

Applying Advanced Uncertainty Methods to Improve Forecasts of Vehicle Ownership and Passenger Travel in China^{*}

Paul Natsuo Kishimoto[†]

July 6, 2012

Abstract

Transportation in China is characterized by unprecedented, rapid growth in all metrics, including energy use and emissions. Uncertainty in data, as well as in Chinese parameterization of transport growth models, makes forecasting these trends a challenge. A new application of the Markov chain Monte-Carlo sampling method yields large sets of parameters for models which predict passenger vehicle ownership and passenger travel volume, conditioned on historical data. The resulting suite of forecasts helps characterize the uncertainty in the posterior distribution of these macroscopic transport metrics. The need and prospects for wider application of the uncertainty methodology are discussed.

^{*}N.B. per the instructions at <http://www.usaee.org/usaee2012/paperawards.html>, this single-author student paper is submitted as a candidate for the Dennis J. O'Brien USAEE Best Student Paper Awards. The abstract is submitted for presentation at the conference even if no award is received.

[†]Research Assistant, MIT Joint Program on the Science & Policy of Global Change, +1 617 302 6105 <pnk@mit.edu>

1 Introduction

The expansion of personal transport in China and other rapidly developing nations has been characterized as “faster, sooner and more simultaneously” than the historical pattern in OECD nations (Marcotullio et al., 2005). One key growth area is the switch from low-speed modes to cars and light trucks that accompanies increasing household income (Schäfer et al., 2009). This trend has broad implications for the world’s largest market of vehicles, as well as for refined oil demand, a keen energy security concern in a country heavily and increasingly dependent on petroleum imports—in 2011, China imported 56.7% of the oil it consumed (Xinhua News Agency, 2012).

The paper proceeds as follows. Section 2 lays out the method, including a brief statement of the Markov chain Monte-Carlo (MCMC) algorithm and subtypes used (2.1), the construction of simplified models for vehicle ownership and passenger travel volume based on the literature (2.2), and sources of the historical transport data upon which the models are conditioned (2.3). In Section 3, the essential process of model tuning (selecting burn-in periods and verifying aperiodicity) is described and followed by a discussion of results from both models. Finally, Section 4 gives some general conclusions, with avenues for extension of the work and its potential use in support of other kinds of transport modelling.

1.1 Uncertainty in Forecasts of Chinese Transport Demand

Annual vehicle sales in China have recently grown by as much as 38% year-on-year—a pace unmatched elsewhere in the world, faster than in the period before 2007, and beyond the growth rate of the general economy (Wang et al., 2011). This rapid motorization is also characterized by *great uncertainty* arising from at least three sources. First, while the *data* produced by the National Bureau of Statistics of China (NBSC) are improving with time, the quality and accuracy of past figures are still in question, notwithstanding revised calculations and generally improving coverage over the past decade (Guan et al., 2012; Sinton, 2001). As a result, it is difficult to specify historical trends with precision, or to assess the soundness of using these data to develop forecasts. Second, the decisions of central and local governments

create *policy uncertainty* as to whether vehicle supply may be restricted, or demand redirected to other modes, i.e. public transport (Hao et al., 2011). Thirdly, the future *economic context* which will constrain consumer choices is unclear. Population projections from the United Nations bracket an enormous 450-million range in 2050 (UN Population Division, 2010); as a multiplier for per-capita ownership, this considerably affects vehicle stocks. Will the economic expansion continue apace, or will it settle—to what level, and how soon? Urbanization will create hundreds of cities of 1-2 million people, while at the same time possibly decreasing density in the directly-governed megacities (Nam and Reilly, 2012)—what transport services can the government make available to city dwellers, and how will they use them?

Together, these aspects of uncertainty in Chinese transport reduce the value of single point estimates. One response is to develop a range of scenarios which bracket uncertain outcomes of interest, for example as in Kishimoto et al. (2012); Wang et al. (2006); Huo et al. (2011); but scenario analysis, like a point forecast, is “always wrong” (de Neufville and Scholtes, 2011, pp.22–23), so a need remains for thorough characterization of the uncertainty in transport and energy projections.

2 Methodology

The paper will detail the construction of an MCMC sampler to generate forecasts which incorporate, rather than discard, uncertainty in historical data and the parameters of models commonly used to represent macroscopic trends in transportation. In doing so, it extends these analyses from econometric approaches yielding single, high-confidence forecasts to a suite of forecasts whose statistical properties yield additional, useful information about possible outcomes.

Important components of an MCMC algorithm are an underlying model, a parameter updating scheme, and accept-reject criteria (Rubenstein and Kroese, 2007).

Structural uncertainty—competing explanations or models for the same transport trends—is not addressed by this approach, but two separate functional forms are considered for the underlying model. The first is a S-shaped Gompertz curve describing *per-capita vehicle ownership*,

a form commonly used to model the adoption technologies from cellphones to alternative fuel vehicles and entire fleets (Dargay et al., 2007). The second shows exponential growth in *total passenger distance travelled*, a form also used in literature (Schäfer, 2006); within this, additional parameters track the mode share of passenger vehicles.

Each functional form has a distinct set of parameters to be iteratively updated by the Gibbs method, a subtype of MCMC which relies on conditional probabilities among the parameters. Appropriate distributions are chosen for candidate parameter values, and domestic Chinese data from the NBSC is employed as a basis for calculating historical residuals as accept-reject criteria¹ and candidate distributions for sampling. Once constructed, the models are calibrated with regard to *burn-in* (the number of iterations required until the model output becomes stationary) and transition stride sizes (the amount of change between parameter values in subsequent samples) in order to ensure the usefulness of the outputs.

2.1 Markov-chain Monte Carlo using Gibbs sampling

A *Markov chain* is a system which undergoes transitions between different states $\mathbf{x} \in \mathbf{X}$ with transition probabilities $q(x_{s+1} | x_s)$.² *Markov chain Monte-Carlo* is a methodology for sampling from complex probability distributions $f(\mathbf{x})$ for which probability density functions are not directly available. Instead, a Markov chain is constructed so that, after a period of initial *burn-in*, successive states approximate a stationary distribution $\pi(\mathbf{x}) = f(\mathbf{x})$ and may be treated as quasi-independent *samples* from $f(\mathbf{x})$. In the present application, the continuous state space \mathbf{X} consists of parameter vectors for transport growth models and each MCMC sample is an equally-likely set of parameter values for the given model, which allows statistical examination of $f(\mathbf{x})$, the distribution of the model outputs.

The most general expression of the methodology is the Metropolis-Hastings algorithm, which involves four looping over steps:

1. Begin with a state \mathbf{x}_s .

¹The general form of which is the residual of back-cast model values versus historical data. The probability of accepting a new parameter value is inversely related to the residual.

²In the following descriptions, t is always a year, and s is the current sampler iteration.

