Journal of Cleaner Production 172 (2018) 3028–3041

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Assessment of energy consumption and modeling of output energy for wheat production by neural network (MLP and RBF) and Gaussian process regression (GPR) models

Morteza Taki [a], Abbas Rohani [b,*], Farshad Soheili-Fard [a], Abbas Abdeshahi [c]

[a] Department of Agricultural Machinery and Mechanization, Ramin Agriculture and Natural Resources University of Khuzestan, Mollasani, Iran
[b] Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
[c] Department of Agricultural Economics, Ramin Agriculture and Natural Resources University of Khuzestan, Mollasani, Iran

A R T I C L E I N F O

Article history:
Received 18 May 2017
Received in revised form
18 July 2017
Accepted 15 November 2017
Available online 20 November 2017

Keywords:
Neural networks
Spread parameter
Energy
Wheat production

A B S T R A C T

The objective of this study was to predict the irrigated and rainfed wheat output energy with three soft computing models include Artificial Neural Network (MLP and RBF models) and Gaussian Process Regression (GPR) for the first time, in Shahreza city, Isfahan province, Iran. Data were collected from an extensive research on wheat farms, including 120 irrigated and 90 rainfed wheat farms (totally 210 questionnares) at three levels (small: < 2ha, medium: 2–4 ha and large: > 4ha) using with face to face questionnaire method. Results of energy analysis showed that diesel fuel was the most influential factor on energy consumption in irrigated wheat production and also for medium and large lands of rainfed wheat production, but for small rainfed lands, total of fertilizers and poisons had the highest impact on total energy consumption. Results of output modeling showed that ANN-RBF model is more accurate than MLP-ANN and GPR models. RMSE and MAPE for irrigated and rainfed output modeling for ANN- RBF were 63.12–72.30 MJ and 0.05–0.14%, respectively. The results of selecting the best spread factor (one of the best parameters on RBF model performance) showed that for irrigated wheat with LM algorithm and at training phase (irrigated-LM-training) and irrigated wheat with BR algorithm and at training phase (irrigated-BR-training), this factor is equal to 7 and 4, respectively. The ANN-RBF model developed was capable of predicting irrigated and rainfed wheat output energy under different land size and using input energies.

1. Introduction

Wheat (*Triticum aestivum* L.) is one of the oldest and most cultivated agricultural crops and has the second rank after maize in the world cereal output production. This plant is very important because uses as a main food for more than half of the world's people, therefore, should be considered as a strategic crop (Taki et al., 2012a). Iran with 15.03 trillion tons of wheat production was the 12th country among the biggest wheat producers in 2010. (Khoshnevisan et al., 2013). Irrigated wheat covers only one-third of the total wheat area, thus the bulk of the wheat crop depends on rainfall. Most of the rainfed wheat crop is concentrated in the west and northwestern regions of Iran. General reports show that the Iran government, plans to improve irrigation by introducing modern irrigation systems to 450000 ha of farmland. So the researches on the innovative methods for increasing the wheat production and also optimizing the input energy should be developed and continued (Abdi et al., 2012).

Energy is one of the important elements in modern agriculture as it depends deeply on fossil and other energy sources (Safa and Samarasinghe, 2011). Energy consumption in agriculture has been increasing in response to the limited supplies of arable land, population growth, changes in technology and an attention for higher standards of living (Kizilaslan, 2009). In the other hands, energy prediction and modeling is an interesting subject for engineers and researchers who are concerned with energy consumption and production with related to environmental hazards (Al-Ghandoor et al., 2009). In the energy subject, a different range of models have been used; from geological models in research on physical resources to modeling future energy request (Safa et al., 2009).

* Corresponding author.
E-mail address: arohani@um.ac.ir (A. Rohani).

https://doi.org/10.1016/j.jclepro.2017.11.107
0959-6526/

Nomenclature

ANN	Artificial Neural Networks
RMSE	Root Mean Squared Error
GHG	greenhouse gas
MAPE	Mean Absolute Percentage Error
R^2	coefficients of determination
EF	Model Efficiency
MLR	Multiple Linear Regression
SFP	Seed, Fertilizer and Pesticide
ANFIS	Adaptive Neuro-Fuzzy Inference System
LM	levenberg-Marquardt backpropagatio
RBF	Radial Bias Function
BR	Bayesian Regularization backpropagatio
MLP	Multilayer Perceptron
OLS	Orthogonal least square
GPR	Gaussian Process Regression
GP	Gaussian process
BB	Basic Back-propagation
LCA	Life Cycle Assessment

During the past 15 years there has been an essential increase in the interest on Artificial Neural Networks (ANN). The basis of ANN modeling methods is biological neuron activities. (Pachepsky et al., 1996; Taki et al., 2012b). In agricultural sector, a few scientists have used ANN model to predict the future yield and energy of wheat production. Khoshnevisan et al. (2013), presented ANN to predict wheat production yield and (greenhouse gas) GHG emissions on the basis of energy inputs for some region of Isfahan province, Iran. Several ANN models were developed and the prediction accuracy of them was evaluated using the statistical parameters. Results showed that the ANN model with 11-3-2 structure was the best one for predicting the wheat yield and GHG emissions. The coefficients of determination (R^2) of the best topology were 0.99 and 0.998 for wheat yield and GHG emissions, respectively. Safa and Samarasinghe (2011), developed ANN models for determining energy consumption in wheat production. They compared ANNs with Multiple Linear Regression (MLR) and found that ANNs can predict energy consumption better than MLR. Khoshnevisan et al. (2014) developed an intelligent system based on Adaptive Neuro-Fuzzy Inference System (ANFIS) for predicting wheat grain yield on the basis of energy inputs in Fereydonshahr region, Iran. The results illustrated that ANFIS model can predict the yield more precisely than ANN. Nabavi-Pelesaraei et al. (2017) developed mathematical models to optimize total energy consumption and environmental pollution for paddy production. Optimization results indicated that energy savings by employing mathematical models are 21.15%–71.63%, respectively. They indicated that the enhancement of the efficiency of paddy production is mainly in terms of toxins and chemical fertilizers. All the prediction models focused on ANN and MLR models and the results of these researches always concluded that ANN models are good tools to solve the nonlinearity relation between inputs and outputs of agricultural production and solve the prediction equations.

With respect to above literature, there is a lack of study on irrigated and rainfed wheat production in Shahreza city (one of the important plains of wheat production in Iran). So, the main goals of this study are to energy assessment and prediction of wheat production at different farm size groups (small: < 2ha, medium: 2–4 ha and large: > 4 ha) and different irrigation system (irrigated and rainfed) basis of input energy using ANN with two training algorithm: Multilayer Perceptron (MLP) and Radial Bias Function (RBF) and also a new soft computing model: Gaussian Process Regression (GPR). Energy survey of wheat production in this region, application of GPR model for the first time and using RBF learning algorithem are the main novelity of this research. Therefore, in this study, the above models are developed and qualitative parameters are utilized to predict the models accuracy to estimate output energy for wheat production. The results of this research can help the farmers to estimate total output energy (total yield) to choice a good decision for marketing and other finicial subjects.

2. Methodology

2.1. Selection of case study region and data processing

The Isfahan province is located within 30–42° and 34-30° north latitude and 49-36° and 55-32° east longitude. Shahreza is located 508 km south to Tehran city and about 80 km south west to Isfahan city and Zard Kooh mountain chain runs from north-west to south-east of the city, enjoying a cold climate. In this study, two wheat farms were considerate; irrigated and rainfed. For each category, three samples were selected, i. e small farms (<2 ha), medium farms (2–4 ha) and large farms (>4 ha). The data were collected from 120 irrigated and 90 rainfed wheat farms in Shahreza city in Isfahan province using face to face questionnaire method for yield period of 2014–2015. The questionnaire included several questions about the use of various inputs (fuel, electricity, fertilizers, biocides, etc.), the amount of land cultivated, wheat yield per year, total working hours of labor over the total stages (from land preparation to wheat harvest), total working hours of machinery and equipment, etc. A brief summary of the sample questionnaire is provided in Table 1.

The objective farms were randomly selected from the rural communities in the research region. The sample size was calculated based on the derivation of Neyman technique as presented by Eq. (1) (Taki et al., 2012c):

$$n = \frac{\sum N_s P_s}{N^2 D^2 + \sum N_s P_s^2} \tag{1}$$

where n is the required sample size; N is the number of total population; N_s is the number of the population in the s stratification; P_s is the standard deviation in the s stratification, P_s^2 is the variance in the s stratification, D^2 is equal to $\frac{d^2}{z^2}$; d is the precision, $(\bar{x} - \bar{X})$ (5%) is the permissible error and z is the reliability coefficient (1.96, which represents 95% reliability). Thus the sample size was found to be 120 irrigated and 90 rainfed wheat farms. Table 2 shows the summary of the sampling wheat farms according to size and type of them.

To survey energy consumption in the production of various agricultural products, it should be revealed whether the energy consumption value covers the stages of production and supply. For this reason, a clear boundary for the production system should be defined and the energy consumption value for the total activities done within the determined boundary of the production system should be achieved. In this study, these activities comprise tillage and land preparation, all activities during the growing season resulting in increased product yield including fertilization, spraying, etc., and activities that lead to energy consumption for the eventual wheat harvest (Taki et al., 2013). Energy consumption beyond the system boundaries, such as for processing after harvest, is not calculated in this study. Also, this research has considered only the energy applied in wheat production, without calculating the natural sources of energy (radiation, wind, rain, etc.). In the

Table 1
A brief summary of the sample questionnaire (Nabavi-Pelesaraei et al., 2016).

Questionnaire No: …
Data: 2015/ … / …
The total area under cultivation (ha): …
Duration of the production: …
Crop yield per hectare (kg): …
Number of fixed labors: …
Number of daily labors: …
Daily working hours: …
Machinery operation used: …
Types of machinery used: …
Total weight of machinery per year (kg): …
Total diesel consumption (L): …
Types of chemical fertilizers: …
Total weight of chemical fertilizers from each type (kg): …
Types of chemical biocides: …
Total weight of biocides from each type (kg): …
Total weight of FYM use (kg): …

Table 2
Summery of wheat farms types and number of sampling for each one.

