Deep Learning Aided Interpolation of Spatio-Temporal Nonstationary Data

Sayako Kodera*, Florian Römer*, Eduardo Pérez†*, Jan Kirchhof†, Fabian Krieg†*‡
*Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany,
†Technische Universität Ilmenau, Germany,
‡Hochschule für Technik und Wirtschaft des Saarlandes, Saarbrücken, Germany

***Abstract*—Despite the growing interest in many fields, spatio-temporal (ST) interpolation remains challenging. Given ST nonstationary data distributed sparsely and irregularly over space, our objective is to obtain an equidistant representation of the region of interest (ROI). For this reason, an equidistant grid is defined within the ROI, where the available time series data are arranged, and the time series of the unobserved points are interpolated. Aiming to maintain the interpretability of the whole process while offering flexibility and fast execution, this work presents a ST interpolation framework which combines a statistical technique with deep learning. Our framework is generic and not confined to a specific application, which also provides the prediction confidence. To evaluate its validity, this framework is applied to ultrasound nondestructive testing (UT) data as an example. After the training with synthetic UT data sets, our framework is shown to yield accurate predictions when applied to measured UT data.**

***Index Terms*—Spatio-temporal interpolation, Geostatistics, Kriging, Deep learning, Ultrasound NDT, Manual measurements**

I. Introduction

There has been growing interest to analyze and predict spatio-temporal (ST) data in many fields and applications, such as meteorology [1], earth science [2], medical imaging [3], among others. Although different types of ST processes exist, this study concerns the ST processes which are nonstationary in both space and time with non-separable covariance functions.

One example for such processes is ultrasonic testing (UT). UT is a nondestructive evaluation (NDE) method to localize the flaws within test objects by insonifying the objects under test. Although the test object may be time invariant, UT data exhibit dynamic temporal behavior and thus can be regarded as ST nonstationary. Furthermore, their temporal behavior depends on the time of flight between the sensor and scatterers, making UT data non-separable in ST domain.

This work was supported by the Fraunhofer Internal Programs under Grant No. Attract 025-601128. E. Pérez and F. Krieg are funded by DFG under the project "CoSMaDu".

To improve the inspection quality of manual UT, an assistance system called SmartInspect is being developed [4]. With SmartInspect, the inspection area is discretized into equidistant grid points. The measured signals, each of which is a time series, are then assigned to their nearest grid points and subsequently postprocessed. Since the data is collected manually, it is sparsely and irregularly sampled in space, impairing the reconstruction quality. As a countermeasure interpolation has been shown to be beneficial for postprocessing [5].

In order to incorporate an interpolation process into SmartInspect, the following requirements should be satisfied. It must be (a) capable of capturing dynamic temporal behavior, (b) robust against sparse and irregular sampling and (c) fast enough for on-site execution. Moreover, it is preferable to maintain the interpretability of the whole process and to be able to quantify the prediction confidence.

However, handling ST nonstationary and non-separable data still remains a challenging task [1, 6]. Kriging, a classical interpolation technique originating from the geostatistical field, is essentially a minimum mean squared error (MMSE) predictor. With the proper statistical modeling, Kriging is well known to be the best linear unbiased predictor. As a byproduct it can also provide the prediction confidence. Yet, to properly capture the ST statistics we should take into account the joint dependency of the space and time domains. This easily leads to large dimensionality of the sample covariance matrix, often inhibiting the use of Kriging for ST data [7].

Our previous work demonstrates that an autoencoder model can predict unknown UT signals from randomly sampled measurements in their neighborhood [5]. Yet, this method is sensitive to the change in measurement settings such as the measurement duration or the sampling period. Moreover, as it is based solely on DNNs, it lacks the interpretability, which may limit its application. For interpolating irregularly sampled meteorological data, the authors of [1] propose to decompose the ST data into temporally referenced basis functions and spatial

coefficients. While the temporal basis is determined via principal component analysis (PCA), the spatial coefficients are estimated using DNNs, preserving the interpretability. However, it still requires *a priori* knowledge of the model order, which is unrealistic for many applications.

Another approach for ST decomposition is to transform ST data into other domains [6]. By transforming it into space-frequency (SF) domain, the temporal dependency of ST data can be decorrelated. This results in a significant reduction in the dimensionality of the sample covariance matrix, making Kriging feasible. By utilizing the knowledge of the distribution of each Fourier coefficient, the spatial statistics in SF domain can be estimated parametrically [8]. Nevertheless, in many scenarios such *a priori* knowledge may not be available, or the distribution of the signal Fourier coefficients cannot be well modeled.

