad \begin{cases} U^T U = UU^T = I \\ S_{jk}^2 = 0,\ j \neq k \end{cases}, \tag{9}$$

The matrices S, U are found by calculating the eigenvalues $\left(S^2\right)$ and eigenvectors (U) [11] of $\mathrm{cov}(q)$. This transformation makes the marginal pdfs $f_k(q_k)$ to differ substantially from the univariate pdfs ϕ_k of the independent but scaled parameters $S\tilde{q}_k$,

$$f_k(q_k)=\int\phi_1\left([Uq]_1\right)\phi_2\left([Uq]_2\right)\cdots\phi_k\left([Uq]_k\right)\cdots\phi_n\left([Uq]_n\right)dq_1\cdots dq_{k-1}dq_{k+1}\cdots dq_n\neq\phi_k(q_k),\ \text{if } U\neq I. \tag{10}$$

All ϕ_k are hence mixed according to U. Dependencies are thus difficult to account for. One rare exception is provided by the multinomial distribution [9]. It is often better to assign the pdfs to the canonical parameters in the original instead of the canonical basis. The transformation then reads $\tilde{q}: q=U^T SU\tilde{q}$. As required, it leaves cov(q) invariant. The marginalization in Eq. 10 changes accordingly, $U\to SU^T S^{-1}U$. Since the transformation $U^T SUS^{-1}$ of $S\tilde{q}_k$ contains cancelling operations U, U^T and S, S^{-1}, it is generally less distorting than U^T. Indeed, if the commutator $[S,\ U^T]\equiv SU^T-U^T S$ vanishes, $U^T SUS^{-1}=I$. The transformation U^T must satisfy the stronger criterion $U=I$ to avoid mixing. For any transformation $q\to Wq$, an indicator of mixing of the components of q is given by,

$$\Psi(W)\equiv\frac{1}{n}\sum_{r=1}^{n}\left(1-\frac{\max_c|W_{rc}|-\min_c|W_{rc}|}{\|W_{r,:}\|}\right)\in[0,1],\quad \|W_{r,:}\|\equiv\sqrt{\sum_{c=1}^{n}|W_{rc}|^2}. \tag{11}$$

A simple example illustrates that the mixing effect can be considerable, even for minute correlations. Assume a model has two parameters with a covariance matrix,

$$\operatorname{cov}(q)=\begin{pmatrix}0.90 & 0.10\\ 0.10 & 0.90\end{pmatrix}\Leftrightarrow U=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}, S^2=\begin{pmatrix}1 & 0\\ 0 & 0.8\end{pmatrix}\Rightarrow\begin{cases}\phi_1(S\tilde{q}_1) & = & \text{UNI}\left([0,1]\right)\\ \phi_2(S\tilde{q}_2) & = & \text{UNI}\left(\left[0,\sqrt{0.8}\right]\right)\end{cases}. \tag{12}$$

Large rotations are required because the canonical variances S_{jj}^2 are similar, i.e. cov(q) is almost *degenerate*. As shown in Fig. 1, the large rotations mix the assigned pdfs $\phi_k(S\tilde{q}_k)$ to marginal pdfs $f_k(q_k)$ beyond recognition for the transformation U^T but not for $U^T SUS^{-1}$.

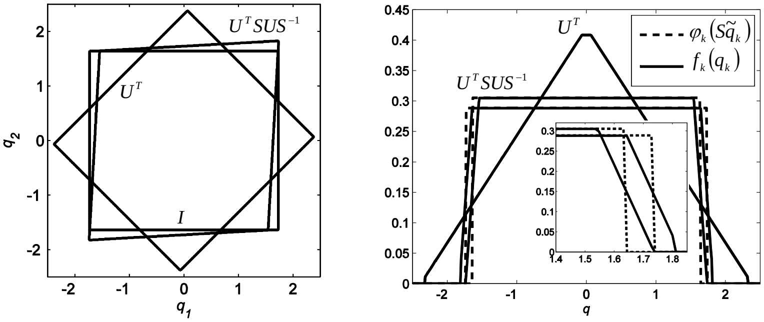

Figure 1. Left: The sample space of independent scaled parameters $(I: q_k=S\tilde{q}_k)$ (Eq. 12), and of the two transformations (U^T) (rotated) and $(U^T SUS^{-1})$ (skewed and tilted). Right: Assigned pdfs $\phi_k(S\tilde{q}_k)$ (dashed) and obtained margin-

al pdfs $f_k(q_k)$ (solid)) with mixing $\Psi(U^T)=1.00$ and $\Psi(U^T SUS^{-1})=0.058$, and magnified upper transition region (inset).

Specifying both marginal probability distributions and covariance is either redundant or inconsistent, as the latter is uniquely determined by the former. Nevertheless, this reflects the typical available information for signal processing applications. The moments can be accurately determined [4] for sufficiently large data sets but the joint distribution $f(q)$ is hardly ever known with any precision. Some of its properties are usually assigned, with varying degree of confidence. For instance, the allowed maximal range $M^{(\infty)}$ of the parameters of digital filters is given by stability constraints. The transformation technique above is well adapted to these facts, since the covariance is prioritised. The transformation $q=U^T SU\tilde{q}$ will be utilized in section 5.2 to include correlations with limited mixing of the statistics assigned to independent normalized canonical parameters $\tilde{q}$.

3.3. Refinements of Monte Carlo

To increase the efficiency of MC, the original brute force sampling technique has been further developed in mainly two directions: model simplification and sample distribution improvement. In response surface methodology (RSM) [18], the model is replaced by a simple approximate surrogate model. A model of order v may be found by applying linear (with respect to C) regression at *collocation points* [15] $\hat{q}^{(k)}=\mu^{(k)}$,

$$H(\mu)\approx R(\mu)C, \quad \begin{cases} R_{kj} &= R_j\left(\mu^{(k)}\right), \quad j=1,2,\ldots,v \\ H_k &= H\left(\mu^{(k)}\right) \end{cases}, \quad \begin{cases} C &= \left(C_1 \quad \cdots \quad C_v\right)^T \\ \mu &= \left(\mu^{(1)} \quad \cdots \quad \mu^{(m)}\right)^T \end{cases}, \quad m\geq v, \tag{13}$$

where $R_j(q)$ is basis function j. Since it may be non-linear, RSM allows for non-linear propagation of uncertainty and may give a substantially different and more accurate result than LIN. If only linear basis functions are used $R_j(q)=q_j$, RSM becomes equivalent to LIN. The best least square approximation is directly obtained from Eq. 13 [19],

$$C=\left(R^T R\right)^{-1} R^T H \tag{14}$$

Let RSM(r) utilize a complete set of mixed polynomial basis functions up to order r. Its least number (v) of collocation points grows rapidly with both the number of parameters (n) and polynomial order (r) [12],

$$v=\sum_{k=0}^{r} w(n,k): \quad w(n,k)=\sum_{j=1}^{\min(n,k)} \binom{n}{j} \cdot w(j,k-j), \quad w(j,0)=1. \tag{15}$$

In practice, $r>3$ often yields an unacceptable number of samples, see table 1.

	$n=2$	$n=5$	$n=10$	$n=20$
$r=1$	3	6	11	21
$r=2$	6	21	66	231
$r=3$	10	56	286	1771

Table 1. Efficiency $\rho=v$ for RSM(r), for selected polynomial orders r and numbers n of parameters.

The distribution of samples may be improved with stratification, as in Latin Hypercube sampling (LHS) [18]. By dividing the sample space into intervals, or stratas representing equal probability the need for large ensembles is reduced. In LHS, each parameter is sampled exactly once in each of its stratas giving a generalized *latin square* [20]. This selection pushes the samples away from each other and distributes them more evenly. To illustrate the improvement with stratification, sample one parameter $q\sim$NRM(0, 1). After division into m intervals of equal probability, samples are found with the inverse transform method described in section 3.2 (Eq. 8). As seen in Fig. 2, the convergence improves dramatically. Still, even for $m=100$ samples the second moment (left) varies noticeably. The convergence is generally poorer for higher order moments $M^{(k)}$, as shown for $k=4$ (right).