Type of wheat farms	Small (<2 ha)	Medium (2–4 ha)	Large (>4 ha)
Irrigated wheat farms	50	40	30
Rainfed wheat farms	20	40	30
Total	70	80	60

agriculture sector, the input and output energy sources are limited and all have standard energy coefficients. Accordingly, the energy amount of each input and output can be calculated by multiplying the physical amount with the energy coefficient. In this research, the energy input coefficients were achieved from literature, which are given in Table 3. These coefficients are fixed and not related to product type. For example, human labor and diesel fuel in the production of various products are of the same nature and have constant coefficients for the conversion to their energy consumption equivalents. Thus, the difference in energy consumption for agricultural production is the input value. For this reason, determining the energy consumption for wheat production using standard coefficients shows a certain pattern of energy consumption in wheat production. It also enables comparing the results with other studies using standard energy coefficients for the calculation of energy consumption.

The energy requirements of the inputs and outputs are calculated by multiplying the quantity of each item (unit ha^1) with their energy equivalents ($MJunit^1$) from Table 3. For calculating the epitomized energy in machinery section, it was assumed that the energy consumed for the production of the tractors and other agricultural equipments be amortized during their economic life time (Taki et al., 2012c); so, the machinery input energy was calculated using the following Eq (Abdi et al., 2013):

$$ME = \frac{G \times MP \times t}{T} \tag{2}$$

where ME is the machinery energy per unit area (MJ ha^{-1}); G is the machine mass (kg); MP is the production energy of machine (MJ kg^{-1}); t is the time that machine used per unit area (h ha^{-1}) and T is the economic life time of machine (h).

The energy equivalent of human labor is the muscle power used in field operations of crop production. Pesticides and chemical fertilizers energy equivalents means the energy consumption for producing, packing and distributing the materials and they are given on an active ingredient basis. Farmyard manure is regarded as a source of nutrients, so the energy equivalent of farmyard manure is equated with that of mineral fertilizer equivalents corresponding to the fertilization effect of the applied manure. Also, the energy sequestered in diesel fuel mean their heating value (Enthalpy) and the energy needed to make their energy available directly to the farmers. Moreover, the seed energy is the energy used in the production of a crop and the grain energy is the gross energy content determined from laboratory bomb calorimeter tests (Bahrami et al., 2011). The energy equivalent of water for irrigation input means indirect energy of irrigation consist of the energy consumed for manufacturing the materials for the dams, canals, pipes, pumps, and equipment as well as the energy for constructing the works and building the on-farm irrigation (Ranjbar et al., 2013).

2.2. Artificial neural network

ANNs are regarded as an analytical method of simulating system performance and were inspired by the principles of data processing in the brain. The method relies on experimental data used to 'train' the ANN so that it can precisely predict system performance (Najafi et al., 2009). Prior to any ANN training process with the trend free data, the data must be normalized over the range of [0, 1]. For ANN models, this is necessary for the neurons' transfer functions, because a sigmoid function is calculated and consequently these can only be performed over a limited range of values. If the data used with an ANN are not scaled to an appropriate range, the ANN will not converge on training or it will not produce meaningful results.

The most commonly employed method of normalization

Table 3
Energy coefficients of different inputs used and output in wheat production.

Inputs		Unit	Energy equivalent (MJ $unit^{-1}$)	Reference
1.Human labor	Man	h	1.96	Ranjbar et al., 2013
	Woman	h	1.57	
2.Chemical fertilizer and farmyard manure	N	kg	47.1	Taki et al., 2012c
	P_2O_5	kg	15.8	
	K_2O	kg	9.28	
	farmyard manure	kg	0.3	
3.Chemical poison	Pesticides	kg	101.2	Ranjbar et al., 2013
	Herbicide	kg	238	
4.Agricultural machinery	Tractor	kg year	93.61	Abdi et al., 2012
	Implement and machinery	kg year	62.7	
	Combine and mower	kg year	87.63	
5. Seed	Improved seed or hybrid	kg	25	Abdi et al., 2013
	General seed	kg	15.7	
6.Diesel fuel		L	47.8	Ranjbar et al., 2013
7.Electricity		kWh	10.59	Taki et al., 2012b

involves mapping the data linearly over a specified range, whereby each value of a variable x is transformed as follows (Rohani et al., 2011):

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \times (r_{max} - r_{min}) + r_{min} \tag{3}$$

where x is the original data, x_n the normalized input or output values, x_{max} and x_{min} are the maximum and minimum values of the concerned variable, respectively. r_{max} and r_{min} an correspond to the desired values of the transformed variable range. A range of 0.1–0.9 is appropriate for the transformation of the variable onto the sensitive range of the sigmoid transfer function.

2.2.1. MLP algorithm

MLP is a feed-forward layered network with one input layer, one output layer, and some hidden layers. Every node computes a weighted sum of its inputs and passes the sum through a soft nonlinearity. The soft nonlinearity or activity function of neurons should be non-decreasing and differentiable. In this research, two transfer functions were used; hyperbolic tangent sigmoid transfer function (Taki et al., 2016a):

$$f(\theta) = \frac{1}{1 + e^{-\theta}} \tag{4}$$

And log-sigmoid transfer function:

$$f(\theta) = \frac{2}{1 + e^{-2\theta}} - 1 \tag{5}$$

The network is in charge of vector mapping, i.e. by inserting the input vector, x^q the network will answer through the vector z^q in its output (for $q = 1, 2, \ldots, Q$). The aim is to adapt the parameters of the network in order to bring the actual output z^q close to the corresponding desired output d^q (for $q = 1, 2, \ldots, Q$). Training algorithm is based on minimization of a suitable error cost function (Taki et al., 2016b). In this research, Basic Back-propagation (BB) algorithm with two training roles (levenberg-Marquardt backpropagation and bayesian regularization backpropagation) was employed. No transfer function was used for the first layer. For the hidden layers the sigmoid functions were used, and for the output layer a linear transfer function was applied as desired for estimating problems.

2.2.2. RBF algorithm

One type of ANN is the radial basis function (RBF) neural network, which uses radial basis functions as activation functions. This ANN is a linear combination of radial basis functions. RBF networks form a special architecture of neural networks that present important advantages compared to other neural network types, including simpler structure and faster learning algorithms (Iliyas et al., 2013). RBF is a feed-forward neural network model with good performance and it has already proven its universal approximation ability with no local minima problem (Iliyas et al., 2013). An RBF has a single hidden layer. Each node of the hidden layer has a parameter vector called center. This center is used to compare with the network input vector to produce a radially symmetrical response. Responses of the hidden layer are scaled by the connection weights of the output layer and then combined to produce the network output. For a single output, where the outputs of the nonlinear activation are combined linearly with the weight vector (β) of the output layer, y_m can be calculated (Iliyas et al., 2013):

$$y_m = \sum_{i=0}^{M} \beta_i \varphi_i \tag{6}$$

In which β_i is the joint weighted value of the *ith* basis function. The most commonly used radial base is the Gaussian function given as (Haykin, 2009):

$$\varphi_i(x) = \exp\left(-\frac{\|x - c_i\|^2}{\sigma_i^2} \right) \tag{7}$$

where c_i and σ_i are center and spread of the *ith* RBF node.

Training the radial basis function neural network is done in two steps. In the first step centers are selected from the training data (without training) or constructed by clustering the training data. The second step is basically a linear estimation of one weighting vector using ordinary least squares. RBF is an interpolating network. It can be built using all the available training points, or it can be built using reduced number of points. The selection of centers can then be performed by clustering the training data. There are three types of learning strategies used in selecting RBF centers; fixed randomly selected centers, self-organized center selection, and supervised selection of centers. A number of training algorithms have been developed for training of RBF networks. Orthogonal least square (OLS) (Taki et al., 2016a) techniques are self-organized technique, and have been employed to select centers so that adequate and efficient RBF network can be obtained. The OLS uses gram-Schmidt algorithm for center selection and center updating of RBF network, while adaptive gradient descent procedure, described in Haykin (2009), was used to adapt the weights. The network parameters are found such that they minimize a cost function:

$$\min J = \sum_{i=1}^{Q} \left(|y_{mi} - y_{di}|^2 \right) \tag{8}$$

where Q is the number of training pattern, while y_m and y_d are the network output and desire target output, respectively.

In this study, two algorithms for ANN model were used as below: Bayesian regulation backpropagatio (BR) and Levenberg-Marquardt backpropagation (LM). The main problem with implementing regularization is setting the correct values for the objective function parameters. The Bayesian framework for neural networks is based on the probabilistic interpretation of network parameters. That is, in contrast to conventional network training where an optimal set of weights is chosen by minimizing an error function, the Bayesian approach involves a probability distribution of network weights. As a result, the predictions of the network are also a probability distribution (Sorich et al., 2003; Xu et al., 2006). In the training process, a common performance function is used for computing the distance between real and predicted data. This function can be expressed as follows (Kayri, 2016):

$$F = E_D(D|\omega, M) = \frac{1}{N} \sum_{i=1}^{n} \left(\hat{t}_i - t_i \right)^2 \tag{9}$$

here, E_D is the mean sum of squares of the network error; D is the training set with input-target pairs. M is a neural network architecture that consists of a specification of the number of layers, the number of units in each layer, and the type of activation function performed by each unit. E_D is a criterion for early stopping to avoid over fitting; it is used in MATLAB for many training algorithms. Therefore, early stopping for regularization seems to be a very crude method for complexity control (Mackay, 1996; Kayri, 2016).

However, although the early stopping regularization can reduce the variance it increases the bias. Both can be reduced by BR (Kelemen and Liang, 2008). In a BR network, the regularization adds an additional term and then an objective function to penalize large weights that may be introduced in order to obtain smoother mapping. In this case, a gradient-based optimization algorithm is preferred for minimizing the objective (Gianola et al., 2012; Okut et al., 2013):

$$F = \beta E_D(D|\omega, M) + \alpha E_W(\omega|M) \tag{10}$$

where, $E_W(\omega|M)$ is $E_W = \frac{1}{n}\sum_{i=1}^{n}\omega_j^2$, the sum of squares of network weights, α and β, are hyperparameters that need to be estimated function parameters. The last term, $\alpha E_W(\omega|M)$ is called weight decay and α is also known as the decay rate. If $\alpha \ll \beta$ then the training algorithm will make the errors smaller. If $\alpha \gg \beta$, training will emphasize weight size reduction at the expense of network errors, thus producing a smoother network response (Shaneh and Butler, 2006).