Motivated by the work of [7], in this study we propose the Deep Learning aided SF Kriging (DLSFK), where full length time series data are predicted based on their neighborhood measurements. With DLSFK, interpolation is performed via MMSE prediction in SF domain [7], whereas the spatial statistics of SF domain are estimated non-parametrically using a DNN to accommodate their frequency-dependent variability. Our approach is similar to [1], where we use Fourier transform instead of PCA as temporal basis functions.

DLSFK has the following advantages. First, it does not require any knowledge regarding the distribution of the signal Foruier coefficients, making it generic and applicable to various signals. Second, the same network can be used for all frequencies, providing the flexibility for the measurement setup. Third, as a class of DNN-aided inference [9], DLSFK preserves the interpretability of the interpolation process. Fourth, the prediction confidence can be estimated. Lastly, the network can be trained with a synthetic data set.

II. Spatio-Temporal Data and Process Modeling in Space-Frequency Domain

A. Spatio-Temporal Data

In this work, we consider a ST process which is observed within a spatial region of interest (ROI) $D_s \subset \mathbb{R}^d$ and an observational time $D_t \subset \mathbb{R}_{0+}$. Although the actual ST process may be continuous in space and time, we assume that the positional and temporal information associated to the observations is limited due to a finite precision of measuring instruments, such as an optical tracking system or a signal recorder. This follows an assumption that D_s and D_t are finite sets of equidistant points in $\mathbb{R}^d$ and $\mathbb{R}_{0+}$, respectively, i.e. regular lattices. The data collected on lattices are denoted as lattice data [6].

A time series data for $t \in D_t$ at a position $\boldsymbol{s} \in D_s$ is modeled as a sum of bandlimited signals and measurement noise as

$$\begin{aligned} x(\boldsymbol{s},t) &= \sum_{k=1}^{K} a_k(\boldsymbol{s},t) + n(\boldsymbol{s},t) \\ &= a(\boldsymbol{s},t) + n(\boldsymbol{s},t) \end{aligned} \tag{1}$$

Here, $a(\cdot)$ denote the signal of interest, and $n(\cdot)$ is regarded as Gaussian noise with a finite variance, independent of $a(\cdot)$. The quantity K is the signal model order, which may or may not be known. In SF domain, the Fourier coefficients of (1) are expressed as

$$\begin{aligned} X(\boldsymbol{s},\omega) &:= \int_{D_t} x(\boldsymbol{s},t) e^{-j\omega t} \\ &= A(\boldsymbol{s},\omega) + N(\boldsymbol{s},\omega), \end{aligned} \tag{2}$$

where $\omega = 2\pi f$, and $A(\cdot)$ and $N(\cdot)$ are the SF spectra of the signal and the noise, respectively.

Based on these settings, the objective of this work is set as follows. Suppose D_s consists of L equidistant points over a region in $\mathbb{R}^d$. Within the same region, several measurements are taken at irregularly spaced locations. The measurement positions are available with a limited precision, which are arranged onto the nearest points of D_s and denoted as a set of distinct positions $S_N = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \dots \boldsymbol{s}_N\} \subset D_s$. At each position of S_N a time series data of length M is recorded, resulting in the collection of measurements represented as a matrix $\boldsymbol{X}_N = [\boldsymbol{x}_1\, \boldsymbol{x}_2 \cdots \boldsymbol{x}_N] \in \mathbb{R}^{M\times N}$. Given the data set $\boldsymbol{X}_N$ and the measurement positions S_N, our goal is to predict the missing $(L-N)$ time series data $\boldsymbol{X}_{L-N} \in \mathbb{R}^{M\times(L-N)}$.

B. Spatial Statistics Based Process Modeling

For the sake of simplicity, in this section we consider a scenario where an unknown time series data of the position $\boldsymbol{s}_0 \in \{D_s \setminus S_N\}$ is to be predicted based on the observations in its neighborhood. Instead of predicting the time series data directly, prediction is carried out in SF domain, since Fourier coefficients of different frequencies are uncorrelated and can be predicted separately [7].

