

# Motivated Support Vector Regression with Structural Prior Knowledge

Wei Zhang, Yao-Yu Li, Yi-Fan Zhu, Qun Li, and Wei-Ping Wang

**Abstract**—It's known that incorporating prior knowledge into support vector regression (SVR) can help to improve the approximation performance. Most of researches are concerned with the incorporation of knowledge in form of numerical relationships. Little work, however, has been done to incorporate the prior knowledge on the structural relationships among the variables (referred as to Structural Prior Knowledge, SPK). This paper explores the incorporation of SPK in SVR by constructing appropriate admissible support vector kernel (SV kernel) based on the properties of reproducing kernel (R.K). Three-levels specifications of SPK are studies with the corresponding sub-levels of prior knowledge that can be considered for the method. These include Hierarchical SPK (HSPK), Interactional SPK (ISPK) consisting of independence, global and local interaction, Functional SPK (FSPK) composed of exterior-FSPK and interior-FSPK. A convenient tool for describing the SPK, namely Description Matrix of SPK is introduced. Subsequently, a new SVR, namely Motivated Support Vector Regression (MSVR) whose structure is motivated in part by SPK, is proposed. Synthetic examples show that it is possible to incorporate a wide variety of SPK and helpful to improve the approximation performance in complex cases. The benefits of MSVR are finally shown on a real-life military application, Air-to-ground battle simulation, which shows great potential for MSVR to the complex military applications.

**Keywords**—admissible support vector kernel, reproducing kernel, structural prior knowledge, motivated support vector regression

## I. INTRODUCTION

**I**N non-linear regression, Support Vector Regression(SVR) [1] has been proven to possess many advantages, e.g. no local optima, good ability of generalization, intrinsic regularization and the sparseness of support vectors, etc. [2], and consequently be utilized in various applications [2]–[6].

More recently, the research on the incorporation of prior knowledge in SVR is well concerned. It's shown that prior knowledge can be incorporated by replacing the bias term of the kernel expansion with a combination of a set of independent basis functions [7]. The prior knowledge in the form of multiple polyhedral sets can also be incorporated into a reformulation of SVMs for improving the kernel approximation [8], [9]. A prior knowledge on the derivatives can be incorporated into the quadratic (QP) [9] and linear programming(LP) [10] formulation of the SVR. A tutorial can

be seen in [10]. All the related methods, however, which allow the use of prior knowledge for SVR, focus on the knowledge in form of numerical correlations (referred as to Numerical Prior Knowledge, NPK), such as equality, inequality and simple IF-THEN rules on numerical values [8], [11], etc. The strength of the incorporation of NPK lies on its simplicity and generality as it amounts to the addition of some equality or inequality constraints to the QP or LP of SVR. It can handle problems where conventional data may be few or not available and improve the approximation performance.

It's known that SVR is to learn an unknown function based only on a training set of  $N$  input-output pairs  $\{x_i, y_i\}, i = 1, \dots, N$  in a black box modeling approach [10](as shown in Fig. 1(a)). Sometimes, however, modelers and analysts may prefer polynomials to SVR due to its inconvincibility on explaining how the results have been obtained. For example, high-level decision-makers, especially in military fields, avoid making decisions based on models that they do not fully understand, especially when they partially know the phenomena, such as the dimensions or given cause-effect correlation among some factors. The weakness of SVR, i.e. lack of comprehensibility and interpretability in the structure of regression model, hinders its applicability in complex problems.

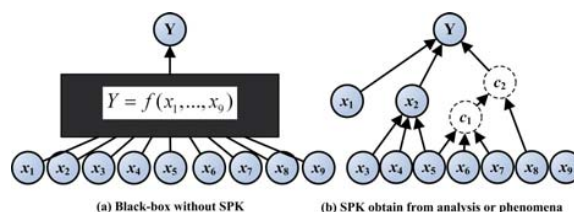


Fig. 1. An illustration of SPK in modeling

Little work, however, has been done to incorporate the prior knowledge on the structural relationships among the variables (which is called as Structural Prior Knowledge, SPK, in this paper) into SVR. Actually, in real world applications, such as system identification and military simulations, some information is usually known beforehand. For example, the model has one output  $Y$ , and modelers may handpick several inputs of most interest from dozens of even hundreds of factors, of which there are 9(the  $x$ 's). Sometimes, some important information, such as the hierarchy, interaction, and some knowledge on the analytical forms, can be obtained from problem analysis or phenomenological knowledge (as shown in Fig1(b), which will be discussed in more detail later).

Compared with NPK, the SPK can only be incorporated into SVR by configuring the kernel function, which includes

Wei Zhang is with the College of Information System and management, National University of Defense Technology (NUDT), Changsha, Hunan, 410073, China (e-mail: the\_ant@163.com).

Yao-Yu Li is with the College of Information System and Management, NUDT, Changsha, Hunan, China. (e-mail: Garett\_1984@hotmail.com).

Yi-Fan Zhu is with the College of Information System and management, NUDT, Changsha, Hunan, China (e-mail: nudtzyf@hotmail.com).

Qun Li is with the College of Information System and management, NUDT, Changsha, Hunan, China.

Wei-Ping Wang is with the Graduate School, NUDT, Changsha, Hunan, China, (e-mail: wangwp@nudt.edu.cn).

the modification of structure of inputs and the configuration of an appropriate kernel function. Some researchers with the same concerns would argue for improving the sophistication of conventional SV algorithm, and much is probably possible by doing so. For example, Schölkopf *et al.* [12] showed that the knowledge on invariance and image locality can be used to design kernel functions in image classification tasks. However the prior knowledge still belongs to NPK and is hard to extend to the regression tasks. Furthermore, conventional admissible support vector kernels (SV kernels) [13] perform an equal treatment for all the input dimensions. That is regarded as advantageous in many respects, especially when the prior knowledge is of lack or inconsistency, because it corresponds to “allowing the data to speak”, rather than biasing results with one or another theory [14]. Such disregard for the SPK in training SVR, however, may generate redundant computation and uninterpretability of results, and consequently mislead the high-lever analysis.

Reproducing kernel, which is proven to be a SV kernel [15], has many important properties, e.g. unique existence, positive definiteness, convergence, projection etc. in reproducing kernel Hilbert space (RKHS) [16]. Complex R.Ks can be composed by some simpler ones through the direct sum and tensor product, which is usually employed in convolution kernel [17] to handle the sets whose elements are discrete structures, such as strings, trees and graphs. The composition approach is more general that it encompasses the methods for compositing SV kernels through nonnegative linear combination or product [18], [19]. It's possible, with the properties of direct sum and tensor product, to group the inputs and make differential treatment of each group with a specific R.K based on SPK.

This paper describes the hierarchy of SPK specifications and proposes a way for the incorporation of SPK into SVR (referred as to *Motivated Support Vector Regression*, MSVR) by compositing appreciate SV kernels based on R.K. Finally, a systematic comparison between the new R.Ks designed by the proposed method and the conventional SV kernels, i.e. polynomial kernel (PK) and Gaussian kernel (RBF) are presented under three synthetic problems and an Air-to-ground battle simulation. The experimental results illustrate that the MSVR outperform conventional SVR with any single SV kernel in terms of accuracy and efficiency.