After the data are taken with the Gaussian additive noise assumed in target values, the posterior distribution of the ANN weights can be updated according to Bayes' rule:

$$P(\omega|\mathrm{D}, \alpha, \beta, \mathrm{M}) = \frac{P(D|\omega, \beta, \mathrm{M})\cdot \mathrm{P}(\omega|\alpha, \mathrm{M})}{P(D|\alpha, \beta, \mathrm{M})} \tag{11}$$

Therefore, the BR includes a probability distribution of network weights and the network architecture can be identified as a probabilistic framework (Sorich et al., 2003). In Equation (6), D is the training sample and the prior distribution of weights is defined as: $\mathrm{P}(\omega|\alpha, \mathrm{M}) = \left(\frac{\alpha}{2\pi}\right)^{\frac{m}{2}}\exp\left\{-\frac{\alpha}{2}\omega'\omega\right\}$M is the particular ANN used and w is the vector of networks weights. $\mathrm{P}(\omega|\alpha, \mathrm{M})$ states our knowledge of weights before any data is collected, $P(D|\omega, \beta, \mathrm{M})$ is the likelihood function which is the probability of the occurrence, giving the network weights. In this Bayesian framework, the optimal weights should maximize the posterior probability $P(\omega|\mathrm{D}, \alpha, \mathrm{P}, \mathrm{M})$. Maximizing the posterior probability of w is equivalent to minimizing the regularized objective function ($\mathrm{F} = \beta \mathrm{E}_D + \alpha E_\omega$):

$$P(\alpha, \beta|\mathrm{D}, \mathrm{M}) = \frac{P(D|\alpha, \beta, \mathrm{M})\cdot \mathrm{P}(\alpha, \beta|\mathrm{M})}{P(D|, \mathrm{M})} \tag{12}$$

According to Mackay (1996) it is:

$$P(D|\alpha, \beta, \mathrm{M}) = \frac{P(D|\mathrm{w}, \beta, \mathrm{M})\cdot \mathrm{P}(\mathrm{w}|\alpha, \mathrm{M})}{P(w|\mathrm{D}, \alpha, \beta, \mathrm{M})} = \frac{Z_F(\alpha, \beta)}{(\pi/\beta)^{n/2}(\pi/\alpha)^{m/2}} \tag{13}$$

where n and m are the number of observations and total number of network parameters, respectively. Equation (13) (Laplace approximation) produces the following equation:

$$Z_F(\alpha, \beta) \propto \left|H^{MAP}\right|^{\frac{-1}{2}} \exp\left(-F\left(\omega^{MAP}\right)\right) \tag{14}$$

where H^{MAP} is the Hessian matrix of the objective function and MAP stands for maximum a posteriori. The Hessian matrix can be approximated as:

$$H = J'J \tag{15}$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to network parameters. J calculated by:

$$J = \begin{bmatrix} \frac{\partial e_1(\omega)}{\partial \omega_1} & \frac{\partial e_1(\omega)}{\partial \omega_2} & \cdots & \frac{\partial e_1(\omega)}{\partial \omega_n} \\ \frac{\partial e_2(\omega)}{\partial \omega_1} & \frac{\partial e_2(\omega)}{\partial \omega_2} & \cdots & \frac{\partial e_2(\omega)}{\partial \omega_n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & & \cdot \\ \frac{\partial e_N(\omega)}{\partial \omega_1} & \frac{\partial e_N(\omega)}{\partial \omega_2} & \cdots & \frac{\partial e_N(\omega)}{\partial \omega_n} \end{bmatrix} \tag{16}$$

The Gauss-Newton approximation versus the Hessian matrix ought to be used if the LM algorithm is employed to replace the minimum of F (Shaneh and Butler, 2006), an approach that was proposed by (Mackay, 1996). In LM, algorithm parameters at l iteration are updated as:

$$\omega^{l+1} = \omega' - \left[J^T J + \mu I\right]^{-1} J^T e \tag{17}$$

where μ is the Levenberg's damping factor. μ is adjustable for each iteration and leads to optimization. It is a popular alternative to the Gauss-Newton method of finding the minimum of a function (Souza, 2015).

A typical MLP (for irrigated wheat) and RBF network (for rainfed wheat) configurations are shown in Fig. 1. For RBF model, levenberg-Marquardt backpropagation and bayesian regularization backpropagation algorithem were used for training. A typical flowchart of MLP and RBF neural network for prediction of irrigated and rainfed wheat production and all the analytical process is shown in Fig. 2.

2.3. Gaussian process regression model (GPR)

GPR works under the probabilistic regression framework, which takes as input a training data set $\mathrm{D} = \{(\mathrm{y_n}, \mathrm{x_n}),\ \ \mathrm{n} = 1, 2, 3, \ldots, \mathrm{N}\}$ of N pairs of vector input $\mathrm{x_n \hat{I} R^L}$ and noisy scalar output yn, and constructs a model that generalizes well to the distribution of the output at unseen input locations. The noise in the output models uncertainty due to factors external to x, such as truncation or observation errors. Here we assume that noise is additive, zero-mean, stationary and normally distributed, such that:

$$\mathrm{y} = \mathrm{f(x)} + \varepsilon, \quad \varepsilon \text{»} \mathrm{N}\left(0, \mathrm{s}^2_{\mathrm{noise}}\right) \tag{18}$$

where $\mathrm{s}^2_{\mathrm{noise}}$ is the variance of the noise (Wan and Sapsis, 2017). The primary idea behind GPR is to use a Gaussian process (GP) to represent f, referred to as latent variables. The input x plays the role of indexing these latent variables such that any finite collection $\{\mathrm{f(x_1)}, \ldots, \mathrm{f(x_k)}\}$ with unique indices follow a consistent Gaussian distribution. In this way, we limit ourselves to only looking at functions whose values correlate with each other in a Gaussian manner. In Bayesian framework, this is equivalent to putting a GP prior over functions. Due to the consistency requirement, we are able to make inference on function values corresponding to unseen inputs conveniently using a finite set of training data.

A major advantage for using the Gaussian prior assumption is that functions can be conveniently specified by a mean function m(x) and a covariance function $\mathrm{k(x, x')}$:

$$m(x) = E[f(x)], \quad k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] \tag{19}$$

where E[.] denotes expectation. The form of the mean function is

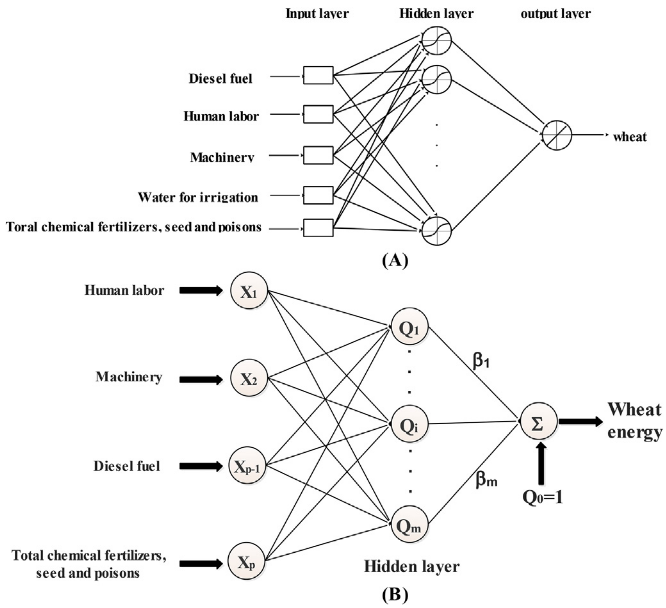

Fig. 1. A typical MLP (A) and RBF (B) network configurations for prediction of irrigated and rainfed wheat output energy (Iliyas et al., 2013).

important only in unobserved region of the input space and usually set to zero. The properties of the process are then entirely dictated by the covariance function, which is by definition symmetric and positive semi-definite when evaluated at any pair of points in the input space (Rasmussen and Williams, 2005). The covariance function typically contains a number of free parameters called hyper parameters which define the prior distribution on f(x). The most commonly used is the squared exponential covariance function:

$$k(x, x') = q_1 \exp\left(\frac{\|x - x'\|}{2q_2}\right) \tag{20}$$

where, $\|.\|$ is a norm defined on the input space. Note that this covariance function decays rapidly when evaluated at increasingly distant pairs of input x and x′, indicating weak correlations between f(x) and f(x′). (q1) is a hyper parameter specifying the maximum allowable covariance. (q_2) is a strictly positive hyper parameter defining rate of decay in correlation as points become farther away from each other. Another hyper parameter (q_3) which is not expressed explicitly in (Eq. (19)), is used to represent the unknown variance s^2_{noise} of Eq. (18). The hyper parameters $\{q_1, q_2, q_3\}$ are grouped together as a vector (q) treated as the realization of a random vector (Q). The realization that is most coherent with the data set is selected using training data and then used to make predictions. Assuming that the hyper parameters are known, inference is easily made. Denoting the vector of training latent variables by f and the vector of test latent variables by f^*we have the following joint Gaussian distribution:

$$p(f, f^*) = N\left(0, \begin{bmatrix} K_{f,f} & K_{*,f} \\ K_{f,*} & K_{*,*} \end{bmatrix}\right) \tag{21}$$

K is the symmetric covariance matrix whose ijth entry is the covariance between the ith variable in the group denoted by the first subscript and the jth variable in the group denoted by the second subscript (* is used in place of f* for short), computed using covariance function $k(.,.)$ in Eq. (21) and corresponding hyper parameters (Wan and Sapsis, 2017). The prediction framework is

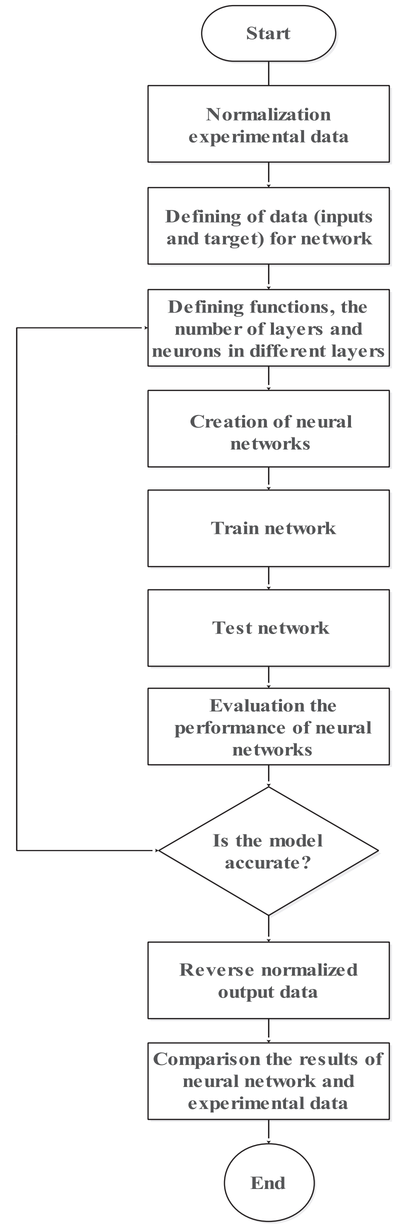

Fig. 2. Procedure of modeling and process of data preparing in the ANN (MLP and RBF) models.

shown in Fig. 3.