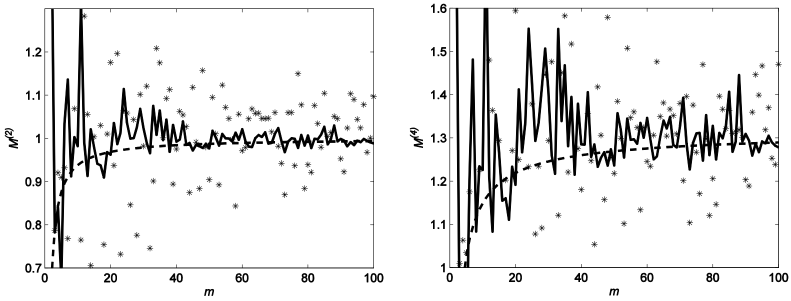

Figure 2. The second $M^{(2)}$ (left) and fourth $M^{(4)}$ (right) moments for stratified (solid) and brute force sampling ($*$) of $q\sim$NRM(0, 1), compared to a fixed grid (dashed).

In this case, it is questionable if 100 samples are sufficient to represent as few as four moments $M^{(1)}-M^{(4)}$. The probabilistically evenly distributed fixed grid (dashed) converges more rapidly to the proper statistics. Despite the prevailing tradition, there is no absolute requirement of using a random generator to represent statistical information. Fixed grids are examples of deterministic sampling. Stratification provides an interesting intermediate type of sampling since it is partially deterministic – the strata are constructed deterministically but the samples within each stratum are generated randomly. The construction of a fixed grid requires focus on the most relevant features. To reproduce $M^{(1)}-M^{(4)}$ *exactly*, a very sparse grid or few deterministic samples are needed,

$$\hat{q}=\begin{cases}\left(\pm 1.376 \quad \pm 0.325\right) & q\sim \mathrm{UNI}\left(0,1\right)\\ \left(\pm 1.732 \quad 0\left(\times 4\right)\right) & q\sim \mathrm{NRM}\left(0,1\right)\end{cases}. \tag{16}$$

If the problem at hand only depends on these moments, the exact solution will be obtained. The size of such small ensembles must be fixed, no matter how they are generated. Adding, or perturbing a single sample would modify the statistics substantially.

4. Deterministic sampling

Deterministic sampling (DS) of uncertain systems is a viable alternative to random sampling (RS). Instead of using random generators, specific DS rules are devised to generate appropriate, but still statistical (Fermi's notation, see section 3.2) ensembles. A rudimentary example illustrates the principle: Assume a model $y(q)$ depends on one parameter q with mean $\langle q\rangle$ and variance $\langle \delta q^{2}\rangle$. To estimate the mean $\langle y\rangle$ and the variance $\langle \delta y^{2}\rangle$ of the model, the samples (filter parameters) $\hat{q}^{(1,2)}=\langle q\rangle \pm \sqrt{\langle \delta q^{2}\rangle}$ are appropriate since they satisfy the desired statistics, $\langle \hat{q}\rangle_{E}=\langle q\rangle$ and $\langle \delta \hat{q}^{2}\rangle_{E}=\langle \delta q^{2}\rangle$. The formula for $\hat{q}^{(1,2)}$ constitutes the sampling rule and $\hat{q}^{(1,2)}$ is the statistical ensemble containing only two model samples. By paying the computational cost of using more samples and improving the sampling rule, additional moments $\langle \delta q^{k}\rangle$, $k>2$ or other statistical features can be accounted for.

In deterministic sampling the model evaluations involve no approximations and are *non-invasive*. In many respects, deterministic sampling is constructed and optimized for quantification of modeling uncertainty: Minimal ensembles allow for evaluation of the most numerically demanding models. The model evaluations are exact and non-invasive to fully respect non-linear deeply hidden parameter dependences. Only vaguely known statistics of the model is approximated.

4.1. Concepts of deterministic sampling

DS does not per se specify the goal of sampling, e.g. given mean and covariance of the parameters. In the example at the end of section 3.3, the primary target was the joint pdf of the parameters. In section 4.2, the target is $M^{(2)}(q)$. In section 5, this will be complemented with additional requirements. DS can also be utilized for direct evaluation of confidence intervals [12]. The targets of various DS methods may differ but the focus on the most influential statistical aspect and customization is shared. In stark contrast, almost without exception RS targets the joint pdf of the parameters and ignores the final utilization. Adaptation and fixed ensemble sizes provides the principal means to improve the efficiency of sampling.

4.2. Propagation of covariance in the standard unscented Kalman filter

The reference will be the specific variant of DS used for propagating covariance in what will be referred to as the standard unscented Kalman filter (UKF) [21-23]. The ensemble consists of 2n samples, or *sigma-points*,

$$\hat{q}^{(s,k)} \equiv \langle q \rangle + s \cdot \sqrt{n} \cdot \Delta_{:k}, \quad \Delta\Delta^T = \mathrm{cov}(q), \quad k = 1,2,\ldots n, \quad s = \pm \tag{17}$$

where $\Delta_{:k}$ denotes the k-th column of Δ. The sampling rule is manifested in the square root calculation of the covariance matrix (Δ). As suggested [23] it may be found with a Cholesky factorization [19]. The square root matrix is not unique though – the Cholesky root is upper triangular and thus asymmetric. A more symmetric standard alternative is to evaluate the matrix square root in a canonical basis [24] Uq where cov(Uq) is diagonal. The canonical variations $U\delta\hat{q}^{(s,v)}$ will be unit vectors in the n positive and negative directions of the *principal* axes of the covariance matrix, amplified by the marginal standard deviations and most importantly, $\sqrt{n}$. For many parameters with large covariance, the scaling with $\sqrt{n}$ may cause the UKF to fail since the scaling is not related to the variability of the parameters, only their total number. A possible solution to the scaling problem is provided by the scaled unscented transformation [25]. However, it is based on Taylor expansions and thus suffers from an approximation problem of the model.

5. Sampling with conservation of moments

One class of methods of deterministic sampling conserves a limited number of statistical moments. The model parameters are sampled to satisfy these moments and collected in ensembles, similar to how parameters are sampled to fulfill probability distributions in RS.

5.1. Principle

The constraints of satisfying statistical moments constitute an infinite system of equations for the samples $\delta\hat{q}_i^{(v)}$. It can formally be viewed as sampling ($\hat{=}$) of the joint pdf $f(q)$,

$$
\begin{array}{lclclcl}
0=\left\langle \delta q_i \right\rangle & \equiv & \int \delta q_i f\left(q\right)dq & \hat{=} & \frac{1}{m}\sum_{v=1}^{m}\delta\hat{q}_i^{(v)} & \equiv & \left\langle \delta\hat{q}_i \right\rangle_E \\
\left\langle \delta q_{i1}\delta q_{i2} \right\rangle & \equiv & \int \delta q_{i1}\delta q_{i2} f\left(q\right)dq & \hat{=} & \frac{1}{m}\sum_{v=1}^{m}\delta\hat{q}_{i1}^{(v)}\delta\hat{q}_{i2}^{(v)} & \equiv & \left\langle \delta\hat{q}_{i1}\delta\hat{q}_{i2} \right\rangle_E \\
\left\langle \delta q_{i1}\delta q_{i2}\delta q_{i3} \right\rangle & \equiv & \int \delta q_{i1}\delta q_{i2}\delta q_{i3} f\left(q\right)dq & \hat{=} & \frac{1}{m}\sum_{v=1}^{m}\delta\hat{q}_{i1}^{(v)}\delta\hat{q}_{i2}^{(v)}\delta\hat{q}_{i3}^{(v)} & \equiv & \left\langle \delta\hat{q}_{i1}\delta\hat{q}_{i2}\delta\hat{q}_{i3} \right\rangle_E \\
\vdots & \equiv & \vdots & \hat{=} & \vdots & \equiv & \vdots
\end{array}
\tag{18}
$$