## II. PRELIMINARY

### A. Formulation of Standard Support Vector Regression

Given an training set  $\mathcal{D} = \{(x_i, y_i), i = 1, \dots, l\} \subset \Omega \times \mathbb{R}$ , where  $\Omega$  denotes the space of the input data (e.g.  $\Omega = \mathbb{R}^d$ , where  $d$  denotes dimension of input). SVR aims at training a model of the form  $y = \langle w, \phi(x) \rangle + b$ , which minimizes a general risk function as follows:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l L(y_i, f(x_i)) \quad (1)$$

where  $w$  controls the flatness of the model,  $\phi(x)$  is a mapping function,  $b$  is the bias,  $\langle \cdot, \cdot \rangle$  denotes the dot product, constant  $C > 0$  determines the trade-off between error minimization and the maximization of the function flatness. In this

paper, the  $\varepsilon$ -insensitive loss function  $L_\varepsilon$  [1] is used, i.e.,

$$L_\varepsilon(y, f(x)) = |y - f(x)|_\varepsilon = \max\{0, |f(x) - y| - \varepsilon\} \quad (2)$$

where  $\varepsilon \geq 0$  is a constant controlling the noise tolerances.

It's well-known that SVR can be formulated as the following quadratic programming (QP) problem [13] which can be solved efficiently by many well-documented optimization algorithms:

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \\ & + \sum_{i=1}^l (\alpha_i + \alpha_i^*)\varepsilon - \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)y_i \\ \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C], i = 1, \dots, l \end{aligned} \quad (3)$$

Consequently, the regression model takes a form as follows:

$$f(x) = \sum_{i \in SV} (\bar{\alpha}_i - \bar{\alpha}_i^*)K(x_i, x) + b \quad (4)$$

where  $i \in SV$  denotes the indices of support vectors (SVs), i.e.  $x_i$  with nonzero Lagrange multiplier  $\bar{\alpha}_i$  or  $\bar{\alpha}_i^*$ ,  $K(\cdot, \cdot)$  is the kernel function, which can be cast in terms of dot products of a mapping function, i.e.  $K(s, t) = \langle \Phi(s), \Phi(t) \rangle$ , where  $\Phi(\cdot)$  is a mapping function.

Obviously, the complexity of (4) depends only on the amount of SVs (ASV) and SV kernel rather than the dimensionality of the input space  $\Omega$ . In practice, the SVs, which depend on the selection of kernel and coefficients of SV algorithm [20], can be automatically extracted by SV algorithm. In other words, the major task of the SV algorithm lies in the selection of its kernel [19].

### B. Definition of Reproducing Kernel

The theory of reproducing kernel Hilbert space (RKHS) has been developed for years [16] before the SVM was introduced [1]. It is a rigorous and effective framework for smooth multivariate interpolation of arbitrarily scattered data [21] and accurate approximation of general multidimensional functions [22]. RKHS owes the name to the so-called reproducing kernel (R.K). In this section, some basic concepts are introduced briefly. For more details on RKHS see e.g. [16], [23], [24].

*Definition 1:* Let  $\Omega \subseteq \mathbb{R}^d$  be an arbitrary nonempty set,  $\mathcal{H}$  is a Hilbert space of function  $f : \Omega \rightarrow \mathbb{R}$  (short for  $f \in \mathbb{R}^\Omega$ ). We call that  $\mathcal{H}$  is a reproducing kernel Hilbert space (RKHS) if there exists  $K : \Omega \times \Omega \rightarrow \mathbb{R}$ , satisfies the following:

- (i)  $\forall x, K_x(y) = K(y, x)$  as a function of  $y$  belongs to  $\mathcal{H}$ .
- (ii) The reproducing property:  $\forall x \in \Omega$ , and  $\forall f \in \mathcal{H}$ ,

$$f(x) = \langle f, K_x \rangle \quad (5)$$

(iii)  $\mathcal{H}$  is spanned by  $K$ , that is,  $\mathcal{H} = \overline{\text{span}\{K_x(\cdot) | x \in \Omega\}}$ . Here,  $\mathcal{H}$  is called the native space [25] of  $K$  (short for  $\mathcal{H}_K(\Omega)$ ).

*Definition 2:*  $K : \Omega \times \Omega \rightarrow \mathbb{R}$  is called a R.K of  $\mathcal{H}$ , if it satisfies the conditions (i) and (ii) in 1.

Note that applying (5) to function  $K_x$  at  $y$ , we get

$$K(x, y) = K_x(y) = \langle K_x, K_y \rangle, \forall x, y \in \Omega \quad (6)$$

which implies that the nonlinear map function takes the form of R.K with any fixed  $x$  or  $y$ . Whereas the conventional SV kernels are incapable of presenting the map functions explicitly. As a result, the features of images of data can be analyzed further.

Aronszajn [16] proved that simpler R.Ks can be contributed to compose more complex R.Ks by sum and product operators, that is, let  $K_i (i = 1, 2)$  is the R.K of the RKHS  $\mathcal{H}_i$  in same set  $\Omega$  with the norms  $\|\cdot\|_i$ , then

*Property 1:*  $K = K_1 + K_2$  is the R.K of a RKHS  $\mathcal{H}$  of all functions  $f = f_1 + f_2$  with  $f_i \in \mathcal{H}_i, i = 1, 2$ , and with the norm defined by  $\|f\|^2 = \min\{\|f_1\|_1^2, \|f_2\|_2^2\}$ , i.e. the minimum taken for all the decompositions  $f = f_1 + f_2$  with  $f_i \in \mathcal{H}_i, i = 1, 2$ .

Note that the property can be extended to the case where  $K = \sum_{i=1}^n K_i$ . In addition, the difference of R.Ks is also a R.K; more details see [16] for reference as well.

*Property 2:* The direct product of  $\mathcal{H}_1$  and  $\mathcal{H}_2$  possesses a R.K  $K(x_1, x_2, y_1, y_2) = K_1(x_1, y_1)K_2(x_2, y_2)$ .

Similarly to Property 1, the product property can be extended to the case where  $K = \prod_{i=1}^n K_i$ . More detail see [16] for reference.

### C. Relations between SV Kernel and Reproducing Kernel

It's necessary to discuss the relations between various kernels to validate that the R.K can be used as a SV kernel. It is hoped that the discussion here would help to bridge the conceptual gap between some familiar kernels, e.g. positive (semi-)definite kernel (PDK), Mercer kernel and R.K, whereas some of the observations are not new or profound.

*Definition 3:* Let  $\Omega$  be a subset of  $\mathbb{R}^n, n \in \mathbb{N}, K : \Omega \times \Omega \rightarrow \mathbb{R}$  is symmetric and positive (semi-)definite (PD), if and only if for arbitrary finite sets  $\{x_1, \dots, x_m\} \subseteq \Omega$ , the matrix  $\mathbf{K} = (K(x_i, x_j))_{1 \leq i, j \leq m}$  is symmetric and positive definite, i.e.  $\forall m \in \mathbb{N}, \forall c_i \in \mathbb{R}$ , for any  $x_1, \dots, x_m \in \Omega, i = 1, \dots, m, K$  satisfies the following inequality

$$\sum_{i,j=1}^m c_i c_j K(x_i, x_j) \geq 0 \quad (7)$$

*Theorem 1:*  $K : \Omega \times \Omega \rightarrow \mathbb{R}$  is a SV kernel iff  $K$  is a PDK.

