TMIP Exploratory Modeling and Analysis Tool (TMIP-EMAT) Beta Test Results

DECEMBER 2019

U.S. Department of Transportation
Federal Highway Administration

TMIP
Better Methods. Better Outcomes.

1. Report No. FHWA-HEP-20-015	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle TMIP Exploratory Modeling and Analysis Tool (TMIP-EMAT) Beta Test Results		5. Report Date December, 4, 2019
		6. Performing Organization Code
7. Authors Martin Milkovits, Rachel Copperman, Jeffrey Newman, Jason Lemp, Thomas Rossi		8. Performing Organization Report No.
9. Performing Organization Name and Address Cambridge Systematics, Inc. 101 Station Landing, Suite 410 Medford, MA 02155		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DTFH61-16-D00016L
12. Sponsoring Agency Name and Address United States Department of Transportation Federal Highway Administration 1200 New Jersey Avenue SE Washington, DC 20590		13. Type of Report and Period Covered February 2019 to October 2019
		14. Sponsoring Agency Code HEPP-30
15. Supplementary Notes FHWA Project Manager: Sarah Sun.		

16. Abstract

The objective of this project is to demonstrate and motivate the use of regional travel demand models in an exploratory/experimental manner, as opposed to the traditional single point predictive approach, specifically for analyzing the impacts of new technology. The study contributes to our understanding of the impact that the rapid technological evolution has on the movement of people and goods on surface transportation system. It also evaluates the applicability of the robust decision making process on transportation planning by identifying and addressing hurdles in the application of an exploratory analysis to support real-world planning analysis.

This report describes the successful deployment of Travel Model Improvement Program Exploratory Modeling and Analysis Tool (TMIP-EMAT) at three beta-test sites: Greater Buffalo-Niagara Regional Transportation Council (GBNRTC), Oregon Department of Transportation (ODOT), and San Diego Association of Governments (SANDAG). All three beta-testers developed scopes specific to the agency interests, developed the API and model extensions and completed sufficient runs on agency workstations to support an analysis workshop. The beta-tester core models represent three of the major travel demand model software programs and the applications range from regional, sub-regional and corridor level analyses. Each of the beta-testers utilized their calibrated, official models or subcomponents of the official production model. This report describes the beta-test process, summarizes the applications and highlights lessons learned and areas for future improvement in the TMIP-EMAT application process and in the tool itself.

17. Key Words	18. Distribution Statement
TMIP-EMAT, Exploratory Modeling and Analysis, Travel Demand Modeling, Regional Models, Corridor Studies, uncertainty analysis, Monte Carlo simulation	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	97	N/A

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TMIP Exploratory Modeling and Analysis Tool (TMIP-EMAT) Beta Test Results

Original: **November 2019**
Final: **December 2019**

Prepared for:

Federal Highway Administration

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

- ABM: Activity-Based Model
- API: Application Programming Interface
- ATDM: Active Transportation and Demand Management
- AV: Autonomous Vehicle
- BRT: Bus Rapid Transit
- GBNRTC: Greater Buffalo-Niagara Regional Transportation Council
- GHG: Greenhouse Gas
- GPR: Gaussian Process Regression
- HOT: High-Occupancy / Toll
- HOV: High-Occupancy Vehicle
- ITS: Intelligent Transportation System
- IVTT: In-Vehicle Travel Time
- KnR: Kiss and Ride
- LHS: Latin HyperCube Sampling
- LR: Linear Regression
- MaaS: Mobility as a Service
- MPO: Metropolitan Planning Organization
- ODOT: Oregon Department of Transportation
- OMSC: Oregon Modeling Steering Committee
- NFTA: Niagara Frontier Transportation Authority
- PEV: Personal-Electric Vehicle
- PnR: Park and Ride
- POE: Point of Entry
- PRIM: Patient Rule Induction Method
- RTD: Regional Transportation District
- SANDAG: San Diego Association of Governments
- SEMA: Smart-Lane Enhanced Mobility Arterials
- SOV: Single-Occupancy Vehicle
- TMIP: Travel Model Improvement Program

- TMIP-EMAT: TMIP Exploratory Modeling and Analysis Tool
- TNC: Transportation Network Companies
- TOD: Transit Oriented Development
- TRL-H: Technology Readiness Level for Highway Research
- TSP: Transit Signal Priority
- VE: VisionEval
- VHT: Vehicle-Hours Traveled
- VMT: Vehicle-Miles Traveled
- VOT: Value of Time

1.0 Introduction

1.1 Acknowledgments

The FHWA would like to acknowledge the assistance of the three agencies who committed time, resources, and their models to conduct the test. Without the generous contributions of the Greater Buffalo-Niagara Regional Transportation Council (GBNRTC), Oregon Department of Transportation (ODOT), and San Diego Association of Governments (SANDAG), this work would not have been possible.

1.2 TMIP-EMAT Project Overview

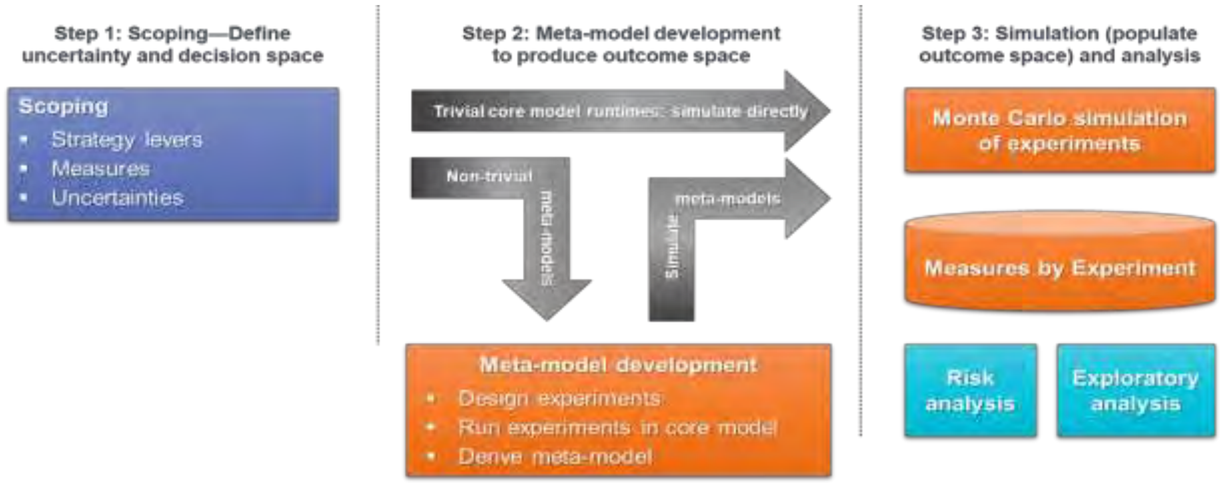
1.2.1 Project Background

The TMIP Exploratory Modeling and Analysis Tool (TMIP-EMAT) utility was developed as part of the FHWA Exploratory Modeling and Simulation Study, whose purpose it is to focus on exploratory, rather than predictive, modeling of future transportation systems with particular attention to the impact of new technology. The goal of this project is to develop a tool that will help agencies manage uncertainties by illuminating interactions between transportation supply and demand on urban surface transportation system through exploratory modeling and simulation; provide insights of potential, possible, plausible, probable or preferred futures; and support robust regional transportation planning decision-making incorporating principles of risk management.

1.2.2 TMIP-EMAT Overview

The FHWA TMIP Exploratory Modeling and Analysis Tool (TMIP-EMAT) is a utility that can be integrated with existing travel demand models to facilitate the application of those models in an exploratory, rather than predictive, manner. It builds upon the evolving sensitivity and risk analysis approaches utilizing travel demand models to forecast uncertainty, can be used to understand the effect of future mobility impacts on travel patterns, and incorporates exploratory-type visualizers and optimization search tools to present and analyze the results.

TMIP-EMAT is designed to integrate with a **core model**, which is any application/region specific transportation model. The typical user of TMIP-EMAT is envisioned to be a planner/modeler who is familiar with the capabilities, and limitations, of the core model. The three major steps to working with TMIP-EMAT are defined in Figure 1.



Source: FHWA

Figure 1: TMIP-EMAT Process Flow

During **Scoping** the user identifies the strategies to be analyzed and the measures to evaluate. The user also considers uncertainties that may affect the outcome of the measures and can be represented by parameters of or inputs to the core model.

For cases where the core model run time is non-trivial, TMIP-EMAT utilizes **meta-models**, which are regression models that estimate the core model outputs. The meta-models run very quickly (microseconds) and thus can be used to produce measures comprehensively across the uncertainty space. Where the core model run time is trivial, as in a sketch model, TMIP-EMAT utilizes the core model directly.

To populate the outcome space, TMIP-EMAT utilizes Monte Carlo **Simulation** to sample across the uncertainty distributions. For each simulation run, the associated value of each uncertainty is set in the model (meta-model or core model), and the measure estimate is recorded in a database. The user can then examine and analyze the effects of their strategy levers on measures of interest utilizing risk analysis and/or exploratory analysis approaches.

1.2.3 Motivation and Purpose of Beta-Tests

The Technology Readiness Level for Highway Research (TRL-H)[1] scale measures where a highway-related technology is on a scale of 1 - Basic Research to 9 – Implementation. The advantage of the scale is to indicate the sort of development that may be necessary and/or additional tests to conduct. The scale also serves as a guide to structure discussion to uncover technical gaps/questions and outline next steps in development.

At the outset of the beta-test, TMIP-EMAT was assessed to be at a TRL-H Level 5: *Integrated Components Demonstrated in a Laboratory Environment*. This level was based on the completed and documented proof of concept application as well as the end user and integration plan documentation. By conducting beta-tests, TMIP-EMAT is a demonstrated prototype in a relevant

[1] https://www.fhwa.dot.gov/publications/research/ear/17047/17047.pdf

and operational environment (TRL-H Levels 6 and 7). The TRL-H assessment is presented in Section 3.4.

The beta-test process provided a great opportunity to evaluate the existing documentation and examples and identify what areas to pursue to improve TMIP-EMAT and to extend the applications. Beta-testers required hands-on support by the FHWA consultants. The support points and materials are documented in this report as well as the TMIP-EMAT online documentation.

Most importantly, the beta-test process broadened the TMIP-EMAT user community. New users of TMIP-EMAT now have the experiences and advice of multiple agencies to help them determine how to best apply the utility for their applications.

1.3 Beta-Test Results

TMIP-EMAT was successfully deployed with a custom API at each of the three beta-testers and sufficient runs completed on agency workstations to support the analysis workshop. Each agency specified seven levers and exogenous uncertainty variables, thus requiring 70 runs of the core model. A summary of the beta tester attributes is shown in Table 1.

Table 1: Beta-Test Summary

Beta-Tester	Model Software Platform	Model Type	Test Application
ODOT	VISUM	Activity-Based Model	Regional analysis of transit, and parking strategies with uncertainty on new vehicle technology
SANDAG	Emme	Tour-Based Sub-Component	Sub-regional analysis of border access highway policies and transit investments with uncertainty around changing demographics, land use, border access, and new vehicle technology
GBNRTC	TransCAD	Trip-Based Model	Corridor-level analysis of transit and complete streets-type improvements with uncertainty on land use, new mobility services, and weather impacts

The beta-tester core models represent three of the major travel demand model software programs, and the applications range from regional to subregional and corridor level analyses.

Each of the beta-testers utilized its calibrated, official model or subcomponents of the official production model. However, the model inputs were not necessarily representative of actual projects or forecasts but instead were developed to demonstrate the model capabilities and ensure a significant response in the model.

For all beta-testers, the API deployment required outside consultant support, at least to develop extensions to the model platform to allow for programmatic control of the model sequence or represent the scoped levers / uncertainty variables.

1.4 Report Structure

Section 2 describes the beta testing process that was conducted by GBNRTC, ODOT, and SANDAG including the beta-test recruitment and selection, scoping workshop, API development and model testing, and final analysis workshop.

Section 3 presents a summary of the beta-testing experiences and results. TMIP-EMAT updates are described along with the lessons learned and next steps for the utility development and deployment. Section 3 concludes with an assessment of the TMIP-EMAT TRL-H level.

Details from each beta-test are included in the appendix. The references section provides links to the utility script repository and online documentation.

2.0 Beta-Testing Process

The beta-tests consisted of four major steps. The steps are described briefly below and discussed in more detail in the remainder of this section.

1. **Initial on-site scoping workshop**: FHWA consultants met with the Agency staff to:
   - o Conduct the scoping process to identify uncertainties, metrics and strategy levers;
   - o Discuss the technical details and work plan to deploy TMIP-EMAT; and
   - o Engage and educate other staff within the Agency on exploratory modeling and analysis.
2. **Development of Application Programming Interface (API)**: The integration, meta-modeling, and simulation process are supported in TMIP-EMAT through a set of components and data repository formats that are agnostic to a specific core model or application. Deploying TMIP-EMAT, however, requires the development of an API to the core model. Through the API, the tool programmatically defines scenarios, launches, retrieves errors and status, and imports measures from the core model. The Agency, or Agency consultant, was responsible for developing this API. The API and core model were run through a series of univariate sensitivity tests to assure reasonable responses and correct operation. FHWA consultants supported development by answering questions, providing examples, and helping to troubleshoot issues that arose.
3. **Utilizing TMIP-EMAT to evaluate proposed strategy or uncertainty.** The Agency completed the model runs to evaluate the proposed strategy and uncertainty under the guidance of the FHWA consultants.
4. **Final on-site analysis workshop**: FHWA consultants met with the Beta-Tester staff to gather lessons learned, value of the exploratory modeling process, and general feedback on the beta-testing experience.

Partner agencies to assist with testing were identified in February and March of 2019, with scoping workshops held in March and May. The model sensitivity tests and API development took place through the spring and early summer. By late summer, all the beta-testers were conducting core model runs through TMIP-EMAT and were able to support the analysis workshops in late September and early October 2019.

2.1 Partner Agency Readiness and FHWA-Support

Throughout the beta-tests, the role of FHWA and its consultants was to offer guidance, examples, review assistance, and troubleshooting. The purpose of this was to enable the beta-testers to do each of the steps necessary to deploy TMIP-EMAT. As much as possible, documented materials were developed and shared to respond to requests for support that would be transferable to other users. Discussions with the prospective testing partners were held to confirm readiness based on:

- **Calibrated and validated core model**. The Agency must be confident that its model produces reasonable results and must be willing to conduct further sensitivity testing as part of the beta-test.

