TECHNOLOGICAL AND ECONOMIC DEVELOPMENT OF ECONOMY
ISSN 2029-4913 print/ISSN 2029-4921 online
2013 Volume 19(Supplement 1): S157–S175
doi:10.3846/20294913.2013.876690

FORECASTING TRANSPORTATION INFRASTRUCTURE IMPACTS OF RENEWABLE ENERGY INDUSTRY USING NEURAL NETWORKS

Kasthurirangan GOPALAKRISHNAN[a], Konstantina "Nadia" GKRITZA[b]

[a]*Department of Civil, Construction & Environmental Engineering, 354 Town Engineering Building, Iowa State University, Ames*
[b]*Department of Civil, Construction & Environmental Engineering, 404 Town Engineering Building, Iowa State University, Ames*

Received 10 April 2012; accepted 06 April 2013

Abstract. Iowa is a state rich in renewable energy resources, especially biomass. The successful development of renewable energy industry in Iowa is concomitant with increase in freight traffic and is likely to have significant impacts on transportation infrastructure condition and increased maintenance expenses for the state and local governments. The primary goal of this paper is to investigate the feasibility of employing the Neural Networks (NN) methodology to forecast the impacts of Iowa's biofuels and wind power industries on Iowa's secondary and local road condition and maintenance-related costs in a panel data framework. The data for this study were obtained from a number of sources and for a total of 24 counties in clusters in Northern, Western, and Southern Iowa over a period of ten years. Back-Propagation NN (BPNN) using a Quasi-Newton second-order training algorithm was chosen for this study owing to its very fast convergence properties. Since the size of the training set is relatively small, ensembles of well-trained NNs were formed to achieve significant improvements in generalization performance. The developed NN forecasting models could identify the presence of biofuel plants and wind farms as well as large-truck traffic as the most sensitive inputs influencing pavement condition and granular and blading maintenance costs. Pavement deterioration resulting from traffic loads was found to be associated with the presence of both biofuel plants and wind farms. The developed NN forecasting models can be useful in identifying and properly evaluating future transportation infrastructure impacts resulting from the renewable energy industry development and thus help Iowa maintain its competitive edge in the rapidly developing bioeconomy.

Keywords: neural networks, renewable energy, biofuel plants, wind farms, forecasting, panel data, infrastructure impacts.

Reference to this paper should be made as follows: Gopalakrishnan, K.; Gkritza, K. "N." 2013. Forecasting transportation infrastructure impacts of renewable energy industry using neural networks, *Technological and Economic Development of Economy* 19(Supplement 1): S157–S175.

JEL Classification: C45, C80, O18, R42.

Corresponding author Kasthurirangan Gopalakrishnan
E-mail: *rangan@iastate.edu*

Introduction

Iowa has not only become a leading producer of sustainable biofuels such as grain-based ethanol and soy-based biodiesel, but is also becoming a leading center of wind energy generation (Kedron, Bagchi-Sen 2011). While trying to meet the U.S. Energy Independence and Security Act of 2007 (EISA) goal of 36 billion gallons per year of renewable transportation fuels by 2022 (U.S. DOE 2012) Iowa's renewable energy industry is now expanding into developing second generation of biofuels based on cellulosic feedstock such as wheat and rice straw, switch grass, etc. (Wakeley *et al.* 2008). All these could have significant implications on Iowa's transportation infrastructure system as a result of heavier, more frequent and large-sized vehicular traffic (Fu *et al.* 2008).

According to recent statistics reported by the Iowa Department of Economic Development (IEDA 2012), Iowa ranks 1st in the U.S. in ethanol production, 4th in biodiesel production, and 2nd in wind generation output. In addition, 25% of national ethanol production comes from Iowa, which is also the leading producer of raw biomass in the U.S. Maps of ethanol plants, biodiesel plants (as of November 16, 2010), and wind farms in Iowa are displayed in Fig. 1.

Since Iowa's thriving renewable energy industry involves production/collection and distribution of raw materials as well as finished products, its success heavily depends on the quality of service that the transportation infrastructure can provide. Like the traditional agricultural products grown in Iowa today, new energy crops will likely originate along the 90,000 miles of rural county roads that make up the state's secondary system. The increased traffic, especially resulting from highly transportation-intensive biofuel distribution, will not only likely impact the physical condition of the transportation infrastructure, but will also result in increased maintenance expenses for state and local governments (Haddad *et al.* 2009).

Given that most biofuel plants or wind farms are located very close to a county road or the secondary road system which falls under the jurisdiction of the county, this adds an additional burden of maintenance to the counties which are already overloaded with more roads than what they can maintain (Bai *et al.* 2011). The secondary system accounts for 79% of all state, county and municipal roads distributed throughout Iowa; 74% are gravel, 21% are paved, and 5% are unmaintained dirt roads (Gkritza *et al.* 2011). In addition, Iowa roads and secondary systems were not originally designed to accommodate so long and heavy vehicles that are required to carry wind turbines and other parts from manufacturing plants to the wind turbine construction site. Recent projections estimate that $23.4 billion will be needed to maintain and upgrade the secondary system over the next 20 years, while only $10.9 billion in projected revenues are forecast to address these needs during the same period (Gkritza *et al.* 2011).

In order to document the current physical and fiscal impacts of Iowa's existing biofuels and wind power industries on transportation infrastructure, Gkritza *et al.* (2011) conducted a statewide survey to identify counties with existing biofuel production plants, wind farms, as well as, to obtain infrastructure and financial data related to these renewable energy facilities. Based on the survey results, a total of 24 counties in clusters in Northern, Western, and Southern Iowa that represent a variety of soil, terrain, and environmental conditions were selected. Information related to maintenance-related costs, indicator of road pavement condition, renewable energy plant capacity and years of operation, as well as environmental and agricultural factors were collected for the 24 counties during the period 1999–2008 and assembled in a panel data framework.

The primary goal of this paper is to investigate the feasibility of employing Neural Networks (NNs) in developing forecasting models that would predict transportation infrastructure impacts from Iowa's renewable energy industry based on existing traffic volumes and other related inputs.