2. From some *proposal distribution* $q(\mathbf{y} | \mathbf{x}_t)$, sample a candidate \mathbf{y} .
3. Compute $\alpha(\mathbf{x}, \mathbf{y}) = \min\left(1, \frac{\pi(\mathbf{y})q(\mathbf{x}|\mathbf{y})}{\pi(\mathbf{x})q(\mathbf{y}|\mathbf{x})}\right)$
4. With probability α , accept \mathbf{y} as \mathbf{x}_{s+1} ; otherwise $\mathbf{x}_{s+1} = \mathbf{x}_s$.

Here two variants of the Metropolis-Hastings algorithm are used, both based on the subtype called Gibbs sampling. In a Gibbs sampler, the proposal distribution for each state variable (parameter) is a distribution that is conditional on the current values of all other state variables. Each variable is updated in sequence, α is 1 (*i.e.* all samples are accepted) and new samples are used immediately to condition the distributions for the remaining state variables. A new sample is produced once each state variable has been updated.

For vehicle ownership, the Gibbs update procedure is used with a random-walk Metropolis-Hastings proposal distribution $q(\mathbf{y} | \mathbf{x}) = q(|\mathbf{x} - \mathbf{y}|)$ for some parameters, based on the difference between the current and proposed parameter value. As well, *simulated annealing* is used to allow the sampler to proceed quickly through the state space during the burn-in period, yet remain in the likely region once it is reached.

The algorithm trades the need to know the probability density of $f(\mathbf{x})$ for some skill required in implementation. In particular, the chain must be constructed to be *aperiodic* (states must not repeat in a fixed pattern) and *positive recurrent* (the expected return time to any state must be finite).

2.2 Models

2.2.1 Vehicle Ownership

Dargay et al. (2007) used an S-shaped Gompertz curve to forecast long-run effects of vehicle ownership, and included a detailed formulation for the saturation level to facilitate comparison of the effects of urbanization, population density, and asymmetry in growth pathways. Here a simpler model is used, with the form:

$$\hat{V}(t, \mathbf{x}) = P(t) \cdot \hat{v}(t) = p(t) \cdot \alpha \exp(\beta \exp(\gamma \cdot t)) \quad (1)$$

Where:

$P(t)$ = population, historical for $t < 2010$, else projected

$\hat{v}(t)$ = per-capita vehicle ownership

$x_0 = \alpha$ = saturation level of ownership

$x_1 = \beta$ = t -wards shift parameter (more negative \rightarrow further right) < 0

$x_2 = \gamma$ = growth rate parameter (more negative \rightarrow faster growth) < 0

The state variable is thus $\mathbf{x} = [\alpha \beta \gamma]$, with the latter two parameters constrained negative. As historical data are available for $V(t)$, $P(t)$ in the interval $t_0 = 1978 < t < t_1 = 2010$, the root-sum-of-squares residual between the back-cast stock $\hat{V}(t < t_1)$ and actual is defined:

$$R(\mathbf{x}) = \sqrt{\sum_{\tau=t_0}^{t_1} (V(\tau) - \hat{V}(\tau, \mathbf{x}))^2} \quad (2)$$

The first update in a Gibbs iteration is to α . The saturation ownership level is constrained to the range between the lowest (0.45 vehicles per capita) and highest (0.8) observed in industrialized countries with peak or post-peak ownership. A discrete probability mass function for $\alpha \in (0.45, 0.8)$ is constructed by calculating $R([\alpha \beta_{s-1} \gamma_{s-1}])$ across this range, subtracting from $\max_{\alpha}(R)$ and then normalizing so that the mass is lowest where the residual is greatest. To accomplish simulated annealing, the sampled α is mixed with the existing value by weighting it with the reciprocal of the iteration count to produce α_s .

Next, β and γ are updated in sequence, using the same process for each. An ‘optimal’ β^* (or γ^* , etc.) which minimizes $R([\alpha_s \beta \gamma_{s-1}])$ conditional on the new α and old γ is calculated by the Nelder-Mead simplex algorithm.³ If β^* is positive, the optimum (although not a sample) is rejected and β_{t-1} kept. If negative, it used to construct a gamma distribution on $(0, \infty)$ for

³As implemented in SciPy, described in Jones et al. (2001–2012).

$-\beta$. The gamma distribution is used because it has the simple, two-parameter expectation $E[\Gamma(k, \theta)] = k\theta$ and is supported on the desired domain. The shape parameter k is chosen to produce a sharp peak in the probability density, and increased gradually with the number of iterations. The scale parameter is then set by:

$$\theta = \frac{E[-\beta]}{k} = \frac{-\beta^*}{k}$$

From the resulting distribution with an expectation of $-\beta^*$, a single sample is drawn and accepted as the next value β_s . The process is repeated for γ_s , and the three new values $[\alpha_s \beta_s \gamma_s]$ comprise the new sample and next state of the Markov chain. The effect of this procedure is to produce parameter values which are randomly distributed, but conditioned on the quality of the match with historical data.

2.2.2 Passenger travel volume

Schäfer et al. (2009) illustrated several important long-term trends in passenger mobility. With improving wealth and technological advances, higher-speed modes become more affordable, leading to a higher average travel speed within a fixed time budget of ~ 1.1 hours per day and a corresponding increase in passenger distance travelled per capita. When examined as percentages or shares of travel distance, this manifests as a declining share for low-speed modes including public transport; an increasing share for civil aviation and high speed railways, and a remainder for road transport that peaks and, in some industrialized nations, has begun to decline (Schäfer, 2006, Figure 4, page 28).

The model for passenger travel volume embodies these trends, by representing as a declining S-curve the share of passenger travel volume on railways (3); as an increasing S-curve the share for civil aviation (4), and a two-parameter exponential growth in PDT per capita (5):

$$\hat{d}_r = 1 - \alpha_r \exp(\beta_r \exp(\gamma_r \cdot t)) \quad (3)$$

$$\hat{d}_a = \alpha_a \exp(\beta_a \exp(\gamma_a \cdot t)) \quad (4)$$

$$\hat{d} = \alpha_d \beta_d^t \quad (5)$$

Neglecting the share for marine passenger transport, as explained below, the share for highway (road) travel is implicitly $\hat{d}_h = 1 - \hat{d}_a - \hat{d}_r$, so that a projection of total passenger distance travelled on roads (highways) is given by:

$$\begin{aligned} \hat{D}_h &= D \cdot \hat{d}_h = P(t) \cdot \hat{d} \cdot (1 - \hat{d}_a - \hat{d}_r) \\ &= P(t) \cdot \alpha_d \beta_d^t \cdot (\alpha_r \exp(\beta_r \exp(\gamma_r \cdot t)) - \alpha_a \exp(\beta_a \exp(\gamma_a \cdot t))) \end{aligned}$$

