Up-scaling Agent-Based Discrete-Choice Transportation Models using Artificial Neural Networks

Lance E. Besaw§
Graduate Research Assistant
Civil and Environmental Engineering
University of Vermont
213 Votey Hall
33 Colchester Ave, Burlington, VT, 05405
Telephone: (802) 656-4595
Fax: (802) 656-5838
Email: lbesaw@cems.uvm.edu

Michael B. Pellon
Graduate Research Assistant
Dept. of Computer Science
University of Vermont
204 Farrell Hall, 210 Colchester Ave
Burlington, VT, 05405
Telephone: (802) 656-4595
Fax:(802) 656-5838
Email: Michael.Pellon@uvm.edu

Donna M. Rizzo
Associate Professor
School of Engineering
University of Vermont
213 Votey Hall, 33 Colchester Ave
Burlington, VT, 05405;
Telephone: (802) 656-1495
Fax: (802) 656-8446
Email: drizzo@cems.uvm.edu

David K. Grover
Undergraduate Research Assistant
Civil and Environmental Engineering
University of Vermont
204 Farrell Hall, 210 Colchester Ave
Burlington, VT, 05405
Telephone: (802) 656-4595
Fax: (802) 656-5838
Email: David.Grover@uvm.edu

Margaret J. Eppstein
Associate Professor
Dept. of Computer Science
University of Vermont
327 Votey Hall, 33 Colchester Ave
Burlington, VT, 05405
Telephone: (802) 656-1918
Fax: (802) 656-5838
Email: Maggie.Eppstein@uvm.edu

Jeffrey. S. Marshall
Professor
School of Engineering
University of Vermont
231A Votey Hall, 33 Colchester Ave
Burlington, VT, 05405
Telephone: (802) 656-3826
Fax: (802) 656-3358
Email: Jeff.Marshall@uvm.edu

§Corresponding author

ABSTRACT
Agent based models (ABMs) can be used for simulating consumer transportation discrete choices, while incorporating the effects of heterogeneous agent behaviors and social influences. However, the application of ABMs at large-scales may be computationally prohibitive (*e.g.*, for millions of agents). In an attempt to harness the modeling capabilities of ABMs at large scales, we develop a recurrent artificial neural network (ANN) to replicate nonlinear spatio-temporal discrete choice patterns produced by a spatially-explicit ABM with social influence. This particular ABM has been developed to model consumer decision making between purchasing a Prius-like hybrid or plug-in hybrid electric vehicle (PHEV) for a given geographic region (*e.g.*, city or town). Our goal is to see if an ANN trained at the city scale can operate as a "fast function approximator" to estimate nonlinear dynamic response functions (*e.g.*, fleet distribution, environmental attitudes, etc.) based on city-wide attributes (*e.g.*, socio-economic distributions). Recurrent feedback connections were added to the ANN to leverage the temporal history and correlations and improve forecasts in time. Outputs from the city-scale ABM, run for a variety of population sizes and initial and input conditions, were used to train and test the ANN. Initial results suggest the ABM may be replaced by ANNs that interact with each other and other agents (*e.g.*, manufacturing agents) to investigate PHEV penetration at the national scale.

INTRODUCTION AND MOTIVATION

Regulatory actions by federal, state and local governments can play a critical role in influencing the transportation energy market. Due to the high degree of interdependency between various governing market factors, it is difficult to predict the market consequences and sensitivity to any given regulatory change.

Discrete choice decision models are commonly used to study transportation and travel phenomena, including vehicle choice behavior (1), hybrid choice behavior (2), and travel mode and destination choice (3). Agent based models (ABMs) with social influence are capable of modeling nonlinear spatio-temporal discrete choice patterns. ABMs have been used as an alternative methodology to model discrete-choice decisions in applications like driver route choice (4) and pedestrian walking behavior (5). Incorporating social influences in an agent-based discrete choice models make them more representative of real-world decision-making process (6), but drastically elevates the complexity of the simple binary choice model (7). While ABMs have proven useful for modeling behavior in complex systems (8) and demonstrate utility in the field of transportation (9), they can require large amounts of computation when implementing complicated decision with numerous agents.

The ABM in this work (10) has been developed to simulate the consumer discrete-choice decision process between purchasing a hybrid or plug-in hybrid (PHEV) vehicle type for a given demographic region (*e.g.*, city or town), with given socioeconomic characteristics. As a proof of concept, our ABM model uses synthetic data to reveal the importance of social influence on the vehicle purchasing behavior of its agents. However, this consumer discrete-choice model with numerous agents requires an alternative modeling strategy to replicate the behavior of the ABM when considering regulatory policies across scales larger than individual towns or cities (*e.g.*, larger than the state level).

In this research, we use an artificial neural network (ANN) to learn the dynamic behavior of the socially-influenced agent-based discrete-choice model and map its behavior for a wide range of synthetic demographic regions (towns) and socioeconomic characteristics. The ANN, operates as a fast function approximator, and requires recurrent feedback connections to allow outputs at one time step to be used as inputs for the next time step. We look to answer the question: can a simple ANN be used to replicate the behavior of a complex agent-based consumer discrete choice model with social influence? If proven successful, these ANNs will be surrogates for the ABM in a national-scale simulation that investigates alternative policies to influence PHEV fleet penetration.

BACKGROUND

Artificial neural networks (ANNs) are nonparametric statistical tools that can be viewed as universal approximators (11). They were developed as large parallel-distributed information processing systems in attempt to model the learning processes of the human brain. Many different types of ANN have been introduced over the years for problems in pattern recognition and function approximation. ANNs specialize in mapping nonlinear relationships given extremely large datasets (12). They have a relatively simple computational architecture, which makes them extremely powerful and computationally efficient. The ANN known as feedforward backpropagation (FFBP) has been used in numerous transportation modeling studies, for example, predicting travel time under transient traffic conditions (13), among others. However, FFBP does have some drawbacks, namely it can become trapped in local minima, requires optimization of parameters (*e.g.*, number of hidden layers and nodes, the learning coefficient and momentum) and can take extended periods of training to converge to an adequate solution.