(a)

(b)

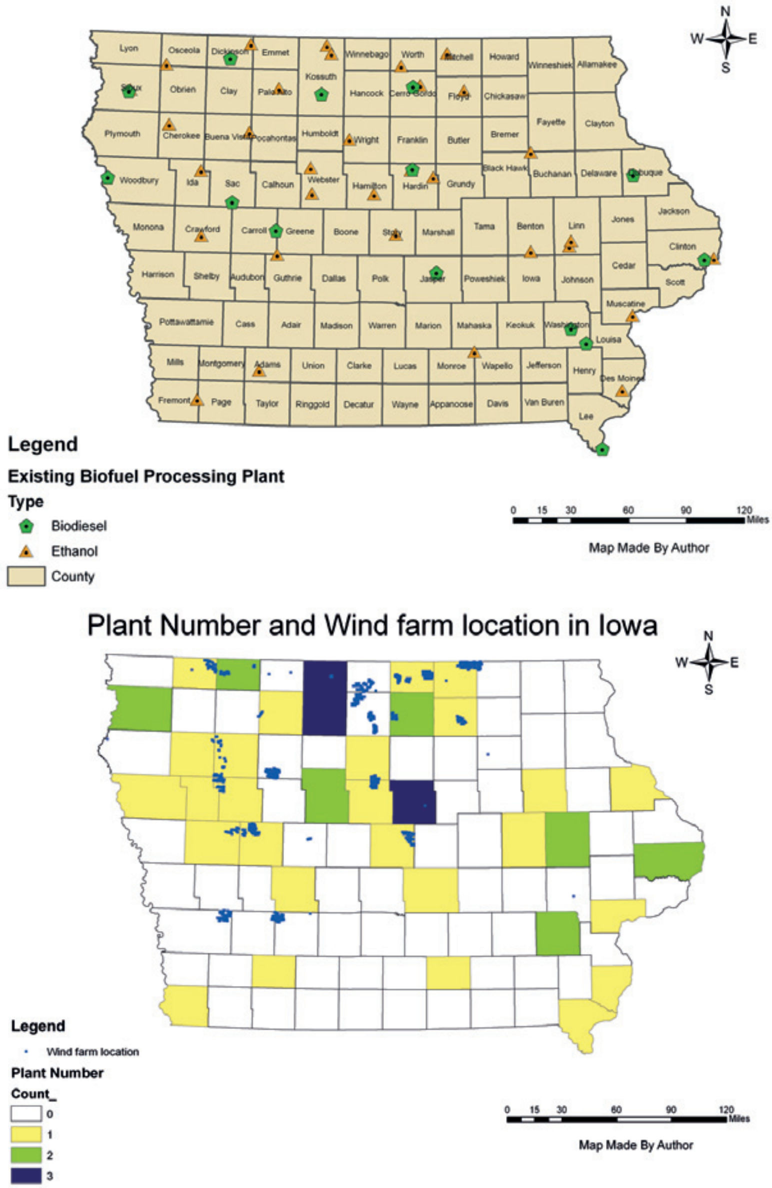

Fig. 1. (a) Map of Iowa ethanol; (b) Map of Iowa wind farms
(Source: Iowa Department of Economic Development; Iowa DOT)

1. Objective and scope

In this paper, NN modeling is proposed as an alternative tool for forecasting transportation infrastructure impacts of renewable energy industry in Iowa using panel data. Since panel data are generally characterized by high number of cross-sections (regions) and a limited number of years for which the data is available, they do not easily lend themselves to the application of standard econometric techniques (Autant-Bernard 2012; Platoni *et al.* 2012). Econometric techniques adapted for panel data analysis are nowadays quite easily implemented in standard statistical packages. However, such techniques may not be able to produce reliable short-term forecasts owing to the constraints imposed for estimating the parameters (Longhi *et al.* 2005). Neural networks, owing to their ability to model arbitrary and complex functions of the data quite well, without a priori knowledge of the underlying phenomenon, have become a popular forecasting tool (Monteiro *et al.* 2013; Huang, Tian 2013; Ye *et al.* 2012). The choice of using NN as a forecasting tool in this study stems from its success in improving macroeconomic forecasts and achieving superior business forecasts as documented by several studies (Adya, Collopy 1998; Swanson, White 1997; Chen 2011; Qi, Chang 2011).

2. Neural networks review

In recent years NNs have been applied to complex engineering problems in various civil engineering areas such as pavement and geotechnical engineering, structural engineering, transportation, water resources and environmental engineering (Adeli 2001; Sun *et al.* 2012; Gupta, Cohn 2012).

NNs are computational intelligence systems that simulate the behavior of the human brain and nervous system. The basic element in the NN is a processing element, called an artificial neuron or node. Each neuron contains a very limited amount of local memory and performs basic mathematical operations on data passing through them. These neurons are highly interconnected in layers such as an input layer, an output layer and one or more hidden layers. The computational power of NN comes from this interconnection which makes input data concurrently processed in artificial neurons (TRB Circular 1999).

An artificial neuron receives information (signal) from other neurons, processes it, and then relays the filtered signal to other neurons (Tsoukalas, Uhrig 1997). The receiving end of the neuron has incoming signals (x_1, x_2, x_3….and x_n). Each of them is assigned a weight (w_{ji}) that is based on experience and likely to change during the training process. The summation of all the weighted signal amounts yields the combined input quantity (I_j) which is sent to a preselected transfer function (*f*), sometimes called an activation function. A filtered output (y_j) is generated in the outgoing end of the artificial neuron (*j*) through the mapping of the transfer function. The parameters can be expressed in the form of following equations:

$$I_j = \sum_{i=1}^{n} w_{ji} x_i; \tag{1}$$

$$y_j = f(I_j)\,. \tag{2}$$

There are several types of transfer functions that can be used, including sigmoid, tangent hyperbolic, threshold, and Gaussian functions. The sigmoid function is the most commonly used transfer function because of its differentiability. The sigmoid function can be represented by the following equation:

$$f(I_j) = \frac{1}{1+\exp(-\varphi I_j)}, \tag{3}$$

where j = positive scaling constant, which controls the steepness between the two asymptotic values 0 and 1 (Tsoukalas, Uhrig 1997).

The hyperbolic tangent function (*tanh*) is also a commonly used (sigmoid) nonlinear activation function for which the amplitude of the output lies in the range −1 £ $f(I_j)$ £ 1 and is expressed as follows:

$$f(I_j) = \frac{\exp(\varphi I_j) - \exp(-\varphi I_j)}{\exp(\varphi I_j) + \exp(-\varphi I_j)}. \tag{4}$$

NN performs two major functions: learning (training) and testing. A training data set and an independent testing data set are prepared for these functions. Inputs from a training data set are presented to the input layer to start the propagation of data. Inside the network, weights are adjusted when data pass between artificial neurons along the connections. Since interconnected neurons have the flexibility to adjust the weights, NN has the ability to analyze complex problems. It uses a learning rule to find a set of weights such that the error is minimum. This process is called "learning" or "training" (Shahin *et al.* 2001). The learning mechanisms used by NN are of three primary types (TRB Circular 1999):

- Supervised learning: system/weight is adjusted by comparing the network output with a given or desired output;
- Unsupervised training: the network is trained to form categories based on similarity among the data and identify irregularities in data;
- Reinforcement learning: the network attempts to learn the input-output vectors by trial and error through maximizing a performance function. The system can identify whether a given output is correct or not but cannot estimate the exact output.

