REGIONAL DYNAMIC TRAFFIC ASSIGNMENT FOR REAL WORLD TRAVEL DEMAND MODELS – TRYING IT OUT IN SAN FRANCISCO

Michalis Xyntarakis
Parsons Brinkerhoff
Albuquerque, N.M., U.S.A.
Xyntarakis@pbworld.com

Elizabeth Sall*
Senior Transportation Planner
San Francisco County Transportation Authority
Telephone: 415 522-4810
Elizabeth.sall@sfcta.org

Jim Hicks
Parsons Brinkerhoff
Albuquerque, N.M., U.S.A.
hicksji@pbworld.com

Billy Charlton
Deputy Director for Technology Services
San Francisco County Transportation Authority
Telephone: 415 522-4816
Billy.charlton@sfcta.org

Submission Date – August 1, 2009

4878 words + 4 figures x 250 words + 0 tables x 250 words = 5878 words

Paper submitted for consideration for Presentation and Publication
Original Submission date: August 1, 2009

*Corresponding author

ABSTRACT

Over the past few years many travel demand modelers have come to the realization that their static traffic assignment algorithm is inadequate. As the demand portion of our models become more sophisticated, we are able to confidently model and compare many policy scenarios. Embarrassingly, within the limits of static traffic assignment we cannot calculate such simple measures as travel time between an origin and destination without a string of caveats attached. Nor can we provide volume-to-capacity ratios that traffic engineers won’t laugh at. Unfortunately, many of these simple “skims” that we are so unsure about in static assignment are fed back in to our demand models. How well can our sophisticated demand models do when they are being fed wrong inputs from the start? All signs seem to be pointing to simulation-based Dynamic Traffic Assignment (DTA) as the solution to our static assignment ailments, but no regional modeling agency has (to our knowledge) taken the plunge to actively explore DTA as a possible replacement for static assignment.

To date, several examples of simulation-based equilibrium DTA models coupled with regional travel demand models exist. However, extraordinary run times and complexity have diminished their usefulness outside of research or for some very specific projects. This paper shares our experience to date: strategies, approach, and discoveries in applying DTA in San Francisco to evaluate its feasibility to be a part of our regional demand model as we explored various DTA packages.

REGIONAL DYNAMIC TRAFFIC ASSIGNMENT FOR REAL WORLD TRAVEL DEMAND MODELS – TRYING IT OUT IN SAN FRANCISCO

INTRODUCTION

To date, most regional travel demand models assign vehicle trips to their roadway network using static, deterministic user equilibrium (SUE) methodology. This methodology has several benefits: deterministic methodologies result in the same answer the each time it is run with the same inputs; static user equilibrium is relatively quick and easily distributable; years of research and use have led to widespread understanding of the underlying methodology and assumptions; the network and validation data needs are relatively small and have been collected for a long period of time to allow for easy “backcasting”; the algorithm is relatively cheap and included in most (if not all) current transportation modeling packages. Historically, the aggregate approach of SUE has been sufficient for evaluating capacity-increasing projects such as freeway expansion and does a fairly good job of predicting volumes and travel times on suburban-style low congestion networks. However, as our cities’ transportation networks becomes more and more congested, and our ability to match travel demand with increased capacity diminishes, travel modelers need to be able to confidently evaluate the efficacy and feasibility of alternative mitigations and planning measures other than adding lanes for single occupancy vehicles.

Several problems plague realistic analysis with SUE: (1) aggregate link-based travel time functions (i.e. Akcelik or 1964 Bureau of Public Roads) are calculated irrespective of upstream and downstream congestion and ignore the effects of bottlenecks, intersection geometry and delay, transit vehicle interaction, and queuing; (2) links can be assigned more vehicles than their ultimate capacity resulting in impossible volume-to-capacity ratios; (3) aggregate representation of travel conditions over an entire time period implies that every vehicle traveling over the same link for a particular time period will experience the same travel time on that link, and that travel time is affected by every other vehicle who traverses that link during that time period. As demand increases for peak periods, peak periods become longer, making the reliability of a single travel time representing an entire period suspect; (4) aggregate representation of demand makes it difficult and inefficient to represent variation among individual travelers such as distributed values of time (VOT). These drawbacks to SUE are especially apparent when analyzing San Francisco, CA, with a congested, signal-controlled, grid network interacting with many transit vehicles and parked cars. Furthermore, the inability of SUE to simulate individual choices based on their value of time makes it difficult to confidently model High-Occupancy Toll (HOT) lanes, and congestion pricing, just two of the major policies being tested in the San Francisco Bay area.