- **Commitment of technical and planning staff and computational resources.** The Agency is willing to dedicate resources to develop the necessary enhancements to their model to integrate with TMIP-EMAT, planning staff to conduct the scoping and evaluation, and computational resources to run the core model through TMIP-EMAT on the order of 10's of times.
- **Practical application for TMIP-EMAT**. The Agency has a proposed strategy and uncertainties for evaluation using TMIP-EMAT. The Agency will conduct this evaluation utilizing the outputs of TMIP-EMAT and will work with the FHWA consultants to identify areas of value and potential improvements to the tool.

2.2 Recruitment and Application

FHWA and its consultants recruited partner agencies to assist with testing through a TMIP webinar[2] and a presentation at the TRB annual meeting. Interested agencies were asked to provide the following information:

- Proposed application for TMIP-EMAT, specifically:
  - What is the policy/strategy or groups of policies/strategies that will be analyzed?
  - What are the general types of uncertainties that may affect the proposed policy/strategy?
  - Briefly describe the core model that will be used (type of model; model platform; run times)
  - What measures (model output) will be used to assess the performance of the policy/strategy?
- Staff and staff representatives who would attend the workshops
- Explanation of how API development will be conducted (in-house or using consultants)
- Explanation of how model runs will be completed
- Schedule for:
  - On-site scoping workshop and complete scoping
  - API development
  - Completion of model runs and production of meta-models
  - On-site analysis workshop

A sample schedule for the beta-test is shown in Table 2.

[2] https://tmip.org/content/tmip-webinar-introducing-exploratory-modeling-and-analysis-tool-tmip-emat.

Table 2: Example Beta-Test Schedule

Period Ending:	5/7/2019	5/21/2019	6/4/2019	6/18/2019	7/2/2019	7/16/2019	7/30/2019	8/13/2019	8/27/2019	9/10/2019	9/24/2019
Task 1: Define the model scope											
Experiment Definition	■	■									
Metric Definition	■	■									
Univariate Sensitivity Tests		■	■	■	■						
Task 2: API Development											
Environment Configuration	■										
Model Programmatic Support	■	■									
Python Development	■	■	■								
Test API				■	■						
Task 3: Run the model											
System Configuration and Run						■	■				
Develop Meta-models								■			
Task 4: Analysis											
Visualization									■	■	
Scenario Discovery									■	■	
Robust Search									■	■	
Analysis Workshop											•

There were eight agencies that expressed interest in becoming beta-testers. The selected agencies were chosen to represent a variety of regions, travel demand model software, and applications (see Table 1).

2.3 Scoping Workshop

The purpose of the scoping workshop was to orient the agency modelers and planners to think about the model from an exploratory, rather than predictive, process and to specify how TMIP-EMAT and the agency core model could be used to help build a robust set of strategies.

There are four major steps to the scoping process:

- Part 1: Develop High-Level Scopes. High-level scopes translate from the agency goals and objectives to levers that could support the goal, an enumeration of the metrics that would assess the value of the lever, and exogenous uncertainties that may affect the efficacy of the lever.
- Part 2: Identify Model Functionality. This step takes the high-level scope and matches it against the capabilities of the existing core model regarding the levers and the exogenous uncertainties that can be represented and the metrics that can be produced. The result of this step is a description of how the existing model capabilities will be utilized and/or what new functionality will be developed.

- Part 3: Select and Define Scope. The core model to use and specific levers, metrics, and uncertainties that will be represented in the model are specified as part of this step.
- Part 4: Implementation Workplan. The levers, metrics, and uncertainties are specified in the TMIP-EMAT formatted scope file. Enhancements to the model and extensions of the TMIP-EMAT API are developed.

2.4 Deployment Support

Between the scoping and analysis workshop, the FHWA consultants hosted bi-weekly conference calls and reviewed progress against the task list shown in Table 3. Each of these tasks are described in more detail in this section.

Table 3: TMIP-EMAT Beta-Test Task List

Task 1: Define the model scope	
1.1	Experiment Definition
1.1.1	Specify model variables (levers and uncertainties)
1.1.2	Specify variable ranges / categories
1.1.3	Define programmatic mechanism to set variable
1.2	Metric Definition
1.2.1	Specify metrics
1.2.2	Define programmatic mechanism to report metric
1.3	Univariate sensitivity tests
Task 2: API Development	
2.1	**Environment Configuration**
2.1.1	Install Anaconda and Python Packages
2.1.2	Confirm correct operation through RoadTest example
2.1.3	Develop scope YAML file
2.2	**Model Programmatic Support**
2.2.1	Script / batch file to programmatically run model
2.2.2	Macro to restore defaults
2.2.3	Macro to set experiment variables
2.2.4	Macro to produce metrics
2.3	**Python Development**
2.3.1	Create instantiation of FilesCoreModel class
2.3.2	Implement run method to launch model
2.3.3	Implement method to set experiment variables
2.3.4	Implement method to produce metrics
2.3.5	Implement method to archive model
2.4	**Test API**
2.4.1	Design univariate sensitivity tests
2.4.2	Run model and compare to manual runs
Task 3: Run the model	
3.1	System Configuration and Run
3.1.1	Identify run system workloads and archive location
3.1.2	Configure model and EMAT on each system
3.1.3	Design experiments on each system
3.1.4	Run subset of experiments on each system
3.2	Develop Meta-models
3.2.1	Identify analysis system and install EMAT
3.2.2	Design experiments and import from archive
3.2.3	Derive meta-models and review parameters
Task 4: Analysis	
4.1	Visualization
4.2	Scenario Discovery
4.3	Directed and Robust Search

2.4.1 Task 1: Define the Model Scope

The scoping workshop produced a rough scope specification that needed to be refined as the agency investigated the model operation in more detail and conducted sensitivity tests. The first task then of the beta-test was to fill in any remaining details of the scope. The details to consider in a scope specification are:

- **The exact model input variables** that will be changed to represent the scoped levers and uncertainty variables. Input variables can have categorical values, binary values, or continuous values. Note that each individual input variable cannot vary in more than one dimension, e.g., land use changes would have to be translated from a single variable or set of categories (low growth, medium growth, high growth); representing independent development areas in a region would require multiple input variables in TMIP-EMAT to be represented.
- **The range and categories for the input variables.** This may involve conducting a literature review (e.g., for the potential impact of new technology) or summary of historical data. For example, SANDAG examined the historical border crossing wait times in order to specify the minimum and maximum values. Where data is not available, the range should be determined by discussing the expected impacts with local experts and stakeholders. Ranges should be set in consideration of the current model input data. For example, GBNRTC reviewed the available parcels along the Bailey corridor and calculated the maximum possible population and employment land use.
- **The uncertainty variable distributions**. The input variable distributions represent some information about how likely the uncertainty variable will take a given value within the defined range. If there are areas of the range that are less likely than others, setting the input distribution will produce output statistics (e.g., confidence intervals) that incorporate this information. However, if there is deep uncertainty about a given variable, then any value within the range is equally likely and a uniform distribution is most appropriate—in these cases, the risk-analysis type output statistics are not really applicable. The shape of the input variable distribution influences how experiments are designed to support the meta-models and simulation results. Note that lever inputs default to a uniform distribution.
- **The metrics to be captured from the model.** The metrics identified should represent the expected impacts of the input variables. An expanded set of metrics may be defined to test for correct model functionality, i.e. to report outputs that should not respond to changes in the input variables.

  Once the variables, input ranges, and metrics are specified, univariate sensitivity testing may begin. The default univariate tests are run by walking each variable through the extreme values or each categorical value while all other variables are set to the default value. TMIP-EMAT users should expect that the univariate sensitivity tests will prompt some revisions to the input variable ranges, input variables themselves, and/or metrics:

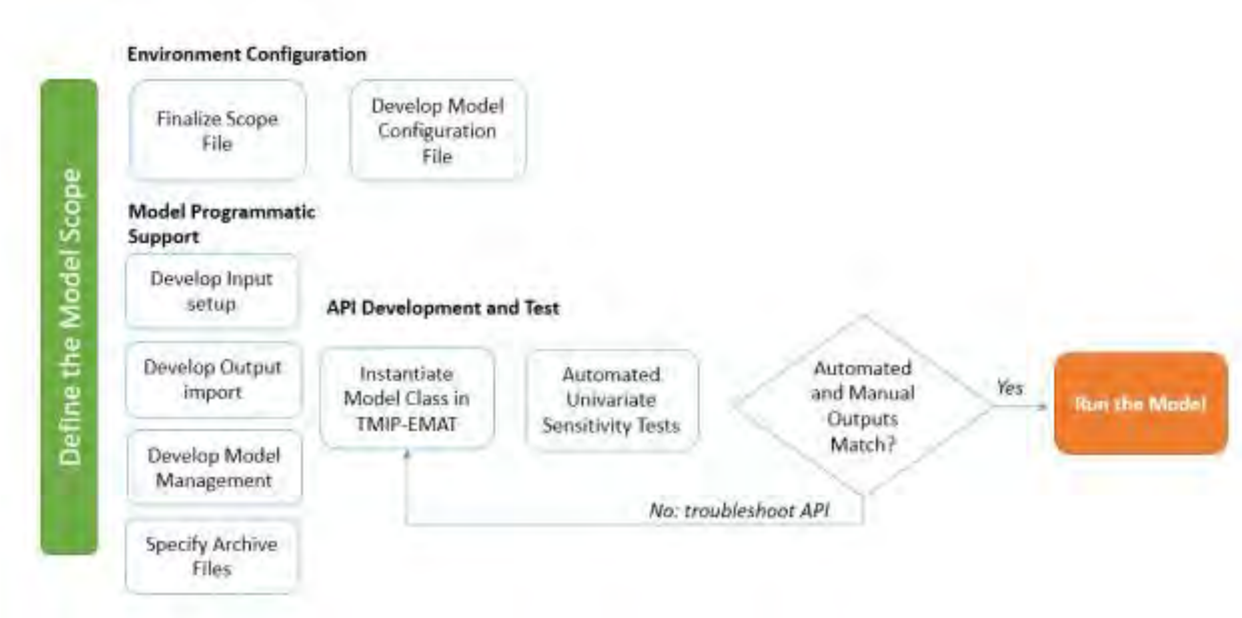
- If the model response is not reasonable, the variable or range may be changed.
- If the metric does not adequately capture the model response, a refined metric may be developed.

The univariate tests are a critical part of the process. If the model results are not reasonable and meaningful in response to changes in individual variables, a combined set of changes will only obfuscate the responses and could lead to incorrect interpretations.

The other major purpose of this initial task is to sketch out the steps required to programmatically set the input variables and generate the metrics. It is advisable to begin this role early in the deployment process to identify variables or metrics that may be particularly difficult to access in a programmatic fashion. Notably, the majority of the script development effort conducted during the beta-tests was in the programmatic mechanisms to specify input variables and collect metrics.

2.4.2 Task 2: API Development

The API development task covers both the installation and configuration of EMAT as well as the development of the model functionality to be programmatically controlled. This task is summarized in Figure 2.



Source: FHWA

Figure 2: TMIP-EMAT Deployment Steps

Environment Configuration

Each beta-tester began this task using the demonstration "Road Test" model that is included in TMIP-EMAT to verify that their Anaconda environment was set up with the correct package versions. The installation process and guidelines on how to "fork" the repository are described in the user documentation.

The scope file can be developed in advance of the programmatic support being developed. The advantage to building a scope file in advance is so that the variable and metric names are defined and can be referenced through the script development. TMIP-EMAT utilizes the string names from the scope file to pass the input variable values through the API and specify the metrics to import.

Model Programmatic Support

The areas requiring programmatic support are:

- To launch and run the model, with the capability to return completion status (success or error);
- To initialize a model run, in the case that a single scenario is modified each time the default values must be restored between experiments;

- To set each input variable (levers and uncertainties); and
- To generate the metrics in a format that can be parsed by TMIP-EMAT (csv file format).

This functionality can be implemented in a programming language that is most convenient to the core model platform. For example, GBNRTC extended their TransCAD GISDK macros with this functionality while SANDAG implemented python scripts to launch and control the Emme model.

Python Development

This section encompasses the script development within the TMIP-EMAT API. The beta-testers all utilized core models that operate against a set of file inputs and generate file outputs, as opposed to a model that is entirely contained within Excel, for example. Therefore, each beta-tester needed to instantiate the FilesCoreModel Abstract Base Class specified in TMIP-EMAT. This class contains several methods that need to be connected to the associated model programmatic support:

- A setup method that will call the process to restore defaults in the core model and set the input variable values;
- A run method that will launch the model;
- A method to produce metrics in format that can be parsed by TMIP-EMAT (the method to import the metrics is generic to core models and only the file name and metric row/column location need be specified in the API); and
- A method to archive the model outputs. The archive step is included to support troubleshooting as well as allow for the later definition of new metrics to be imported.

2.4.3 Task 3: Run the Model

At this point, the beta-tester moves into production and runs the set of experiments through TMIP-EMAT, potentially on multiple workstations, and archives the results. Note that the experiment design (subtask 3.1.3) and meta-model development (subtask 3.2) are automated utilities of TMIP-EMAT and are listed as separate subtasks to identify when they are able to be executed.

Meta-models are derived within TMIP-EMAT for the purpose of supporting the Monte Carlo simulation and exploratory analysis tools. The parameters from the meta-models can also be analyzed directly to evaluate the reasonableness of the travel demand model. For example, in a linear regression where the dependent variable is vehicle-miles traveled and an independent variable (policy lever/uncertainty variable) is highway capacity, we would expect the coefficient estimate for that variable to be positive with reasonable magnitude representing sensitivity to highway capacity.

The meta-model goodness of fit statistics (linear regression R-square and Gaussian Process Regression (GPR) cross-validation) can also be utilized to identify cases where an input variable is not behaving as expected or a metric is not well defined.

2.4.4 Task 4: Analysis

The beta-testers first confirmed the ability to generate the visualization, scenario discovery, and robust search utilizing their meta-models. The actual analysis was conducted as part of the Analysis workshop and is discussed in the following section.

2.5 Feedback and Analysis Workshop

The purpose of this workshop was to bring back together the stakeholders from the scoping workshop and demonstrate the analysis capabilities of TMIP-EMAT. The group then applied these capabilities to the beta-test results and assessed whether the outputs produced new insights for the policy questions defined. FHWA and their consultants also collected general feedback on the overall beta-test process.

