NISS

Assessing Uncertainties in Traffic Simulation: A Key Component in Model Calibration and Validation

M.J. Bayarri, James O. Berger, German Molina,
Nagui M. Rouphail, Jerome Sacks

Technical Report Number 137
July, 2003

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Acknowledgment and Disclaimer

This material is based upon work supported by the National Science Foundation under Grant No. DMS-00739520

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

Assessing Uncertainties in Traffic Simulation: A Key Component in Model Calibration and Validation

M.J. Bayarri[1], James O. Berger[2], German Molina[2], Nagui M. Rouphail[3], Jerome Sacks[2,4]

[1]University of Valencia, [2]Duke University, [3]North Carolina State University, [4]National Institute of Statistical Sciences

Abstract

Calibrating and validating a traffic simulation model for use on a transportation network is a process that depends on field data that is often limited, but essential for determining inputs to the model and for assessing its reliability. A quantification and systemization of the calibration/validation process exposes statistical issues inherent in the use of such data. Our purpose is to elucidate these issues and describe a methodology to address them.

The formalization of the calibration/validation process leads naturally to the use of Bayesian methodology for assessing uncertainties in model predictions arising from a multiplicity of sources (randomness in the simulator, statistical variability in estimating and calibrating input parameters, inaccurate data and model discrepancy). We exhibit the methods and the approach on an urban street network, using the micro-simulator

CORSIM, while calibrating the demand and turning movement parameters. We also indicate how the process can be extended to deal with other model parameters as well as with the possible misspecification of the model. While the methods are described in a specific context they can be used generally, inhibited at times by computational burdens that must be overcome, often by developing approximations to the simulator.

Key words and phrases: Bayesian Analysis; Posterior Distribution; Stochastic Model Approximation; Traffic Simulation; Model Validation; Calibration; CORSIM

0. Introduction

Calibrating and validating traffic simulation models for use on a transportation network is inherently a complex process that is commonly treated informally and through a mix of ad hoc methodologies. Explicit ingredients of the process are field data that are often limited and expensive to acquire but essential for determining inputs to the simulation model and for assessing the reliability of the model. A quantification and systemization of the calibration/validation process exposes statistical issues inherent in the use of such data for assessing the validity of a model. Our purpose here is to elucidate these issues and describe methodology to address them.

A clear statement of what "validation" means is rarely set forward. Usually, the question is put as "does the model faithfully represent reality?" But, the answer to this question is simple: no, models are not perfect. But models can make useful, reliable predictions in particular settings; they may be useful for some purposes, useless for others. We can state this more formally as

$$\Pr\left[\left|\text{"reality"} - \text{simulation prediction}\right| < \delta\right] > \alpha ,$$

where we must specify δ = tolerable difference (how close) and α = level of assurance (how certain), say what is meant by "reality" and what is needed to make sense of Pr and how to compute the probabilities involved.

What we mean by "reality" is, operationally, a feasible measure of actual performance of a particular network. For example, it may be a system queue time measure in an urban traffic network under a current signal plan or, perhaps, under a proposed timing plan. To compare actual performance with simulation prediction will require access to field data and simulation output that relate to a performance measure. A review of the literature indicates that little attention has been paid to the characterization of the *uncertainty* in simulation model *inputs*. Rather the focus appears to be on the analysis of the stochastic outputs of the performance measures derived from various models (see for example Benekohal and Abu Lebdeh (1994); Tian et al. (2002)).

A general framework for treating validation is set forth in Bayarri et al. (2002). The framework calls for

- description of the inputs to the simulation model and knowledge about them
- a specification of performance measures (also called evaluation functions) to be used as described in the last paragraph
- data collection: both field data and simulator output
- analyses, including calibration that account for a multiplicity of uncertainties.

Accomplishing the last while making sense of (Pr) is the primary focus of this paper. In doing so we begin with examining more closely the nature of inputs to the model. Our comments and methods are general but we will freely utilize as a test bed, the simulation model, CORSIM, (Federal Highway administration, 1997) and its use on a traffic network in Chicago. The network is described succinctly in Section 1, especially Figure 1; a detailed account of the network, the data and their use in a signal study is in Sacks et al. (2002).

Inputs to simulators, such as CORSIM, come in various forms. Some, such as geometric inputs (lane widths, bus stops, etc.) are readily supplied by accurate measurement or through documented sources. Others can be obtained by calibration that is, through the use of field data. These latter can be classified as

(1) parameters that can be directly estimated, albeit with error, from field data (vehicle mix, arrival rates, turning percentages,...)

(2) parameters not directly measurable (e.g., degree of driver aggressiveness); possibly by choice (e.g., discharge headway distribution)

(3) tuning parameters (e.g., free flow speed, lost time) that are not "real" but are required by the model.

Some parameters in (2) and (3) may be set to default values and thus removed from the calibration process. The remaining parameters in (2) and (3) are typically treated by adjusting them until the model output seems to 'fit' the field data, data from the real traffic system being modeled. Sometimes the adjustment is informal; sometimes more formal by optimizing the parameters, using least squares or other optimization tools to minimize the difference between features of the field data and predictions from the simulator (see for example Jha et al., 2003, and Hourdakis and Michalopoulos, 2003).

There are two fundamental problems with such approaches to calibration. First, there may be an identifiability problem that allows the model to be tuned in many ways and masks possible large uncertainties in the tuned parameters with resulting inaccuracies in model predictions. Second is "over-tuning" whereby tuning hides model imperfections that may exist and results in models that are potentially highly inaccurate outside the range of the observed field data.

