Probabilistic kernel machines for predictive monitoring of weld residual stress in energy systems

Alamaniotis, M, Mathew, J, Chroneos, A, Fitzpatrick, M & Tsoukolas, L

Original citation & hyperlink:
Alamaniotis, M, Mathew, J, Chroneos, A, Fitzpatrick, M & Tsoukolas, L 2018, 'Probabilistic kernel machines for predictive monitoring of weld residual stress in energy systems' *Engineering Applications of Artificial Intelligence*, vol 71, pp. 138-154 https://dx.doi.org/10.1016/j.engappai.2018.02.009

DOI 10.1016/j.engappai.2018.02.009
ISSN 0952-1976

Publisher: Elsevier

Probabilistic Kernel Machines for Predictive Monitoring of Weld Residual Stress in Energy Systems

Miltiadis Alamaniotis[a1], Jino Mathew[b], Alexander Chroneos[b], Michael E. Fitzpatrick[b], and Lefteri H. Tsoukalas[a]

[a]School of Nuclear Engineering, Purdue University, West Lafayette, IN, USA
{malamani,tsoukala}@ecn.purdue.edu
[b]Faculty of Engineering, Environment and Computing, Coventry University, Priory Street, Coventry CV1 5FB, UK
{ab9932,ab8104,ab6856}@coventry.ac.uk

Abstract

Predictive monitoring supports the *a priori* scheduling of critical component maintenance and contributes significantly in attaining a safe yet economic operation and management of complex energy systems by mitigating the risk of accidents and minimizing the number of operational pauses. The current work studies the learning ability of probabilistic kernel machines, and more particularly of Gaussian Processes (GP) equipped with various kernels for the estimation of weld residual stress profiles of stainless steel pipe welds. The GP models are tested on experimentally-obtained data of axial and hoop residual stresses in two different stainless-steel pipes. The results exhibit the ability of GP to accurately predict the weld residual stress profile in the axial and hoop direction by providing a predictive distribution, i.e., mean and variance values. Furthermore, performance of GP is compared to a non-probabilistic kernel machine, such as support vector regression (SVR) equipped with the same kernels, and to multivariate linear regression (MLR). Comparison results exhibit the robustness of GP over SVR and MLR with respect to prediction

[1] **Corresponding author.** Address: 400 Central Dr., West Lafayette, IN 47907, USA. Tel.: +1 765-496-9696.
E-mail address: malamani@ecn.purdue.edu (M. Alamaniotis)

accuracy of weld residual stress in terms of root mean square error. With respect to a second metric, namely, correlation coefficient between measured and predicted values, GP is superior to SVR and MLR in the majority of the cases.

Keywords: Probabilistic Kernel Machines, Gaussian Process Regression, Machine learning, Welding, Residual stresses, Support Vector Regression.

Highlights:

- Application of probabilistic kernel learning machines for weld stress prediction
- Adoption of sixteen different kernel functions for prediction
- Performance is assessed by RMSE and correlation coefficient
- GP is shown to be robust and advantageous compared to SVR and MLR
- Practical implementation of predictive monitoring of energy systems

1. Introduction

Satisfaction of the growing demand for electrical energy necessitates the continuous operation of power generation plants that provide "base-load" electricity generation (McKoy *et al*., 2013). Part of the overall power plant operational management is the predictive maintenance that encompasses the *a priori* scheduling of maintenance and replacement of close-to-failure mechanical components (Atoui *et al*., 2015), aiming to minimize the number of operational pauses and enhance the overall system safety (Lei *et al.*, 2009; Ebersbach & Peng, 2008). To that end, estimation – within narrow uncertainty bounds – of the remaining life of vital components in energy systems (Liu *et al*., 2008) will contribute to the adoption of predictive monitoring techniques (Wootton et al, 2017). Such techniques may realize cost effective maintenance strategies that lead to enhanced system safety and performance (Hashemian & Bean, 2011; Alamaniotis *et al*., 2014; Mathias *et al.*, 2014).

Essential part in the overall operation of every energy installation are pipes: in thermal power plants, these pipes carry steam at high temperatures and pressures. It has been identified that residual stresses introduced as a consequence of welding processes is a fundamental factor that can lead to the initiation of cracks whose growth may lead to component failure (Castellanos *et al*., 2011; Withers, 2007; Babu *et al*., 2009). The thermomechanical effects of the welding process result in plastic misfits that lead to the generation of elastic residual stresses (Bouchard, 2007; Withers & Bhadeshia, 2001; Francis *et al*., 2007). However, this can be further exacerbated by repair processes (Edwards *et al*., 2004).

Accurate prediction of residual stresses requires the development of analytical models encompassing a high number of interacting factors, detailed knowledge of the welding parameters and materials properties, such as cyclic hardening behavior and the thermal-dependence. It is the

inherent complexity and, often, lack of necessary information in sufficient detail that makes quantification of the residual stress distribution a challenging problem (Stone *et al*., 2008). Experimental techniques such as neutron diffraction (Hutchings *et al.*, 2005), deep hole drilling (George *et al*., 2002), and the contour method (Kartal *et al*., 2013) have been developed to measure residual stress distribution, but come at a high financial cost, cannot be deployed *in situ*, and are at least semi-destructive of the component (Mahmoudi *et al.*, 2009; Prime, 2001). In addition, computational and data-driven methods have been applied in predicting residual stress in energy systems. For example, finite element models (Schwane *et al*., 2014; Xie et al., 2017) and artificial neural networks (Mathew *et al.*, 2014) are used despite their limitations in modeling of highly complex non-linear and interacting processes that introduce a high degree of uncertainty (Tsoukalas & Uhrig, 1997).

Machine learning (Liu *et al*., 2015) has been recently identified as a domain that potentially offers solutions to a large range of predictive problems in materials (Yildiz, 2013; Balachandran *et al*., 2015). Machine learning tools are exposed to a set of known datasets, which consist of experimental or simulated data, in order to evaluate their parameters: a process known as *learning* or *training* (Alamaniotis *et al.*, 2012). The trained models are able to provide predictions for components that are exposed to similar conditions (Bishop, 2006). Predictions may be performed either in the form of interpolation or extrapolation depending on the specifics of the application under study (Alamaniotis *et al*., 2010; Babu *et al*., 2010). To that end, machine learning tools such as artificial neural networks have been applied to predict weld residual stress profiles where the predictions from ensemble networks were used to develop a prediction interval (Mathew *et al*., 2013; Mathew *et al*., 2017a). Further, Dhas & Kumanan (2016) proposed an evolutionary fuzzy support vector regression method that predicts weld residual stresses, while a non-dominated moth

flame optimization technique is proposed by (Savsadi & Tawhid, 2017). In addition, a method that integrates neural networks and evolutionary computing in weld stress prediction is introduced by Dhas & Kumanan (2014), and a hybrid method that utilizes experimental data and neural networks by Mathew et al. (2017b). The adoption of advanced statistical learning methods for weld prediction was presented in (Lewis et al., 2017 July), and the use of an iterative substructure method (ISM) for weld stress prediction in pressurized water reactor (PWR) in (Maekawa et al., 2016). Several weld residual stress prediction techniques are based on the finite elements method in conjunction with various simulated or experimental conditions as introduced in (Wang et al., 2015), (Mondal et al., 2017), (Jiang et al., 2015) and (Afsari et al., 2016).

In this paper, probabilistic kernel machines (Fricke, 2001) are utilized for predicting weld stress profiles of power plant components (Pilania *et al.*, 2013). In particular, the machine learning aspect of Gaussian Processes (GP) (Rasmussen, 2006; Quiñonero-Candela & Rasmussen 2005) modeled with a variety of kernel functions is studied for predicting residual stress, and its performance is compared to non-probabilistic kernel machines. Overall, the set of research objectives contains:

i) Study of the learning ability of various forms of GP to predict the weld stress profiles,

ii) Comparison of GP prediction performance to non-probabilistic driven tools,

iii) Application of GP on recent experimentally obtained datasets of stainless steel pipe welds.

In addition, the practical implications of the current study involve the implementation of predictive monitoring of weld stress in energy systems, and the development of automated maintenance techniques. Additionally, GP predictions of weld stress allow a timely yet low cost replacement of critical system components.

In the next section, Gaussian processes are presented in the context of probabilistic kernel machines, while section 3 proposes the application of Gaussian process for weld residual stress prediction and lastly, section 4 concludes the paper and summarizes its findings.

2. Probabilistic Kernel Machines

2.1. Kernel-based Gaussian Process Regression

With the exception of linear regression methods where the output is a single value that is computed by a set of weighted inputs, there exists a class of methods that make predictions in the function space (Gregorčič & Lightbody, 2009); such methods are the probabilistic kernel machines. In the realm of machine learning, probabilistic kernel machines are recognized as the Bayesian extension of simple kernel machines (Bishop, 2006; Rasmussen, 2006). Here, we implement the notion of a kernel-based Gaussian process with a joint distribution that is modeled as a function of a kernel.

A kernel, which is denoted as $k(x_1, x_2)$, is a valid mathematical function that can be written as (Bishop, 2006):

$$k(x_1, x_2) = f(x_1)^T f(x_2) \tag{1}$$

where $f(x)$ is any valid analytical function. Expressing an analytical model as a function of a kernel is called the "kernel trick" (Bishop, 2006). A Gaussian process expressed as a function of Eq. (1) and utilized for prediction making in regression problems is simply called *Gaussian process regression* (GPR).

Derivation of GPR has an initial point analogous to the simple linear regression, whose vector form is given below:

$$y(x) = \boldsymbol{w}^T \varphi(x) \tag{2}$$

where $\boldsymbol{w}$ is the regression coefficient vector and $\varphi(x)$ the vector containing the basis functions. The basis function may be nonlinear in relation to x, but still y remains a linear combination of basis functions. Next, a prior distribution over regression coefficients w is set, as shown in Eq. (3):

$$p(\boldsymbol{w}) = N(\boldsymbol{w}|0, \sigma^2 I) \tag{3}$$

where the mean value is zero (a convenient choice since we have little or no prior information on weights), I is the identity matrix, and σ^2 denotes the variance associated with the regression coefficients.

A Gaussian process is defined by two parameters, namely, its mean $m(x)$ and covariance $C(x', x)$ function. Therefore, a Gaussian process is given by:

$$GP \sim N(m(x), C(x', x)) \tag{4}$$

where $m(x)$ is taken as equal to zero and $C(x', x)$ is replaced by a kernel function, i.e., kernels are inserted into the *GP* as covariance functions. As a result, the prior distribution over the output y is:

$$P(y) = N(0, \boldsymbol{K}) \tag{5}$$

with **K** being the so-called Gram matrix, whose elements are taken as:

$$K_{ij} = k(x_i, x_j) = \sigma_{ij}^2 \sum_p \sum_n \varphi_p(x_i)\varphi_n(x_j)\ . \tag{6}$$

In general, a measured value contains the real value y added by noise:

$$t_n = y(\boldsymbol{x}) + \varepsilon_n, \tag{7}$$

where t_n is the n^{th} datapoint of y and ε_n is uncorrelated Gaussian noise of zero mean and variance σ_n^2. Hence, we obtain a Gaussian distribution over the observed targets:

$$P(\boldsymbol{t}) = N(0, \mathbf{C}) = N(0, \mathbf{K} + \sigma_n^2 I), \tag{8}$$

where the elements of the covariance matrix are:

$$C_{km} = k(\mathbf{x}_k, \mathbf{x}_m) + \sigma_n^2 \delta_{km} \tag{9}$$

with δ_{km} representing the Dirac delta function. Considering the population of available datapoints

being equal to N, then Eq. (8) can be used to predict the target value t_{N+1} by utilizing the previously observed vector $\mathbf{t}_N$ because the joint probability density $P(t_{N+1}, \mathbf{t}_N)$ is Gaussian as well. Utilizing the property that the posterior distribution of t_{N+1} at $\mathbf{x}_{N+1}$ is also Gaussian, we obtain:

$$P(t_{N+1}|\mathbf{t}_N) \propto exp\left[-\frac{1}{2}[\mathbf{t}_N t_{N+1}]\mathbf{C}_{N+1}^{-1}\begin{bmatrix}\mathbf{t}_N^T \\ t_{N+1}\end{bmatrix}\right]. \quad (10)$$

Next, we perform the trick of splitting the covariance matrix $\mathbf{C}_{N+1}$ into four submatrices:

$$\mathbf{C}_{N+1} = \begin{bmatrix}[\mathbf{C}_N] & [\mathbf{k}] \\ [\mathbf{k}^T] & [k]\end{bmatrix}, \quad (11)$$

where: i) matrix $\mathbf{C}_N$ is a square matrix of dimension NxN; ii) $\mathbf{k}$ is a vector of length N with elements evaluated by the kernel $k(\mathbf{x}_m,\mathbf{x}_{N+1})$ with $m=1,\ldots,N$; iii) $\mathbf{k}^T$ is the transposed vector of $\mathbf{k}$; and iv) k is a scalar equal to output of the kernel $k(\mathbf{x}_{N+1},\mathbf{x}_{N+1})$.