Although (2) is deterministic, our information regarding its behavior within D_s is limited due to inadequate sample size. A common strategy to account for those uncertainties is to model such a process as a random process [6]. Since it is complex-valued, in this work we call the process model of (2) a complex-valued spatial random field (CSRF), which is denoted as $\{\tilde{X}(\boldsymbol{s},\omega) : \boldsymbol{s} \in D_s\}$. With the CSRF process modeling, measurements are generally considered as a single realization of the process, denoted with small letters as $\tilde{x}(\boldsymbol{s},\omega)$ in this work.

Let us consider the prediction of the Fourier coefficient of a frequency $\omega_m = 2\pi \frac{m}{M} f_S$ with M and f_S being the number of frequency bins

and the sampling frequency, respectively. In general, a linear prediction of the Fourier coefficient $\tilde{X}_m(\boldsymbol{s}_0) \coloneqq \tilde{X}(\boldsymbol{s}_0, \omega_m)$ can be expressed as

$$\hat{\tilde{X}}_m(\boldsymbol{s}_0) = \boldsymbol{w}_m^{\mathrm{T}}(\boldsymbol{s}_0) \cdot \tilde{\boldsymbol{x}}_m, \tag{3}$$

where $\boldsymbol{w}_m \in \mathbb{C}^N$ is the weight vector, and $\tilde{\boldsymbol{x}}_m = [\tilde{x}_m(\boldsymbol{s}_1), \tilde{x}_m(\boldsymbol{s}_2), \ldots, \tilde{x}_m(\boldsymbol{s}_N)]^{\mathrm{T}} \in \mathbb{C}^N$ is a vector representation of the Fourier coefficients of the available samples. To achieve a linear MMSE predictor, the optimal weights of (3) can be attained using *frequency variograms* (FVs) [7]. FVs are based on the assumption of intrinsic stationarity (IS). Let the spatial incremental process of the CSRF be $\tilde{Y}^{[\boldsymbol{h}]}(\boldsymbol{s}, \omega) = \tilde{X}(\boldsymbol{s}, \omega) - \tilde{X}(\boldsymbol{s} + \boldsymbol{h}, \omega)$, with $\boldsymbol{h} \in \mathbb{R}^d$ being a spatial separation, typically called a lag. Under the assumption of IS, the incremental process $\tilde{Y}^{[\boldsymbol{h}]}(\boldsymbol{s}, \omega)$ is assumed to be zero-mean and shift invariant, allowing the actual process to be non-stationary [6].

Based on these assumptions, a FV is defined as

$$2\gamma_{\tilde{X}}(\boldsymbol{h}, \omega) \coloneqq \mathrm{Var}\left\{\tilde{Y}^{[\boldsymbol{h}]}(\boldsymbol{s}, \omega)\right\}, \tag{4}$$

which is the expected value of the periodogram of the incremental process $\tilde{Y}^{[\boldsymbol{h}]}(\boldsymbol{s}, \omega)$ [8]. Upon interpolation (3), a FV (4) for each frequency is to be estimated based on the sample spatial statistics. The FV of $\tilde{\boldsymbol{x}}_m$ is defined as [8]

$$g_{\tilde{\boldsymbol{x}}}(\boldsymbol{h}_{ij}, \omega) = \frac{1}{N(\boldsymbol{h}_{ij})} \sum_{\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}_{ij}} |\tilde{x}(\boldsymbol{s}_i, \omega) - \tilde{x}(\boldsymbol{s}_j, \omega)|^2, \tag{5}$$

where $N(\boldsymbol{h}_{ij})$ is the number of location pairs whose lag is $\boldsymbol{h}_{ij}$. In this study, we call (5) a sample FV.

III. Method: Data Driven Frequency Variogram Estimation

A. Problem Fromulation for DLSFK

As described in Section II-A, our ST data are lattice data, consisting of N vector-valued samples and $L - N$ unknown vectors to be predicted. Predictions of those time series data are carried out in SF domain. Hereafter, the ROI D_s is assumed to be small enough to ensure IS within D_s.

The whole interpolation process is depicted in Figure 1 (a). Utilizing the orthogonality of Fourier kernels, interpolation is performed individually for each frequency of interest. As shown in Figure 1 (b), the actual interpolation procedure for a single frequency is divided into three parts. First, the sample FV of the measurements $\boldsymbol{X}_N$ are computed. Second, the FV of the ROI D_s are estimated based on the sample FV. Lastly, the Fourier coefficients of the unknown time series data $\boldsymbol{X}_{L-N}$ are predicted via (3), where the weights are determined based on the estimated FV.