Bayesian analysis (described in Sections 2, 3) provides an attractive path to accomplish the calibration of the parameters of type (1) and, at the same time, the calibration of the parameters of type (2) and (3). Such analysis determines the posterior (or summary) distribution of model parameters and inputs, given the observed field data. The resulting distribution will then reflect the actual uncertainty in the parameters and inputs, and will be considerably more resistant to over-tuning. Moreover, the Bayesian approach can, at the same time, deal with the possible presence of model bias and hence be used for validation. In Bayarri et al. (2002) this is done in the context of specific deterministic simulation codes. In principle these methods can be used for simulation models such as CORSIM. But, there are complicating features arising from the stochasticity of CORSIM, the possibly large number of calibration parameters and, as in the test bed example below, a shortage of field data. However, we shall show how this process can be carried out with the calibration of parameters of type (1) and the resulting effect on model predictions (Sections 2 and 3). We will also describe a pathway to incorporating the calibration of type (2) and (3) parameters and to validation (Section 4).

1. The Test Bed Model, Network and Data

The simulator we use is CORSIM (US Federal Highway administration, 1997), release 4.3.2, a computer model of street and highway traffic. (Since the conduct of this research

FHWA has released two updates, 5.0 and 5.1). There is intrinsic randomness in CORSIM resulting from vehicles arriving at random, turning at intersections with prescribed probabilities and movements or processes governed by (other) specified probability distributions.

The traffic network we use is a 29-intersection neighborhood in Chicago schematically depicted in Figure 1. The network is a principal area connecting an important freeway (at the west end, U and 21 in Figure 1) and major arterials (running north/south through the network) with the central business district (east and south of the network).

[FIGURE 1 GOES HERE]

Running CORSIM requires a large number of inputs, of which there are three basic kinds: *fixed inputs*, such as the network geometry, bus routes, speed limits, etc.; *controllable inputs* such as the cycle length, phasing and traffic signals' offsets; and unknown *random* inputs, such as the distributions of bus inter-arrival and dwell times, distributions of the proportion of different types of vehicles (trucks, cars, pedestrians), distributions of different types of driver behaviors, turning, and traffic demand distributions.

We focus on the two most significant (unknown) inputs to CORSIM: (i) Demand and (ii) Turning probabilities. *Demand, D*, consists of parameters that determine the numbers of vehicles that enter the system from external streets, while the *turning probabilities, P,* refer to the probabilities that a vehicle turns right, turns left, or goes through a given intersection. Demand and turning probabilities are street and intersection specific, so that

D is actually a vector of 16 numbers (for the studied system), while P is an 84-dimensional vector of probabilities. These will be determined from observational data, consisting of counts, C, made on the real-world traffic network.

1.1 The data and their adjustment

The data, C, is a vector of counts of vehicles. Let C_{ijk} denote the count of vehicles arriving at intersection j from intersection i, and proceeding to intersection k. Thus, referring to Figure 1, C_{673} is the count of vehicles arriving at intersection 7 from the North, and turning left. It is also convenient to define C_{ij} as the total observed number of vehicles coming from intersection i towards intersection j. The counts fall into three classes of data:

Demand counts: These are counts, made over a one-hour period, by observers placed on the streets entering the traffic neighborhood in Figure 1 (such as I,K,M), and correspond to certain of the C_{ij} above. Manual observation is highly imperfect and these counts are suspected to be inaccurate. We will denote the vector of all the observed demand counts by C_D.

Turning counts: These are counts, made by observers over shorter time intervals (typically 10 to 20 minutes), of the numbers of right-turn, left-turn, and through vehicles at each intersection. Some of the counts are missing and all are subject to error. Denote the vector of all observed turning counts by C_T.

Video counts: At the intersections in Figure 1 that lie within the central dashed rectangle, video cameras perched atop nearby buildings were placed that recorded all vehicles passing through the intersection over the one-hour period. The recordings were later analyzed to determine the numbers of right-turn, left-turn, and through vehicles from each direction, at each of the intersections. These counts can be treated as exact.

The demand and manual turning counts, because of human error, may not be consistent with the video counts. It will be necessary to take this into account in subsequent analyses. At first, we will employ a commonly used tactic: adjust, by hand, a few counts until the observer counts are compatible with each other and with the video counts. Take these as accurate and proceed from there (we do so in the next Section). In Section 3, however, we recognize the observer error and treat matters there accordingly and, as we shall see, with different results.

2. Analysis When Vehicle Counts Are Assumed to Be Accurate

In this section we assume that <u>all</u> vehicle counts are accurate, as if the entire network had video (or accurate sensor) data. This section will serve to contrast the effect of <u>not</u> accounting for the uncertainty in estimating the CORSIM input parameters (even with wholly accurate data the demand and turning parameters cannot be determined exactly) as well as contrasting (in Section 3) with the additional accounting for observer error. At the same time, this section provides a simplified, useful context in which to explain the Bayesian analysis and how it handles uncertainties.

The basics of Bayesian analysis consist of combining, via Bayes theorem, the data density, given the demand parameters D and turning probabilities P, with a prior distribution for D and P. The resulting posterior distribution in (2.2) or (2.3), denoted by $\pi(D, P|C)$, is simply the conditional distribution of the parameters D and P, given the data. This posterior distribution indicates the uncertainty that is present about D and P after the data C are collected.

Uncertainty resulting from simulator predictions can then be assessed by treating $\pi(D,P|C)$ as the ``random input distribution" for the simulator, and making repeated runs of the simulator, initialized by draws from this distribution. This is not significantly more expensive than running a stochastic simulator, such as CORSIM, for a fixed set of inputs. For a stochastic simulator, the distribution of predictions can only be ascertained through repeated runs, and starting each run with D and P chosen from $\pi(D,P|C)$ is virtually as cheap as starting each run with a fixed D and P.

