Transactions C: Chemistry and Chemical Engineering
Vol. 16, No. 1, pp. 41–53

SCIENTIA IRANICA

An Optimal Radial Basis Function (RBF) Neural Network for Hyper-Surface Reconstruction

A. Shahsavand[1]

Abstract. *Data acquisition of chemical engineering processes is expensive and the collected data are always contaminated with inevitable measurement errors. Efficient algorithms are required to filter out the noise and capture the true underlying trend hidden in the training data sets. Regularization networks, which are the exact solution of multivariate linear regularization problems, provide an appropriate means to perform such a demanding task. These networks can be represented as a single hidden layer neural network with one neuron for each distinct exemplar. Efficient training of a regularization network requires the calculation of linear synaptic weights, selection of isotropic spread (σ) and computation of an optimum level of regularization (λ^*). The latter parameters (σ and λ^*) are highly correlated with each other. A novel method is presented in this article for the development of a convenient procedure for de-correlating the above parameters and selecting the optimal values of λ^* and σ^*. The plot of λ^* versus σ suggests a threshold σ^* that can be regarded as the optimal isotropic spread for which the regularization network provides appropriate model for the training data set. It is also shown that the effective degrees of freedom of a regularization network is a function of both regularization levels and isotropic spread. A readily calculable measure of the approximate degrees of freedom of a regularization network is also introduced, which may be used to de-couple λ^* and σ.*

Keywords: *Neural network; Regularization network; Function approximation; Optimum spread; Degrees of freedom.*

INTRODUCTION

Despite the tremendous increase in application of neural networks in many fields, such as electrical, electronics, civil, and control engineering, they were practically unknown to many chemical engineers until the 1990's. Serious efforts to apply neural networks to the simulation and optimization of chemical, biochemical and mineral processes have only begun since the late 1980's [1-3].

Artificial neural networks are usually classified as recurrent and feed-forward. The former networks are mostly used for the empirical modeling of control problems and the latter are well suited for classification or function approximations. The close relationship between the function approximation problem and feed-forward artificial neural networks was explored earlier [4]. Within this framework, neural networks may be viewed as approximation techniques for reconstructing input-output mappings in high-dimensional spaces [5]. In neural network parlance, training (learning) is equivalent to finding a hyper-surface in a multidimensional space that provides the best fit for the training data and generalization is equivalent to the use of this hyper-surface to interpolate within the domain of data where there are no examples [6].

Kernel based neural networks such as Radial Basis Function Networks (RBFNs) have been shown to possess good approximation capabilities. Poggio and Girosi [7,8] reported that among all feed-forward networks, RBFN enjoy the best approximation property [6]. Hunt et al. [9] presented further theoretical support for RBF networks.

The training of RBF networks with specified non-linearities (centers and spreads) reduces to an over-determined set of linear equations, which can be solved by a variety of highly stable techniques such as SVD (Singular Value Decomposition) or Kalman filtering [10]. These networks have a close connection

1. *Department of Chemical Engineering, Ferdowsi University of Mashhad, Mashhad, P.O. Box 91775-1111, Iran. E-mail: shahsavand@um.ac.ir*

Received 17 March 2007; received in revised form 8 December 2007; accepted 1 July 2008

with the well studied subject of multivariate function approximation and enjoy a firm theoretical foundation.

Due to the complexity of chemical engineering processes, the acquired data are always contaminated with heavy measurement errors. As a result, the employment of advanced noise filtering techniques is crucial to avoid over-fitting (or fitting the noise) phenomena [11]. RBF networks that have originated from a multivariate regularization theory can distinguish and filter-out the noise. These networks are ideal for capturing the true underlying trend from a set of heavily contaminated chemical engineering data.

It was shown [7,8] that the solution of the multivariate regularization problem can be represented as a single hidden layer network, known as the regularization network. Regularization networks are traditionally constructed using isotropic Gaussian basis functions. An illustrative example is employed to investigate the effect of the regularization level and the value of the isotropic spread, σ, on the performance of a regularization network. A convenient procedure is introduced for selecting the appropriate value of the isotropic spread, σ. To the best of our knowledge, this approach has not been addressed previously and makes a significant improvement in the performance of the regularization networks.

STRICT INTERPOLATION PROBLEM

We start by considering the strict interpolation problem and the remarkable theorem due to Michelli [12], which pinpoints the importance of radial basis functions in multivariate function approximation and neural networks. The strict interpolation problem may be stated as: "Given a set of N input s ($\underline{x}_i \in \Re^P, i = 1, \cdots, N$), and the corresponding outputs, (y_i, $i = 1., \cdots, N$), find a continuous multivariate function $F(\underline{x})$ which maps the inputs to the output and satisfies the following N interpolating conditions":

$$F(\underline{x}_i) = y_i, \qquad i = 1, \cdots, N. \tag{1}$$

Let us expand $F(\underline{x})$ as a linear summation of N radial basis functions, each centered at a distinct data point:

$$F(\underline{x}) = \sum_{j=1}^{N} w_j \phi_j \left(\| \underline{x} - \underline{x}_j \| \right). \tag{2}$$

Here $\phi_j(r)$ represents a radial function whose argument is a measure of the (Euclidean) distance from the known centre located at $\underline{x}_j$, $r = \| \underline{x} - \underline{x}_j \|$ and the w_j's are the weighting coefficients. Equations 1 and 2 can be combined and stated in the compact form:

$$\underline{F} = \Phi \underline{w}, \tag{3}$$

where $\underline{F} = [F(\underline{x}_1), F(\underline{x}_2), \cdots, F(\underline{x}_N)]^T$, $\underline{w} = [w_1, \cdots, w_N]^T$ and Φ is the $N \times N$ interpolation matrix with elements,

$$\phi_{ij} = \phi \left(\| \underline{x}_i - \underline{x}_j \| \right). \tag{4}$$

The theorem originally proved by Michelli [12] for multiquadrics basis functions states that: "if $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_N$ are N distinct points in $\Re^P$, the interpolation matrix, Φ, is non-singular". All that is required is for the input vectors to be distinct and the elements of the interpolation matrix to be based on radial functions centered on the known distinct data points. In exact or infinite precision arithmetic, the system in Equation 3 can always be solved to obtain the unique optimal weights:

$$\underline{w}^* = \Phi^{-1} \underline{y}. \tag{5}$$

The continuous function in Equation 2 with the optimal weights will satisfy the N interpolating conditions (Equation 1) exactly. It should be pointed out here that in practice we always deal with finite rather than infinite precision arithmetic and as a result, the interpolation matrix, Φ, may turn out to be numerically singular. This does not violate Michelli's theorem; it only tells us that we must use higher precision for our calculations. The strict interpolation problem always has a solution in terms of radial basis functions centered at the data points, irrespective of the size of the data set N or the dimensions of the input vector, $\underline{x}$.

A large class of RBFs are covered by the Michelli theorem [6,13,14]. The functions shown in Table 1 are of particular interest in the study of multivariate regularization and its implementation as RBF networks. Michelli has proved that for localized basis functions such as Gaussian and Inverse Multiquadric, the interpolation matrix, Φ, is positive definite. The Multiquadric and Thin Plate Spline functions are global and for these basis functions, the interpolation

Table 1. Various radial basis functions.

Radial Basis Functions	Formula
Gaussian	$\phi(\underline{x}; \underline{t}, \sigma) = \exp(-\frac{\|\underline{x}-\underline{t}\|^2}{\sigma^2})$
Multiquadrics	$\phi(\underline{x}; \underline{t}, \sigma) = [\|\underline{x} - \underline{t}\|^2 + \sigma^2]^{\frac{1}{2}}$
Inverse multiquadrics	$\phi(\underline{x}; \underline{t}, \sigma) = \frac{1}{[\|\underline{x}-\underline{t}\|^2+\sigma^2]^{\frac{1}{2}}}$
Thin plate splines	$\phi(\underline{x}; \underline{t}, \sigma) = \left[\frac{\|\underline{x}-\underline{t}\|}{\sigma}\right]^{2n} \ln\left(\frac{\|\underline{x}-\underline{t}\|}{\sigma}\right)$

matrix is indefinite with $N-1$ negative and one positive eigenvalues [6]. In principle, global radial basis functions construct a smoother fit and local RBFs can extract more specialized features due to their locality.

