TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Data-driven system identification of thermal systems using machine learning

Citation for published version (APA):
Nechita, S. C., Tóth, R., & van Berkel, K. (2021). Data-driven system identification of thermal systems using machine learning. *IFAC-PapersOnLine*, *54*(7), 162-167. https://doi.org/10.1016/j.ifacol.2021.08.352

Document license:
CC BY-NC-ND

DOI:
10.1016/j.ifacol.2021.08.352

Document status and date:
Published: 01/07/2021

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 09. Oct. 2026

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 54-7 (2021) 162–167

Data-driven System Identification of Thermal Systems using Machine Learning

Ștefan-Cristian Nechita * **Roland Tóth** *,** **Koos van Berkel** ***

* *Department of Electrical Engineering Eindhoven University of Technology, Eindhoven University of Technology, Eindhoven, The Netherlands (e-mail: { s.c.nechita, r.toth}@tue.nl).*

** *Systems and Control Laboratory, Institute for Computer Science and Control, Kende u. 13-17, H-1111 Budapest, Hungary.*

*** *Research Department, ASML Netherlands B.V., 5504 DR Veldhoven, The Netherlands (e-mail: koos.van.berkel@asml.com).*

Abstract: The paper addresses the identification of spatial-temporal mirror surface deformations as a result of laser-based heat load within the lithography process of integrated circuit production. The thermal diffusion and surface deformation are modeled by separation of the spatial-temporal effects using data-driven orthogonal decomposition. A novel *tree adjoining grammar* (TAG) and sparsity enhanced symbolic-regression-based learning methods are deployed to discover temporal dynamics that connect the spatial variation. The resulting data-driven procedure is applied to automatically synthetise a compact model representation of synthetic thermal effects induced mirror surface deformations.

Copyright © 2021 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0)

Keywords: Spatial-temporal System Identification, Separation of Variables, Machine Learning, Genetic Programming, MIMO System Indentification, Tree Adjoining Grammar, Equation Discovery, Gaussian Proccesses.

1. INTRODUCTION

In *extreme ultraviolet lithography* (EUVL), a set of guiding mirrors align a laser beam to print chip patterns over silicon wafers. These mirrors are exposed to significant thermal loads and build up thermal-induced surface deformation over time. If not corrected for, these deformations affect the laser beam alignment and can lead to errors in the printed features. In order to perform correcting actions upon the mirrors, the heat-induced deformation should be precisely modeled. As first principle modeling has proven to be ineffective in describing the process due to the modalities of each physical system, an identified spatial-temporal model of the process is required. As such, the goal of the present paper is to propose an automated data-driven strategy based on known numeric and symbolic regression methods that captures the behavior of a thermal mechanical deformation system.

The heat-induced deformation model can be seen as a coupled *partial differential equation* (PDE) model, where the first PDE describes the heat-flux to temperature diffusion and the second PDE describes the temperature diffusion to surface deformations. In the literature, several distributed-parameter system identification methods have been developed (e.g. see (Li and Qi, 2011)). By Following the separation of variables principle, the large scale spatial-temporal identification problem is often split into a spatial model reduction problem and a temporal system identification problem. The spatial model reduction problem is thoroughly discussed in Boyd (2000) where the spectral Galerkin method is proposed to transform the dense numerical representation of the system into a smaller, approximated form that can be used for computer-aided simulations. Using this technique, the spatial distribution of heat, temperature and deformation signals are described through a set of *spatial basis functions* (SBFs). The remaining temporal system identification problem is seen as the identification of two *multi-input multi-output* (MIMO) temporal models.

A main challenge posed by identifying the above-mentioned MIMO system is the unknown dynamic structure that generates the data. There are two main approaches to learn these structures from data. The first one, known as non-parametric identification, proposes candidate models under a flexible function approximation approach. These methods usually utilize large, fine tuned, dynamic structures that can offer reliable candidate models. Some downsides of these approaches are the existence of a large number of interconnected parameters as in *artificial neuron network* (ANN - (Goodfellow et al., 2016)) or the fact that the entire model is constructed on and grows with acquired data, as in Gaussian Process based kernel methods (Chiuso and Pillonetto, 2019). These non-parametric methods often lack the interpretability in the sense that it is difficult to extract the significant dynamic modes representing each SBF. As opposed to this, parametric learning procedures such as *tree adjoining grammar guided genetic programming* (TAG3P - (Khandelwal, 2020), (Nechita and Tóth, 2021)) and *equation qiscovery* (ED - (Simidjievski et al., 2020)) can evolve or construct a compact candidate model that yields the most dominant dynamic modes active in

⋆ This research was supported by the Ministry of Innovation and Technology NRDI Office within the framework of the Artificial Intelligence National Laboratory Program and by the Dutch Organization for Scientific Research (NWO, domain TTW, grant: 13852) which is partly funded by the Ministry of Economic Affairs of The Netherlands.

2405-8963 Copyright © 2021 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2021.08.352

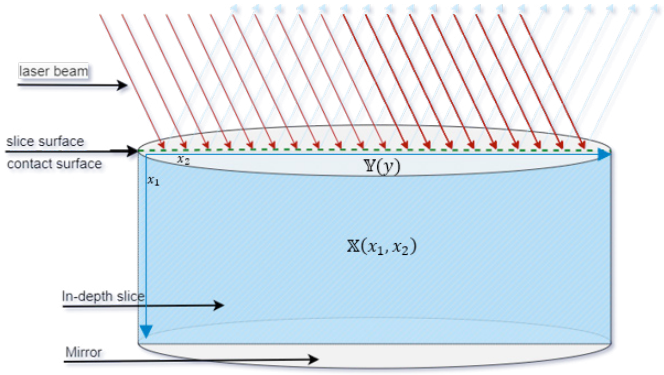

Fig. 1. Schematics of the guiding mirror. $\mathbb{X}$: in-depth slice. $\mathbb{Y}$, $\mathbb{X}_s$: surface of the in-depth slice (dotted green).

the data. Usually, the parametric solutions are described directly in the time-domain as a mathematical equation and the dynamic properties, such as input delay time or interconnection of SBFs can be directly observed in the model. Moreover the user can define the maximum complexity of the candidate model and indicate the model search space. The remainder of the paper is organized as follows. Section 2 discusses the coupled PDE modelling framework. Section 3 elaborates on the MIMO learning approaches that are applied for identifying the synthetic thermal deformation model described in Section 4. Finally the identification results are reported and discussed in Section 5.