In this work, we use a generalized regression neural network (GRNN) to forecast discrete consumer choices using socioeconomic, social influence and market condition descriptors as inputs. The GRNN has many advantages. It relaxes many of the assumptions required by traditional parametric statistical methods (*e.g.*, does not require an assumption of multivariate normality;

allows binary or categorical data). Unlike the FFBP network, the GRNN has one-pass training and guaranteed convergence. In transportation studies, the GRNN has been used to forecast daily trip flows (14), predict the hazardousness of intersection approaches (15), model travel mode choice (16), predict CO_2 fluxes (17), predict real-time driver fatigue (18) and real-time video traffic modeling (19,20).

METHODS

Agent Based Model

When selecting a vehicle to purchase, real-world consumers may compare a variety of characteristics, including fuel efficiency, seating and cargo capacity, safety, reliability, brand-loyalty, public perception, etc. Our model currently assumes that our agents are the subset of *new* vehicle consumers who have already narrowed their choice down to a less-expensive Prius-like hybrid and a higher-premium Prius-like PHEV.

Since PHEVs are not yet available in the commercial marketplace, we based PHEV price premiums, battery recharge requirements, electric assist ranges, and mileage with and without electric assist, from reported specifications for the Hymotion PHEV conversion kit for the Prius (21). The hybrid's fuel economy is 45 mpg while the PHEV's is 105 mpg when running on all-electric mode (45 mpg otherwise), with an all-electric range of 35 miles and 5.5 hour charging time at 5 kWH. We assume otherwise identical specifications and features between the two vehicles with only gas mileage and price premium differing.

Several major assumptions have been made in the development of our ABM to simplify the modeled processes. Our goal was not to exactly model real-world vehicle purchasing behavior, but rather to grossly approximate it and investigate the impact of social influence and regulatory policies. Here, we present a brief overview of the ABM; for more details, please refer to Pellon et al. (10).

Each agent represents a single vehicle consumer (not a household) and drives only one car (with no specification of vehicle purpose). Agents' socioeconomic characteristics are drawn randomly from distributions based on National Household Travel Survey (NHTS) data (22), including annual driving distance and durations of vehicle ownership. Agents are randomly distributed, yet clustered in space with an urban center and four suburban peripheral town-centers. Agents with similar annual salary have been loosely clustered in space using a 2-D turning bands method (23). Several agent attributes (*e.g.*, age, driving distance, number of years they typically own a vehicle, and willingness to consider adopting the new PHEV technology) are also positively or negatively correlated to salary, in varying degrees. The threshold for willingness to consider new PHEV technology was initialized so that roughly one half of new car buyers were willing to consider being PHEV early-adopters, consistent with recent survey results that indicated that 46% of potential consumers reported that they were some chance they might purchase a PHEV, depending on the price premium (24).

Agents have specified social and spatial neighbors that make up their social and geographical networks. The spatial network is defined on the physical proximity of the agents, while the social network is based on physical proximity and similar socioeconomic characteristics. These networks affect the agents' decision-making process, as agents look in their networks to see what vehicles other agents currently own; in addition, agents' attitudes can be influenced by other agents in their social networks. Heterogeneity in agent locations, social networks and driving distances cause different agents to be exposed to different vehicle fleets. The proportion of PHEVs within an agent's observed fleet influences their willingness to consider adopting this new technology. This "threshold" concept is found to be a very important feature in social influence models (7; 25).

Agents stochastically decide when to purchase a vehicle based on their current vehicle's age, the number of years they expect to own a car and how much more attractive a new vehicle is compared with their current vehicle. Once an agent decides to purchase a vehicle, they compare the relative financial costs and environmental benefits to determine whether the PHEV (assuming their threshold has been met) or hybrid vehicle best suits their needs, and then purchase the best vehicle for their circumstance. This process is repeated every year, for all agents over the 15 year simulation time period.

In addition to agents deciding to purchase vehicles, some of their internal characteristics are updated every year. For example, their environmental attitude (or greenness) and the time period over which they compute potential fuel savings may both be increased by social influence, depending on their social susceptibility. Greenness is a weighting factor ranging between 0 and 1 that determines how much an agent values the perceived environmental benefits of the PHEV (in this paper, proportion of gas saved by the PHEV relative to the HEV) vs. the perceived financial benefits of the HEV (proportion of cost savings of the HEV relative to the PHEV); a greenness value of 0 implies the decision is based solely on perceived financial benefits. In estimating relative costs, some agents ignore potential fuel savings and simply look at the price premium of the PHEV, some also compute projected fuel costs over a period of 1 year, and others compute projected fuel costs over the entire duration for which they anticipate owning their next car. Thus, some agents are more forward-thinking than others, and these agents can influence others in their social network to become more forward-thinking (i.e., to consider projected fuel costs over longer periods), resulting in interesting dynamics in discrete-choice behavior. Note that if an agent's projected fuel savings exceed the PHEV price premium and the PHEV will be thus perceived as the cheaper vehicle, the agent will purchase the PHEV, regardless of their greenness value (assuming their threshold has been met). See (10) for more details.

Our model assumes there is no shortage of either vehicle type (*i.e.*, no waiting period). There are several additional exogenous inputs: PHEV price premium as well as future projections of gas prices and current national electricity prices (26).

Generalized Regression Neural Network

Developed as a nonlinear, non-parametric extension of multiple linear regression, the GRNN is a memory-based network capable of estimating continuous variables (27). The GRNN consists of four layers: input, pattern, summation, and output (Figure 1). Each layer is fully connected to the adjacent layers by a set of weights between the nodes. Some output, $\hat{Y}(t)$, is predicted based on a set of input variables, $\mathbf{x}$, defined by some non-linear function $Y = f(\mathbf{x})$, captured by the training data. Training data consist of a set of input vectors, $\mathbf{x}$, and corresponding output, Y (input-output pairs). For this application, we use the algorithm to predict two output variables (Y^1 and Y^2), which represent median greenness and proportion of PHEVs in a given town, respectively.