2.1 Bayesian Analysis of Turning Probabilities

We first concentrate on the vector of turning probabilities $\mathbf{p}$ for a single intersection. For example, the vehicles entering intersection 2 from K, can either turn left or go through the intersection. Let p_L and p_T be the respective probabilities. Since $p_T = 1 - p_L$ it suffices to consider one of, these, say $p = p_L$ (dropping the subscript L for convenience). Let $c = \{ c_L , c_T \}$ denote (again only in this subsection) the corresponding manual counts, that is, the observed number of cars going left and through at intersection 2 when arriving from *K*. If, as in CORSIM, the cars turn independently of one another and of their previous path, then, letting $n = c_L + c_T$ denote the total number of vehicles entering intersection 2 from *K,* the distribution of *c,* given *n,* is given by $c_L \sim Binomial(n,p),$ which has density

$$f(c_L \mid p, n) = \binom{n}{c_L} p^{c_L} (1-p)^{n-c_L} . \qquad (2.1)$$

The conventional way of dealing with the unknown p is to estimate it by $\hat{p} = c_L / n$, the maximum likelihood estimate (MLE), and use this estimated turning probability in CORSIM, ignoring the uncertainty in the estimate.

Bayesian analysis requires a prior distribution, $\pi(p)$ for p. This can reflect genuine prior information (e.g.,, earlier traffic counts) if available, or it can be chosen in an objective fashion. The traditional objective prior for the Binomial model is a special case of a Beta distribution (for more details on the Bayesian nomenclature and technical terms used in this paper, see Bernardo and Smith, 2001):

$$\pi(p) = Beta(p \mid 0.5, 0.5) = \frac{1}{3.14...} p^{-0.5} (1-p)^{-0.5}$$

(It might seem more natural to use the uniform distribution as the objective prior for p, but the above prior, called the Jeffreys prior, has superior properties.) Combining the prior with the likelihood via Bayes theorem produces the posterior distribution

$$\pi(p \mid c_L, n) = \frac{\pi(p) f(c_L \mid p, n)}{\int_0^1 \pi(p) f(c_L \mid p, n) dp} = \frac{\Gamma(n+1)}{\Gamma(c_L + 0.5)\Gamma(n - c_L + 0.5)} p^{c_L - 0.5}(1-p)^{n - c_L - 0.5}, \quad (2.2)$$

where $\Gamma(\cdot)$ is the gamma function. This posterior density can be recognized as the $Beta(p \mid c_L + 0.5, n - c_L + 0.5)$.

The posterior in (2.2) takes into account all the uncertainty about p. The mean of (2.2) is $[c_L + 0.5] / (n+1)$ which for moderate n and c_L will be very close to $\hat{p}$; indeed, when n is large, (2.2) will be tightly concentrated about $\hat{p}$ (the variance of the posterior distribution is $(c_L + 0.5)(n - c_L + 0.5)(n+1)^{-2}(n+2)^{-1}$. However, when n is not large, (2.2) will be more dispersed and Bayesian predictions will reflect this uncertainty; the traditional approach won't. To propagate this uncertainty to simulator predictions CORSIM is fed with independent generations from $\pi(p \mid c_L, n)$, instead of the fixed value $\hat{p} = c_L / n$.

Example 2.1 For the intersection mentioned above (vehicles arriving at 2 from K) in the period 09:00–10:00am, a total of $n = 375$ cars were observed, $c_L = 50$ of which went left at the intersection. The posterior density in (2.2) of p, the probability of turning left, is then $Beta(p \mid 50.5, 375.5)$. $\hat{p} = 0.118$, the posterior also has mean 0.118 and its standard deviation is 0.0156.

If 10 runs of CORSIM are required, the traditional approach will use $\hat{p} = 0.118$ as the value for p in each of the runs, with variability coming only from the stochasticity in CORSIM, but not in the inputs. The Bayesian approach will use 10 independent values generated from the posterior distribution, for the 10 runs; for instance, random generation from the posterior gave $\{0.142, 0.152, 0.118, 0.131, 0.109, 0.109, 0.126, 0.129, 0.101, 0.153\}$.

For many intersections, vehicles entering from a specific direction can either turn right, turn left, or go through. In this case, the vector **p** has three components, $\mathbf{p} = (p_R, p_L, p_T)$, and, given the total number, n, of vehicles entering from the specific direction, the respective counts, (c_R, c_L, c_T), follow a multinomial distribution. The standard objective prior for **p** is a *Dirichlet* (1/2, 1/2, 1/2) distribution, a multidimensional generalization of the *Beta* distribution.

If, as in CORSIM, all turning movements of vehicles are treated independently, the analysis for all turning probabilities, P, simply proceeds by performing the analysis given above for each intersection, and then multiplying together (because of the independence) the resulting posterior distributions. Recalling that C_T denotes all turning data, this product is the posterior distribution, $\pi_T(p|\, C_T)$ of all turning probabilities, given the data.

2.2 Bayesian Analysis of Entry Demand Parameters

First consider the demand at a single input location of the system for instance, the number of cars entering the system at E and going towards intersection 20. Let λ be the rate at which cars enter per hour and let c be the observed count of cars entering in a given one-hour period. Assume the inter-arrival times of such vehicles have an exponential distribution, so that probability density of c is Poisson:

$$f(c \mid \lambda) = \frac{1}{c!}\lambda^c e^{-\lambda}.$$

The objective prior density for λ in this model is $\pi(\lambda) = \lambda^{-0.5}$. (Formally, this is called the Jeffreys prior. It is not a proper density i.e., does not integrate to 1, but its use has strong justification, Bernardo and Smith, 2001.) Formal application of Bayes theorem produces the posterior density for λ:

$$\pi(\lambda \mid c) = \frac{\lambda^c e^{-\lambda}\lambda^{-0.5}}{\int_0^\infty \lambda^c e^{-\lambda}\lambda^{-0.5} d\lambda} = \frac{1}{\Gamma(c+0.5)}\lambda^{c-0.5}e^{-\lambda}, \qquad (2.3)$$

recognizable as the *Gamma*$(\lambda \mid c + 0.5, 1)$ density. The mean $(c + 0.5)$ and mode $(c - 0.5)$ of (2.2) are very close to the MLE $\hat{\lambda} = c$. The variance is $(c+0.5)$, which can be quite appreciable, and the resulting uncertainty should be incorporated into runs of the simulator, by simulating demand rates using (2.3), as was done for turning probabilities, using (2.2).