2. PRELIMINARIES

2.1 Coupled PDE modeling framework

This section presents the first-principles-based modeling concept of the spatial-temporal heat-induced temperature diffusion and mechanical surface deformation of an in-depth 2D mirror slice. A visual representation of the mirror is indicated by Figure 1. The 2D $\mathbb{X} = [0, \mathbb{L}_{x_1}] \times [0, \mathbb{L}_{x_2}]$ space, with the sample grid sizes Δx_1, Δx_2, represents the in-depth 2D mirror slice and the $\mathbb{Y} = [0, \mathbb{L}_y]$ space, with the sample grid size Δy, represents the 1D surface side of the in-depth slice of the mirror. Since these spaces represent the same physical object, $\mathbb{L}_y = \mathbb{L}_{x_2}$ but $\Delta y \neq \Delta x_2$. Consider the space $\mathbb{X}_s \subset \mathbb{X}$, $\mathbb{X}_s = [0, \mathbb{L}_{x_2}]$ as the subspace that represents the surface of the in-depth slice. The laser generated heat flux $\mathcal{Q}(\mathbb{X}, k)$ is the spatial-temporal signal that excites the model via space $\mathbb{X}$. The heat flux $\mathcal{Q}(\mathbb{X}, k)$ generates a 2D in-depth diffusion of temperature denoted as a spatial-temporal signal $\mathcal{T}(\mathbb{X}, k)$. The mechanical surface deformation $\mathcal{D}(\mathbb{Y}, k)$ is a direct result of the surface temperature $\mathcal{T}(\mathbb{X}_s, k)$ and inner temperature diffusion. The heat-flux, temperature and deformation signals are considered to be measurable over the $\mathbb{X}$ and $\mathbb{Y}$ spaces at the grid points delivered by by Δx_1, Δx_2 and Δy. In practice, the grid sampling between signals can be different due to the different used sensors (e.g. infrared camera for temperature diffusion and laser Doppler vibrometer for deformation). The heat flux - deformation phenomenon can be written in the form:

$$\mathcal{D}(\mathbb{Y}, k) = F\left(\mathrm{q}, \mathcal{Q}(\mathbb{X}, k)\right), \tag{1}$$

where q is the forward time-shift operator, k is the time-sample and $F(\cdot)$ described a discrete time PDE, i.e it represents a function in $\mathcal{Q}(\mathbb{X}, k)$ and its shifted ...xxxxx... $\mathrm{q}^{-i}\mathcal{Q}(\mathbb{X}, k)$. The deformation is often considered to be a direct result of the temperature distribution. Moreover, temperature diffusion is a result of the heat-flux, thus (1) can be described by

$$\mathcal{T}(\mathbb{X}, k) = F_{\text{in}}\left(q, \mathcal{Q}(\mathbb{X}, k)\right), \tag{2a}$$
$$\mathcal{T}(\mathbb{Y}, k) = F_{\text{inter}}\left(\mathcal{T}(\mathbb{X}_s, k)\right), \tag{2b}$$
$$\mathcal{D}(\mathbb{Y}, k) = F_{\text{out}}\left(q, \mathcal{T}(\mathbb{Y}, k)\right), \tag{2c}$$

where F_{in} and F_{out} are PDEs, F_{inter} is an interpolation function (further described in Section 4.1).

2.2 Temporal modeling framework

+Usually, the signals described through PDEs are considered to be infinite dimensional, due to the spatial component. A reduced order approximation of the signals, done via the *separation of variables principle* (see Chapter 3 in (Boyd, 2000)). In short, the separation of variable principle decomposes the complex spatial-temporal signal in two components: temporal coefficients and spatial variation captured by SBFs. Consider the discrete-space discrete-time heat-flux $\mathcal{Q}(\mathbb{X}, k)$, temperature $\mathcal{T}(\mathbb{X}, k)$, $\mathcal{T}(\mathbb{Y}, k)$ and deformation $\mathcal{D}(\mathbb{Y}, k)$ signals and their respective collections of SBFs $\Psi_{\mathcal{Q}}(\mathbb{X})$, $\Psi_{\mathcal{T}}(\mathbb{X})$, $\Phi_{\mathcal{T}}(\mathbb{Y})$ and $\Phi_{\mathcal{D}}(\mathbb{Y})$. Also applying truncation to finite expansions by the Galerkin method, the spatial-temporal signals can be described as

$$\begin{aligned}\mathcal{Q}(\mathbb{X}, k) &\cong \sum_{i=1}^{r_{\mathcal{Q}}} q_i(k)\psi_{\mathcal{Q},i}(\mathbb{X}), \mathcal{T}(\mathbb{X}, k) \cong \sum_{i=1}^{r_{\mathcal{T}}} a_i(k)\psi_{\mathcal{T},i}(\mathbb{X}),\\ \mathcal{T}(\mathbb{Y}, k) &\cong \sum_{i=1}^{r_{\mathcal{T}}} t_i(k)\varphi_{\mathcal{T},i}(\mathbb{Y}), \mathcal{D}(\mathbb{Y}, k) \cong \sum_{i=1}^{r_{\mathcal{D}}} b_i(k)\varphi_{\mathcal{D},i}(\mathbb{Y}),\end{aligned} \tag{3}$$

Collect

$$\begin{aligned}Q(k) &= [q_i(k)]_{i=1}^{r_{\mathcal{Q}}}, A(k) = [a_i(k)]_{i=1}^{r_{\mathcal{T}}},\\ T(k) &= [t_i(k)]_{i=1}^{r_{\mathcal{T}}}, B(k) = [b_i(k)]_{i=1}^{r_{\mathcal{D}}}\end{aligned} \tag{4}$$

and

$$\begin{aligned}\Psi_{\mathcal{Q}}(\mathbb{X}) &= [\psi_{\mathcal{Q},i}(\mathbb{X})]_{i=1}^{\top r_{\mathcal{Q}}} \quad \Psi_{\mathcal{T}}(\mathbb{X}) = [\psi_{\mathcal{T},i}(\mathbb{X})]_{i=1}^{\top r_{\mathcal{T}}}\\ \Phi_{\mathcal{T}}(\mathbb{Y}) &= [\varphi_{\mathcal{T},1}(\mathbb{Y})]_{i=1}^{\top r_{\mathcal{T}}} \quad \Phi_{\mathcal{D}}(\mathbb{Y}) = [\varphi_{\mathcal{D},1}(\mathbb{Y})]_{i=1}^{\top r_{\mathcal{D}}}\end{aligned} \tag{5}$$