REGULARIZATION NETWORK WITH ISOTROPIC SPREADS

In a different approach, Poggio and Girosi [7,8] illustrated that the solution of a multivariate regularization problem can be represented as:

$$(G + \lambda I_N)\underline{w}_\lambda = \underline{y}, \tag{6}$$

where G is the $N \times N$ symmetric Green's matrix with elements $G_{ij} = G(\underline{x}_i, \underline{x}_j)$ and λ is the regularization parameter. In practice, we may always choose λ sufficiently large to ensure that the matrix $(G + \lambda I_N)$ is positive definite and, hence, invertible. Equation 6 can be symbolized as the network shown in Figure 1. The network consists of a single hidden layer with N neurons and the activation function of the jth hidden neuron is a Green function, $G(\underline{x}, \underline{x}_j)$, centered at a particular data point, $\underline{x}_j$. The influence of the regularization parameter, λ, is embedded in the unknown synaptic weights, $w_j's$.

Poggio and Girrosi [7,8] also pointed out that for a special choice of stabilizing operator, Green's function reduces to a multidimensional factorizable isotropic Gaussian basis function, which is both translationally and rotationally invariant, having an infinite number of continuous derivatives [6].

$$G(\underline{x}, \underline{x}_j) = \exp\left[-\frac{\left\|\underline{x} - \underline{x}_j\right\|^2}{2\sigma_j^2}\right]$$

$$= \prod_{k=1}^{p} \exp\left[-\frac{(x_k - x_{j,k})^2}{2\sigma_j^2}\right]. \tag{7}$$

The σ_j appearing in Equation 7 denotes the *isotropic* spread of the jth Green function, which is assumed identical for all input dimensions. The performance of

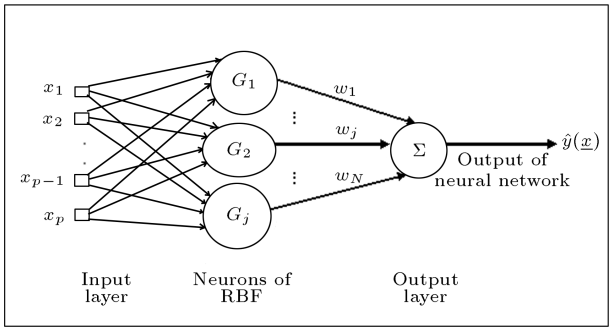

Figure 1. The regularization network.

the regularization network strongly depends on both the appropriate choice of the isotropic spread and the proper level of regularization [15]. The leave-one-out cross validation criterion [16] is used for efficient computation of the optimum regularization parameter, λ^*, for a given σ.

$$CV(\lambda) = \frac{1}{N}\sum_{k=1}^{N}\left[\frac{\underline{e}_k^T(I_N - H(\lambda))\underline{y}}{\underline{e}_k^T(I_N - H(\lambda))\underline{e}_k}\right]^2, \tag{8}$$

where N is the number of both the training exemplars and neurons of the regularization network, $\underline{e}_k$ is the kth unit vector of size N, I_N is the $N \times N$ unit matrix and $H(\lambda)$ is the smoother matrix originally defined by Hastie and Tibshirani [17]. (If we focus on the fit at the observed data points $x_1, x_2, \cdots, x_N$ a linear smoother can be written as $f = Sy$ where $S = \{S_{i,j}\}$ is an $N \times N$ smoother matrix [17].)

It is already known that, by definition, the regularization network is a linear smoother as developed by Poggio & Girrosi [7,8]. The generalization performance of such a network can be simply computed from $f = G\underline{w}_\lambda$. Replacing $\underline{w}_\lambda$ from Equation 6 results in:

$$f = G(G + \lambda I)^{-1}\underline{y}. \tag{9}$$

Therefore, by definition of the smoother matrix, $H(\lambda)$ for the regularization network can be computed from:

$$H(\lambda) = S = G(G + \lambda I)^{-1}. \tag{10}$$

The effective number of parameters or degrees of freedom (df) of a linear smoother is defined as the trace of the smoother matrix that is equal to the sum of its eigenvalues, $df = tr(H(\lambda))$ (or sum of its diagonal elements). Evidently, the number of degrees of freedom is a function of the span and the predictor values in the data set; it is not a function of the response (y) [17].

The evaluation of $H(\lambda)$ and hence $CV(\lambda)$, at each trial value of λ, requires the inversion of the $N \times N$ matrix $(G + \lambda I)$ and may prove too time-consuming. This can be avoided by resorting to the similarity transformation technique [4]. The basic idea was initially presented by Golub et al. [18] for ridge regression.

Equations 7 to 9 show that $CV(\lambda)$ is a complex function of both λ and σ. Therefore, the optimal value of the regularization parameter, λ^* (which minimizes $CV(\lambda)$), is highly correlated with the value of the isotropic spread, σ. In other words, the appropriate value of λ^* greatly depends on σ for a specific data set with a fixed level of noise. The strong correlation between these two parameters (λ^* and σ) is extremely complicated and may not be described directly in analytical form. A simple indirect procedure is proposed in this article for the decoupling of such

powerful dependency and for finding the optimum value of spread and the corresponding optimal level of regularization for the given noisy data set. The motivation behind such a decoupling procedure is to train the best optimal network for reconstructing the true hyper-surface embedded in the bunch of a noisy data set. It is clearly shown that such an optimum trained network can successfully filter out the noise and provide the finest generalization performance.

An illustrative example presented in the next section demonstrates the strong correlation between λ^* and σ. A significant contribution of this article is the development of a convenient procedure for decorrelating these parameters and selecting the optimal values of λ^* and σ^* for the specified data set.

AN ILLUSTRATIVE EXAMPLE

To illustrate the capabilities and probable shortcomings of the regularization network, a three-dimensional (bivariate) synthetic example is presented in this section. The main reason for the selection of a 3D case study is its perfect visualization feature. Two-dimensional examples are too simple to mimic the real world and 4D$^+$ case studies are hard to visualize. The following function describes the true underlying relationship between two input variables ($0 \leq x_1, x_2 \leq 1$ radians) and one response (output) variable, as shown in Figure 2.

$$Z = 100\left[(\mathrm{Sin}(5.5x)\mathrm{Cos}(3y))^2 + 0.15\right] \exp\left[-\left(\frac{x-0.5}{0.4}\right)^2\right] \exp\left[-\left(\frac{y-0.5}{0.3}\right)^2\right]. \quad (11)$$

Since all experimental data are inevitably associated with some measurement errors (noise), the training exemplars generated from the above surface are contaminated with the random noise of uniform distribution and known width in the following examples. The random vectors with zero mean and identity covariance matrices were generated using MATLAB software. To produce such random vectors, the following procedure may be employed: Suppose that a random vector, $\underline{X}$, has a covariance matrix, Q. Since this matrix is Hermitian symmetric and positive semidefinite, by the spectral theorem from linear algebra, we can diagonalize or factor the matrix in the following way:

$$Q = E\Lambda E^T, \quad (12)$$

where E is the orthogonal matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues.

To whiten such a random vector $\underline{X}$ with mean μ and covariance matrix Q, we may transform it to a white vector $\underline{W}$ as follows:

$$W = \Lambda^{\frac{-1}{2}} E^T(\underline{X} - \mu). \quad (13)$$

It can be simply verified that the expectation of such a vector is zero (i.e. $E(\underline{W}) = 0$) and the expectation of its covariance matrix is equal to the identity matrix (i.e. $E(\underline{W}\,\underline{W}^T) = I$). Thus, with the above transformation, we can whiten the random vector to have zero mean and an identity covariance matrix [4].

As a first trial, four hundred training data were sampled randomly from the input domain of $0 \leq x_1, x_2 \leq 1$ (see Figure 3) and the corresponding true underlying responses were computed using Equation 8. The normalization of input space does not create any limitation for practical applications, but extremely enhances the selection of isotropic spreads. The true underlying responses were then contaminated with 10, 20 and 50 percent noise levels and the corresponding noisy outputs are shown in Figure 4. The blue line indicates that $y = x$.

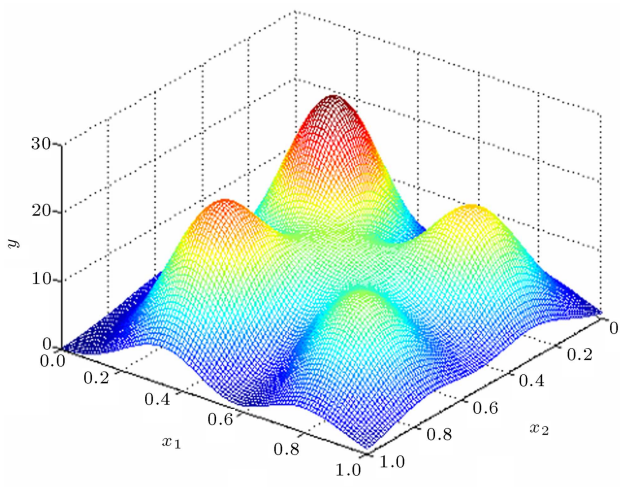

Figure 2. 3D plot of the bivariate example.