The sample size used in this study was 120 irrigated and 90 rainfed wheat farms. Embedded energies (including human labor, total chemical fertilizers with seed and poisons, diesel fuel, irrigation water and machinery for irrigated farms and all of above inputs except of irrigation water for rainfed farms) were chosen as inputs while total wheat grains was selected as output of the three soft computing models. For MLP, RBF and GPR models, a computer code developed in MATLAB software.

2.4. Performance evaluation criteria

To evaluate the performance of a model some criteria have been used from the literature. These criteria include: Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Model Efficiency (EF). This statistical parameters are defined as follows:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{j=1}^{n}\left|\frac{d_j - p_j}{d_j}\right| \times 100 \tag{22}$$

$$\mathrm{RMSE} = \sqrt{\sum_{j=1}^{n}\left(\mathrm{d_j} - \mathrm{p_j}\right)^2} \tag{23}$$

$$\mathrm{EF} == \frac{\sum_{j=1}^{n}\left(d_j - \overline{d}\right)^2 - \sum_{j=1}^{n}\left(p_j - d_j\right)^2}{\sum_{j=1}^{n}\left(d_j - \overline{d}\right)^2} \tag{24}$$

where d_j is the *ith* component of the desired (actual) output for the *jth* pattern; p_j is the component of the predicted (fitted) output produced by the network for the *jth* pattern; $\overline{d}$ and $\overline{p}$ are the average of the whole desired (actual) and predicted output and n is the number of variable outputs. A model with the smallest RMSE, MAPE and largest EF is considered to be the best (Taki et al., 2016b).

3. Results and discussion

3.1. Results of energy consumption

Table 4 shows a summary of energy consumption (MJ ha^{-1}) in three sizes of wheat farm in Shahreza region. Total energy consumption as inputs during irrigated and rainfed wheat production and also output energy were 23406, 9351 and 90528, 44131 38 MJ ha^{-1}, respectively.

The share of each input on total energy output is shown in Fig. 4. As it can show, in all farm types, diesel fuel has the highest impact on energy output except small rainfed farms (<2 ha-rainfed). In those farms, because of small lands, total chemical fertilizer and poisons had the highest impact on total output energy. Khoshnevisan et al. (2013) reported that total input energy for wheat production in Isfahan province (Fereydunshahr city) were 80.4, 79.29 and 81.11 GJ ha^{-1} for small (<1 ha), medium (1–3 ha) and large (>3 ha) farms that this results are very higher than our results but the authors did not explain about this fact. In that research, electricity and chemical fertilizer had the most important effects on output energy (49% and 29% respectively). In a similar study, Safa et al. (2009) reported that total input energy for irrigated and rainfed (dry) wheat production in New Zealand were 25.6 and 17.45 GJ/ha^{-1}. In that research, chemical fertilizer and electricity played the most important role in irrigated farms output with 10.19 and 3.43 GJ/ha^{-1}. In Shahreza city, some of farmers used diesel fuel engines for water pumping and so we calculated that energy for diesel fuels, so the result of diesel fuel consumption in this research is more than other researches. In this region, because of increasing of drought period and also existing of deep wells, farmers have to use much energy for water extraction. Similar results were reported by Abdollahpour and Zaree (2009). They showed that total input and output energy for irrigated wheat production was 25.67 and 21.01 GJ/ha^{-1} at west of Iran (Kermanshah province). In another study in Iran, reported by Tabatabaeefar et al. (2009), the least energy consumed for wheat production was 8.8 MJkg^{-1} in no-till fields and 11.8 MJkg^{-1} in fields with moldboard plow plus roller plus drill. They reported that energy consumed for tillage using moldboard plow plus roller plus drill was 32.5% of the total energy. Taghavifar and Mardani (2015) reported that total energy input and output for wheat production in west Azerbaijan, Iran, were 30626.4 and 53480.4 MJha^{-1}, respectively. It was disclosed that the greatest shares of input energy with 11984 MJha^{-1} and 6824.2 MJha^{-1} corresponded to the diesel fuel and Nitrogen, respectively.

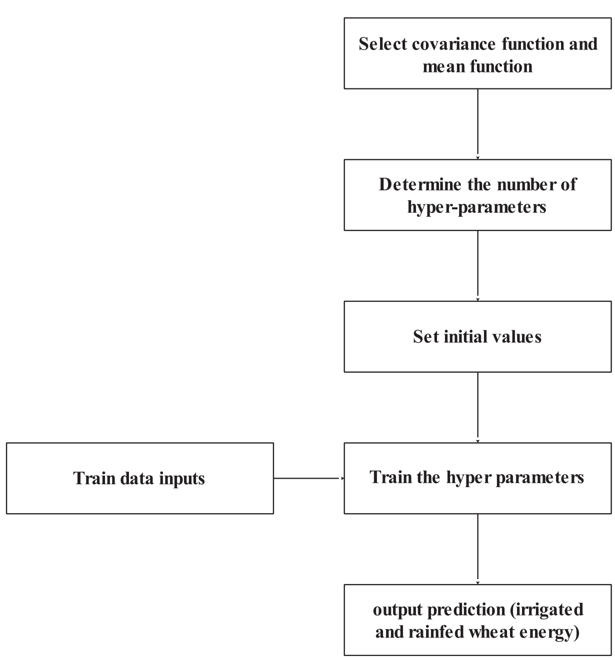

Fig. 3. The prediction framework based on the GPFR model (Liu et al., 2013).

The results of this study revealed that in irrigated large farms (up to 4 ha), total input energy was higher than small farms but the total output energy in small farms was higher than large farms. This fact shows that using of technology was not effective in wheat farms and will be worse by increasing the area of lands. This fact is not novel because we know that this problem. After White Revolution in 1963–1978, farm lands were divided between farmers and so the area of lands were small than last. Now, we consuming energy more than needing of standard wheat production but we cannot harvest equal output energy. One of the good (best) ideas for solving this problem is using of cooperative company for using of total capital of farmers for culturing in large lands. The government should invest in this subject and especially by increasing in drought period in this region; this problem can show other hazards such as increasing in final price of crops and some of social problems (migration and dispute). The results illustrated that the consumption of electricity and chemical fertilizers was high for wheat production in the region. The accurate utilization of chemical fertilizers and application of new electrical pumps for water pumping can improve the energy use efficiency without any dramatic decline in profitability or yield. Table 5 shows the total energy consumption only for operation. As it can show, by changing the farm sizing (small to large), the total machinery energy consumption will increase. This table can complete the above claim about farm sizing in this region and maybe at total farm lands of Iran. Some researchers reported similar results (Ghorbani et al., 2011; Beheshti Tabar et al., 2010; Singh et al., 1999). In recent years, due to the highly mechanized agricultural system in Iran, increasing in total fuel and machinery energy consumption is a sensaible fact.

The results of Table 5 show that, harvesting is the main part at the consuming of total energy at machinery operation. At harvesting section, combine has the main share than other machineries. It can prove above claim about land sizing, again. Using of big machinery such as combine at small farms can increase total energy consumption. In the other hands, using of labors at small farms can increase the final price of crop. For better showing of this subject, future studies should focus on relation between energy

Table 4
Energy inputs and output for wheat production (MJha^{-1}).

Item		Irrigated wheat farms						Rainfed wheat farms					
		Small (<2 ha)		Medium (2–4 ha)		Large (>4 ha)		Small (<2 ha)		Medium (2–4 ha)		Large (>4 ha)	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Inputs	Machinery	658	198	838	241	825	196	311	81	412	157	486	171
	Water for irrigation	3187	3514	4538	4527	5464	4683	0	0	0	0	0	0
	Human labor	203	114	187	92	128	79	143	79	103	72	47	53
	Diesel fuel	6519	1529	7640	2091	7216	1739	3689	561	4302	1107	4438	1021
	Total chemical Fertilizer	5544	871	5210	958	6335	923	3769	844	3177	826	2850	666
	Pesticides	97	102	440	126	460	86	80	101	86	101	87	58
	Herbicides	371	199	156	80	148	91	214	242	249	240	269	239
	Fungicides	42	53	80	54	62	53	58	34	48	38	58	40
	Seed	4857	3021	4680	3155	4335	2662	1080	839	1095	318	1003	558
	Total input energy	21478	–	23769	–	24973	–	9344	–	9472	–	9238	–
Inputs	Average of input energy (for each farm type)	23406						9351					
Output	Output energy (Grain + straw)	95864	44531	94170	38668	81551	35587	54947	14180	45761	19958	31685	13600
Output	Average of output energy (for each farm type)	90528						44131					