Using the above splitting trick, it can be shown (Gibbs, 1997; Mackay, 1998) that the GPR provides a predictive distribution at $\mathbf{x}_{N+1}$ that follows a Gaussian with mean and variance taken as:

$$m(\mathbf{x}_{N+1}) = \mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{t}_N\ , \quad (12)$$

$$\sigma^2(\mathbf{x}_{N+1}) = k - \mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{k}\ . \quad (13)$$

with $\sigma = \sqrt{\sigma^2}$ being the standard deviation. Overall, we observe from the above equations that predictions made by GPR depend upon the form of the selected kernel function.

2.2. Parameter Evaluation

As mentioned before, a valid kernel function, which is evaluated as a covariance function in the GPR framework, can significantly affect the prediction process (Wang et al., 2013). Importantly, each kernel function may be expressed as a function of one or more parameters. The kernel parameters are evaluated using the maximum likelihood method on the available datasets. In particular, the log-likelihood form of the GPR predictive distribution (Bishop, 2006), i.e.,

$$\ln p(t|\theta) = -\left(\frac{1}{2}\right)\ln(|\mathbf{K}_N|) - \left(\frac{1}{2}\right)\mathbf{t}^T\mathbf{K}_N^{-1}\mathbf{t} - \left(\frac{N}{2}\right)\ln(2\pi) \quad (14)$$

is maximized with respect to a set of parameters that is denoted with the Greek letter θ. In case a kernel does not contain any explicit parameters, then utilization of the maximum likelihood expression in Eq. (14) does not occur. Needless to say, parameter evaluation occurs prior to using GPR for prediction making, otherwise the model is not complete, and no prediction can be made.

3. Weld Residual Stress Profile Prediction

3.1. Kernel models

Gaussian process regression is utilized here for prediction of weld residual stress profiles by employing historical experimental residual stress data. In the present study, the following set of kernels has been applied to the available datasets (Bishop, 2006; Rasmussen, 2006):

i) *Constant Kernel:*

$$k(x_1, x_2) = 1/\theta \quad (15)$$

This kernel is a constant value obtained as the ratio of the unit over a parameter θ equal to the variance of the training data.

ii) *Linear Kernel:*

$$k(x_1, x_2) = \theta x_1^T x_2 \quad (16)$$

This kernel simply implements the inner product of the datapoints x_1 and x_2, multiplied by a scale parameter θ.

iii) *Matérn Kernel*

$$k(x_1, x_2) = \left(2^{1-\theta_1}/\Gamma(\theta_1)\right)\left[\sqrt{2\theta_1}|x_1 - x_2|/\theta_2\right]^{\theta_1} K_{\theta_1}\left(\sqrt{2\theta_1}|x_1 - x_2|/\theta_2\right) \quad (17)$$

The *Matérn* kernel is comprised of two parameters θ_1, θ_2. Here, θ_1 = 3/2 (see Rasmussen (2008) for more details), $\Gamma(\,)$ is the gamma distribution, while $K_{\theta_1}(\,)$ is a modified Bessel function.

iv) *Neural-Network-based kernel:*

$$k(x_1, x_2) = \theta_0 sin^{-1}\left(\frac{2\tilde{x}_1^T \Sigma \tilde{x}_2}{\sqrt{(1+\tilde{x}_1^T \Sigma \tilde{x}_1)(1+\tilde{x}_1^T \Sigma \tilde{x}_2)}}\right) \quad (18)$$

where $\tilde{x} \triangleq (1, x_1, \ldots, x_D)^T$, Σ is the covariance matrix and θ_0 is a scale parameter.

v) *Sum of linear and neural network kernels:* This kernel is the sum of kernels (ii) and (iv).

vi) *Sum of linear and Matérn kernels:* This kernel is the sum of kernels (ii) and (iii).

vii) *Sum of Matérn and neural network kernels.* This kernel is the sum of kernels (iii) and (iv).

viii) *Sum of constant and linear kernels.* This kernel is the sum of kernels (i) and (ii).

ix) *Sum of constant and Matérn kernels.* This kernel is the sum of kernels (i) and (iii).

x) *Sum of constant and neural network kernels.* This kernel is the sum of kernels (i) and (iv).

xi) *Sum of constant, linear and Matérn kernels.* This kernel is the sum of kernels (i), (ii), and (iii).

xii) *Sum of constant, Matérn and neural network kernels:* This kernel is the sum of kernels (i), (iii), and (iv).

xiii) *Sum of linear, Matérn and neural network kernels:* This kernel is the sum of kernels (ii), (ii), and (iii).

xiv) *Sum of constant, linear and neural network kernels:* This kernel is the sum of kernels (i), (ii), and (iiv).

xv) *Sum of constant, linear and Matérn kernels:* This kernel is the sum of kernels (i), (ii), and (iiv).

xvi) *Sum of constant, linear, Matérn and neural network kernels:* This kernel is the sum of kernels (i), (ii), (iii) and (iiv).

Notably, the kernel models expressed in (i)-(xvi) are valid kernels that are used within the GPR framework to make predictions of the weld stress profile. The Gaussian process models are trained (for evaluation of the process parameters) by optimizing the log-likelihood (refer to Eq. (14)) using the *Polak-Ribiere line search method* (Luenberger, 1984).

The sixteen probabilistic kernel models (i)-(xvi) are trained and validated on experimentally-obtained residual stress data. The data were divided into two groups: 1) the training data which are used for training the models prior to any prediction; and 2) validation data which are used to assess the performance of the models against datasets not previously seen. The overall process of predicting the weld stress profiles is depicted in Fig. 1.

3.2. Benchmark Methods

3.2.1. Support Vector Machines

The support vector machine (SVM) emerged from the Vapnik-Chervonenkis theory and consists of a generalization of the Portrait Algorithm (Cristianini & Shawe-Taylor, 2000; Bishop,

2006). It has gained significant popularity in the last decaddes and has been widely applied to several problems.

Support vector machines (SVM) is also a class of kernel machines that are commonly used for classification and regression problems (Cristianini & Shawe-Taylor, 2000). In the latter case, it is called support vector regression (SVR) and provides a set of sparse solutions. In contrast to Gaussian processes, SVM does not provide posterior probabilities and therefore is a non-probabilistic kernel machine, while its model parameters are determined via a convex optimization problem (Vapnik, 2013). Furthermore, SVM does not adopt prior probabilities in its formulation as opposed to GPR.

SVR is used to make predictions by minimizing a regularized error function:

$$C\sum_{i=1}^{N} E_{\varepsilon}\left(y(x_i)-t_i\right)+(1/2)\|w\|^2 \tag{19}$$

where $E_{\varepsilon}(\,)$ is an ε-insensitive error function, C is a regularization parameter adopted to control the predicted values, and w is a penalty term. The error function in (19) takes the form that is given in Eq. (20):

$$E_{\varepsilon}(y(x)-t)=\begin{Bmatrix} 0, & if\ |\,y(x)-t\,|<\varepsilon \\ |\,y(x)-t\,|-\varepsilon, & otherwise \end{Bmatrix} \tag{20}$$

where ε is a positive constant, $y(x)$ is the predicted and t the target value. Equation (20) is known as the *ε-insensitive* error function, with those data-points that lie on the boundary or outside of the *ε-insensitive* region to be called *support vectors*. Therefore, the formulation of SVR allows only a subset of the available data to be considered for prediction making; to make it clear, only those datapoints identified as suppor vectors contribute to the final prediction.

Assuming a set of N available datapoints, SVR takes the form of the *v*SVR, where a portion of datapoints equal to v lies outside the *ε-insensitive* region. In particular, at most $v*N$ points lie

outside the *ε-insensitive* region, and at least *v*N* datapoints are identified as support vectors. In this formulation, a solution is determined using the Lagrange multipliers *a and b*:

$$L(a,b) = -(1/2)\sum_{i=1}^{N}\sum_{j=1}^{N}(a_i - b_i)(a_j - b_j)k(x_i, x_j) + \sum_{i=1}^{N} t_i (a_i - b_i) \tag{21}$$

$$0 \le a_i \le (C/N) \tag{22}$$

$$0 \le b_j \le (C/N) \tag{23}$$

$$\sum_{i=1}^{N}(a_i - b_i) = 0 \tag{24}$$

$$\sum_{i=1}^{N}(a_i + b_i) \le vC \tag{25}$$

where *t* is the target value, and *k*( ) is a kernel function. Furthermore, we observe that the Lagrange coefficients are imposed to a set of box and linear contraints, thus defining a single objective constrained optimization problem. Once a solution is found, then prediction may be obtained by using the following expressions:

$$\beta = t_i - \varepsilon - \sum_{l=1}^{N}(a_l - b_l)k(x_i, x_l) \tag{26}$$

$$y(x) = \sum_{j=1}^{N}(a_j - b_j)k(x, x_j) + \beta \ . \tag{27}$$

From Eq. (27), we observe that SVR (and generally the SVM) does provide a single output value for a specific input; in other words, it gives a point estimate, i.e., a single value, as opposed to a Gaussian process, which provides a predictive distribution, i.e., mean and variance values. Notably, the output of the SVR model depends on the form of the selected kernel, that should be carefully done by the modeler taking into consideration the problem at hand. Therefore, the advantage of the SVR is that it consolidates the problem of regression down to constraint

optimization problem (Eq. 21-25). The interested reader is referred to Bishop (2006) for a concise and comprehensive introduction to SVR.

In this manuscript, SVR is also applied for prediction-making using the kernels (i)-(viii) presented in the previous subsection. Furthermore, we adopt the vSVR models with v=0.2, and regularization parameter C=10. The SVR parameters are evaluated after performing a cross validation approach of the SVR models on the available training data. The dual of the aforementioned values was found to be give the best prediction performance pertained to the cross-validation of the training data (cross-validation by leaving out 20% of the available data at each iteration). Given that there is no widely accepted method for choosing the above SVR values for unknown datasets, then cross-validation, yet time consuming, accommodates selection of parameter values (Bishop, 2006).

3.2.2. Multivariate Linear Regression

Statistical tools have been widely used for data analysis, estimation and prediction. The most common tool in statistics in the simple linear regression (SLR). The SLR formula is given by:

$$y = b_1 x + b_0 \tag{28}$$

Where y is the predicted value, b_1 is the slope, and b_0 is the intercept. The independent value x is also known as the regressor. In case, where there is more than one regression, then SLR takes the form of a multivariate linear regression (MLR):

$$y = b_0 + \sum_{n=1}^{N} b_n x_n \tag{29}$$

where b_0 is the intercept and b_n the regression coefficients. In the case of Eq. (29) the population of regressors (variables) is equal to N. Therefore, the predicted value y is the result from contributions coming from multiple parameters.

Regression coefficients are evaluated by minimizing the distance of the curve defined by (Eq. (29) from the available data. This distance is expressed in term of a mean square error between the available datapoints and the values of *y* (Bishop, 2006).

In the current work, MLR is applied on the available data in order to evaluate is coefficients and subsequently to predict the weld stress profile. Notably the number of regressors is N=3; the variables of MLR are the variables shown in Eq. (30).

3.4. Benchmark datasets

The GPR, SVR and MLR models were developed using residual stress data in stainless steel pipe welds collated in Bouchard (2007). These measurements were undertaken by diverse measurement techniques such as deep hole drilling, neutron diffraction, and block removal splitting and layering as part of a UK nuclear power industry research programme. A detailed description of the residual stress data and the experimental techniques can be found in Bouchard (2007). Training data are comprised of weld residual stress profiles from eight different pipe components fabricated from austenitic stainless steel. A schematic diagram defining the stress components and the geometry of a welded pipe is shown in Fig. 2. The measurement database covers a wide range of welding heat input *Q* in the range [0.8-2.2] kJ/mm, wall thickness (*t*) in the range [16-110] mm, and mean radius-to-wall-thickness ratio (*R*/*t*) in the range [1.8-25.0], which are considered to be the key input parameters controlling the residual stress distribution in circumferentially welded pipes (Song *et al*., 2015). The validation datasets are obtained from two

butt-welded pipe components also fabricated from austenitic stainless steel of type 316L, with dimensions 320mm long, 250mm diameter and 25mm thickness, measured using neutron diffraction (Mathew *et al.,* 2017). In the present study, the input parameters to the tested models, are 3*x*1 vectors with entries of 1) the radius to thickness ratio *R*/*t*; 2) the thickness *t*; and 3) the net heat input *Q*, as given in Eq. 30:

$$\text{Input} = [R/t \quad t \quad Q]^{\mathrm{T}} \tag{30}$$

while the output of the model is the predicted residual stress profiles expressed as the through-wall position *x*/*t*, which is the ratio of position from the inner surface of the pipe wall to pipe thickness. It should be noted that in the current work, we use two different validation datasets, that we name them dataset 1 and dataset 2 respectively, to assess the predictive performance of the adopted tools. Each validation dataset contains experimental measurements undertaken using neutron diffraction for measuring both axial and hoop residual stresses.