The infinite number of equations requires an infinite number of samples. However, it is implicitly assumed that relatively few moments are known and significantly influence the result of interest. Only a few moments then needs to be accurately represented by $\{\delta\hat{q}^{(v)}\}$. Typically, $\langle\delta q_i\rangle$ and $\langle\delta q_{i1}\delta q_{i2}\rangle$ are estimated when models are identified [4,7]. In addition, the range $M^{(\infty)}$ or another higher diagonal moment can generally be determined from underlying physical constraints like stability. Clearly, any sampling rule must generate a fixed number of samples and create them simultaneously. The samples are consequently strongly dependent. One obvious sampling method is to solve Eq. 18 numerically for a sufficiently large number of samples $\delta\hat{q}$, as in Eq. 16. Due to the strong non-linearities, this is quite difficult for a large number of moments but may be feasible for a few moments.

5.2. The excitation matrix

The UKF (section 4.2) utilizes DS with conservation of all first $(\langle\delta q_i\rangle)$ and second $(\langle\delta q_{i1}\delta q_{i2}\rangle)$ statistical moments. The invariance in its *formulation* allows for any additional 'half' unitary transformation $\Delta\rightarrow\Delta V,\ V:VV^T=I$. This results in another equally valid matrix $\widetilde{\Delta}$, since $\widetilde{\Delta}\widetilde{\Delta}^T=\Delta V[\Delta V]^T=\Delta VV^T\Delta^T=\Delta\Delta^T$. Since the transformation V is allowed and influences the result, the result of applying the UKF is not unique. The matrix V condenses this invariance and provides practical means to manipulate the UKF ensemble. A key feature of V is the absence of constraints on V^TV. That makes it possible to stretch V 'horizontally' (as long as $VV^T=I$). That corresponds to adding samples (sigma-points). The improved transformation U^TSU (section 3.2) can be applied by also combining U with V. The square root of the covariance matrix will then read $\Delta=U^TSUV$ instead of $\Delta=U^TSV$,

$$
\Sigma_{n\times m}\equiv\langle q\rangle\cdot 1_{1\times m}+U^TSU\hat{V},\quad \hat{V}\equiv\sqrt{m}\cdot V_{n\times m},\quad VV^T=I,\quad V\cdot 1_{m\times 1}=0. \tag{19}
$$

The samples $\hat{q}^{(k)}$ are here collected in columns of the ensemble matrix Σ. The matrix $\mathrm{cov}(q)=\Delta\Delta^T=U^TS^2U$ is diagonalized,$S^2=\mathrm{eig}(\mathrm{cov}(q))$, with the unitary transformation U [24]. The normalization factor$\sqrt{m}$ is included in the *excitation matrix* $\hat{V}$ to satisfy the correct covariance, $\langle\Sigma\Sigma^T\rangle=\Delta\Delta^T$, just as the factor $\sqrt{n}$ was included in Eq. 17 (the ensemble is here expanded from n to m samples). The excitation matrix controls the sampling *beyond* the first and second moments, e.g. the range of the samples. Row k of the matrix $S\hat{V}$ can be interpreted as deterministic samples of the pdf $\phi_k(S\tilde{q}_k)$, assigned to canonical parameters in RS, see section 3.2. All ensembles will be described with a unique excitation matrix$\hat{V}$.

The adopted transformation $U^T SUS^{-1}$ of $S\hat{V}$ distorts all higher moments than the second. This mixing effect is indicated by the index $\Psi(U^T SUS^{-1})$ defined in Eq. 11. To diagonalize large matrices cov(q) many efficient techniques have been developed. This should not cause any difficulties even for $n \sim 1000$, especially since cov(q) usually is either very sparse, or rank deficient for models with many parameters.

5.3. Elimination of singular values

The rationale for applying the reduction to be presented is that any model is derived from a limited set of experiments, resulting in a usually moderate rank of cov(q). If the number of parameters is large, it is thus often (nearly) rank deficient. The widely practiced singular value decomposition (SVD) [19] may then be used to reduce the excitation matrix and hence the number of samples. The most general form of SVD cannot be used here since it renders an asymmetric decomposition $\mathrm{cov}(q) = U^T S^2 W$, where $UU^T = U^T U = WW^T = W^T W = I$ and $S^2_{jk} = \delta_{jk} S_{kk} \geq 0$. Different matrices U, W allow for decomposition of an arbitrary matrix. For the symmetric matrix cov(q), a symmetric SVD $U = W$ can be found with the less general eigenvalue decomposition [24], according to the spectral theorem [11]. As cov(q) is positive definite, all its eigenvalues S^2_{kk} fulfill the requirement of being positive. This is required to directly obtain a real-valued matrix square root $\Delta = U^T SUV$.

The ensemble may now be reduced by elimination of singular values (ESV). Choose a threshold α and remove row r and column r from S and row r from U for all r such that,

$$\left|S_{rr}\right| < \alpha \cdot \max_k \left|S_{kk}\right|, \quad \alpha << 1. \tag{20}$$

Proceeding as in many applications of SVD, this reduction (indicated by tilde below) will not change the result significantly, if α is small enough. Accordingly, samples are eliminated using the alternative decomposition of the square root of cov(q),

$$\Delta_{n\times m} = \left(U^T\right)_{n\times n} S_{n\times n} V_{n\times m} \approx \left(\tilde{U}^T\right)_{n\times r} \tilde{S}_{r\times r} \tilde{V}_{r\times \mu} = \tilde{\Delta}_{n\times \mu}. \tag{21}$$

Unfortunately, the less distorting transformation $U^T SUS^{-1}$ of $S\hat{V}$ advocated in section 3.2 do not allow for $r < n$ rows of the matrix V. The increase in distortion of $M^{(k>2)}$ indicated by $\Psi(U^T SUS^{-1}) \to \Psi(U^T)$ is less important for the intended use though. Signal processing models with large numbers of parameters are typically non-parametric and usually describe samples of signals like impulse responses, or noise signals. The required LR property of the system then implies LP. The propagation of covariance is then linear and only the undistorted first and second moments need to be encoded.

5.4. Correlated sampling of non-parametric models

A major difference between parametric and non-parametric models is the dimensionality. A conceptual dissimilarity is that non-parametric models usually refer to correlated signals, rather than abstract model structures. The parameters may describe discrete samples of input noise [7], or an impulse response [6]. A common parametric pole-zero model may contain 20 parameters, while a non-parametric model can be expressed in perhaps 1000 parameters. The ensembles of non-parametric models often need to be reduced drastically.

Due to limited resolution, the correlation times of any signal or impulse response is finite. Their 'memory' is thus finite so sample variations may be regenerated or repeated, as long as the time between repetitions exceeds the correlation time. This *correlated sampling* (CRS) provides efficient and accurate reduction of the ensembles. The minimal number of parameters n is then set by the correlation time of the model. Most importantly, the size of the ensemble becomes independent of the size of the model (the length of the signal).