Then, compactly we can write (2a)-(2c) as

$$\begin{aligned}\mathcal{T}(\mathbb{X}, k) &\cong A(k)\Psi_{\mathcal{T}}(\mathbb{X}),\\ \mathcal{D}(\mathbb{Y}, k) &\cong B(k)\Phi_{\mathcal{D}}(\mathbb{Y}).\end{aligned} \tag{6}$$

Following the above approximation, the temporal evolution of the spatial-temporal signals $\mathcal{T}(\mathbb{X}, k)$ and $\mathcal{D}(\mathbb{Y}, k)$ is represented solely by the evolution of the temporal coefficients sets $A(k)$ and $B(k)$. Further, we can assume that the temporal evolution of the sets $A(k)$ and $B(k)$ is described by two temporal dynamic models such as:

$$A(k) = \mathcal{F}_{\text{in}}(\{A(k-i)\}_{i=1}^{n_a}, \{Q(k-i)\}_{i=1}^{n_q}), \tag{7}$$
$$B(k) = \mathcal{F}_{\text{out}}(\{B(k-i)\}_{i=1}^{n_b}, \{T(k-i)\}_{i=1}^{n_t}). \tag{8}$$

with n_a, n_q, n_t and n_b are finite time-shift values and $A(k) = A_{\text{m}}(k) + \Xi_a(k)$, $T(k) = T_{\text{m}}(k) + \Xi_t(k)$ and $B(k) = B_{\text{m}}(k) + \Xi_b(k)$ are measured temporal expansion coefficients of $\mathcal{T}(\mathbb{X}, k)$,$\mathcal{T}(\mathbb{Y}, k)$ and $\mathcal{D}(\mathbb{Y}, k)$ signals respectively. Thus, the heat-flux - temperature - deformation spatial-temporal phenomenon is described by two temporal (MIMO) dynamic models in a serial connection.

2.3 Spatial basis functions

Within the Galerkin method, the spatial distribution of the signals is represented through a set of SBFs. These SBFs can be a priori constructed based on analytical expresions or computed based on data. Furthermore we describe two methods to construct the SBFs.

Cosine Fourier basis Due to the finite 1D space $\mathbb{V}$ in terms of the interval $[0, \mathbb{L}]$, the orthonormal cosine Fourier SBF set $\Phi_F(\mathbb{V}) = \{\varphi_{\mathbb{V}}\}_{i=1}^{\infty}$ is generated by

$$\varphi_{\mathbb{V}}(x) = \sqrt{\frac{2}{\mathbb{L}}} \cos\left(\frac{i\pi x}{\mathbb{L}}\right), \tag{9}$$

where $x \in \mathbb{V}$. Due to the finite 2D space $\mathbb{W} = [x_1 \times x_2]$ in terms of the interval $[0, \mathbb{L}_1] \times [0, \mathbb{L}_2]$, the orthonormal cosine Fourier SBF set $\Phi_F(\mathbb{W}) = \{\varphi_{\mathbb{W}_{i,j}}\}_{i=1,j=1}^{\infty,\infty}$ is generated by convolution of 1D cosine basis:

$$\varphi_{\mathbb{W}_{i,j}}(x_1, x_2) = \frac{2}{\sqrt{\mathbb{L}_1 \mathbb{L}_2}} \cos\left(\frac{i\pi x_1}{\mathbb{L}_1}\right) \cos\left(\frac{j\pi x_2}{\mathbb{L}_2}\right). \tag{10}$$

Considering this convolutive nature, in order to compute a finite approximation form of the 2D spatial-temporal signal $S(\mathbb{W}, t)$, one have to select first $\mathrm{r}_{x_1}, \mathrm{r}_{x_2} \in \mathbb{N}$ 1D cosine Fourier basis that span the $[0, L_1]$ and $[0, L_2]$ spaces. The resulting $\Phi_F(\mathbb{W})$ SBF set is

$$\begin{aligned} \Phi_F(\mathbb{W}) &= \left\{\varphi_{\mathbb{W}_{i,j}}(x_1, x_2)\right\}, \\ \mathrm{Card}(\Phi_F(\mathbb{W})) &= ||\Phi_F(\mathbb{W})||_0 = \mathrm{r}_{x_1} \times \mathrm{r}_{x_2}. \end{aligned} \tag{11}$$

As a consequence, the approximation of the signal $S(\mathbb{W}, t)$ requires $\mathrm{r}_{x_1} \times \mathrm{r}_{x_2}$ temporal expansion coefficients. Since the spatial distribution of the SBFs is not optimal with respect to the signal $S(\mathbb{W}, t)$, the required reduction orders r_{x_1} and r_{x_2} are often high (see (Li and Qi, 2011)). Moreover, this considerable number of temporal coefficients implies that the dynamic relations (7) and (8) are governed by large MIMO systems. Thus, using non-optimal SBFs increases the difficulty of the system identification task. On the other hand, these SBFs are used to simulate the thermal PDE equation under various material properties and spatial distributions of the heat-flux signal.

2.4 Proper Orthogonal Decomposition (POD) basis

A data-driven alternative to find an optimal number of SBFs is by computing the *singular value decomposition* (SVD) of the "snapshot" matrix of the spatial signal. Consider the signal $S(\mathbb{W}, k)$ sampled both in time and space denoted by $\mathrm{S}(\bar{x}_i, k\mathrm{T_s})$ and a snapshot matrix $[\mathrm{W}]_{i,j} = \mathrm{S}(\bar{x}_i, j)_{i=1,j=1}^{\mathrm{M,N}}$, where $\bar{x}_i$ are the discrete space points and j are discrete time samples of period $\mathrm{T_s}$. Let $d = min(\mathrm{M}, \mathrm{N})$ then, the SVD

$$\mathrm{U}\Sigma\mathrm{V}^{\top} = \mathrm{W}, \tag{12}$$

where $\mathrm{U} \in \mathbb{R}^{\mathrm{M}\times\mathrm{M}}$, $\mathrm{V} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ and $\Sigma = \mathrm{diag}\{\sigma_i\}_{i=1}^{d} \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$ is diagonal matrix filled with the singular values of W, $\mathrm{U} = [\mathrm{u}_i]_{i=1}^{\mathrm{M}}$ and $V = [v_i]_{i=1}^{\mathrm{N}}$ are the sets of orthonormal right and left singular vectors of W. Based on the properties of SVD (see (Antoulas, 2005)) and the values of σ_i, we can select the first r columns of U or a linear combination of the first r columns of V as a set of orthonormal SBFs $\Phi_{\mathrm{POD}}(\mathbb{W})$. If there are more time samples than spatial samples($\mathrm{N} > \mathrm{M}$), then

$$\Phi_{\mathrm{POD}}(\mathbb{W}) = \{\varphi_{\mathrm{POD}_i}(\mathbb{W}) = \mathrm{u}_i\}_{i=1}^{\mathrm{r}} \tag{13}$$

else when ($\mathrm{N} \leq \mathrm{M}$)

$$\Phi_{\mathrm{POD}}(\mathbb{W}) = \{\varphi_{\mathrm{POD}_i}(\mathbb{W}) = \frac{1}{\sigma_i}\mathrm{W}\mathrm{v}_i\}_{i=1}^{\mathrm{r}}, \tag{14}$$