3.3. Results

The probabilistic kernel machines equipped with kernels (i)-(xvi) are utilized for prediction of weld residual stress profiles in the two benchmark datasets. Likewise, SVR is also equipped with the same kernel models, while both benchmark methods, i.e., SVR and MLR are applied to the same data as GPR. The predictive performance between the predicted and the experimentally-obtained profiles in the validation dataset is evaluated using the root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(E_n - P_n)^2} \tag{31}$$

where E_n is the experimental value, P_n is the predicted value and N is the total number of measurements. In particular, we examine the predicted profiles for the axial and hoop components of the residual stresses, which were determined by the neutron diffraction measurements. It should

be noted that in this study the output data was not normalized, and therefore RMSE is expected to be anywhere in the range [0 +oo].

The RMSE results are presented in Table 1 and Table 2, and provide a quantitative measure of the difference between the predicted profile and the actual measurements for axial and hoop stress profile prediction respectively. In particular, each Table presents a set of 66 RMSE values, i.e., 33 taken with respect to dataset 1 (16 values from GP, 16 values from SVR and 1 from MLR); and 33 taken with respect to dataset 2 (16 values from GP,16 values from SVR and 1 from MLR). It should be noted that MLR is not a kernel machine – keeps the same form – and thus provides only one value.

With regard to the axial predictions provided in Table 1, we observe that the two kernel based tested methods provide very close RMSE values, with their absolute differences being very small in the majority of the cases. It should be noted that by taking into consideration all tested kernels, GP provides more accurate results for 19 out of 32 kernel functions (11 for dataset 1 and 8 for dataset 2) compared to SVR with respect to axial stress profile prediction. Furthermore, we observe that there is no dominant kernel machine model in predicting the axial profile: in dataset 1 the lowest RMSE is provided by SVR equipped with the sum of Matérn and Neural Network kernel, while in dataset 2 the best performance is taken with the GP-Linear kernel model: i.e., two different models give the best performance for two different datasets. With regard to MLR, we observe that the value of RMSE for dataset 1 is equal to 102.434. This value indicates that MLR performs worse than all GP models with the exception of GP-Constant kernel. Furthermore, in comparison to SVR, MLR outperforms 5 out of the 16 tested models for predicting axial performance of dataset 1. With respect to dataset 2, MLR performance is pretty close to most of the GP models (i.e., RMSE values

around 135), but it is clearly outperformed by the GP-Linear kernel. By comparing MLR to SVR, we observe that MLR provides better prediction in 11 out of 16 cases.

The RMSE values imply that both probabilistic and non-probabilistic kernel machines exhibit similar performance. However, there are three factors that make GP superior to SVR. Firstly, all GP models provide RMSE values within a narrow range of values, while the SVR fall within a much wider range. In particular, for dataset 1 the GP range is [95-145] (if excluding constant kernel then it becomes [95-99.7]), while the respective SVR range is [91-304]. Likewise, we also observe that GP models provide RMSE values that fall within a smaller interval than those by SVR. Those observations imply that the GP predictions are more robust and more stable than SVR, promoting the adoption of GP models for axial weld stress prediction; independently of the kernel function, GP will give low RMSE as opposed to SVR that may provide much higher error. Given that we do not know *a priori* which kernel function will be the best fit, then probabilistic kernel machines are preferable to non-probabilistic ones. A second advantage of GP is the absence of modeler-defined parameters: on one hand, SVR requires the *a priori* determination of the hyperparameters v and C in addition to any kernel parameters, a process that requires extra datasets and may be time consuming, while there is no guarantee that the optimal dual can be identified. On the other hand, the GP formula has no user defined hyperparameters that need to be explicitly defined by the modeler (note: in GP only kernel parameters, if any, need to be evaluated), and thus, it ensures the best possible prediction-making (Chatzidakis *et al.*, 2014; Alamaniotis *et al.*, 2011). The third factor refers to the ability of GP to provide a predictive distribution, i.e., a mean value and variance around the mean value: SVR does provide a simple point estimation, i.e., no variance parameter. In such complex systems implementing maintenance strategies necessitates the evaluation of variance before any decision-making.

The predicted axial stress profiles obtained by GP and SVR with each kernel are depicted in Fig. 3 (a and b) and 4(a and b) respectively, along with the actual measurements of dataset 1 used for testing. In Fig. 3, a shaded band is shown that is defined as the interval of values in the range [m-σ, m+σ]. Likewise, the axial profiles computed by GP and SVR for kernels (i)-(iv) pertaining to dataset 2 are provided in Fig. 5.

For the hoop stress profiles, our observations are not any different to the axial stress profiles. For both datasets, the RMSE values taken by GP and SVR are very close. However, in this case SVR provides better accuracy than GP in 18 out of 32 cases. Observations made are the same as in the axial prediction case, and also exhibit the robustness of GP in contrast to SVR. In contrast to the axial case, in the hoop stress prediction there is a single model that provides the highest accuracy in both tested datasets, namely, the SVR-Matérn model. Based on that, it should be noted that the datasets were obtained with the neutron diffraction method and we may state that Matérn kernel provides good results for neutron diffraction measured data. Comparison with MLR exhibits that a high number of GPR models provides a more accurate prediction than MLR (only three models outperformed by MLR). The same behavior is also observed between MLR and SVR. Thus, as a general conclusion, on the average GPR is a better predictor than MLR.

Visualizations of predicted hoop stress profiles with kernels (i)-(iv) plotted against the real values of datasets 1 and 2 are given in Fig. 6 and 7 respectively. The GP profiles are visualized in terms of [m-σ, m+σ] bands. In addition, he profiles taken with MLR for both axial and hoop profiles of datasets 1 and 2 are depicted in Fig. 8, while the obtained RMSE values for both axial and hoop stress datasets are plotted in Fig. 9, to exhibit the variance of those values. Visual inspection of Fig. 9 confirms that the RMSE from the GP prediction varies within a narrower band than that from SVR in all four cases.

In addition to the root mean square error (RMSE) we have adopted a second performance metric: the correlation coefficient, which is computed by the respective predicted weld residual profile and the measured values of each test dataset. The average of the two dataset computed correlation values are provided in Fig. 10 and 11 for axial and hoop stress profiles respectively. In Fig. 10, we observe that the GP model provide the highest correlation in the vast majority of the tested kernels with regard to axial stress profile prediction. In particular, GP provides the best average correlation in 9 out of 16 cases, SVR in 4 out of 16, whilst in 3 cases correlation is exactly the same. With regard to hoop stress profiles, Fig. 11 shows that GP provides the highest correlation in the most of the cases: specifically, in 10 cases. SVR provides the highest correlation in 2 cases, while in the remaining 4 cases the average correlation coefficient coincides for both models. In general, we observe that a) correlation values are very close to each other, and b) correlation coefficients are very high (>0.75 in all cases), demonstrating the ability of kernel machines to provide accurate prediction of weld residual stress profiles. Furthermore, Table 3 provides the average correlation coefficient taken between real values and predicted profiles by MLR. By comparing table 3 values with Fig. 10 and 11, we observe that GPR provides profiles that are more similar to real values than MLR for both axial and hoop profile.

As a general finding, we observe that the complexity of the kernel is not correlated with the prediction performance. From the cases presented here, we see that composite kernels do not necessarily provide more accurate profiles than simpler kernels. Overall, it is the modeler's responsibility to select an appropriate kernel.

4. Conclusion

We have studied and compared the application of sixteen probabilistic kernel models in predicting weld residual stress profiles, and benchmarked them against non-probabilistic kernel machines, and against a widely used tool such as linear regression. In particular, we used sixteen Gaussian process (GP) models, where each is equipped with a different kernel function, and compared them to Support Vector Regression (SVR) models also equipped with the same kernel functions. As an additional benchmark method, we selected the multivariate linear regression. We see that the Gaussian Process models provide robust predictions compared to the SVR and MLR models. In addition, probabilistic kernel machines are able to provide a predictive distribution, i.e., mean and standard deviation, instead of single point estimation. Thus, GP provide a band of possible values and implicitly quantify the uncertainty over the predicted value. From a practical point of view, GP accommodate the implementation of predictive monitoring techniques in complex energy systems, and allow the timely as well the low-cost maintenance and replacement of critical components.

Further, we conclude that a highly sophisticated kernel (sum of two or more simple kernels) does not prevail in prediction-making for this problem compared to simple kernels, and this implicitly denotes that the role of the modeler in selecting the form of kernel used in prediction is critical in determining the overall outcome. This is a clear advantage of the current approach since it allows a significant degree of flexibility in model selection (i.e., the form of the kernel) to the modeler. Future work will be follow two directions. Firstly, we perform testing of the GP models in higher variety of weld residual stress datasets; several experiments have been scheduled regarding the stresses in harsh environments such as nuclear power plants. Secondly, we intend to adopt other valid kernel functions beyond the set of 16 ones that were used in the current work.

The second direction, i.e., the extensive testing of several other kernels, will give us the opprotunity us to build ensemble GP models and applying them in weld stress profile prediction making.

Acknowledgements

The authors are grateful for funding from the Lloyd's Register Foundation, a charitable foundation helping to protect life and property by supporting engineering-related education, public engagement and the application of research. They would like also to acknowledge Professor John Bouchard and Dr. Richard Moat of The Open University for their valuable help in obtaining the experimental datasets.

References

Afshari, D., Sedighi, M., Karimi, M. R., & Barsoum, Z. (2016). Prediction of residual stresses in resistance spot weld. *Aircraft Engineering and Aerospace Technology: An International Journal*, *88*(4), 492-497.

Alamaniotis, M., Grelle, A., & Tsoukalas, L. H. (2014). Regression to fuzziness method for estimation of remaining useful life in power plant components. *Mechanical Systems and Signal Processing*, *48*(1), 188-198.

Alamaniotis, M., Ikonomopoulos, A., & Tsoukalas, L. H. (2012b). Probabilistic kernel Approach to online monitoring of nuclear power plants. *Nuclear Technology*, *177*(1), 132-145.

Alamaniotis, M., Tsoukalas, L. H., & Ikonomopoulos, A. (2011). On-line Surveillance of Nuclear Power Plant Peripheral Components using Support Vector Regression. *Proceedings of the International Symposium on Future I&C for Nuclear Power Plants, Cognitive Systems*

Engineering on Process Control, and International Symposium on Symbiotic Nuclear Power Systems (ICI 2011), Daejeon, Korea, August 2011, pp. 1230(1-6).

Alamaniotis, M., Ikonomopoulos, A. & Tsoukalas, L.H., (2010). Gaussian Processes for Failure Prediction of Slow Degradation Components in Nuclear Power Plants", *Proceedings of the European Safety and Reliability Conference 2010* (ESREL 2010), Rhodes, Greece, September, 2096-2102.

Atoui, M. A., Verron, S., & Kobi, A. (2015). Fault detection with Conditional Gaussian Network. *Engineering Applications of Artificial Intelligence*, *45*, 473-481.

Babu, K. V., Narayanan, R. G., & Kumar, G. S. (2010). An expert system for predicting the deep drawing behavior of tailor welded blanks. *Expert Systems with Applications*, *37*(12), 7802-7812.

Babu, K. V., Narayanan, R. G., & Kumar, G. S. (2009). An expert system based on artificial neural network for predicting the tensile behavior of tailor welded blanks. *Expert Systems with Applications*, *36*(7), 10683-10695.

Balachandran, P. V., Theiler, J., Rondinelli, J. M., & Lookman, T. (2015). Materials Prediction via Classification Learning. *Scientific reports*, *5*.

Bishop, C. M. (2001). Bishop Pattern Recognition and Machine Learning, Berlin: Springer-Verlag Publications.