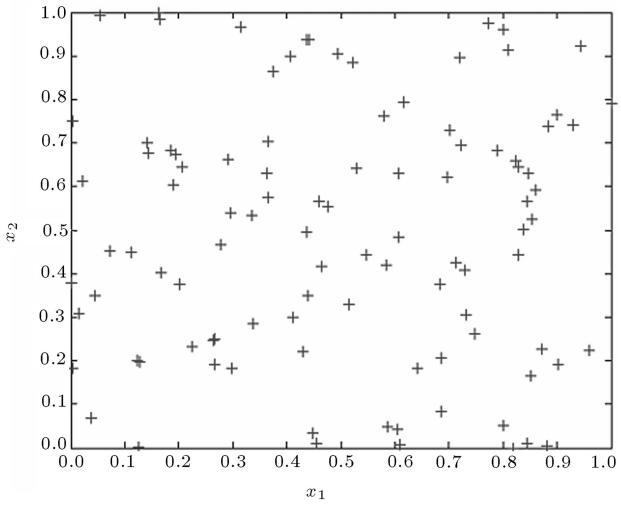

Figure 3. Random input data of the bivariate example.

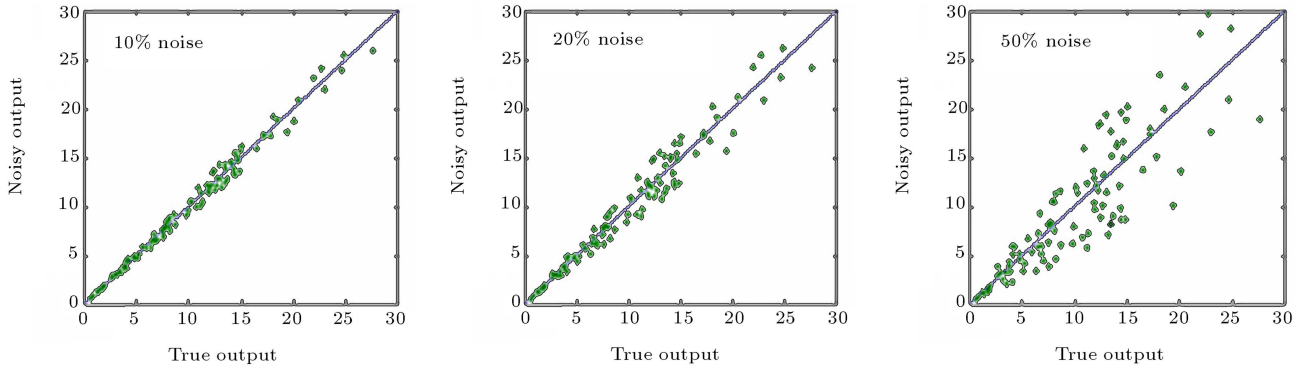

Figure 4. Training data with different noise levels.

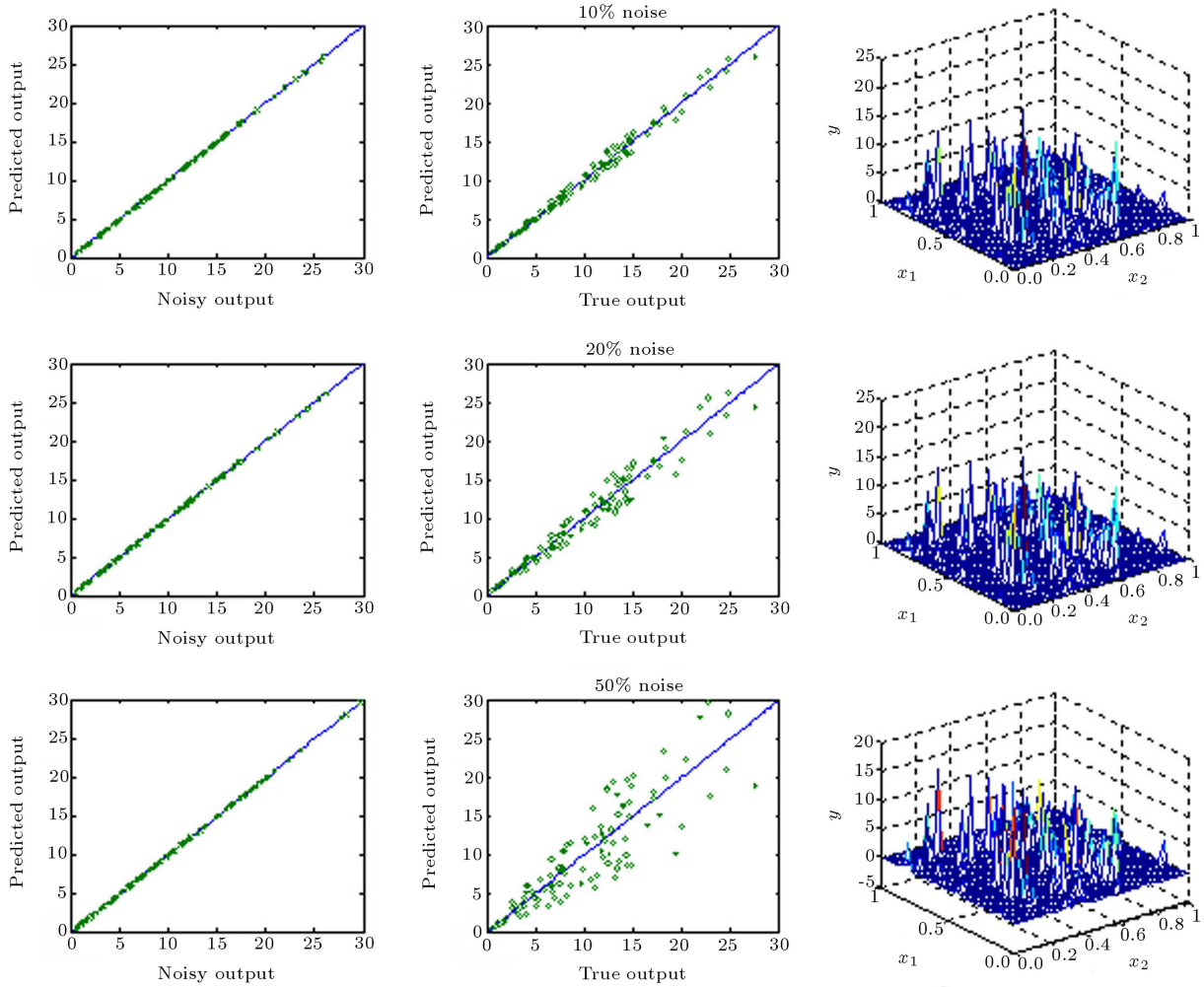

Figure 5a. Generalization performance of regularization network for various noise levels in the absence of regularization ($\sigma = 0.01$).

The above noisy data sets were used for training three separate regularization networks with different isotropic spreads, each with 100 Gaussian basis functions centered at the training data points. Figures 5a to 5c show the 3D generalization performance of these networks on a 50×50 uniform grid for the isotropic spreads of $\sigma = 0.01, 0.1$ and 0.5 and noise levels of 10%, 20% and 50%, in the absence of regularization ($\lambda = 0$). All 2D diagrams correspond to recall performances of regularization networks for the training data set. As before, the blue line indicates that $y = x$.