The proof is obvious. Refer to e.g. [15], [16].

*Theorem 2:*  $K : \Omega \times \Omega \rightarrow \mathbb{R}$  is a Mercer kernel iff  $K$  is a PDK.

Proof: if  $K$  is a Mercer kernel, i.e. there exists a map function  $\Phi$  such that  $K(t, s) = \langle \Phi(s), \Phi(t) \rangle$ . Then,

$$\begin{aligned} \sum_{i,j=1}^m c_i c_j K(x_i, x_j) &= \sum_{i,j=1}^m c_i c_j \langle \Phi(x_i), \Phi(x_j) \rangle \\ &= \left\| \sum_{i=1}^m c_i \Phi(x_i) \right\|^2 \geq 0 \end{aligned}$$

thus,  $K$  is a PDK according to (7).

For the converse, if  $K$  is a PDK,  $K$  is a Mercer kernel according to Theorem 1 and Mercer's Theorem [26], which completes the proof.

*Theorem 3:*  $K : \Omega \times \Omega \rightarrow \mathbb{R}$  is a Mercer kernel iff there exists a RKHS  $\mathcal{H}$  with R.K  $K$ , i.e.  $\mathcal{H}_K(\Omega)$ .

Proof: According to Moore–Aronszajn Theorem [16], any PDK  $K$  is associated with a space  $\mathcal{H}_K(\Omega)$  and vice versa. Note that the Theorem 2 holds if  $K$  is a PDK, that is,  $K$  is a Mercer kernel, which completes the proof.

## III. MOTIVATED SUPPORT VECTOR REGRESSION WITH STRUCTURAL PRIOR KNOWLEDGE

### A. Motivation

Decision-makers, especially working on high-level issues, often seek a robust logic for their choices, which makes sense to themselves and can be explained to others. They may prefer a “roughly right” formula that displays issues transparently. In other word, the regression models can “tell a story”, which explains why the model behaves as it does [14], that is, the model should be physically meaningful and interpretable. The “story” can appear with many forms, such as history data, logic, rules, phenomena, even expertise or highly subjective and personal insights. Some of them, which can be expressed in the form of data, are included in the NPK, while some are tacit knowledge. The latter, sometimes, is highly personal and hard to formalize, making it difficult to employ in regression directly.

Since the conventional SVRs are mathematical constructs, often with little if any intuitive value to decision-makers. In this section, a new SVR, namely motivated support vector regression (MSVR), whose structure is motivated in part by phenomenological considerations, is presented. The SPK, such as knowledge based on simplified physical reasoning and dimensional analysis, are employed to aid to postulate a structural form for the regression model. Therefore, it's necessary to present the specifications of SPK firstly.

### B. Specifications of Structural Prior Knowledge and Description Matrix

As shown in Fig.2, derived from Fig.1(b), there are three levels of SPK, that is,

#### Level 2: Hierarchical SPK (HSPK).

HSPK, shown in Fig.2(a), is the highest level of SPK. It roughly describes the relations between inputs at different levels of resolution without detailed interaction. The inputs may come from the same object model, but differ in the abstraction level or perspective with which they describe a problem, i.e. they usually appear in a multi-resolution, multi-perspective model [27] used for military or political-military decision aids, defense planning and machine intelligence. For example, in an analysis of air-to-ground attack, inputs may blend some higher resolution variables (e.g. efficiency of read weapons in initial phase ( $x_3$ ), efficiency of read weapons in final phase ( $x_4$ ), time of suppression of enemy air defenses (SEAD) ( $x_5$ )) with a lower-level variable, e.g. average efficiency of weapon ( $x_2$ ), i.e. the amount of Blue target destroyed per Red attack unit per time unit(day). Therefore, it's necessary to distinguish the inputs at different levels before training regression model in order to avoid the inconsistency. In other word, the inputs should contain  $x_2$  or  $x_3, x_4$  and  $x_5$  exclusively depending on the analysts' needs.

#### Level 1: Interactional SPK (ISPK).

ISPK is a kind of SPK in the middle level who describes the dependency among inputs and the ones between input and output. It can bring us two types of information, i.e., 1) grouping information of inputs; 2) independence information between input and output. For instance,  $x_6$  to  $x_8$  should be

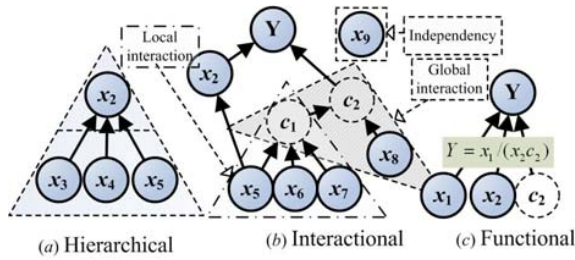


Fig. 2. Typical types of SPK

divided into two different groups related with output  $Y$ , while  $x_9$  is another group which is independent to  $Y$ , as shown in Fig.2(b). This type of SPK can be derived from dimension analysis, sensitivity analysis, correlation analysis, or a single-subject experiment. For example, if the output being calculated is a distance, and inputs include various times and an average speed, then it is reasonable to construct composite variables (i.e.  $c_1$  and  $c_2$  shown in Fig.2(b), which can be regarded as aggregation fragments as a HSPK) with the dimensions of distance, e.g. times multiplied by an average speed, to be candidate regression variables. Consequently, ISPK can be divided into three sublevels, i.e.,

1) **Independence** (e.g.  $x_9$ );

2) **Global interaction**: the interaction between the potential composite factors or that between composite factor and independent variable (e.g. the interaction between composite factor  $c_1$  and  $x_8$ , or the interaction among  $x_1, x_2, c_2$ );

3) **Local interaction**: the interaction between the independent variables (e.g. the interaction among  $x_6, x_7$  and  $x_8$ , which can be replaced by a composite factor  $c_1$ );

ISPK can, in quantity sense, sort or reduce the input variables, and group them to proceed to the further design of appropriate SV kernel. Consequently, it can, in some cases, indeed greatly improve "the story" and avoid errors associated with the nonlinearities within the black-box models [14]. It also somewhat improves average accuracy of the model and reduced the computation effect, but that was less important.

#### Level 0: Functional SPK (FSPK)

FSPK is in the lowest level which describes the relatively detailed correlations between inputs and output. It can be expressed in form of some specific mathematical function (e.g. polynomial), statistical function (e.g.  $\min$ ,  $\max$ ), piecewise function (which can identify important branches) or some information which can be expressed by familiar functions (e.g. periodicity v.s.  $\sin$ , approximability v.s.  $\exp$ , etc.).

FSPK can be divided into two sub-levels, i.e.

1) **Exterior-FSPK**: function relations among all groups to the output;

2) **Interior-FSPK**: function relations between a single group and the output;

The former contributes to obtain a global regression model which can reflect the phenomena (e.g. as shown in Fig.2(c), the output  $Y$  is of a form:  $Y = x_1/(x_2 c_2)$ ), and the latter helps to construct a proper SV kernel for a local regression model (e.g. a local model in form of  $c_1 = f_{\text{local}}(x_5, x_6, x_7)$ ).

To assist in getting more clear understanding of the SPK

with a systematic and graphical overview of its essential types, a description matrix of SPK is presented as shown in Table I. The horizontal axis identifies the main levels of SPK and corresponding detailed sub-levels.