The amount of information captured by the i^{th} POD basis is represented by the magnitude of the i^{th} singular value. The error measure η_r [%] based on selecting the first r POD basis to represent the signal $\mathrm{S}(\bar{x}_i, k\mathrm{T_s})$ is

$$\eta_r = (1 - \frac{\gamma_r}{\gamma_d})100, \;\; \gamma_r = \sum_{i=1}^{\mathrm{r}} \sigma_i, \gamma_d \geq \gamma_{d-1} \geq ... \geq \gamma_1. \tag{15}$$

For an arbitrary value r, a reduced order signal captures η_r % amount of information of a spatial-temporal signal. Since this is an optimal data-driven method to obtain the set of SBFs, most of the signal information is captured by a compact set of POD basis under the assumption that the trajectories of the system will have also similar spatial distribution. This reduces the difficulty of the dynamic system identification task.

3. SYSTEM IDENTIFICATION PROBLEM

3.1 Heat-Temperature Identification problem

In van den Hurk et al. (2018), the authors show that the 2D thermal diffusion PDE equation can be represented as a reduced order linear-time invariant state space (LTI-SS) model by applying the Galerkin method. Thus, the LTI-SS temporal model captures and describes the temporal evolution of the thermal diffusion PDE. Following this idea, for a real heat-flux to temperature distribution set-up, the identification problem of the temporal MIMO dynamic model depicted in Equation (7) can be simplified by considering the discrete-time LTI structure (16)

$$A(k) = \mathcal{F}_{\mathrm{LTI}}(\{A(k-i)\}_{i=1}^{n_a}, \{Q(k-i)\}_{i=1}^{n_a}) + \Xi_{\mathrm{a}}(k), \tag{16}$$

where $\Xi_{\mathrm{a}}(k)$ is the measured output noise of the temporal expansion coefficients $A(k)$. Therefore the heat to temperature temporal model identification problem turns into a Output Error (OE) model identification problem described as the following minimization problem

$$\min_{\hat{a}_i(k)} \quad \frac{1}{\mathrm{r}_{\mathcal{T}}} \sum_{i=1}^{\mathrm{r}_{\mathcal{T}}} \sqrt{\frac{1}{\mathrm{N}} \sum_{k=1}^{\mathrm{N}} (a_i(k) - \hat{a}_i(k)))^2} \tag{17}$$

where $\{\hat{a}_i(k)\}_{i=1}^{\mathrm{r}_{\mathcal{T}}} = \hat{A}(k) = \hat{\mathcal{F}}_{LTI}(\{\hat{A}(k-i)\}_{i=1}^{n_a}, \{Q(k-i)\}_{i=1}^{n_a})$ is the multi-channel simulation model output. The input and output dimensions of the system (16) are determined by the arbitrary values $\mathrm{r}_{\mathcal{Q}}$ and $\mathrm{r}_{\mathcal{T}}$. To solve the minimisation problem (17) a prediction error state space method can be deployed. The `ssest()` Matlab function performs a gradient descent parameter optimization over an LTI state space structure that is a priori initialised by a subspace identification (detailed in Van Overschee and De Moor (1994)). To obtain a candidate model we used the `ssest()` function together with the `n4sid` algorithm and *canonical variate algorithm* weighting scheme for initializing the state space model. Another method to obtained a compact representation of system (16) is by evolving a candidate model through genetic programming. The TAG3P identification framework, with grammar $\mathrm{G_{LTI}}$, proposes, by automated structure selection, candidates that can capture the LTI dynamic structure of the model (16) under a compact discrete time-domain representation. For MIMO system identification problems, this method is further detailed in (Nechita and Tóth, 2021).

3.2 Temperature-Deformation Identification problem

In Section 3.3 of (van den Hurk et al., 2018) the authors show that under certain material properties and grid sampling conditions "the dynamics of the thermal diffusion process are approximately a factor 10^8 times slower than the mechanical elasticity waves". Therefore, the elastic deformation can be considered static if you view it from the time frame of the thermal dynamics. Thus the value of the deformation temporal expansion coefficient $B(k)$ at any time-sample k, is determined by the values of $T(k)$

alone. Following this idea, in order to scale down the identification problem of the reduced order temporal model depicted in Equation (8), we consider that a candidate temporal model is defined by a discrete-time static MIMO function. Thus, Equation (8) can be written as

$$B(k) = \mathcal{F}_{\mathrm{St}}(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}) + \Xi_b(k), \tag{18}$$

where $T(k) = \{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}$ and $\Xi_b(k)$ is the measured output noise of the temporal expansion coefficients $B(k)$. Hence, the evolution of $B(k)$ is not influenced by past values $\{B(k-j)\}_{j=1}^{n_b}$ and $\{T(k-j)\}_{j=1}^{n_t}$. Therefore the MIMO ($\mathrm{r}_\mathcal{T}$ to $\mathrm{r}_\mathcal{D}$) static relation depicted in Equation (18) can be written as

$$B(k) = \begin{bmatrix} b_1(k) \\ \vdots \\ b_{\mathrm{r}_\mathcal{D}}(k) \end{bmatrix} = \begin{bmatrix} f_{1,\mathrm{St}}(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}) + \xi_{b,1}(k) \\ \vdots \\ f_{\mathrm{r}_\mathcal{D},\mathrm{St}}(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}) + \xi_{b,\mathrm{r}_\mathcal{D}}(k) \end{bmatrix} \tag{19}$$