2.5.1 Overview and Beta-Test Experience

The analysis workshop opened with an overview of TMIP-EMAT to serve as a reminder on the process, how the meta-models are constructed, and how TMIP-EMAT can produce thousands of simulated experiments. The group next discussed the beta-test deployment including a review of the scope that was defined and how it was refined, model extensions required, and sensitivity test results. The sensitivity tests demonstrated the response of the model to changes for each individual lever and uncertainty variable.

2.5.2 Visualization

The analysis section of the workshop began with a review of the Latin HyperCube Sampling (LHS) experiment results and derived meta-models. The group reviewed scatter plots generated from the LHS experiments to investigate correlation and variation of each metric by input variables. The correlations are summarized at a higher level with through the feature scoring table. The meta-model statistics were reviewed to confirm that the subsequent simulation and analysis would be based on well fit models.

Next, the meta-models were utilized to generate 5,000 simulated results. The group then inspected scatterplots of the meta-model outputs and identified where the variable correlations and variations were more clear due to the higher density of outputs. TMIP-EMAT's visualization also supports an exploratory analysis through setting constraints on the input variables and output metrics.

At this point, the workshop began the demonstration and application of EMA Workbench utilities: scenario discovery and directed/robust search.

2.5.3 Scenario Discovery

The scenario discovery Patient Rule Induction Method (PRIM) is applied to identify variable constraints to meet user defined thresholds, i.e. to define "boxes" with different shares of results that meet the defined threshold or not. Reviewing the resulting boxes gives insight into which variables are most important and the tipping points necessary to meet a threshold.

Essentially, PRIM demonstrates the trade-off of type 1 errors (false positives) with type 2 errors (false negatives) occurring for each performance metric threshold. At the beginning of the discovery process, the coverage is high (no type 2 errors) but density (i.e. probability of the performance metric threshold being reached) is low (lots of type 1 errors). The process of PRIM reduces the type 1 errors but increases the type 2 errors until the density is maximized.

The scenario discovery process identifies subsets of the uncertainty and decision spaces. The subset space does not have a probability associated with it, and so it does not necessarily follow that the constrained area is less likely to come to pass. But the subset areas do require a limiting of the scoped lever and uncertainty ranges and imply that it could be "harder" to bring about a future that meets the threshold.

2.5.4 Directed and Robust Search

The directed search algorithm and visualization identifies the impact of different lever combinations on performance metrics and displays the ones that are a Pareto optimal solution in a parallel coordinates plot. The directed search algorithm also can be directed at the exogenous uncertainties to identify the worst-case scenario outcomes given a set of levers. This application allows for an understanding of which scenarios would produce the worst set of metrics.

Another application of the directed search algorithm is as a robust search, where many uncertainty variable scenarios are defined (instead of one) and the policy variable space is searched over for optimal solutions of the performance metrics.

3.0 Beta-Test Experience, Lessons Learned, and TRL-H Assessment

This section presents a summary of the beta-testing effort, the necessary updates to TMIP-EMAT as well as the overall feedback, lessons learned and next steps for the utility development and deployment. The section concludes with an assessment of the TMIP-EMAT TRL-H level.

3.1 Agency Level of Effort

The level of effort to deploy TMIP-EMAT is highly dependent on the structure of the core model and the ease with which functions can be developed to programmatically control the core model. The beta-testers reported levels of effort to develop the API between 100 hours and 400 hours. However, a repeat of the process with a new set of input variables and metrics is anticipated to only require 40-80 hours.

3.2 TMIP-EMAT Updates to Support Beta-Tests

Over the course of the beta-tests, there were many enhancements made to the TMIP-EMAT supporting materials, including: workshop guidance; installation process facilitation; example Jupyter notebooks; and visualization and analysis improvements. Notably, there were no significant or structural changes required in the TMIP-EMAT API design. This demonstrates that the API is sufficiently flexible to support a variety of core model software platforms and applications.

The improvements to the utility Enhancements prompted several new versions of EMAT over the course of the beta-test, from v0.1.2 to v0.2.5.[3]

In some cases, the beta-testers defined policy levers that did not fit well into the existing design, specifically inputs that would best be represented by a discrete-continuous variable type. As an example, a user may want to have a policy lever to add a new transit line or not, and if added, what level of service will be provided on the new transit line. This can be approximated by two levers instead of one, but it may be easier to manage and interpret if only one lever is used. This also could be useful in automating experimental design; for example, for instances of the "no build" experiments, the level of service lever is meaningless, and so fewer "no build" experiments than "build" experiments are needed.

An interim solution to this kind of problem is to use two parameters: one Boolean parameter controlling the build/no build status, and another continuous variable to represent the level of service, which is effectively ignored in a model run if "no build" is selected. The interpretation of these two levers must be done in the bespoke model API or implicitly in the core model itself. To control the design of experiments to over-weight the "build" solution (more experiments are needed to explore the "build" space that contains more diversity than the "no build"), the binary setting for the distribution can be set in the scope with a *p_true* value larger than 0.5.

[3] https://github.com/tmip-emat/tmip-emat/blob/dev/CHANGELOG.md

3.3 Feedback, Lessons Learned and Next Steps

The common beta-tester's feedback and lessons learned, recommended improvements, and next steps are listed here. A more complete set of experiences, lessons learned, recommendations and next steps by beta-tester is provided in the appendix.

TMIP-EMAT Experience Feedback and Lessons Learned

- **Iterative approach**: TMIP-EMAT applications are best conducted as iterative processes with the expectation that the scoped levers, uncertainty variables, and metrics will be refined in response to the sensitivity tests and core model runs. For example, sensitivity tests may show an unreasonable or insubstantial response to a changing input, or the model response may inspire an interest to investigate another aspect of the outputs more closely. All beta-testers reported that they would make further adjustments to their scoped levers, uncertainties, and metrics given more time. Moreover, incorporating TMIP-EMAT at the beginning of the planning process would ensure a higher consistency between the levers to be tested and the modeled representation.
- **Improved modeler confidence:** The beta-testers reported that they were more confident in their model following the execution of the model runs; this implies that the beta-test parameters and visualization features were suitable to isolate and understand how the model responded to a multi-variate change in inputs. This is particularly valuable when working with models with many components, such as ABMs.
- **Effective presentation of outputs.** TMIP-EMAT was found to effectively present the information from the core model and allow analysts to think about the implications of the model assumptions on the results. During the analysis workshop, there was some hesitation to draw real conclusions from the results because there were dominating assumptions. The limitations of these assumptions were more obvious than they would have been with a simpler scenario planning approach. This also highlights the need to proceed with exploratory analysis iteratively, as once the importance of these dominating assumptions became clear, it was obvious that they should be accounted for explicitly in the analysis.
- **Land use challenges.** Representing land use through a continuous uncertainty variable is challenging. Regional models represent land use as segmented spatial data, and defining a continuous change variable required simplifying assumptions and did not always produce insightful results. Instead of a proportional, even growth, it is more likely that certain areas will develop first and other spots will lag behind. Unless sufficient time can be spent to properly develop and test a continuous variable that can handle these intricacies, land use is better represented through multiple variables and/or as a categorical variable with a few land use scenarios.
- **Metrics should be as specific as possible.** In traditional model sensitivity testing and planning applications only a small number of model scenarios are run. Therefore, more aggregate metrics are ideal because there are fewer variations in the input data and insufficient points to understand all of the inner-workings. In a TMIP-EMAT application there are many points across all dimensions of uncertainty and levers, and thus a higher resolution of the model results is available. More disaggregate and highly specified metrics can be analyzed in detail utilizing TMIP-EMAT. These more detailed metrics both better inform how the model is working and supports a more detailed analysis within model application.
- **Define a broad base of metrics.** It is better to have more metrics available for the sensitivity testing and visualization and then restrict them in analysis and application, rather than starting with fewer metrics from the beginning. Adding metrics to an existing scope does not require

more core model runs but still requires effort to develop the summary scripts. In addition, care must be taken to ensure the metrics captured from the model are appropriate, particularly when running a subset of the model. For example, SANDAG only ran their cross-border component but had originally identified region-wide emissions as a metric. Some of the variables would have an effect on regional emissions, but because the entire model was not run, this was not fully represented.

- **Potential use as a default application.** One beta-tester theorized that, similar to the standard forecast that is run for all projects, TMIP-EMAT could be utilized with a standard set of uncertainties with which to test all projects. The advantage of this approach is that a "no-build" set of core model runs could be complete in advance, thus reducing the core model runs necessary when testing an individual project.
- **Run time challenges.** The beta-testers all expressed some degree of concern about the model run time. This was alleviated, somewhat, by the use of multiple systems and TMIP-EMAT's automated processes that allowed for a "set and let run" approach to the entire set of Core Model runs.

Recommendations for TMIP-EMAT Improvements

- **Scoping guidance.** The beta-testers found the scoping workshop and process to be very beneficial but concluded that they highly benefited from the guidance of FHWA and the FHWA consultants to understand the best approach to scoping. An overview or guide, including an example of the scoping process, would be useful for agencies without prior experience in exploratory scoping, to be in the right mindset prior to beginning this critical part of the work.
- **More accessible tutorials.** The learning curve for Jupyter notebooks was steep for new users. Screencasts or similarly accessible tutorials were recommended to complement the materials available through the online documentation.
- **Simplified visualizations.** A common piece of feedback from the beta-testers was to produce a simpler graphic that could be taken to a higher-level decision maker (e.g., MPO board member). The challenge is that these graphics would be highly specific to the application and "story" coming out of the utility and therefore may not be ideal or possible to develop within TMIP-EMAT itself.
- **GIS-based presentation of metrics**. Existing scenario tools have geospatial visualization capabilities. It would be useful to connect TMIP-EMAT metrics to these spatial tools. It may be desirable to set-up metrics such that they could support separate plots/conversations for different geographic levels or population subsets (e.g., urbanized areas, low income, etc.).
- **Risk-based visualizations.** The beta-tests primarily focused on exploratory analysis. Features and examples to support a risk analysis, including reports of confidence intervals and probabilities, would also be useful.
- **Reference scenario in outputs.** Presenting a reference scenario as the anchor of the analysis and visualizations would help orient planners and modelers accustomed to a single point forecast to put the range of TMIP-EMAT outputs into context.
- **Synthesized metric of "acceptable solution".** The directed search outputs would benefit from a calculated metric that represents the threshold of multiple metrics as a measure of "acceptable solution". Although constrained optimization in this manner is possible using the existing tools, additional documentation is required to demonstrate this capability in a way that illustrates multiple-metric constraints.

- **Better control on Scenario Discovery utilities.** The scenario discovery trajectory might start with the reference scenario (likely not meeting the set threshold), rather than the all options implemented scenario (100% in the box), and increasingly add in changes to incrementally add more dimensions in order to get into the desired 'box' range, and then further increase the coverage and density within the box. Also, it would be helpful to set multiple constraints in the scenario discovery process to define scenarios within a subset of the scoped universe of outputs.
- **Procedure to constrain scope in analysis.** The visualizers and analysis utilities key off of the scope file. During the analysis process, users may want to modify input ranges or drop levers, uncertainties, and/or metrics from the original scope. A defined procedure to make the changes to the scope and reinitialize the visualizer would be helpful.
- **Better control on Directed Search utilities display.** It would be useful to constrain the levers and uncertainty ranges for the parallel coordinate plot visualizations. In the parallel coordinate plot, rather than a simple toggle on-off of uncertainties, the user might set the desired elements/levels of exogenous uncertainty to include.
- **Utilize parallel coordinate plots for meta-model results.** An addition use of the parallel coordinate plots would be plotting the Monte Carlo simulation results of the meta-models as an alternative, or in addition, to the interactive histograms and scatter plots. A further enhancement would be an additional interconnected visual tool to the histograms and scatter plots. In other words, a tool that initially shows all the scenarios in the parallel coordinate plot and lets the user limit them. This could allow a deep exploration of the simulation outputs without involving an optimization method.
- **Reduce core model run requirements.** Run time was a major concern for two of the beta-testers. Methods that would reduce the number of required model runs for variables that have little variance would be advantageous.
- **Tighter integration with travel demand model software.** The TMIP-EMAT API is generic to model software, which implies a greater need to develop custom coding in the TDM environment to support the API. An approach that reduces the need for custom coding would require tighter integration of TMIP-EMAT with the travel demand model software packages, or development of the meta-model procedures within the TDM software.
- **Test model uncertainty**. As a complement to the feedback that running TMIP-EMAT provided more confidence to modelers, tests that delve into the uncertainty of the model structures and parameters may be more beneficial than tests varying the input data.

Agency Next Steps with TMIP-EMAT

- **Application with strategic model**. Two of the beta-testers reported that they are interested in running TMIP-EMAT applications with strategic models. Strategic models are attractive because of their reduced run time. A strategic model could be used to develop the preferred scenario for running through the full travel demand model.
- **Continued use with regional model.** Two of the beta-testers reported that they will continue to use TMIP-EMAT with their regional model and intend to integrate the utility into their planning process. The agency that does not foresee further application of TMIP-EMAT with their regional model cited concerns about model run time.

3.4 TMIP-EMAT Technology Readiness Assessment

The Technology Readiness Level for Highway Research (TRL-H) scale measures where a highway-related technology is on a scale of 1 - Basic Research to 9 – Implementation. The advantage of the scale is to indicate the sort of development that may be necessary and/or additional tests to conduct. The scale also serves as a guide to structure discussion to uncover technical gaps/questions and outline next steps in development.

Prior to the beta-test, TMIP-EMAT was assessed to be at a TRL-H Level 5: *Integrated Components Demonstrated in a Laboratory Environment*. This level was based on the completed and documented proof of concept application as well as the end user and integration plan documentation.

The purpose of the beta-tests was to bring TMIP-EMAT through the next two levels:

- TRL-H 6: Prototype Demonstrated in Relevant Environment; and
- TRL-H 7: Prototype Demonstrated in Operational Environment.

As is demonstrated by the discussion below, TMIP-EMAT meets the TRL-H 6 and 7 levels.

3.4.1 TRL-H 6: Prototype Demonstrated in Relevant Environment

The key term of TRL-H 6 is "Relevant Environment," which refers to: *"that of an MPO or consultant project, using real data on a full-size model. The model's recommendations are not yet being used for decision-making."*

Each of the TRL-H 6 questions is discussed below.