Example 2.2 For the demand discussed above (the number of cars entering the system at E and going towards intersection 20), the total of cars arriving in the period 09:00–10:00am was c =540. The resulting *Gamma*(540.5, 1) posterior density for λ has mean 540.5 and standard deviation $\sqrt{540.5}$. If 10 runs of the simulator are required, then 10 Bayesian inputs into CORSIM for λ, arising as random draws from this distribution, are

{516.41, 534.07, 557.40, 519.52, 506.03, 553.95, 544.05, 532.00, 551.43, 541.51}. The traditional approach will, instead, always input $\hat{\lambda} = 540$ as the value for λ.

For the full system, let λ denote all the vehicle inter-arrival rates $\{\lambda_{ij}\}$ at the entry points of the network. (We earlier denoted the counts by D). If all demands are assumed to be independent, analysis of the full system simply proceeds by replicating the analysis at each demand input and multiplying together the resulting posterior distributions to get the posterior distribution, $\pi(\lambda| C_D)$, of all the inter-arrival rates. Later, we will also need to refer to $f_D(\lambda,C_D)$, the joint density of λ and C_D , which will simply be the product of all the Poisson densities for the $\{C_{ij}\}$ (given λ) times the product of all the prior densities $\pi(\lambda_{ij}) = \lambda_{ij}^{-0.5}$. We sample from $\pi(\lambda| C_D)$ to select λ's to feed CORSIM.

2.3 Bayesian Analysis of the Full System with Accurate Counts

The demands counts are not independent of the turning counts, but it can be shown (see Molina, et. al., 2003) that the posterior distribution of all turning probabilities P and inter-arrival rates λ can be found by multiplying together the posterior densities for P and λ. The joint posterior distribution of all unknowns is then

$$\pi(\quad, P \mid C_D, C_T) = \pi_D(\quad \mid C_D)\pi_T(P \mid C_T)$$

The effect of incorporating such uncertainty into CORSIM is indicated in Figure 2.

[FIGURE 2 GOES HERE]

The objective here is the prediction of system queue time (SQT). The bold curve (H1) is the histogram of SQT obtained from the output of 200 independent CORSIM runs with input parameters "plugged-in" as the MLE estimates of P and λ after informal tuning of the data to make the observed counts compatible. The thinner curve (H2) is also the result of 200 independent CORSIM runs but based on the Bayesian analysis just described where the P, λ inputs for each CORSIM run are randomly drawn from the posterior distribution $\pi(\quad, P \mid C_D, C_T)$. The uncertainty in the input parameters leads to a more widely spread SQT distribution, as expected. Increased variability could have strong implications when the calibrated model is used for prediction purposes. Note that the dashed curve results from the analysis (H3) are described in Section 3.

3. Methodology Addressing Errors in the Data

The simple analysis of the previous section is valid when there are no errors in the counts (and no missing counts), an unlikely situation when data are manually collected. Also, the power of the Bayesian methodology is not fully realized in the context of Section 2 – we could have used the distribution of the MLE's of P and λ, or an approximation to that distribution, to generate random inputs to feed to the simulator and achieve similar results. However, we need to introduce a more sophisticated analysis to handle observer error and the Bayesian approach adapts itself immediately to that end albeit with complications in computation but neither in interpretation nor principle. While the methodology is basically the same the introduction of models for observer error results in complexity that prevents writing simple closed form expressions for the needed posterior distributions. Consequently, posterior probabilities must be obtained by numerical methods or by simulation. The most commonly used Bayesian computational technique is called Markov Chain Monte Carlo (MCMC); see Robert and Casella (1999) for a thorough description.

We first address the modeling of measurement error in the manual counts. An additional ingredient, (exact) video data, introduces constraints that complicate the analyses and we treat that in 3.2 along with an approximation that makes computations feasible; the resulting posterior distribution is described in 3.3

3.1 Modeling the Error in Manual Counts

To model errors in the counts, we introduce notation for the *real*, unobserved counts. Let N_{ijk} denote the true number of vehicles going from intersection i to j to k, and let N_{ij} denote the true (demand) count of vehicles entering the network at j from i. Such true, but unobserved, variables are often called `latent counts.' The latent counts are related to the observed counts as follows:

Demand counts: Assume that the observed demand count C_{ij} (entering j from i) is Poisson distributed with mean $b_{ij} N_{ij}$, where $b_{ij} > 0$ and $b_{ij} - 1$ is the unknown *bias* of the observer doing the counting. The resulting Poisson density is

$$\frac{(b_{ij}N_{ij})^{C_{ij}} e^{-b_{ij}N_{ij}}}{C_{ij}!}$$

If different observers are utilized at different input intersections (as in the Chicago study) then we assume independence of the count measurement errors. The joint density of all observed demand counts C_D is found by multiplying together all the Poisson densities from the various intersections. Denote this joint density $f(C_D \mid N_D, \boldsymbol{b})$ where N_D is the vector of true latent demand counts and $\boldsymbol{b}$ is the vector of observer biases.

We specify a prior distribution for the unknown $\boldsymbol{b}$ by taking the b_{ij} as independent identically distributed (i.i.d) with a *Gamma*(α,β) distribution, where α and β are positive but unknown and will also be given a prior distribution. Assuming the observer biases are i.i.d. from some population is reasonable, and choosing a Gamma distribution for positive variables is natural. Let $\pi(\boldsymbol{b}|\alpha,\beta)$ denote the product of these Gamma densities.