In order to track the performance of the network, the Mean Squared Error (MSE) (or any other performance measure) is calculated at the end of each epoch. An epoch is defined as one full presentation of all training vectors to the network. It would be expected that the MSE would decrease almost monotonically with respect to the increase in the number of epochs until it reaches an optimum number. It would ultimately level off (converge) at which point the network has 'fully learnt'.

Once the training phase of the model has been successfully accomplished, the network performance is verified by presenting independent testing datasets to the NN. This process is called "testing." Details regarding the theory and mathematics behind the NN is widely available (Aleksander, Morton 1990; Fausett 1994; Haykin 1998; Bishop 1995; Swingler 1996).

There are many different types of NN used in many areas of engineering. These differ in the arrangement and degree of connectivity of their neurons, the types of calculations performed within each neuron, the degree of supervision they receive during training, the determinism of the learning process, and the overall learning theory under which they operate (Mehra, Wah 1992). However, certain types of NN are more repeatedly used, either because they are broadly applicable to a wide variety of problems or ideally suited for a narrow range of problems (TRB Circular 1999). These include hopfield nets, adaptive resonance theory (ART) networks, self-organized feature maps (SOFM), backpropagation neural networks (BPNN), feedback (sequential) neural networks (FBNN), counter propagation networks, radial basis function network (RBF), and generalized regression neural networks (GRNN).

Multi-Layer Perceptron (MLP) BPNN is one of the most preferred NN in civil engineering related applications of NN because of its powerfulness, versatility, and simplicity (TRB Circular 1999; Adeli 2001). BPNN can be taught a mapping from one data space to another using a representative set of patterns/examples to be learned. BPNN refers to a multi-layered, feed-forward neural network trained using an error backpropagation algorithm (Fig. 2). The algorithm, pseudo-code, and theory behind BPNN are well-documented (Adeli, Hung 1995; Haykin 1998).

Cheng and Titterington (1994) provide an interesting review of NN from a statistical perspective. The reported applications of NNs for analyzing panel data are rather limited (Lin 1992). Longhi *et al.* (2005) and Patuelli *et al.* (2006) developed a NN tool for forecasting regional employment patterns based on employment data collected for 327 West German regions over a period of fourteen years. Giovanis (2008) developed a NN model to examine all factors that impact greenhouse effect in fifteen countries of the European Union from 1990 to 2005.

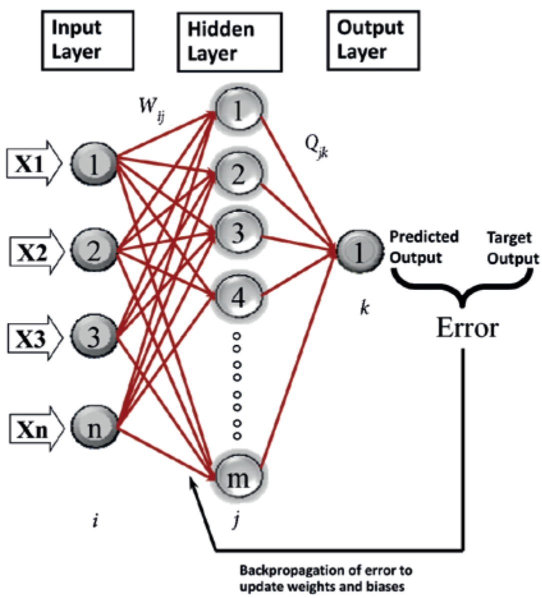

Fig. 2. Schematic of a typical feed-forward Multi-Layered Perceptron (MLP) BackPropagation Neural Network (BPNN) (Note: this architecture with a single hidden layer is denoted as A-B-C, where A represents the number of inputs (n); B represents the number of hidden neurons (m); and C represents the number of outputs (1))

3. Description of data

The data for this study come from a number of sources including a local agency survey conducted to study the physical and fiscal impacts to the transportation system due to renewable energy facility construction and operations in Iowa (Gkritza *et al.* 2011). The dataset is structured as a panel of 24 cross-sections (24 counties in the northern, western, and southern parts of Iowa) and ten time periods from 1999 to 2008. The data include maintenance-related costs, Pavement Condition Index (PCI), renewable energy plant capacity, years of plant operation, traffic volumes, and environmental and agricultural factors.

The data source for maintenance-related costs is the annual county expense reports prepared by county engineers and filed annually with the Iowa DOT. The Iowa Environmental Mesonet (ISU 2012) provided the environment-related data for the study. The annual corn production and soybean production data per county was obtained from the U.S. Department of Agriculture, National Agriculture Statistic Service (USDA 2012). The Iowa DOT Geographic Information Management System (GIMS) provided the traffic data, while the pavement condition data were extracted from the Iowa DOT Pavement Management Information System (PMIS).

The original intention was to use all maintenance-related cost variables, which include asphalt and concrete pavement maintenance costs, granular and blading costs, winter maintenance costs, as well as bridge maintenance costs. Preliminary NN analysis revealed that it would not be possible to model the impact of Iowa's renewable energy industry operations on asphalt and concrete pavement maintenance costs, winter maintenance costs and bridge maintenance costs using the limited data available for this study. Therefore, the maintenance-related cost category includes only granular and blading costs related to gravel road maintenance.

A brief summary of the data utilized in this study is provided in Table 1. The input and output variables selected for developing the forecasting models were based on comprehensive literature review, county surveys, as well as existing information (Gkritza *et al.* 2011). The goal is to forecast the gravel road maintenance-related costs and PCI in each region r (with r ranging from 1 to 24) in year t, given the traffic volumes, renewable energy plant capacity, years of plant operation, and environmental and agricultural factors in the previous years.

There are certain limitations to the collected data which need to be acknowledged upfront. The PCI data was collected for both primary and secondary road systems. Northern counties in Iowa, however, had not collected PCI data for their secondary road system since 2004 (Gkritza *et al.* 2011). As a result, PCI for the primary road system alone was utilized in this study. Further, owing to these data limitations, maintenance costs reported in the annual county expense reports were used as an indirect indicator of pavement condition and deterioration over time. It should be noted that the maintenance-related costs are based on available funds rather than actual needs. As such, the magnitude of impacts could be higher than that reflected in the maintenance costs.