The state space for this model has eight dimensions, $\mathbf{x} = [\alpha_r \beta_r \gamma_r \alpha_a \beta_a \gamma_a \alpha_d \beta_d] = x_i, i \in 0..7$. To allow a more direct application of the Gibbs sampling methodology, conditional distributions for all parameters are constructed as follows. First, for each of (3), (4) and (5), a traditional least-squares fit is found.⁴ This yields not only the fitted values of each parameter—for instance, $\bar{\alpha}_r, \bar{\beta}_r,$ and $\bar{\gamma}_r$ for the share of railway travel—but also their covariances of the parameters, in the form (again for instance):

$$\Sigma_b = \begin{bmatrix} \sigma_\alpha^2 & \text{Cov}(\alpha, \beta) & \text{Cov}(\alpha, \gamma) \\ \text{Cov}(\beta, \alpha) & \sigma_\beta^2 & \text{Cov}(\beta, \gamma) \\ \text{Cov}(\gamma, \alpha) & \text{Cov}(\gamma, \beta) & \sigma_\gamma^2 \end{bmatrix}$$

Then assuming each of the model's parameters are jointly multivariate normal distributed with the fitted values as means, it is equivalent to combine the covariance matrices for each of the three sub-model equations to produce a block-diagonal form:

⁴Using the Levenberg-Marquadt algorithm as implemented in SciPy, described in Jones et al. (2001–2012).

2.3 Data sources

The NBSC's *China Statistical Yearbook*, 2011 edition, is the source for data series for the transport sector (National Bureau of Statistics of China, 2011). Tables include historical national figures, and in some cases the current year regional disaggregation. **Table 1** gives a summary

Table	Quantity	Units	Notation
16.07	Passenger travel volume, total		D
	Rail, total		D_r
	National, local, joint venture	10^8 km/year	
	Highways		D_h
	Waterways		D_w
	Civil aviation		D_a
16.24	Civil vehicles		
	Passenger vehicles, total		V
	Large, medium, small, minicar	10^4 units	
	Trucks, total		
	Heavy, medium, light, mini		
Others			
16.25	Private vehicles		
	Passenger vehicles, total		(unused)
	Large, medium, small, minicar	10^4 units	
	Trucks, total		
	Heavy, medium, light, mini		
Others			

Table 1: Data series from the China Statistical Yearbook 2011.

of the data available and notation for the series used below; capital symbols denote absolute values or national totals, while lower-case symbols denote fractional shares or per-capita levels.

One feature of Chinese transport which causes some challenges in the comparison of literature is government ownership of a large, though declining, fraction of the total stock—for instance, 48% of passenger vehicles in 2002 and 19% in 2010. Just as the classification of SUVs as light trucks in the U.S. brought a large portion of the latter category under the umbrella of light duty (passenger) vehicles, government-owned vehicles in China affect figures on ownership and mileage because their usage patterns differ from privately-owned vehicles. In this work, the total *civil* (including both government and private) passenger vehicle fleet is the subject of the forecast.

The passenger travel model used here discards the waterways or marine fraction D_w of total passenger distance traveled D . As shown in **Figure 1**, this was small (5% in 1980) and has since grown smaller in both absolute terms and share, making only one quarter of one percent of total travel distance in 2010.

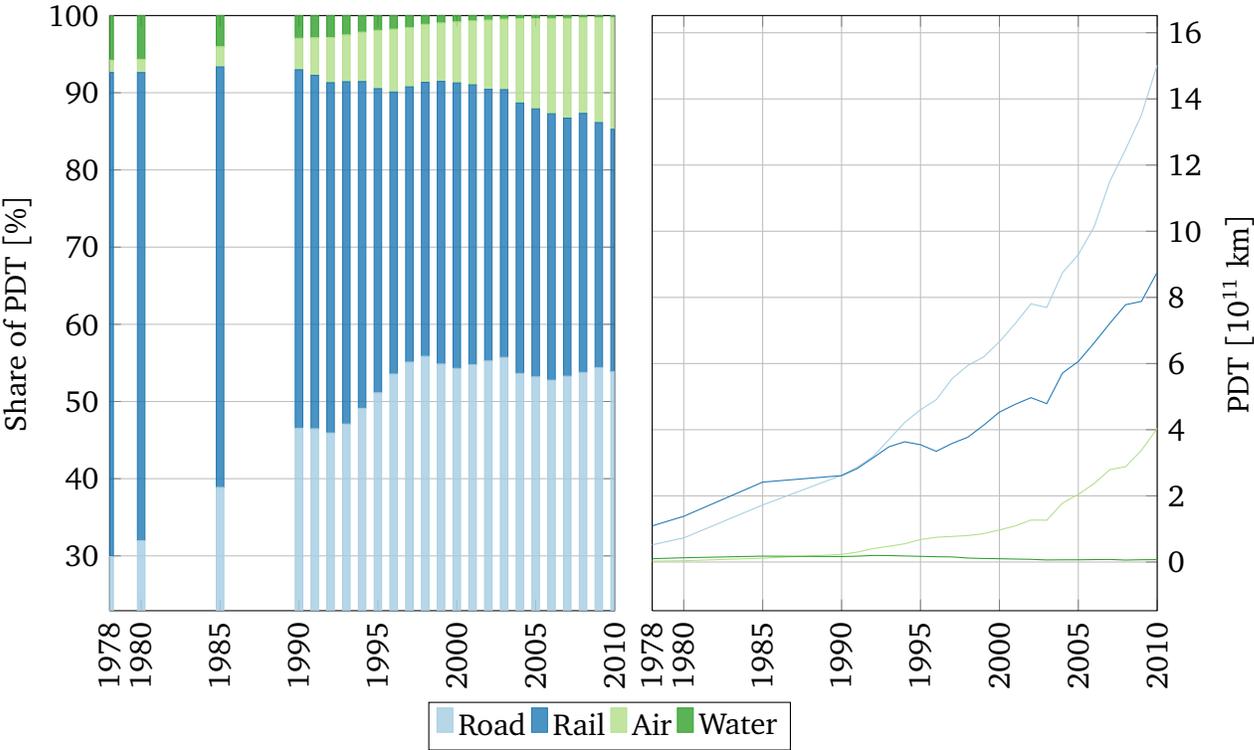


Figure 1: Mode share of passenger transport in China by distance travelled, 1978–2010 (National Bureau of Statistics of China, 2011).

The United Nations Population Division’s *World Population Prospects*, 2011 revision, gives estimated population to 2010 and projections to 2100 (UN Population Division, 2010). The median projections are themselves based on a Bayesian hierarchical model representing key factors including changing fertility across the demographic transition, mortality, and international migration (Alkema et al., 2011). However, the published data for the low-fertility and high-fertility variants represent fixed offsets, which do not vary across countries, from the median fertility, and not any range in the unpublished, per-country Bayesian outputs. Rather than attempt to reproduce the UN projections here, the medium-fertility variant is used directly.