We take as prior distribution for α,β a constant density on the region $\alpha < 2\beta$. This constraint implies that the prior mean bias (which is $\alpha/(\alpha+\beta)$ for the Gamma distribution) is restricted to be less than 100%, a rather mild restriction. Write this prior density as $1_{\alpha<2\beta}$. Putting everything together and using the rules of probability, the joint distribution of the observed counts, the biases and the parameters α,β given the latent counts is

$$f_M(C_D, \ , \alpha, \beta \mid N_D) = f(C_D \mid N_D, \ )\pi(\ \mid \alpha, \beta) 1_{\alpha<2\beta}. \quad (3.1)$$

Modeling the Latent Demand Counts: Since N_{ij}, at an input intersection, is the actual number of vehicles arriving with exponential inter-arrival rate λ_{ij}, it has a Poisson distribution with mean λ_{ij} . Thus we can use the same model and prior as in Section 2.2 where the C_{ij} were assumed accurate. Then, the joint density of the latent counts and inter-arrival rates is a product of Poisson densities and prior densities of the $\{\lambda_{ij}\}$ and is of the form $f_D(\boldsymbol{\lambda}, N_D)$, as in Section 2.2.

Modeling the Latent Turning Counts: While we could follow a similar path for turning counts as for demand counts we choose to simplify the analysis and accept C_T as exact. Fuller discussion and justification can be found in Molina et al. (2003); the rationale for the simplification lies in the comparatively diffuse (so-termed, vague) posterior distributions that result for the turning probabilities, making the introduction of additional measurement error less compelling. (The simplification, of course, is accurate on the inner network where the video data gives exact turning counts.) The turning counts are then modeled exactly as was done in subsection 2.2, leading to the same posterior density for P as given there, namely

$$\pi_T(P \mid C_T). \qquad (3.2)$$

We will keep track of true latent counts but we will use (3.2) for the posterior distribution needed later.

3.2 Video Counts and a Probabilistic Network Approximation

The implementation of the analysis outlined above requires counts other than entry demand counts (such as internal link or exit counts). For the test bed problem we have video counts on the inner network outlined in Figure 1, which we treat as exact. Let N_{I} denote the vector of video counts of vehicles that enter and leave the inner network. As before, we denote individual counts by N_{ij}, the number of vehicles going from intersection i to intersection j.

Our goal is to produce the posterior probability distribution for all the parameters λ, P, α, β, $\boldsymbol{b}$, N_D, N_T given the data C_D, C_T, and taking into account the video data. Actually we only need the posterior distribution for λ and P to then feed values to the simulator, but it is convenient to carry all the parameters along in a Bayesian analysis and pick out what is needed at the end. If we could run the simulator hundreds of thousands of times we could use the simulator to produce a density $f_{sim}(N|\lambda, P)$ for the joint distribution of all the latent counts on the network, put video count values in and get a 'likelihood' for the parameters λ, P. This likelihood can then be used to complete a Bayesian analysis. However, this approach is infeasible: a CORSIM run for the Chicago network takes 1-3 minutes and would lead to months or more of computation.

The path we follow instead is to approximate the simulator (in this case, CORSIM) by developing a *probabilistic network* that approximates the needed key features of the simulator. (See Tebaldi and West, 1997, and Tebaldi et al., 2002, for a description of such networks for traffic modeling.) This network has the same intersections, inputs, turning moves, etc., but treats vehicles as passing instantaneously through the network. Many simulator features are omitted (e.g., vehicle waiting times), but the approximation is, arguably, accurate in terms of its representation of the number of vehicles passing through links, as long as the system is viewed to be in steady state. The steady state assumption is inappropriate during times of changing traffic demand, but is reasonable for the one-hour period during rush hour in our test bed study.

In the probabilistic network we get, exactly as in Section 2 the density $f_T(N_T \mid P)$

of all true turning counts, the product of multinomial probabilities. For the demand parameters we have the density $f_D(\lambda, N_D)$ for the true latent demand counts N_D and the demand parameters λ, just as in Section 2 (a product of Poisson densities and Gamma priors).

It is important to note that there are numerous constraints on the N_{ijk} and N_{ij}. For instance, the total number of vehicles entering an intersection must equal the number leaving the intersection. Furthermore (and of crucial effect), the video counts lead to known values of some of the N_{ij}, and these known values induce other constraints. Let N denote the region implied by all the constraints, and 1_N be the indicator function on this region.

Put all together, we get the density, given *P,* for the true latent counts and the demand parameters λ

$$f_T(N_T \mid P) f_D(\quad, N_D) I_{\mathcal{N}}. \tag{3.3}$$

3.3 The Posterior Distribution

The overall posterior distribution of all parameters given the data can be shown (see Molina et al., 2003) to be proportional to the product of terms developed in Sections 3.1 and 3.2:

$$\begin{aligned} &\pi(\quad P, \quad, \alpha, \beta, N_D, N_T \mid C_D, C_T) \propto \\ &\quad f_M(C_D, \quad, \alpha, \beta \mid N_D) \pi_T(P \mid C_T) f_T(N_T \mid P) f_D(\quad, N_D) I_{\mathcal{N}}. \end{aligned} \tag{3.4}$$

Dealing with a complicated posterior distribution, as in (3.4), is done with Markov Chain Monte Carlo (MCMC) analysis (see Chen et al., 2000, and Robert and Casella, 1999) adapted in Molina et al. (2003) to account for the constraints induced by the relationships between vehicles entering and leaving an intersection. To use in the simulator, simply record the collection of λ, P values and use these values as inputs to the simulator, one sample value for each run. In practice, only a few hundred values of the (λ, P) vectors will be utilized in a prediction, whereas the MCMC run will typically result in, say, 100,000 (dependent) realizations of (λ, P). Hence, one need only save, say, every 500th realization of the (λ, P)'s from the MCMC run. The resulting 200 realizations will also be much less dependent and be adequate for use as inputs to CORSIM.

4. Results, Discussion and Extensions

4.1 Results of the Bayesian analyses

In Figure 3 we see the effect of observer error on the estimate of demand counts. The observed demand counts C_{E20} , C_{F15} at two adjacent intersections were incompatible with video data. An ad hoc adjustment was made to change the observer count C_{F15} from 145 to 82 and leave the observer count C_{E20} at 540 ("untuned"). The histogram in 3a is the posterior density of the true count, N_{E20}; the histogram in 3b is that of N_{F15}.