Bouchard, P. J. (2007). Validated residual stress profiles for fracture assessments of stainless steel pipe girth welds. *International Journal of Pressure Vessels and Piping*, *84*(4), 195-222.

Castellanos, V., Albiter, A., Hernández, P., & Barrera, G. (2011). Failure analysis expert system for onshore pipelines. Part–I: Structured database and knowledge acquisition. *Expert Systems with Applications*, *38*(9), 11085-11090.

Chatzidakis, S., Alamaniotis, M., & Tsoukalas, L. H. (2014). Creep Rupture Forecasting: A Machine Learning Approach to Useful Life Estimation. *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR), 2*(2), 1-25.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Dhas, J. E. R., & Kumanan, S. (2014). Neuro evolutionary model for weld residual stress prediction. *Applied Soft Computing, 14*, 461-468.

Dhas, J. E. R., & Kumanan, S. (2016). Evolutionary fuzzy SVR modeling of weld residual stress. *Applied Soft Computing*, *42*, 423-430.

Ebersbach, S., & Peng, Z. (2008). Expert system development for vibration analysis in machine condition monitoring. *Expert Systems with Applications*, *34*(1), 291-299.

Edwards, L., Bouchard, P. J., Dutta, M., Wang, D. Q., Santisteban, J. R., Hiller, S., & Fitzpatrick, M. E. (2005). Direct measurement of the residual stresses near a 'boat-shaped'repair in a 20mm thick stainless steel tube butt weld. *International Journal of Pressure Vessels and Piping*, *82*(4), 288-298.

Francis, J. A., Bhadeshia, H. K. D. H., & Withers, P. J. (2007). Welding residual stresses in ferritic power plant steels. *Materials Science and Technology*, *23*(9), 1009-1020.

Fricke, S., Keim, E., & Schmidt, J. (2001). Numerical weld modeling—a method for calculating weld-induced residual stresses. *Nuclear Engineering and Design*, *206*(2), 139-150.

George, D., Kingston, E., & Smith, D. J. (2002). Measurement of through-thickness stresses using small holes. *The Journal of Strain Analysis for Engineering Design*, *37*(2), 125-139.

Gibbs, M. N. (1998). *Bayesian Gaussian processes for regression and classification* (Doctoral dissertation, University of Cambridge).

Gregorčič, G., & Lightbody, G. (2009). Gaussian process approach for modelling of nonlinear systems. *Engineering Applications of Artificial Intelligence*, *22*(4), 522-533.

Hashemian, H. M., & Bean, W. C. (2011). State-of-the-Art Predictive Maintenance Techniques. *Instrumentation and Measurement, IEEE Transactions on*, *60*(10), 3480-3492.

Hutchings, M. T., Withers, P. J., Holden, T. M., & Lorentzen, T. (2005). *Introduction to the characterization of residual stress by neutron diffraction*. Boca Raton, FL: CRC press.

Jiang, W., Luo, Y., Wang, B., Woo, W., & Tu, S. T. (2015). Neutron diffraction measurement and numerical simulation to study the effect of repair depth on residual stress in 316L stainless steel repair weld. *Journal of Pressure Vessel Technology*, *137*(4), 041406.

Kartal, M. E., Liljedahl, C. D. M., Gungor, S., Edwards, L., & Fitzpatrick, M. E. (2008). Determination of the profile of the complete residual stress tensor in a VPPA weld using the multi-axial contour method. *Acta Materialia*, *56*(16), 4417-4428.

Lei, Y., He, Z., & Zi, Y. (2009). Application of an intelligent classification method to mechanical fault diagnosis. *Expert Systems with Applications*, *36*(6), 9941-9948.

Lewis, J. R., Brooks, D., & Benson, M. L. (2017, July). Methods for Uncertainty Quantification and Comparison of Weld Residual Stress Measurements and Predictions. In *ASME 2017 Pressure Vessels and Piping Conference* (pp. V06BT06A069-V06BT06A069). American Society of Mechanical Engineers.

Liu, R., Kumar, A., Chen, Z., Agrawal, A., Sundararaghavan, V., & Choudhary, A. (2015). A predictive machine learning approach for microstructure optimization and materials design. *Scientific reports*, *5*.

Liu, X., Xuan, F. Z., Si, J., & Tu, S. T. (2008). Expert system for remnant life prediction of defected components under fatigue and creep–fatigue loadings. *Expert Systems with Applications*, *34*(1), 222-230.

Luenberger, D. G., & Ye, Y. (1984). *Linear and nonlinear programming* (Vol. 2). Reading, MA: Addison-Wesley.

Mahmoudi, A. H., Hossain, S., Truman, C. E., Smith, D. J., & Pavier, M. J. (2009). A new procedure to measure near yield residual stresses using the deep hole drilling technique. *Experimental Mechanics*, *49*(4), 595-604.

Mathew, J., Moat, R. J., & Bouchard, P. J. (2014, July). Optimised Neural Network Prediction of Residual Stress Profiles for Structural Integrity Assessment of Pipe Girth Welds. In *ASME 2014 Pressure Vessels and Piping Conference* (pp. V005T11A028-V005T11A028). American Society of Mechanical Engineers.

Mathew, J., Moat, R. J., & Bouchard, P. J. (2013, July). Prediction of Pipe Girth Weld Residual Stress Profiles Using Artificial Neural Networks. In *ASME 2013 Pressure Vessels and Piping Conference* (pp. V06BT06A075-V06BT06A075). American Society of Mechanical Engineers.

Mathew, J., Moat, R. J., Paddea, S., Fitzpatrick, M. E., & Bouchard, P. J. (2017a). Prediction of residual stresses in girth welded pipes using an artificial neural network approach. *International Journal of Pressure Vessels and Piping*, *150*, 89-95.

Mathew, J., Moat, R. J., Paddea, S., Francis, J. A., Fitzpatrick, M. E., & Bouchard, P. J. (2017b). Through-Thickness Residual Stress Profiles in Austenitic Stainless Steel Welds: A Combined Experimental and Prediction Study. *Metallurgical and Materials Transactions A*, *48*(12), 6178-6191.

MacKay, D. J. (1998). Introduction to Gaussian processes. *NATO ASI Series F Computer and Systems Sciences*, *168*, 133-166.

Mathias, M. H., De Medeiros, E. C., & Gonçalves, V. D. (2012, December). Design of a LabVIEW System Applied to Predictive Maintenance. In *Applied Mechanics and Materials*, 249, 208-212).

Maekawa, A., Kawahara, A., Serizawa, H., & Murakawa, H. (2016). Prediction of Weld Residual Stress in a Pressurized Water Reactor Pressurizer Surge Nozzle. *Journal of Pressure Vessel Technology*, *138*(2), 021401.

McCoy, K., Alamaniotis, M., & Jevremovic, T. (2013). A Conceptual Model for Integrative Monitoring of Nuclear Power Plants Operational Activities Based on Historical Nuclear Incidents and Accidents. *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, *1*(1), 69-81.

Mondal, A. K., Biswas, P., & Bag, S. (2017). Experimental and FE analysis of submerged arc weld induced residual stress and angular deformation of single and double sided fillet welded joint. *International Journal of Steel Structures*, *17*(1), 9-18.

Pilania, G., Wang, C., Jiang, X., Rajasekaran, S., & Ramprasad, R. (2013). Accelerating materials property predictions using machine learning. *Scientific reports*, *3*.

Prime, M. B. (2001). Cross-sectional mapping of residual stresses by measuring the surface contour after a cut. *Journal of Engineering Materials and Technology*, *123*(2), 162-168.

Quinonero-Candela, J., & Rasmussen, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *The Journal of Machine Learning Research*, *6*, 1939-1959.

Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for machine learning* (Vol. 1). Cambridge: MIT press.

Savsani, V., & Tawhid, M. A. (2017). Non-dominated sorting moth flame optimization (NS-MFO) for multi-objective problems. *Engineering Applications of Artificial Intelligence, 63*, 20-32.

Schwane, M., Kloppenborg, T., Khalifa, N. B., Jäger, A., & Tekkaya, A. E. (2014, February). Finite Element Based Determination and Optimization of Seam Weld Positions in Porthole Die Extrusion of Double Hollow Profile with Asymmetric Cross Section. In *Key Engineering Materials* (Vol. 585, pp. 95-102).

Song, S., Dong, P., & Pei, X. (2015). A full-field residual stress estimation scheme for fitness-for-service assessment of pipe girth welds: Part I – Identification of key parameters. International Journal of Pressure Vessels and Piping, 126–127, 58–70. http://doi.org/10.1016/j.ijpvp.2015.01.002

Stone, H. J., Bhadeshia, H. K., & Withers, P. J. (2008, April). In situ monitoring of weld transformations to control weld residual stresses. *Materials Science Forum*, 571, 393-398.

Tsoukalas, L. H., & Uhrig, R. E. (1997). *Fuzzy and neural approaches in engineering*. New York: John Wiley & Sons, Inc..

Vapnik, V. (2013). *The nature of statistical learning theory*. Berlin: Springer science & business media.

Wang, X., Gong, J., Zhao, Y., & Wang, Y. (2015). Prediction of Residual Stress Distributions in Welded Sections of P92 Pipes with Small Diameter and Thick Wall based on 3D Finite Element Simulation. *High Temperature Materials and Processes*, *34*(3), 227-236.

Wang, L., Liu, Z., Chen, C. P., Zhang, Y., Lee, S., & Chen, X. (2013). Fuzzy SVM learning control system considering time properties of biped walking samples. *Engineering Applications of Artificial Intelligence*, *26*(2), 757-765.

Withers, P. J. (2007). Residual stress and its role in failure. *Reports on progress in physics*, *70*(12), 2211.

Withers, P. J., & Bhadeshia, H. K. D. H. (2001). Residual stress. Part 2–Nature and origins. *Materials science and technology*, *17*(4), 366-375.

Wootton, A. J., Butcher, J. B., Kyriacou, T., Day, C. R., & Haycock, P. W. (2017). Structural health monitoring of a footbridge using Echo State Networks and NARMAX. *Engineering Applications of Artificial Intelligence*, *64*, 152-163.

Xie, X. F., Jiang, W., Luo, Y., Xu, S., Gong, J. M., & Tu, S. T. (2017). A model to predict the relaxation of weld residual stress by cyclic load: Experimental and finite element modeling. *International Journal of Fatigue*, *95*, 293-301.

Yildiz, A. R. (2013). Comparison of evolutionary-based optimization algorithms for structural design optimization. *Engineering applications of artificial intelligence*, *26*(1), 327-333.

Highlights:

- Application of probabilistic kernel learning machines for weld stress prediction
- Adoption of sixteen different kernel functions for prediction
- Benchmark of GP against Support Vector Regression
- Performance is measured by RMSE and Correlation Coefficient
- GP is shown to be robust and advantageous compared to SVR

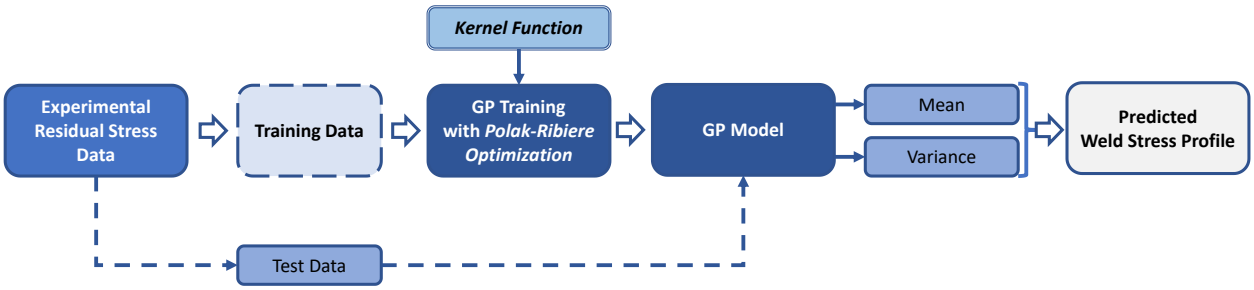

Fig. 1. Weld stress profile prediction method using kernel modeled Gaussian Process (GP).