Is the operational environment fully known (i.e. user community, physical environment, and input data characteristics as appropriate)?

The key aspect of the operational environment for TMIP-EMAT is the existence of a core model. This core model defines the physical environment and input data.

TMIP-EMAT "input data" is essentially the scope file contents that specify the input variables and their characteristics as well as the output metrics.

The scope file syntax and structure are defined in the utility and script documentation and, with the completion of the beta-tests, there are now several examples for future users to consider.

Metrics output from the core model that TMIP-EMAT can input are single values. The metrics are accessible to TMIP-EMAT either as a row and column index that the utility parsers can use to import from a CSV file or through a custom import definition that could be developed within the API structure. The characteristics of the metrics, however, are dependent on the core model and thus outside the consideration of TMIP-EMAT.

The TMIP-EMAT user community consists of direct core model users and the immediate consumers of the core model outputs. Effective application of TMIP-EMAT requires a detailed understanding of the core model functionality, both to define an appropriate scope and to interpret the relationships in the outputs.

Was the prototype tested in a realistic environment outside the laboratory (i.e. relevant environment)?

Yes, TMIP-EMAT was tested at three agencies utilizing their production model or components of the production model with inputs typical of a production model application.

Does the prototype satisfy all operational requirements when confronted with realistic problems?

TMIP-EMAT handled the seven input variables defined in each of the beta-tests and supported running core model experiments on independent systems, which was crucial for beta-testers with longer model run times to complete the core model runs in advance of the analysis workshop. The analysis tools have been demonstrated with 5,000 simulated outputs and showed that the visualizers, scenario discovery, and robust search tools are able to reasonably respond on a laptop computer.

The primary challenge in the deployment of TMIP-EMAT is to implement programmatic controls over land use or other core model inputs. These controls, however, are outside the scope of TMIP-EMAT and are specific to the core model.

Is the prototype able to assess the relevant policy questions?

The visualization and analysis tools supported a nearly comprehensive investigation of the simulated metrics. Limitations on the ability to assess policy questions rest on the underlying core model functionality.

Are the input data to the models readily available?

As discussed above, TMIP-EMAT input data consists of the scope definition and the core model metrics.

The scope definition includes the range and distribution of variables. Defining a "reasonable" range may be difficult, especially for variables that have deep uncertainty (e.g. impact on capacity due to autonomous vehicles). However, the advantage of TMIP-EMAT is that the ranges can be adjusted during analysis to reflect new information and use of a range reduces the importance of getting specific values correct.

The metrics are, by definition, readily available because they are produced by a core model that is a prerequisite to deploying TMIP-EMAT. The API does require a non-trivial development effort (between 2-3 months depending on the structure of the core model and complexity of the scoped input variables and metrics).

Can the models be adequately calibrated and validated? How do the results compare with the results of traditional models?

The primary validation statistics from the meta-models are the R-squared values from the linear regression (LR) and cross-validation statistics from the GPR. TMIP-EMAT does not support calibration of the meta-models. Rather, a poor LR or GPR statistic indicates an issue with the output metric or unreasonable response within the core model. Because the meta-models include all input variables that are changing in the core model, there are no other unobservable effects to be fit.

Are the requirements for supporting software and hardware reasonable?

TMIP-EMAT requires the installation of the latest version of Anaconda Python and can be setup as an environment within Anaconda. This allows the user to have different Python package versions on the same machine. All software required is available as open-source, although

Administrative access may be required to complete the installation. TMIP-EMAT can be run on Windows and Linux operating systems.

The core model is likely to have a greater hardware requirement than TMIP-EMAT. Analyses for the beta-testers were all conducted on mid-range laptops without issue.

3.4.2 TRL-H 7: Prototype Demonstrated in Operational Environment

As opposed to TRL-H 6, level 7 requires an "Operational Environment", which *"is provided when the system is 'owned' and run by the MPO/State DOT. Those running the model have sufficient technical expertise and are using real-data on a full-size model with the intent of using results of the model for decision-making."*

Each of the TRL-H 7 questions is discussed below.

Are available components representative of production components?

TMIP-EMAT is currently run through Jupyter Notebooks and will continue to be operated in this manner through production. The major steps are all complete (scoping; meta-model development; and experiment design) and operate as they will in production. There may be other supporting features included (e.g., a method to port experiment outputs to a new scope definition automatically), but these will operate in a similar manner through Jupyter Notebooks.

Visualizations based in Jupyter Notebooks are recently developed and under continuous improvement. TMIP-EMAT is an open-source utility and thus leverages open-source visualization and analytical tools. It is likely that continued use of TMIP-EMAT will prompt the further development of new visualizations.

Is the fully integrated prototype demonstrated in an operational environment (i.e. real world conditions, including the user community)? Has it been run by an end user (MPO, State DOT) on real data?

The beta-tests all involved production models and were run by Agency modelers on their local workstations. The input data was developed specifically for this beta-test but was consistent with real world data structures and values. Specifically:

- The input data for ODOT was fabricated to produce a more interesting model result (forecast growth was increased). But the models and data formats are all exactly as they would be under operational circumstances.
- SANDAG does not typically operate only their cross-border model component, but the variables and metrics of the beta-test were appropriate for running only this component, and the analysis would not have changed significantly if the full model had been run.
- GBNRTC ran the full model for the beta-test. The shared mobility and weather impacts variables were extensions of the normal operation of the model.

Are all interfaces tested individually under stressed and anomalous conditions?

Utility initiation (load scope, create the database), experiment design, core model runs, and meta-model development have been run on workstations in at least four different firms at this point.

The analysis notebooks have not been as widely exercised; these were run on FHWA consultant laptops for the workshops although the notebooks have all been provided to the beta-testers, and no issues have been reported.

The analysis process operates well with 5,000 simulated experiments. A large increase in the number of experiments would cause the visualizations to run more slowly and potentially halt the system. The parallel coordinates plots operate well for about 1-2 hours of interaction. There have been occasional lock ups that require a restart of the Jupyter notebook kernel and a reloading of the notebook. The directed and robust search can and have been run with caching, which allows the modeler to quickly reload the parallel coordinates plots without having to rerun the optimization, which can take 5-30 minutes.

Have model configuration, inputs, and outputs to run in operational environment been documented?

The scope definition (configuration) and API (inputs and outputs) are documented through the online documentation, as part of the Python scripts, and by way of example implementations available on the GitHub repository. Furthermore, the notebooks to conduct all steps of the utility operation, including restoring from a failed core model run and running independent experiments on distributed systems, are available through the online documentation.

4.0 References

TMIP-EMAT Source Code Repository: https://github.com/tmip-emat/tmip-emat

TMIP-EMAT Online Documentation: https://tmip-emat.github.io/source/emat.intro.html

5.0 Appendices

A.1 *Greater Buffalo Niagara Regional Transportation Council Test Summary*

Disclaimer: Greater Buffalo Niagara Regional Transportation Council (GBNRTC) conducted the TMIP-EMAT beta-test using the GBNRTC's four-step travel demand model to evaluate policy and investment uncertainties associated with improvements along a corridor. To help achieve that purpose a realistic, but fictitious, set of transit improvements and other inputs was developed. As part of this beta test, a couple of sensitivity testing oddities with regard to transit improvements were investigated. Information in this data and analysis serves as an example for how to use TMIP-EMAT using realistic data. This dataset and analysis should not be used to draw any specific conclusions about transportation policy's impact on system performance and outcomes.

Greater Buffalo Niagara Regional Transportation Council (GBNRTC) conducted a TMIP-EMAT beta-test to explore the possibility of using GBNRTC's four-step travel demand model in conjunction with TMIP-EMAT to evaluate investments along a corridor, with the Bailey Avenue Corridor as a prime example.

Buffalo's Bailey Avenue corridor extends for 7.5 miles across a variety of places, including dense residential neighborhoods, a major university and veterans hospital, manufacturing and other industries, and a food terminal with rail line connections. Like many corridors in "Rust Belt" cities, Bailey Avenue has experienced commercial decline, areas of disinvestment and vacancy, and aging and overbuilt infrastructure. Sections of the corridor can be unsafe for pedestrians, cyclists and drivers. Some parts are congested at peak travel times, which causes delays for all users—including public transit buses. And while this is one of the highest ridership, highest frequency Niagara Frontier Transportation Authority (NFTA)-Metro bus routes in the region, buses often experience delays at peak times, and transit enhancements are needed—including those that address first-last mile connections.

However, a number of opportunities are arising. The new regional transportation plan, Moving Forward 2050, developed a framework for next generation corridors like Bailey Avenue that integrates technology to improve mobility. In proximity to several light rail stations, Bailey Avenue would benefit from an expected expansion of the region's light rail system. NFTA-Metro has identified Bailey Avenue as a possible Bus Rapid Transit (BRT) corridor. Additionally, Empire State Development, New York State's economic development organization, identifies Bailey Avenue as a priority investment corridor. And local community development organizations and foundations are investing in schools, housing, and façade improvements along Bailey Avenue.

GBNRTC seeks to identify a set of transportation and land use investment recommendations for long-term, holistic approaches to improving the corridor in terms of economic and community development, safety, workforce access and overall mobility. These recommendations should utilize emerging technology and new mobility options—even if not yet necessarily "proven" or fully tested—and should address new types of funding and new forms of governance. The final product will help stakeholders and decision makers identify the best alternatives for implementation.

GBNRTC intends for the Bailey Avenue corridor project to be a transformative project that improves mobility, provides a host of safety improvements for all road users, dramatically improves the streetscape, supports community and economic development, and contributes to the region's revitalization. This corridor should serve as a model for other initiatives in the Buffalo Niagara region, as well as for other post-industrial Rust Belt regions.

The GBNRTC model was utilized during for the TMIP-EMAT proof of concept, which demonstrated the application with regional uncertainties. For this beta-test, GBNRTC was interested in enhancing their corridor-level project analysis through the application of TMIP-EMAT and in developing the skills to run TMIP-EMAT in house. GBNRTC wanted to test planning policies (levers) such as improved transit and mobility options against uncertainties such as land use, emerging technology and inclement weather.

A.1.1 FHWA Workshops and Support

FHWA produced two workshops to support EMAT scoping and analysis as well as bi-weekly calls to answer questions, lay out next steps, share information across the beta-testers, and collect information on the experience using TMIP-EMAT.

Scoping Workshop

The scoping workshop was held at GBNRTC's office in Buffalo, NY on May 7, 2019 from 8:00AM to 3:00PM. The workshop participants are identified in Table 4: GBNRTC Scoping Workshop Participants. The participants included representatives from GBNRTC's modeling and planning groups as well as their modeling consultant.

Table 4: GBNRTC Scoping Workshop Participants

Person	Affiliation
Matthew Grabau	GBNRTC
Amy Weymouth-Michaux	GBNRTC
Michael Davis	GBNRTC
Lisa Kenney	GBNRTC
Kimberley Smith	GBNRTC
Hal Morse	GBNRTC
Andrew Bartlett	Niagara International Transportation Technology Coalition (NITTEC)
John Lewis	GBNRTC Modeling Consultant
Sarah Sun	FHWA
Maria Chau	FHWA
Martin Milkovits	FHWA TMIP-EMAT Consultant
Rachel Copperman	FHWA TMIP-EMAT Consultant

Analysis Workshop

The analysis workshop was held at GBNRTC's office in Buffalo, NY on October 8, 2019 from 8:00AM to 3:00PM. The workshop participants are identified in Table 16. The participants included representatives from GBNRTC's modeling and planning groups.

Table 5: GBNRTC Analysis Workshop Participants

Person	Affiliation
Matthew Grabau	GBNRTC
Amy Weymouth-Michaux	GBNRTC
Michael Davis	GBNRTC
Lisa Kenney	GBNRTC
Kimberley Smith	GBNRTC
Hal Morse	GBNRTC
Sarah Sun	FHWA
Maria Chau	FHWA
Martin Milkovits	FHWA TMIP-EMAT Consultant
Rachel Copperman	FHWA TMIP-EMAT Consultant
Jeffrey Newman	FHWA TMIP-EMAT Consultant

Bi-weekly Calls

FHWA and their consultants hosted calls with GBNRTC at least every two weeks between the scoping and analysis workshops. Matthew Grabau was the main GBNRTC participant on these calls and reported progress against the task list shown in Table 3.

A.1.2 Model

The GBNRTC model is a traditional four-step model with a time-of-day and vehicle availability component. GBNRTC uses their model to evaluate highway and transit projects in their region. The GBNRTC model is implemented in TransCAD 6.0 and runs on a standard workstation in about 2 hours.

A.1.3 Scope

This section presents the materials developed during the scoping workshop and concludes with the selected EMAT scope.

High-Level Scoping

In the workshop, the group started with the four-part focus areas defined in the Moving Forward 2050 RTP:[4] *"Moving Forward 2050 aims to use transportation investments to strengthen communities and focus growth where we already have infrastructure, create economic development, and support workforce access. The plan also looks at ways to improve mobility using technology, and aims to protect our natural environment by using sustainable materials and innovative design features."*

During the scoping these high-level goals were translated into policy lever outcomes that would support each goal. The outcomes were selected with the mind of what could be tested through the regional model:

- Strengthen Communities
    - Increase accessibility to influence land use
    - Increase access to services for general population and communities of concern (high-poverty zip codes)
    - Increase multi-modal access to neighborhood services
    - Increase active transportation options
    - System safety for all modes
    - Improve access to parks, greenways, waterfronts
- Create Economic Development and Support Workforce Access
    - Reduce freight delays
- Protect the Natural Environment
    - Decrease lane-miles with under-utilized capacity
    - Decrease impervious surfaces
    - Decrease VMT
- Improve Mobility using Technology and Innovation
    - Increase lane miles of connected corridors
    - Improve reliability

The group then discussed a number of policy, project, treatment, and strategy levers that would support these outcomes along a corridor or within the region.