[FIGURE 3 GOES HERE]

The histograms exhibit the uncertainties in the values of N_{E20} and N_{F15} with reductions to both of the incompatible counts.

The effect on estimates of turning percentages can be seen in Figure 4. The histogram on the left depicts the posterior distribution for the probability of going through from intersection E to 20, the one on the right is for the left turn probability. The vertical line shows the MLE estimates for each. The indication is that through traffic is well described by the MLE estimate, while the left turn shows considerable bias.

[FIGURE 4 GOES HERE]

The combination of ingredients of the type just described leads to histogram H_3 in Figure 2, formed from the SQT output of 200 runs of CORSIM using inputs drawn from the posterior density of (3.4). The extra variability over H_2 is a measure of the effect of observer error. Perhaps more to the point is that the comparison of the two Bayesian histograms (H_2 and H_3) reveals the value of high-quality data: H_3incorporates observer error and is substantially more variable than H_2 which assumes all entry demand data are accurate (as if video were available everywhere). It provides an interesting representation of the "benefit" of collecting accurate data in terms of *decreased uncertainty* in the model outcome.

For many individual links there will be substantial differences (as implied by Figure 2) between the ad hoc tuned and "full" Bayesian results. For others there may be only a small, even opposite effect. The Bayesian analysis can also produce posterior distributions of other unknowns of interest. For instance, Figure 5 presents histograms of the posterior distributions of several observer biases. These were estimated for entry counts taken at E, 20 (top left); M, 4 (top right); R, 24 (bottom left); and H, 5 (bottom right). Not surprisingly, bias (= b−1) can be considerably greater than 0 for some observers and considerably negative for others.

[FIGURE 5 GOES HERE]

4.2 Missing Counts

The complete Bayesian analysis can readily handle missing counts in the data (whether or not measurement errors are present). Details are in Molina et al. (2003). The traditional solution of inputting guesses for the missing counts, totally ignoring the uncertainty in such guesses, results in underestimation of the variability of the outputs. The Bayesian approach assures the incorporation of the inherent uncertainty.

4.3 Calibration of Other Parameters

Calibration or tuning of other model parameters for example, saturation flow rate or lane changing behavior distributions can be pursued following a similar line of attack. Such parameters would be given a prior distribution and then incorporated into the Bayesian framework. However, a serious complication will be encountered. In order to engage these parameters we will necessarily require many runs of the simulator. In Section 3 in order to deal with demand counts, we circumvented matters by replacing the simulator with a probabilistic network approximation, but these new model parameters are intrinsic to the simulator and not manageable by such a device. A different type of approximation to the simulator must be found that will allow rapid evaluations of output when values of these new input parameters are varied (varying the values of free-flow speed would not affect outputs in the probabilistic network approximation of Section 3.3). This is an open question. The possibility of employing the approximation methods used in Bayarri et al., (2002) remains to be explored. Once such an approximation is deployed the Bayesian approach can produce the needed uncertainty assessments and predictions.

For simulators that employ route-choice algorithms the key input of an O-D table is subject to considerable uncertainty stemming in large part from incomplete data and uncertainties in survey data. The Bayesian approach can feasibly treat the uncertainty in the O-D table provided the O-D variability can be parameterized with a few parameters rather than a whole table of parameters.

4.4 Validation

Validation in the presence of calibration and data uncertainties brings yet another layer of complexity. In particular, a validation process must introduce the possibility of model discrepancy or model bias that is, the difference between what the model predicts and reality. Accounting for this potential bias, which is intertwined with calibration/ tuning parameters, adds a level of uncertainty that must be treated. To see how the Bayesian approach can respond to these issues we refer to Bayarri et al. (2002) where the methods are applied to deterministic simulators. Extending this approach to stochastic simulators like CORSIM is, in principle, doable, subject to finding reasonable approximations that make computations feasible.

5. Acknowledgments

This research was supported by the U.S. National Science Foundation, Grant DMS-0073952, made to the National Institute of Statistical Sciences. We also thank Byungkyu (Brian) Park and Baohong Wan for assistance with the computations.

References

Bayarri, M.J., Berger, J.O., Higdon, D., Kennedy, M., Kottas, A., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J.C., Lin, C-H., Tu, J. (2002) "A Framework for Validation of Computer Models". *Technical Report 128, NISS, www.niss.org/*

Benekohal, R.F. and Abu-Lebdeh, G. (1994). "Variability analysis of traffic simulation outputs: practical approach for TRAF-NETSIM". *Transportation Research Record* **1457**, 198-207

Bernardo, J.M., and Smith, A.F.M. (2001), *Bayesian Theory* (second edition), New York, Wiley.

Chen, M.H., Shao, Q.M. and Ibrahim, J.G. (2000), *Monte Carlo Methods in Bayesian Computation*, New York: Springer-Verlag.

Hourdakis, J. and P. Michalopoulos (2003), Practical Methodology for Microscopic Simulation Calibration, Presented at the Workshop on Traffic Modeling: Trends and Challenges, Sitges, Spain, June 2-4.

Jha, M., Darda, D., Tomer, T. and M. Ben-Akiva (2003), Joint Estimation of Time Dependent O-D and Calibration of Microscopic Simulation Model, Presented at the Workshop on Traffic Modeling: Trends and Challenges, Sitges, Spain, June 2-4.

Molina, G., Bayarri, M.J. and Berger, J.O. (2003). Statistical Inverse Analysis for a Network Microsimulator, *ISDS Working Paper Series* 03-03.

Robert, C.P. and Casella, G. (1999), *Monte Carlo Statistical Methods*, New York: Springer-Verlag.

Sacks, J., Rouphail, N., Park, B., Thakuriah, P. (2002). Statistically based validation of computer simulation models in traffic operations and management. *J. Transp. Stat.* **5** 1-24

Tebaldi, C., and West. M. (1997). Bayesian Inference on Network Traffic using Link Count Data (with discussion), *J. American Statistical Association*, 93, 557-576.