Table 1. Summary description of NN inputs and outputs

Category	Variable	Description/Comments
Pavement Condition (Output)	Pavement Condition Index (PCI)	A numerical index between 0 (very poor) and 100 (excellent) widely used to indicate pavement condition and deterioration over time based on measurements of roughness and surface distress. PCI for the primary road system only was available for this study.
Maintenance-related Costs (Output)	Granular and Blading Cost (dollars)	Maintenance costs were used as an indirect indicator of pavement condition and deterioration over time. The amount spent on maintenance is based on available funds rather than actual needs and, as such, the magnitude of impacts could be higher than that reflected in the maintenance costs.
Traffic Volume Information (Input)	Primary Rural VMT	Annual Vehicle Miles Traveled (VMT) for all vehicles for primary rural road systems. Traffic volume was used as a measure of repeated traffic loading.
	Primary Urban VMT	Annual VMT for all vehicles for primary urban road systems. Traffic data were extracted from annual VMT reports provided by Iowa DOT.
	Primary Rural LVMT	Annual Large-truck VMT (LVMT) for primary rural road systems. Large truck is defined here as defined as single- or multiple-trailer trucks with four or more axles.
	Primary Urban LVMT	Annual LVMT for primary urban road systems.
	Secondary LVMT	Annual LVMT for secondary road systems.
	Secondary VMT	Annual VMT for secondary road systems.
	Local LVMT	Annual LVMT for local road systems.
	Local VMT	Annual VMT for local road systems.
Renewable Fuel Production and Energy Plant Information (Input)	Ethanol Plant Present	1 if present, 0 otherwise
	Biodiesel Plant Present	1 if present, 0 otherwise
	Wind Farm Present	1 if present, 0 otherwise
	Number of Ethanol Plants	Two categories of impacts associated with renewable fuel production (ethanol and biodiesel) plants are considered: heavy vehicle traffic transporting grain to the biofuel plant and the finished products to retail markets.
	Number of Biodiesel Plants	
	Years of Operation of Ethanol Plant	
	Years of Operation of Biodiesel Plant	
	Number of Wind Turbines	Impacts associated with the transportation of turbines, parts and materials with oversized vehicles from manufacturing plants to the wind turbine construction site are considered. The operation of wind farm is not expected to generate ongoing heavy vehicle traffic like the biofuel plants.
	Years of Operation of Wind Farm	
	Capacity of Ethanol Plant (million gallons)	Currently, Iowa has the capacity to produce more than 3.28 billion gallons of ethanol annually which accounts for more than 25% of the entire U.S. ethanol production (IEDA 2012).
	Capacity of Biodiesel Plant (million gallons)	

Continued Table 1

Category	Variable	Description/Comments
Agri-cultural Factors (Input)	Corn Production (bushel)	Annual corn and soybean production information by county was obtained from U.S. Department of Agriculture (USDA) and National Agriculture Statistic Service.
	Soybean Production (bushel)	
Environ-mental Factors (Input)	Snow Depth (inch)	Both temperature and precipitation influence the strength of the pavement layers. The effect of temperature was modeled using a freezing index measured in degree-days below freezing. Snow depth and rainfall depth were used as proxies to model the effects of precipitation levels on pavement deterioration. Environment-related data were obtained from Iowa Environmental Mesonet.
	Rainfall Depth (inch)	
	Freezing Index in northern counties (degree-days)	

4. Neural network forecasting models

As seen in Table 1, the panel data set consists of 24 inputs (traffic volumes, renewable fuel production and energy plant related information, agricultural factors, and environmental factors) and two outputs (PCI and granular and blading costs).

In panel data regression modeling, the effect of specific regional and time characteristics are taken into account by means of regional dummies and time dummies, respectively. In developing the NN forecasting models, a slightly modified approach was used in accounting for these effects as suggested by Longhi *et al.* (2005). The regional effects were modeled using a discrete variable ("County") computed as (1/R)*r, where r ranges from 1 to 24, and R = 24 (counties). Similarly, the effect of time-specific characteristics was modeled using a discrete variable ("Year") computed in the same way with r ranging from 1 to 9 and R = 9 (years). The data for the last year (2008) was completely set aside for testing the NN model's forecasting performance. This method of accounting for regional effects and time effects eliminates a number of additional explanatory variables required in the NN modeling equal to the number of counties and the number of years, which will adversely affect the forecasting performance. Note that both variables, County and Year, were included as two additional inputs in NN modeling, thus increasing the total number of inputs to 26.

The NN model development, parameter tuning, and performance evaluation were carried out in STATISTICA® and MATLAB® software environment. The available data from years 1999 to 2007 for all 24 counties were used in training the NNs and identifying the optimum network architectures. The best-performance NNs were then tested on year 2008 data to obtain ex-post forecasts which were used in evaluating the NN forecasting model's generalization properties. As mentioned previously, individual forecasting models were developed for PCI and gravel road maintenance costs as shown in Table 1. All input and output variables were scaled using linear transformations in the range of 0 to 1 to prevent network saturation effect.

The choice of learning algorithm, number of hidden layers and hidden neurons, and other network parameters need to be carefully selected through a sensitivity analysis to achieve best-performance settings. If the NN architecture consists of too many hidden layers and/or hidden neurons, it will most likely result in "memorization" of input-output patterns resulting in poor generalization performance. On the other hand, if the network is too simple, it may not learn the mapping at all, once again leading to poor forecasting performance. A single hidden layer was used in this study. The optimum number of hidden neurons was chosen by varying the number of hidden neurons in the range of 6 to 20 and comparing its effect on the performance for a large number of network architectures. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) or Quasi-Newton second-order training algorithm was chosen owing to its very fast convergence properties. The sigmoid transfer function (Eq. (3)) was used as the non-linear activation function in the hidden layer of the network.

Best-performance networks were selected for each transportation infrastructure impact indicator (PCI and granular and blading maintenance cost) based on a search of over 500 network configurations and trials. In general, ensembles of well-trained NNs are formed to achieve significant improvements in generalization performance, especially when the size of the training set is small. In this study, five best-performance member networks were used to form ensemble NN prediction models for each output. The five best-performance member networks for each of the forecasted variable are identified in Table 2 under second column. Here, the designation *MLP x-y-z* for each member network refers to Multi-Layered Perceptron (MLP) neural network architecture with x input nodes, 1 hidden layer with y hidden neurons and z output nodes. Ensembles can be conceived as collection of best-performance neural networks that cooperate in providing a prediction. The member networks' predictions are typically averaged (or voted in the case of a classification problem) to obtain the ensemble outputs after weighting the average using the networks' membership weights. This provides a relatively simple way to reduce model variance without increasing model bias. The results reported and discussed in the next section pertain to ensemble NN predictions.