3 Results

Before sample sets can be drawn from the models, it must be verified that they have ‘burned in’ to an ergodic state. **Figure 2** demonstrates the progress of the vehicle ownership sampler

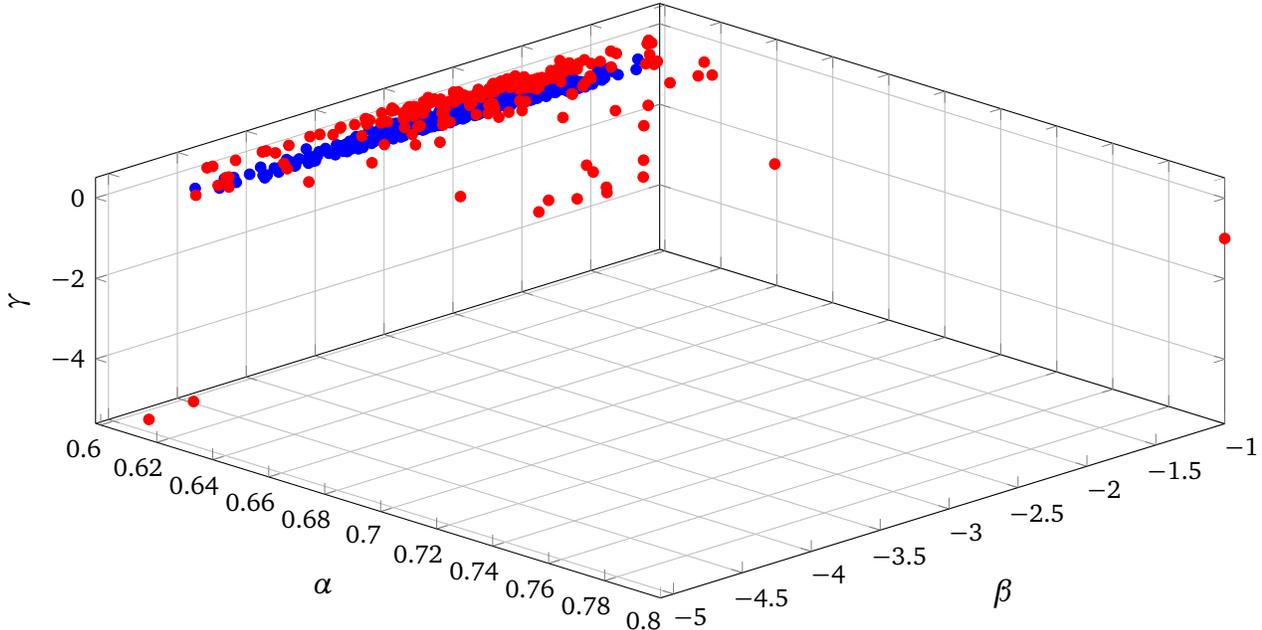


Figure 2: Sample chain in parameter space for 200 burn-in samples (red) and 400 usable samples (blue), vehicle ownership model.

from an arbitrary initial point $(\alpha, \beta, \gamma) = (1, -1, -1)$ to the likely region of the parameter space. A total of 600 samples are shown; the first two hundred, in red, are discarded as burn-in samples.⁵ The remaining 400, in blue, comprise the set, or *chain*, used for the following analysis. **Figure 3** shows a similar distinction made after 300 burn-in samples for the passenger travel volume model. Note that the distance model, having eight parameters, samples within an eight-dimensional space which is not easily visualized; the algorithm also progresses from initial points to a likely region in the other dimensions. Also, the assumption of a multivariate normal distribution for the travel volume (rather than the imposed conditioning, via residual, in the ownership model) produces a more neatly distributed chain in the sample space.

With proper operation of the algorithm confirmed, further samples can be drawn to ex-

⁵It should be noted that, having determined the posterior mean value for each parameter, it is straightforward to restart the algorithm from an initial point closer to—or precisely at—the mean, reducing the number of samples to be discarded for burn-in. To better illustrate the methodology, points further afield are used in both Figure 2 and Figure 3.

amine the distribution of future vehicle ownership and passenger-distance travelled in China. Through drawing a large number of samples, the extrema, centrality and expected values of the posterior probability distribution in each indicator—as well in parameter space, when the parameter values have economic meaning—may be examined. Below, the results of the two models are examined in turn.

3.1 Vehicle ownership

Quantity	$E[\bullet] \pm \sigma$	10%	90%
Ownership [$\text{vehicles}/\text{capita}$]			
2030	0.279 ± 0.188	0.0469	0.566
2050	0.402 ± 0.193	0.0963	0.610
Stock [10^6vehicles]			
2030	389 ± 261	65.3	797
2050	521 ± 249	125	790
Maximum < 2050	”	”	819
Year of peak[0]	>2050	>2050	2039
Parameters	Median		
Saturation ownership $\alpha [=] \text{v}/\text{cap.}$	0.614	0.610	0.616
Time shift $\beta [=] 0$	-2.70	-3.61	-2.04
Growth rate $\gamma [=] 0$	-0.0542	-0.160	-0.0171

Table 2: Parameter values and posterior distribution characteristics for vehicle ownership model. The columns contain expectations, sample standard deviations, first and ninth deciles respectively.

Table 2 and **Figures 4** and **5** give results from the vehicle ownership model. In **Table 2**, note that the saturation ownership level varies little across the distribution, but there is a wide range in projected ownership in 2050. The difference lies in the larger uncertainty in the growth rate, the response of the sampler to the recent, rapid growth in ownership. Thus, only in the 9th-decile projection does a peak in auto ownership come before mid-century; after this point, steady, saturated ownership is overcome by the effect of a decline in the medium-fertility population forecast from a 2027 peak.