Tebaldi, C., West, M. and Karr, A.F. (2002). Statistical Analyses of Freeway Traffic Flows, *Journal of Forecasting*, 21, pp. 39-68.

Tian, Z.Z., Urbanik, T, II, Engelbrecht, R., Balke, K. (2002). Variations In Capacity and Delay Estimates from Microscopic Traffic Simulation Models, *Transportation Research Record* **1802**, pp. 22-31.

US Federal Highway Administration (1997), ``CORSIM User's Manual", FHWA, U.S. Department of Transportation Office of Safety and Traffic Operation R&D, Intelligent Systems and Technology Division, McLean, VA.

List of Figures

Figure 1

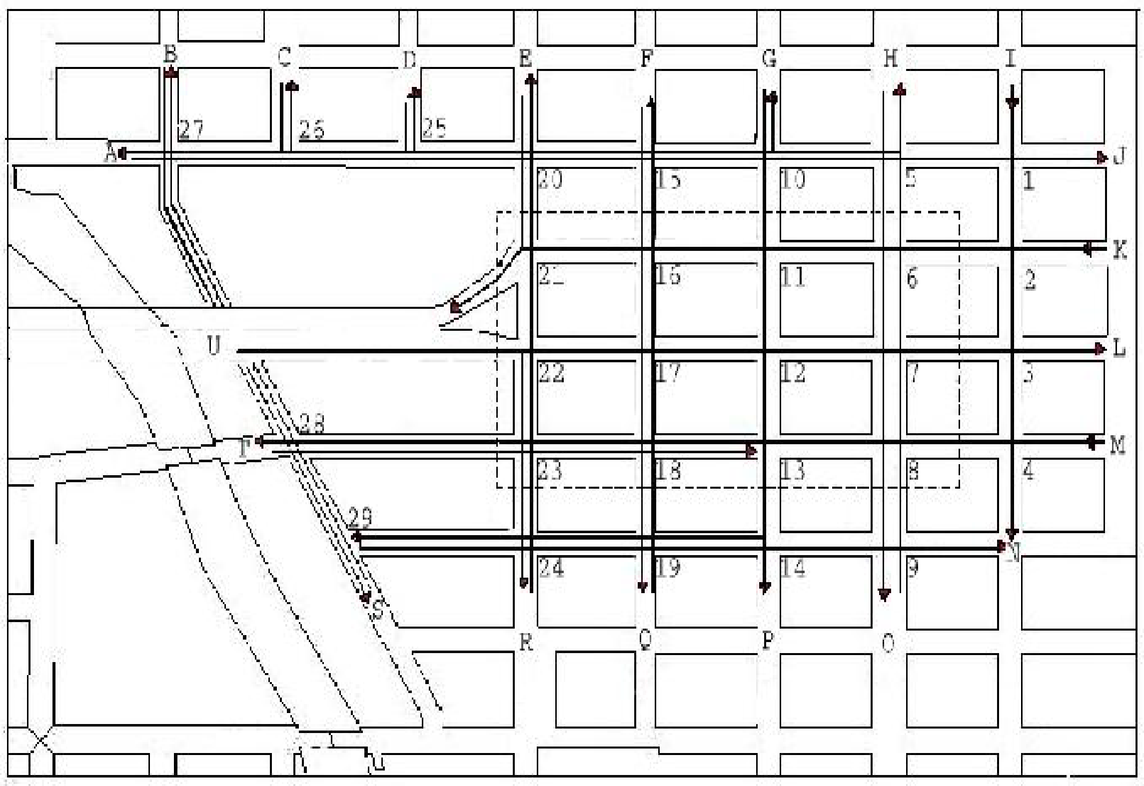

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

Figure 2

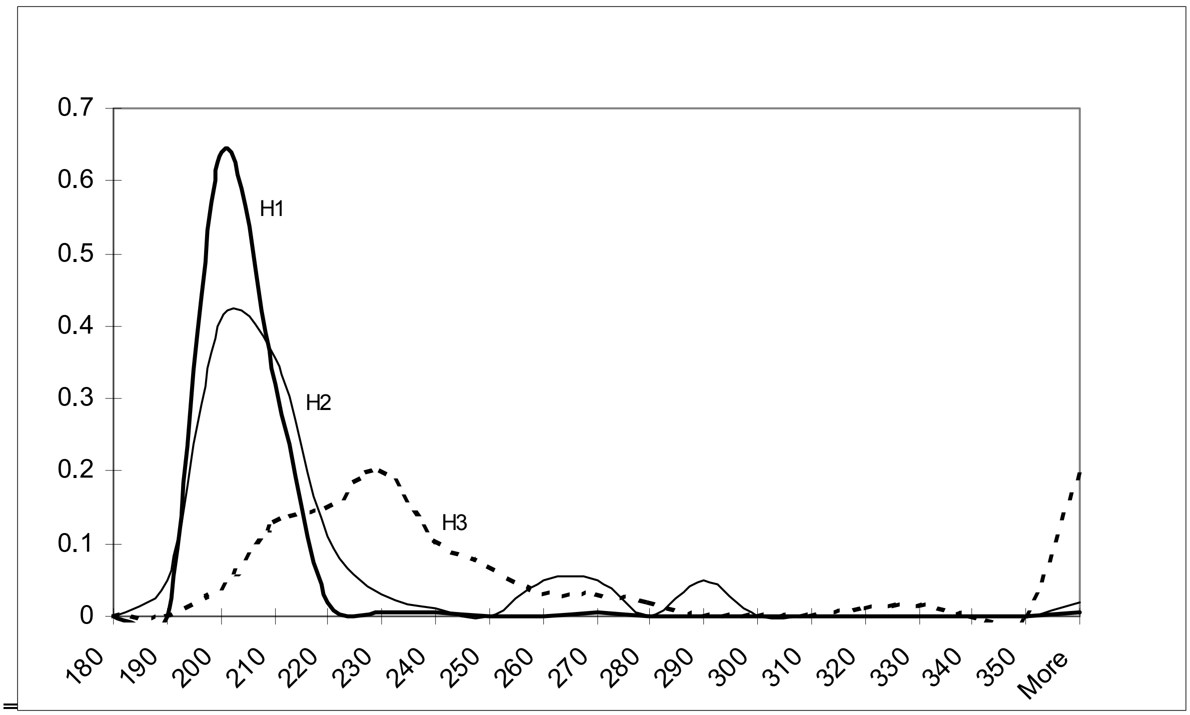

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
H1
H2
H3
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
More