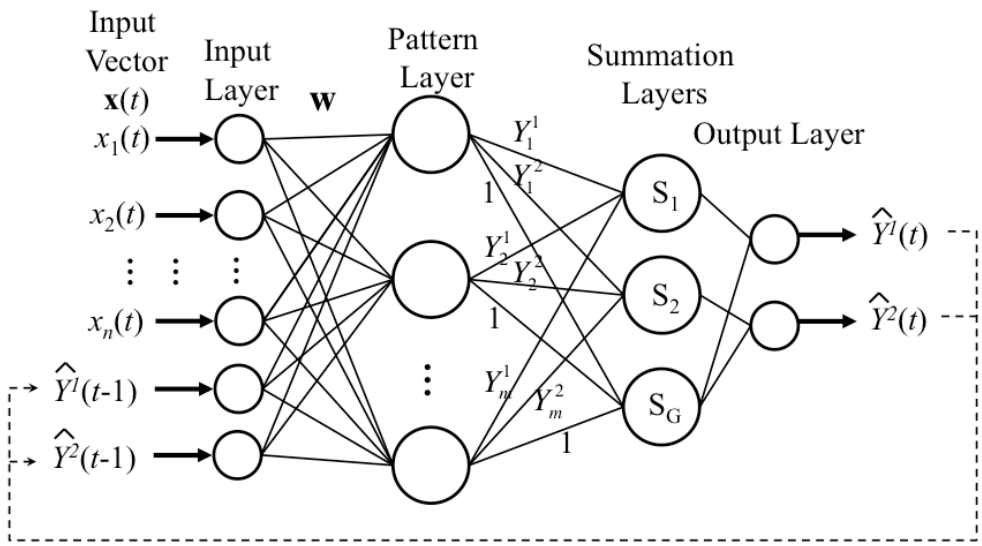

Figure 1. Architecture of a general regression neural network with recurrent feedback connections for two outputs variables Y^1 and Y^2.

The user specifies the number of nodes in the input layer by determining what predictor variables best represent the system being modeled. The pattern layer has one node for each of the *m* training patterns. The weights on the left side of the pattern layer nodes store (*e.g.*, are set equal to) the input training vectors, **x**. Each node in the pattern layer is connected to the summation layer nodes, S_1, S_2 and S_G. The weights linking the pattern layer nodes with summation nodes S_1 and S_2 store the model outputs for the two variables being predicted (*e.g.*, Y_i^1, Y_i^2) for all input-output training patterns (*i*=1, 2,…*m*). The weights from the pattern layer nodes to summation node S_G are set equal to 1.

Once the weights are set, the GRNN may predict each of the two outputs. A new input vector for which a prediction is desired, **x**, is presented to the pattern layer. The Euclidean distance is computed between the input vector and all pattern weight vectors, $\mathbf{w}_i$, where *i*=1,2,…*m* as: $D_i^2 = (\mathbf{w}_i - \mathbf{x})^T(\mathbf{w}_i - \mathbf{x})$. The distance, D_i^2, is passed to the summation layers and a prediction for variable $\hat{Y}^1$ is computed as:

$$\hat{Y}^1 = \frac{S_1}{S_G} = \frac{\sum_{i=1}^{m} Y_i^1 \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^{m} \exp\left(-\frac{D_i^2}{2\sigma^2}\right)},$$

where σ^2 is a smoothing parameter that is pivotal for estimating $\hat{Y}^1$ (this process is also executed for predicting $\hat{Y}^2$). Large values of σ^2 smooth the regression surface and produce estimates that approach the sample mean; while small values produce a surface with greater chance of discontinuity resulting in nearest neighbor estimates. Intermediate values of σ^2 produce well behaved estimates that approximate the joint probability density function of **x** and *Y* (27). The prediction, $\hat{Y}^1$, is a weighted average of all stored response observations ($\hat{Y}_1^1$, $\hat{Y}_2^1$, …$\hat{Y}_m^1$), where each response is weighted exponentially according to its Euclidean distance from input vector $\mathbf{x}_i$ (the same computations are computed for output variable $\hat{Y}^2$).

In addition, the GRNN has been modified to allow for recurrent feedback connections (dashed line of Figure 1). Recent predictions for $\hat{Y}^1$ and $\hat{Y}^2$, are passed back to the input layer and used to predict $\hat{Y}^1$ and $\hat{Y}^2$ at the next time step. For more detail, please refer to Besaw et al. (28). The GRNN algorithm described in this work was written in MatLab V. 7.4.0.287 (R2007a).

ABM and GRNN implementations

The ABM was run under several scenarios to provide a range of town socioeconomic and vehicle fleet distributions (Table 1). Prior experimentation revealed similar ABM trends when run for 10,000 or 1,000 agents. To save on computation, our scenarios use 1,000 agents. The hypothetical region of interest consisted of one larger (population of 726 agents) and 4 smaller towns (populations ranging from 55 to 79 agents) (Figure 2a). The socioeconomic distributions of these areas where generated pseudo-randomly, different from town to town and included spatial cross-correlation of salary and other inter-attribute correlation. Agents in the respective towns had different social networks based on the proximity of neighboring agents and their socioeconomic characteristics. In addition, the ABM simulations use 10 initial random seeds to account for potential stochastic noise in the results.

Table 1. List of ABM parameters that were varied during the generation of training, validation and prediction datasets.