The temperature to deformation temporal model identification problem turns into a $\mathrm{r}_\mathcal{D}$ × MISO ($\mathrm{r}_\mathcal{T}$ to 1) identification problems in terms of (20)

$$\min_{\hat{b}_i(k)} \quad \sqrt{\frac{1}{\mathrm{N}} \sum_{t=0}^{\mathrm{N}} \left(b_i(k) - \hat{b}_i(k) \right)^2} \tag{20}$$

where $\{\hat{b}_i(k)\}_{i=1}^{\mathrm{r}_\mathcal{D}} = \{\hat{f}_{i,\mathrm{St}}(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}})\}_{i=1}^{\mathrm{r}_\mathcal{D}}$. The set of static polynomial terms of an arbitrary order p is large enough to represent most of the core structures that form a wide range of signals. Therefore, we assume that the static functions $\hat{f}_{i,\mathrm{St}}$ can be well described by a linear combination of polynomial terms. For a predefined *input polynomial* (IP) basis set $\mathcal{P} = \{P_l\}_{l=1}^{\mathrm{n_p}}$ of order p, the candidate models for each $\hat{f}_{i,\mathrm{St}}$ can be defined by:

$$\hat{f}_{i,\mathrm{St}}(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}) = \sum_{l=1}^{n_p} c_{l_i} P_{l_i}(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}) \tag{21}$$

where $n_p = \mathrm{Card}(\mathcal{P})$ is the number of polynomial basis. Thus, the minimisation problem (20) can be written as

$$\min_C \sqrt{\frac{1}{\mathrm{N}} \sum_{t=0}^{\mathrm{N}} \left(b_i(k) - \sum_{l=1}^{n_p} c_{l_i} P_{l_i} \left(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}} \right) \right)^2} \tag{22}$$

where $C = \{c_{l_i}\}$ is a coefficient array. Equation (22) can be interpreted as: which subset of $\mathcal{P}$, approximates the best the signal $b_i(k)$. Considering the possible large dimension $\mathrm{r}_\mathcal{T}$ of signal $T(k)$, its finite, but arbitrary large time shift values $\{T(k-j)\}_{j=1}^{n_t}$ and the polynomial order p, the set $\mathcal{P}$ contains a considerable amount of polynomial terms $P_l\left(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}\right)$. The task of selecting the polynomial terms that together form a model for approximating each of the b_i signal is called *equation discovery* (ED). This can be completed through at least two methods: enhancing model sparsity starting from a large set $\mathcal{P}$ or evolving the polynomial terms via TAG3P.

Equation discovery via enhancing model sparsity To select the subset $\mathcal{P}_{\min} = \{P_s \in \mathcal{P}\}$ of polynomial terms with minimum cardinality and from a predefined library of polynomial terms $\mathcal{P}$, the problem (22) turns

$$\min_C \sqrt{\frac{1}{\mathrm{N}} \sum_{t=0}^{\mathrm{N}} \left(b_i(k) - \sum_{l=1}^{n_p} c_{l_i} P_{l_i}\left(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}\right) \right)^2} + ||C||_0 \tag{28}$$

where $||\cdot||_0$ is the l_0 pseudo norm. In order to solve the problem of sparse signal recovery, which is NP hard,

Algorithm 1 Sparse signal recovery ($\mathcal{P}_{\min,i}, \mathcal{C}_{\min,i}$, $\hat{b}_i(k)$)

Define $C = \{c_l\}_{l=1\dots n_p}$ ▷ (parameter array)
Define $\mathrm{D}^0 = \mathrm{I}_{n_p}$ ▷ (initial weight matrix)
Define ν ▷ (weight regularization factor [0,1])
Define ε ▷ (non-zero parameter)
Define μ ▷ (parameter threshold)
Define m ▷ (Number of maximum iterations)
for $j = 1,\ j+1,\ j \le \mathrm{m}$ **do**
Solve:

$$C_i^{(j)} = \min_{c_{l,i}^{(j)}} \sqrt{\frac{1}{\mathrm{N}} \sum_{t=0}^{\mathrm{N}} \left(b_i(k) - \sum_{l=1}^{n_p} c_{l,i}^{(j)} P_{l,i}(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}) \right)^2} + \frac{1}{\nu} ||\mathrm{D}^{(j-1)} C_i^{(j)}||_1 \tag{23}$$

Update:

$$\mathrm{D}^j = \mathrm{diag}\left\{ \frac{1}{c_{l,i}^{(j)} + \varepsilon} \right\} \tag{24}$$

end for
Construct

$$\mathcal{I}_i = \left\{ s \in \{1 \dots n_p\} \,|\, c_{s,i}^{(\mathrm{m})} > \mu \right\}, \quad n_i = \mathrm{Card}(\mathcal{I}_i) \tag{25}$$

Re-optimize

$$\mathcal{C}_{\min,i} = \min_{\{c_{s,i}\}_{s\in\mathcal{I}_i}} \sqrt{\frac{1}{\mathrm{N}} \sum_{t=0}^{\mathrm{N}} \left(b_i(k) - \sum_{s\in\mathcal{I}_i} c_{s,i} P_{s,i}(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}) \right)^2} \tag{26}$$

Construct

$$\hat{b}_i(k) = \hat{f}_i(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}) = \sum_{s\in\mathcal{I}_i} c_{s,i} P_{s,i}(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r}_\mathcal{T}}) \tag{27}$$

in (Candès et al., 2007), the authors showed that the sequential solving of a weighted l_1 problems leads to the solution of the l_0 regularization form. Algorithm 1 is built around the guidelines in (Candès et al., 2007). In order to identify the evolution of the entire $\hat{B}(k)$ deformation temporal coefficients, we ran Algorithm 1 for each $i = 1 \dots r_\mathcal{D}$, using a pre-defined set of polynomials $\mathcal{P}$ and determined $\mathrm{r}_\mathcal{D}$ MISO temporal models.

Equation Discovery via TAG3P The main goal of this strategy is to construct the candidate solution via genetic programming that solves the problem (22) by evolving a population of candidates. This genetic population evolution has the goal of exploring the polynomial search space and selecting the terms that minimize the cost function (22). In contrast to the ESS method, the search space within MIMO TAG3P method is not defined by the polynomial order p but by the number of possible combinations of finite number of auxiliary trees up to a given maximum limit. The TAG3P with grammar G_{IP} aims to construct candidate models that minimize problem (20). In order to identify the evolution of all deformation temporal coefficients $\hat{B}(k)$, we ran the TAG3P with G_{IP} for each $i = 1 \dots r_\mathcal{D}$.