All the columns should cover the specifications of SPK in every level and sub-levels.

1) The first column should cover the detailed hierarchy of all the inputs of interest and specify which hierarchy is the following analysis on.

2) The next three columns cover the ISPK in the sequence of independence, global interaction and local interaction, whose independent variables in regression analysis should be in the same hierarchy specified in the HSPK column. The corresponding contents should cover the input names and the information whether composite variables should be used. If so, the matrix should describe why and how.

3) The last two columns indicate the FSPK for exterior and interior information. They are the most important part of MSVR for postulating structural form motivated in part by phenomenological considerations. Therefore, they should be described relatively in detail. The corresponding contents should indicate the specifications of function form and conditions, as well as the SV algorithm.

TABLE I  
DESCRIPTION MATRIX OF SPK

Problem No.	Level 2: HSPK	Level 1: ISPK			Level 0: FSPK	
		Indep.	Global	Local	Exterior	Interior
A(output) <sup>a</sup>						

<sup>a</sup>Footnote: describe the symbols used in the matrix for facilitating the sharing and communication.

Applying the matrix is a means to make a complete inventory of the SPK are located in regression tasks and how they can be typified in terms of SPK levels. one should be aware that some information can manifest itself in various forms simultaneously. For example, the global interaction may describe the local interaction (e.g. interaction between  $x_5, x_6$  and  $x_7$ ), while it is higher level interaction among the composite variables and between composite variable and independent variable (e.g. interaction between  $c_1$  and  $x_8$  or  $c_2$  and  $x_2$ ). In filling in the matrix, one should be also aware that the content of SPK should facilitate the sharing and communication (e.g. make use of a specific document or footnote).

#### C. New method for composition SV kernel with SPK

As stated in previous section, the R.K can be used as a SV kernel, and possesses some good properties for composing complex R.Ks. However, there is usually no R.K as free lunch for the regression problems to be dealt with. For example, the R.K(6) can only be used in the SV algorithm with input space  $\Omega'$ , where  $\Omega' \subseteq \Omega$ , and the dimension of  $\Omega'$  is equal to that of  $\Omega$ .

The question raises now is, whether a R.K can be employed when the dimension of  $\Omega'$  is bigger than that of  $\Omega$ . The direct sum case of two R.Ks in the Property 1 is firstly considered:

**Theorem 4:**  $K(x, y) \triangleq a_1 K_1(x_1, y_1) \oplus a_2 K_2(x_2, y_2)$  is a SV kernel, where  $x, y \in \Omega$ ,  $x = (x_1, x_2)$ ,  $y = (y_1, y_2)$ ,  $x_i, y_i \in \Omega_i$ ,  $\Omega_1 \oplus \Omega_2 = \Omega$ ,  $a_i \geq 0$ , and  $K_i$  is a R.K,  $i = 1, 2$ .

**Proof:** Noted that  $K_s$  is a PDK since it is a R.K from the Theorem 2, 3 and Definition 3. In other words,  $\forall m \in \mathbb{N}$ ,  $x_i^{(1)}, \dots, x_i^{(m)} \in \Omega_i, \forall c_j \in \mathbb{R}, j = 1, \dots, m$ , then

$$\sum_{i,j=1}^m c_i c_j K_s(x_s^{(i)}, x_s^{(j)}) \geq 0 \quad (8)$$

Since  $\Omega_1 \oplus \Omega_2 = \Omega$ , then there is a unique  $x^{(k)} = (x_1^{(k)}, x_2^{(k)}) \in \Omega$ , thus

$$\begin{aligned} \sum_{i,j=1}^m c_i c_j K(x^{(i)}, x^{(j)}) &= \left( \sum_{i,j=1}^m c_i c_j \sum_{s=1}^2 a_s K_s(x_s^{(i)}, x_s^{(j)}) \right) \\ &= \left( \sum_{s=1}^2 a_s \sum_{i,j=1}^m c_i c_j K_s(x_s^{(i)}, x_s^{(j)}) \right) \geq 0 \end{aligned} \quad (9)$$

Thus,  $K(x, y)$  is a PDK according to (8), which proves that it is a SV kernel from Theorem 1.

It's clear that the theorem can be extended to the following case, which results from Theorem 4 immediately.

**Corollary 1:**  $K(x, y) \triangleq a_1 K_1(x_1, y_1) \oplus \dots \oplus a_n K_n(x_n, y_n)$  is a SV kernel, where  $x = (x_1, \dots, x_n)$ ,  $y = (y_1, \dots, y_n)$ ,  $x_i, y_i \in \Omega_i$ ,  $a_i \geq 0$ ,  $K_i$  is a R.K,  $i = 1, \dots, n$ .

Note that, the inputs are divided into  $n$  groups. Each group is defined in different domain  $\Omega_i$  so as to be mapped into different feature space by specific SV kernel  $K_i$ . Therefore, the impact on output  $Y$  of the different groups can be computed separately by appropriate SV kernel based on the SPK. For example, suppose there are some SPKs show that the relations between inputs  $x_i, x_j$  and output  $y$  are linear and exponential respectively, then it's believed that it's better to choose linear kernel and exponential kernel for the  $i^{th}$  and  $j^{th}$  groups.

In addition,  $a_i \geq 0$  is a weight, which can be considered to be a parameter for controlling the impact of a certain group of input on the total performance of SVR. It's of profound theoretical and practical significance that different parts of input can be analyzed by setting different value of artificially, while the conventional SV kernels (e.g. RBF) perform an equal treatment for all the input dimensions. Furthermore, the weights can also be computed evolutionally in evolutionary algorithms, e.g. genetic algorithm, to determine the contribution of the components to total results dynamically.

Similarly, the composition method of SV kernel can be applied to tensor product case.

**Theorem 5:** The kernel  $K(x, y) \triangleq K_1(x_1, y_1) \otimes K_2(x_2, y_2)$  is a SV kernel, where  $x, y \in \Omega$ ,  $x = (x_1, x_2)$ ,  $y = (y_1, y_2)$ ,  $x_i, y_i \in \Omega_i$ ,  $\Omega_1 \oplus \Omega_2 = \Omega$ , and  $K_i$  is a R.K,  $i = 1, 2$ .

**Proof:** In fact that  $K_i$ , ( $i = 1, 2$ ) is a Mercer kernel, since  $K_i$  is a R.K. From Mercer's theorem,  $\forall m \in \mathbb{N}$ , the kernel Gram matrix  $\mathbf{K}_i$  of  $K_i$  to  $x_i^{(1)}, \dots, x_i^{(m)} \in \Omega_i$  is as follows:

$$\mathbf{K}_i := \left( K_i(x_i^{(j)}, x_i^{(k)}) \right)_{j,k=1}^m \quad (10)$$

is positive (semi-)definite.

Using a classical Schur product theorem [28], it is easy to prove that the kernel Gram matrix  $\mathbf{K}$  of  $K(x, y)$  is also a positive (semi-)definite matrix. Then,  $K(x, y)$  is a Mercer kernel, and  $K(x, y)$  is a SV kernel from Theorem 2 and 3, which completes the proof.