Model Parameter	Training & Validation Dataset	Prediction Dataset
Median Social Susceptibility	[0.01, 0.09, 0.33, 0.49]	[0.17, 0.45]
Proj. Gas Price	Low, Medium, High	Low-mid, Mid-High
PHEV Price Premium	$5k, $10k, $15k	$7k, $13k
Town Identification	1, 2, 3, 4, 5	1, 2, 3, 4, 5
Region Population	1,000	1,000
Initial Random Seeds	1 to 10	1 to 10

For this proof-of-concept, the agent thresholds for social susceptibility varied from 0 to 1, making a broad distribution of potential early adopters and non-conformists. Social susceptibility distributions were stochastically generated between 0 (not socially susceptible) to 1 (very susceptible). The medians of these distributions were varied to generate agent populations with large susceptibility variations (Table 1 and Figure 2b). Agent initial greenness distributions were also stochastically generated but remained statistically similar between scenarios. Exogenous model inputs (*i.e.*, projected gas prices (Figure 2c) and PHEV price premium were also varied to produce representative scenarios. U.S. Energy Information Administration (26) provided reasonable projections of high and low gas prices (medium was the average of high and low). Different PHEV-premiums were used to investigate the potential impact of price incentives. Greenness was stochastically initialized to range from 0 to 1, with an initial median value of approximately 0.17; as shown for a representative distribution in Figure 2d.

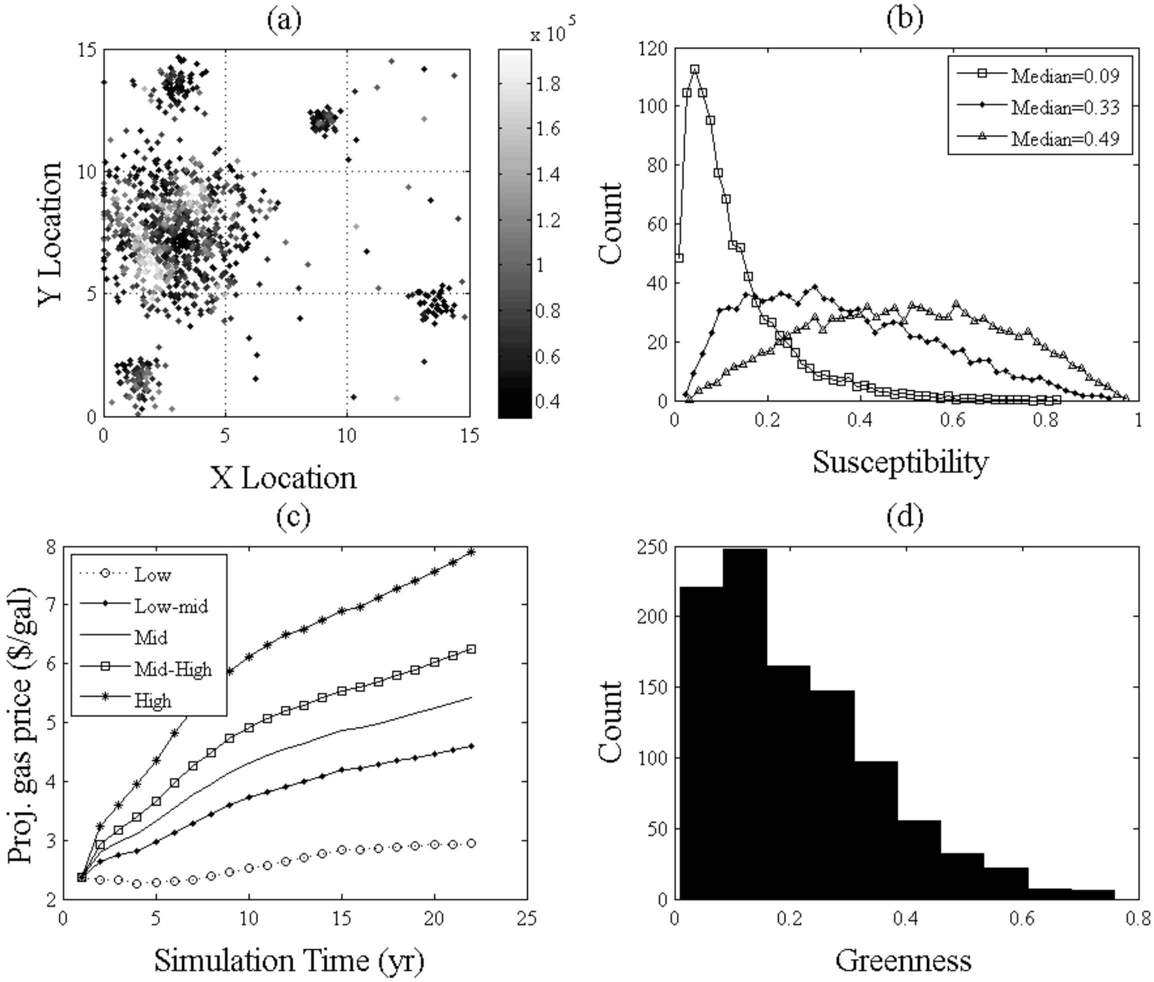

Figure 2. (a) Spatial coordinates of agents and their annual income shown in grayscale. (b) Example distributions of social susceptibilities with different medians. (c) Fifteen year projections of gas prices (high and low data from EIA). (d) Representative distribution of agents' initial greenness.

Descriptive statistics (median and interquartile range) of the ABM simulations were computed at each time step for the numerous agent distributions. These were then averaged over the 10 random seeds for each town and simulation setup. This resulted in a total of 4,320 different town simulations used to train the GRNN (comprising a 15-year time series when PHEVs are introduced in year 2). As is typical in ANN applications, this dataset was separated into training and validation datasets. The training dataset comprised 3,000 town simulations (~80%) to set the GRNN weights. The validation dataset was comprised of the remaining 1,320 town simulations (~20%) and was used to optimize the GRNN smoothing parameter σ^2 (via trial and error).

GRNN inputs include many of the socioeconomic descriptors incorporated in the ABM, including a statistical descriptors agents' age, income and social susceptibilities, PHEV thresholds, driving distances and car replacement age. In addition exogenous inputs include projected gas prices, and PHEV-price premium. All inputs were normalized such that their minimum and maximum values were 0 and 1 respectively. Outputs from the GRNN are statistical descriptors of agent greenness (median) and the proportion of the fleet that are PHEVS in a particular town. As this is a recurrent GRNN, these outputs are fed back into the GRNN and used as inputs in the next time step.