Just as more sophisticated activity- and tour-based travel demand models have shown to have better sensitivity to non-standard transportation improvements compared to traditional four-step models, dynamic traffic assignment (DTA) provides the sensitivity and robustness needed to analyze three current items of great importance in the San Francisco Bay Area: pricing, diversion analysis, and reliability.

To date, there are three major barriers to a full regional DTA application: required computing time, required network detail, and consistent, good, and believable solution quality metrics. This study attempts to evaluate the potential of a regional DTA application on the basis of being able to overcome these three problems. The Study Team evaluated three DTA software packages with the dedicated support of their developers. Other software packages were considered, but ultimately time requirements necessitated focusing on three: Dynameq, Vista, and DynusT.

This paper discusses the feasibility of using DTA on a regional scale with the context of large-scale travel demand forecasting. It begins with a discussion of the analysis framework for evaluating DTA. Next it discusses some of the open-source tools developed for this project in order to overcome data management and transferability problems. Third, it discusses the analysis results for a small test network, the San Francisco County network, and extrapolates these to a regional network. The final section presents conclusions and recommendations.

SIMULATION-BASED DTA OVERVIEW

All three packages being evaluated in San Francisco use simulation-based equilibrium dynamic traffic assignment (DTA) methodologies that yield an approximate dynamic user-equilibrium (DUE) solution using an iterative solution method. The solution can be interpreted as a prediction of drivers' pre-trip route choices under conditions of relative familiarity with the network and prevailing traffic conditions.

These DTA models consist of three primary component methodologies: a *time-dependent shortest path* mechanism, a *mesoscopic traffic simulation* based traffic flow model, and a feedback-based dynamic uer-optimal equilibration procedure. Combined, these components seek to obtain a time-varying extension of the Wardrop (1) user-equilibrium conditions for static assignment: for any given departure interval and origin and destination, a driver cannot improve his travel time by unilaterally changing paths. Each iteration consists of one or more executions of the path-building model followed by one run of the traffic simulation model (covering the entire time period under study). The path-building model and traffic simulation model sequence computes the time-varying assignment (mapping of origin-destination demands to path demands). The feedback and equilibration step determines the time-varying *experienced* path travel times realised by the traffic simulation model as a function of the previous iteration. The traffic simulation model determines the arrival times of vehicles to links along paths through the network such that at convergence of the algorithm, the equilibrium conditions are met.

The simulation model – from which the path travel times are determined for each new set of path flows – is a detailed model that more realisticaly represents traffic flow conditions than the BPR curves do in static user equilibrium. The traffic simulation model moves each vehicle from its origin to its destination on a link or lane-based (2) representation of the network using a simplified car-following model(3) or cell transmision model (4,5), combined with lane changing and gap-acceptance models. In contrast to the widely used BPR functions used to relate link flow and link travel time the DTA simulation models respect the triangular (simplified linear) version of the flow vs. density relationship (fundamental diagram) and as such are able to effectively model spillback effects, weaving, and lane drops. Gap-acceptance is used to resolve uncontrolled conflicts at intersections. Sophisticated heurisicts are often applied for modelling a driver's lane selection behavior under various traffic conditions (2). These heuristics often include a "look-ahead" feature that ensures that downstream information (traffic conditions) that are available to the driver (in reality), either due to driver familiarity or because they are directly visible, is used in an appropriate and realistic way.

The time-varying assignment consists of determining the path proportions for each O-D pair for each departure-time interval. By computing the assignment for all intervals based on the complete travel-time information from the previous iteration, the interactions between vehicles starting their trips in different departure intervals is explicitly taken into consideration. The proprtions of O-D flow assigned to available paths are determined by the equilibration procedure implemented in the DTA model.

No claims can be made about the existence or the uniqueness of a DUE solution. The equilibrium principle is used as a guide in computing an approximate solution. Moreover, as the higher level of realism leads to greater sensitvity to model inputs, and the use of stochastic traffic simulation can introduce a certain amount of noise in the results, it should be appreciated that typical values of relative gap (after convergence to equilibrium) are often considerably higher than those obtained with static assignment models. One of the objectives of this study was to see what convergence and stability properties woulod be observed in the San Francisco DTA model.