The above figure shows that regularization net-

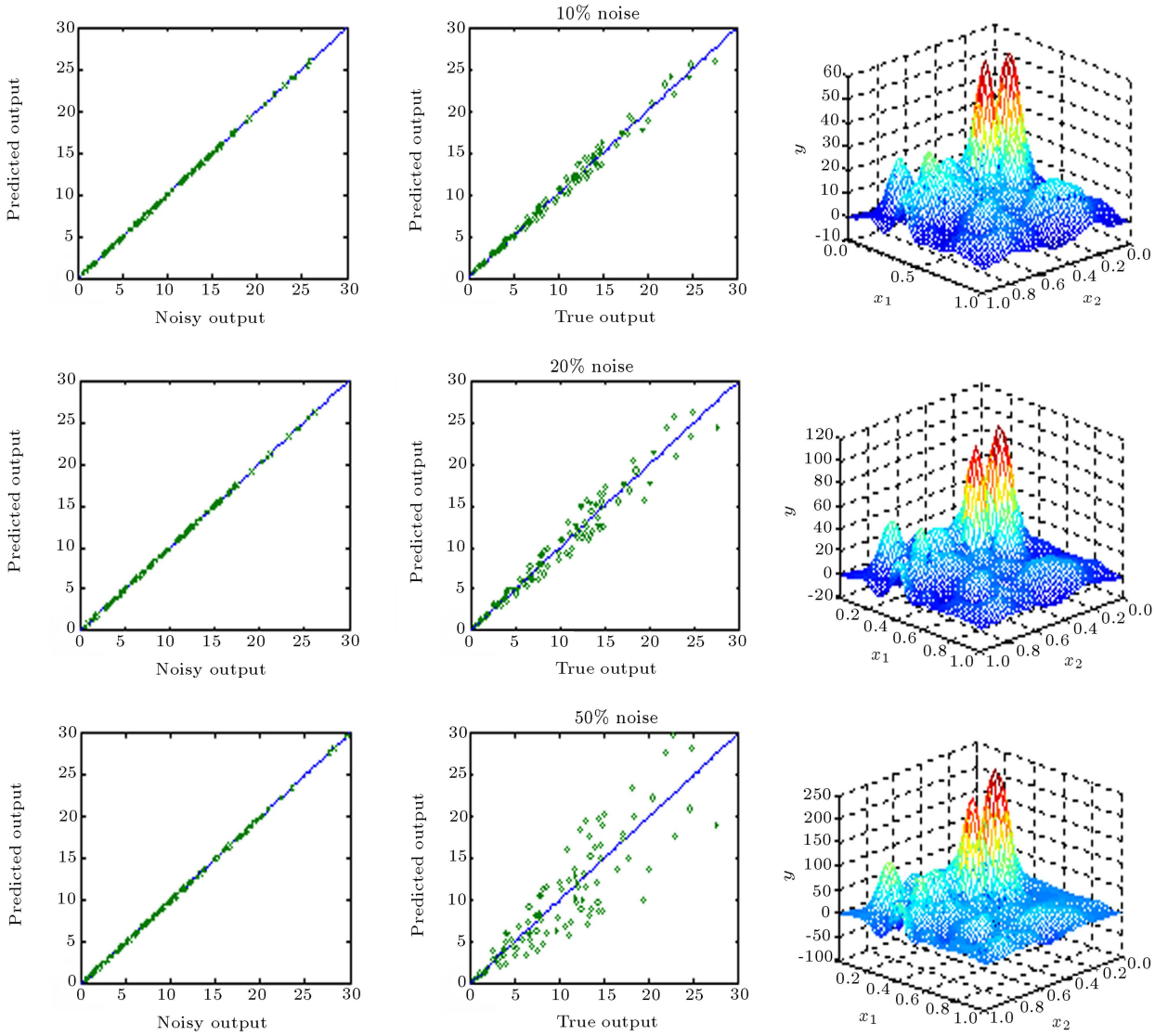

Figure 5b. Generalization performance of regularization network for various noise levels in the absence of regularization ($\sigma = 0.1$).

works with extremely small spreads ($\sigma = 0.01$) will fit the noise in the absence of regularization. Such networks exactly reproduce the noisy data but perform inadequately between the training data points. The reason for this is that the Green matrix of the regularization network tends to an $N \times N$ identity matrix for very small spreads. Such matrices are very well behaved and invertible. Since both the Green matrix and its inverse are identical, the network exactly recovers the training data points. On the other hand, the Gaussian basis functions are exceedingly narrow and, as can be seen, the network cannot produce a sufficient response between the training data points. Increasing the value of the isotropic spread to 0.1 in the absence of regularization, although fitting the noise, leads to a smoother hyper-surface with large oscillations (see the magnitude of Y on the 3D plots of Figure 5b). Figure 5c shows that a similar situation can happen for the case of $\sigma = 0.5$. The reason for these large oscillations is the ill-conditioning of the Green matrix. This phenomenon occurs because of the excessive overlap between adjacent centers at relatively large spreads. Inversion of such a nearly singular matrix can lead to extremely oscillatory synaptic weights and produces a fairly smooth but unreliable surface with tremendously large responses.

Employing the regularization technique can alleviate the ill-conditioning problem of the Green matrix, due to the overlap of large spreads, but will not cure the inadequacy of the model (network) for very small spreads. This issue is clearly demonstrated by using the same data set for training the previous networks with an optimum level of regularization (λ^*). The three-dimensional plots of Figures 6a to 6c illustrate

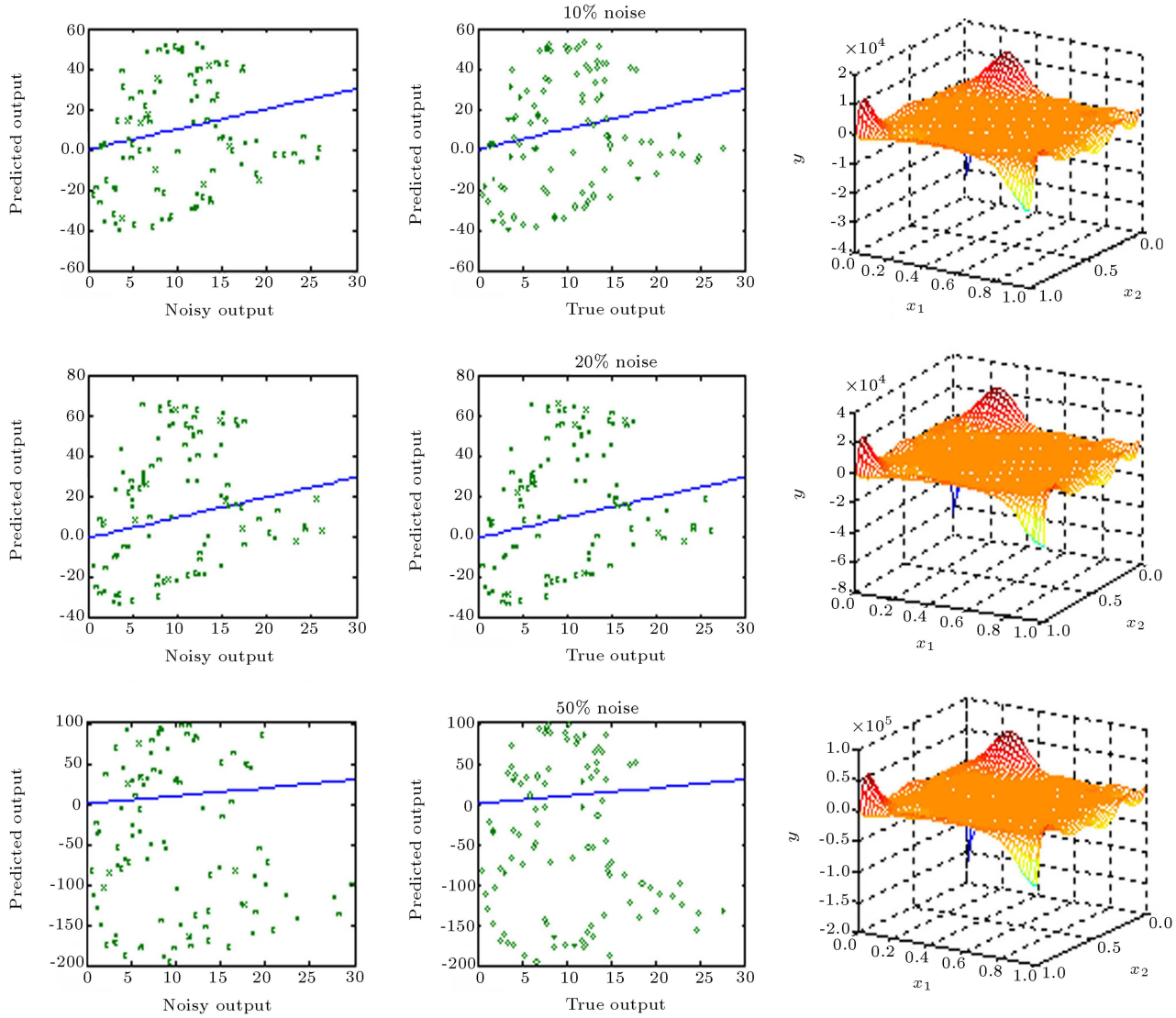

Figure 5c. Generalization performance of regularization network for various noise levels in the absence of regularization ($\sigma = 0.5$).

the generalization performances of these networks on a 50×50 uniform grid for $\sigma = 0.01, 0.1$ and 0.5. The Leave-One-Out (LOO) Cross Validation (CV) criterion was used to compute the optimum level of regularization. As mentioned earlier, the 2D plots show the recall performances of the regularization networks.