A finite correlation length $\tau \in N$ of any model $\delta x(t)$ is normally inferred from the decay of its autocorrelation function$C(t,\ T)$, where t denotes the lag and T refers to a non-stationary variation. Here, a global τ will be defined through its l^2-norm and determined for a relative truncation threshold β (argmin returns the minimizing argument),

$$\tau = \max_T \left[\arg\min_\tau \left| \sum_{t=\tau+1}^{\infty} \left|C(t,T)\right|^2 - \beta^2 \sum_{t=0}^{\infty} \left|C(t,T)\right|^2 \right| \right], \quad C(t,T) \equiv \left\langle \delta x\left(T+\frac{t}{2}\right) \delta x\left(T-\frac{t}{2}\right) \right\rangle, \quad \beta << 1. \tag{22}$$

If the model is expressed as a convolution $\delta x(t){=}h\,(t) * w(t)$ of an impulse response $h\,(t)$ and time-dependent white noise $w(t)$ as in section 6.2,

$$C(t,T) = \sum_{u=0}^{\infty} \eta^2\left(T-\left(u+\frac{t}{2}\right)\right) h(u)h(u+t) \approx \eta^2(T) \sum_{u=0}^{\infty} h(u)h(u+t), \quad \eta(t) = \sqrt{\mathrm{var}(w)}, \quad \langle w \rangle = 0. \tag{23}$$

By padding the model to an integer multiple of $\gamma \geq 2\tau$ samples, it is always possible to choose an excitation matrix partitioned to block-diagonal form,

$$\hat{V}_{n\times m} = \begin{pmatrix} c\tilde{V}_{\gamma\times\gamma} & 0_{\gamma\times\gamma} & 0_{\gamma\times\gamma} & \cdots \\ 0_{\gamma\times\gamma} & c\tilde{V}_{\gamma\times\gamma} & 0_{\gamma\times\gamma} & \cdots \\ 0_{\gamma\times\gamma} & 0_{\gamma\times\gamma} & c\tilde{V}_{\gamma\times\gamma} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \tilde{V}_{\gamma\times\gamma} \tilde{V}^T_{\gamma\times\gamma} = \gamma \cdot I, \quad c = \sqrt{\frac{m}{\gamma}}, \tag{24}$$

where $\widetilde{V}_{\gamma\times\gamma}$ is any allowed deterministic sub-ensemble. The factor c accounts for the change from γ samples of $\widetilde{V}_{\gamma\times\gamma}$ to the $m>\gamma$ samples of $\hat{V}_{n\times m}$. By violating the normalization constraint $\hat{V}_{n\times m}\hat{V}^{T}_{n\times m}=m$, the size of the ensemble can be 'compressed' from m to γ samples by moving all sub-matrices $c\widetilde{V}_{\gamma\times\gamma}$ to the first block-column and skipping all zeros. The introduced constant c drops out as $m\rightarrow\gamma$,

$$\hat{V}_{n\times\gamma}\equiv\begin{pmatrix}\tilde{V}_{\gamma\times\gamma}\\ \tilde{V}_{\gamma\times\gamma}\\ \tilde{V}_{\gamma\times\gamma}\\ \vdots\end{pmatrix},\quad \hat{V}_{n\times\gamma}\left(\hat{V}_{n\times\gamma}\right)^{T}=\gamma\begin{pmatrix}I_{\gamma\times\gamma} & I_{\gamma\times\gamma} & I_{\gamma\times\gamma} & \cdots\\ I_{\gamma\times\gamma} & I_{\gamma\times\gamma} & I_{\gamma\times\gamma} & \cdots\\ I_{\gamma\times\gamma} & I_{\gamma\times\gamma} & I_{\gamma\times\gamma} & \cdots\\ \vdots & \vdots & \vdots & \ddots\end{pmatrix}\neq\gamma\cdot I_{n\times n}. \tag{25}$$

Accordingly,

$$\operatorname{cov}\left(x\right)_{jk}\rightarrow\left(U^{T}SUVV^{T}U^{T}SU\right)_{jk}=\begin{cases}\operatorname{cov}\left(x\right)_{jk}, & \left|j-k\right|\leq\gamma/2\\ \operatorname{cov}\left(x\right)_{j,k+n\gamma}, & \begin{cases}\left|j-k\right|>\gamma/2\\ n=\arg\min\limits_{l\in Z}\left|j-k-l\gamma\right|\end{cases}\end{cases}. \tag{26}$$

The consequence of violating the normalization constraint is that only a limited diagonal band of $\operatorname{cov}(x)$ is correctly reproduced. If a non-parametric model of a signal is propagated through a system model with impulse response h of correlation length $\sigma\leq\gamma/2$ this will nevertheless *not* result in any error of $\operatorname{var}(h)$, as it is independent of all faulty elements $\operatorname{cov}(v)_{jk}$, $|j-k|>\gamma/2\geq\sigma$. To correctly evaluate $\operatorname{cov}(h)_{uv}$ though, the size of the sub-ensembles $\widetilde{V}_{\gamma\times\gamma}$ of correlated sampling must fulfill the stronger size constraint $\gamma\geq2(\max(\tau,\sigma)+|u-v|)$. The symmetry of convolutions implies a corresponding result when the non-parametric model describes the impulse response h of the system, rather than a signal.

5.5. Combining covariance

A signal processing model generally includes both parametric and non-parametric sources of uncertainty. For instance, a device (parametric system model) may be fed with a signal corrupted with noise (non-parametric noise model). The question then arises how the two sources q_k, x_k of uncertainty can be combined. For propagation of uncertainty through LP models, the combined covariance is given by the Gauss approximation formula [16],

$$\operatorname{cov}\left(y\right)\overset{\mathrm{LP}}{=}\sum_{k}\operatorname{cov}\left(y^{(k)}\right), \tag{27}$$

where $\mathrm{cov}(y^{(k)})$ is the propagated covariance of q_k, x_k. This will seize to apply for non-LP models. There exists no general non-linear summation rule for propagated covariance. A method of summation can be given though, if different ensembles are combined as in RS.

To combine ensembles of parametric (q) and non-parametric models (x), collect all parameters, $q \rightarrow (q^T \quad x^T)^T$, and diagonalize the enlarged covariance matrix, $\mathrm{cov}(q) = U^T S^2 U$. Build $\hat{V}$ with two blocks and use CRS (section 5.4) for the non-parametric model,

$$\hat{V}_{(n+k\gamma)\times(m+v)} \equiv \begin{pmatrix} \sqrt{1+c^{-1}} \cdot \hat{V}_{n\times m} & 0 \\ 0 & \sqrt{1+c} \cdot \hat{V}_{\gamma\times v} \\ 0 & \sqrt{1+c} \cdot \hat{V}_{\gamma\times v} \\ \vdots & \vdots \\ 0 & \sqrt{1+c} \cdot \hat{V}_{\gamma\times v} \end{pmatrix}, \quad c = \frac{m}{v}, \quad \hat{V}_{(n+k\gamma)\times(m+v)} \left(\hat{V}_{(n+k\gamma)\times(m+v)} \right)^T = (m+v) \cdot I. \tag{28}$$

The scaling $\sqrt{1+c^{\pm 1}}$ may cause a similar scaling problem as the factor $\sqrt{n}$ in the UKF (section 4.2). Using extended excitation matrices these factors can be eliminated,

$$\hat{V}_{(n+k\gamma)\times c} \equiv \begin{pmatrix} \hat{A}_{n\times c} \\ \hat{B}_{\gamma\times c} \\ \hat{B}_{\gamma\times c} \\ \vdots \\ \hat{B}_{\gamma\times c} \end{pmatrix}, \quad \hat{E}_{(n+\gamma)\times c} = \begin{pmatrix} \hat{A}_{n\times c} \\ \hat{B}_{\gamma\times c} \end{pmatrix}, \quad \hat{E}_{(n+\gamma)\times c} \left(\hat{E}_{(n+\gamma)\times c} \right)^T = c \cdot I, \quad c = \max(m, v) > (n+\gamma). \tag{29}$$

A disadvantage of this summation is that the same type of ensemble must be used for all parameters. Both alternatives combine the statistics of the two models non-linearly. The uncertainties are propagated and combined by evaluating the model for all samples and calculating the desired statistics, just as if the combined ensemble described one model.