To generate the prediction dataset, an entirely new spatial distribution of agents and their corresponding socioeconomic characteristics were generated (Table 1). Other initial model parameters were changed as well, including: social susceptibility, projected gas prices and PHEV-premium. Again, descriptive statistics were computed for each town at every time step and averaged over the 10 random seeds. This combination of parameters resulted in prediction scenarios that were similar to, yet significantly different from, those upon which the GRNN was trained.

RESULTS AND DISCUSSION

We have summarized the accuracy of the GRNN predictions for both the validation and prediction datasets (Table 2) for both PHEV penetration and one of the changing agent attributes (greenness). We have used the coefficient of determination (R^2), and in some cases the root-mean-square error (RMSE) between the GRNN predictions and the ABM results as accuracy metrics. The means and standard deviations of R^2 are computed for each town over the given number of scenarios. The optimized smoothing parameter (σ^2=0.004) was used to predict PHEV fleet proportion and greenness in each of the five towns for 171 validation and 16 prediction scenarios.

GRNN predictions of ABM PHEVs and greenness

Although the time horizon over which individual agents projected fuel costs for prospective vehicles was allowed to vary dynamically due to social influence, current implementation of the GRNN does not feed this back as a recurrent input. Despite this known source of error, the coefficients of determination for the validation dataset (Table 2) show that the GRNN was able to map the general behavior of the discrete-choice ABM. When predicting PHEV-fleet proportion and greenness in the validation dataset, the GRNN was most accurate when predicting larger towns (*e.g.*, towns 1 and 3, Table 2). The GRNN predictions were least accurate in town 4 (PHEV average and standard deviation of R^2 were 0.57 and 0.33) where the number of agents was least.

The GRNN performed better when predicting PHEV fleet proportion than when predicting greenness in all validation scenarios. In addition, there exist large amounts of variability between the coefficients of determinations computed in the validation dataset (for both PHEV and greenness). These effects are most likely caused by the large variation of social susceptibility in the training and validation datasets (Table 1). Figure 2c provides a visual comparison of some of these distributions. When social susceptibilities are so drastically different, the trajectories of the greenness, and to a lesser extent PHEV fleet proportion, are very different. It appears there may have been too much variation in our distributions of social susceptibility in the training and validation datasets that limited the GRNNs ability to accurately learn all of the 3,000 scenarios. Future experiments will add the distribution of time horizons for fuel cost projections as a recurrent input into the GRNN to see if this improves the situation.

The R^2's computed for the predictions dataset indicate the GRNN has successfully learned the relationships from the training dataset (including the influence of social influence/networks) and was able to accurately predict PHEV fleet proportion and greenness. The scenarios used for prediction were generated using entirely new datasets with respect to the projected gas prices, PHEV price premium, spatial distribution and social susceptibility (Table 1). However, we did not introduce as much variation in the social susceptibility as was introduced in the training and validation datasets (medians of 0.17 and 0.45).

As a result, for the particular scenarios presented here, the GRNN was more accurate predicting on the prediction dataset than on the validation dataset. The estimated PHEV-fleet proportion in the validation and prediction datasets had R^2's of 0.8 and 0.97 respectively (0.75 and

0.87 for greenness). Similar to the validation dataset, we predicted PHEV-fleet proportion more accurately than greenness in the prediction dataset, 0.97 and 0.87 respectively. The greater accuracies are most likely due to the less variable distributions of social susceptibility used in the prediction dataset.

Table 2. Table of summary statistics for the coefficient of determination (R^2) computed for the validation and prediction datasets (a mean R^2 of 1.0 indicates the ANN predictions perfectly match the ABM results).

Output Variable	Dataset	Statistic	Town 1	Town 2	Town 3	Town 4	Town 5	All Towns
Number of Agents	Validation & Prediction	Mean	74	66	726	55	79	1000
PHEV-Fleet Proportion	Validation (n=171)	Mean R^2	0.89	0.83	0.88	0.57	0.83	0.80
		Std. Dev. R^2	0.22	0.34	0.24	0.33	0.25	0.28
	Prediction (n=16)	Mean R^2	0.97	0.98	0.99	0.94	0.96	0.97
		Std. Dev. R^2	0.02	0.01	0.01	0.03	0.02	0.02
Greenness	Validation (n=171)	Mean R^2	0.85	0.82	0.87	0.47	0.76	0.75
		Std. Dev. R^2	0.27	0.31	0.24	0.29	0.29	0.28
	Prediction (n=16)	Mean R^2	0.83	0.74	0.97	0.89	0.91	0.87
		Std. Dev. R^2	0.09	0.06	0.01	0.04	0.01	0.04

Further analysis into two prediction scenarios

To further highlight the GRNN capabilities to learn the behavior of this discrete-choice ABM, we have selected two scenarios (call them I and II) from our possible 16 prediction scenarios and plotted the time-series of PHEV proportion and greenness for towns 2, 3 and 4. The social susceptibility distribution was with only parameter that varied between scenario I and II (Table 3).

Scenarios I and II demonstrate several different ABM phenomena that we would like our GRNN to be able to replicate, including: differential rates of PHEV adoption, different final adoption proportion as well as linear and non-linear dynamics of adoption. These interesting dynamics have arisen out of the different distributions of social susceptibility used in these two scenarios. The time series plots provided below assume the PHEV was been introduced in year 0 (with year -1 being our initial model conditions).

Table 3. Parameter values for the two representative scenarios I and II.

Parameter	Scenario I	Scenario II
PHEV premium	$13k	$13k
Social Susceptibility Median	*0.17*	*0.45*
Proj. Gas Price	Low-mid	Low-mid
Representative Simulation	Figure 3	Figure 4

In scenario I, we observe that roughly 50% of the agents in towns 2 and 3 (Figures 3a and 3b) have adopted PHEVs. Due to the low economic status of agents in town 4 (Figure 3c), a much lower fraction of these agents have adopted PHEV. These three towns have different PHEV adoption trajectories. In town 2 (Figure 3a) we see highly nonlinear behavior. There is slow adoption at first, followed by an increased rate of adoption. In town 3 (Figure 3b), the adoption rate remains fairly constant over the simulation. Finally, town 4 (Figure 3c) shows a nonlinear jump in adoption.