Table 2. Forecasting performance of NN models

Output	Network	MAE	MAPE	MSE	RMSE
PCI	MLP 26-6-1	3.601	0.153	19.942	4.466
	MLP 26-10-1	3.810	0.130	18.511	4.302
	MLP 26-7-1	3.461	0.151	17.296	4.159
	MLP 26-7-1	3.003	0.143	18.366	4.286
	MLP 26-7-1	3.796	0.143	19.146	4.376
	Ensemble	3.112	0.142	13.158	3.627
Granular and Blading Costs ($)	MLP 26-13-1	237,454	0.228	7.4E+10	272,380
	MLP 26-11-1	240,171	0.250	7.5E+10	273,479
	MLP 26-7-1	213,391	0.236	6.8E+10	261,008
	MLP 26-11-1	243,110	0.233	7.6E+10	276,046
	MLP 26-10-1	226,401	0.295	7.1E+10	266,299
	Ensemble	227,410	0.237	6.6E+10	256,318

Note: MLP x-y-z refers to Multi-Layered Perceptron Neural Network with x input nodes; 1 hidden layer with y hidden neurons; and z output neuron.

5. Results and discussion

The goodness-of-fit statistics for the NN predictions were computed using ex-post forecasts for the year 2008 and statistical indicators such as the Mean Absolute Error (MAE), Mean Absolute Percentage Error (*MAPE*), *MSE* and Root-Mean Squared Error (*RMSE*). However, such indicators are calculated over all 24 counties for one year due to the panel structure of the data. Thus, the forecasting errors computed as the difference between actual and NN model forecasts for all counties are summed up to obtain the global forecasting error. As a result, the variability across regions, rather than time, is captured by the statistical indicators. The statistical indicators are defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i^t - y_i^p\right| \; ; \tag{5}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^t - y_i^p}{y_i^t}\right| \; ; \tag{6}$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i^t - y_i^p)^2}{n} \; ; \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i^t - y_i^p)^2}{n}} \; , \tag{8}$$

where: y_i^t and y_i^p are the target (actual) and forecasted values, respectively; $\overline{y_i^t}$ and $\overline{y_i^p}$ are the mean of the target and forecasted values corresponding to *n* patterns, respectively. Here *n* represents the total number of counties (24). MAE, MAPE, MSE, and RMSE are commonly used statistical indicators reported in the time-series literature and studies related to NN modeling of panel data (Swanson, White 1997; Longhi *et al.* 2005). In general, smaller values of these errors indicate more accurate forecasting results. All four statistical indicators were used in assessing the overall prediction accuracy of the forecasting models with respect to the test data.

The NN models' forecasting performance is summarized in Table 2. It can be observed that the use of ensemble NNs provide lower error magnitudes across all statistical indicators. Lewis (1982) provided a scale for interpreting the forecast accuracy based on MAPE values as follows:

– Highly accurate forecast: MAPE < 0.1 (10%);
– Good forecast: 0.1 (10%) < MAPE < 0.2 (20%);
– Reasonable forecast: 0.2 (20%) < MAPE < 0.5 (50%);
– Inaccurate forecast: MAPE > 0.5 (50%).

Thus, applying the scale developed by Lewis (1982) on the results summarized in Table 2, it can be concluded that NN-based PCI prediction model provides good forecasting accuracy. With the availability of data for more years, the NN forecasting accuracy can be significantly improved. A scatterplot of observed and (ensemble) NN predicted PCI values is shown in Fig. 3 revealing the closeness of the graph to the 45-degree line. Similar results for NN predicted granular and blading maintenance costs are displayed in Fig. 4. A MAPE of 24% (Fig. 4) for the ensemble NN granular and blading maintenance cost model indicates that it can provide reasonable prediction accuracy. The histograms of network (error) residuals for ensemble PCI and granular and blading maintenance costs are displayed in Figs 5 and 6, respectively, confirming that the residuals are close to a normal distribution with a mean value of zero.

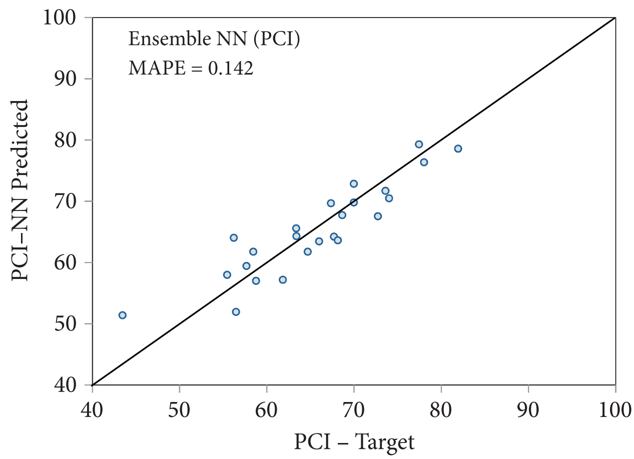

Fig. 3. Scatterplot of observed versus ensemble NN predicted PCI

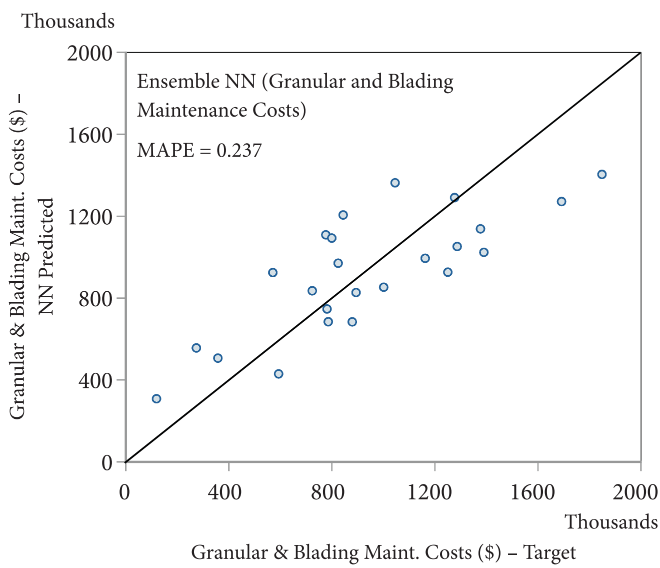

Fig. 4. Scatterplot of observed versus ensemble NN predicted granular and blading maintenance costs