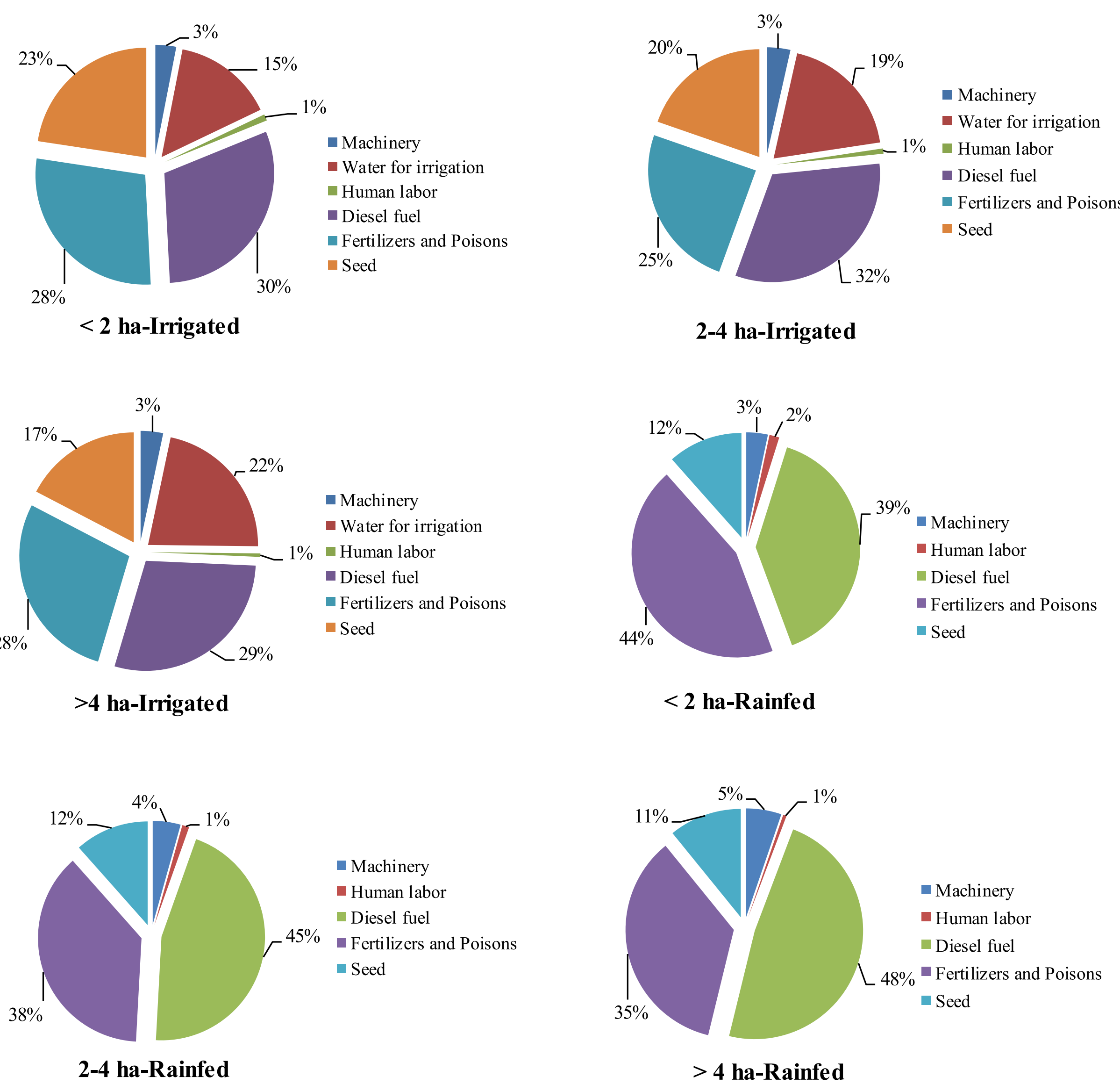

Fig. 4. Share of all inputs for irrigated and rainfed wheat production at different farm size.

consumption, final price of wheat production and sizing of farm lands.

3.2. Evaluation soft computing models

Because of diversity and dispersion of management methods and climatic conditions on the farms, the data sets of energy from them are very different and the standard deviations are very changeable. This makes it impossible to use regression models for energy modeling. The classical regression methods are based on statistical hypotheses, including the normality of the data set. Therefore, the nature of the energy data set has led researchers to use soft computing methods such as artificial neural networks. An important advantage of soft calculation methods is learning the behavior of output variations according to how input changes without taking into account statistical assumptions. So in this research, soft computing methods including MLP, RBF and GPR were used. MLP has been used in most studies (Taki et al., 2012c; Khoshnevisan et al., 2013; Nabavi-Pelesaraei et al., 2016) and RBF is also used in some sources (Taki et al., 2016a; Najafi et al., 2009) for energy modeling. But so far, based on available information, the GPR model has not been used for this purpose. Therefore, one of the objectives of this research is to introduce GPR and compare these three methods together. The limitation of these methods (estimated models) is their validity in the range of variations of the studied variables. Therefore, in order to have a global model, we need to have a complete data set in order to help the modeling of output energy for each product. Therefore, the most important limitation of these models is the lack of a guarantee of outsourcing. This is unthinkable due to lack of access to the entire statistical community.

Two algorithms (BR and LM) for ANN model were used. BR is a mathematical process which converts a nonlinear regression into a statistical issue in the form of series regression. The LM algorithm is a curve-fitting method and the most highly used optimization algorithm that applies the minimum of a several function which is considered as the sum of squares of the errors among the function and the measured data parts (Taki et al., 2016a). For transfer function, logsig and the tansig were used. These methods are very popular and applicable and many researches were used them. For MLP model, we analyzed this method with different hidden layers. The results of this analysis is shown in Fig. 5.

The results of MLP model showed that the best transfer function and learning algorithm are hyperbolic tangent sigmoid and bayesian regularization backpropagation, respectively. The number of neurons in hidden layer for irrigated and rainfed models prediction, are 21 and 17, respectively.

The second model is RBF-ANN. Hidden size (Number of neurons in the hidden layer), spread parameter and learning algorithm are the effective factor for RBF efficiency. So, the next step for RBF model is selecting the best number of neurons in hidden layer and also determining the spread factor. In this case, the output of i-th hidden unit with center μ_i and spread σ_i is given as follows (Esmaeili and Mozayani, 2009):

$$\phi_i(x) = \phi(\|x-\mu_i\|;\sigma) = e^{\left(\frac{\|x-\mu_i\|^2}{2\sigma_i^2}\right)}, \forall_i \tag{25}$$

Training an RBF network consists of finding the values for these parameters, such that the overall approximation or classification error is reduced. The values chosen for the centers and the spread of the radial functions have a great effect on the generalization abilities of the network. Many algorithms have been proposed for finding the centers of an RBF network in literature such as: random subset selection, various clustering algorithms such as K-means and vector quantization, sequential growing and most recently a training algorithm in which the locations and the number of hidden units are tried to be optimized using particle swarms. In all of these works the spread of the radial functions of the hidden units are set either to a fixed value or a value obtained using heuristic methods (Esmaeili and Mozayani, 2009). In this research, the best spread factor with the best number of hidden layer was selected. Fig. 6 shows the results of this analysis. Analysis of selection the best spread factor and hidden size was done for training and testing section. As it can be seen from Fig. 6, the interactions between spread factor and hidden size can be effective on RBF performance. The best combined of these factors (spread factor and hidden size) can be seen with dash line at Fig. 6.

Mostly at this situation, correlation of determination (R^2) will be increased with the number of hidden layers. For irrigated wheat, 17 neurons were selected for both of LM and BR algorithm but for rainfed wheat, 19 and 21 neurons were selected for LM and BR algorithm, respectively. The results of selecting the best spread factor

Table 5
Energy consumption ($MJha^{-1}$) based on operations in wheat production.

Machinery	Irrigated wheat farms						Rainfed wheat farms					
	Small (<2 ha)		Medium (2–4 ha)		Large (>4 ha)		Small (<2 ha)		Medium (2–4 ha)		Large (>4 ha)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Tractor	256	100	254	72	250	81	143	32	153	46	140	39
Moldboard plough	42	11	43	16	42	22	40	11	44	13	42	12
Dick plough	5	11	9	17	12	18	0	0	0	0	0	0
Field cultivator	17	9	15	9	13	10	18	2	17	5	16	4
Disk Hiller	3	3	4	7	4	2	0	0	0	0	0	0
Ditcher	1	2	1	2	3	2	0	0	0	0	0	0
Land Leveler	1	6	9	14	14	19	0	0	0	0	0	0
Grain drill	3	10	15	19	12	15	0	0	0	0	0	0
Combinat	0	0	1	9	3	11	0	0	0	0	0	0
Broadcaster (seeder)	2	5	3	7	6	10	15	3	17	3	18	4
Broadcaster (fertilizer)	3	9	11	17	20	17	0	0	27	13	21	6
Boom-type sprayer	12	11	17	12	20	6	20	7	15	5	19	6
Combine	205	211	372	182	404	158	298	99	341	62	370	80
Mower	27	32	9	22	4	16	51	10	49	11	49	10
Baler	8	27	31	47	8	22	0	0	0	0	0	0
Stationary Thresher	73	68	43	43	32	33	48	12	50	13	40	8
Total	658	198	838	241	847	196	633	81	713	157	715	171

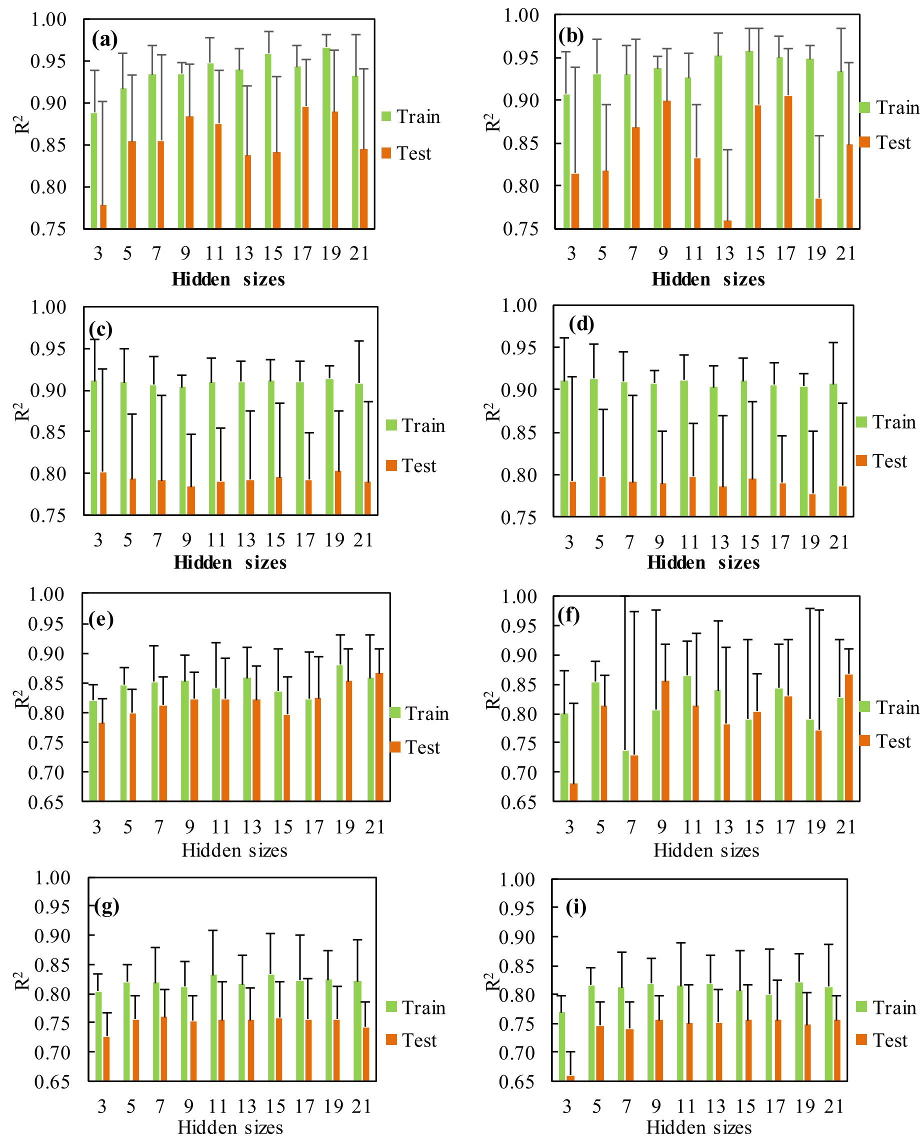

Fig. 5. Changing in R^2 by different hidden layer size (a: Irrigated, LM-Logsig, b: Irrigated, LM-Tansig, c: Irrigated, BR- Logsig, d: Irrigated, BR-Tansig, e: Rainfed, LM-Logsig, f: Rainfed, LM-Tansig, g: Rainfed, BR-Logsig, h: Rainfed, BR-Tansig).