For accurately predict the unobserved signals, estimation of FVs plays a crucial role. With D_s being regular lattices, the available lags within D_s form a

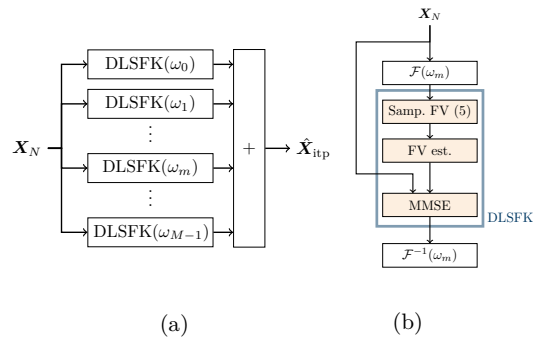

Figure 1: Block diagram of DLSFK, the proposed framework for spatio-temporal interpolation: (a) entire framework and (b) space-frequency domain interpolation for a single frequency ω_m. The prediction process (DLSFK) is highlighted in (b). Here, $\boldsymbol{X}_N \in \mathbb{R}^{M \times N}$ and $\hat{\boldsymbol{X}}_{\mathrm{itp}} \in \mathbb{R}^{M \times L}$ are the measurement data, i.e. samples, and the interpolated data, respectively. The notation $\mathcal{F}$ and $\mathcal{F}^{-1}$ respectively represent the forward and the inverse Fourier transform.

finite set with N_h known elements. As a result, for each frequency we only need the FV values for these particular lags. This implies that we do not need to estimate a continuous function (4). Instead, only the estimate of a vector $\boldsymbol{\gamma}_{\tilde{X}} \in \mathbb{R}_{0+}^{N_h}$ is sufficient.

Understanding the implication of FVs is also helpful to formulate the estimation problem. In the same manner as the conventional variograms, the structure of FVs represent the underlying spatial statistics of SF domain. Specifically, if the values of the ROI evolve over space, i.e. spatially correlated, the corresponding FV gradually increases with lags. Contrarily, a FV converges very quickly, if the values are not correlated. Furthermore, when a frequency ω_m is within the signal spectra, its Fourier coefficients over D_s are highly spatially correlated, whereas those of noise are not. This suggests that the structure of a FV varies depending on the frequency.

Based on these considerations, we formulate the FV estimation problem as a regression problem where the structure of a FV is modeled for each frequency. To accommodate their frequency dependence, the regression is performed nonparametrically using a DNN. This allows us to estimate FVs without restricting the signal type or requiring *a priori* knowledge about the distribution of the signal Fourier coefficients.

B. Network Architecture and Training for UT Application

For estimating a FV of a single frequency ω_m, a DNN is designed, such that the same network can be applied to all frequencies. As the network inputs, two attributes are selected. One is the distribution of the available lags within the samples as a probability $\boldsymbol{p}(h|S_N) \in \mathbb{R}_{[0,1]}^{N_h}$. The other one is

the sample FV of ω_m (5) which is smoothed over distance and normalized with its maximal value. Considering that neighboring bins are similar in their FV structures, the FVs of two neighboring bins, namely ω_{m-1} and ω_{m+1}, are also fed into the network to provide further insights and reduce the effect of the noise. The network outputs an estimate of the normalized FV $\hat{\boldsymbol{\eta}} \in \mathbb{R}_{0+}^{N_h}$, which is subsequently denormalized. The selected network is composed of a single 1D convolutional layer followed by four dense layers. However, it has to be highlighted that other architectures would also be suitable for the proposed framework.

The training data are based solely on synthesized data. Aligned to our scenario, UT data are modeled as a sum of clean UT signals and measurement noise. The clean UT signals are generated using a delay-and-sum based forward model [5]. The noise is modeled as additive white Gaussian noise, whose SNR is selected based on our actual UT measurements and set to 20 dB.

We first generate the groundtruth signals $\boldsymbol{X}_L \in \mathbb{R}^{M \times L}$ which are computed for all L grid points in D_s. This follows random selection of the sampling positions S_N from L points in D_S, and the measurements $\boldsymbol{X}_N \in \mathbb{R}^{M \times N}$ are subsampled from $\boldsymbol{X}_L$ accordingly. The network is trained in a supervised manner, where the network aims to minimize the mean squared error between the normalized FV of the groundtruth $\boldsymbol{\eta} \in \mathbb{R}_{0+}^{N_h}$ and the network output $\hat{\boldsymbol{\eta}}$. To diversify the training data set, we vary the number of the scatterers, i.e. K in (1), and their positions, the center frequency and the bandwidth of the UT pulse, and the number of samples N and their spatial distribution.