Obviously, some weights can also be added into the SV kernel in Theorem 5, only with the requirement that the product of these weights must be positive, i.e.  $a_1 a_2 > 0$ . Furthermore, Theorem 5 also can be extended to the tensor product of multi-kernels, that is,

**Corollary 2:** Suppose  $K_i : \Omega_i \times \Omega_i \rightarrow \mathbb{R}$ , ( $i = 1, \dots, n$ ) is a R.K, then

$$K(x, y) = a_1 K_1(x_1, y_1) \otimes \dots \otimes a_n K_n(x_n, y_n) \quad (11)$$

is a SV kernel, where  $x = (x_1, \dots, x_n)$ ,  $y = (y_1, \dots, y_n)$ ,  $x_i, y_i \in \Omega_i$ ,  $\prod_{i=1}^n a_i > 0$ .

The proof can be done by the complete induction.

#### D. Methodology of Motivated Support Vector Regression

In Fig. 3, the methodology of MSVR is presented.

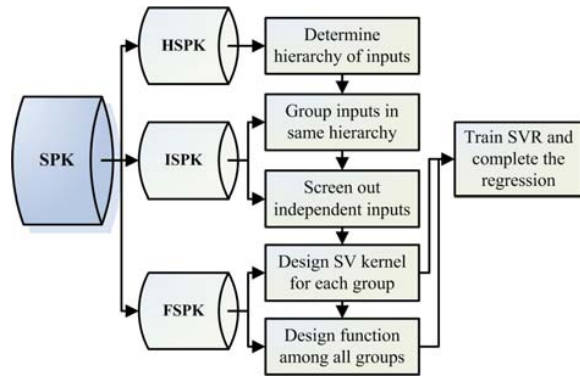


Fig. 3. Flow chart of MSVR based on well-designed SV kernel

##### Step1: Determine the hierarchy of all inputs of interest based on HSPK

As the complexity of problem, diversity of existing models and documents and limitations of human cognition, the inputs of interest may be hierarchical. It may, sometimes, produce large error and misleading information for analysts to train a SVR in black-box pattern without considering the hierarchy. Therefore, it's necessary to analyze the hierarchy of inputs based on HSPK if at all possible.

##### Step 2: Group the dealt inputs and remove the independent inputs based on ISPK

The step is the most important prerequisite of MSVR in this paper. It's usually that, in many applications, the analysts are concerned with a certain part of problem or finding critical components. However, the original models, which may be large and complex, often contain a great many variables. Some of them could be omitted as they turn out to be "no use" (or, at least, average to a constant over quite a range of cases) to the output of interest, and some of them are combined to act on the output. Group the input in same hierarchy can help to clarify the problem at work and consequently to contribute to the comprehensibility and interpretability. Dimension analysis, sensitive analysis and correlation analysis are useful tools in this step.

##### Step 3: Design appropriate SV kernel for each group based on interior-FSPK and postulate a structural form for the integrated regression model based on exterior-FSPK.



This step is the foundation part of MSVR. As stated previously, R.K is regarded as a SV kernel, which can perform a specialized treatment for different dimensions of inputs based on the SPK. In addition, the exterior-FSPK can integrate the groups of inputs more reasonable and meaningful. Because the interest in this paper is largely in models that are understandable, rather than complex mathematics, it's believed that it can greatly improve the "story" of the integrated regression model and accuracy and efficiency as additional product, given that SPK is known partially about the real-world systems being described.

**Step 4: Train the SVR with the well-designed SV kernel and complete the integrated model based on SV algorithm and exterior-FSPK**

The first part of this step is same as the conventional SVR. The differences between them are 1) the SV kernel is designed with SPK, and 2) the integrated structure may be totally different from the kernel expansion in conventional SVR, e.g. integral, statistical functions, and even piecewise functions.

Overall, it's believed that MSVR may possess the following advantage, which will be illustrated in next section:

- (1) Improve the comprehensibility and interpretability for cognition needs;
- (2) identify the critical components and branches;
- (3) improve the accuracy and efficiency;

#### IV. TEST PROBLEMS AND TEST SCHEMES

##### A. Synthetic Problems

In order to illustrate the performance of the MSVR, three synthetic examples and a military simulation example with and without prior knowledge are compared. The test functions are as shown in Table II, which including one output and two or three inputs.

TABLE II  
LIST OF SYNTHETIC PROBLEMS

No.	Approximating function forms
P1	$y = f(x) = 0.3x_1 + 0.5(\sin x_1)^2 - \exp(-0.3x_2^2)$
P2	$y = f(x) = 0.3x_1 + 0.5x_1x_2 + 0.5x_1^2 - 0.2x_2^2 - e^{-0.3x_3^2}$
P3	$y = f(x) = (0.3x_1 + 0.5x_1x_2 + 0.5x_1^2 - 0.2x_2^2) \times (e^{-0.3x_3^2})$

Consider a training data set of  $N = 100$  points in which the input data point  $x_i, i = 1, 2, 3$  is picked uniformly from the interval  $[-2, 2]$ , and the target  $y$  is generated by an additive noise process  $y = f(x) + \varepsilon$ , where  $\varepsilon$  is white Gaussian noise. Additionally, assumed some PRK about these problems are known as filled in Table III.

##### B. Air-to-ground Battle Simulation

Simple, low-resolution models are needed in this test scheme for high-level reasoning and rapidly adaptive calculations. It's noted that the original large and complex object models, which are regarded as reasonably valid, are time-consuming and hopelessly opaque for strategic analysis. Furthermore, the original object models are high resolution models consisting of a large number of variables and uncertainties.

TABLE III  
DESCRIPTION MATRIX OF SYNTHETIC PROBLEMS

Problem No.	Level 2: HSPK	Level 1: ISPK			Level 0: FSPK	
		Indep.	Global	Local	Exterior	Interior
P1	none	none	none	none	$\oplus^a$	$x_1 : P$ $x_2 : E^b$
P2	none	none	$c_1 : x_1, x_2;$ $x_3$	$x_1, x_2$	$\oplus$	$x_1 : P$ $x_2 : E$
P3	none	none	$c_1 : x_1, x_2;$ $x_3$	$x_1, x_2$	$\otimes^c$	$x_1 : P$ $x_2 : E$

<sup>a</sup> $\oplus$  means direct sum for the SV kernels

<sup>b</sup> $x_1 : P$  indicates  $x_1$  is with polynomial relationship with output, and  $x_2 : E$  indicates  $x_2$  is with exponential relationship

<sup>c</sup> $\otimes$  means tensor product for the SV kernels

It's impractical to analysis all the variables with the high resolution models, which is called "disaster of dimension".

In addition, A large number of runs for all the scenario are required to eliminate the errors caused by uncertainties. Therefore, the training samples of Air-to-ground simulation are relatively smaller. This character makes it's an opportunity for SVR to train a lower-level regression model, since SV algorithm is a machine learning method which possesses advantages for resolving small-sample, non-linear and high dimension problems.