The GRNN accurately predicts the time-series of PHEV adoption for towns 2 and 3 (R^2 of 0.99 and 1, respectively). This is encouraging because one trajectory is nonlinear while the other is linear. The GRNN does relatively poor when predicting the PHEV-fleet proportion for town 4. As in the training and validation dataset, town 4 continues to be the least well predicted of the five towns (Table 2). However, results demonstrate the GRNN has learned both linear and nonlinear dynamics of this ABM.

As illustrated with the training and validation datasets, the GRNN does not predict greenness (Figures 3d, 3e and 3f) as wells as it predicts PHEV-fleet proportion. In Town 2, the GRNN accurately predicts greenness (R^2=0.97). However, the coefficients of determination and RMSE show that the GRNN does not predict greenness in towns 3 and 4 very well. This is most likely due to the high variability of social susceptibility in the training dataset, upon which greenness is primarily based.

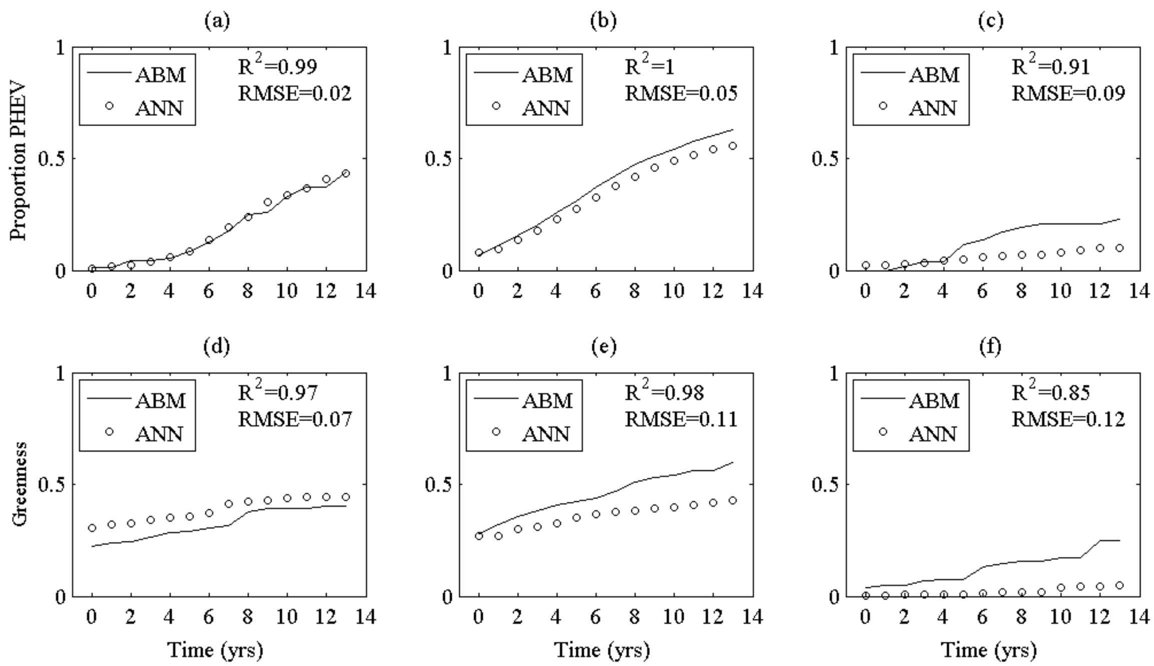

Figure 3. Representative GRNN predictions (for scenario I in which the PHEVs are adopted by ~50% of the consumer agents). GRNN predictions of PHEV fleet proportion versus time for (a) town 2, (b) town 3 and (c) town 4. (d-f) GRNN predictions of greenness for towns 2, 3 and 4 respectively.

In scenario II, we observe a greater proportion of the agents adopting PHEVs in towns 2 and 3 (Figures 4a and 4b). Fewer PHEVs were adopted by town 4 (Figure 4c), indirectly due to its low annual incomes (Figure 2a, rightmost town) which were correlated with longer vehicle ownership times and higher PHEV adoption thresholds. We see non-linear adoption rates in towns 2 and 3. Adoption is relatively slow initially, but increases with time. In this scenario, the GRNN again accurately predicts PHEV-fleet proportion for towns 2 and 3 (R^2 of 0.98 and 1 respectively). The GRNN greenness predictions for towns 2 and 3 are more accurate than scenario I indicating that the GRNN may be slightly biased, as well as more accurate, when predicting using larger social susceptibility medians. This could be remedied easily using more training scenarios with lower social susceptibility medians and by allowing the time horizon for fuel cost projections, which are also affected by social susceptibility, to be recurrent.