The Dynamic Traffic Assignment and Simulation System (DynusT) is a simulation-based DTA modeling package employing a mesoscopic simulation approach based on the Anisotropic Mesoscopic Simulation model that omits inter-vehicle car-following details while maintaining realistic macroscopic traffic properties (6).

The Visual Interactive System for Transportation Algorithms (Vista) is a simulation-based DTA modeling software package (7,8). A critical capability of VISTA is the ability to achieve Dynamic User Equilibrium (DUE) by employing a mesoscopic traffic/transit simulation using high fidelity time-dependent shortest path algorithms.

Similar to the two DTA packages mentioned above Dynameq is a simulation-based DTA model that is based on an event-based car following model (Vista and DynusT use a time-step approach) that simulates vehicles on lanes rather than links that the previous two models do (2, 9, 10).

STUDY SCOPE

San Francisco Network

The San Fransisco network shown in Figure 1 is an ideal large-scale test network for several reasons. There are only four major entry points into the city, making it easy to cut off the network there and not worry about diversions. The four major entry points are U.S. 101 North (the Golden Gate Bridge) to Marin County, I-80 East (the Bay Bridge) to Treasure Island and Oakland, and U.S. 101 and I-280 South, to the San Francisco peninsula and Silicon Valley/San Jose. San Francisco mainly consists of a tight grid network, which can showcase the benefits of DTA due to the wide variety of alternative paths and queuing potential.

Every single highway, street, alleyway, and turn penalty is already coded into the street network within San Francisco, making additional coding minimal. The total number of links is 24,000 and 8,000 nodes; however, many of the centroid connectors are currently coded directly to intersections; this has raised several problems when importing the signal settings these intersections have from Synchro. There are 1,200 signals in San Francisco (red dots in Figure 1). While it would be onerous to code these one by one, they are available in electronic format allowing the Team to develop software libraries that convert them to various DTA formats. The current 2008 PM three-hour peak period assignment has 550,000 vehicle trips and 1000 traffic analysis zones.