show that for irrigated wheat with LM algorithm and at training phase (irrigated-LM-training) and irrigated wheat with BR algorithm and at training phase (irrigated-BR-training), this factor is equal to 7 and 4, respectively.

The next step for RBF model is selecting the best learning algorithm between LM and BR. The result of this analysis was shown in Fig. 7. As it can be seen, the BR algorithm is more accurate than LM. Between the BR and LM training methods, the BR obtained the highest correlation of determination in terms of predictive ability. Similarly, Okut et al. (2011), proved that the BR training algorithm was the most effective method in terms of predictive ability. Okut et al. (2013) investigated the predictive performance of BR and scale conjugate gradient training algorithms. In that study, they found that the BR gave slightly better performance, but not significantly so. In many studies (Saini, 2008; Ticknor, 2013), the BR training algorithm has given either moderate or the best performance in terms of comparison with other training algorithms. BR-ANNs have some important advantages, such as choice and robustness of model, choice of validation set, size of validation effort, and optimization of network architecture (Burden and Winkler, 2008). Bayesian methods can solve the overfitting problem effectively and complex models are penalized in the Bayesian approach. In contrast to conventional network training, where an optimal set of weights is chosen by minimizing an error function, the Bayesian approach involves a probability distribution of network weights (Wayg et al., 2005). This, combined with its advantage of having the potential ability to capture nonlinear relationships, means it can be used in quantitative studies to provide a robust model.

The third model is GPR. For more comparison, the results of all three models (MLP-RBF and GPR) to predict irrigated and rainfed wheat production output energy are shown in Table 6. As it can show MAPE and RMSE factors for RBF model is very lower than MLP and GPR models. In another hands, EF factor for RBF is higher than MLP and GPR. The GPR method was not applied at any similar literature to predict output energy for agriculture production until now. Application of GPR model in this research is an exactly new one, but as the results can show, this model can predict the output energy similar to MLP model and not higher than RBF. So it is recommended to use this model for other crop productions and

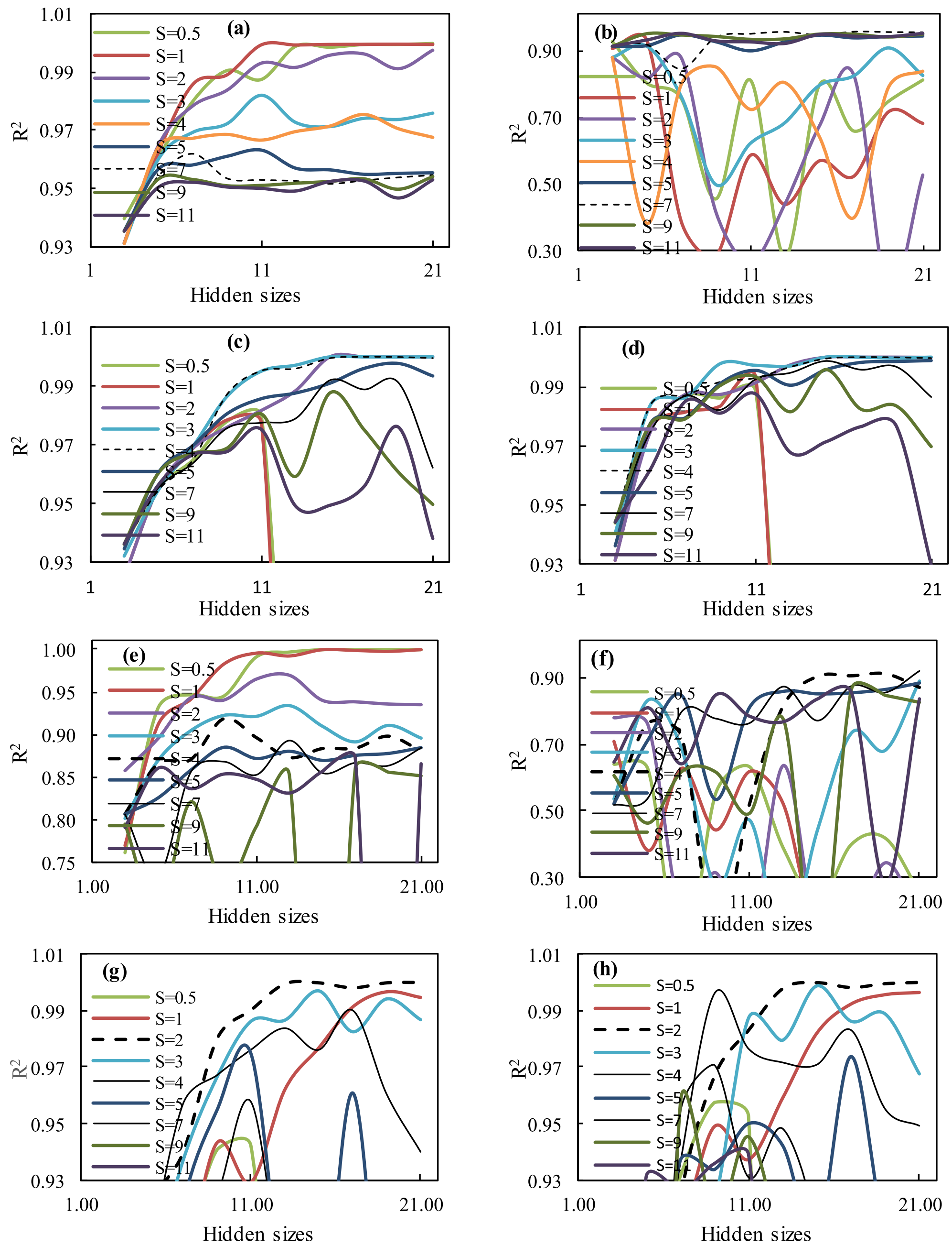

Fig. 6. Changing in R^2 by different hidden layer size (a: Irrigated, Lm-training phase, (b): Irrigated, Lm-test phase, (c): Irrigated, Br, training phase, f(d): Irrigated, Br-test phase, (e): Rainfed, Lm-training phase, (f): Rainfed, Lm-test phase, (g): Rainfed, Br-training phase, (h): Rainfed, Br-test phase).

estimated the ability of this soft computing model. At this time, the results of this model cannot compare with any similar models.

The so called *t-test* was used to compare the means of both series of data (original and predicted) for all three models. It was also assumed that the variance of both samples could be considered equal. The obtained *p* values were greater than the threshold (0.05); hence the null hypothesis cannot be rejected in this case. The variance was analyzed using the *F-test*. Here, a normal distribution of samples was assumed. Again, the *p* values confirm the null hypothesis in all cases ($p > 0.13$). Finally, the *Kolmogorov–Smirnov test* also confirmed the null hypothesis. From statistical point of view, again, the *p* values confirm the null hypothesis in all cases.

As shown in Fig. 8, energy consumption estimated by the RBF-ANN accounted for 99% of the actual variability in energy use in validation and training data of irrigated and rainfed wheat production, respectively. Correlation between the observed and predicted energy consumption was very high for both training and testing phase. Therefore, this network has been able to explain the relationship between inputs and outputs witt high accuracy and have a good capability.

Fig. 9 shows the actual and predicted values (estimated by the best selected RBF-ANN model) for irrigated and rainfed wheat production. As it can be seen, the line of actual and predicted data was coinciding. Many researchers had good results from ANN models for prediction of output energy, yield and other complex function of agricultural productions. Khoshnevisan et al. (2013) developed ANN to prediction of output energy and greenhouse

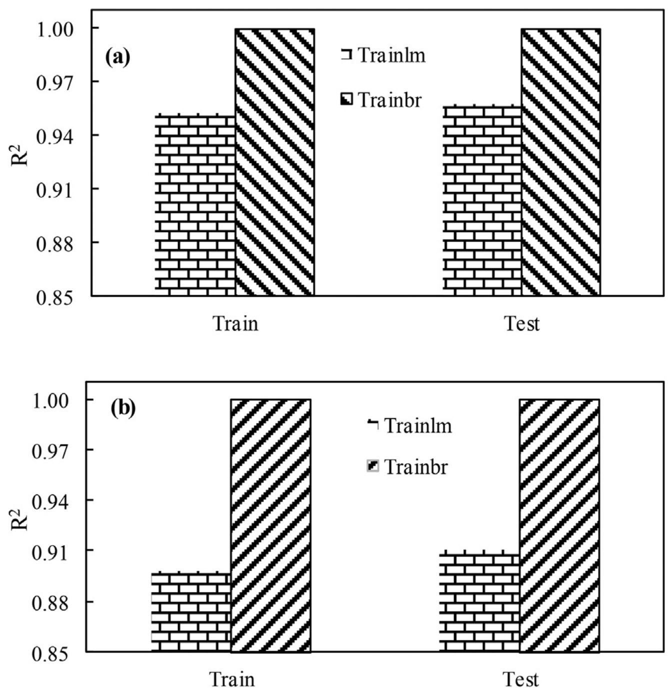

Fig. 7. Comparison between LM and BR algorithm for training and testing the network (**a:** Irrigated wheat, **b:** Rainfed wheat).

gas (GHG) emissions for wheat production in Fereydunshahr, Isfahan. The results of modeling showed that the ANN model with 11-3-2 LM structure was the best one for predicting the wheat yield and GHG emissions. The coefficients of determination (R^2) of the best topology were 0.99 and 0.998 for wheat yield and GHG emissions, respectively. Pahlavan et al. (2012) reported that a model consisted of an input layer with seven neurons, two hidden layers with 20 neurons in each one and one neuron in the output layer was the best one for predicting basil production in Isfahan province of Iran. Safa and Samarasinghe (2011), developed an ANN model based on a modular neural network with two hidden layers that can predict energy consumption based on farm conditions (size of crop area), social factors (farmers' education level), and energy inputs (N and P use, and irrigation frequency). Their results showed the ability of ANN model to predict energy consumption in wheat production using heterogeneous data. In another study, an ANN model with a 13-4-2 LM structure was developed to model energy consumption and GHG emissions of orange production in Guilan Province (Nabavi-Pelesaraei et al., 2016). According to the results the ANN can model yield and GHG emissions based on energy inputs for the horticultural crops of Guilan Province with high accuracy and low error.