The above figure clearly shows that the inadequacy of the model (due to small isotropic spreads) cannot be alleviated by the optimum level of regularization. As shown in Figure 7, the Cross Validation (CV) criterion does not possess any minima for various noise levels with $\sigma = 0.01$. On the other hand, when the isotropic spread is large enough to provide an appropriate model for the data set, the Green matrix may become ill-conditioned and the CV criterion shows a clear minimum. The optimum level of regularization eliminates the ill-conditioning problem and leads to a reasonable generalization performance, as shown in Figures 6b and 6c.

It is interesting to note that a regularization network with a proper choice of spread and an optimum level of regularization can filter out the noise (instead of following it) and capture the true response from the highly noisy data sets (see 2D plots of Figure 6c for 50% noise). It is also clearly visible that, with a reasonable value of isotropic spread, the regularization network is able to reconstruct the general features of the true underlying surface hidden in a set of noisy data. It should also be noted that the predicted surface is still oscillatory and cannot provide the exact true responses. Therefore, a mechanism should be devised to find the optimal value of isotropic spread (σ^*) for a given data set.

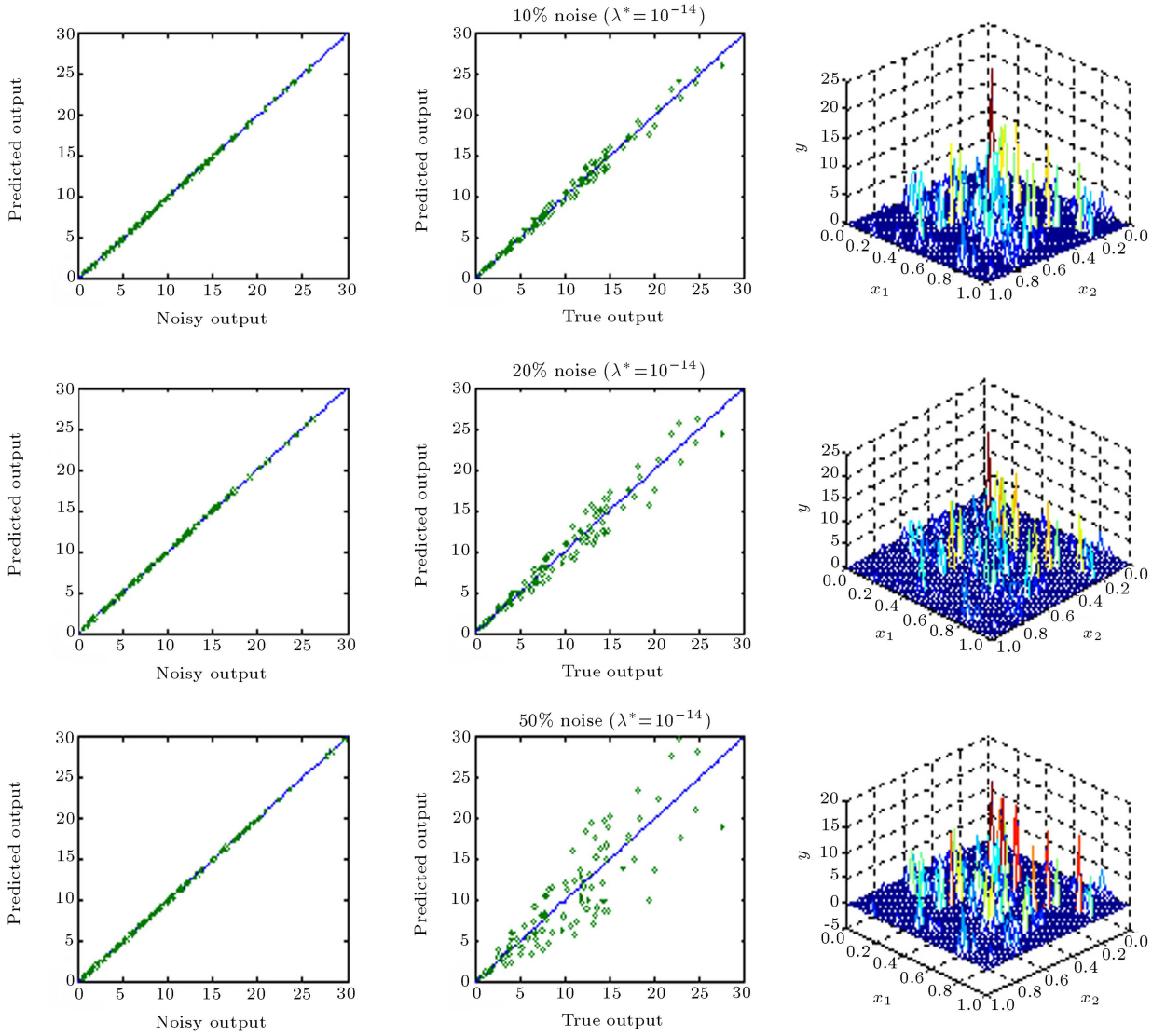

Figure 6a. Generalization performance of regularization network for various noise levels at the optimum level of regularization ($\sigma = 0.01$).

Closer examination of Figures 6a to 6c reveals that the optimum level of regularization, (λ^*), is strongly correlated with the value of the isotropic spread. A plot of λ^* versus σ for various noise levels on Figure 8 shows a clear maximum; the corresponding value of the isotropic spread can be regarded as the optimum value of spread (σ^*) for the regularization network under consideration. On the other hand, the cross validation criterion decreases monotonically and does not provide any maximum or minimum, as shown in Figure 8. Evidently, such criterion cannot be used to select the optimum value for the isotropic spread. Figure 9 illustrates the generalization performance of the regularization networks at optimum values of isotropic spread and the corresponding optimum level of regularization for various noise levels.

As can be seen in Figure 8, the optimum value of the spread is independent of noise level, but further investigation shows that it is strongly dependent on the number of training data points.

It is also interesting to note that the trained network with an optimal isotropic spread at an optimum level of regularization does not follow the noise, but tries to capture the true underlying trend hidden in the remarkably noisy training sets (see Figure 9). The above discussion proves that regularization is essential to prevent spurious oscillation caused by overfitting of the noisy data. More significantly, using an appropriate value of σ and an optimal level of regularization determined by the CV criterion recovers the true underlying surface remarkably well from the above set of limited and noisy data.