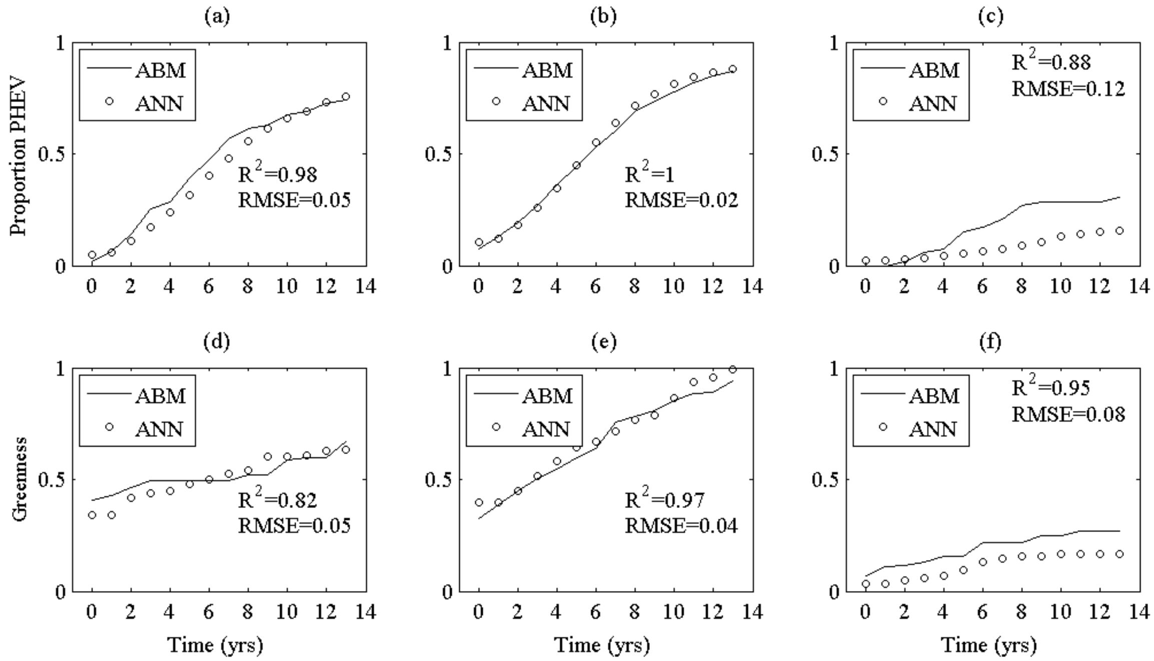

Figure 4. Representative GRNN predictions (for scenario II in which the PHEVs are adopted by ~80% of the consumer agents). GRNN predictions of PHEV fleet proportion versus time for (a) town 2, (b) town 3 and (c) town 4. (d-f) GRNN predictions of greenness for towns 2, 3 and 4 respectively.

Scenarios I and II present a good examples of the impact of social influence and networks on the discrete-choice ABM. In scenario I, the social susceptibility distribution was skewed left causing the general population to be less influenced by their social network (median = 0.17). Conversely, the social susceptibility distribution of scenario II was less skewed (median = 0.45), resulting in agents that were more influenced by their social network. The PHEV-fleet proportion curves of scenario II (greater social influence) had a larger number of PHEVs adopted in towns 2 and 3 than in scenario I (roughly 50% and 80% respectively). This is due to the difference in contribution of social influence in the two scenarios.

The predictions for these two scenarios show the GRNN was able to accurately predict PHEV-fleet proportion under these two different conditions (less and more social susceptibility). This is an important contribution of this work; because accurate predictions of PHEV adoption will allow us to use the GRNNs as a surrogate for the ABM when predicting PHEV adoption under many different types of scenarios. It also demonstrates that the GRNN has been able to learn the importance of social influence and networks generated by the ABM.

These two scenarios demonstrate the GRNN has learned and accurately predicted important ABM phenomena, including: differential rates of PHEV adoption, different final adoption proportion, as well as, linear and non-linear dynamics of adoption.

Computational Speedup

The speed of computation is another important consideration of this work. Our objective is to use the GRNN as a surrogate for the ABM for large-scale simulations (*e.g.*, nation scale). To be a viable surrogate, the GRNN must be computationally faster than the ABM for large-scale simulations.

The ABM scales super linearly with increasing number of agents (N). This is due to the large amounts of computation performed at every time step for every agent. The ABM takes: 11.5 seconds for 1000 agents, 70 seconds for 10000 agents, and continues to rise. Example computations include the threshold to consider purchasing a PHEV based on social and geographic networks (which scale super linearly with N). Because of this super linear scaling, performing large scale simulations with the ABM may not be feasible. However, as demonstrated previously, the GRNN is capable of learning the non-linear dynamics of the discrete-choice ABM.

The production of the GRNN training and validation datasets also scale non-linearly with N, due to the same computations discussed above. It took approximately 24 hours to produce the training and validation dataset (4,320 scenarios) used in this proof-of-concept. However once these datasets are developed, the GRNN takes relatively little time to train (~4 hours in this work). Once the GRNN is trained, it takes very little time to run (~0.4 seconds for any number of agents), no matter how large of a town or city is being simulated. This is due to the fact that the GRNN must only march through 15 time steps, independent of the town's population. Thus, the major time required for the GRNN is the production of the training and validation datasets by the ABM. Thus, for the 4,320 scenarios (each with 1000 agents) in this small demonstration comparable timings for the ABM and ANN are 13.8 hours versus 0.48 hours, respectively.

These demonstration timings were generated on a 3 GHz Intel Core 2 Duo processor with 3 GB of RAM running MatLab V. 7.3.0.267 (R2006b).

CONCLUSIONS AND FUTURE WORK

This work was originally conceived based on the premise that the GRNN can operate as a fast function approximator of a discrete-choice agent-based transportation model with social networks and influences. For training and validation, we produced a large dataset that exhibited spatio-temporal dynamics of this simplified consumer vehicle choice ABM. The inclusion of social influence introduced through social networks, adoption thresholds and susceptibility resulted in nonlinear agent behavior.

This proof-of-concept GRNN has proven capable of learning the spatio-temporal dynamics of the discrete-choice ABM with social influence. By incorporating a recurrent feedback connection, the easy-to-train GRNN was able to adequately replicate the behavior of ABM at the town scale. Adding in additional recurrency for other dynamically changing attributes (in this case, time horizon for fuel cost projections) is expected to improve results even further. Although it does take significant amounts of time to generate the training and validation datasets and optimize the GRNN's smoothing parameter, once trained, it operates as a fast function approximator that demonstrated tremendous speedup time relative to the ABM. The combined effects of accurate approximation and dramatic speedup will allow us to simulate large-scale dynamics more computationally efficiently than running a large-scale ABM.