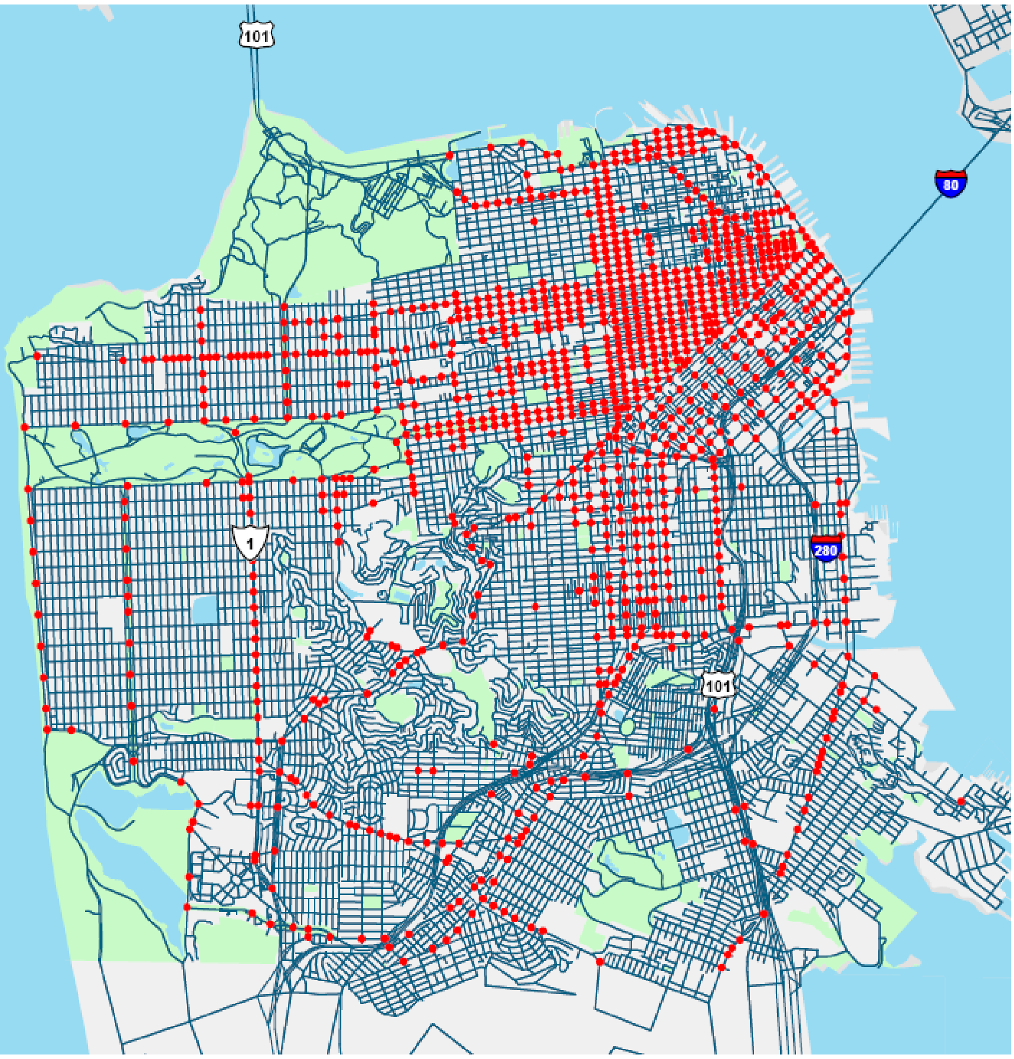

FIGURE 1 San Francisco County Network

San Francisco Bay Area Regional Network

The future goal, is to accomplish DTA on the San Francisco Regional Network, of which the San Francisco Network is a subset. The Regional Network has over 2,400 zones and is expected to carry over 30 million trips by 2030. Throughout, the team was mindful that all exercises and tests carried out on the San Francisco network are at the small side of this future goal.