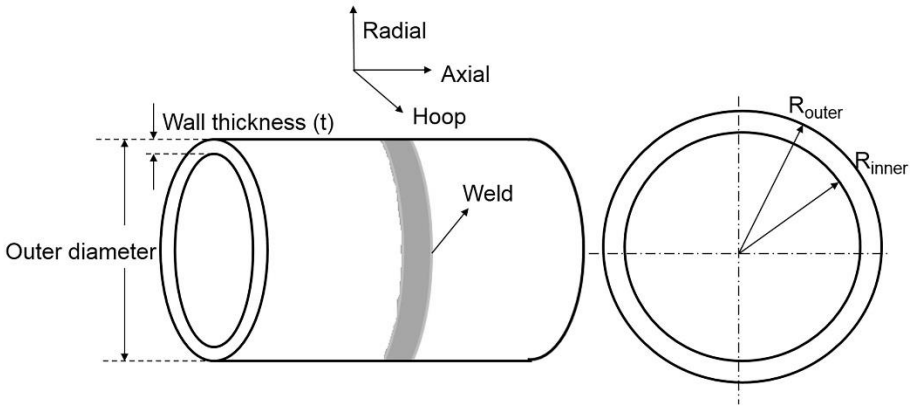

Fig. 2. Schematic illustration of a welded pipe showing the residual stress components.

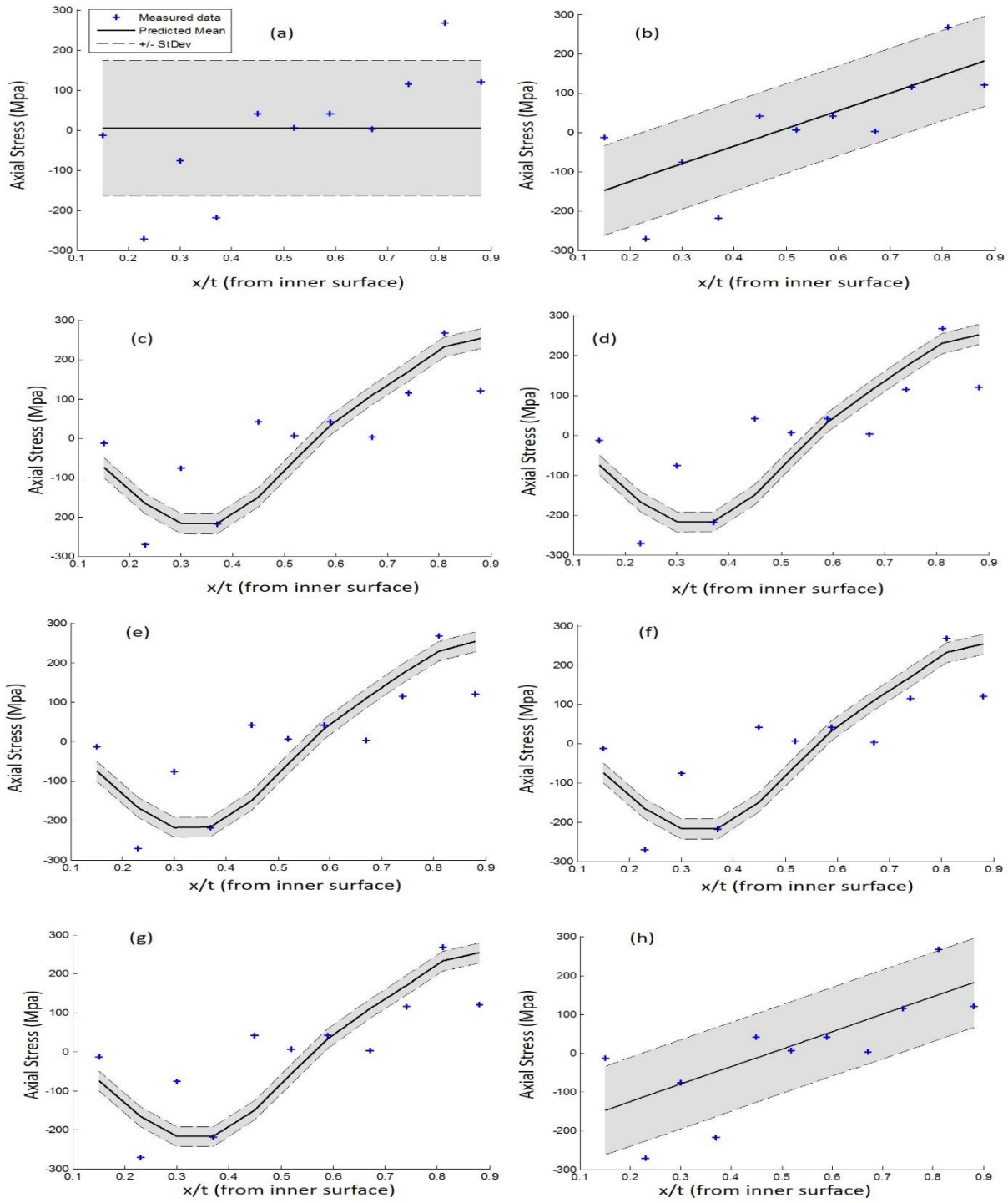

Fig. 3a. Axial residual stress profile prediction for dataset 1 obtained by a Gaussian Process (GP) equipped with a) Constant kernel, b) Linear kernel, c) Matérn kernel, d) Neural Network kernel, e) Sum of Linear and Neural Network kernel, f) Sum of Linear and Matérn kernel, g) Sum of Neural Network and Matérn kernel, h) Sum of Constant and Linear kernel,

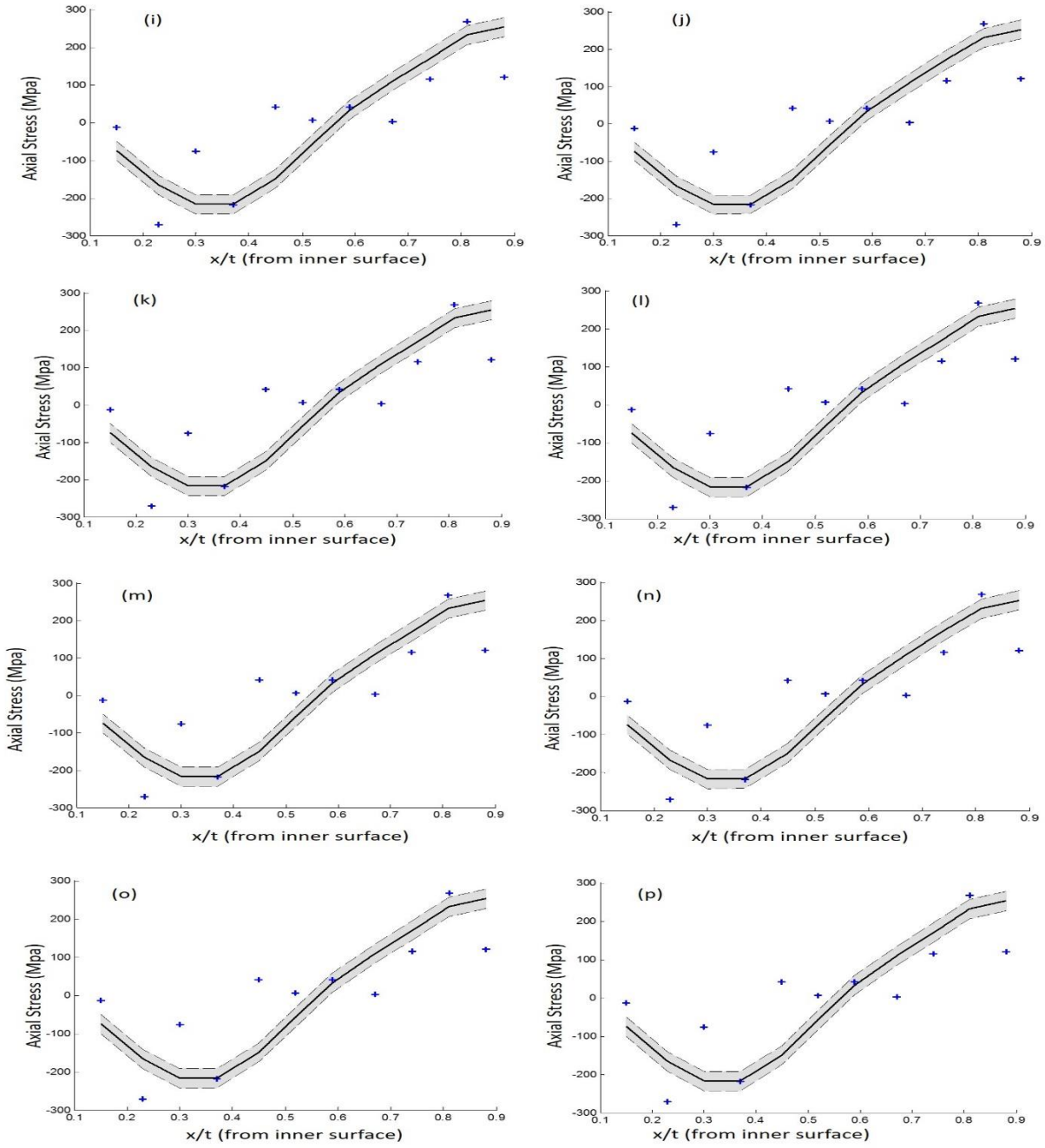

Fig. 3b. Axial residual stress profile prediction for dataset 1 obtained by a Gaussian Process (GP) equipped with i) Sum of Constant and Matérn, j) Sum of Constant and Neural Network, k) Sum of Constant, Linear and Matérn, l) Sum of Constant, Neural Network and Matérn, m) Sum of Linear, Neural Network and Matérn, n) Sum of Constant, Linear and Neural Network, o) Sum of Constant, Linear and Matérn, and p) Sum of Constant, Linear, Neural Network and Matérn.

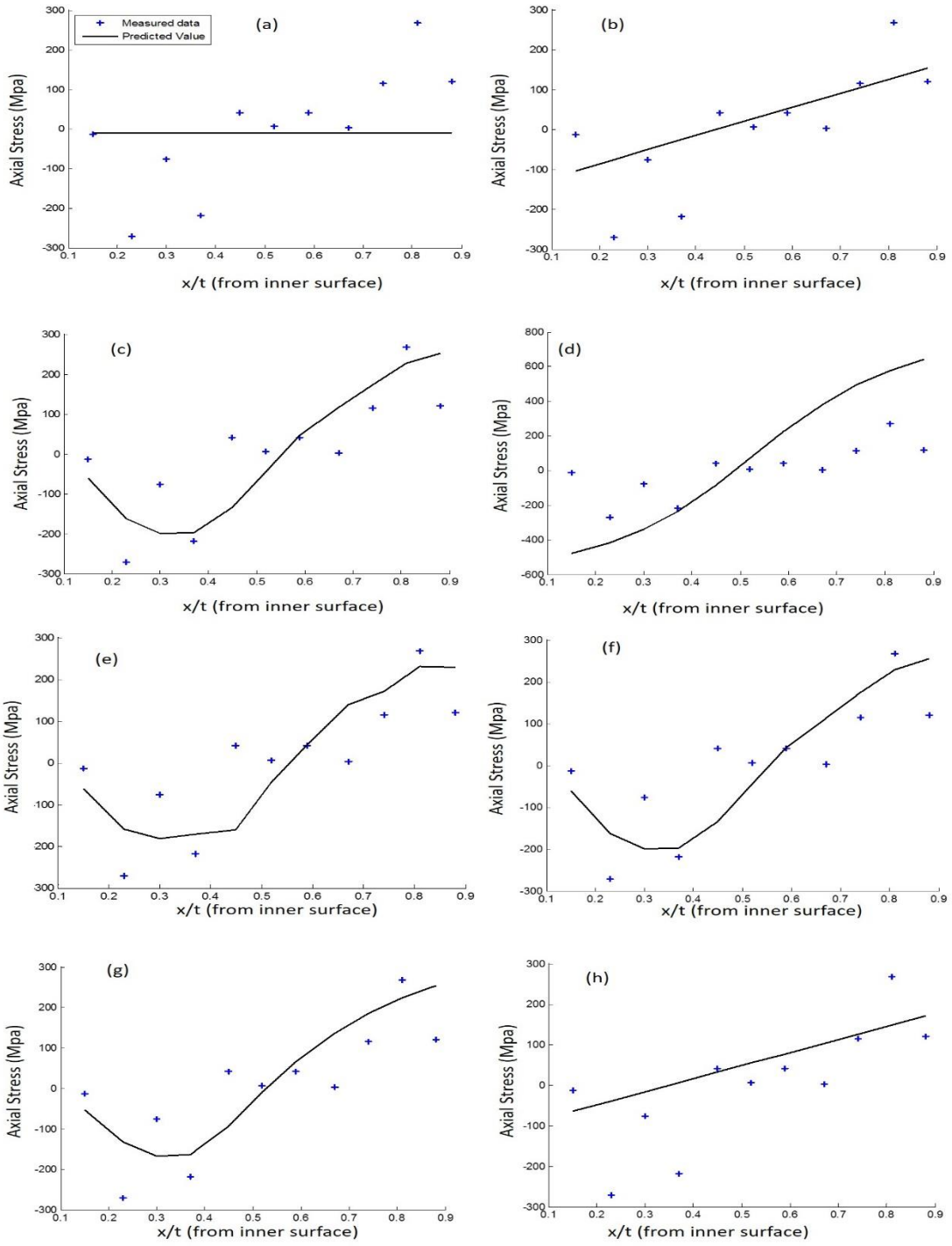

Fig. 4a. Axial residual stress profile prediction for dataset 1 obtained by Support Vector Regression (SVR) equipped with a) Constant kernel, b) Linear kernel, c) Matérn kernel, d) Neural Network kernel, e) Sum of Linear and Neural Network kernel, f) Sum of Linear and Matérn kernel, g) Sum of Neural Network and Matérn kernel, h) Sum of Constant and Linear kernel.