TABLE IV  
VARIABLES IN AIR-TO-GROUND SIMULATION

Variable	Description
Variable Inputs	
$N$	Amount of Blue targets to be destroyed
$B_0$	Initial inventory of Red weapon before the battle
$T_{SEAD}$	Time required to suppress Red air defenses
$L_0$	Red loss ratio in SEAD stage
$L_1$	Red loss ratio after SEAD stage
$P_{avg}$	Average efficiency of Red weapon during the battle
$P_0$	Initial efficiency of Red weapon in beginning
$P_1$	Final efficiency of Red weapon at last
Variable Outputs	
$T$	Required time of an air attacking force (Red) to destroy a specific amount of Blue targets

Here, an actual military simulation problem is presented. Suppose there are a high resolution simulation, which describes the required time of an air attacking force (Red) to destroy a specific amount of Blue targets, can generate "data"(15 samples) from the overall input space and collected the variables shown in Table IV through a number of runs (1000). Suppose the initial amount of Blue targets is  $N$ , the initial inventory of Red weapon is  $B_0$ , which continues to loss until all the  $N$  targets are destroyed.  $T_{SEAD}$  denotes the time required to suppress Red air defenses. Suppose the pre-loss ratio is  $L_0$ , the initial loss rate valid for time  $T_{SEAD}$ , and a subsequent post-loss ratio is  $L_1$  valid thereafter. The average

efficiency of Red weapon is  $P_{avg}$ , which indicates the amount of Blue target destroyed per Red attack unit per time unit (day), and in fact changes over time. The initial efficiency is  $P_0$  and final efficiency is  $P_1$ .

### C. Parameter Selection and Metrics for Performance Measures

Parameter selection is a notorious problem since SV algorithm is very sensitive to the adequate choice of parameter values [29], which makes it hard for non-experts. It is, however, a combinatorial optimization problem, and also a NP-hard problem, to select a segment from thousands of their infinite combinations. Lots of papers have shown that genetic algorithm (GA) [30], [31] is useful to solve the combinatorial problem without prior knowledge. Because the strategy of setting parameters is not our research focus, the GA based on GAOT toolbox with its standard settings is used to obtain the best parameters evolutionally [32].

Furthermore, similarly to previous work [15], two qualitative criteria and corresponding five quantitative metrics are used to measure the performance of SVR. They are the fitting precision and efficiency for qualitative measures as well as R square ( $R^2$ ), relative average absolute error (RAAE), relative maximum absolute error (RMAE), modeling time (MT), amount of SVs (ASV) for quantitative measures. For more detail see [15].

The  $R^2$ , RAAE and RMAE are defined in (12)-(14) respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^l (y_i - \hat{y}_i)^2}{\sum_{i=1}^l (y_i - \bar{y})^2} \quad (12)$$

$$RAAE = \frac{\sum_{i=1}^l |y_i - \hat{y}_i|}{\sum_{i=1}^l |y_i - \bar{y}|} \quad (13)$$

$$RMAE = n \times \max\{|y_i - \hat{y}_i|\}_{i=1}^l \bigg/ \sum_{i=1}^n |y_i - \bar{y}| \quad (14)$$

where  $\hat{y}_i$  denotes the corresponding predicted value for observed value  $y_i$ ;  $\bar{y}$  denotes the mean of the observed values.

For the convenience of defining the fitness function in GA, a new measure, *Integration Precision* (IP), is introduced:

$$IP = \alpha(\beta R^2 + (1 - \beta)/RAAE) + (1 - \alpha)/RMAE \quad (15)$$

where  $\alpha, \beta \in [0, 1]$  are weights. In this paper,  $\alpha = 0.9, \beta = 0.5$  to indicate that  $R^2$  and RAAE are more important than RMAE. It is obvious that the larger the IP, the more precise the SVR. Furthermore, the optimal results mentioned latter imply the computation result with the “best” parameters when IP is largest.

## V. SIMULATION RESULTS AND ANALYSIS

### A. Synthetic Problems

Figs. 4-6 show the quantitative comparison results for the problems listed in Table II, where the three legends of figures correspond to the MSVR proposed in this paper, conventional SVR based on polynomial kernel (SVR-PK) and Gaussian

kernel (SVR-RBF) without any SPK respectively (for the convenience of comparison, all the results in the Radar Charts are  $\lg$ -transformed values).

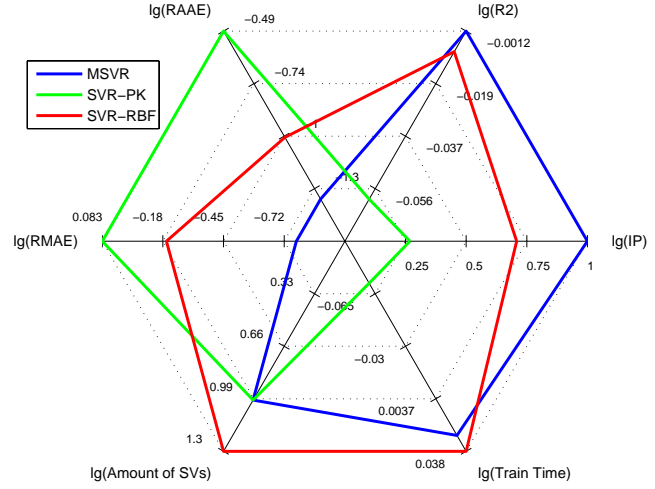


Fig. 4. Radar Chart of Comparison Results for Problem 1

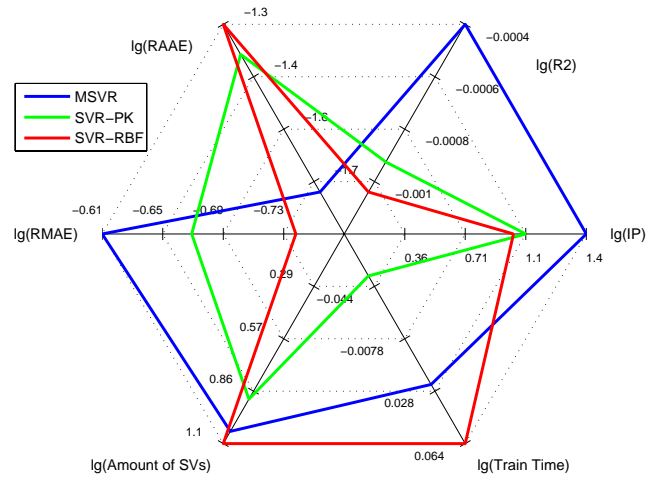


Fig. 5. Radar Chart of Comparison Results for Problem 2

It's obviously that, MSVR is superior to the other two SVR in the fitting precision for all the problems, which indicates that incorporating SPK into SVR can extremely improve the accuracy and robustness of regression tasks. Furthermore, MSVR is close (strictly speaking slightly inferior) to polynomial-based SVR in ASV, while superior to RBF, which shows that MSVR possesses as good ability of generalization as polynomial-based SVR. In other words, the SV kernel used in MSVR is a global kernel as polynomial kernel rather than a local kernel [18]. In addition, the modeling time, i.e. time for training and validating SVR, used by MSVR is less than RBF-based SVR, while relatively more than polynomial-based SVR.

Overall, it's believed that the superiority of MSVR in fitting precision and generalization ability can completely offset against the inferiority in time as the much smaller magnitude

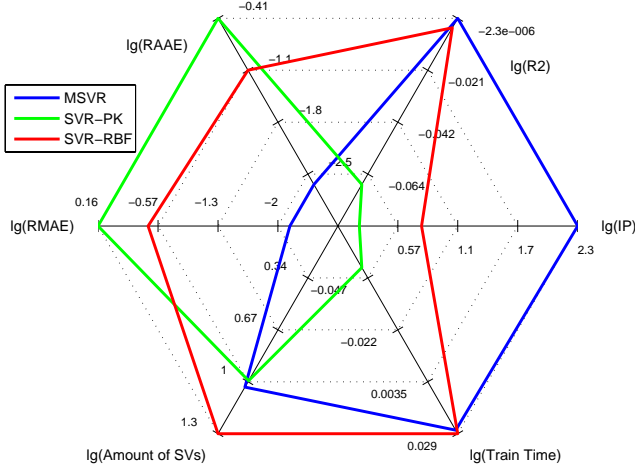


Fig. 6. Radar Chart of Comparison Results for Problem 3

of  $MT$  than that of precision and the desired for good ability in replicative, predictive and structural validity [33].