In **Figure 4** the contrast between the narrow distribution for saturation ownership and more uncertain growth rate is visible in the clustering of samples near $0.6 \text{ vehicles}/\text{capita}$ in 2050. Samples below the first decile reflect history up to ~ 2005 more so than recent growth, representing

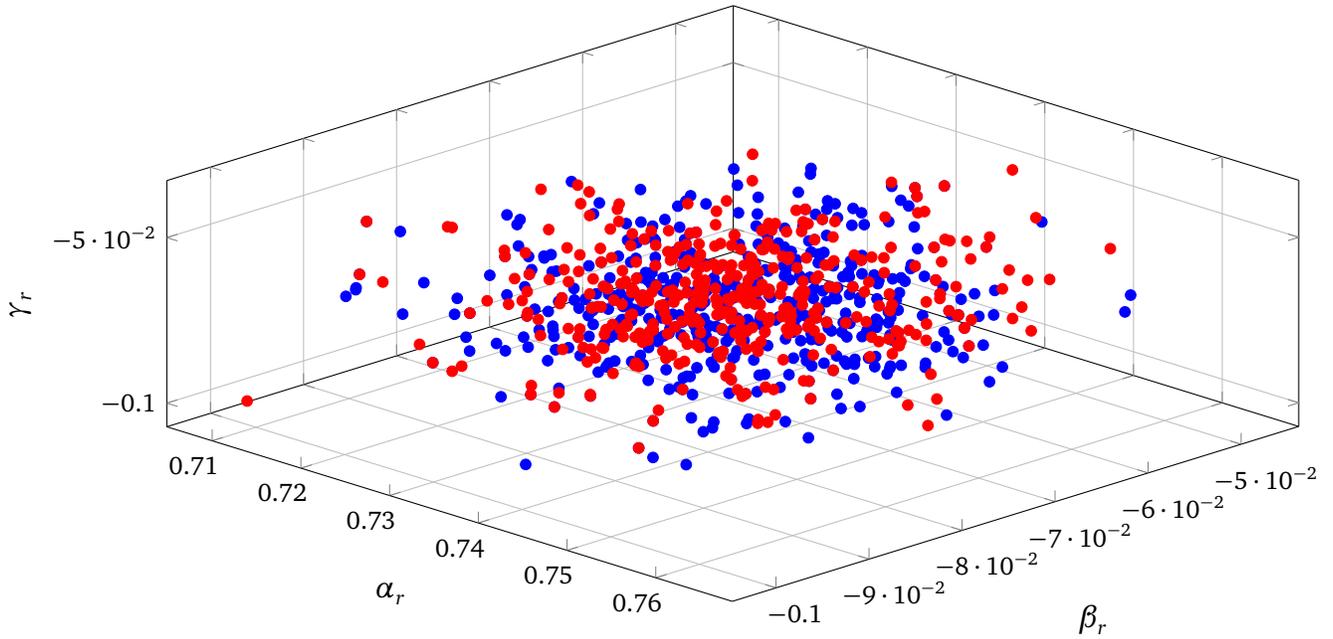


Figure 3: Sample chain in partial parameter space for 400 burn-in samples (red) and 400 usable samples (blue), passenger travel volume model.

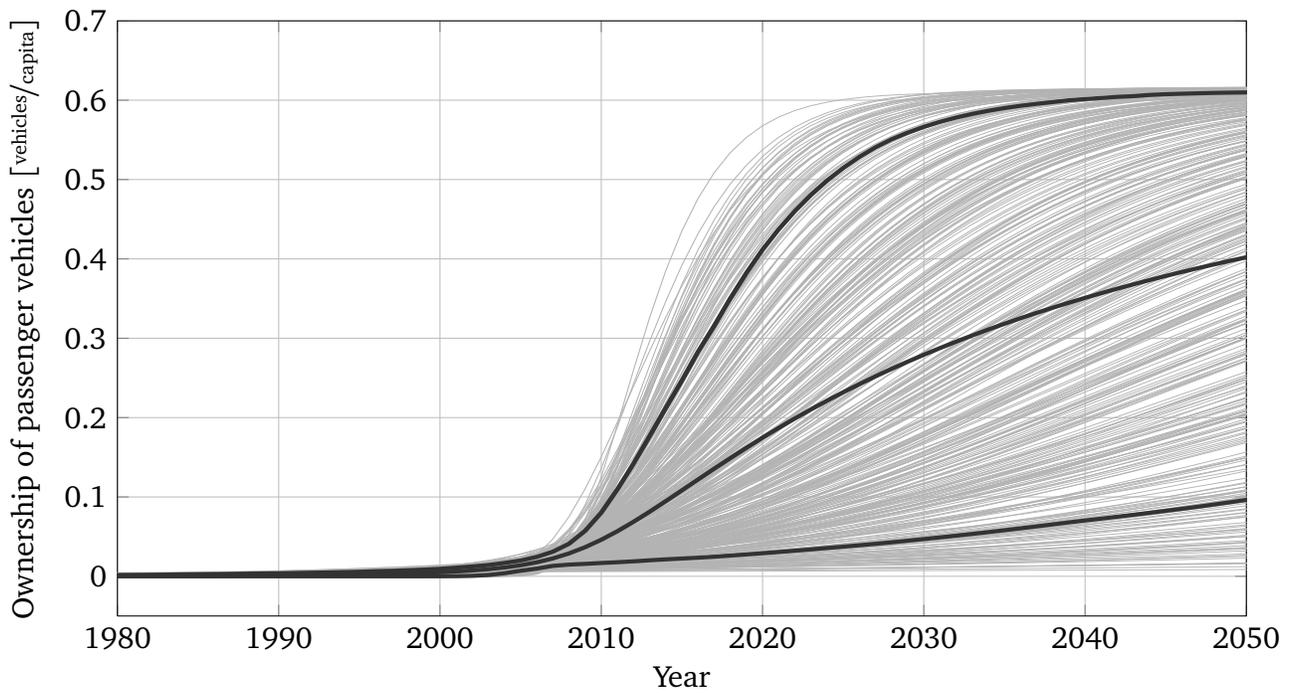


Figure 4: Four hundred projections of passenger vehicle ownership, with 1st decile, mean and 9th decile (black lines).

a possibility that growth is transient and will soon end. In contrast, **Figure 5** makes the effect of the underlying population trend on actual stocks clearly visible.

As stated, this model is forecasting civil passenger vehicles, rather than all vehicles with four or more wheels as in Dargay et al. (2007). However, it is notable that the saturation value of the Gompertz ownership is much lower than that study's long-run maximum of about 0.8,⁶ and while the total vehicle stock of 390 million in 2030⁷ matches the 389 ± 261 million measured from the posterior distribution here, the sample standard deviation indicates that the agreement is less consequential than the very large range across forecasts.

3.2 Passenger travel volume

Quantity	E [•] ± σ	10%	90%
Passenger distance travelled [10 ¹² km]			
2030			
Total	15.0 ± 10.4	5.58	27.7
Highways	10.0 ± 6.93	3.74	18.4
2050			
Total	87.5 ± 117	13.6	193
Highways	58.1 ± 77.8	9.07	128
Share of highways = $1 - d_a - d_r$			
2030	0.667 ± 0.0146	0.686	0.649
2050	0.663 ± 0.0165	0.685	0.642
PDT per capita [km/capita]			
2030	10,799 ± 7,437	4,000	19,868
2050	67,502 ± 90,150	10,530	149,323
Parameters	Median		
PDT annual growth rate = $1 - \beta_d$	7.60%	4.94%	10.6%
Saturation share of civil aviation α_a	0.0857	0.0603	0.107
Saturation share of railways = $1 - \alpha_r$	0.264	0.250	0.278
Saturation share of highways = $1 - \alpha_a - (1 - \alpha_r)$	0.653	0.624	0.676

Table 3: Parameter values and posterior distribution characteristics for passenger travel volume model.