In order to investigate potential regulatory policies that can influence the adoption of PHEVs (*e.g.,* price incentives, rebates) a large-scale model is currently under development. This large-scale model incorporates additional types of agents (*e.g.*, vehicle manufacturing and electricity producing agents). With the GRNN able to operate as a surrogate of the consumer ABM, GRNNs representing neighboring towns will interact with each other and with other large-scale agents. With this framework, we can explore potential regulatory policies that may impact the decisions and behavior of large-scale agents.

ACKNOWLEDGEMENTS

This work was funded in part by the United States Department of Transportation through the University of Vermont Transportation Research Center. We gratefully acknowledge computational resources and expertise provided by the Vermont Advanced Computing Center, supported in part by NASA (NNX 06AC88G).

REFERENCES

[1] Train, K.E. and Winston, C., Vehicle choice behavior and the declining market share of U.S. automakers. *Int. Eco. Rev.*, Vol. 48 No.4 2007, pp 1469-1496.

[2] Bolduc, D., Boucher, N. and Alvarez-Daziano, R. Hybrid choice modeling of new technologies for car choice in Canada. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2082, 2008, pp. 63-71.

[3] Foehich, P., 2008. Changes in Travel Behavior of Commuters Between 1970 and 2000. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2082, 2008, pp. 35-42.

[4] Dia, H., An agent-based approach to modelling driver route choice behaviour under the influence of real-time information. *Transportation Research Part C: Emerging Technologies*, Vol. 10, No. 5-6, 2002, pp. 331-349.

[5] Antonini, G., Bierlaire, M. and Webber, M. Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, Vol. 40, No. 8, 2006, pp. 667-687.

[6] Tversky, A. and Kahneman, D. Rational choice and the framing of decision. *Journal of Business*, Vol. 59, No. 4, 1986, pp. 251-278.

[7] Watts, D. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, Vol. 99, No. 9, 2006, pp. 5766-5771.

[8] Axelrod, R.M., *The complexity of cooperation: agent-based models of competition and collaboration*. Princeton University Press, Princeton, NJ, 1997

[9] Davidson, P., Henesey, L., Ramstedt, L., Tornquist, J. and Wernstedt, F. An analysis of agent-based approaches to transport logistics. *Transportation Research Part C: Emerging Technologies*, Vol. 13, No. 4, 2005, pp. 255-271.

[10] Pellon, M., Eppstein, M.J., Marshall, J.M., Rizzo, D.M. and Besaw, L.E., An agent-based model for estimating PHEV penetration. *Transportation Research Board Annual Conference (in review)*, 2009.

[11] Haykin, S., 1998. *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Englewood Cliffs, NJ, 1998

[12] Rumelhart, D.E. and McClelland, J.L., *Parallel Distributed Processing*. Massachusetts Institute of Technology Press, Cambridge, MA, 1988.

[13] Mark, C.D. and Sadek, A.W., Learning Systems for Predicting Experiential Travel Times in the Presence of Incidents: Insights and Lessons Learned. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1879, 2004, pp. 51-58.

[14] Celikoglu, H.B. and Cigizoglu, H.K., Public transportation trip flow modeling with generalized regression neural networks. *Advances in Engineering Software*, Vol. 38, 2007, pp. 71-79.

[15] Kumara, S.S.P., Chin, H.C. and Weerakoon, W.M.S.B., Identification of Accident Causal Factors and Prediction of Hazardousness of Intersection Approaches, *Transportation Research*

Record: Journal of the Transportation Research Board, Vol. 1840, 2003, pp. 116-122.

[16] Celikoglu, H.K., Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling. *Mathematical and Computer Modelling*, Vol. 44, 2006, pp. 640-658.

[17] Coutts, A.M., Beringera, J. and Tappera, N.J., Characteristics influencing the variability of urban CO2 fluxes in Melbourne, Australia. *Atmospheric Environment*, Vol. 41, No. 1, 2006, pp. 51-62.

[18] Ji, Q., Zhu, Z. and Lan, P., Real-Time Nonintrusive Monitoring and Prediction of Driver Fatigue. *IEEE Transactions on Vehicular Technology*, Vol. 53, No. 4, 2004, pp. 1052-1068.

[19] Doulamis, A.D., Doulamis, N.D. and Kollias, S.D., An Adaptable Neural-Network Model for Recursive Nonlinear Traffic Prediction and Modeling of MPEG Video Sources. *IEEE Transactions on Neural Networks*, Vol. 14, No. 1, 2003, pp. 150-166.

[20] Gharavol, E.A., Khademi, M. and M.-R. Akbarzadeh-T., A New Variable Bit Rate (VBR) Video Traffic Model Based on Fuzzy Systems Implemented Using Generalized Regression Neural Network (GRNN), *IEEE International Conference on Fuzzy Systems*, 2006, pp. 2142-2148.

[21] Hymotion. http://www.a123systems.com/hymotion, 2009.

[22] Hu, P.S. and Reuscher, T.R., . *Summary of Travel Trends, 2001 National Household Travel Survey*, Federal Highway Administration, 2004

[23] Emery, X., 2008. A turning bands program for conditional co-simulation of cross-correlated Gaussian random fields. *Computers & Geosciences*, Vol. 34, 2008, pp. 1850-1862.

[24] Richard Curtin, Yevgeny Shrago, and Jamie Mikkelsen, Plug-in Hybrid Electric Vehicles, Reuters / University of Michigan Surveys of Consumers, The University of Michigan, 2009, http://www.sca.isr.umich.edu/main.php.

[25] Granovetter, M., 1978. Threshold models of collective behavior. Am. J. Sociol., Vol. 83, No. 6, 1978, pp. 1420-1443.

[26] Energy Information Administration. *Annual Energy Outlook 2009*, US Department of Energy, 2009

[27] Specht, D.F., A general regression neural network. *IEEE Transactions on Neural Networks,* Vol. 2, No. 6, 1991, pp. 568-576.

[28] Besaw, L.E., Rizzo, D.M., Bierman, P.R. and Hackett, W.R., Advances in ungauged streamflow prediction using artificial neural networks. *Journal of Hydrology (in review)*, 2009.