### B. Air-to-ground Simulation

At beginning, a high-resolution computer model that runs at limited scenarios for predicting the time  $T$  to accomplish the task is used to generate “data”. Conventional SVR to emulate the computer model as a black-box model might predict  $T$  based on all the variable inputs, i.e.  $N$ ,  $B_0$ ,  $T_{SEAD}$ ,  $L_0$ ,  $L_1$ ,  $P_{avg}$ ,  $P_0$ ,  $P_1$  equally. in this paper, the conventional SVR based on Gaussian kernel is treated as the baseline against which MSVR is compared.

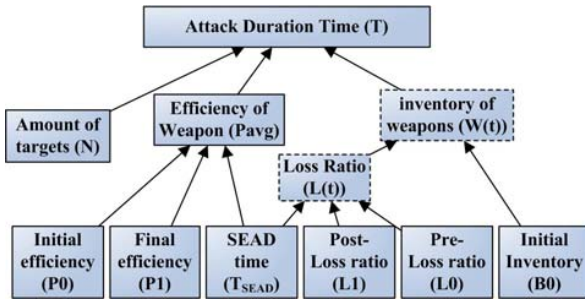


Fig. 7. Hierarchy of Variables for Air-to-ground Simulation

Through analysis, however, a hierarchy is shown in Fig.7 for visualization understanding. The SPK, which should be filled in the description matrix, are listed as follows since the space is limited:

1) **HSPK**:  $P_{avg}$  is in higher hierarchy than others in Table IV, and the regression analysis lies on the lower hierarchy;

2) **ISPK**: (a) there are no independent variables; (b) three composite variables, i.e. a Loss ratio  $L(t)$  whose components are  $T_{SEAD}$ ,  $L_0$ ,  $L_1$ , an inventory of weapon  $W(t)$  which is composed of  $L(t)$ ,  $B_0$  and a similar variable  $P(t)$  replaces  $P_{avg}$  can be introduced. The symbol  $(t)$  indicates these variables are time-related. There exists two global interactions, i.e. one is between  $B_0$ ,  $L(t)$ , the other is among  $N$ ,  $P(t)$ ,  $W(t)$ ; (c)

there are two local interactions, one is among  $P_0$ ,  $P_1$ ,  $T_{SEAD}$ , and the other is among  $T_{SEAD}$ ,  $L_0$ ,  $L_1$ ;

3) **FSPK**: (a) *exterior-FSPK*:

$$\begin{aligned} T &= \frac{N}{W(t)P(t)} \text{ (rough)} \\ N &= \int_0^T W(s)P(s)ds \text{ (detailed)} \end{aligned} \quad (16)$$

(b) *interior-FSPK*:

$$\begin{aligned} (a) \quad P(t) &= \begin{cases} \frac{P_0+P_1}{2} + C_1, & \text{(rough)} \\ P_1 + (P_0 - P_1)(t - T_{SEAD})_+^0, & \text{(detailed)} \end{cases} \\ (b) \quad W(t) &= W(t-1)(1 - L(t)), W(0) = B_0 \\ (c) \quad L(t) &= \begin{cases} \frac{L_0+L_1}{2} + C_2, & \text{(rough)} \\ L_1 + (L_0 - L_1)(t - T_{SEAD})_+^0, & \text{(detailed)} \end{cases} \end{aligned} \quad (17)$$

$$\text{where } (t - T_{SEAD})_+^0 = \begin{cases} 1, & t \geq T_{SEAD} \\ 0, & t < T_{SEAD} \end{cases}$$

Note that the cumulation loss of Red weapon, that is the inventory of weapon is a certain function of time  $T$ , while independent to  $P$ . For the rough SPK, the kernel can be selected nonlinear polynomial kernel for both  $W$  and  $P$ . However, it's more complex to incorporate the detailed SPK, because the piecewise functions and statistical function should be considered. Obviously, the efficiency and loss rate are piecewise function. Being more careful consideration, two possible time  $T_1$  and  $T_2$  are defined as follows, depending on whether the time  $T$  turns out to be greater than the  $T_{SEAD}$ .

$$N = \int_0^{T_1} P_0 B_0 (1 - L_0)^s ds \quad (18)$$

$$N = \int_0^{T_{SEAD}} P_0 B_0 (1 - L_0)^s ds + \int_{T_{SEAD}}^{T_2} P_1 B_0 (1 - L_1)^s ds \quad (19)$$

Then  $T = T_1$  if  $T_1 < T_{SEAD}$  and  $T = T_2$  otherwise, and  $T = \min\{T_1, T_2\}$ . Inspired by the results solved from (18)-(19), two complex nonlinear functions can be defined as follows:

$$\Phi_1 = \ln \left( \frac{N \ln(1 - L_0)}{B_0 P_0} + 1 \right) / \ln(1 - L_0) \quad (20)$$

$$\Phi_2 = \ln \left\{ \frac{(1 - L_1)^{T_{SEAD}} + \left[ \left( N - \frac{B_0 P_0}{\ln(1 - L_0)} \left( (1 - L_0)^{T_{SEAD}} - 1 \right) \right) \times \ln(1 - L_1) / B_0 P_1 \right]}{\ln(1 - L_1)} \right\} \quad (21)$$

Then the SV kernels can be defined as

$$K_1(x, y) = \langle \Phi_1(x), \Phi_1(y) \rangle \quad (22)$$

$$K_2(x, y) = \langle \Phi_2(x), \Phi_2(y) \rangle \quad (23)$$

which completes the incorporation of SPK. Fig. 8 shows the comparison results among MSVR with rough SPK (MSVR-R), MSVR with detailed SPK (MSVR-D) and the conventional SVR based on Gaussian kernel (SVR-RBF). It's obvious that



MSVR greatly outperforms conventional SVR without incorporating any SPK for fitting precision. Note that the more the structural prior knowledge, the more precise the prediction, whereas the more the training and predicting time.