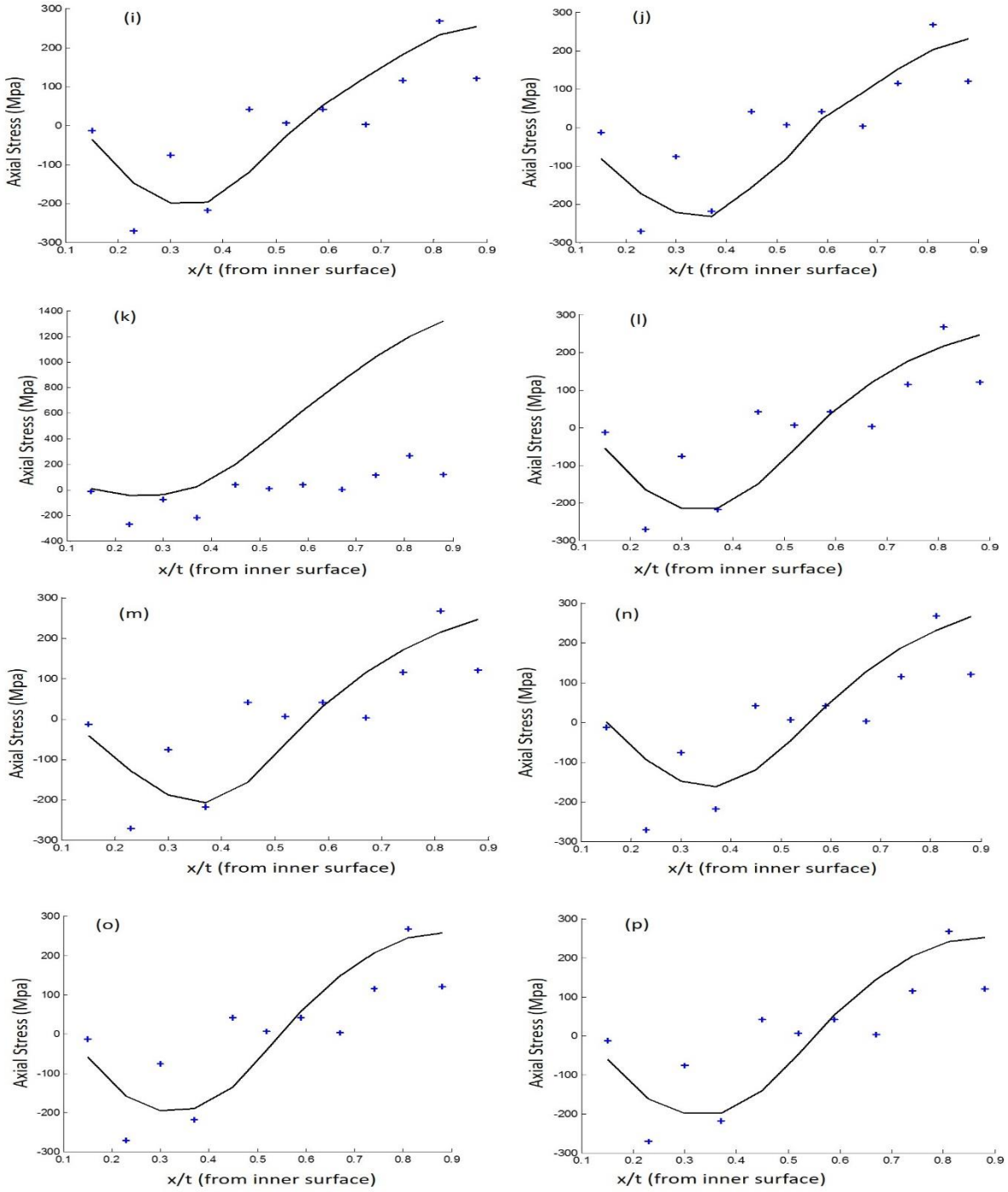

Fig. 4b. Axial residual stress profile prediction for dataset 1 obtained by Support Vector Regression (SVR) equipped with i) Sum of Constant and Matérn, j) Sum of Constant and Neural Network, k) Sum of Constant, Linear and Matérn, l) Sum of Constant, Neural Network and Matérn, m) Sum of Linear, Neural Network and Matérn, n) Sum of Constant, Linear and Neural Network, o) Sum of Constant, Linear and Matérn, and p) Sum of Constant, Linear, Neural Network and Matérn.

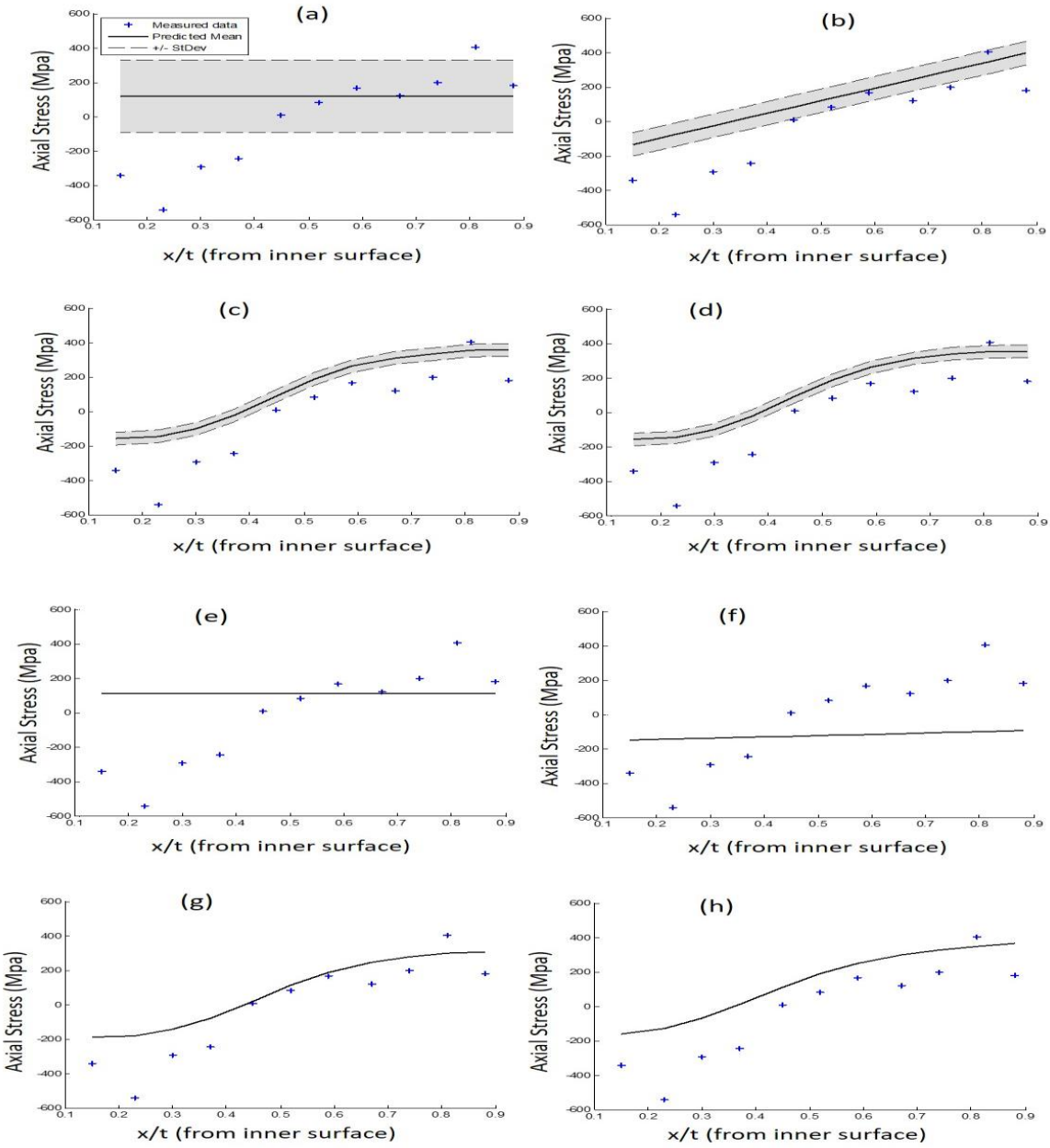

Fig. 5. Axial residual stress profile prediction for dataset 2 obtained by Gaussian Process (GP) equipped with a) Constant kernel, b) Linear kernel, c) Matérn kernel, and d) Neural Network kernel, compared to those obtained by Support Vector Regression with e) Constant kernel, f) Linear kernel, g) Matérn kernel, and h) Neural Network kernel.

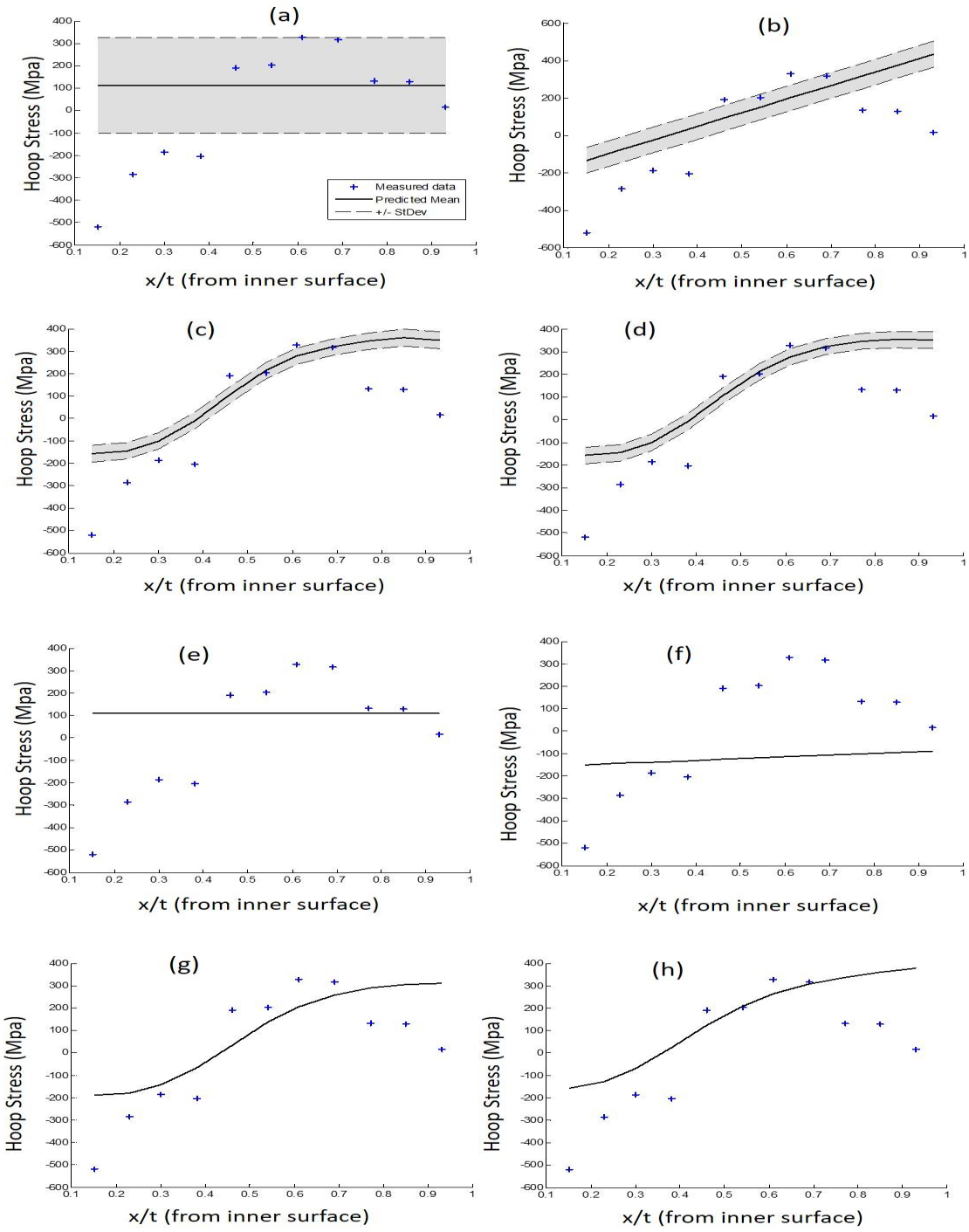

Fig. 6. Hoop residual stress profile prediction for dataset 1 obtained by Gaussian Process (GP) equipped with a) Constant kernel, b) Linear kernel, c) Matérn kernel, and d) Neural Network kernel, compared to those obtained by Support Vector Regression with e) Constant kernel, f) Linear kernel, g) Matérn kernel, and h) Neural Network kernel.