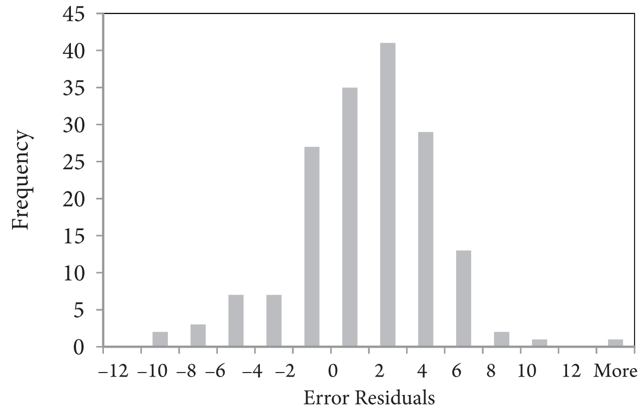

Fig. 5. Histograms of ensemble NN PCI error residuals (predicted – target) using both training and test data

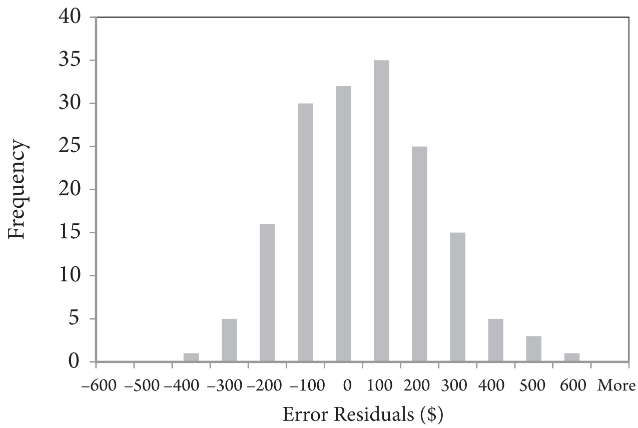

Fig. 6. Histograms of ensemble NN granular and blading maintenance cost error residuals (predicted – target) using both training and test data

A Global Sensitivity Analysis (GSA) was carried out to investigate the relative importance of variables used in NN modeling. GSA determines the effect of changes in input variables on network predictions and the prediction error rates. Thus, the prediction error will not increase much if an unimportant variable is changed or removed. Fig. 7 displays the Global Sensitivity Ratios (GSRs) obtained using the ensemble NN PCI model by dividing the network error with a given input omitted by the network error with the input available. Variables with GSR equal to or less than 1.0 are likely to be less important in terms of their contribution to network performance. The presence of both biofuel plants and wind farms, as well as rural vehicle miles traveled (VMT) seem to be significant factors impacting pavement deterioration (PCI).

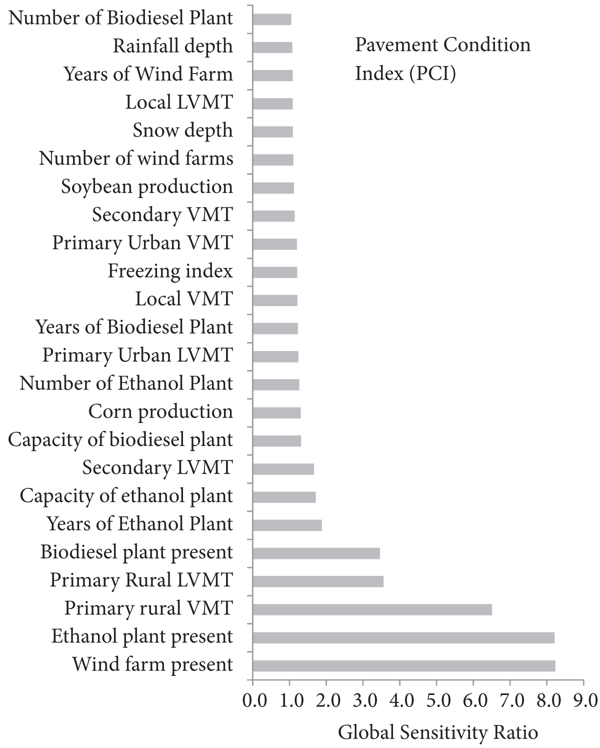

Fig. 7. Global sensitivity analysis results using Ensemble NN PCI model

Similar GSRs obtained using the ensemble NN granular and blading maintenance cost model are displayed in Fig. 8. Again, the presence of ethanol plants and biodiesel plants followed by secondary VMT and large-truck VMT (LVMT) have significant influence on granular and blading maintenance costs. This is also consistent with the findings reported by Gkritza *et al.* (2011) who observed an increasing trend in maintenance-related costs in the year a biofuel plant was constructed as well as after it became operational. Fig. 8 also seems to indicate that the effect of biodiesel/ethanol plants is higher on gravel roads maintenance-related costs than that of wind farms. This is because once the wind turbines are installed, there might be no further road deterioration due to the operation of wind turbines as opposed to ethanol and biodiesel plants that require regular transportation of raw material and final products during operation.

As mentioned before, the primary goal of this paper was to investigate the feasibility of employing the NN methodology to forecast the impacts of Iowa's biofuels and wind power industries on Iowa's secondary and local road condition and maintenance-related costs. The developed NN forecasting models are expected to become more robust in their predictive accuracy as the database expands with data collected over many years. These models would be useful in evaluating isolated impacts of renewable energy industry on Iowa's transportation infrastructure. For instance, One-at-A-Time (OAT) local sensitivity analysis could be

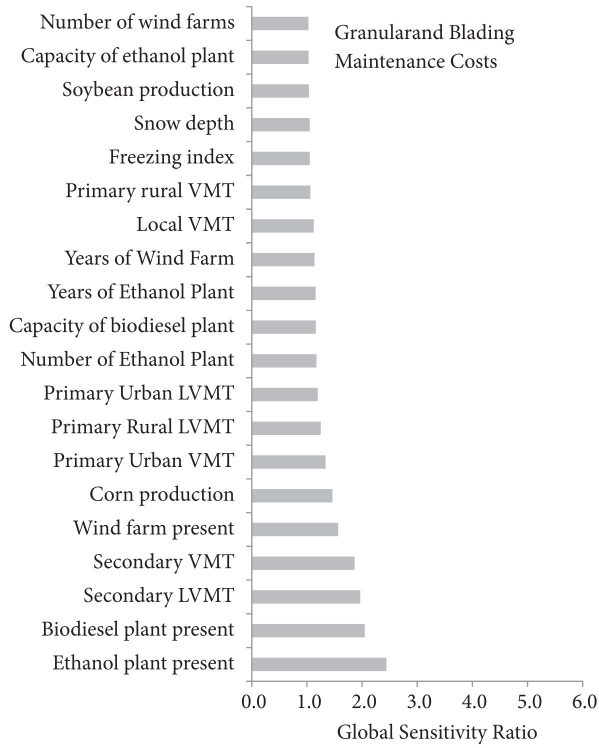

Fig. 8. Global sensitivity analysis results using Ensemble NN granular and blading maintenance costs model

carried out to study the effect of input factors such as the presence of biofuel plants and wind farms, categorized VMT, etc. on pavement deterioration (PCI). The input factors could then be triaged qualitatively (very sensitive, sensitive, and not sensitive) based on their expected effects on pavement deterioration. Similarly, the ensemble NN models could be used to run a large number of GSA simulations to provide predictions of pavement deterioration at random discrete locations in the problem domain which could then be fitted using a continuous Response Surface Model (RSM).