Both selection and genetic programming solutions provide candidate models that can be used to generate predictions

Model enhancement via Gaussian process modeling Regardless of the chosen parametric identification strategy to propose a candidate for the temperature-deformation temporal model (18), there will be uncaptured signal modes due to imposed search space restrictions (limited computational memory or limited run-time). Therefore, to compensate for the uncaptured dynamics consider the approximation error signal

$$E(k) = B(k) - \hat{B}(k). \tag{29}$$

We can further model each error signal $e_i(k)$ as a Gaussian process ($\mathcal{GP}$) with the regression model

$$e_i(k) = f_i(T(k)) + \xi_e(k) \tag{30}$$

where $f_i(T(k)) \cong \mathcal{GP}_i(m, \mathrm{K})$ is a Gaussian process defined by the mean function m and covariance matrix K and $\xi_e(k)$ is the modeling error or measured noise. For simplicity, consider the target data points $e_i(k) = e_{i,k}$ and input data points $T(k) = T_k$. For a collection of data points $\mathrm{D} = \{\mathrm{e_i} = \{e_{i,k}\}_{k=1}^{\mathrm{N}}, \mathrm{T} = \{T_k\}_{k=1}^{\mathrm{N}}\}$, a prior distribution is constructed upon the mean values $m(\mathrm{T}) = \mathbb{E}(f_i(\mathrm{T}))$ and the kernel $[\mathrm{K}]_{j_1,j_2} = \mathrm{cov}(\mathrm{T}_{j_1}, \mathrm{T}_{j_2}) = \mathbb{E}(f_i(\mathrm{T}_{j_1}) - m(\mathrm{T}_{j_1}))\mathbb{E}(f_i(\mathrm{T}_{j_2}) - m(\mathrm{T}_{j_2}))$. Based on the joint distribution of the prior and the known data:

$$\begin{bmatrix} \mathrm{e_i} \\ f_i(\mathrm{T}_*) \end{bmatrix} = \mathcal{N}\left(0, \begin{bmatrix} \mathrm{K}(\mathrm{T},\mathrm{T}) + \sigma_{\mathrm{n}}^2 I_{\mathrm{n}} & \mathrm{K}(\mathrm{T},\mathrm{T}_*) \\ \mathrm{K}(\mathrm{T}_*,\mathrm{T}) & \mathrm{K}(\mathrm{T}_*,\mathrm{T}_*) \end{bmatrix}\right) \quad (31)$$

where T_* is a set of new data points, prediction values over a test point T_j can be obtained by applying the *maximum a posteriori estimation*(MAP)

$$f_i(T_j)|\mathrm{D}, T_j \sim \mathcal{N}\left(\bar{f}_i(T_j), \mathrm{cov}(f_i(T_j))\right) \quad (32)$$

where

$$\begin{aligned} \bar{f}_i(T_j) &= \mathrm{K}_j^\top \left[\mathrm{K} + \sigma_{\mathrm{n}}^2 I_{\mathrm{n}}\right]^{-1} \mathrm{e_i} \\ \mathrm{cov}(f_i(T_j)) &= \mathrm{K}_{j,j} - \mathrm{K}_j^\top \left[\mathrm{K} + \sigma_{\mathrm{n}}^2 I_{\mathrm{n}}\right]^{-1} \mathrm{K} \end{aligned} \quad (33)$$

with $\mathrm{K}_j = \mathrm{K}(\mathrm{T}, T_j)$, $\mathrm{K} = \mathrm{K}(\mathrm{T},\mathrm{T})$, $\mathrm{K}_{j,j} = \mathrm{K}(T_j, T_j)$. For a full description and derivation of the Gaussian processes modeling and inference method see Chapter 2 in Rasmussen and Williams (2006). To compute the $\mathcal{GP}$ models we used the `gpml` Matlab Toolbox introduced in Rasmussen and Williams (2006) with polynomial kernels of order up to 4.

4. MIRROR MODEL

4.1 Heat-Deformation data-generating system

We have tested the proposed identification solutions on a simplified synthetic mirror model. The first part of the model simulates the 2D thermal dissipation and it is based on the continuous-time thermal diffusion PDE over space $\mathbb{X}$ described in Equation (34).

$$\rho c \frac{\partial}{\partial t}\mathcal{T}(\mathbb{X}, t) = k\left(\frac{\partial^2}{\partial x_1^2}\mathcal{T}(\mathbb{X}, t) + \frac{\partial^2}{\partial x_2^2}\mathcal{T}(\mathbb{X}, t)\right) + \mathcal{Q}(\mathbb{X}, t) \quad (34)$$

with Dirichlet boundary conditions, where ρ, c, and k are material density, heat capacity and thermal conductivity. The material coefficients were selected to match a natural thermal diffusion evolution. In order to simulate the thermal diffusion (34) we applied the separation of variable framework using the 2D cosine Fourier basis set $\Phi_F(\mathbb{X})$ described in (10) with $\mathrm{r}_{x1} = r_{x_2} = 40$. As described in van den Hurk et al. (2018) the outcome of this is a continuous-time state-space linear model. The reduced orders $\mathrm{r}_{x1} = r_{x_2} = 40$ correspond to a trade-off between representation error and model complexity. Discrete-time spatial temperature values $\mathcal{T}(\bar{x}_j, k\mathrm{T_s})$, sampled at $\mathrm{T_s} = 500$s are obtained by applying form (6). For the considered thermal behavior the sampling frequency $\mathrm{f_s} = 0.025$Hz was chosen to be at least two times faster than the bandwidth of the state-space linear model.

The second part of the model represents a simplified model for the static thermal deformation on the surface of the in-depth mirror slice. The spatial deformation model F_{poly} is assumed to be second order polynomial mapping of the temperature signal: (35)

$$\mathcal{D}(y, k) = F_{\mathrm{poly}}(\mathcal{T}(y, k)), \forall y \in \mathbb{Y} \quad (35)$$

Using the Fourier cosine 1D SBFs, the temperature signal $\mathcal{T}(\mathbb{Y}, k)$ is obtained by re-mapping the temporal coefficients defined on $\mathbb{X}_s$ to $\mathbb{Y}$. Thus,

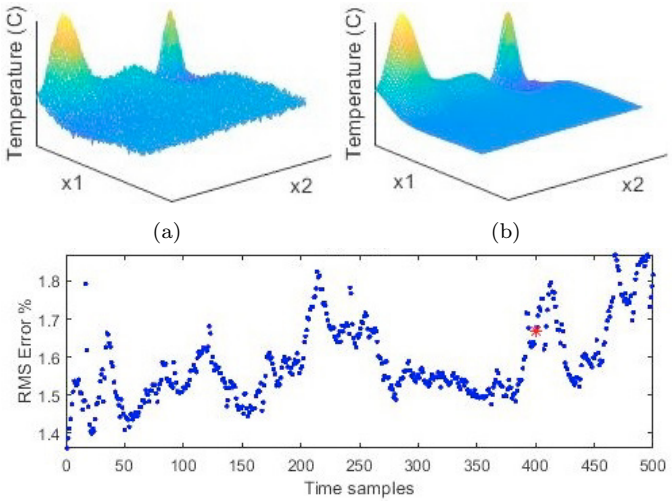

Fig. 2. (2a) The synthetic noisy temperature profile at $k = 400$, (2b) The subspace identified model simulated temperature profile at $k = 400$, (2c) Spatial error $\mathbb{S}_{\mathrm{s}}\left(\hat{\mathcal{T}}(\bar{x}_j, k)_{j=1}^{\mathrm{M}}\right)$ over $\mathbb{D}_{\mathrm{v}}(\mathrm{N_{id}})$.