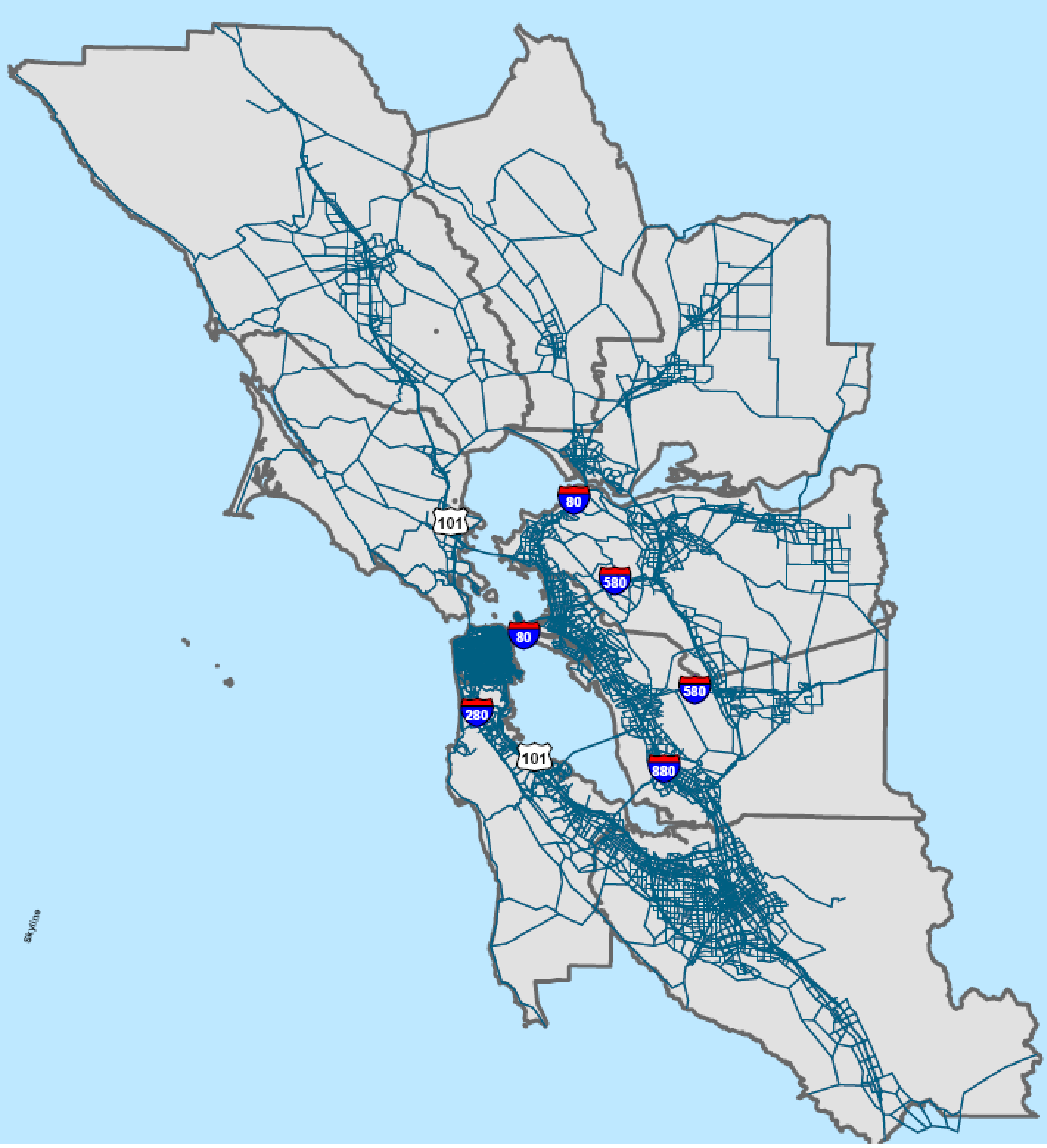

FIGURE 2 San Francisco Bay Area Regional Network

DTA DATA MANAGEMENT

The study team chose to begin developing the DTA model using DynusT. As data became available and was cleared from errors using DynusT as a test bed model it was transferred to both Vista and Dynameq. The following sections describe the data conversion processes that were used.

Network Data

Considerable effort has been expended in the assembly, testing, and validation of model inputs. Separate networks have been constructed for the AM and PM peak periods, using the regional network as the starting point. The highway network underwent rigorous validation to ensure proper coding and connectivity. Checks included searches for asymmetrical link properties, improperly coded link distances, zones that could not be reached, inconsistent link attributes (e.g., a freeway link miscoded as a major arterial), ramp and lane directionality, and other coding issues. Centroid connectors for the Dynameq and Vista networks were rebuilt so that they connect at mid-block locations, rather than intersections.

Signal Data

Increased network fidelity, almost at the same level of detail a traffic engineer works with, is required for DTA modeling. The most important prerequisite of a DTA model is the assembly of traffic signal control information and turn bay locations that accurately reflect the base year conditions. Additional fidelity can be gained, and it is often necessary, by coding the stop or yield signs and the merge and diverge bays at freeway junctions. A primary source of signalized data can be the Synchro models agencies use to optimize signals in a grid or along a corridor. However, more often than not, the planning network a DTA modeler starts with and the one or more Synchro networks available are incompatible. The two main reasons being: (1) Synchro networks usually contain significantly more detail than is usually found in planning networks; and (2) Synchro networks are not always coded to scale and thus do not reflect actual distances or relate well spatially to the planning network. If they are properly maintained Synchro networks are an excellent source of information because besides the indispensable timing plans, information can be found about the number and length of turn bays, the presence of prohibited movements, and the hourly flows used for the signal optimization process.

The fundamental elements of a Synchro network are nodes, links, movements, timeplans and phases. All these elements are defined in Synchro’s Universal Data Transfer Format (11) and are present in an ASCII file. Synchro time plans can be subdivided into three categories: (1) pre-timed plans in which both the phase timing and the phase sequencing are predefined and do not change during the signal’s operation; (2) single-ring actuated signals for which the phase sequencing is pre-determined but the timing is influenced by the vehicle flows during operation; (3) dual-ring actuated signals in which both the timing and the sequencing of the phases can vary based on the conditions on the ground. In contrast, signal representation in the DTA packages under study is much more simplified. Dynameq and Vista DTA can currently model only pre-timed signals and only DynusT is able to simulate single-ring actuated signals. As a result, signal conversion from Synchro to the DTA models is a time-consuming and error prone procedure. From our experience, an experienced traffic engineer can convert thirty to forty signals a day. Given that a metropolitan region has many hundred and often thousands of signals and that signal timing usually varies by time period it becomes evident that an automation procedure that imports and validates the input signals is highly desirable. Nevertheless, manual inspection and intervention will always be necessary as many special cases will arise in the conversion process that cannot always be handled with pre-defined logic embedded in a script.

The study team developed a hierarchical procedure that translates Synchro network elements to the corresponding DTA ones. Overall, out of the 1200 signals found in the study area the study team successfully converted 900 using the following steps. The timing parameters for the remaining 300 of them were calculated using Webster's method (i.e., optimal timings for each signal as if it were independent and not affected by adjacent signals).