Summary and conclusions

U.S. continues to be a leading producer of biofuels and wind energy. In order to meet the U.S. Energy Independence and Security Act of 2007 (EISA) goal of 36 billion gallons per year of renewable transportation fuels by 2022, a dramatic increase in ethanol and biodiesel production and distribution is expected. Consequently, the stress on the transportation infrastructure resulting from production and shipping of raw materials as well as finished products is expected to be significant. While semi-trucks are used mainly to transport corn and soybean from farm to biofuel plants, longer and heavier vehicles are required to move wind turbine blades and other parts from manufacturing sites to the wind farms. In coming years, these drivers are expected to significantly impact Iowa's rural transportation infrastructure, much of which is near or beyond its original design life. At the same time, public sector maintenance and rehabilitation costs associated with Iowa's rural pavements and unpaved roadways are expected to escalate to sustain the renewable energy industry-oriented traffic.

In this paper, NN modeling was proposed as an alternative tool for forecasting transportation infrastructure impacts of renewable energy industry in Iowa using panel data. The data for this study was collated from a number of sources aimed at documenting the physical and fiscal impacts to the transportation system due to renewable energy facility construction or operations in Iowa. The data is structured as a panel of 24 cross-sections (24 counties in the northern, western, and southern part of Iowa) and ten time periods from 1999 to 2008, and includes maintenance-related costs, Pavement Condition Index (PCI), renewable energy plant capacity, years of plant operation, traffic volume information, and environmental and agricultural factors.

Back-propagation NN models using a Quasi-Newton second-order training algorithm were developed to compute forecasts. Ensembles of well-trained NNs were formed to achieve significant improvements in generalization performance. NN models seem to offer good forecasts of pavement condition (PCI) and reasonable forecasts of gravel road maintenance-related costs for the year on which the models were tested. Further, the NN models could identify the presence of biofuel plants and wind farms as well as large-truck traffic as the most sensitive parameters impacting pavement condition and maintenance-related costs. Pavement deterioration resulting from traffic loads are associated with both the biofuel plants and wind farms. The effect of biodiesel/ethanol plants is higher on gravel roads maintenance-related costs than that of wind farms. This is because the operation of wind turbines may not impact roads at a regular basis as opposed to ethanol and biodiesel plants that require regular transportation of raw material and final products during operation.

Due to the availability of limited data, it was difficult to estimate the number of years (time periods) required for proper training, validation, and testing. As a result, the developed models could only be tested for one year. The availability and use of longer time series is a promising avenue to validate this work.

In developing the NN forecasting models, the effect of specific regional and time characteristics were taken into account using a slightly modified approach rather than using regional dummies and time dummies as typically done in panel data regression modeling. An alternative way of modeling regional-specific characteristics is to group the 24 counties in three directional bins (north, south, west), in which case r ranges from 1 to 3, and R = 3. It is recommended that future studies validate the efficacy of this approach.

References

Adeli, H. 2001. Neural networks in civil engineering: 1989-2000, *Computer-Aided Civil and Infrastructure Engineering* 16(2): 126–142. http://dx.doi.org/10.1111/0885-9507.00219

Adeli, H.; Hung., S. L. 1995. *Machine learning: neural networks, genetic algorithms, and fuzzy systems*. New York: Wiley. 211 p.

Adya, M.; Collopy, F. 1998. How effective are neural networks at forecasting and prediction? A review and evaluation, *Journal of Forecasting* 17(5–6): 481–95. http://dx.doi.org/10.1002/(SICI)1099-131X(1998090)17:5/6<481::AID-FOR709>3.3.CO;2-H

Aleksander, I.; Morton, H. 1990. *An introduction to neural computing*. New York: Van Nostrand Reinhold Co. 284 p.

Autant-Bernard, C. 2012. Spatial econometrics of innovation: recent contributions and research perspectives, *Spatial Economic Analysis* 7(4): 403–419. http://dx.doi.org/10.1080/17421772.2012.722665

Bai, Y.; Hwang, T.; Kang, S.; Ouyang, Y. 2011. Biofuel refinery location and supply chain planning under traffic congestion, *Transportation Research Part B* 45: 162–175. http://dx.doi.org/10.1016/j.trb.2010.04.006

Bishop, C. M. 1995. *Neural networks for pattern recognitio*n. New York: Oxford University Press Inc. 482 p.

Chen, M. Y. 2011. Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches, *Computers* & *Mathematics With Applications* 62(12): 4514–4524. http://dx.doi.org/10.1016/j.camwa.2011.10.030

Cheng, B.; Titterington, D. M. 1994. Neural networks: a review from a statistical perspective, *Statistical Science* 9(1): 2–30. http://dx.doi.org/10.1214/ss/1177010638

Fausett, L. V. 1994. *Fundamentals of neural networks*. 1st ed. Prentice Hall, NJ. 461 p.

Fu, G.; Feng, J.; Dekelbab, W.; Moses, F.; Cohen, H.; Mertz, D. 2008. Impact of commercial vehicle weight change on highway bridge infrastructure, *Journal of Bridge Engineering* 13(6): 556–564. http://dx.doi.org/10.1061/(ASCE)1084-0702(2008)13:6(556)

Gupta, D.; Cohn, L. F. 2012. Intelligent transportation systems and NO2 emissions: predictive modeling approach using artificial neural networks, *Journal of Infrastructure Systems* 18(2): 113–118. http://dx.doi.org/10.1061/(ASCE)IS.1943-555X.0000073

Giovanis, E. 2008. *A panel data analysis for the greenhouse effects in fifteen countries of European Union* [online]. Munich Personal RePEc Archive (MRPA) Paper No. 10321. Available from internet: http://mpra.ub.uni-muenchen.de/10321/.