Corridor-level levers:

[4] https://www.gbnrtc.org/movingforward2050

- Transit and non-motorized improvements
  - Complete streets with potential higher-transit service (BRT)
  - Mobility hubs
  - TSP and bus priority
  - Cycle track
- Land use
  - Encourage redevelopment
  - Densification of land use
- Freight
  - Bi-national, green, autonomous, freight-corridor
  - Distribution centers
- Roadway improvements
  - Adaptive signal control in coordination with highways (e.g. 190)
  - Support for new vehicle technology (fully autonomous / mixed / connected)

Region-wide:

- Increase pre-trip information
- Shared-use mobility services
- Regional cycle network
- Mobility as a Service (MaaS)

Three corridors were then discussed in more detail with regard to the specific set of investments:

- Twin City
  - Roadway reconfiguration: combine 425 with Division or repurpose 425 for multi-modal use
  - Promote non-motorized modes - Twin City corridor is the regional connection for cycle network (Albany to Buffalo)
  - Enhance transit and multi-modality on corridor
- Bailey Avenue
  - Enhanced Transit
    - BRT – with lane (headway and travel time)
    - Without lane (partial dedicated)
    - No change (lower level of travel time)
    - Transit mobility hubs

  - Parking changes (restricted on-street and system controlled)
- Smart-Lane Enhanced Mobility Arterials (SEMA)
  - Support new vehicle technology
  - Operate at higher roadway speeds
  - Transit priority
  - Flexible curb-space

The group then discussed the uncertainties that could affect the ability of each set of corridor investments would have on reaching the high-level goals and performance metric outcomes:

- Land Use: shift in employment locations and density of new development
- Demographics: aging population and income distributions
- Vehicle Technology: mix of connected, automated technology available and capabilities
- Mobility Services including TNCs and micro-mobility
- Climate and weather impact on non-motorized modes
- International travel demand in response to change in the currency exchange and how the border crossings will operate.

Identify Model Functionality

The next step in the workshop was to identify the existing or needed model functionality to represent each lever and uncertainty variable in the model. Table 6 identifies all potential model variables for each lever and uncertainty. As part of the discussion, the group discussed the level of effort involved in developing new model functionality.

Table 6: GBNRTC Model Variable Identification by Lever

Lever	Potential Model Variables to Represent Lever
Roadway reconfiguration	• Highway network geometry • Lane configuration • Functional class and capacity attributes • Speeds • Intersection delay
Promote non-motorized modes	• Make links available for walk and bike modes • Centroid connector number and location • Non-motorized speeds (micro-mobility)
Mobility Hubs	• Add PnR availability at key stops

Lever	Potential Model Variables to Represent Lever
Transit Enhancements	• Travel time improvements (TSP and BRT) • Headway improvements • Stop frequency • Access / Egress improvements • Add new transit route
Parking Policies	• Terminal times (represent less on-street parking)

Table 7: GBNRTC Model Variable Identification by Exogenous Uncertainty

Exogenous Uncertainty	Potential Model Variables to Represent Uncertainty
Land Use / Demographics	• Employment level by segment (retail, wholesale, manufacturing, government, service, office). • Household size and income segmentation • Number of households by location along corridor • School / university enrollment • Development at key areas (e.g. Genesee node on Bailey)
Vehicle Technology	Supply side changes • Roadway capacity • Intersection delay • Parking costs and terminal times to represent self-parking • Electric vehicle reductions in operating costs Demand side changes • In-vehicle travel time sensitivity • Zero Occupancy Vehicle travel generated as separate trip table • New mobility services represented through vehicle availability levels

Exogenous Uncertainty	Potential Model Variables to Represent Uncertainty
Climate / Weather	• Decrease walk speed • Non-motorized distance threshold • Increase transit IVT • Decrease roadway speeds / capacities • Reduce parking capacity
International	• Increased shopping trips across border • Change enplanements at BUF airport • Border crossing availability / capacity / delay

Select and Define Scope

After considering the work required to implement each lever, uncertainty variable and metric, the group selected the model variables that would be leveraged through TMIP-EMAT keeping in mind that the total number of policy-levers and independent uncertainty variables have a linear relationship with the required number of Core Model runs, for example: 10 policy-levers/uncertainty variables require 100 Core Model runs, while the number of metrics has no impact on the number of Core Model runs required.

The scope was revised through the subsequent steps. The main scoping change was to focus only on the Bailey Avenue Corridor. GBNRTC identified four levers associated with the corridor: 1) Improved Transit Headway; 2) Micro-mobility options; 3) Mobility hubs; and 4) Reduced Parking. They also identified four uncertainties: 1) Land-Use; 2) Self-parking vehicles; 3) Shared mobility; 4) Inclement weather. Table 21 summarizes the selected levers and exogenous uncertainties. Originally, nine levers and uncertainties were scoped. Through the development and testing process, two levers were dropped because of unreasonable or insubstantial responses in the model.

Table 8: GBNRTC Selected Levers and Uncertainty Variables

Policy-Lever/Uncertainty Variable	Minimum	Default	Maximum	Distribution (applies to Exogenous Uncertainties only)	Unit/Correlations/Other Notes
LEVER: Transit Headway	True	False		NA	Half the headway on Bailey Ave. routes
LEVER: Micro-mobility		False	True	NA	Improved access to transit stops; Higher density of transit stops along Bailey Corridor
LEVER: Mobility Hubs		False	True	NA	Every other stop on Bailey Ave. is a PNR lot
LEVER: Reduced Parking		False	True	NA	Parking on Bailey is moved to side streets by increasing terminal time for auto
EX. UNCERTAINTY: Bailey Land Use	0	0	1	uniform	0 = base 2025 forecast; 1 = Full build out of vacant lots along corridor
EX. UNCERTAINTY: Self Parking	False	False	True	binary	True = all terminal times are set to zero; False = base model terminal times related to land-use density
EX. UNCERTAINTY: Shared Mobility	0	0	1	uniform	0 = calibrated distribution of zero and insufficient vehicle households; 1 = all households are treated as having sufficient vehicles
EX. UNCERTAINTY: Weather Impacts	0	0	1	Binary w/ 90% = 0; 10% = 1	0 = base capacity and walk speed; 1 = 75% decrease in highway capacity, walk speed

The scoped metrics and their optimization goal are shown in Table 22. The metrics are a mix of corridor level outputs, based on the links representing Bailey Avenue or by segmenting the trip table to capture trips to and from Bailey, and regional level outputs.

Table 9: GBNRTC Scoped Metrics

Metric	Optimization Hint
Bailey VMT	minimize
Bailey VHT	minimize
Bailey Delay AM	minimize
Bailey Delay PM	minimize
Bailey Delay MD	minimize
Bailey Delay NT	minimize
Regional trips to/from Bailey	maximize
Bailey corridor route ridership	maximize
Bailey Transit share	maximize
Bailey NonMotorized share	maximize
Employment < 20 transit mins from Bailey	maximize
Region-wide VMT	minimize
Total Transit Boardings	maximize
Peak Transit Share	maximize
Peak NonMotorized Share	maximize
Off-Peak Transit Share	maximize
Off-Peak NonMotorized Share	maximize
Daily Transit Share	maximize
Daily NonMotorized Share	maximize

A.1.4 Implementation and Model Run Experience Report and Feedback

This section summarizes the TMIP-EMAT API development, ABM model extensions, and conducting the core-model runs.

API Development

The GBNRTC model was utilized for the TMIP-EMAT proof of concept so an API was already developed and the additional work was to develop the model functionality to set the new levers and uncertainty variables and to summarize the new metrics. GBNRTC leveraged their model development consultant to develop the new model functionality. GBNRTC found the process of utilizing a consultant familiar with their model to be a great benefit and vital to a successful beta-test deployment.

Core Model Runs

GBNRTC conducted the sensitivity tests and core model runs on a single machine. A subset of model runs were identified that could complete during off-work hours (overnight and weekend).

In preparing for the analysis workshop, GBNRTC and FHWA concluded that the performance of the Transit TSP and BRT strategies were skewing the scenario discovery and directed search tools enough to be counter-productive for the workshop. The issues with these levers were noted during the univariate sensitivity testing, but the LHS core model runs were initially conducted with the levers included. Therefore, a new set of core model runs was kicked off with insufficient time to complete prior to the workshop. However, the statistical fit of the meta-models was high enough to conduct the analysis, even though only 75% (53 of 70) of the core model runs were complete.

Run and Setup Issues

The GBNRTC micro-mobility, transit TSP, and BRT levers were represented through changes in the transit and highway networks (note that the transit TSP and BRT levers were removed from the final analysis due to issues with the transit representation). Representing the different permutations of these three levers required four different networks (TSP and BRT were exclusive). Coordinating changes across all the networks and managing the files was an opportunity for error and required careful testing. Another challenge related to the different networks was in calculating the metrics. Changes in link or transit routes between networks meant that the metric needed to handle all cases.

GBNRTC experienced an issue with the archive step not locating the network drive. Jupyter notebook must be launched with the same privileges that mapped the network drive.

A.1.5 Analysis Feedback

This section summarizes GBNRTC's feedback and discussion by analysis type.

Univariate Sensitivity Test Review

The univariate sensitivity tests were useful as a first test in assessing the reasonableness of the levers, uncertainties, and metrics. If something did not look right during the sensitivity tests GBNRTC went back and adjusted inputs and variables.

The univariate sensitivity tests were helpful to identify low model sensitivity (walk speed) and unreasonable results (transit TSP and BRT levers).

Visualization and Feature Scoring using Meta-Model Results

As discussed above, only 75% of the core model runs were complete by the analysis workshop, but the meta-model statistics showed a high goodness of fit and it was concluded that the analysis could proceed with the available data.

Overall GBNRTC found Feature Scoring to be a useful visualizer. The visualizations and feature scoring were useful for showing the "30,000 foot" perspective of how the levers and uncertainties affected the performance metrics. It helps explain what lever/uncertainty is dominating the results. The group discussed that it may be useful to "turn off" the dominating lever/uncertainty to demonstrate how other levers/uncertainties are affecting the performance metrics more clearly. GBNRTC recommended using the feature scoring results to inform a larger story that would be told to decision-makers, rather than using it directly.

The "Lasso" feature was presented to GBNRTC that allowed for the user to select a set of scenarios and scroll through different plots to see where they lie within the range of other metrics, uncertainties, and levers. This feature allowed for assessing best/worst case outcomes and outliers, understand what other uncertainties or levers have the biggest impact on metrics, and assess which levers/uncertainties are driving the interaction w/ other levers/uncertainties (i.e. driving outliers).

GBNRTC found the scatterplots and histograms as potential visualizations that could be pulled out and shown to policy-makers and the public. GBNRTC suggested that it may be useful to set the histogram sliders to narrower ranges of levers and uncertainties and redo plots with only the narrower range (i.e. no blue bars and dots). They did find the scatterplots to be useful in understanding the interactions between uncertainties and levers which cannot be done in traditional planning.

Scenario Discovery and Directed / Robust Search

These tools were impressive, but not found to be directly informative for the corridor level measures. In part because of the strong correlation between the input variables and metrics and dominance of the shared mobility uncertainty.

A.1.6 Level of Effort

Developing new model inputs (networks) and the supporting macros required approximately 60 hours of the GBNRTC model consultant time. GBNRTC spent approximately 12 hours developing the maximum land use level, a process that required examining the available parcels along the corridor.

GBNRTC estimated that about 40 hours were spent outside of the workshops to refine the scope file process, experiment with Jupyter notebook, and oversee the model runs.

A.1.7 Lessons Learned and Overall Feedback

Overall GBNRTC found TMIP-EMAT to be a useful tool that they would utilize again. They described TMIP-EMAT as being a valuable tool in their toolbox (that contains other tools as well). Ultimately, the TMIP-EMAT process and application resulted in GBNRTC having a better understanding of the regional model, which is in and of itself beneficial. It also provided insight on how the GBNRTC model can be improved during the next model update (i.e. what functions/capabilities should be added to the model).

Scoping

GBNRTC concluded that TMIP-EMAT would be most useful on a planning project if used from the beginning to ensure that the scoping matches the project goals, objectives, policies, and metrics closely. That way what is analyzed with TMIP-EMAT truly matches what the planners and policy-makers are interested in understanding.

The scoping exercise made very clear the capabilities of the regional model. The scoping workshop was effective at identifying the model capabilities and aspects that cannot be represented in the model.

GBNRTC suggested that a set of uncertainties could be scoped and then used in default application for a variety of projects. Region-wide and high-level uncertainties are similar across various transportation investments and policies, and so having these same uncertainties analyzed across these projects would be worthwhile to support a robust-decision making process.

GBNRTC also thought it would be valuable to have scoping that focused mainly on alternative land-use scenarios given the uncertainty that surrounds development in a region.

Set-up and running TMIP-EMAT

GBNRTC found that there is a steep learning curve for utilizing and working with Jupyter Notebooks. GBNRTC recognized the usefulness of Jupyter Notebooks and believes that this understanding is within their capabilities, but making TMIP-EMAT more user-friendly would reduce the initial hurdle.

For future applications, GBNRTC would again have their model development consultant help to develop the model functionality and be available for assistance as needed for further refinements and analysis.

Land Use Challenges

GBNRTC invested a substantial amount of time to develop a maximum land use level for Bailey corridor. However, the intermediate values generated through the TMIP-EMAT process were interpolations between the base 2025 forecast and the maximum values were not necessarily representative of how GBNRTC expected the corridor may develop. Instead of a proportional, even growth, it is more likely that certain areas will develop first and other spots will lag behind. It is a challenge to define these non-linear patterns through a linear variable, as is required by TMIP-EMAT.

Analysis

GBNRTC stressed that the results are accompanied by a number of caveats. The inputs into the model have a number of embedded assumptions that need to be understood and recognized when reviewing the TMIP-EMAT outputs and visualizations.

They saw great value in being able to use TMIP-EMAT to back-up the qualitative assumptions and outcomes from traditional planning and knowledge with quantifiable results and visualizations. It also helped GBNRTC to think differently about the planning process and how the regional model is utilized.

Suggestions for improvements to TMIP-EMAT

Documentation

TMIP-EMAT could be improved through additional “how-tos”, video tutorials, and more documentation. It would help to have more examples of scoping and scripting, so that developers can review and modify existing scripts.

Improved visualizations

It would be beneficial to include a clearly laid out glossary of uncertainties and levers for each scope that is easy to see or find when analyzing the visualizations.

It would be beneficial to have a process for developing simpler visualizations that could be taken to stakeholders. For example, the Robust search line graphs are hard to follow, so finding a way to clearly narrate what is going on or find a different way to depict the results would be desirable. A narrative associated with other visualizations would be useful as well.

It would be ideal to have GIS-based visualizations and maps, such as select link analysis for each of the core runs. Reviewing how the select link analysis changes between runs would be very informative and valuable.