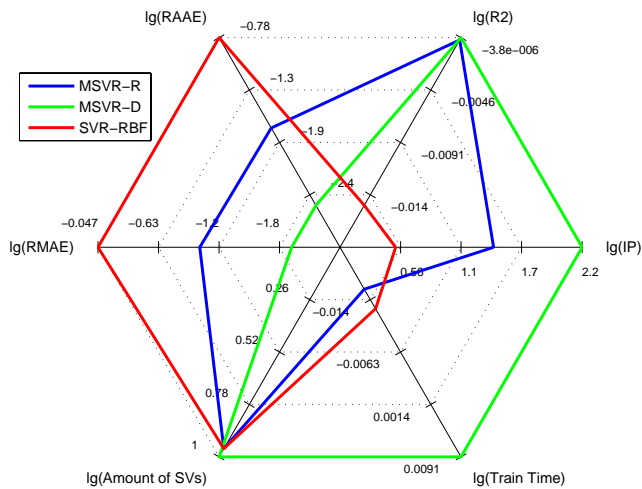


Fig. 8. Radar Chart of Comparison Results for Air-to-ground Simulation

## VI. CONCLUSION

This paper proposes an new method for incorporating structural prior knowledge into SVR. It's known that there are a significant problem in conventional SVR, that is the ubiquitous SPK are neglected in the SVR. It results in the lack of comprehensibility and interpretability in the structure of regression model and consequently hinders the applicability of SVR in complex problems, even if they are reasonable accurate "on average" sometimes. Therefore, this paper is concerned with suggesting ways to improve the quality of SVR by striking a synthesis between conventional SV algorithm and more structural prior knowledge, i.e. motivated support vector regression (MSVR). Subsequently, the specifications of SPK are summarized and an exploratory tool for describing the SPK, that is a Description Matrix of SPK, is proposed as the enabling technologies. Furthermore, a method for composing more complex SV kernel based on R.K is introduced. It possesses many advantages, e.g. ability of considering inputs separately, set weights depending on the analysts' needs, etc. Finally, some synthetic problems and an actual miliary simulation are used to valid the performance of MSVR compared with the conventional SVR based on polynomial kernel and Gaussian kernel. The numerical results indicate the MSVR can indeed improve the quality of conventional SVR in fitting precision and efficiency. It shows great potential for MSVR to the complex military applications.

The problem that examined in this paper was in some respects narrow. significantly more thinking will be necessary to extend the ideas to other classes of problems, for example, how to incorporate the NPK and SPK at the same time; how to use the computerized aid (including the tools used in data mining) to discover more physical insights that could then be

used to MSVR. Such topics would be appropriate subjects of further research.

## ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the financial support of the National Defense Pre-Research Foundation of China (Grant No. 9140C640505), the National Natural Science Foundation of China (No. 60974073 and No. 60974074).

## REFERENCES

- [1] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [2] G. Bloch, F. Lauer, G. Colin, and Y. Chamaillard, "Support vector regression from simulation data and few experimental samples," *Information Sciences*, vol. 178, pp. 3813–3827, 2008.
- [3] J.-B. Gao, S. R. Gunn, and C. J. Harris, "Mean field method for the support vector machine regression," *Neurocomputing*, vol. 50, pp. 391–405, 2003.
- [4] M. A. Mohandes, T. O. Halawani, S. Rehman, and A. A. Hussain, "Support vector machines for wind speed prediction," *Renewable Energy*, vol. 29, no. 6, pp. 939–947, 2004.
- [5] W.-W. He, Z.-Z. Wang, and H. Jiang, "Model optimizing and feature selecting for support vector regression in time series forecasting," *Neurocomputing*, vol. 73, no. 3, pp. 600–611, 2008.
- [6] F. Pan, P. Zhu, and Y. Zhang, "Metamodel-based lightweight design of b-pillar with twb structure via support vector regression," *Computers and Structures*, vol. 88, pp. 36–44, 2010.
- [7] A. J. Smola, T. Friess, and K.-R. Miller, "Semiparametric support vector and linear programming machines," *Advances in neural information processing systems*, vol. 11, pp. 585–591, 1998.
- [8] O. L. Mangasarian, J. Shavlik, and E. W. Wild, "Knowledge-based kernel approximation," *Journal of Machine Learning Research*, vol. 5, pp. 1127–1141, 2004.
- [9] M. Lzaro, F. Prez-Cruz, and A. Arts-Rodriguez, "Learning a function and its derivative forcing the support vector expansion," *IEEE Signal Processing Letters*, vol. 12, pp. 194–197, 2005.
- [10] F. Lauer and G. Bloch, "Incorporating prior knowledge in support vector regression," *Mach. Learn.*, vol. 70, pp. 89–118, 2008.
- [11] O. L. Mangasarian and E. W. wild, "Nonlinear knowledge in kernel approximation," *IEEE Transactions on Neural Networks*, vol. 18, pp. 300–306, 2007.
- [12] B. Schölkopf, P. Simard, A. J. Smola, and V. Vapnik, "Prior knowledge in support vector kernels," in *Advanced in Kernel Method-Support Vector Learning*, C. B. A. S. Schölkopf, B., Ed. Cambridge, England: MIT Press, 1998.
- [13] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [14] P. K. Davis and J. H. Bigelow, "Motivated metamodels: Synthesis of cause-effect reasoning and statistical metamodeling," RAND, Tech. Rep. MR-1570, 2003.
- [15] W. Zhang, X. Zhao, Y.-F. Zhu, and X.-J. Zhang, "A new composition method of admissible support vector kernel based on reproducing kernel," in *International Conference on Computer and Applied Mathematics*, Rio de Janeiro, March 2010.
- [16] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.
- [17] D. Haussler, "Convolution kernels on discrete structures," University of California, Sana Cruz, CA., Tech. Rep. UCSC-CRL-99-10, 1999.
- [18] G. F. Smits and E. M. Jondaan, "Improved svm regression using mixtures of kernels," in *Proceeding of the 2002 international joint conference on neural networks*, vol. 3. Honolulu, Hawaii, USA: IEEE, 2002.
- [19] Y. Tan and J. Wang, "A support vector machine with a hybrid kernel and minimal vapnik-chervonenkis dimension," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, pp. 385–395, 2004.
- [20] B. Schölkopf, K.-K. Sung, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with gaussian kernels to radial basis function classifiers," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2758–2765, 1997.
- [21] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.
- [22] J. Stöckler, "Multivariate bernoulli splines and the periodic interpolation problem," *Constr. Approx.*, vol. 7, pp. 105–120, 1991.

- [23] G. Wahba, "Support vector machines, reproducing kernel hilbert spaces and randomized gacv," in *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf, C. J. Burges, and A. J. Smola, Eds. Cambridge, England: MIT Press, 1999, pp. 69–88.
- [24] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Boston, Dordrecht, London: Kluwer Academic Publishers Group, 2003.
- [25] R. Schaback, "A unified theory of radial basis functions native hilbert spaces for radial basis functions ii," *Journal of Computational and Applied Mathematics*, vol. 121, pp. 165–177, 2000.
- [26] J. Mercer, Ed., *Functions of positive and negative type and their connection with the theory of integral equations*, ser. Philosophical Transactions of the Royal Society, London, 1909, vol. A, 209.
- [27] P. K. Davis, "Introduction to multiresolution, multiperspective modeling (mrmpm) and exploratory analysis," RAND, Tech. Rep. WR-224, 2005.
- [28] R. Horn, *Topics in Matrix Analysis*. Cambridge, MA: Cambridge University Press, 1994.
- [29] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
- [30] Y.-J. Park, "Application of genetic algorithms in response surface optimization problems," Doctor of Philosophy, Arizona State University, December 2003.
- [31] S. S. Chaudhry and W. Luo, "Application of genetic algorithms in production and operations management: A review," *International Journal of Production Research*, vol. 43, pp. 4083–4101, 2005.
- [32] C. R. Houck, J. A. Joines, and M. G. Kay, "A genetic algorithm for function optimization: A matlab implementation," Tech. Rep. NCSU-IE TR 95-09, 1995.
- [33] B. P. Zeigler, H. Praehofer, and T. G. Kim, *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*, second edition ed. CA, USA: Academic Press, 2000.