The plot of λ^* versus σ on Figure 8 provides

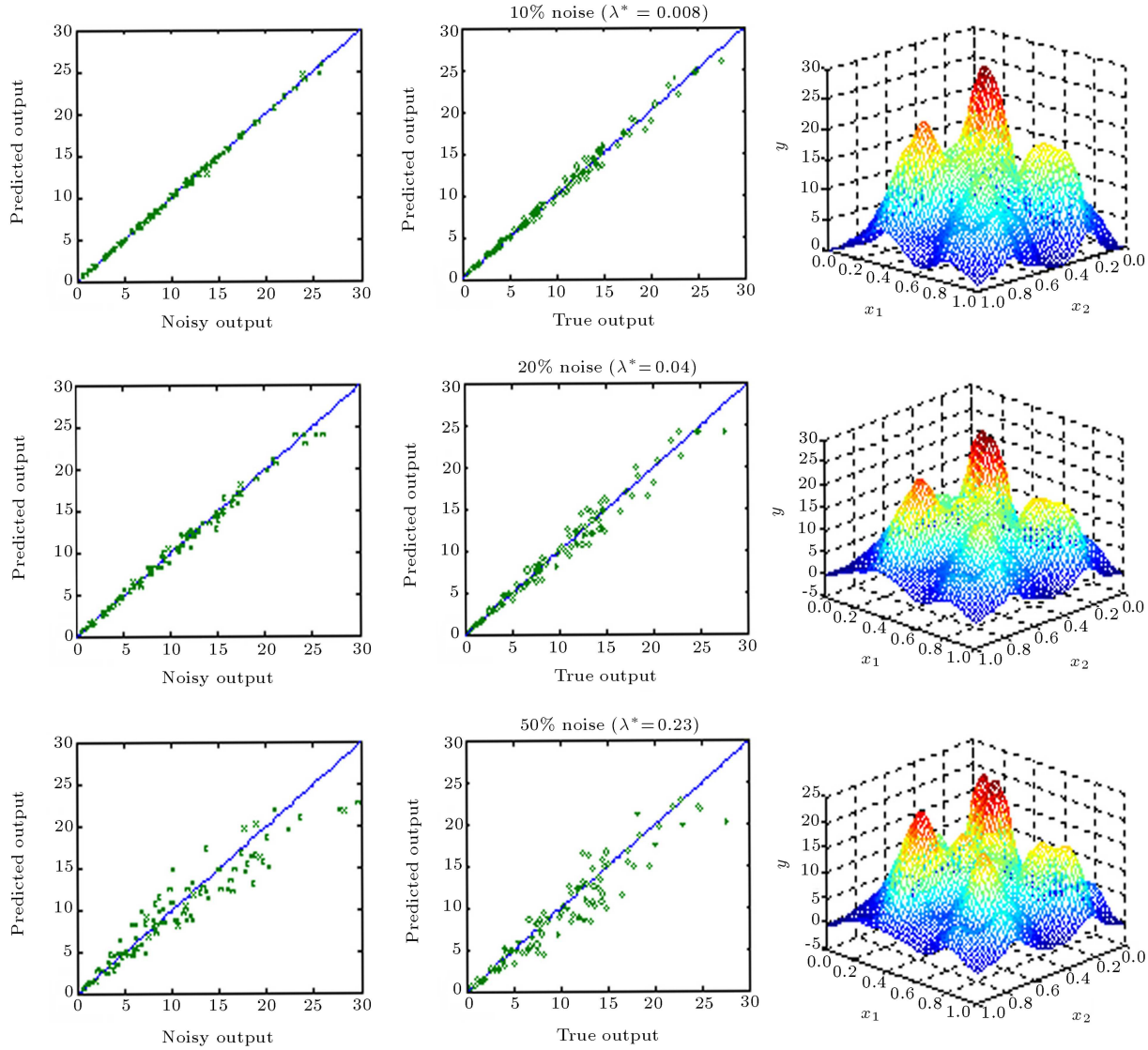

Figure 6b. Generalization performance of regularization network for various noise levels at the optimum level of regularization ($\sigma = 0.1$).

two local maxima for relatively low noise levels. Figure 10 shows the generalization performance of the regularization networks with isotropic spreads of 0.03 (corresponding to first local maxima) for 10 and 20 percent noise levels. Evidently, the first maxima cannot de-correlate the isotropic spread and level of regularization.

We close this discussion of regularization networks by giving a justification for the existence of a threshold value for the isotropic spread, σ, based on an approximate measure of the degrees of freedom that can be sustained by the data. For a model with M parameters representing given N observations, the effective degrees of freedom are equal to $N - M$. Using the definition of the smoother matrix as [17]:

$$H(\lambda) = G(G + \lambda I)^{-1}. \tag{14}$$

For model comparisons, the approximate degrees of freedom, which give an indication of the amount of fitting that H does, are defined as the $tr(H)$ (sum of the eigenvalues of matrix H). Figure 11 illustrates the variations of the optimal levels of regularization, λ^*, and the corresponding approximate degrees of freedom $df(\lambda^*)$= tr $(H(\lambda^*))$, with the isotropic spread of the Gaussian basis functions used in a regularization network.

It is clear that threshold σ occurs near the point where the approximate degrees of freedom have a minimum. Using threshold σ, therefore, enables us to select the approximate degrees of freedom required to fit the underlying surface. Smaller σ introduces larger degrees of freedom leading to spurious oscillations, while larger σ limits the degrees of freedom and leads to over-smoothing.

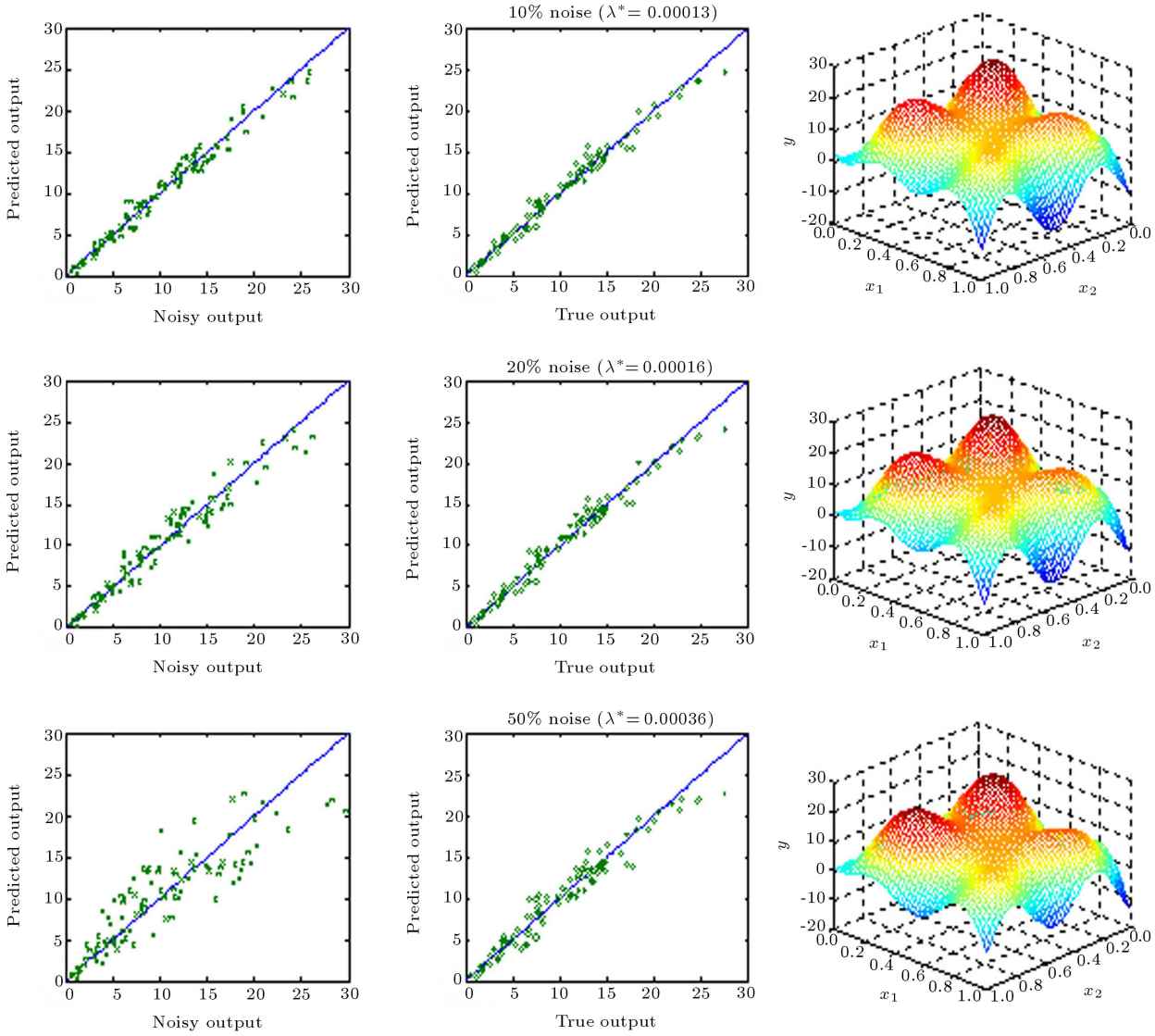

Figure 6c. Generalization performance of regularization network for various noise levels at the optimum level of regularization ($\sigma = 0.5$).

CONCLUSION

Chemical engineering data are expensive to collect and are always contaminated with some level of noise. Efficient algorithms are required to filter out the noise and capture the true underlying trend from the noisy data sets. Regularization networks are inherently equipped with proper means to perform such a demanding task. This paper was aimed exclusively at an important class of feed-forward neural networks with a single hidden layer (Regularization Networks), which have a solid mathematical foundation. In the majority of reported applications, the regularization network has been employed with Gaussian radial basis functions with a constant isotropic spread.

It is shown that the optimal value of regularization parameter, λ^*, is highly correlated with the isotropic spread, σ, an obvious point that has received surprisingly little attention to date. An illustrative example was used to clearly demonstrate the strong correlation between λ^* and σ. A significant contribution of the present article is the development of a convenient procedure for de-correlating these parameters and selecting the optimal values of λ^* and σ^*.