GBNRTC would also be interested in quantifying the impacts in a risk analysis fashion (i.e. 85% probability of project impacting traffic by X amount). So, additional visualizations that focus on risk, and not just exploratory, analysis would be very useful.

Suitable Applications of TMIP-EMAT

GBNRTC highlight the following potential applications of TMIP-EMAT:

- Alternative land-use scenarios
- Utilizing the same set of uncertainties across a range of projects

A.2 Oregon Department of Transportation Test Summary

Disclaimer: Oregon Department of Transportation (ODOT) conducted the TMIP-EMAT beta-test primarily to evaluate ODOT's new Activity Based Model (ABM); specifically the ability of the ABM to provide information about emerging technologies. To help achieve that purpose a realistic, but fictitious, set of regional ABM inputs was developed. At the end of this beta test, several flaws in the performance measure creation and methodology were noted as potential improvements for future analysis, but were not corrected in this dataset and resulting analysis. The information in this data and analysis serves as an example for how to use TMIP-EMAT using realistic data. This dataset and analysis should not be used to draw any specific conclusions about transportation policy's impact on system performance and outcomes.

ODOT is deploying Oregon's first regional (MPO scale) ABM. As part of the deployment, ODOT in partnership with the OMSC (Oregon Modeling Steering Committee) wishes to test the abilities and functionality of the ABM in regards to future technologies and new behavior patterns, such as; automated vehicles, light-weight personal electrics, and mobility-as-a-service. In coordination with the OMSC and the TMIP-EMAT team ODOT defined a set of reasonable test scenarios in relation to future technologies and related performance measures and completed a sensitivity assessment around the new ABM's capabilities regarding future technologies.

Additionally, there has been a lot of local (MPO) interest and questions around the ABM's capabilities, responsiveness, and measures regarding transit scenarios. ODOT's interest in using TMIP-EMAT was also to develop a series of responsible transit scenarios, potentially along with land use changes to better support transit use, to test and report on the ABM's ability to test transit scenarios.

A.2.1 FHWA Workshops and Support

FHWA produced two workshops to support EMAT scoping and analysis as well as bi-weekly calls to answer questions, lay out next steps, share information across the beta-testers, and collect information on the experience using TMIP-EMAT.

Scoping Workshop

The scoping workshop was held at ODOT's district office in Salem, OR on March 20, 2019 from 8:30AM to 2:30PM. The workshop participants are identified in Table 16.

The participants included representatives from ODOT's modeling and planning groups as well as their model development consultant.

Table 10: ODOT Scoping Workshop Participants

Person	Affiliation
Alex Bettinardi	Oregon DOT
Tara Weidner	Oregon DOT
Brian Dunn	Oregon DOT
Michael Rock	Oregon DOT

Person	Affiliation
Adam Argo	Oregon DOT
Joel Friedman	ODOT Model Development Consultant
Sarah Sun	FHWA
Martin Milkovits	FHWA TMIP-EMAT Consultant
Rachel Copperman	FHWA TMIP-EMAT Consultant

Analysis Workshop

The analysis workshop was held in two parts. The first part was held at ODOT’s district office in Portland, OR on September 17, 2019 from 3:30PM to 5:30PM and the second part at ODOT’s office in Salem, OR on September 18, 2019 from 8:00AM to 2:00PM. The workshop participants are identified in Table 16.

The participants included representatives from ODOTs modeling and planning groups as well as modelers from Oregon MPOs.

Table 11: ODOT Analysis Workshop Participants

Person	Affiliation
Alex Bettinardi	Oregon DOT
Tara Weidner	Oregon DOT
Brian Dunn	Oregon DOT
Michael Rock	Oregon DOT
Adam Argo	Oregon DOT
Peter Bosa	Portland Metro
Ray Jackson	MWVCOG
Stephanie Nappa	OCWCOG
Sarah Sun	FHWA
Martin Milkovits	FHWA TMIP-EMAT Consultant
Rachel Copperman	FHWA TMIP-EMAT Consultant
Jeffrey Newman	FHWA TMIP-EMAT Consultant

Bi-weekly Calls

FHWA and their consultants hosted calls with ODOT at least every two weeks between the scoping and analysis workshops. Alex Bettinardi was the main ODOT participant on these calls and reported progress against the task list shown in Table 3.

A.2.2 Model

The core model is an ABM using the CT-RAMP platform. This ABM covers two small MPOs in the Southern part of Oregon; Middle Rogue MPO and Rogue Valley MPO. In total the model area covers a population of approximately ~275,000 people in the base year (2010) and closer to 400,000 people in the forecast year, 2045. The scenario run for TMIP-EMAT tested a model region population of approximately 500,000 which would represent a forecast year greater than 2050. CT-RAMP (the core ABM) runs in Java. A batch file process runs the full model iteration loop. The full loop includes:

- Python to launch Visum where input data and skims are stored
- Python to develop the three tiered zone / skim system that the ABM uses.
- R to operate the external and commercial vehicle models
- Several feedback iterations in CT-RAMP and Visum assignment/skims with starting with a 10%population sample and ending with 100% sample (full population)
- Finally, a performance metrics summary and visualization.

This process takes approximately 3.5 hours to run the 2010 model year (275,000) and approximately 5 hours to run a test 2045 model year (the year used for the beta-tests). ODOT made several computers available during the beta-test available to run experiments in parallel. The key limitation in adding computers to speed up core model run time was that each machine needed a Visum software license, so the speed was limited by the number of available licenses not available machines.

A.2.3 Scope

This section begins with ODOT's perspective at the outset of the beta-testing process, then presents the materials developed during the scoping workshop, and concludes with the selected EMAT scope.

Motivation

ODOT was motivated to use TMIP-EMAT to support testing future technologies where there is little to no observed data to estimate or validate models. The other source of deep uncertainty for ODOT was around how land use will develop in accordance with land use and transit plans. EMAT allows the definition of a range of plausible assumptions inputs, rather than single point values, which is a more reasonable approach to specifying how new technology and its adoption at different penetration levels will impact travel behavior (demand), system capacity (supply), and interaction with new modes (e.g., car service, personal electric vehicles). In both cases (future technology and transit scenarios) it is helpful to explore a series of varied assumptions to better understand the risk and uncertainty of different futures and their impact on performance measures of interest.

While the technology, land use, and transit assumptions tested are dependent on the capabilities of the model and the beta-test timeline, the goal was to push the model as far as possible to gain insight into the reasonableness and usefulness of the model. ODOT approached this test with a willingness to scale back tests so that a core set of scenarios can be run within the beta-test timeline and an agile approach was used to minimize schedule risk.

Measures were envisioned to be refined in partnership with the OMSC and the TMIP-EMAT team. Initial scoping called for measures such as regional VMT, number of trips by mode, number of auto vehicle trips, regional transit boarding's/trips, and regional VHT.

High-Level Scoping

In the scoping workshop, the group started with the goal of providing equitable and accessible transportation system for all income groups. They identified a variety of levers to support that goal. The potential levers discussed were:

- Transit system enhancements, through the investment in fixed route system and/or a collaboration with private TNC services.
- Incentivizing transit-oriented development (TOD).
- Pricing mechanisms on roadways (road user charge system, toll / managed lane facilities), on TNC/auto-based mobility services, on transit through fare subsidies, or through parking fees.
- Investment in active transportation modes, possibly through micro-mobility programs.
- Mobility as a Service (MaaS).
- Incentives for electric vehicles.

The next step was to enumerate the desired metrics to evaluate the efficacy of these levers. The potential metrics discussed were:

- Accessibility measures, ideally segmented by income, that capture the travel time to employment and services with a multimodal lens.
- Measure of congestion.
- Standard metrics of vehicle miles traveled, person miles traveled, and vehicle hours traveled.
- Mode shares by demographic segment
- Household expenditures on transportation by income segment
- Total time spent traveling
- Out of home activities (number and duration)
- Revenue from user fees / transit
- Transit ridership
- Safety / Reliability / Exposure

Finally, the exogenous uncertainties that could affect the lever efficacy were identified as:

- Vehicle technology (autonomous and connected):
  - Supply side impacts on capacity through more (or less) efficient use of existing roadways.
  - A changing auto ownership model that could support more family sharing and fewer autos per household or even a fully fleet shared paradigm.
  - The demand side disutility of in-vehicle time may decrease as AV passengers are able to use their time productively.
  - Cost changes (vehicle operating and parking).
- User cost and availability of TNCs
- Land use and demographic changes, including total growth, shifting income and age distributions, spatial distribution, and density / zoning changes.
- Changes in the larger economy that would impact household spending power, travel costs, and work habits (more telecommuting).
- Management of curb facilities to facilitate local delivery.
- Freight operational changes for local delivery as well as long / medium haul.

Identify Model Functionality

The next step in the workshop was to identify the existing or needed model functionality to represent each lever and uncertainty variable in the model. As part of the discussion, the group discussed the level of effort involved in developing new model functionality. During this discussion the TOD and electric vehicle related levers were re-categorized as exogenous uncertainties.

Table 12: ODOT Model Variable Identification by Lever

Lever	Potential Model Variables to Represent Lever
Transit System Enhancements	• Transit lines • Transit headways • Transit travel times • Transit bias coefficients • Transit fares • Increase park and ride availability • Restructure walk-connection for mid-range to represent micro-mobility availability • Synthesize transit skim to represent TNC collaboration

Lever	Potential Model Variables to Represent Lever
Pricing	• Income-specific auto operating costs • Facility-specific tolls by occupancy, area, and time of day • Transit fare by route, district-district connections, vary by person attributes • Parking rates for work and non-work • New TNC mode alternative • Park and ride lot fee
Active Transportation	• Increase bike and walk speeds • Change the maximum distance threshold for non-motorized modes • Enhance active network connectivity • Vary non-motorized bias constant
Mobility as a Service	• Allow zero-auto households to use drive-alone modes • Revise treatment of households with fewer vehicles than workers and/or drivers

Table 13: ODOT Model Variable Identification by Exogenous Uncertainty

Exogenous Uncertainty	Potential Model Variables to Represent Uncertainty
Vehicle Technology Impacts on Operations	• Capacity by facility type, intersection vs. lane capacity
Vehicle Technology Penetration	• Simulate as part of synthetic population generation • Incorporate a new model component • Implement average values proportional to penetration rates
Zero-Occupancy Vehicles	• Post processing of trip tables • Post processing of aggregate VMT • Develop new autonomous vehicle routing model
Vehicle Technology Impacts on Behavior	• Modify time and cost coefficient

Exogenous Uncertainty	Potential Model Variables to Represent Uncertainty
Electric Vehicle Impact on Fuel Costs	• Auto operating costs associated with a simulation of electric vehicle availability
Vehicle Technology Impact on Parking costs	• Factor applied to default parking costs associate with a simulation of automated vehicle availability
Vehicle Technology Impacts on Operations	• Capacity by facility type, intersection vs. lane capacity
New Mobility Services and Increased Use of TNCs	• Would require substantial changes to model and was dropped.
Land Use	• Zonal employment • Modify synthetic population (control totals by geography)
Economy	• Modify jobs (zonal employment) and workers (synthetic population) • Vary income distribution in synthetic population • Vary transit level of service • Reduce work tours
Curb Management	• Terminal times • Parking costs • Availability and alignment of centroid connectors
Freight	• Direct changes to heavy truck trip table • Replace simulated personal shopping trips with truck trips • Reduce commercial vehicle and personal shopping to represent drone delivery

Select and Define Scope

After considering the work required to implement each lever, uncertainty variable and metric, the group selected the model variables that would be leveraged through EMAT keeping in mind that the total number of policy-levers and independent uncertainty variables have a linear relationship with the required number of Core Model runs, for example: 10 policy-levers/uncertainty variables require 100 Core Model runs, while the number of metrics has no impact on the number of Core Model runs required.

The scope was revised slightly through the subsequent steps, but not substantially. ODOT did note that they had hoped to have a larger stakeholder group during the scoping. They are now engaging more closely with this group to add more voices and ideas to the process.

Table 21 summarizes the selected levers and exogenous uncertainties. Originally, nine levers and uncertainties were scoped. Through the development and testing process, two uncertainties were dropped.

Table 14: ODOT Selected Levers and Uncertainty Variables

Policy-Lever/Uncertainty Variable	Minimum	Default	Maximum	Distribution (applies to Exogenous Uncertainties only)	Unit/Correlations/Other Notes
LEVER: Transit Everywhere (Synthesize transit skim to represent TNC collaboration)	NA	Current transit system	Transit everywhere replaces fixed-route system	NA	Originally envisioned as a single lever, later segmented into two levers with the Transit LOS continuous variable lever changing Transit ASC and this Boolean lever changing the availability of transit
LEVER: Transit LOS (IVTT equivalent change in transit utility)	-10.0	0	10.0	NA	Applies to both base fixed-route transit as well as transit everywhere
LEVER: parking rate factor on existing parking	0.5	1.0	20	NA	Existing parking rates are factored up by the factor provided.
LEVER: active transport improvements – factor applied to walk/bike speeds	1	1	2	NA	Changes only to speed, maximum distance is maintained; proxy for micromobility penetration
EX. UNCERTAINTY: Interstate (access controlled) Capacity – vehicles per hour per lane	1500	1900	3000	uniform	A proxy for AV penetration and impact on access controlled facility (interstate) capacity.
EX. UNCERTAINTY: Auto operating cost – cents per mile	1.0	12.4	25.0	uniform	Low represent electric vehicle efficiency, high represents fleet AVs and higher pricing (tax) structures
EX. UNCERTAINTY: Household income multiplier	0.5	1.0	1.5	uniform	Was used as a simple method to represent changing jobs, job type, household worker mix, etc.
EX. UNCERTAINTY: Value of time (change in sensitivity to IVTT)	0.5x	1x	1.2x	Would make higher sensitivity less likely	Uncertainty variable was dropped when found to be perfectly correlated with auto operating costs.
EX. UNCERTAINTY:HH densification (% shift distance to the center)	0.5x from the core	1x from the core	1.5x from the core	uniform	Prototype implementation defined rings, chose number of houses to shift by ring, and did a ring jump. Testing following full LHS runs showed unreasonable responses and variable was removed from scope.

The scoped metrics and their optimization hint are shown in Table 15. ODOT plans to extend the metrics to include more spatial and demographic segments.