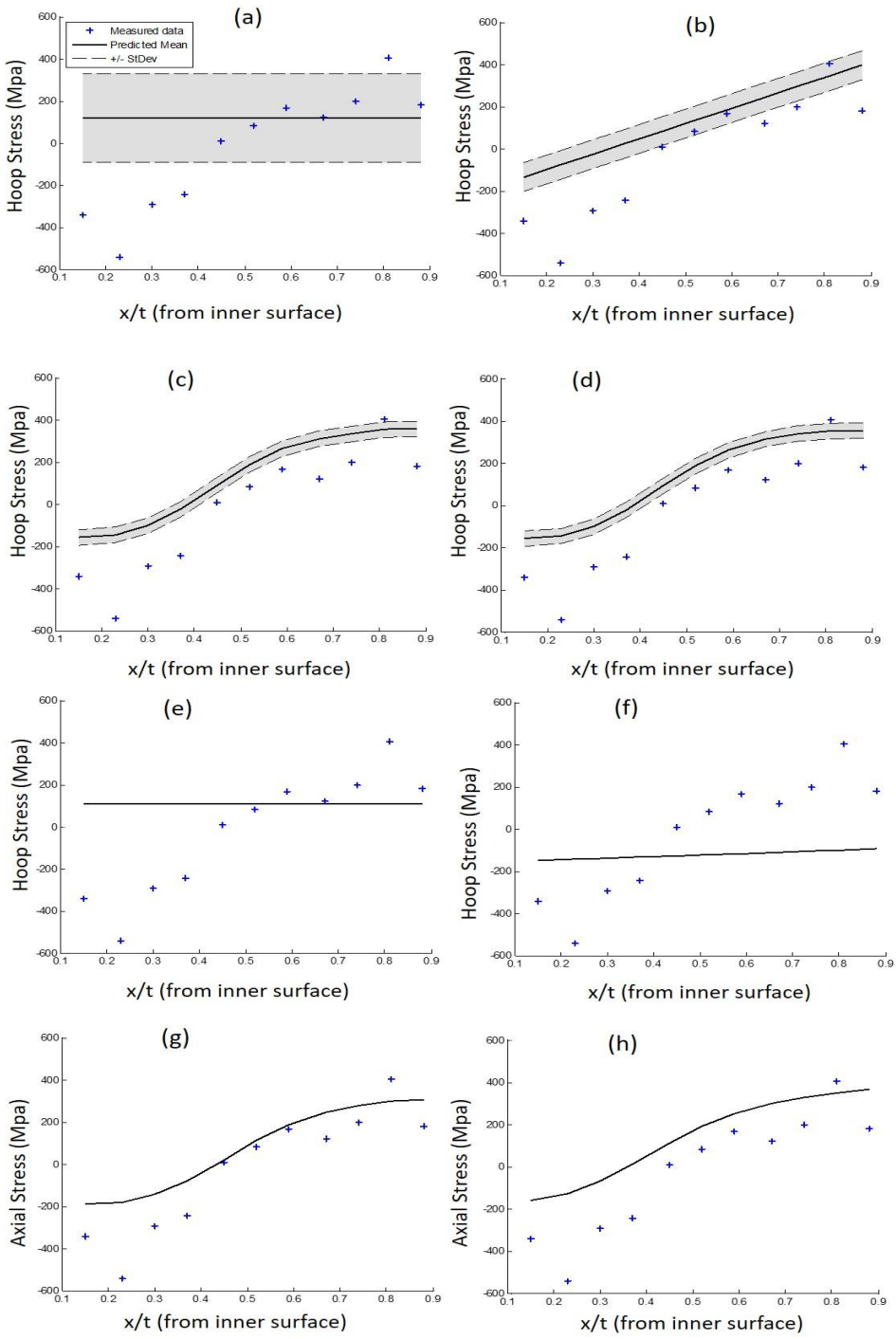

Fig. 7. Hoop residual stress profile prediction for dataset 2 obtained by Gaussian Process (GP) equipped with a) Constant kernel, b) Linear kernel, c) Matérn kernel, and d) Neural Network kernel, compared to those obtained by Support Vector Regression with e) Constant kernel, f) Linear kernel, g) Matérn kernel, and h) Neural Network kernel.

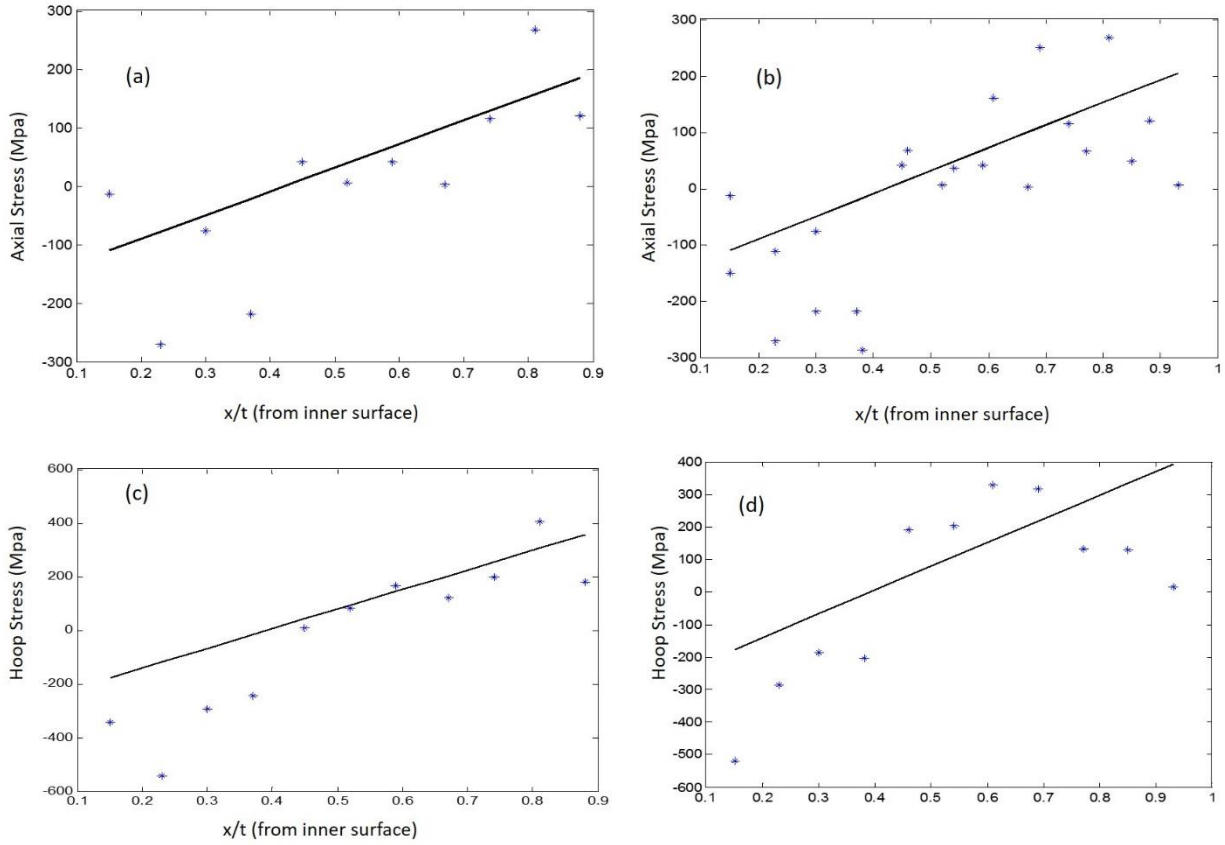

Fig. 8. a) Axial residual stress profile prediction for dataset 1 obtained by Multivariate Linear Regression (MLR), b) Axial residual stress profile prediction for dataset 2 obtained by Multivariate Linear Regression (MLR), c) Hoop residual stress profile prediction for dataset 1 obtained by Multivariate Linear Regression (MLR), d) Hoop residual stress profile prediction for dataset 2 obtained by Multivariate Linear Regression (MLR),

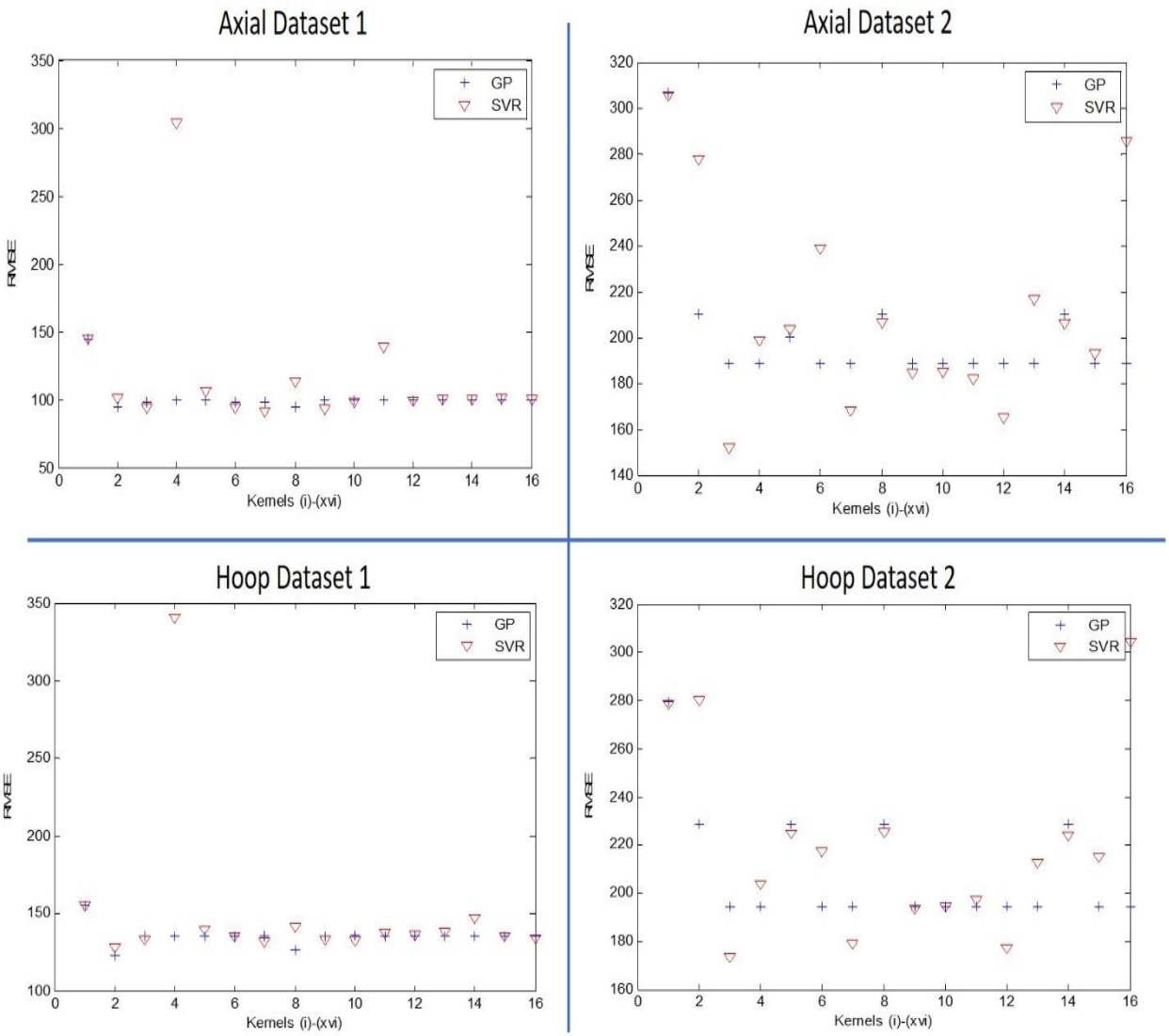

Fig. 9. Plot of RMSE values taken by Gaussian Process (GP) and Support Vector Regression (SVR) for residual stress profile prediction for dataset 1 and 2, aiming at exhibiting the variance of RMSE.