$$\begin{aligned} \mathcal{T}(\mathbb{Y}, k) &= F_{\mathrm{inter}}(T(\mathbb{X}_s, k))) = \sum_{i=1}^{n_{x_1}} a_{\mathbb{X}_s,i}(k)\varphi_i(\mathbb{Y}), \\ a_{\mathbb{X}_s,i}(k) &= \langle T(\mathbb{X}_s, k)), \varphi_i(\mathbb{X}_s)\rangle, \end{aligned} \quad (36)$$

where $\varphi_i(\mathbb{X}_s)$, $\varphi_i(\mathbb{Y})$ are the SBFs described in (9) and formulated on $\mathbb{X}_s$ and $\mathbb{Y}$.

4.2 Spatio-temporal input signal

The input signal $\mathcal{Q}(\mathbb{X}, k)$ was designed to represent a generic laser exposure pattern over an in-depth mirror slice. The input signal is spatially located at the top layer of the 2D in-depth slice, and it is formed by three separated gaussian profile signals $\mathcal{Q}(\mathbb{X}, k) = \{\mathcal{Q}_1(\mathbb{X}, k), \mathcal{Q}_2(\mathbb{X}, k), \mathcal{Q}_3(\mathbb{X}, k)\}$ with centers at $\frac{\mathbb{L}_2}{6}$, $\frac{\mathbb{L}_2}{2}$, $\frac{5\mathbb{L}_2}{6}$ bounded between $\left[0, \frac{2\mathbb{L}_2}{6}\right]$, $\left(\frac{2\mathbb{L}_2}{6}, \frac{4\mathbb{L}_2}{6}\right]$ and $\left(\frac{4\mathbb{L}_2}{6}, \mathbb{L}_2\right]$. All three gaussian profiles have a maximum spatial amplitude of 1 and spatial variance of 0.65. The temporal variantion of the signals is a normal distributed random variable.

In order to determine the POD basis (5) we constructed a data set $\mathbb{D}_{\mathrm{POD}}(\mathrm{N})$ with $\mathrm{N} = 10000$ and applied the SVD procedure described in Section 2.4. For the identification of the heat-temperature (7) and the temperature-deformation (8) relationship we have defined 10 estimation $\mathbb{D}_{\mathrm{e}}(\mathrm{N_{id}})$, testing $\mathbb{D}_{\mathrm{t}}(\mathrm{N_{id}})$ and validation data sets $\mathbb{D}_{\mathrm{v}}(\mathrm{N_{id}})$ with $\mathrm{N_{id}} = 500$. For each data set, the temporal evolution of the heat $\mathcal{Q}(\mathbb{X}, k)$, temperature $\mathcal{T}(\mathbb{X}, k)$, $\mathcal{T}(\mathbb{Y}, k)$ and deformation $\mathcal{D}(\mathbb{Y}, k)$ signals is described by 4 sets of $\mathrm{r}_{\mathcal{Q}} = 3$, $\mathrm{r}_{\mathcal{T}} = 23$, $\mathrm{r}_{\mathcal{T}} = 23$ and $\mathrm{r}_{\mathcal{D}} = 19$ temporal signals respectively. As a consequence, the candidate temporal models $\mathcal{F}_{\mathrm{LTI}}(\cdot)$ and $\mathcal{F}_{\mathrm{St}}(\cdot)$ are described by a 3×23 and nineteen 23×1 input output models respectively. The reduction orders $\mathrm{r}_{\mathcal{Q}}$, $\mathrm{r}_{\mathcal{T}}$ and $\mathrm{r}_{\mathcal{D}}$ were selected based on a accepted representation error $\eta_r = 99.9\%$ as described in Equation (15). Over the temperature signal $\mathcal{T}(\mathbb{X}, k)$ we have added a point-wise $\pm 0.3C^\circ$ measurement noise resulting in a signal to noise ratio SNR $\cong 76$.

5. IDENTIFICATION RESULTS

The spatial-temporal error between two generic signals $\mathcal{S}(\bar{x}_i, k)$ and $\hat{\mathcal{S}}(\bar{x}_i, k)$ is quantified as the spatial root mean square error in percentage $\mathbb{S}$ depicted in Equation (37).

$$\mathbb{S}(\mathcal{S}(\bar{x}_i,k),\hat{\mathcal{S}}(\bar{x}_i,k)) = \frac{\sqrt{\frac{1}{\mathrm{M}}\sum_{i=1}^{\mathrm{M}}\left(\mathcal{S}(\bar{x}_i,k)-\hat{\mathcal{S}}(\bar{x}_i,k)\right)^2}}{\sqrt{\frac{1}{\mathrm{M}}\sum_{i=1}^{\mathrm{M}}\mathcal{S}(\bar{x}_i,k)^2}} \quad (37)$$

where M is the number of spatial sampled points.

5.1 Heat-Tempemperature identification results

The `ssest()` Matlab function was deployed to solve the minimisation problem (17). Based on the initial subspace identification, we set the state number to 19. The identified model achieved an average $\bar{\mathbb{S}}_{\mathrm{s}}\left(\hat{\mathcal{T}}(\bar{x}_j,k)_{j=1}^{\mathrm{M}}\right) = 1.62\%$ over 10 validation $\mathbb{D}_{\mathrm{v}}(\mathrm{N_{id}})$ data sets. Figure 2 shows a comparison between generated noisy temperature profile, simulated temperature profile at $k = 400$ and $\mathbb{S}_{\mathrm{s}}$ error over an entire $\mathbb{D}_{\mathrm{v}}(\mathrm{N_{id}})$ data set.

The TAG3P with the grammar G_{LTI} was used to solve the minimisation problem (17). Within this solution we used the following genetic parameters: Pop = 50, Gen = 200, Complexity = 120. The identified heat-flux to temperature model yields an average spatial error $\bar{\mathbb{S}}_{\mathrm{s}}\left(\hat{\mathcal{T}}(\bar{x}_j,k)_{j=1}^{\mathrm{M}}\right) = 2.14\%$ over 10 validation data sets $\mathbb{D}_{\mathrm{v}}(\mathrm{N_{id}})$.