5.6. Selected ensembles

The standard (STD) ensemble employed in the UKF (as defined in section 4.2) utilizes the perhaps simplest possible excitation matrix,

$$\hat{V}_{\mathrm{STD}} = \sqrt{n} \cdot \left(I_{n\times n} \quad -I_{n\times n} \right), \quad m = 2n. \tag{30}$$

While the ultimate simplicity is its main advantage, the long maximal(!) range $M^{(\infty)}$ is its main disadvantage.

How far the reduction of samples might be driven is illustrated by the minimal simplex (SPX) ensemble,

$$\hat{V}_{\mathrm{SPX}} = \sqrt{n+1}\cdot \perp \left\{ \begin{pmatrix} I_{n\times n} & -1_{n\times 1} \end{pmatrix} \right\}, \quad m = n+1, \tag{31}$$

where the operator $\perp$ performs classical Gram-Schmidt orthogonalization [11] and normalization of rows. The ensemble is constructed from half the STD ensemble, complemented with one sample $1_{n\times 1}$ to cancel the first moments. Since that violates the orthogonality of the rows of V, $\perp$ must be applied to satisfy $V V^{T} = I$. The high efficiency of the SPX ensemble is tarnished by its large skewness, or $M^{(3)}$. This may give considerable bias of propagated covariance for non-LP models, but is irrelevant for LP models.

The binary (BIN) ensemble has minimal range to guarantee allowed samples. By varying all parameters with an equal magnitude of one standard deviation in all samples, the diverging factor $\sqrt{n}$ of the STD is eliminated. Its excitation matrix $\hat{V}_{BIN}$ is fundamentally constructed from a standard binary array, with the difference that the allowed levels are ±1 instead of 0, 1 (see rows 1-3 in Eq. 32). It is then complemented with supplementary rows obtained in two ways, by cyclic shifting and mirror imaging,

$$\hat{V}_{BIN}^{(m)} = \begin{pmatrix} +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & \cdots \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 & \cdots \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & \cdots \\ -1 & +1 & +1 & -1 & -1 & +1 & +1 & -1 & \cdots \\ -1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 & \cdots \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & \cdots \\ -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad m = 2^{\mathrm{ceil}\left(\frac{\mathrm{n}+5}{4}\right)}. \tag{32}$$

Cyclic shifts are applied to all original rows except the first, by a quarter of their periodicity. Mirror imaging of a row is defined to change the sign of its second half and is applied to all original rows except the last two, and all shifted rows except the last. For instance, in Eq. 32 row 4 and 5 are shifted versions of row 2 and 3, while rows 6 and 7 are the mirror images of rows 1 and the shifted row 4. The supplementary rows reduce the size of the ensemble drastically with a corresponding improvement of the efficiency. For $n=20$ parameters, the size drops from roughly 10^6 to 128 samples. That size is acceptable in perspective of the $n+1=21$ samples of the most efficient SPX. Eventually though, the number of samples will grow too large. The BIN can thus only be applied to moderately sized models.

By no means, this brief survey exhausts all possible ensembles. Many criteria for selecting the most appropriate ensemble can be formulated. Here, the first and second moments, parameter ranges and efficiency were in focus.

6. Application — Modeling uncertainty of a dynamic device

The task is to simulate the response of an electrical device such as an amplifier or oscilloscope, in the presence of non-stationary correlated noise on its input. An uncertain LR CT model of the device and its parametric covariance is usually found by applying system identification techniques [4] on calibration measurements [6]. Such a model of the system can be sampled into a digital filter and be described in the pole-zero form in Eq. 2. These standard steps will here be omitted. The system model will instead be assigned to a digital low-pass Butterworth filter, of order 10 and cross-over frequency $f_C = 0.1 f_N$, f_N being the Nyquist frequency and described by parameters K, p_1, p_2, $\ldots p_{10}$, z_1, z_2, z_{10}. The complete correlations of complex-conjugated pole (p) and zero (z) pairs are eliminated by a transformation from $q = z,\ p$ to Re(q), Im(q)≥0, giving $n = 21$ system model parameters,

$$q \equiv \left(K \quad \mathrm{Re}\left(z_1\right) \quad \mathrm{Im}\left(z_1\right) \geq 0 \quad \cdots \quad \mathrm{Re}\left(p_1\right) \quad \mathrm{Im}\left(p_1\right) \geq 0 \quad \cdots \quad \mathrm{Re}\left(p_{10}\right) \quad \mathrm{Im}\left(p_{10}\right) \geq 0\right)^T . \tag{33}$$

To be most general, the non-parametric input noise model is chosen to be correlated/colored and non-stationary. The noise parameter δx_k represents the noise level at time sample k. Its *generating signal* [7] is a Dirac delta function δ_{jk}, centered at time k. The response of a system with impulse response h will be $\delta y_j = \delta x_k \cdot \left(h_j * \delta_{jk}\right) = h_{j-k} \cdot \delta x_k$. In matrix notation, $\delta y = \bar{h}^T \delta x$, where $\bar{h}_{kj} = h_{j-k}$. Hence,

$$\mathrm{cov}\left(y\right) = \left\langle \delta y \delta y^T \right\rangle = \bar{h}^T \left\langle \delta x \delta x^T \right\rangle \bar{h} = \bar{h}^T \,\mathrm{cov}\left(x\right) \bar{h} \tag{34}$$

Since the response is linear in noise parameters, it is sufficient to only capture cov(x).

6.1. Reference ensembles

Traditionally, any method for uncertainty propagation is evaluated by comparisons with the default method of linearization [10,16], and brute-force random sampling (MC) [9] as state-of-the-art. There are several drawbacks of this approach. Linearization is a coarse approximation for LP models and MC suffer from the difficulty of modeling dependencies and low efficiency. An alternative is to construct finite reference ensembles (REF) and by definition let them describe the truth. Their primary advantage is that the finite size of the REF makes it possible to propagate the uncertainty exactly, using all REF samples. A more or less arbitrary REF may be generated randomly, like any MC ensemble. All requirements are also automatically fulfilled since the REF is built of possible realizations. Also, the REF closes the loop as it makes it possible to compare 'true' and approximate samples directly on an equal footing (see Fig. 8). Even though the samples differ substantially, the resulting modeling uncertainties can be similar.

A plausible REF δq_j for the system model realized as a digital filter is created by randomly generating m samples of n parameters q_k from uniform distributionsUNI$(0,\ \sigma_k)$, with σ_k listed in Fig. 3, top left. The joint pdf will have compact support [11], as required to guarantee stability. The mean is subtracted from all samples to remove the bias of the finite random ensemble, $\langle \delta q_j \rangle_E = \langle \delta q_j \rangle = 0,\ \ \forall\ j$. The covariance of the REF will have a desirable more or less random variation for small values of m. If the REF samples are arranged in columns of a matrix $\hat{\Lambda}_{n\times m}$, (as $\hat{V}$) $\mathrm{cov}(q)_{\mathrm{REF}} = 1/m \cdot \hat{\Lambda}\hat{\Lambda}^T$. For $m=31>n=21$, the strong correlations will expose the methods to severe tests with significant transformations $U,\ S$. The mixing Ψ (Eq. 13) using transformation $U^T S U S^{-1}$ was considerable, but less than for U^T, see caption Fig. 3. For the chosen REF, the resulting variations of poles and zeros are displayed in Fig. 3. The obtained variation of the parameters defined in Eq. 33 can be quantified with an averaged correlation index and standard deviation (Fig. 3, bottom left),

$$\xi_k \equiv \sqrt{\frac{1}{n-k}\sum_{j=1}^{n-k}\mathrm{cov}(q)_{j(j+k)}} \Bigg/ \sqrt{\frac{1}{n}\sum_{j=1}^{n}\sigma_j^2}\,,\quad \sigma_j^2 = \mathrm{var}(q_j). \tag{35}$$