Gkritza, K.; Nlenanya, I.; Jiang, W.; Sperry, R.; Smith, D. 2011. Infrastructure impacts of Iowa's renewable energy, *Transportation Research Record, Journal of the Transportation Research Board* 2205: 238–246. http://dx.doi.org/10.3141/2205-30

Haddad, M. A.; Anderson, P. F.; Thol, S.; Hertel, C.; Schmidt, B. 2009. Assessing the future of the bioeconomy in Greene County, Iowa, *Journal of Rural and Community Development* 4(1): 15–33.

Haykin, S. 1998. *Neural networks: a comprehensive foundation*. 2nd ed. Prentice Hall, Upper Saddle River, NJ. 842 p.

Huang, M. T.; Tian, Y. 2013. A novel visual modeling system for time series forecast: application to the domain of hydrology, *Journal of Hydroinformatics* 15(1): 21–37. http://dx.doi.org/10.2166/hydro.2012.158

IEDA 2012. *Renewable fuels* [online]. Iowa Economic Development Authority. Des Moines, IA. Available from Internet: http://www.iowaeconomicdevelopment.com/business/ethanol_biodiesel.aspx

ISU 2012. *Iowa Environmental Mesonet* [online]. Iowa State University Department of Agronomy. Available from Internet: http://mesonet.agron.iastate.edu/GIS/

Kedron, P; Bagchi-Sen, S. 2011. A study of the emerging renewable energy sector within Iowa, *Annals of the Association of American Geographers* 101(4): 882–896. http://dx.doi.org/10.1080/00045608.2011.568875

Lewis, C. D. 1982. *Industrial and business forecasting methods*. London: Butterworths. 143 p.

Lin, C.-F. J. 1992. *The econometrics of structural change, neural network and panel data analysis*: PhD thesis, University of California, San Diego.

Longhi, S.; Nijkamp, P.; Reggiani, A.; Maierhofer, E. 2005. Neural modeling as a tool fort forecasting regional employment patterns, *International Regional Science Review* 28(3): 330–346. http://dx.doi.org/10.1177/0160017605276187

Mehra, P.; Wah, B. W. 1992. *Artificial neural networks: concepts and theory*. IEEE Computer Society Press, Los Alamitos. 667 p.

Monteiro, C.; Ramirez-Rosado, I. J.; Fernandez-Jiminez, L. A. 2013. Short-term forecasting model for electric power production of small-hydro power plants, *Renewable Energy* 50: 387–394. http://dx.doi.org/10.1016/j.renene.2012.06.061

Patuelli, R.; Reggiani, A.; Nijkamp, P.; Blien, U. 2006. New neural network methods for forecasting regional employment: an analysis of German labour markets, *Spatial Economic Analysis* 1(1): 7–30. http://dx.doi.org/10.1080/17421770600661568

Platoni, S.; Sckokai, P.; Moro, D. 2012. Panel data estimation techniques and farm-level data models, *American Journal of Agricultural Economics* 94(5): 1202–1217. http://dx.doi.org/10.1093/ajae/aas072

Qi, C.; Chang, N. B. 2011. System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts, *Journal of Environmental Management* 92(6): 1628–1641. http://dx.doi.org/10.1016/j.jenvman.2011.01.020

Shahin, M. A.; Jaksa, M. B.; Maier, H. R. 2001. Artificial neural network applications in geotechnical engineering, *Australian Geomechanics* 36(1): 49–62.

Sun, S. L.; Huang, R. Q.; Gao, Y. 2012. Network-scale traffic modeling and forecasting with graphical lasso and neural networks, *Journal of Transportation Engineering* 138(11): 1358–1367. http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000435

Swanson, N. R.; White, H. 1997. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks, *The Review of Economic and Statistics* 79(4): 540–550. http://dx.doi.org/10.1162/003465397557123

Swingler, K. 1996. *Applying neural networks: a practical guide*. London, UK: Academic Press. 303 p.

TRB Circular. 1999. *Use of artificial neural networks in geomechanical and pavement systems. Number E-C012.* Transportation Research Board, National Research Council, Washington DC.

Tsoukalas, L. H.; Uhrig, R. E. 1997. *Fuzzy and neural approaches in engineering*. New York: Wiley. 587 p.

USDA. 2012. United States Department of Agriculture. National Agriculture Statistic Service. Available from Internet: http://www.nass.usda.gov/Data_and_Statistics/Quick_Stats/.

U.S. DOE. 2012. *Biomass multi-year program plan* [online]. Office of Energy Efficiency and Renewable Energy (EERE). U.S. Department of Energy (DOE). DOE/EE-0405. Washington, D.C. Available from Internet: http://www1.eere.energy.gov/biomass/pdfs/mypp_april_2012.pdf.

Wakeley, H. L.; Griffin, W. M.; Hendrickson, C.; Matthews, H. S. 2008. Alternative transportation fuels: distribution infrastructure for hydrogen and ethanol in Iowa, *Journal of Infrastructure Systems* 14(3): 262–271. http://dx.doi.org/10.1061/(ASCE)1076-0342(2008)14:3(262)

Ye, Q.; Szeto, W. Y.; Wong, S. C. 2012. Short-term traffic speed forecasting based on data recorded at irregular intervals, *IEEE Transactions on Intelligent Transportation Systems* 13(4): 1727–1737. http://dx.doi.org/10.1109/TITS.2012.2203122

Kasthurirangan GOPALAKRISHNAN. Research Assistant Professor in the Department of Civil Construction, and Environmental Engineering at Iowa State University (ISU), Member of the American Society of Civil Engineers (ASCE) as well as an Editorial Board Member of the *Transport* and *International Journal for Traffic and Transport Engineering*. He received the Dwight D. Eisenhower Transportation Fellowship award in 1999 and he completed his PhD in 2004 at the University of Illinois at Urbana-Champaign, USA. His research interests include the use of bio-inspired computing paradigms in civil engineering informatics, sustainable and green technologies, simulation, imaging, and mechanics of transportation infrastructure systems.

Konstantina "Nadia" GKRITZA. Assistant Professor in the Department of Civil Construction, and Environmental Engineering at Iowa State University (ISU), Associate Director of the Mid-American Transportation Center as well as Associate Editor of the American Society of Civil Engineers (ASCE) Journal of Transportation Engineering. She completed her PhD in 2006 at Purdue University, West Lafayette, USA. She has received several awards including Schafer Award for Excellence in Teaching, Research, and Service at ISU, C.V. Wootan Memorial Award for outstanding PhD dissertation in Policy and Planning. Her research interests include transportation planning and systems evaluation, transportation economics, transportation energy and sustainability, highway safety, and (critical) infrastructure management.