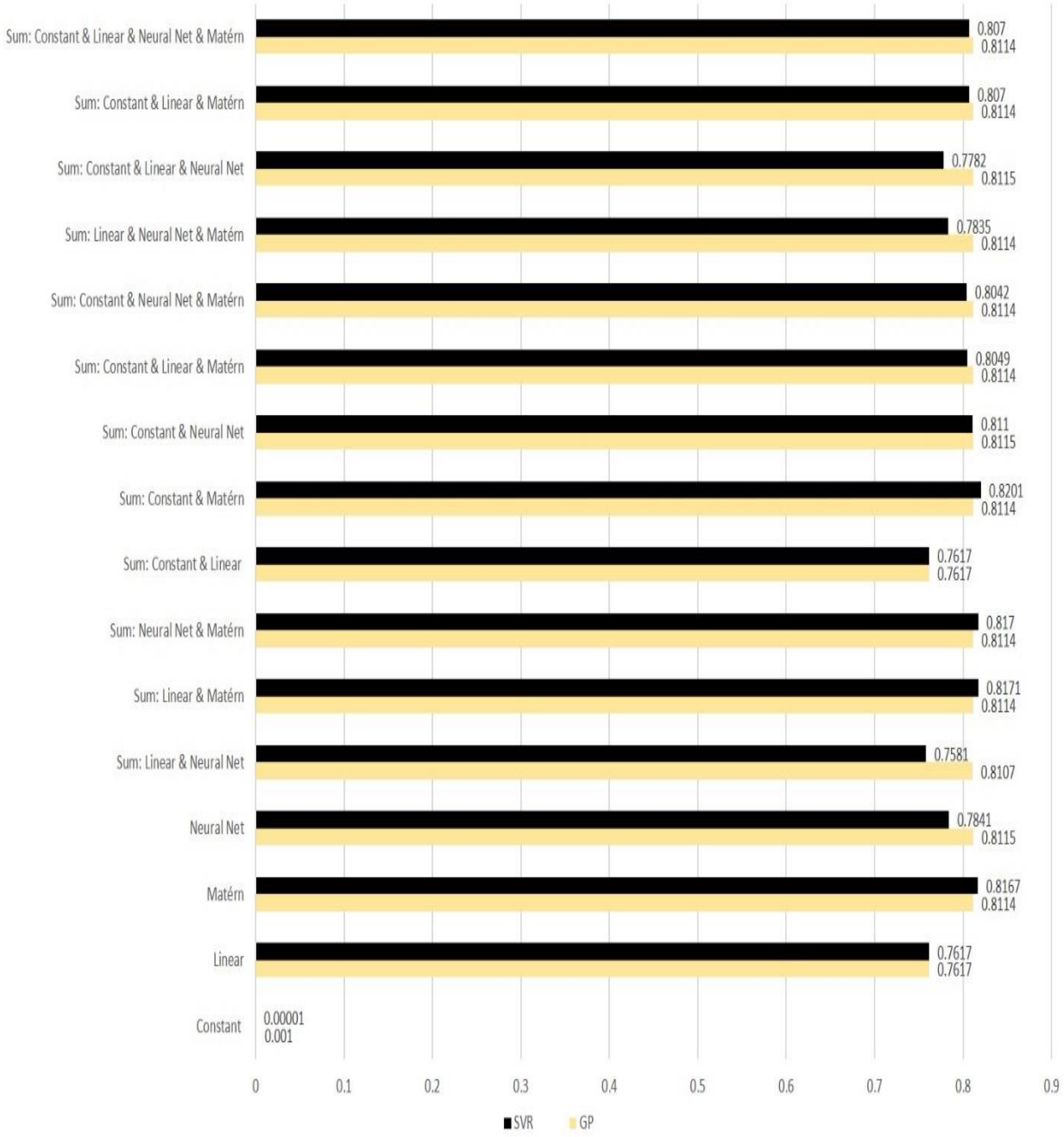

Fig.10. Average Correlation Coefficient values of ***axial*** *weld stress profile prediction* on datasets 1 and 2 obtained by GP and SVR with kernels (i)-(xvi).

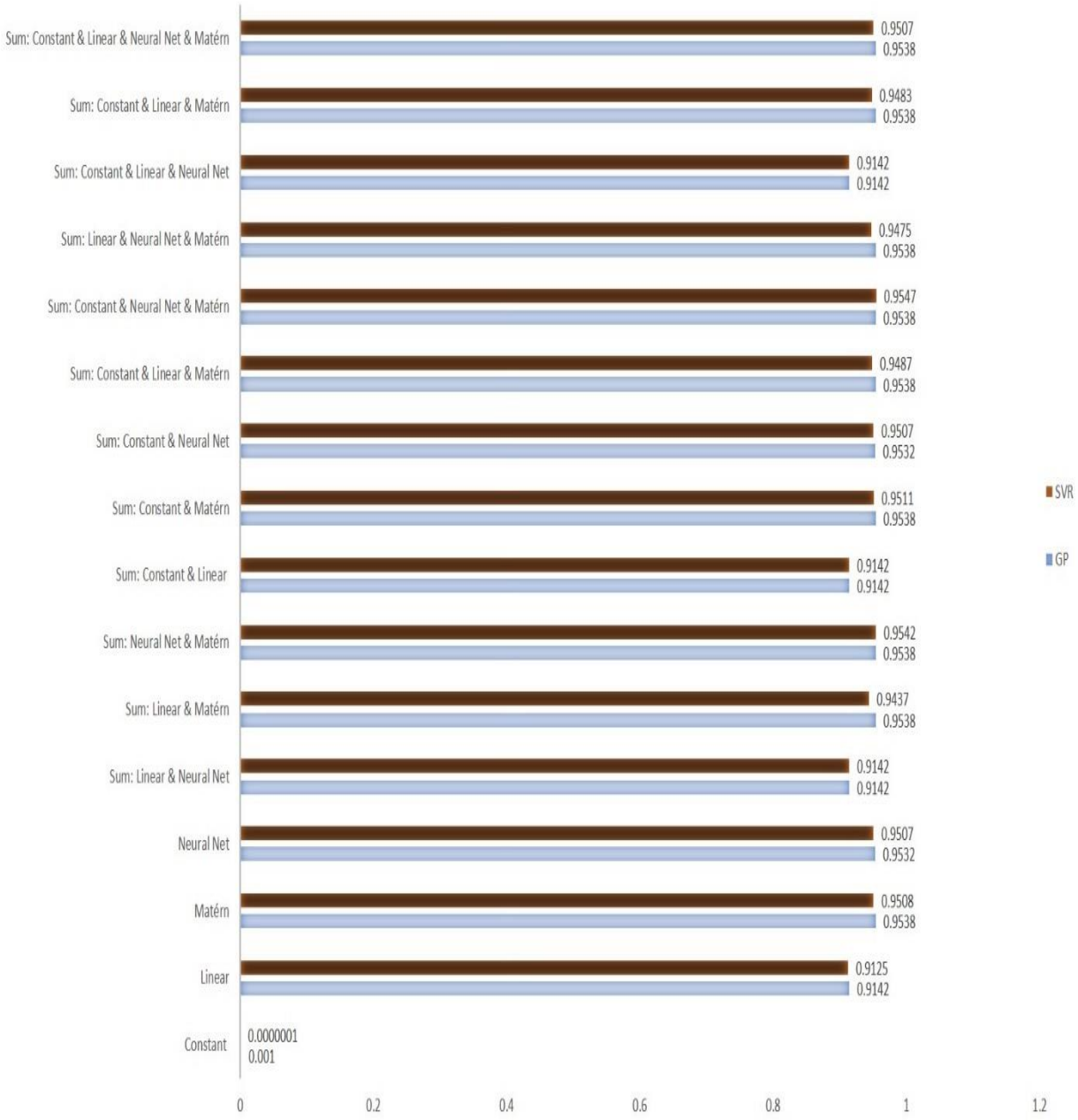

Fig. 11. Average Correlation Coefficient values of ***hoop*** *weld stress profile prediction* on datasets 1 and 2 obtained by Gaussian Process (GP) and Support Vector Regression (SVR) with kernels (i)-(xvi).

Table 1
Root Mean square error (RMSE) of Axial weld stress profile prediction using probabilistic kernel machines, i.e., kernel modeled Gaussian processes (GPs), and non-probabilistic kernel machines, i.e., SVR and MLR.

		RMSE for AXIAL WELD STRESS PROFILE PREDICTION					
ID	Type of Kernel	Dataset 1			Dataset 2		
		Gaussian Process (GP)	Support Vector Regression (SVR)	Multi Linear Regression (MLR)	Gaussian Process (GP)	Support Vector Regression (SVR)	Multi Linear Regression (MLR)
i	Constant	144.770	145.128	102.434	155.449	155.008	135.123
ii	Linear	95.471	101.228		123.026	128.298	
iii	Matérn	98.623	94.41		135.122	133.305	
iv	Neural Net	99.203	304.380		135.178	341.068	
v	Sum: Linear & Neural Net	99.381	106.701		135.277	139.700	
vi	Sum: Linear & Matérn	98.623	94.547		135.708	135.674	
vii	Sum: Neural Net & Matérn	98.622	91.936		135.710	131.954	
viii	Sum: Constant & Linear	95.477	113.398		126.420	141.002	
ix	Sum: Constant & Matérn	99.269	93.451		135.708	133.065	
x	Sum: Constant & Neural Net	99.202	99.116		136.115	132.863	
xi	Sum: Constant & Linear & Matérn	99.269	138.822		135.708	137.971	
xii	Sum: Constant & Neural Net & Matérn	99.269	99.427		135.708	136.069	
xiii	Sum: Linear & Neural Net & Matérn	99.272	100.540		135.699	138.073	
xiv	Sum: Constant & Linear & Neural Net	99.196	100.565		135.277	147.049	
xv	Sum: Constant & Linear & Matérn	99.269	100.889		135.708	135.644	
xvi	Sum: Constant & Linear & Neural Net & Matérn	99.271	100.801		135.708	134.156	

Table 2
Root Mean square error (RMSE) of Hoop weld stress profile prediction using probabilistic kernel machines, i.e., kernel modeled Gaussian processes, and non-probabilistic kernel machines, i.e., SVR, and MLR.

RMSE for HOOP WELD STRESS PROFILE PREDICTION							
ID	**Type of Kernel**	**Dataset 1**			**Dataset 2**		
		Gaussian Process (GP)	**Support Vector Regression (SVR)**	**Multi Linear Regression (MLR)**	**Gaussian Process (GP)**	**Support Vector Regression (SVR)**	**Multi Linear Regression (MLR)**
i	Constant	306.470	305.343	198.74	279.356	278.431	216.53
ii	Linear	210.047	277.428		228.648	279.890	
iii	Matérn	188.629	151.841		194.167	173.315	
iv	Neural Net	188.535	198.611		194.175	203.597	
v	Sum: Linear & Neural Net	200.283	203.224		228.648	224.721	
vi	Sum: Linear & Matérn	188.628	238.462		194.167	217.169	
vii	Sum: Neural Net & Matérn	188.629	168.324		194.164	178.810	
viii	Sum: Constant & Linear	210.047	206.460		228.648	225.375	
ix	Sum: Constant & Matérn	188.629	184.729		194.167	193.463	
x	Sum: Constant & Neural Net	188.537	185.065		194.176	194.573	
xi	Sum: Constant & Linear & Matérn	188.633	182.025		194.175	197.266	
xii	Sum: Constant & Neural Net & Matérn	188.629	165.241		194.167	177.298	
xiii	Sum: Linear & Neural Net & Matérn	188.629	216.577		194.168	212.736	
xiv	Sum: Constant & Linear & Neural Net	210.047	206.073		228.648	223.938	
xv	Sum: Constant & Linear & Matérn	188.626	192.868		194.166	214.937	
xvi	Sum: Constant & Linear & Neural Net & Matérn	188.629	285.655		194.170	304.361	

Table 3
Correlation Coefficient between real values and predicted profile by MLR.

MLR	Dataset 1	Dataset 2
	Average of Axial And Hoop Stress	**Average of Axial And Hoop Stress**
Correlation Coefficient	0.7943	0.9142