A REF signal δx_j for non-stationary correlated noise may conveniently be generated from an autoregressive process (AR) acting on time-dependent zero mean white noise δw, $\delta x = \bar{g}^T \delta w$, where $\bar{g}_{kj} = g_{j-k}$ is the matrix of translated impulse responses g for the AR process defined by parameters α_k. Assigning a square wave time-dependence,

$$\begin{aligned} \sum_k \alpha_k \delta x_{j-k} &= \delta w_j,\quad \left\langle \delta x \delta x^T \right\rangle = \bar{g}^T \left\langle \delta w \delta w^T \right\rangle \bar{g} = \bar{g}^T \mathrm{cov}(w)\bar{g},\quad \mathrm{cov}(w) = \mathrm{diag}\left[\eta(t)\right] \\ \eta(t) &= \mathrm{N}\left[1 + \frac{1}{1+\psi}\theta\left[\cos\left(\frac{2t\pi}{T} + \varphi\right)\right]\right],\quad \theta = \begin{cases} +(-)1 & x > (<)0 \\ 0, & x = 0 \end{cases} \end{aligned} \quad . \tag{36}$$

The exact REF result of modelling covariance of noise is given by combining Eqs. 34 and 36,

$$\sigma_N \equiv \sqrt{\mathrm{Tr}\left(\mathrm{cov}(y)\right)} = \sqrt{\mathrm{Tr}\left(\bar{h}^T \bar{g}^T \mathrm{diag}\left[\eta^2\right]\bar{g}\bar{h}\right)}. \tag{37}$$

An explicit realization of the REF for the noise model is hence not needed. Specifically, a second order system $\alpha = [1 \ \ -0.4 \ \ 0.6]$ with time parameters $\{N,\ \psi,\ T,\ \varphi\} = \{0.05,\ 0.3,\ 2f_C^{-1},\ \pi/8\}$ was

Assigned variation σ		
K	$\mathrm{Re}(z_k)$, $\mathrm{Im}(z_k)$	$\mathrm{Re}(p_k)$, $\mathrm{Im}(p_k)$
2%	0.2%	0.9%

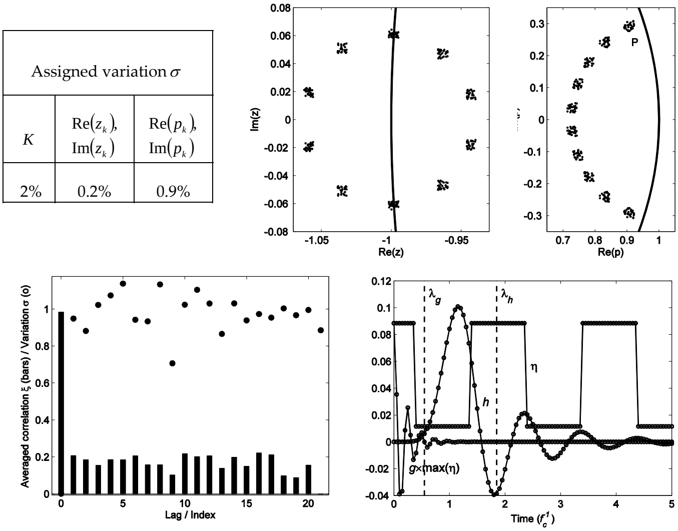

Figure 3. Top: Assigned variations (left) and resulting (z_k middle, p_k right) samples of the REF of the system model. Label P indicates the pole explored in Fig. 8. Bottom, left: Obtained variations σ_k (dots) and correlations ξ_k (bars) of parameters q (Eq. 33), with mixing (Eq. 11) $\Psi(U^T SUS^{-1})=0.22$ (adopted) and $\Psi(U^T)=0.39$. Bottom, right: Impulse responses $h(\langle q\rangle, t)$ and $g(t)$ and time-dependence $\eta(t)$ (Eq. 36) of noise intensity. The correlation lengths λ_h, λ_g were determined according to Eqs. 22-23, for $\beta=0.05$.

chosen. The impulse responses of the AR noise system and the system model, and the variation of the noise model are illustrated in Fig. 3, bottom right.

The 'true' result given by the response for the REFs for the different test signals is shown in Fig. 4. The propagated noise variation σ_N differs substantially from the input square wave η (top left) and is almost opposite in phase, due to the response time of about f_C^{-1}, see delay of μ_S (top, right and bottom). The signal distortion (μ_S) is strongly dependent on the input signal and decreases with increased regularity / differentiability. The propagated covariance σ_S has a more complex variation (top, right and bottom), as it is larger for the more regular Gaussian (bottom, right) than for the triangular pulse (bottom, left).

6.2. Deterministic sampling

The error of the scent and the standard deviation for the STD, SPX and BIN ensembles of the system model is displayed in Fig. 5, for all test signals. The low scent of the REF (left: thin, dotted) suggests the model is close to LP. Despite the relative errors are large they are quite small on an absolute scale. The SPX has the largest errors, for the scent as well as the variance. That is likely caused by its skewness being much larger than that of the REF. The BIN has the lowest errors and is thus the best approximate representation of the REF.

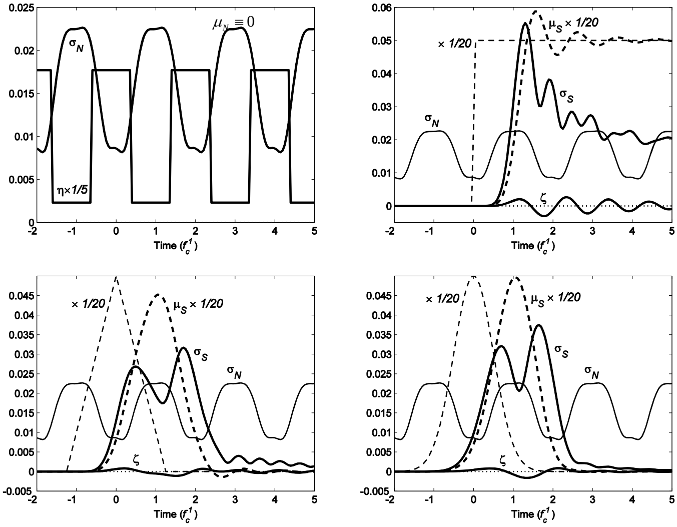

Figure 4. The mean μ_S (dashed), the standard deviations $\sigma_{S,N}$ (solid) and the scent ζ (thin, solid) for the REFs, for the different test signals (thin, dashed). The subscripts refer to the system (S) and noise (N) models. The variation of noise intensity is given by $\eta(t)$ (top, left).