Table 15: ODOT Scoped Metrics

Metric	Optimization Hint
Percentage of Population with Access to 50k Jobs by Car within 20mins in PM	maximize
Bike and Walk Mode Share	maximize
Transit with PNR and KNR Mode Share	maximize
Millions of Person Miles Traveled	minimize
Millions of Vehicle Miles Traveled in PM	minimize
Millions of Auto Miles Traveled	minimize
Millions of Truck Miles Traveled	minimize
Millions of Vehicle Miles Traveled	minimize
Thousands of Vehicle Hours Traveled in PM	minimize
Thousands of Auto Hours Traveled	minimize
Thousands of Truck Hours Traveled	minimize
Thousands of Vehicle Hours Traveled	minimize
Percent of Interstate Miles over 90% V/C Ratio During the PM Peak	minimize
Percent of Principal Arterial Miles over 90% V/C Ratio During the PM Peak	minimize
Percent of Minor Arterial Miles over 90% V/C Ratio During the PM Peak	minimize
Number of Autos Owned Per Household	maximize
Percent of Non-Mandatory Tours	maximize

ODOT refrained from defining all performance metrics during the initial scoping process. It was found that until the lever and variables are fully implemented, it is not obvious what metrics will be worthwhile to capture, particularly for metrics that are focused on a certain area and/or segment of the population.

A.2.4 Implementation and Model Run Experience Report and Feedback

This section summarizes the TMIP-EMAT API development, ABM model extensions, and the core-model runs.

API Development and Model Extensions

ODOT staff had some comfort manipulating and changing python scripts and completed the bulk of the work "in-house." ODOT's modelers were not as familiar with Python as R, so the API implementation had very simple functions in Python within EMAT with the bulk of the functionality occurring in R. This approach is within the design framework of TMIP-EMAT because of the structured API that organizes the interaction points between EMAT and the core model. ODOT also leveraged their model development consultant to assist with development of model-side API functionality to enable a programmatic control of model inputs.

ODOT structured the model and TMIP-EMAT such that a base copy of the model was maintained and copied into a working folder for the active experiment. When the experiment completed, the working folder was renamed to the archive name.

ODOT found the initial Anaconda setup and acquiring the correct Python packages to run TMIP-EMAT challenging. More effort was required to learn and install Anaconda and get Jupyter Notebooks running than initially anticipated.[5]

Core Model Runs

The univariate sensitivity tests were conducted on a single machine. The LHS experiments were run across multiple computers with the scope and results saved in a common SQLite database on a single machine. The results were saved locally on each machine and will be consolidated on a network shared drive archive.

There is an opportunity for errors to be introduced when moving from individual runs to a the full set of experiments. ODOT ended up running the LHS experiments 3 times. The first time, ODOT conducted a review after 20 runs were completed and discovered issues in how the model inputs were set. Next, 80 runs were completed and the review revealed an issue with the land use density variable (see discussion below). Finally, 70 runs were completed and used for the analysis workshop. The use of multiple systems gave ODOT a lot of flexibility and responsiveness to complete the model runs. After the issue with the land use density variable was discovered, ODOT was able to redefine the scope without this variable and regenerate 70 model runs over a long weekend.

Several run and set-up issues occurred. These include:

- The availability of network licenses was an issue at times while running the LHS experiments and ODOT did need to throttle the systems running to allow for other studies to proceed.
- At one point, ODOT did experience a database lock condition that implied a potential network issue. This was not repeated and was attributed to the model configuration.
- For the initial runs, there was an issue with the VISUM model software that caused random crashes that would halt the EMAT process. In these cases, ODOT would manually restart the model in order to take advantage of the several speed feedback loops that had been completed. Once these were complete, the EMAT post processing, metric import, and archive steps were run interactively through Jupyter Notebook. Once those steps were complete, the run_experiments method could be restarted against the pending experiment. The random crash issue was eventually resolved with PTV and the final iteration of LHS experiments completed without error.

A.2.5 Analysis Feedback

This section summarizes ODOT's feedback and the workshop discussion by analysis type.

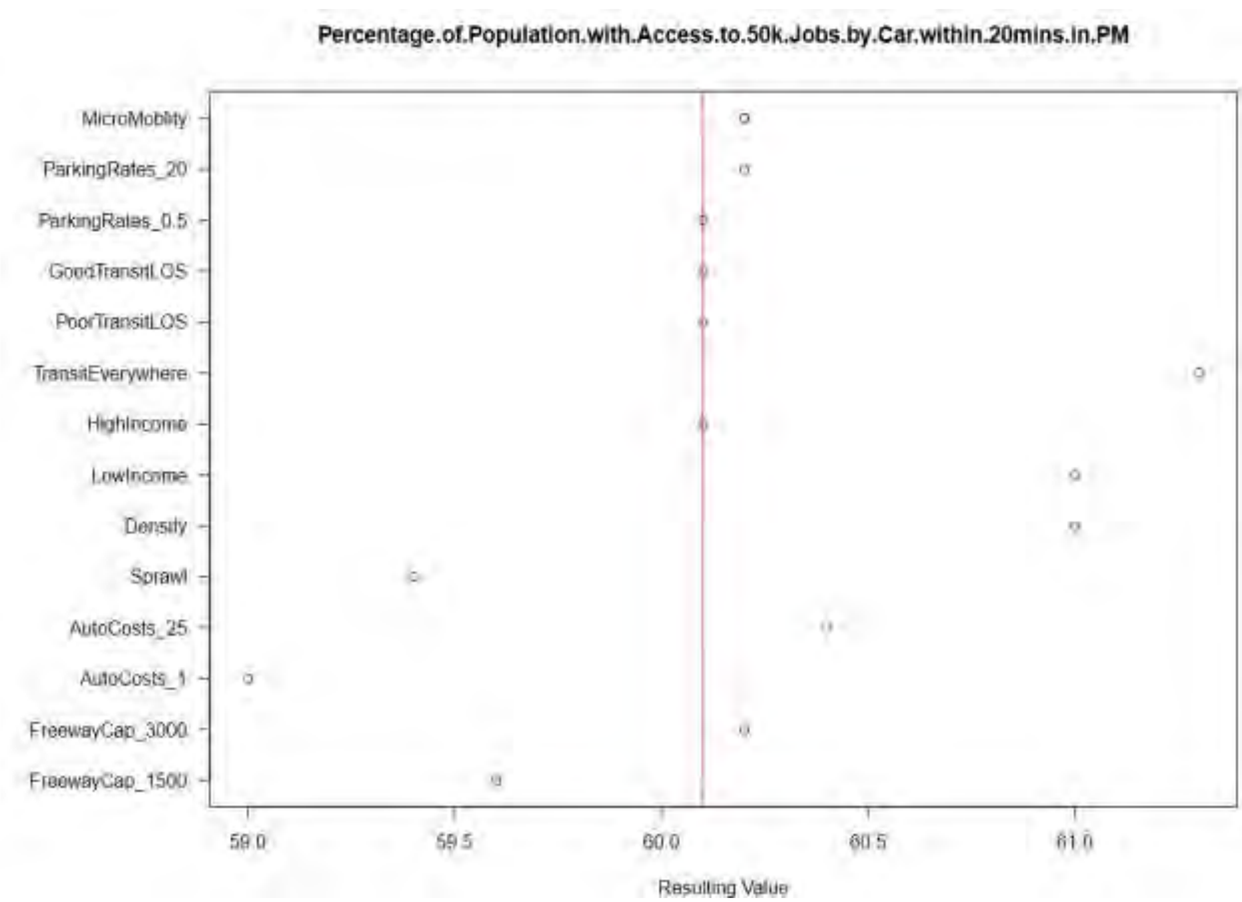
Univariate Sensitivity Test Review

The EMAT univariate experiment design process was a convenient method to methodically walk through each variable and lever to confirm correct operation. However, reviewing the results required specifying a title to each experiment to understand what variation is expected. An

[5] The EMAT installation challenges have been alleviated with the conda-based installation method of EMAT now available.

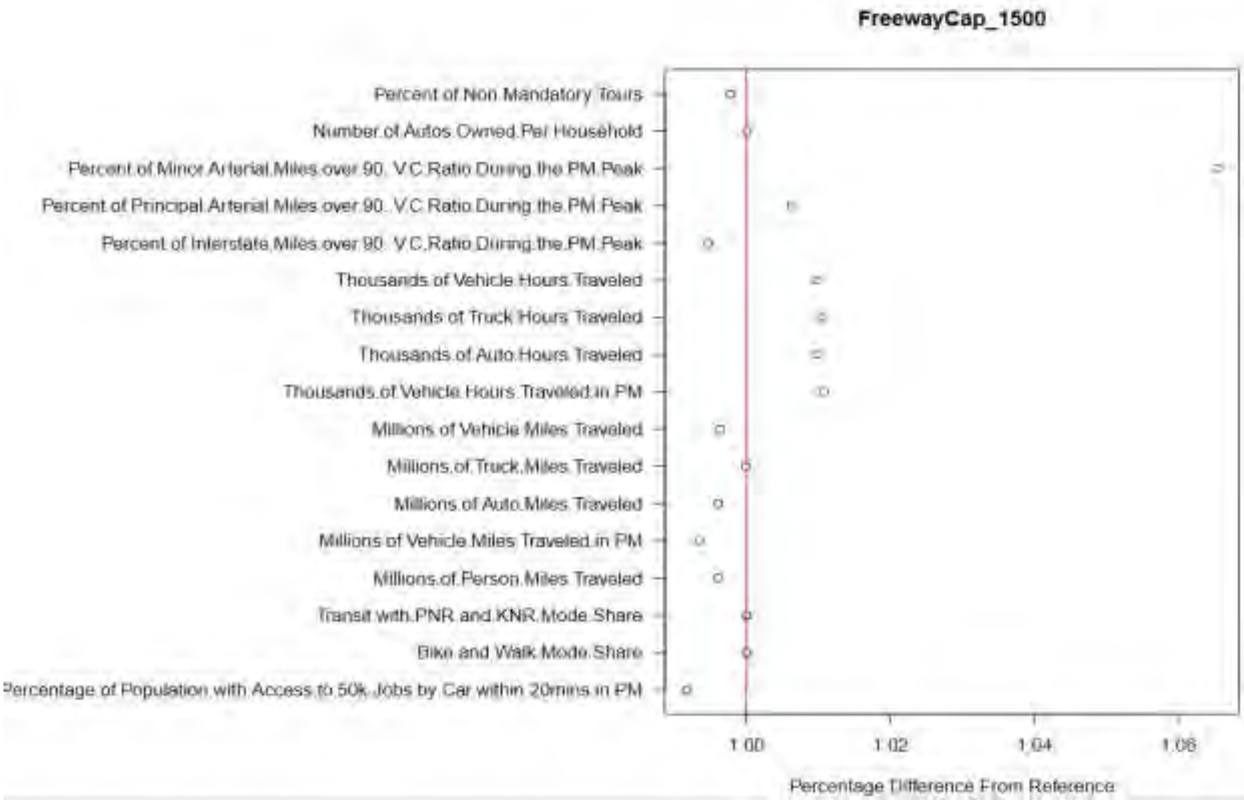
extension that would add a note to each experiment or a post processor that compares the experiment to the scope and identifies what variation is in test would be helpful. But, there are not many experiments conducted in a sensitivity test process, so this is not necessarily a high priority feature.

ODOT produced a set of R summaries that input the experiment dataframe to comprehensively compare all metrics of each experiment against the baseline metrics (Figure 4) and to compare each metric variation by lever and uncertainty variable (Figure 3). The scenario names were defined in a custom R script and indicate the input variable changed and the value set.



Source: ODOT

Figure 3: Comparison of Single Metric Across all Univariate Experiments



Source: ODOT

Figure 4: Comparison of Single Experiment Across all Metrics

As was discovered later, a test of the extreme values does not capture all errors that might occur in the API coding. The land use density variable behaved at the extreme values, but gave unreasonable results at intermediate values.

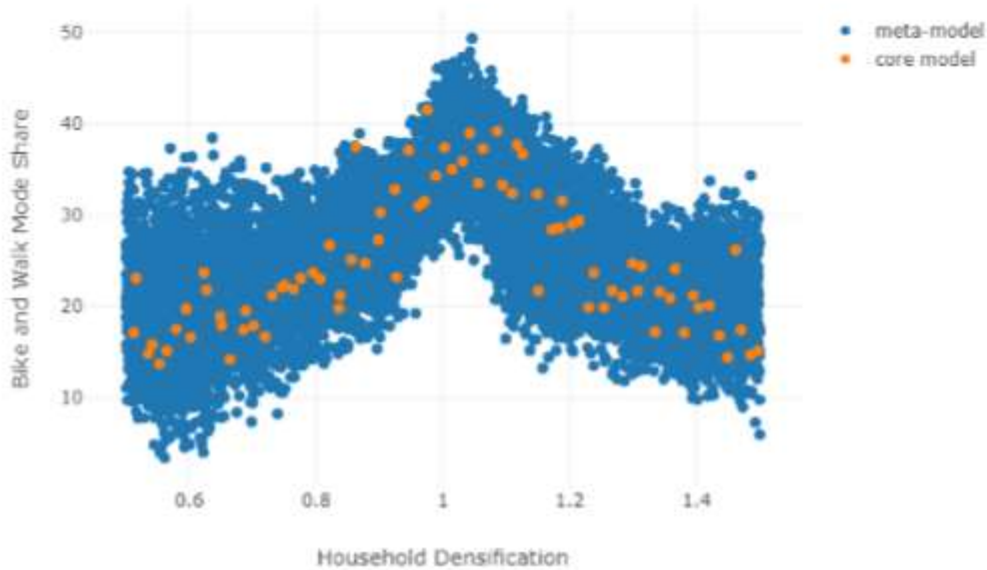
Ultimately, the univariate sensitivity tests were useful as an initial confirmation of the correct operation of the model and to gauge the degree of variation for each metric. It is after the univariate sensitivity tests are complete that we begin to see if the metric selection is appropriate.

LHS Core Model Experiment Review

The review of the LHS core model results was key to identifying issues in the core model. There were two significant issues discovered through this process, one with an uncertainty variable process and the other with a metric definition. Neither of these issues were apparent through the univariate sensitivity tests.

Uncertainty Variable Trouble-Shooting Example

Land use was identified as a key input to the model. A mechanism was developed to move development from the downtown areas to the suburban ring areas. While this mechanism passed the initial sensitivity tests that ran at the extreme value, when the model was run through the LHS experiments the household density produced a non-monotonic and unreasonable response in several performance measures, most notably the non-motorized mode share (Figure 5).