Each Synchro intersection is mapped to a DTA intersection based on its incoming and outgoing street names. A library with string matching algorithms proved necessary for the successful completion of this step because of misspellings and differences in street abbreviations. A match is made only for the intersections with very high similarity for which the possibility of a mistake is minimal.

The rest of the intersections that have not been mapped using street names are being mapped based on their relative distance from the previously mapped intersections and the number of incoming and outgoing links. The two networks are then overlaid using a GIS mapping software to verify the correctness of the procedure. At the end of this step 80 to 90% percent of the nodes have been mapped successfully.

For each adjacent link of a Synchro intersection that has been mapped the corresponding DTA link is being identified based on link orientation. This is a relatively easy step with high success rate in which 95% of the links are associated properly.

For each Synchro intersection that all the previous steps yielded no errors, its timing plan is being copied to the DTA network by creating a DTA phase for each Synchro phase and assigning to it all the movements that correspond to the Synchro movements.

Finally a number of exhaustive checks are applied to the converted signal that ensure that it covers all the movements associated with the intersection and no errors will occur during simulation.

Demand Data

Static estimates of the demand from the SF-CHAMP activity-based demand model were converted into 15-minute intervals using a straightforward process of allocating demand based upon observed diurnal distributions. These distributions were derived from household survey data and observed peaking characteristics from traffic count data across the region. The demand at the external cordon, which includes the Bay Bridge to the west and the Golden Gate Bridge to the north and roadways crossing the County boundary to the south were constrained to observed traffic counts by direction. The distribution of those counts to O/D flows to/from zones within the County was determined by a select link analysis procedure performed using the regional SUE traffic assignment program that is part of the regional model.

The study team observed that the distribution of demand by time period did not allow for traffic flows from the DTA to match very well with the observed link count data from the Synchro files or the other link based counts. They did not appear to match any poorly than the static assignment model, however. Going forward, there is probably a need to perform a matrix adjustment procedure to allow the demand matrices by time period to conform to the observed network flows. An additional requirement will be to devise a forecasting methodology that will produce adjusted demand matrices based on the base year adjustments determined and the forecasted demand from the SF-CHAMP and regional model select link analysis procedures.

POST PROCESING OF DTA RESULTS

All of the DTA packages used in this study are GUI-based and provide adequate facilities for network editing and visualization. However, they are mostly focused towards solving the Dynamic User Equilibrium assignment and they lack many common facilities found in other planning software. In addition, DTA outputs such as paths and travel times are time-varying by nature and cannot be handled by the reporting facilities of the traditional planning software. Consequently, post-processing the DTA output is a necessary step for the modeler and has much greater significance to the planning process than GIS scripting has to the traditional 4-stepped model development.

With the help of the package developers the study team developed a set of object oriented open-source python libraries that read and write data in the following formats: Synchro, DynusT/Dynasmart, Vista DTA, ESRI shapefile, Google KML. Significant effort has been put in place to ensure that good programming standards have been followed during the development and testing phase (12). The reporting generation mechanism described below common for all the networks under study and is built utilizing the common application programming interface shared by each of the network libraries mentioned above. The reports that have been created thus far and used for validation and demonstration purposes are the following:

Validation reports using link or turning movement flows in different time resolutions and their comparisons with observed counts via relative error and percent root mean square error calcualtions.

Validation reports contrasting observed and simulated speeds along a corridor at different times of day.

Speed-plot diagrams that depict the evolution of link speeds across distance and time and are useful in comparing the congestion in a corridor under different scenarios.

Convergence reports that utilize our own time-dependent shortest path algorithm to compare and contrast the quality of the convergence of the various packages.

Google Earth animated KML visuals that can be widely distributed and show how network speeds evolve over time.

EVALUATION METRICS

As a result of performing the exercise of transforming data into proper formats and testing several DTA packages, the SFCTA team has concluded that several general requirements and features are necessary to constitute a successful DTA implementation. These include flexibility, transparency, solution quality, scalability, speed, stability, and usability.

Flexibility is important, recognizing that DTA is a rapidly evolving field, and that the best current DTA implementation may not be the best in the future. The DTA model should be implemented with open source software and be operating system neutral. New and different non-standard operating systems may emerge with the evolution of computers with multiple multi-core CPUs and might provide the best platform for distributed processing and rapid development.

In order to be confident in results, users must understand the assumptions, algorithms, and limitations contained within a DTA. As not every user is an expert in each individual component of DTA, an implementation may be viewed as transparent based on publications of algorithms in addition to being open source. For example, a term such as “relative gap” can have generally the same conceptual meaning but be calculated in varying ways by different DTA software, necessitating clear documentation.