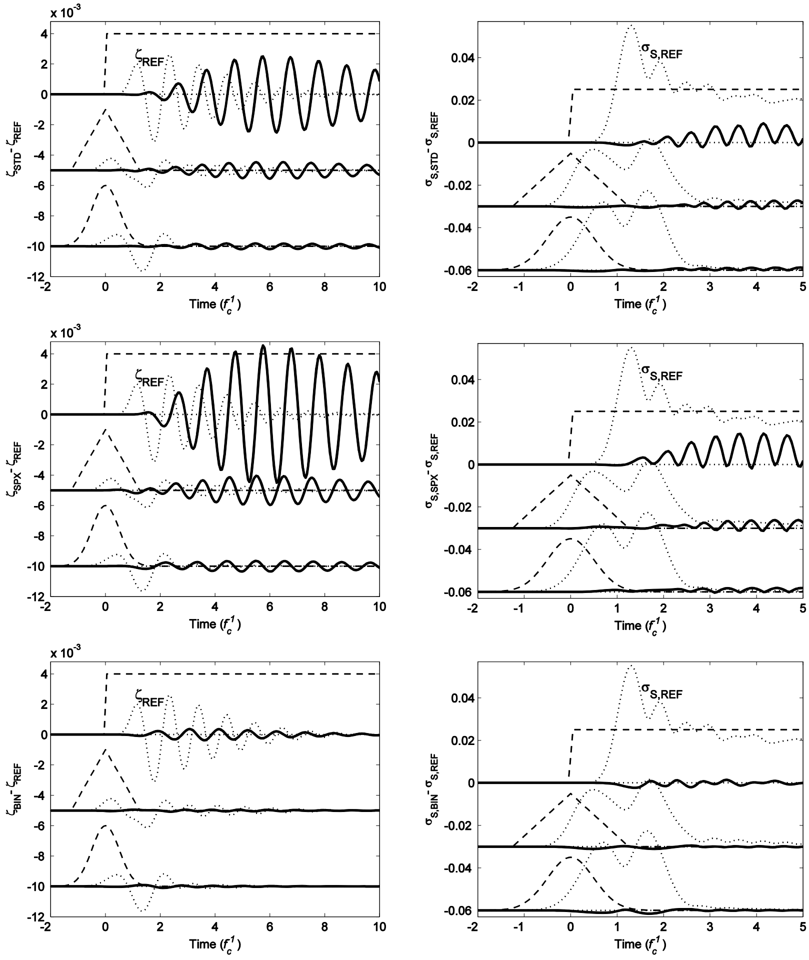

Figure 5. The errors of the scent $\zeta-\zeta_{\text{REF}}$ (left) and the standard deviation $\sigma_S-\sigma_{S,\text{REF}}$ (right) of the system model (solid) for the STD (top), SPX (middle) and BIN (bottom) and the three test signals (thin, dashed). The correct ζ_{REF} (left) and $\sigma_{S,\text{REF}}$ (right) are included for comparison (thin, dotted). The triangular and Gaussian signals are displaced for clarity.

The errors might appear large, considering *all* ensembles are 'correct', i.e. correctly represent (typically) available accurate information (mean and covariance of parameters). The errors reflect ambiguities caused by the ubiquitous lack of information in signal processing, rather than inadequacies of DS. RS can only produce better results by making further *assumptions*.

The result of applying the ESV and the CRS methods to reduce the SPX ensemble for propagating the noise is displayed in Fig. 6. By choosing sufficiently low thresholds α for elimination of singular values (ESV) and β for truncation of the correlation lengths (see Eqs. 20,22), the errors can be made arbitrarily low. As the reduction will decrease accordingly, there is a trade-off between accuracy and efficiency. For the chosen values, CRS is about twice as accurate and twice as efficient as ESV. In contrast to ESV, the number of samples for the CRS method is independent of the number of noise samples. The computational cost thus increases linearly with the length of the noise signal for CRS but quadratically (approximately) for ESV. For ESV to be most efficient, the model covariance needs to be strongly rank deficient. That is not as unlikely as it might appear, since the model usually is derived from a limited amount of experimental results.

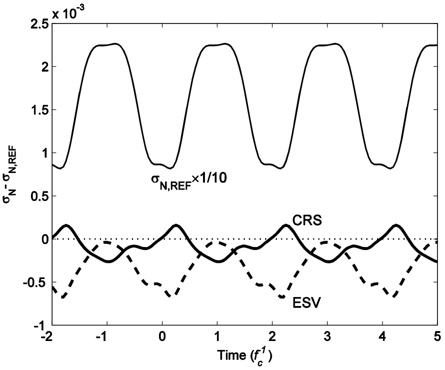

Figure 6. The error $\sigma_N - \sigma_{N,REF}$ of propagated noise, for the ESV (section 5.3) and CRS (section 5.4) ensemble reduction methods, and the correct $\sigma_{N,REF}$ (thin, ×1 / 10). The thresholds were $\alpha = 0.1$ (Eq. 20) for ESV, and $\beta = 0.05$ (Eq. 22) for CRS. That resulted in $m = 142$ samples for ESV and $m = 75$ for CRS, compared to $m = 402$ of the original SPX.

The summation of the noise and the model covariance is illustrated in Fig. 7. The propagation of the covariance of the system model (q) is not LP. The quadratic summation rule (Eq. 27), or Gauss approximation formula [16], is therefore not applicable. Nevertheless, the low scent ζ (Fig. 5, left) suggests that both propagations are close to LP. The summation error (ε) is hence finite, but quite small. It differs qualitatively from both contributions, indicating that the summation is non-trivial.

Finally, the samples of one pole of the derived ensembles are compared to the reference samples of the REF in Fig. 8. The limit ($|z| = 1$) of stability is included to illustrate how close the samples are to be physically forbidden. The construction of the different ensembles is apparent, even though the transformation $T = U^T S U S^{-1}$ distorts the scatter plots (sections 3.2, 5.2), and tilts the principal axes (lines). The samples of the REF are almost evenly distributed. Only four samples of the STD, labelled p_1, $p_{2,_{\wedge}}$ p_3, p_4, deviate significantly from a dense central cluster, as described by the excitation matrix V_{STD} (Eq. 30). It also is evident that SPX originates from half the STD. A small translation required to achieve the correct mean is discernible, while the Gram-Schmidt orthogonalization renders a minor rotation and distortion. The BIN contains comparable variations in all samples and thus has no central cluster and its samples are repelled from the principal directions (lines). The statistical differences to the REF refer to the shape of the joint pdf. Choosing the best ensemble is thus equivalent of selecting the most appropriate pdf in RS. The BIN seems to resemble the REF scatter plot the most, as verified by its low errors in Fig. 5.

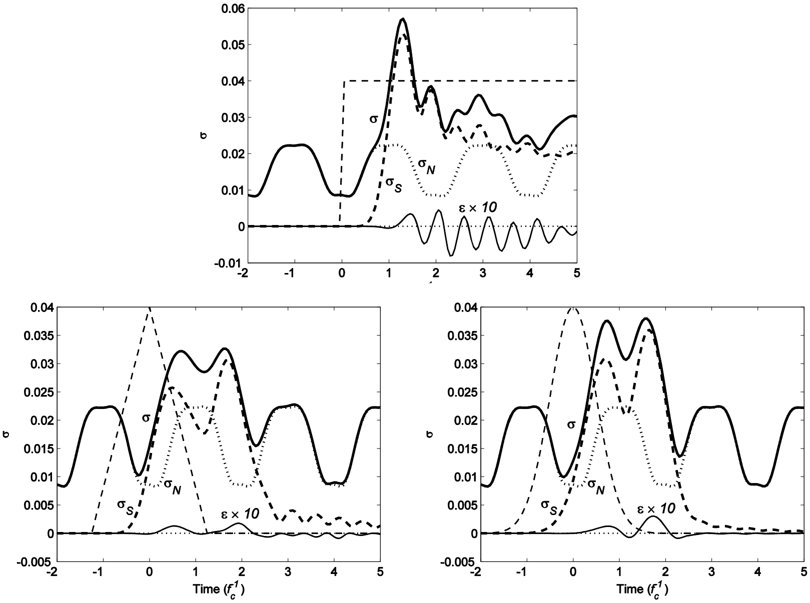

Figure 7. Summation of covariance: Total (solid), system (dashed) and noise (dotted), for the three test signals (thin, dashed), with the error (ε) of square summation (×10) (Eq. 27).