Table 3 and **Figures 6** and **7** give results from the passenger travel volume model. In Table 3 only a subset of parameters with direct economic significance are displayed. The envelope of total passenger travel volume grows at less than 11% per year in the 9th decile,

⁶Dargay et al. (2007), Figure 8, page 17.

⁷*idem.*, Table 3, page 20.

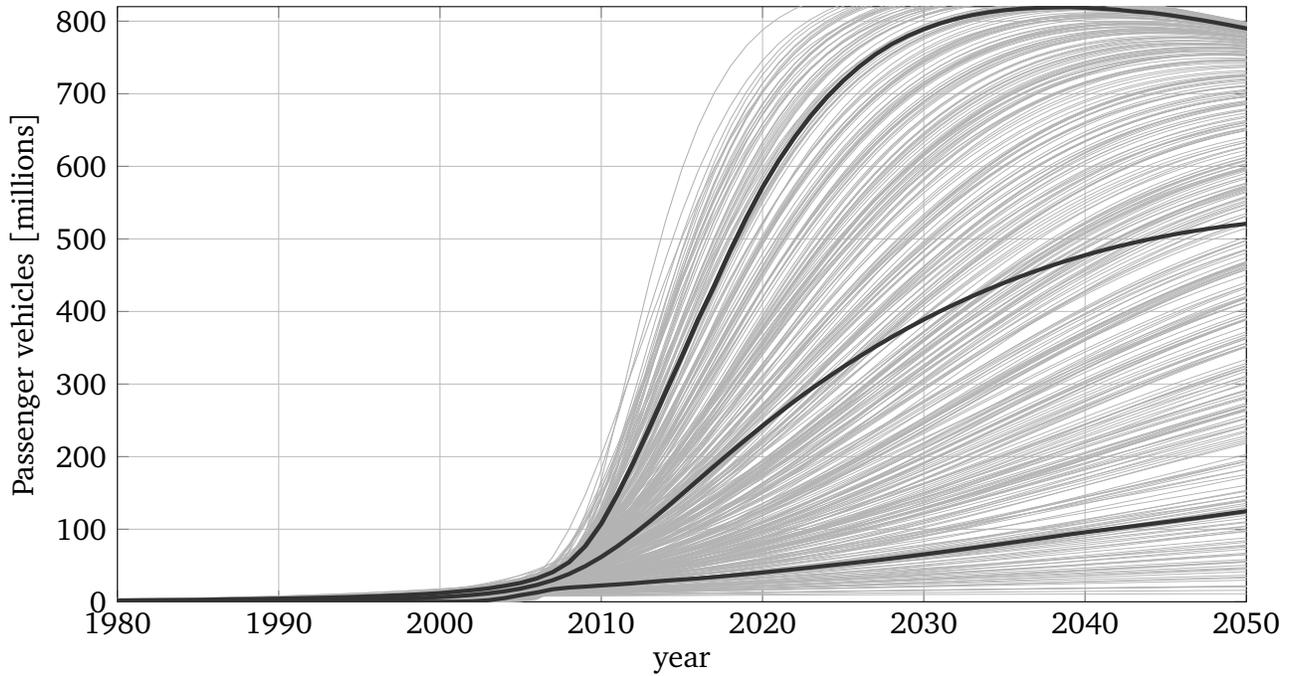


Figure 5: Four hundred projections of total vehicle stock under the medium-fertility population forecast, with 1st decile, mean and 9th decile (black lines).

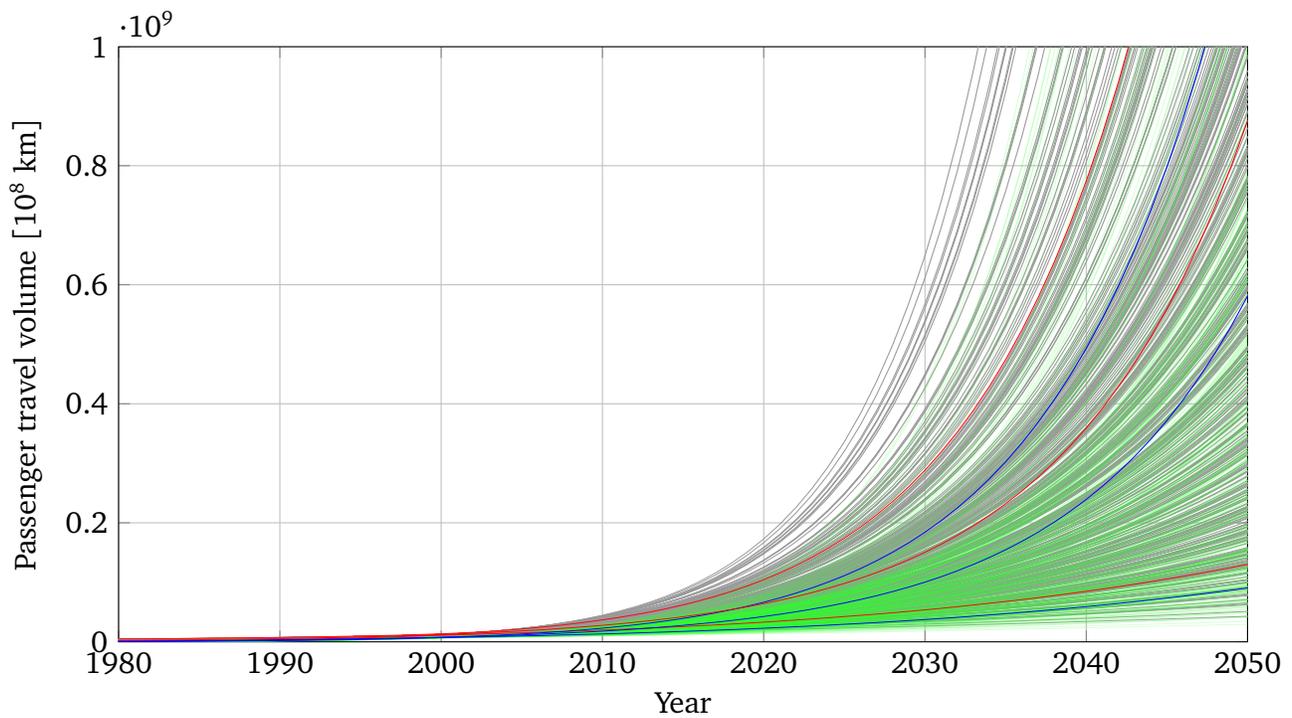


Figure 6: Four hundred projections of total (grey) and highway (green) passenger distance travelled, with 1st decile, mean and 9th decile for highways (blue) and the total (red).