Among measures to demonstrate solution quality are indicators of convergence such as relative gap, travel time variation for similar origin destination pairs and time slices, and simulation model solution – meaning the all vehicles should be loaded into the simulation model and exit from the simulation model within a reasonable amount of time for the period being simulated.

The stability metric can mean several things: stability in solutions or stability in software. Both are desirable. Stability in the solution is indicated by small changes in data resulting in relatively small and justifiable changes in the output. Software stability is also important and is characterized by software that is bug free or that provides useful feedback to users to help them identify errors in their data or their software setup.

In order to be useful in the context of regional demand modeling for SFCTA, the DTA must be able to scale up to do a complete daily DTA of the 2030 Bay Area regional network within an approximately 24-hour runtime.

Finally, an easy-to-use interface is essential to quickly finding and fixing errors, summarizing impact measures and quality of results, and making this tool helpful to non model experts. Usability is also enhanced by tools that ease the process of importing and exporting model data when making full migrations of networks and demand to and from different platforms, allowing the practitioners to easily switch platforms and use the best tool for the question to be answered.

DTA SIMULATION STATUS

The primary objective of the work that has been completed thus far has been to test the feasibility of implementing a DTA model in San Francisco using a process that would allow to move transparently between the three DTA packages under study. A fully validated DTA model for the region that will become an integral part of the SFCTA modeling process will start being built this fall utilizing the wealth of count and travel time data available for the San Francisco County.

An initial validation report can be seen in Figure 3. In this report the average observed route speed in mph is plotted against the average simulated speed for fifty of the most important San Francisco routes spanning 20miles of length. The R-squared value of 0.72 that is obtained is rather typical. It is important that the DTA model has not been calibrated. The traffic flow relationship has not been calibrated to reflect the local conditions, a process similar to the BPR calibration procedure that static models undergo. Also, transit vehicles have not been included and only one class of vehicles is being modeled as trucks have been converted to equivalent passenger car units. For comparison reasons the same plot is being shown in Figure 4 using the speeds from the static model that is currently used by SFCTA and has gone though a calibration process.

The resulting model was run 25 to 30 times using DynusT, including several runs where the inputs were varied in order to ensure that the model was appropriately sensitive to input conditions. DynusT converges in approximately 11 hours on a 3.0Ghz Intel Xeon processor of the latest technology. Initial test runs that have been done using Dynameq without the signals have achieved convergence in about 17 hours on a machine that is 30% slower than the one that is used for the

DynusT runs. The Vista network is still under development and reliable running times cannot be reported.