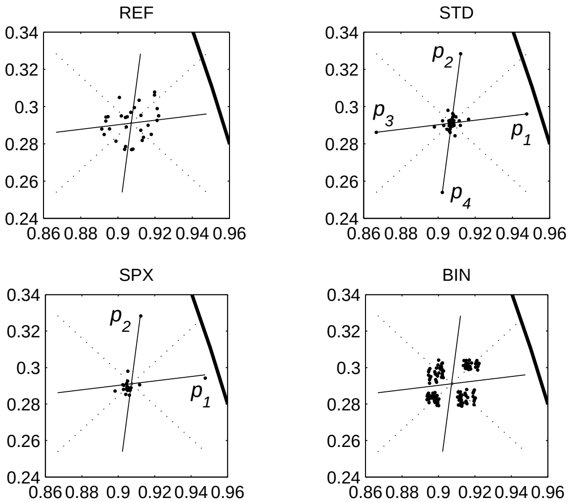

Figure 8. The different samples (dots) of the pole marked 'P' in Fig. 3, of the reference (REF), standard (STD), simplex (SPX) and binary (BIN) ensembles. The limit $|z|=1$ of stability (solid, thick) and lines connecting the primary variations p_1, p_2, p_3, p_4 of the STD as well as lines (dashed) to combined excitations of the BIN are included for reference.

7. Conclusions

Deterministic sampling remains controversial [27] while random sampling has qualified as a preferred state-of-the-art method for propagating uncertainty. Both result in finite *statistical* [17] ensembles, which are approximate finite representations of the primary statistical models. Their sampling strategies and convergence rates are dramatically different. While deterministic sampling humbly aims at representing the most relevant and best known statistical information, random sampling targets complete control of all features of the ensemble. Such detailed information is rarely known and must instead be more or less blindly assigned. The inevitable consequence is that critical computational resources are spent on propagating, at best, vaguely known details. The numerical power of modern computers is better spent on refinements of the signal processing model (longer time series, higher sampling rates, larger systems etc.). Refined methods of random sampling have therefore been proposed which either simplifies the model, or improve the sampling distributions. Compared to deterministic sampling though, their convergence rates remain low.

It is easy to confuse deterministic sampling with experimental design and optimization [28]. Even though any sample could be a possible outcome of an experiment, deterministic ensembles *represent* rather than *realize* (as random ensembles) statistical distributions. Instead of associating a joint distribution to the parameters of an uncertain model, it is possible to directly represent their statistics with a deterministic ensemble. That would eliminate the need of interpreting abstract distributions and result in complete reproducibility. The critical choice of ensemble would be assigned once and for all in the calibration experiment, with no further need of approximation.

The use of excitation matrices made it possible to construct universal generic ensembles. The efficiency of the minimal SPX ensemble is indeed high but so is also its third moment. While the STD maximizes the range of each parameter, the BIN minimizes it by varying all parameters in all samples. The STD is the simplest while the SPX is the most efficient ensemble. In the example, the BIN was most accurate. For non-parametric models with many parameters, reduction of samples may be required. Elimination of singular values (ESV) and correlated sampling (CRS) were two such techniques. The presented ensembles are not to be associated to random sampling as a method. They are nothing but a few examples of deterministic sampling, likely the best ensembles are yet to be discovered.

It is indeed challenging but also rewarding to find novel deterministic sampling strategies. Once the sampling rules are found, the application is just as simple as random sampling, but usually much more efficient. Deterministic sampling is one of very few methods capable of non-linear propagation of uncertainty through large signal processing models.

Author details

Jan Peter Hessling*

Measurement Technology, SP Technical Research Institute of Sweden, Borås, Sweden

References

[1] Kay S. Fundamentals of Statistical signal processing: Estimation Theory. New Jersey: Prentice Hall; 1993.

[2] Hessling JP. Propagation of dynamic measurement uncertainty. Meas. Sci. Technol. 2011; 22 (10) 105105 (13pp).

[3] Hessling JP. Integration of digital filters and measurements. In: Márquez FPG. (ed.) Digital Filters. Rijeka: InTech; 2011. p123-154. Available from http://www.intechopen.com/books/digital-filters/integration-of-digital-filters-and-measurements (accessed 4 July 2012).

[4] Pintelon R, Schoukens J. System Identification: A Frequency Domain Approach. Piscataway, New Jersey: IEEE Press; 2001.

[5] Witteveen JAS. Efficient and Robust Uncertainty Quantification for Computational Fluid Dynamics and Fluid-Structure Interaction. PhD thesis. Delft University of Technology; 2009.

[6] Hale PD, Dienstfrey A, Wang JCM, Williams DF, Lewandowski A, Keenan DA, Clement TS. Traceable Waveform Calibration With a Covariance-Based Uncertainty Analysis. IEEE Trans. Instrum. Meas. 2009; 58 (10) 3554-3568.

[7] Hessling JP. Metrology for non-stationary dynamic measurements. In: Sharma MK. (ed.) Advances in Measurement systems. Vukovar: InTech; 2010. p. 221-256. Available from http://www.intechopen.com/books/advances-in-measurement-systems/metrology-for-non-stationary-dynamic-measurements (accessed 4 July 2012).

[8] Metropolis N, Ulam S. The Monte Carlo Method. Journal of the American Statistical Association 1949; 44 (247) 335-341.

[9] Rubenstein RY, Kroese DP. Simulation and the Monte Carlo Method, 2nd Ed. New York: John Wiley & Sons Inc.; 2007.

[10] Hessling JP. A novel method of evaluating dynamic measurement uncertainty utilizing digital filters. Meas. Sci. Technol. 2009; 20 (5) 055106 (11pp).

[11] Råde L, Westergren B. Beta Mathematics Handbook, 2nd Ed. Lund, Sweden: Studentlitteratur; 1990.

[12] Hessling JP, Svensson T. Propagation of uncertainty by sampling on confidence boundaries, accepted for publication in *International Journal for Uncertainty Quantification*.

[13] Hessling JP. Deterministic sampling for propagating model covariance, submitted for publication.

[14] Lovett T. Polynomial Chaos Simulation of Analog and Mixed-Signal Systems: Theory, Modeling method, Application. Saarbrucken: Lambert Academic Publishing; 2010.

[15] Li H, Zhang D. Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods. Water Resources Research 2007; 43 W09409 (13 pp).

[16] ISO GUM. Guide to the Expression of Uncertainty in Measurement. Geneva: International Organisation for Standardisation; 1995.

[17] Metropolis N. The Beginning of the Monte Carlo Method. Los Alamos Science special issue 1987; 15 125-130.

[18] Helton J, Davis L. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering and System Safety 2003; 81 23-69.

[19] Björk Å. Numerical methods for least squares problems. Philadelphia: Siam; 1996.

[20] Wikipedia: http://en.wikipedia.org/wiki/Latin_square (accessed 3 July 2012):

[21] Julier S, Uhlmann J, Durrant-Whyte HF. A new approach for filtering nonlinear systems. Proc IEEE American Control Conference June 21-23 1995; 1628-1632.

[22] Julier S, Uhlmann J. Unscented filtering and nonlinear estimation. Proceeding IEEE March 2004; 92 (3) 401-422.

[23] Simon D. Optimal State Estimation: Kalman, H∞ and non-linear approaches. New Jersey: Wiley; 2006.

[24] Matlab with Signal Processing Toolbox, The Mathworks, Inc.

[25] Julier S, Uhlmann J. The scaled unscented transformation, Proceedings of the IEEE American Control Conference 8-10 May 2002; 4555-4559.

[26] Hessling JP. Non-linear propagation and summation of covariance using deterministic sampling, in preparation.

[27] Gustafsson F, Hendeby G. Some Relations between Extended and Unscented Kalman Filters. IEEE Trans. sign. proc. 2012; 60 (2) 545-555.

[28] Fischer RA. Statistical Methods, Experimental design and Scientific Inference. New York: Oxford University Press; 1990.