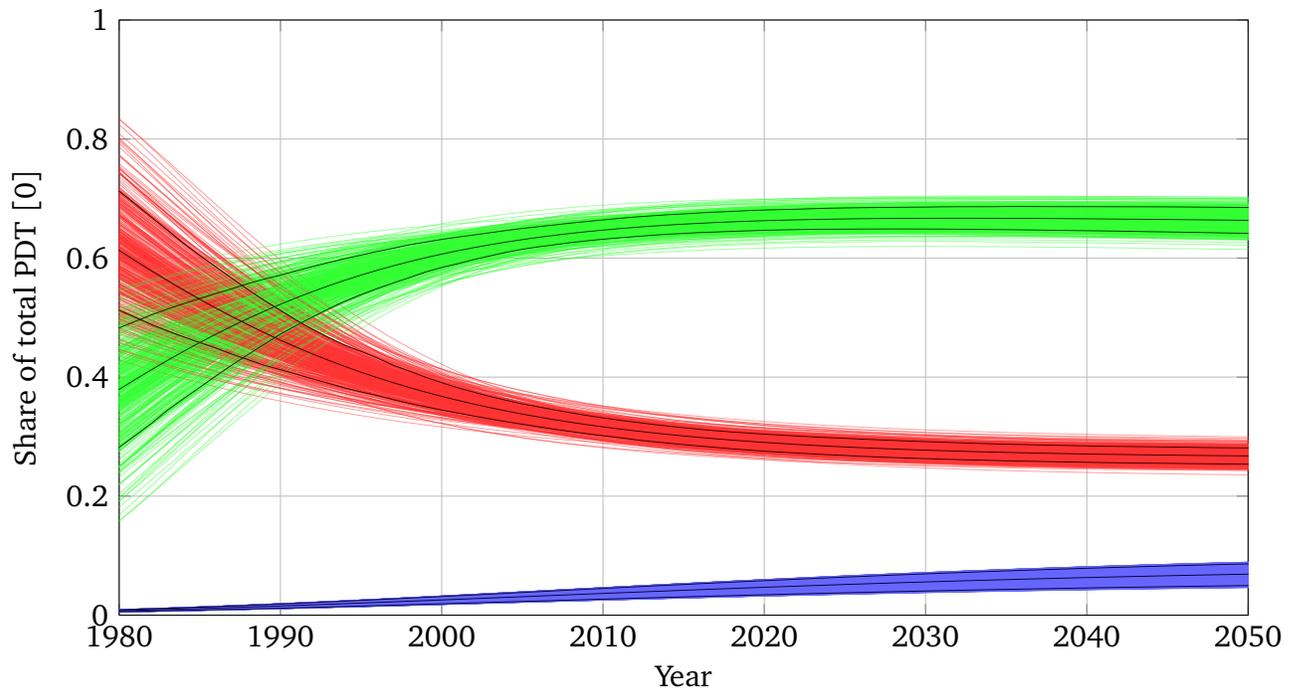


Figure 7: Four hundred projections of travel shares by mode: highway (green), railway (red) and civil aviation (blue), with 1st decile, mean and 9th decile for each (black lines).

or about 7.6% on average. Notwithstanding an increasing share for highway transport and a reported decrease in per-vehicle mileage due to the government-to-private fleet shift, this range of growth rates does not overlap significantly with auto sales growth of 8–17% year-on-year reported in the literature.⁸ However, the increase is still very rapid, with the median representing a doubling time of just under a decade.

In Figure 6, forecasts for passenger vehicle PDT are overlaid on totals. Due to the model form and a small number of very high forecasts, the standard deviations in 2050 for both highway and total travel are larger than the distribution expectations.

While theory suggests a shift from low-speed public transport to passenger vehicles and then aviation with growing income, the time-based sampler does not produce such an outcome within the period of the forecasts. Instead, Figure 7 shows stabilization in the shares for three modes, with civil aviation growing steadily, yet from a low level that prevents it from taking more than 10% of PDT. Rail retains at least 25% of travel in all forecasts, a level higher than other world regions including Western Europe, while the share for passenger vehicles does

⁸Wang et al. (2011), Tables 7–8, page 3302; Table 10, page 3304.

not nearly approach that of North America. The projections emphasize that China will have a transport future, though accelerated, that takes it along a different path and to a different endpoint than countries and regions which motorized in the 20th century.

4 Discussion

Information on the likely distribution of outcomes can aid in the design of robust transportation policy, which will be important in China and other industrializing or developing countries. The successful planning and management of large transport infrastructure projects relies on forecasts of demand. Where there is significant uncertainty, a characterization of the range of future possibilities can allow designers to not only build resilience into systems, but increase project value by preparing to take advantage of changes, for instance by confidently deferring expansions until uncertain outcomes are realized (de Neufville and Scholtes, 2011). Analyses like the present one can contribute to such understanding.

The case of the vehicle ownership saturation levels is also interesting because a high growth rate may accelerate Chinese domestic demand for transport fuels, exacerbate energy security concerns and perhaps also create a price signal which deters ownership, whereas slow growth could allow foreign consumption to press ahead with similar effects. In either situation—by direct triggering, or by allowing events elsewhere to intervene—ownership patterns could be disrupted, with the result that the saturation ownership forecast here is never realized.

4.1 Extensions & integration with CGE models

Although prices signals are not present in the models used here, other forms used in transport economics do contain prices explicitly, and would allow the direct representation of price effects. For example, a Stone-Geary utility function with sampled parameters could be employed to forecast transport demand while incorporating passenger responses to prices, as done econometrically in Meyer et al. (2007).

As mentioned above, work contributing to the U.N. Population Division forecasts fertility

using the phenomenon of the demographic transition that occurs with increasing wealth—a factor which is also separately correlated to transport choices. Information on this relationship would enable a fuller incorporation of uncertainty via a probabilistic $P(t)$ that was jointly distributed with or conditional on other model parameters.

China has a geographically heterogenous energy system; this includes transport energy use. Large subway and bus networks, along with high wealth, and the presence of government in the large, directly-controlled municipalities create a transport environment very different from that in smaller cities and in the countryside, with different scope and prospects for growth and change. Modelling these regional differences in a way that incorporates uncertainty could produce higher quality forecasts.

One of the most promising potential applications of uncertainty methods is in support of computable general equilibrium (CGE) energy-economic modelling, such as the MIT Emissions Prediction and Policy Analysis (EPPA) model (Paltsev et al., 2005) or the China-in-Global- and China Regional Energy & Emissions Models (C-GEM and C-REM) under development by the Tsinghua-MIT China Energy and Climate Project. This class of large models contain detailed representation of transport demand, energy use & policy instruments, and in some instances distinguish between civil aviation and other modes (Winchester et al., 2011), or between powertrain types—internal combustion, plug-in hybrid-electric, battery-electric—in passenger vehicles (Karplus, 2011). Careful application of sampling could yield sets of parameters (elasticities, expenditure shares and preferences) to inform these models where only point estimates are currently used; in turn, CGE contributes tracking of dynamic effects (such as path dependency due to capital accumulation) not accomplished in the present application. With the use of modern computing hardware, replacing the single-equation models used here with full CGE models is entirely feasible even for thousands of samples, and the results can be highly salient to policymaking (Sokolov et al., 2009).