IV. Numerical Results: Spatiotemporal Batch Interpolation

The performance of DLSFK is numerically evaluated using a set of UT measurements as an example. The measurements are obtained with an automated UT measurement system. To mimic the sampling density of manual inspections, this data set is spatially subsampled, and the missing data are interpolated afterwards.

A. Measurement Setup

The measurements are taken in contact mode, where a single transducer is placed directly on the object surface, transmitting an ultrasound pulse and receiving the echoed signals. The test object is made of steel, which has a flat surface and is configured with several flat bottom holes. These holes are regarded as flaws in the specimen. Scan positions are equally spaced with dx = dy = 0.5 mm, which corresponds to 8.7 samples per squared wavelength. As the region of interest (ROI), a batch of 10 dx × 10 dy is chosen. The depth of the ROI is also limited to the range between 69.62 mm and 88.5 mm, where we see the echoes caused by the flaws. In the selected batch, 15 samples are randomly selected as illustrated in Figure 2 (b), which is equivalent to 1.3 samples per squared wavelength. The values of the measurement parameters are summarized in Table I.

Table I: Measurement Parameter Values

	Parameter	Value/range
Specimen	Material	Steel
	Dimension (L × D × H)	200 × 140 × 90 [mm]
	ROI ($N_x \times N_y \times N_z$)	10 × 10 × 512
UT Probe	Diameter	10 mm
	Center frequency	≈ 4.0 MHz
	Bandwidth (−6 dB)	≈ 2.6 MHz
Measurements	Wavelength λ	≈ 1.475 mm
	Spatial spacing	0.5 mm
	Sampling frequency	80 MHz

For this study, two interpolation techniques are compared: inverse distance weighting (IDW) as a benchmark and the proposed method, DLSFK. IDW computes predictions by weighting the available samples inversely proportional to the distance and is widely used for spatial interpolation tasks.

For interpolation, two more parameters are introduced: the maximal neighboring range and the minimal number of samples. The maximal neighboring range sets the boundary from a prediction point, such that the distanced samples are excluded for its prediction. This serves to maintain the prediction accuracy and is set to $\frac{\sqrt{2}}{2} N_x \,\mathrm{dx} = \frac{\sqrt{2}}{2} N_y \,\mathrm{dy} \approx 3.536\,\mathrm{mm}$, which is the longest distance within a batch from its center. The minimal number of samples is the lower bound of samples to initiate FV estimation. Contrary to IDW which can compute a prediction from only one sample, DLSFK requires at least two samples within the neighboring range for FV estimation. We empirically select five as minimal.

Note that a numerical comparison to other state of the art methods is not feasible for our data set, since [5] requires more samples than our scenario and [1] requires *a priori* knowledge of the model order (which would render the comparison unfair). Moreover, the approach of [8] is not applicable, as a parametric model of the Fourier coefficients is required. The authors' assumption of Normality should be verified for the application.

B. Interpolation Results

Figure 2 and 3 show the obtained results in C-Scan and A-Scan representations. Here, C-Scans, which are computed by taking the maximum absolute values along the temporal axis, are intended to demonstrate the capability of capturing the overall trend. On the other hand, the prediction accuracy can be assessed by comparing the A-Scans with the reference data.

In terms of both capturing the trend and predicting missing measurements, the proposed method

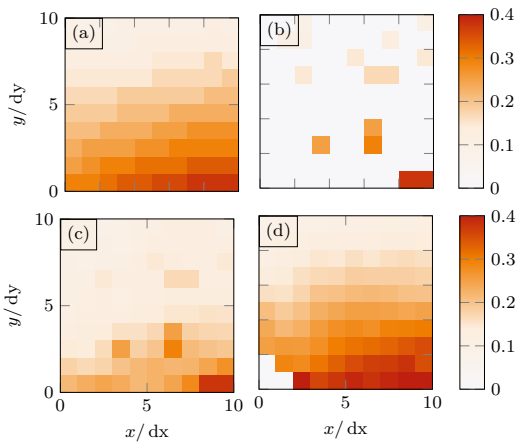

Figure 2: Spatio-temporal interpolation results of a 10×10 batch in C-Scan view. The coverage of the samples is 15%, which is equivalent to 1.3 samples per squred wavelength. The ground truth data are subsampled and interpolated. (a) ground truth, (b) subsampled data, (c) IDW interpolation and (d) interpolation results of the proposed method (DLSFK). The white areas in (b) represents the missing measurements, whereas those in (d) are the regions where no predictions can be made due to the lack of neighboring scans.