Figure 3

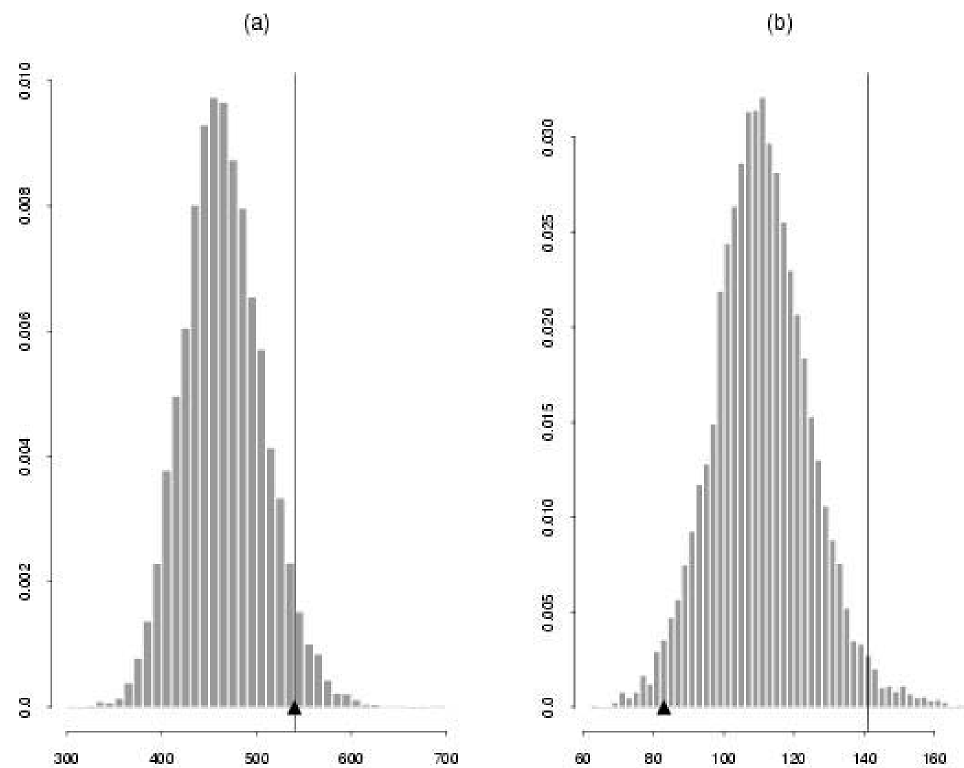
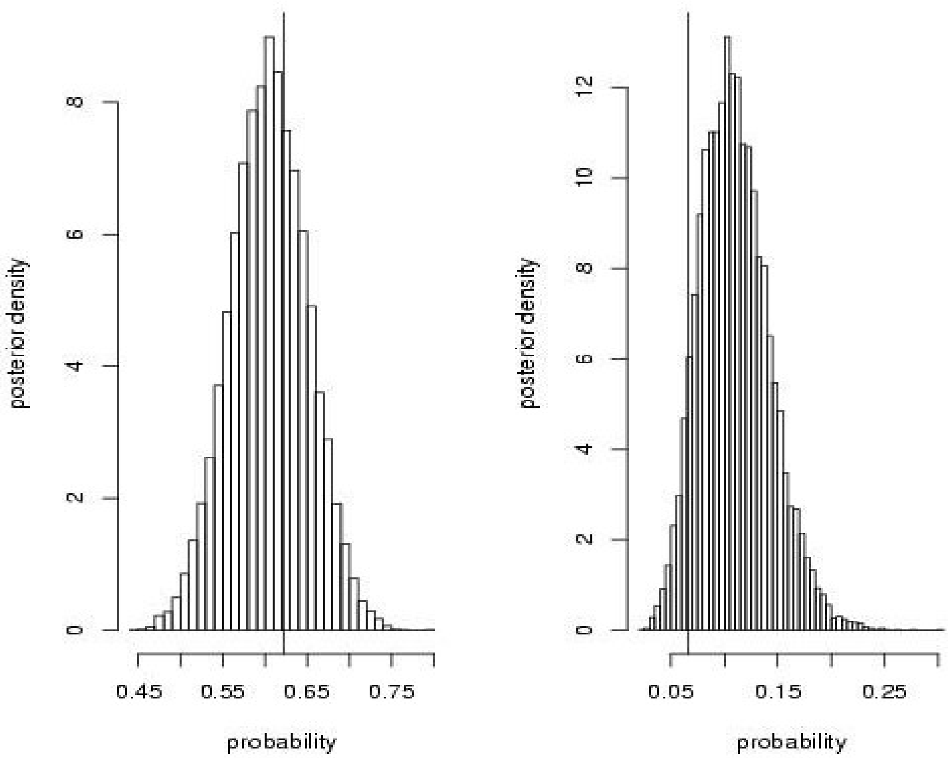

(a)
0.010
0.008
0.006
0.004
0.002
0.0
300
400
500
600
700
(b)
0.030
0.025
0.020
0.015
0.010
0.005
0.0
60
80
100
120
140
160

Figure 4



posterior density
0
2
4
6
8
0.45
0.55
0.65
0.75
probability
posterior density
0
2
4
6
8
10
12
0.05
0.15
0.25
probability

Figure 5