4.2 Conclusion

This paper demonstrated a novel application of existing methods for rigorous treatment of uncertainty to transport forecasting in China, a country whose greatly uncertain transport future will have global implications. In doing so, it also showed how such methods can be used to extend econometric and other models from the literature. Using prior information about vehicle stock and passenger travel evolution, a range of future forecasts is developed which are conditioned, but not dependent, on observed reality. The work advances the important task of better characterizing complex systems such as global transport in order to inform their future design, management and policy.

References

- Leontine Alkema, Adrian E Raftery, Patrick Gerland, Samuel J Clark, François Pelletier, Thomas Buettner, and Gerhard K Heilig. Probabilistic projections of the total fertility rate for all countries. *Demography*, 48(3):815–39, 2011. doi:10.1007/s13524-011-0040-5.
- Joyce Dargay, Dermot Gately, and Martin Sommer. Vehicle ownership and income growth, worldwide: 1960-2030. *The Energy Journal*, 28(4):143–170, 2007.
- Richard de Neufville and Stefan Scholtes. *Flexibility in Engineering Design*. Engineering Systems. The MIT Press, Cambridge, Massachusetts, 2011.
- Dabo Guan, Zhu Liu, Yong Geng, Sören Lindner, and Klaus Hubacek. The gigatonne gap in China’s carbon dioxide inventories. *Nature Climate Change*, 2012. doi:10.1038/nclimate1560.
- Han Hao, Hewu Wang, and Minggao Ouyang. Comparison of policies on vehicle ownership and use between Beijing and Shanghai and their impacts on fuel consumption by passenger vehicles. *Energy Policy*, 39(2):1016–1021, 2011. doi:10.1016/j.enpol.2010.11.039.
- Hong Huo, Qiang Zhang, Kebin He, Zhiliang Yao, and Michael Wang. Vehicle-use intensity in China: Current status and future trend. *Energy Policy (in print)*, 2011. doi:10.1016/j.enpol.2011.09.019.
- Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, documentation, 2001–2012.
- Valerie Jean Karplus. *Climate and Energy Policy for U.S. Passenger Vehicles: A Technology-Rich Economic Modeling and Policy Analysis*. PhD thesis, Massachusetts Institute of Technology, Cambridge, USA, 2 2011.
- Paul N. Kishimoto, Sergey Paltsev, and Valerie J. Karplus. The future energy and GHG emissions impact of alternative personal transportation pathways in China. Kyoto, Japan, 02 2012.
- Peter J. Marcotullio, E. Williams, and Julian D. Marshall. Faster, Sooner, and More Simultaneously: How Recent Road and Air Transportation CO₂ Emission Trends in Developing Countries Differ From Historic Trends in the United States. *The Journal of Environment & Development*, 14(1):125–148, 2005. doi:10.1177/1070496504273716.
- I. Meyer, M. Leimbach, and C.C. Jaeger. International passenger transport and climate change: A sector analysis in car demand and associated emissions from 2000 to 2050. *Energy Policy*, 35(12):6332–6345, 2007. doi:10.1016/j.enpol.2007.07.025.
- Kyung-Min Nam and John Reilly. City-Size Distribution as a Function of Socioeconomic Conditions: An Eclectic Approach to Downscaling Global Population. Joint Program Report Series #213, MIT Joint Program on the Science & Policy of Global Change, Cambridge, MA, USA, March 2012.
- National Bureau of Statistics of China. China Statistical Yearbook 2011. Beijing, China, 2011.

- Sergey Paltsev, John Reilly, Henry Jacoby, R. Eckaus, J. McFarland, M. Sarofim, M. Asadoorian, and M. Babiker. The MIT Emissions Prediction and Policy Analysis (EPPA) Model: Version 4. Report 125, MIT Joint Program on the Science and Policy of Global Change, Cambridge, MA, 2005. URL http://globalchange.mit.edu/files/document/MITJPSPGC_Rpt125.pdf.
- R.Y. Rubenstein and D.P. Kroese. *Simulation and the Monte Carlo Method*, chapter 6. Markov Chain Monte Carlo, pages 167–200. John Wiley & Sons, Inc., Hoboken, NJ, USA., 2 edition, 2007.
- Andreas Schäfer. Long-term trends in global passenger mobility. *The Bridge*, 36(4), 2006.
- Andreas Schäfer, John B. Heywood, Henry D. Jacoby, and Ian A. Waitz. *Transportation in a Climate-Constrained World*. MIT Press, 2009.
- Jonathan E. Sinton. Accuracy and reliability of China's energy statistics. *China Economic Review*, 12(4):373–383, 2001. doi:10.1016/S1043-951X(01)00067-0.
- A. P. Sokolov, P. H. Stone, C. E. Forest, R. Prinn, M. C. Sarofim, M. Webster, S. Paltsev, C. A. Schlosser, D. Kicklighter, S. Dutkiewicz, J. Reilly, C. Wang, B. Felzer, J. M. Melillo, and H. D. Jacoby. Probabilistic Forecast for Twenty-First-Century Climate Based on Uncertainties in Emissions (Without Policy) and Climate Parameters. *Journal of Climate*, 22(19):5175, 2009. doi:10.1175/2009JCLI2863.1.
- UN Population Division. World Population Prospects, the 2010 Revision. Technical report, United Nations Department of Economic and Social Affairs, New York, NY, USA, 10 2010. Online: <http://esa.un.org/unpd/wpp/index.htm>.
- M. Wang, Hong Huo, L. Johnson, and D. He. Projection of Chinese Motor Vehicle Growth, Oil Demand and CO₂ Emissions Through 2050. Technical report, Argonne National Laboratory, Argonne, IL, USA, 12 2006. Report ANL/ESD/06-6.
- Yunshi Wang, Jacob Teter, and Daniel Sperling. China's soaring vehicle population: Even greater than forecasted? *Energy Policy*, 39(6):3296, 2011. doi:10.1016/j.enpol.2011.03.020.
- N. Winchester, C. Wollersheim, R. Clewlow, N.C. Jost, S. Paltsev, J. Reilly, and I.A. Waitz. The Impact of Climate Policy on U.S. Aviation. Report 198, MIT Joint Program on the Science and Policy of Global Change, Cambridge, MA, USA, 4 2011.
- Xinhua News Agency. "China's imported oil dependence continues to rise", May 10, 2012. 2012. URL http://news.xinhuanet.com/english/china/2012-05/10/c_131580559.htm.