For RBF model, the sensitivity analysis was applied to select the effective input parameters. Table 7 shows the results of this analysis. As it can see, all the inputs are effective and cannot be removed. These inputs are the main parameters and can change the output energy. For ANN models, it is common because the output results are much related to the number of data. If we want to select the ineffective parameters based on Table 7, for rainfed wheat production, the arrangement of input parameters is: fuel, machinery, SFP (Seed, Fertilizer and Pesticide) and human labor. For rainfed wheat production, the above results are very common because the main fuel consumption is in irrigated production and farmers use it for water pumping at this region. The results of

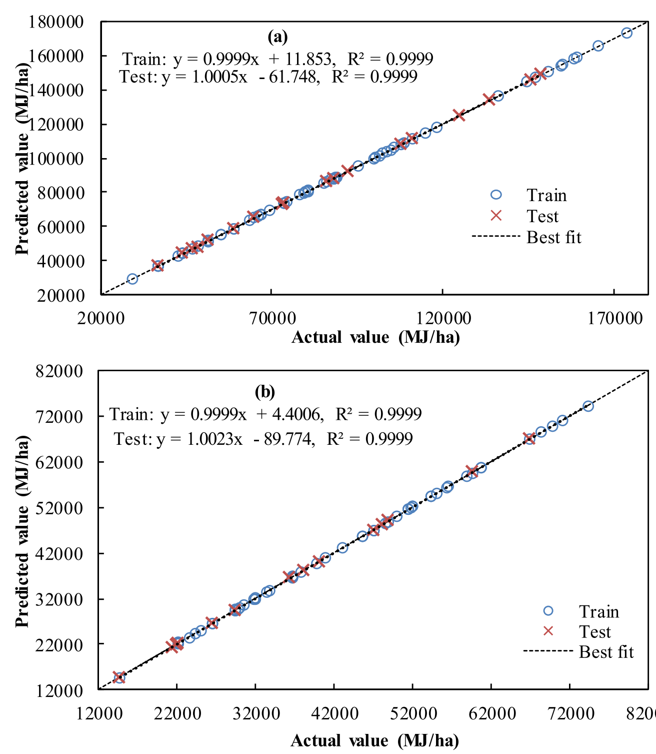

Fig. 8. Cross-Correlation of predicted and actual values of wheat output energy with Br algorithm for testing and training step (a: Irrigated wheat, b: Rainfed wheat).

Table 6
Performance of the best MLP-RBF and GPR topology with for prediction of irrigated and rainfed wheat production output energy.

		MLP[a]			RBF			GPR		
		Train	Test	Total	Train	Test	Total	Train	Test	Total
RMSE (MJ)	Irrigated	9458	14525	10666	49.91	99.79	**63.12**	8981	13705	10104
	Rainfed	5573	7577	6015	67.28	90.14	**72.30**	6890	8153	7153
MAPE (%)	Irrigated	11	12	11	0.04	0.07	**0.05**	10.53	12.12	10.85
	Rainfed	12	23	14	0.12	0.19	**0.14**	15.19	24.18	16.94
EF	Irrigated	0.93	0.83	0.92	0.999	0.999	**0.999**	0.94	0.85	0.92
	Rainfed	0.86	0.74	0.84	0.999	0.999	**0.999**	0.78	0.70	0.77
p-value (Mean)	Irrigated	0.97	0.50	0.81	0.999	0.999	**0.999**	1.00	0.42	0.74
	Rainfed	0.99	0.78	0.92	0.999	0.999	**0.999**	1.00	0.65	0.85
p-value (std)	Irrigated	0.71	0.32	0.51	0.999	0.999	**0.999**	0.8	0.39	0.62
	Rainfed	0.42	0.11	0.17	0.999	0.999	**0.999**	0.36	0.11	0.13
p-value (Distribution)	Irrigated	0.96	0.94	0.94	0.999	0.999	**0.999**	0.96	0.94	0.99
	Rainfed	0.77	0.54	0.74	0.999	0.999	**0.999**	0.45	0.54	0.46

[a] Best topology for MLP model was 5-17-1 and 4-21-1 for irrigated and rainfed wheat production with Br training algorithm.

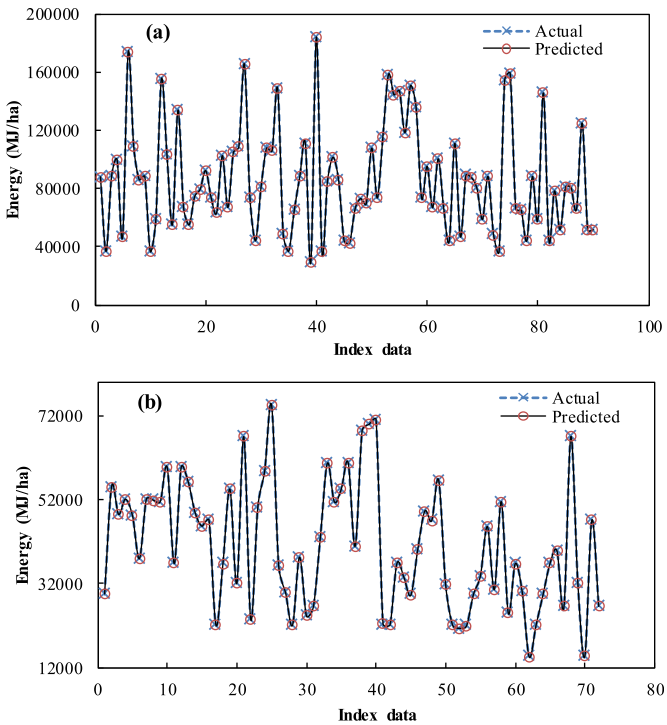

Fig. 9. Comparison between actual and predicted value with Br algorithm (**a:** Irrigated wheat, **b:** Rainfed wheat).

sensitivity analysis for irrigated wheat production show that human labor can affect the output results in this type of farming.

Prediction of crop yield especially strategic plants such as wheat has always been an interesting research area to agro meteorologists, as it is important in national and international economic programming. Dry farming crop production, apart from relationship to the genetic of cultivator, adaphic terms, effect of pests and pathology and weeds, the management and control quality during the growing season and etc. is severely depend to climatic events and can be different in a region. The results of this study showed that application of ANN tool can make models easier and more accuracy from complex natural systems with many inputs. The design of RBF network showed that this model can be used to estimate crop production in long or short term and also with enough and useful data for each area. Furthermore using ANNs can find the most effective factors on crop yield. Therefore some factors that their measurements are difficult and are not cost effective can be ignored. For future research, it is recommended to using dissimilar variables, such as farm conditions and social factors that can improve the ability of decision makers to look at the problem from different perspectives. Also, increasing the number of samples and used some additional variables for a longer period of time (at least 10 years), can help analyze trends in energy consumption in agricultural production and in different lands with different conditions. Also, using Life Cycle Assessment (LCA) can evaluate total environmental pollutions for wheat production in this region and helps farmers to recognize the importance of energy management. Total results of this study can be useful for farmers and governments to estimate the yield of crops before harvesting. So, pricing, marketing and various export policies can be adopted.

4. Conclusion

Summary of conclusions and recommendations can be stated as follows:

1. The average of total input and output energy for irrigated and rainfed wheat production were calculated as 23406-90528 MJha-1 and 9351–44131 MJha-1, respectively. Diesel fuel was the main factor affected on energy consumption for irrigated wheat production and also for medium and large lands of rainfed wheat production, but for small rainfed lands, total of fertilizers and poisons had the highest impact on total energy consumption. Also, the results showed that using of technology was not effective in wheat farms and will be worse by increasing in farm sizing. The results of this study revealed that with respect to the current technological level and agricultural knowledge of the farmers, cultivation in the large farms (>4 ha) can be a good solution to reduce the amount of total energy consumption, increasing the output production and also final income.
2. Results of output energy modeling showed that ANN-RBF model is more accurate than MLP-ANN and GPR models. RMSE and MAPE for irrigated and rainfed output modeling by ANN- RBF were 63.12–72.30 MJ and 0.05–0.14%. The results showed that the best spread factor for irrigated wheat with LM algorithm and at training phase (irrigated-LM-training) and irrigated wheat with BR algorithm at training phase (irrigated-BR-training), was equal to 7 and 4, respectively.
3. It can be inferred from the results that wheat irrigated farming systems in the studied area are significant production methods which are highly efficient and recommendable strategies on the view of energy-related factors. Future researches should apply energy analysis in various low and high input systems along with long term economical, environmental and societal analysis which further explain the suitability and compatibility of production system for establishing sustainable development.

Table 7
Results of sensitivity analysis for removing the ineffective variables to predict irrigated and rainfed wheat output energy by RBF-ANN topology with Br training algorithm.

		RMSE	MAPE	EF
Rainfed	All input	72.30	0.14	0.99
	All exclude Machinery	3002	5.22	0.96
	All exclude Human labor	7850	18.52	0.72
	All exclude Seed, Fertilizer and Pesticide	7023	13.97	0.78
	All exclude Fuel	1287	1.93	0.99
Irrigated	All input	63.12	0.05	0.99
	All exclude Machinery	4084	3.02	0.99
	All exclude Water	2526	2.45	0.99
	All exclude Human labor	4921	4.67	0.98
	All exclude Seed, Fertilizer and Pesticide	3020	3.10	0.99
	All exclude Fuel	3163	2.63	0.99

Acknowledgments

The authors would like to thank the editor in chief and the anonymous referees for their valuable suggestions and useful comments that improved the paper content substantially. This study was supported by Ramin Agriculture and Natural Resources University of Khuzestan, Iran. The authors are grateful for the support provided by this University.