5.2 Temp-Deformation identification results

For the deformation temporal models we have used the average root mean square prediction error:

$$\mathrm{RMS_p} = \frac{1}{\mathrm{r_{\mathcal{D}}}}\sum_{i=1}^{\mathrm{r_{\mathcal{D}}}}\sqrt{\frac{1}{\mathrm{N}}\sum_{k=1}^{\mathrm{N}}(b_i(k)-\hat{b}_i(k))^2} \quad (38)$$

where $\hat{b}_i(k)$ are predicted identified output values.

Equation discovery via enhancing signal sparcity The enhancing model sparsity solution described in Section 3.2.1 was deployed with a predefined polynomial set (39) parameters: $\nu = 0.6$, $\epsilon = 1\mathrm{e}{-200}$, $\mu = 1\mathrm{e}{-4}$, $\mathrm{m} = 5$.

$$\mathcal{P} = \left\{P_{l_{i,j}}\left(\{t_\rho(k)\}_{\rho=1}^{\mathrm{r_\mathcal{T}}}\right) = t_i(k)^{p_1}t_j(k)^{p_2}\right\} \quad (39)$$

where $i,j = 1\ldots r_{\mathcal{T}}$ and $p_1,p_2 \in \{0,1,2\}$.

The resulting model is formed by $\mathrm{r_{\mathcal{D}}} = 19$ functions of form (27) and nineteen $\mathcal{GP}$ models.

Equation discovery via TAG3P The TAG3P with the grammar G_{IP} was deployed to solve the minimisation problem (20). For all $\mathrm{r_{\mathcal{D}}}$ identified MISO model we used the following genetic parameters: Pop = 50, Gen = 1200 Complexity = 65. The agregate model is formed by $\mathrm{r_{\mathcal{D}}} = 19$ MISO models and nineteen $\mathcal{GP}$ models. Figures 3a shows the synthetic deformation profile, predicted deformation profiles produced by the complete solution (TAG3P+$\mathcal{GP}$) at $k = 400$. The spatial error in percentage of the predicted signals are shown in Figure 3b.

Table 1 presents averaged values for predicted spatial deformation error $\bar{\mathbb{S}}_{\mathrm{p}}\left(\hat{\mathcal{D}}(\bar{y}_j,k)_{j=1}^{\mathrm{M}}\right)$ and predicted temporal error $\mathrm{RMS_p}$ over 10 validation data sets $\mathbb{D}_{\mathrm{v}}(\mathrm{N_{id}})$. For both parametric solutions, the nineteen $\mathcal{GP}$ models are an additional enhancement step.

Table 1. Identified deformation model averaged metrics over 10 $\mathbb{D}_{\mathrm{v}}(500)$, ESS + $\mathcal{GP}$ and TAG3P + $\mathcal{GP}$ solutions

	ESS+$\mathcal{GP}$	TAG3P+$\mathcal{GP}$	ESS	TAG3P
$\bar{\mathbb{S}}_{\mathrm{p}}$	2.10%	1.67%	–	–
$\bar{\mathrm{RMS}}_{\mathrm{p}}$[dB]	−35.69	−37.91	−14.78	−15.61

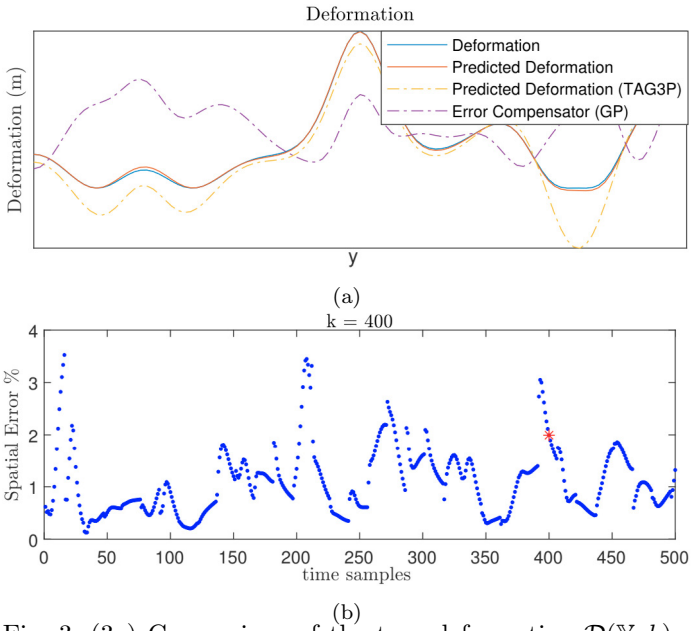

Fig. 3. (3a) Comparison of the true deformation $\mathcal{D}(\mathbb{Y},k)$ and $\hat{\mathcal{D}}(\mathbb{Y},k)$ and the individual TAG3P and $\mathcal{GP}$ components at $k = 400$, (3b) Spatial simulation error $\mathbb{S}_{\mathrm{p}}\left(\hat{\mathcal{D}}(\bar{y}_j,k)_{j=1}^{\mathrm{M}}\right)$ over 500 time samples.

REFERENCES

Antoulas, A.C. (2005). *Approximation of large-scale dynamical systems.* Society for Industrial and Applied Mathematics.

Boyd, J.P. (2000). *Chebyshev and Fourier Spectral Methods.* DOVER Publications.

Candès, E., Wakin, M., and Boyd, S. (2007). Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14.

Chiuso, A. and Pillonetto, G. (2019). System Identification: A Machine Learning Perspective. *Annual Review of Control, Robotics, and Autonomous Systems*, 2.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning.* MIT Press.

Khandelwal, D. (2020). *Automating Data-driven Modelling of Dynamical Systems - An Evolutionary Computation Approach.* Ph.D. thesis.

Li, H.X. and Qi, C. (2011). *Spatio-Temporal Modeling of Nonlinear Distributed Parameter Systems.*

Nechita, S.C. and Tóth, R. (2021). Toolbox for discovering dynamic system relations via tag guided genetic programming. In *19th IFAC Symposium on System Identification SYSID 2021.*

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian processes for machine learning.* MIT Press.

Simidjievski, N., Todorovski, L., Kocijan, J., and Džeroski, S. (2020). Equation discovery for nonlinear system identification. *IEEE Access*, 8, 29930–29943.

van den Hurk, D., Weiland, S., and van Berkel, K. (2018). Modeling and localized feedforward control of thermal deformations induced by a moving heat load. In *2018 SICE International Symposium on Control Systems (SICE ISCS)*, 171–178.

Van Overschee, P. and De Moor, B. (1994). N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1), 75–93.