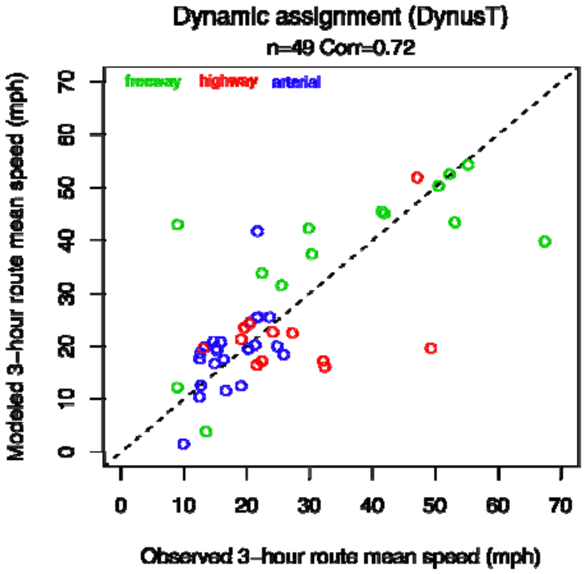

FIGURE 3: Dynamic Assignment Speed Validation

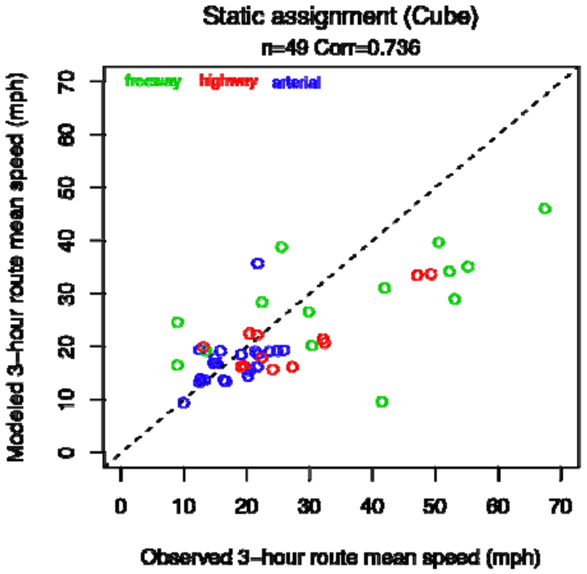

FIGURE 4: Static Assignment Speed Validation

CONCLUSIONS

Simulation-based DTA has the potential to overcome many of the drawbacks that hamper the static assignment including unrealistic volume to capacity ratios, queue spillback, and insensitivity to minor network changes such as signal times. In this study, SFCTA's ongoing efforts were presented to migrate its existing planning network and utilize additional signal timing information found in external sources to build a workable DTA model using more than one DTA package. Initial results obtained from one DTA package suggest that the preliminary un-calibrated DTA model performs no worse than the regional planning model in terms of average speeds. Yet, results from the rest of the models are still pending. Focus of has also been in the development of reusable and easily distributable tools that ease the migration from a traditional static model to a DTA one. Finally, the open-source tools have been developed that provide an interoperability layer between the different network formats and allow the user to work in a network agnostic way.

ACKNOWLEDGEMENTS

The authors would like to thank the various DTA developers who helped with their time and expertise in this project including Yi-Chang Chiu, Michael Mahut, Travis Waller, and Michael Florian. In addition, the authors would like to thank Rick Donnelly.

REFERENCES

1. Wardrop, J. G. Some Theoretical Aspects of Road Traffic Research. In *Proceedings of the Institute of Traffic Engineers*, vol. 1, 1952, pp. 325-378.
2. Mahut, M. Discrete flow model for dynamic network loading. Ph.D. Thesis, Departement of Computer Science and Operations Research, University of Montreal, Canada. 2001.
3. Mahut, M., Florian, M., Tremblay, N., Campbell, M., Patman, D. and McDaniel., Z. Calibration And Application Of A Simulation-Based Dynamic Traffic Assignment Model, In *Transportation Research Record:Journal of the Transportation Research Board.* No. 1876, Tranportation Research Board of the National Academies, Washington, D.C. 2004, pp. 101-111.
4. Daganzo, C.F. The cell transmission model : a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research* 28B (4), 1994, pp. 269-287
5. Ziliaskopoulos, A.K. and S. Lee. A Cell Transmission Based Assignment-Simulation Model for Integrated Freeway/Surface Street Systems. In *Transportation Research Record, Journal of the Transportation Research Board.* No. 1701, Tranportation Research Board of the National Academies, Washington, D.C., 1997, pp. 12-23.
6. Chiu, Y-C, and L. Zhou. An Anisotropic Mesoscopic Traffic Simulation Model: Basic Properties and Numberical Analysis. Presented at 86rd Annual Meeting of Transportation Research Board, Washington D.C.
7. Ziliaskopoulos, A.K. and S.T. Waller. An Internet Based Geographic Information System that Integrates Data, Models, and Users for Transportation Applications, *Transportation Research Part C: Emerging Technologies,* Vol. 8, 2000, pp 427-444.
8. Ziliaskopoulos, A.K., S.T. Waller, Y. Li, and M. Byram. Large Scale Dynamic Traffic Assignment: Implementation Issues and Computational Analysis, *Journal of Transportation Engineering.* Vol.130, 2004, No. 5, pp. 585-593.
9. Florian, M., Mahut, M., Tremblay, N. Application of a simulation-based dynamic traffic assignment model. *Simulation Approaches in Transportation Analysis: Recent Advances and Challenges, R. Kitamura and M. Kuwahara* (eds.). Springer, 2006, pp. 1-22.
10. Mahut, M., Florian, M., Tremblay, N. Comparison of dynamic-equilibrium assignment methods for traffic simulation. Presented at European Transport Conference 2007, 17-19 October, Leeuwenhorst conference centre, near Leiden, The Netherlands. 2007
11. *Synchro Studio 7 Users Guide:Synchro plus SimTraffic and 3D Viewer.* Trafficware, Sugar Land, 2006, chapter 16.
12. Gamma, E., Helm, R., Johnson, R., Vlissides, J. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional Computing Series. 1994