References

Abdi, R., Taki, M., Akbarpour, M., 2012. An analysis of energy input-output and emissions of greenhouse gases from agricultural productions. Int. J. Nat. Eng. Sci. 6 (3), 73–79.
Abdi, R., Taki, M., Jalali, A., 2013. Study on energy use pattern, optimization of energy consumption and CO_2 emission for greenhouse tomato production. Int. J. Nat. Eng. Sci. 7 (1), 01–04.
Abdollahpour, S., Zaree, S., 2009. Evaluation of wheat energy balance under rain fed farming in Kermanshah. J. Sustain. Agric. Knowl. 20 (2), 97–106 [in parsian].
Al-Ghandoor, A., Jaber, J.O., Al-Hinti, I., Mansour, I.M., 2009. Residential past and future energy consumption: potential savings and environmental impact. Renew. Sustain. Energy Rev. 13 (6–7), 1262–1274.
Bahrami, H., Taki, M., Monjezi, N., 2011. Optimization of energy consumption for wheat production in Iran using data envelopment analysis (DEA) technique. Afr. J. Agric. Res. 6 (27), 5978–5986.
Beheshti Tabar, I., Keyhani, A., Rafiee, S., 2010. Energy balance in Iran's agronomy (1990–2006). Renew. Sustain Energy Rev. 14, 849–855.
Burden, F., Winkler, D., 2008. Bayesian regularization of neural networks. Methods Mol. Biol. 458, 25–44.
Esmaeili, A., Mozayani, N., 2009. Adjusting the parameters of radial basis function networks using particle swarm optimization. In: International Conference on Computational Intelligence for Measurement Systems and Applications Hong Kong, China May 11–13.
Ghorbani, Reza, Mondani, F., Amirmoradi, S., Feizi, H., Khorramdel, S., Teimouri, M., 2011. A case study of energy use and economical analysis of irrigated and dryland wheat production systems. Appl. Energy 88 (1), 283–288.
Gianola, D., Manfredi, E., Simianer, H., 2012. On measures of association among genetic variables. Anim. Genet. 43, 19–35.
Haykin, S., 2009. Neural Networks and Learning Machines, third ed. Prentice Hall.
Iliyas, S.A., Elshafei, M., Habib, M.A., Adeniran, A.A., 2013. RBF neural network inferential sensor for process emission monitoring. Control Eng. Pract. 21, 962–970.
Kayri, M., 2016. Predictive abilities of bayesian regularization and levenberg–marquardt algorithms in artificial neural networks: a comparative empirical study on social data. Math. Comput. Appl. 21, 20. https://doi.org/10.3390/mca21020020.
Kelemen, A., Liang, Y., 2008. Statistical advances and challenges for analyzing correlated high dimensional SNP data in genomic study for complex. Dis. Stat. Surv. 2, 43–60.
Khoshnevisan, B., Rafiee, S., Omid, M., Yousefi, M., Movahedi, M., 2013. Modeling of energy consumption and GHG (greenhouse gas) emissions in wheat production in Esfahan province of Iran using artificial neural networks. Energy 52, 333–338.
Khoshnevisan, B., Rafiee, S., Omid, M., Mousazadeh, H., 2014. Development of an intelligent system based on ANFIS for predicting wheat grain yield on the basis of energy inputs. Inf. Process. Agric. 1, 14–22.
Kizilaslan, H., 2009. Input-output energy analysis of cherries production in Tokat province of Turkey. Appl. Energy 86, 1354–1380.
Liu, D., Pang, J., Zhou, J., Peng, Y., Pecht, M., 2013. Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression. Microelectron. Reliab. 53, 832–839.
Mackay, J.C.D., 1996. Comparison of approximate methods for handling hyperparameters. Neural Comput. 8, 1–35.
Nabavi-Pelesaraei, A., Rafiee, S., Hosseinzadeh-Bandbafha, H., Shamshirband, S., 2016. Modeling energy consumption and greenhouse gas emissions for kiwifruit production using artificial neural networks. J. Clean. Prod. 133, 924–931.
Nabavi-Pelesaraei, A., Rafiee, S., Mohtasebi, S.S., Bandbbafha, H., Chau, K.K., 2017. Energy consumption enhancement and environmental life cycle assessment in paddy production using optimization techniques. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2017.06.071.
Najafi, G., Ghobadian, B., Tavakoli, T., Buttsworth, D.R., Yusaf, T.F., Faizollahnejad, M., 2009. Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network. Appl. Energy 86, 630–639.
Okut, H., Gianola, D., Rosa, G.J.M., Weigel, K.A., 2011. Prediction of body mass index in mice using dense molecular markers and a regularized neural network. Genet. Res. Camb 93, 189–201.
Okut, H., Wu, X.L., Rosa, G.J.M., Bauck, S., Woodward, B.W., Schnabel, R.D., Taylor, J.F., Gianola, D., 2013. Predicting expected progeny difference for marbling score in Angus cattle using artificial neural networks and Bayesian regression models. Genet. Sel. Evol. 45, 1–8.
Pachepsky, Y.A., Timlin, D., Varallyay, G., 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. 60, 727–733.
Pahlavan, R., Omid, M., Akram, A., 2012. Energy input output analysis and application of artificial neural networks for predicting greenhouse basil production. Energy 37 (1), 171–176.
Ranjbar, I., Ajabshirchi, Y., Taki, M., Ghobadifar, A., 2013. Energy consumption and modeling of output energy with MLP Neural Network for dry wheat production in Iran. Elixir Agric. 62, 17949–17953.
Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian Processes in Machine Learning. The MIT Press, Cambridge, MA.
Rohani, A., Abbaspour-Fard, M.H., Abdolahpour, S., 2011. Prediction of tractor repair and maintenance costs using Artificial Neural Network. Expert Syst. Appl. 38, 8999–9007.
Safa, M., Samarasinghe, S., 2011. Determination and modelling of energy consumption in wheat production using neural networks: "A case study in Canterbury province, New Zealand. Energy 36 (8), 5140–5147.
Safa, M., Samarasinghe, S., Mohsen, M., 2009. A field study of energy consumption in wheat production in Canterbury, New Zealand. Energy Convers. Manag. 52, 2526–2532.
Saini, L.M., 2008. Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks. Electr. Power Syst. Res. 78, 1302–1310.
Shaneh, A., Butler, G., 2006. Bayesian learning for feed-forward neural network with application to proteomic data: the glycosylation sites detection of the epidermal growth factor-like proteins associated with Cancer as A Case study. In: Lamontagne, L., Marchand, M. (Eds.), Advances in Artificial Intelligence; Canadian AI LNAI 4013. Springer-Verleg, Berlin/Heidelberg, Germany.
Singh, S., Singh, S., Pannu, C.J., Singh, J., 1999. Energy input and yield relations for wheat in different agro-climatic zones of the Punjab. Appl. Energy 63, 287–298.
Sorich, M.J., Miners, J.O., Ross, A.M., Winker, D.A., Burden, F.R., Smith, P.A., 2003. Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-Glucuronosyl transferesa isoforms. J. Chem. Inf. Comput. Sci. 43, 2019–2024.
Souza, D.C., 2015. Neural Network Learning by the Levenberg–marquardt Algorithm with Bayesian Regularization. Available online: http://crsouza.blogspot.com/feeds/posts/default/webcite. (Accessed 29 July 2015).
Tabatabaeefar, A., Emamzadeh, H., GhasemiVarnamkhasti, M., Rahimizadeh, R., Karimi, M., 2009. Comparison of energy of tillage systems in wheat production. Energy 34, 41–45.
Taghavifar, H., Mardani, A., 2015. Energy consumption analysis of wheat production in West Azerbaijan utilizing life cycle assessment (LCA). Renew. Energy 74, 208–213.
Taki, M., Ajabshirchi, Y., Mahmoudi, A., 2012a. Prediction of output energy for wheat production using artificial neural networks in Esfahan province of Iran. J. Agric. Technol. 8 (4), 1229–1242.
Taki, M., Ajabshirchi, Y., Mahmoudi, A., 2012b. Application of parametric and nonparametric method to analyzing of energy consumption for cucumber production in Iran. Mod. Appl. Sci. 6 (1), 75–87.
Taki, M., Mahmoudi, A., Ghasemi-mobtaker, H., Rahbari, H., 2012c. Energy consumption and modeling of output energy with multilayer feed-forward neural network for corn silage in Iran. Agric. Eng. Int. CIGR J. 14 (4), 93–101.
Taki, M., Abdi, R., Akbarpour, M., Mobtaker, H.G., 2013. Energy inputs – yield relationship and sensitivity analysis for tomato greenhouse production in Iran. Agric. Eng. Int. CIGR J. 15 (1), 59–67.
Taki, M., Ajabshirchi, Y., Ranjbar, S.F., Rohani, A., Matloobi, M., 2016a. Heat transfer and MLP neural network models to predict inside environment variables and energy lost in a semi-solar greenhouse. Energy Build. 110, 314–329.
Taki, M., Ajabshirchi, Y., Ranjbar, S.F., Rohani, A., Matloobi, M., 2016b. Modeling and experimental validation of heat transfer and energy consumption 265 in an innovative. Inf. Process. Agric. 3 (3), 157–174.
Ticknor, J.L., 2013. A Bayesian regularized artificial neural network for stock market forecasting. Expert Syst. Appl. 14, 5501–5506.
Wan, Z.Y., Sapsis, P.S., 2017. Reduced-space Gaussian Process Regression for data-driven probabilistic forecast of chaotic dynamical systems. Phys. D. 345, 40–55.
Wayg, Y.H., Li, Y., Yang, S.L., Yang, L., 2005. An in silico approach for screening flavonoids as P-glycoprotein inhibitors based on a Bayesian regularized neural network. J. Comput. Aided Mol. Des. 19, 137–147.
Xu, M., Zengi, G., Xu, X., Huang, G., Jiang, R., Sun, W., 2006. Application of Bayesian regularized BP neural network model for trend analysis. Acidity and chemical composition of precipitation in North. Water Air Soil Pollut. 172, 167–184.