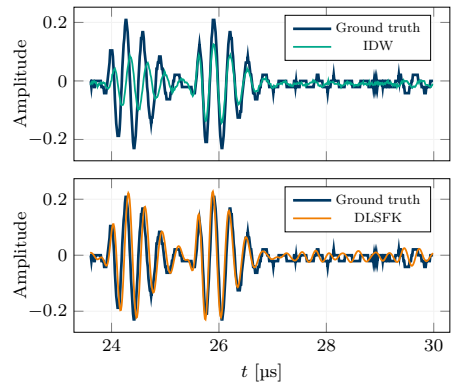

Figure 3: Predictions of a missing A-Scan at $(x, y) = (0\,\mathrm{dx}, 2\,\mathrm{dy})$ via IDW (top) and the proposed method (bottom).

outperforms IDW. IDW tends to average out over the available samples, generating very similar results for two separated prediction positions. This confirms its sensitivity to irregular sampling, which is however inevitable in our UT scenario and many other applications. On the contrary, DLSFK demonstrates its robustness against clustering and irregular sampling. Even for the position where the variance is high due to a lack of samples, their predictions are well representative compared to that of IDW.

There are, yet, some points where no prediction can be made due to a lack of samples within their neighboring range. These points are whitened in Figure 2 (c). In manual UT, such information along with the estimated prediction confidence could provide a feedback to the operator, indicating which region requires more measurements to increase the inspection quality.

V. Summary

In this work, DLSFK, a deep learning aided interpolation framework for spatio-temporal nonstationary data, is presented. Within DLSFK, interpolation is performed in space-frequency (SF) domain via the MMSE predictor, whereas the spatial statistics in SF domain are estimated non-parametrically using a DNN. This yields a fast, interpretable interpolator that is agnostic to measurement parameters such as measurement duration or the sampling period. To evaluate its validity, ultrasonic testing (UT) data are employed as an example. After the training with synthetic UT data, DLSFK is applied to actual UT measurements which are sparsely and irregularly subsampled. DLSFK outperforms the widely used inverse distance weighting, yielding accurate predictions and exhibiting greater robustness to heavy subsampling.

References

[1] F. Amato, F. Guignard, S. Robert, and M. Kanevski. A novel framework for spatio-temporal prediction of environmental data using deep learning. *Scientific Reports*, 10, 2020.

[2] H. Liu, Z. Liu, S. Liu, Y. Liu, J. Bin, F. Shi, and H. Dong. A Nonlinear Regression Application via Machine Learning Techniques for Geomagnetic Data Reconstruction Processing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):128–140, 2019.

[3] Y. Guo, L. Bi, E. Ahn, D. Feng, Q. Wang, and J. Kim. A Spatiotemporal Volumetric Interpolation Network for 4D Dynamic Medical Image. In *Proc. CVPR'20*, 2020.

[4] B. Valeske, A. Osman, F. Römer, and R. Tschuncky. Next Generation NDE Sensor Systems as IIoT Elements of Industry 4.0. *Research in Nondest. Eval.*, 31(5-6), 2020.

[5] R. Pandey, J. Kirchhof, F. Krieg, E. Pérez, and F. Römer. Preprocessing of Freehand Ultrasound Synthetic Aperture Measurements using DNN. In *Proc. EUSIPCO'21*, 2021.

[6] N. A. C. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. WILEY, 2011.

[7] T. S. Rao and G. Terdik. A New Covariance Function and Spatio-Temporal Prediction (Kriging) for A Stationary Spatio-Temporal Random Process. *Journal of Time Series Analysis*, 38(6), 2017.

[8] T. S. Rao and G. Terdik. On the Frequency Variogram and on Frequency Domain Methods for the Analysis of Spatio-Temporal Data. *Journal of Time Series Analysis*, 38(2), 2017.

[9] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis. Model-Based Deep Learning: Key Approaches and Design Guidelines. In *2021 IEEE Data Science and Learning Workshop (DSLW)*, 2021.