It is also clearly demonstrated that the effective degrees of freedom, $df(\lambda, \sigma)$, of a regularization network is a function of both the regularization level, λ, and the isotropic spread, σ. A readily calculable measure of the approximate degrees of freedom of a regularization network was introduced, which may be used to de-couple λ^* and σ. The plot of $df(\lambda^*, \sigma)$ against σ provides a curve which exhibits a minimum. This minimum is an approximate measure of the degrees of freedom which can be reasonably sustained

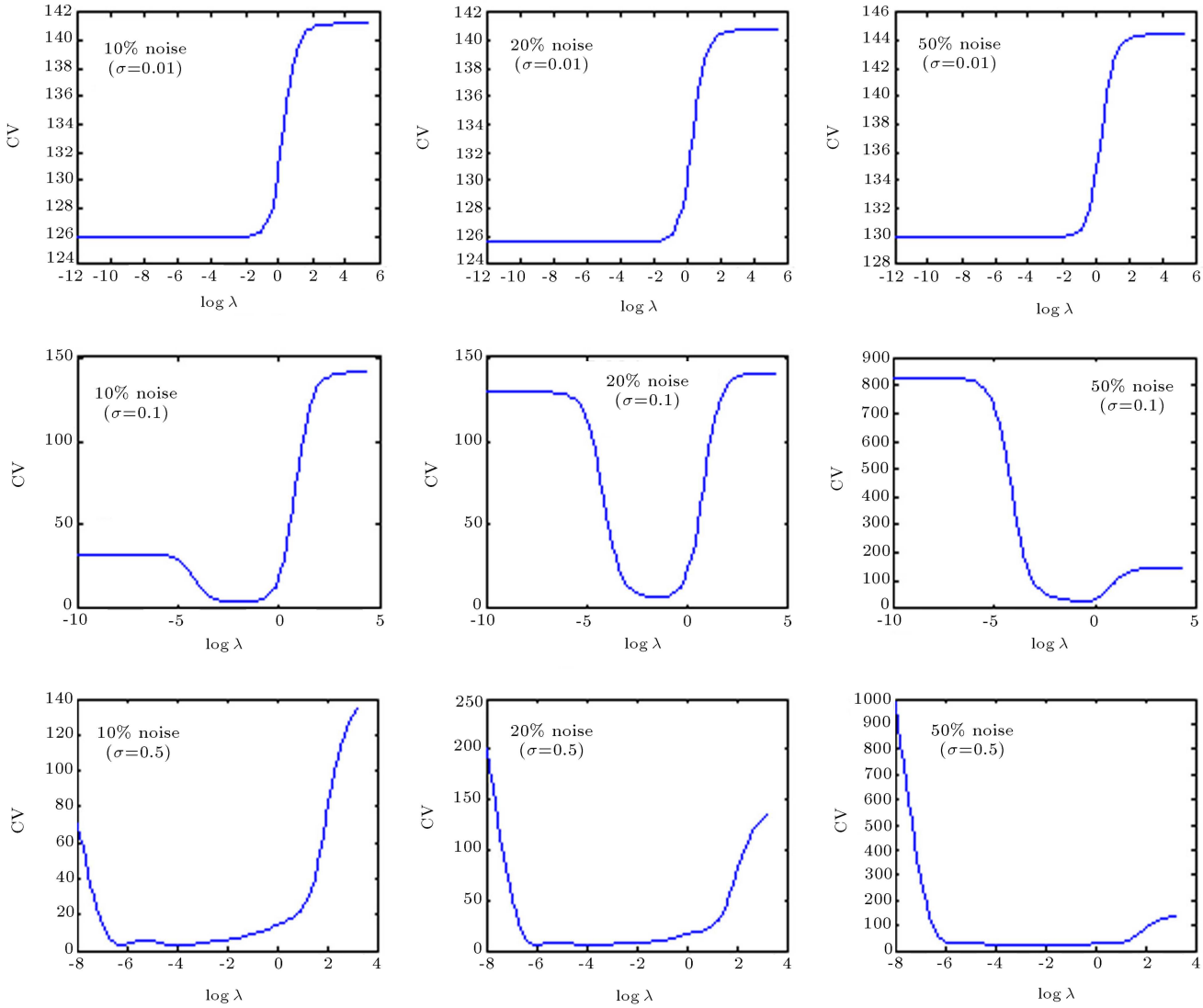

Figure 7. CV criterion vs. level of regularization for various noise levels and spreads.

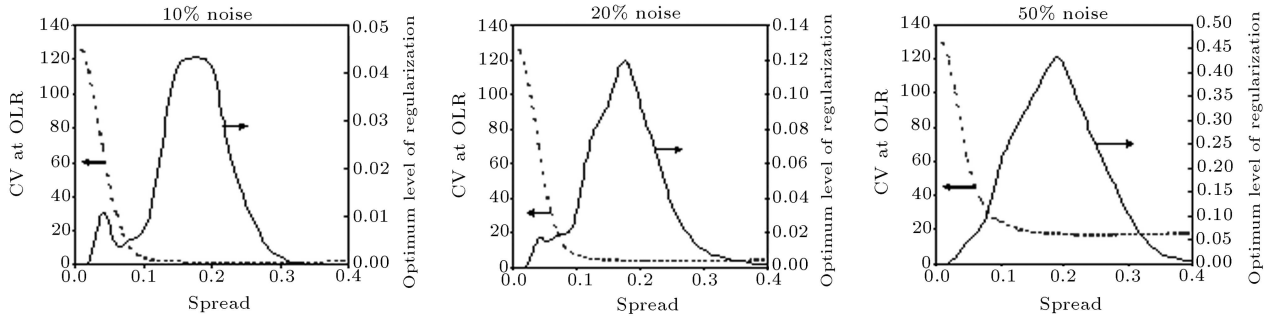

Figure 8. Optimum level of regularization and the corresponding cross validation criterion vs. isotropic spreads for various noise levels.

by the noisy data set and which can be used to provide the best value for the isotropic spread, σ^*. The use of the effective degrees of freedom for this purpose leads to a significant improvement in the performance of the regularization network and, to our knowledge, has not been previously reported. Applications of the above algorithm on several case studies in the fields of characterization and optimization of porous materials and in the modeling of membrane processes are presented elsewhere [19,20].

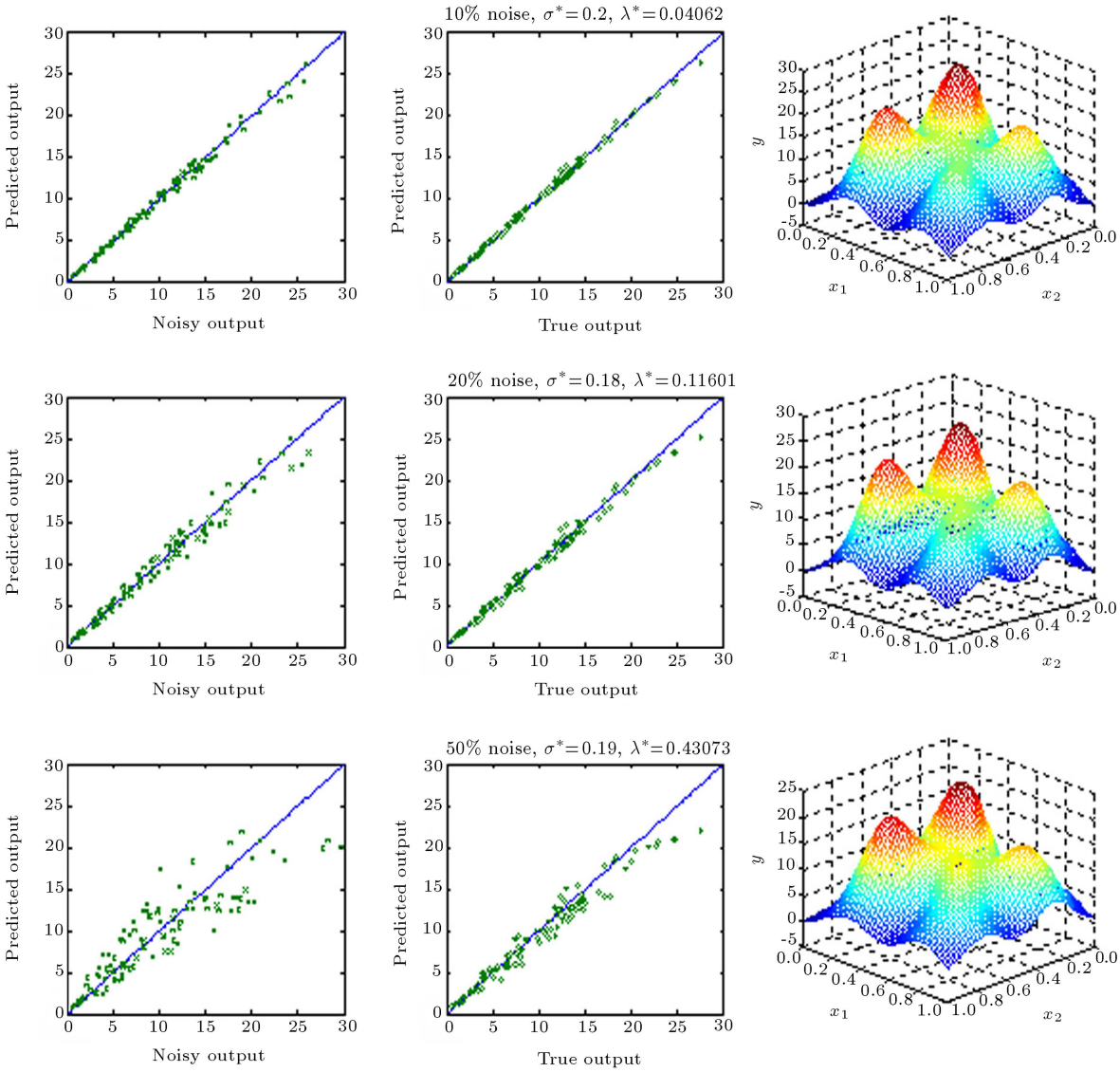

Figure 9. Generalization performance of regularization network at optimum value of isotropic spread and optimal regularization level.

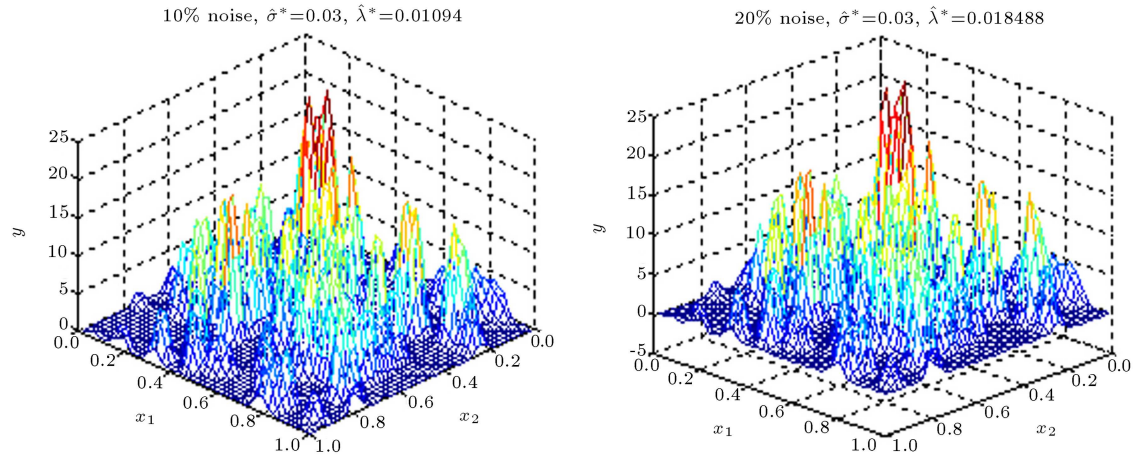

Figure 10. Generalization performance of regularization network with spreads corresponding to the first local maxima at λ^* for relatively low noise levels.

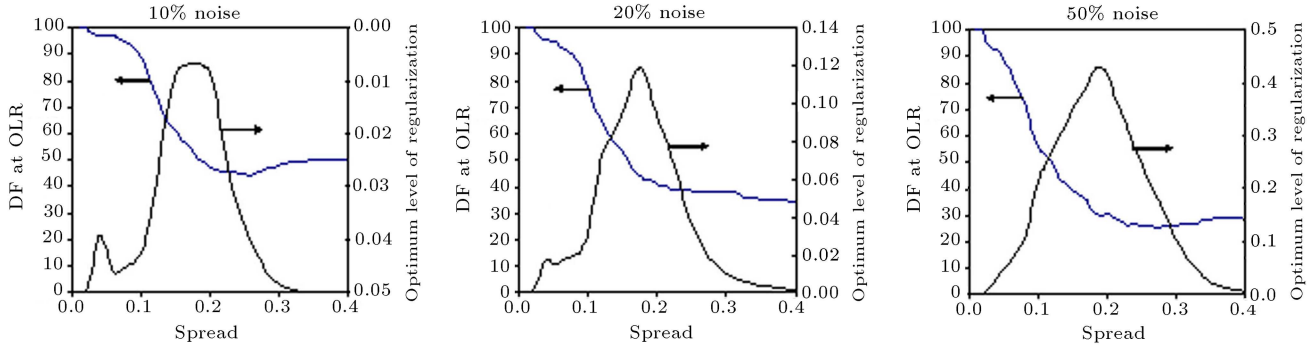

Figure 11. Variations of the degrees of freedom and optimum regularization level with optimum isotropic spread of the Gaussian basis functions regularization network.

REFERENCES

1. Himmelblau, D.M. and Hoskins, J.C. "Artificial neural network models of knowledge representation in chemical engineering", *Computers and Chemical Engineering*, **12**, pp. 881-890 (1988).
2. Venkatasubramanian, V. and Chan, K. "A neural network methodology for process fault diagnosis", *AIChE Journal*, **35**, pp. 1993-2002 (1989).
3. Watanabe, K., Matsuura, I., Abe, M., Kubota, M. and Himmelblau, D.M. "Incipient fault diagnosis of chemical engineering processes via artificial neural networks", *AIChE Journal*, **35**(11), pp. 1803-1812 (1989).
4. Shahsavand, A. "Optimal and adaptive radial basis function neural networks", PhD. Thesis, University of Surrey, UK (2000).
5. Carr, J.C., Beatson, R.K., McCallum, B.C., Fright, W.R., McLennan, T.J. and Mitchell, T.J. "Smooth surface reconstruction from noisy range data", *ACM GRAPHITE 2003 Proceeding*, Melbourne, Australia, pp. 119-126 (2003).
6. Haykin, S., *Neural networks: A Comprehensive Foundation*, Second Ed., New Jersey, Prentice Hall (1999).
7. Poggio, T. and Girosi, F. "Regularization algorithms for learning that are equivalent to multilayer networks", *Science*, **247**, pp. 978-982 (1990).
8. Poggio, T. and Girosi, F. "Networks for approximation and learning", *Proceedings of the IEEE*, **78**, pp. 1481-1497 (1990).
9. Hunt, K.J., Sbarbaro, D., Zbikowski, R. and Gawthrop, P.J. "Neural networks for control systems - a survey", *Automatica*, **28**(6), pp. 1083-1112 (1992).
10. De Nicolao, G. and Ferrari-Trecate, G. "Regularization networks: Fast weight calculation via kalman filtering", *IEEE Trans. Neural Networks*, **12**(2), pp. 228-235 (2001).
11. Carr, J.C., Beatson, R.K., Cherrie, J.B., Mitchell, T.J., Fright, W.R., McCallum B.C. and Evans, T.R. "Reconstruction and representation of 3D objects with radial basis functions", *ACM SIGGRAPH 2001 Proceeding*, CA, USA, pp. 67-76 (2001).
12. Micchelli, C.A. "Interpolation of scattered data: Distance matrices and conditionally positive definite functions", *Constructive Approximation*, **2**, pp. 11-12 (1986.)
13. Powell, M.J.D., *Radial Basis Function for Multivariate Interpolation, a Review*, Clarendon Press, Oxford (1987).
14. Powell, M.J.D. "Radial basis function approximation to polynomials", *Proceedings of 1987 Dundee biennial Numerical Analysis Conference*, pp. 223-241 (1987).
15. Sugiyama, M. and Ogawa, H. "Optimal design of regularization term and regularization parameter by subspace information criterion", *Neural Networks*, **15**(3), pp. 349-361 (2002).
16. Golub, G.H. and Van Loan, C.G., *Matrix Computations*, Johns Hopkins University Press, Baltimore, 3rd Ed. (1996).
17. Hastie, T.J. and Tibshirani, R.J., *Generalized Additive Models*, Chapman and Hall, London, First Ed. (1990).
18. Golub, G.H., Heath, M. and Wahba, G. "Generalized cross validation as a method for choosing a good ridge parameter", *Technometrics*, **21**(2), pp. 215-223 (1979).
19. Shahsavand, A. and Ahmadpour, A. "Application of optimal RBF neural networks for optimization and characterization of porous materials", *Computers and Chemical Engineering*, **29**, pp. 2134-2143 (2005).
20. Shahsavand, A. and Pourafshari, M. "Neural network modeling of hollow fiber membrane process", *Journal of Membrane Science*, **297**, pp. 59-73 (2007).