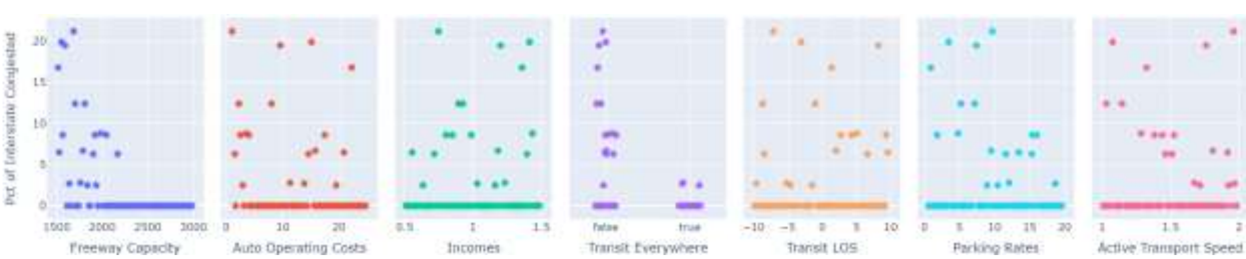
Source: FHWA

Figure 5: ODOT Troubleshooting Example

This problem was initially discovered while reviewing the meta-model statistics. The low R-square of the linear regression models indicated that the model was behaving in an unpredictable manner. Further investigation with the visualization revealed the unreasonable pattern from that variable.

Metric Trouble-Shooting Example

Similar to the uncertainty variable, an issue with the model metrics was discovered through examination of the meta-model statistics. The meta-models returned lower than expected goodness of fit statistics for several of the congestion variables that were defined as percent of congested lane miles. Upon further investigation using the core-model scatter plots, as shown in Figure 6, it was discovered that these metrics had many zero-values for scenarios. The preponderance of zeros hampers the linear regression ability to fit the data. This suggests that a more robust metric definition is needed, for example, with a lower threshold for congestion. Note that this metric performed well in the univariate sensitivity tests, but performed poorly when all experiments were completed.



Source: FHWA

Figure 6: ODOT Congestion Metric Issue

The visualization of the core model results showed experiments as points by lever/variable and metric. The group wanted to identify the specific inputs that were producing a given metric, particularly for the edge conditions[6].

Visualization and Feature Scoring using Meta-Model Results

The visualizations and feature scoring were useful when they did not show a strong correlation between the levers and uncertainties and metrics because this answered the question of whether levers were important factors.

Discussed below in the improvements, ODOT found that they would benefit from seeing a reference base case to gauge the response of the levers and answer the initial question – is it having the appropriate response. This is shown in the univariate sensitivity test summary that ODOT developed.

In the visualization exploration, the group went back to the original goal (providing accessibility to all groups) and investigated metrics that supported the ultimate goal. For example, reducing congestion is an intermediate goal, but improved accessibility may mean more overall auto travel.

Scenario Discovery

ODOT suggested that starting the Scenario Discovery from a reference point and showing the pathway towards meeting the constraint. Distinguishing between exogenous uncertainties and levers would also be useful. Alternatively, PRIM could be set to only include levers.

Directed / Robust Search

Directed and Robust Search utilities revealed tradeoffs between non-motorized and transit shares. ODOT would have set more metrics to “informational” (rather than “minimize” or “maximize”) if they had fully understood the implications for Directed and Robust Search.

The question came up if there were “middle of the road” solutions that were being missed by these utilities. The Directed Search returns the solutions that are optimal for at least one metric – it could be that a solution that fits reasonably well across all metrics exists, but is not optimal for any one metric and thus is not included in a Directed Search result. It was brainstormed that a solution that balances multiple performance measures could also be accommodated with a new performance measure that synthesizes the others into a measure of “acceptable solution”. The optimization could then be conducted on the set of solutions that first ensures this acceptable solution measure is met.

A.2.6 Level of Effort

ODOT was in the process of finalizing the ABM development at the same time as deploying TMIP-EMAT, which made the API development more complicated. ODOT estimates about 20-30 hours per month over the four month deployment was spent primarily on TMIP-EMAT with 80-120 hours spent in total. Breaking the amount spent down by components:

- Moving from the scoping to an understanding of the model impacts and conceptual design took about 20-30 hours

[6] TMIP-EMAT visualizations have been updated to include a mouse-over control that reports the X, Y values and experiment ID.

- Coding of the API and ABM took about 50-70 hours
- Running the model and initial review of results took about 10-20 hours

Now that the API is developed, ODOT anticipates that the addition of new hooks for lever/variable setting and metric collection will be much more straightforward.

Note that ODOT's modeler is an experienced programmer, although with less experience in Python and Jupyter Notebooks. The time required to deploy TMIP-EMAT could be longer with less technical staff.

A.2.7 Lessons Learned and Overall Feedback

The main takeaway from the exercise was that it is an iterative process and is best when treated as such rather than trying to make it perfect the first time. This was concluded through the scoping exercise, model specification, metric refinement, and analysis. ODOT staff emphasized that "you don't get it the first time".

ODOT did not report any one aspect of TMIP-EMAT to be especially challenging. The largest challenge was finding the time in busy schedules to prioritize this work. However, ODOT did acknowledge that FHWA support was necessary to successfully deploy and run the model through TMIP-EMAT with the beta installation procedure and available instructions. Given the automated installation now available and the continued improvement in instructional tools, TMIP-EMAT could be deployed without dedicated FHWA support.

Scoping

It was a good approach to start with the goals that the community care about and then think about what policy levers can be used to improve our goals, what measures exist to understand the improvement (if any), and what uncertainties create risk around how successful they will be at meeting our goals. The RDM approach has already started to improve ODOT's thought process and approach to planning, and they hope to continue to spread this structure into other aspects of planning and modeling work.

Application of this approach with the Oregon steering committee was unbounded and, as such, produced a bit of a laundry list that exceeded model capabilities. Some ways to manage this would be:

- Emphasize tool capabilities, as was done in the TMIP-EMAT scoping workshop with the model consultant
- Manage expectations against tool capabilities, particularly when working with parameters and structures with deep uncertainty

There was benefit to having a model expert included in the scoping process. The model developer contributed toward answering questions and creative thinking about how the tool(s) could be used to represent levers and uncertainties.

ODOT suggested that an overview or an example of the scoping process would be useful for agencies conducting it without prior experience to be in the right mindset prior to scoping.

When scoping for a new model, ODOT's experience suggested it would be best to jump in and start testing the model and iteratively refine the metrics, levers and variables based on the initial model response.

Metric definition and structure

The insight gained from the model is ultimately dependent on the metrics that are produced. The initial metrics ODOT summarized were quite broad and thus limited in the detailed insight that could be gleaned.

Although the metrics ODOT initially collected need to be refined, the potential to produce spatially and demographically segmented metrics, because of the disaggregate nature of the core model, is very interesting and to be able to produce those quickly across a range of levers and uncertainty variables is particularly valuable.

The development of metrics exposed a lot of opportunities. In thinking about the metrics for the scope, it brought up less common metrics. Most of these could be provided by the core model, but some in very complicated (backwards) ways. This process has helped improve thinking around a redesign of how outputs are provided by the core model. Fortunately, adding metrics does not require additional core model runs so the work process would support adding metric definitions to the scope and reprocessing the archive model outputs, which takes substantially less time than having to rerun the core model runs.

Metrics that ODOT intends to further pursue are:

- EJ analysis, social costs, cost of the road improvements, and vehicle and fuel information
- Metrics that are well-associated with the lever. For example, the parking cost lever only impacted a small subset of zones in the region therefore an aggregate metric did not capture well the effect; whereas, a measure of mode share in urban areas would have been less muted by the non-priced areas.
- The analysis and conclusions suggested by the metrics prompted the modelers to think more critically about the lever implementations. For example, the “transit everywhere” demonstrated a great advantage across many metrics, but the lever does not load the transit pod vehicles onto the network and thus the actual impact in VMT and VHT are not realized.

Improved assessment of the core model

The initial testing for the TMIP-EMAT runs identified bug and code issues in the models and exposed questions and areas to correct that likely would not have been discovered without a structured testing processes, i.e. informal model testing would likely not have revealed these issues. Lessons learned during the process suggests that rigorous sensitivity testing and model stress-tests, similar to the processes implemented during TMIP-EMAT application, should be a requirement for testing before a model launch. ODOT found that TMIP-EMAT did a remarkable job in helping to ensure that model is sensitive to the right things and is reacting in the correct way. The TMIP-EMAT process ensures this testing occurs before starting to answering questions with the model.

ODOT also learned that their new ABM structure did not allow for easy access to all the person level information that was desired. The TMIP-EMAT work helped show areas where ABM output simplification would facilitate access to the full set of person level diagnostics desired for project analysis.

Suggestions for improvements to TMIP-EMAT

Reference scenario in visualizations

ODOT hit on the idea to include a reference scenario as anchor of the analysis and to help put the differences into context. Specifically:

- Most or all of the TMIP-EMAT visuals should have an easy option to add the **reference** case to the plot for comparison. Some users might also consider including the base-year scenario as an additional reference case, as in some cases the input (L) or metric (M) might want to be compared to what is happening today (congestion levels, transit funding levels, etc.).
- It would be constructive to include the **reference** future (or even base-year) when identifying which feature moves the needle the most, starting from the reference values.
- More visuals should be designed that explore options to improve measures from the reference case. It would be ideal to be able to toggle the reference and/or base-year scenarios on and off. Alternatively, the visual could ghost these two points on each scale.
- Visuals surrounding how the uncertainties affect the reference point could be added. The single point reference future could be expanded to a cloud by applying various ranges and sets of uncertainties..

Segmentation of exogenous uncertainties and levers

The visualization might separate uncertainties and levers. Having separate plots/conversation about each, or able to toggle uncertainty on-off. ODOT recommended that a consistent color coding of uncertainties, levers, and measures consistently throughout (e.g., in core model & meta-model scatterplots) would be useful.

Metric support

Existing scenario tools have a geospatial visualization. It would be useful to connect TMIP-EMAT metrics to these spatial tools. It may be desirable to set-up metrics such that they could support separate plots/conversations for different geographic levels or population subsets (e.g., urbanized areas, low income, etc.).

Some metrics might be more usefully normalized, e.g., transit ridership or VMT per capita. Then, if inputs/uncertainty change some values (e.g., population growth), the values are still comparable.

Scatter plot extensions

Mouseover in the scatter plot to indicate more than the scenario number. EMAT includes continuous variables, but the range could be split into High-Med-Low levels to make it easier to identify where the scenario falls within the range.

PRIM (scenario discovery) approaches

The PRIM optimization trajectory might start with the Reference scenario (likely not meeting the set threshold), rather than the all options implemented scenario (100% in the box), and increasingly add in changes to incrementally more dimensions in order to get into the desired ‘box’ range, and then further increase the coverage and density within the box. Also, it would be helpful to set multiple constraints in the scenario discovery process to define scenarios within a subset of the scoped universe of outputs. For example, when testing pricing scenarios, it can be important to demonstrate the constraints required to achieve targets without pricing and then to relax the pricing constraint and discover the potentially broader set of scenarios.

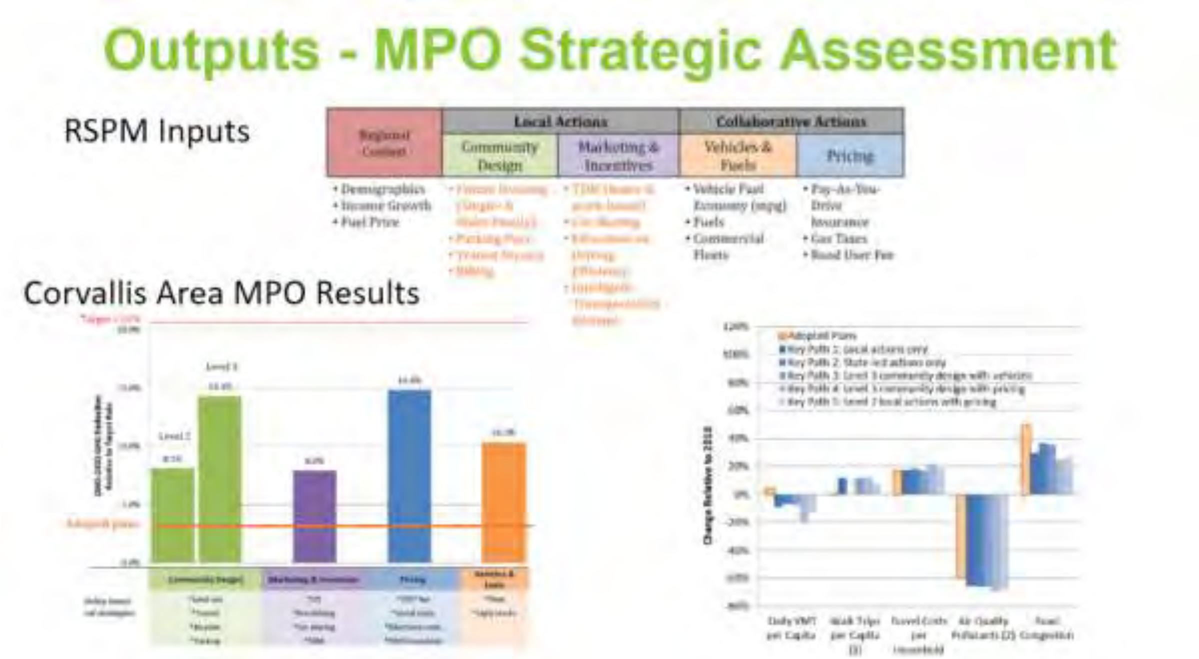
Parallel coordinate visualization extensions

Similar to the scenario discover, it would also be useful to constrain the levers and uncertainty ranges for the parallel plot visualizations. In the parallel plot, rather than a simple toggle on-off of uncertainties, the user might set the desired elements/levels of exogenous uncertainty to include.

Another improvement to the parallel coordinate visualizations would be the ability to plot meta-model runs (5000, 10000…), as an alternative to the interactive histograms and scatter plots. A further improvement would be an additional interconnected visual tool to the histograms and scatter plots. In other words, a tool that initially shows all the scenarios in the parallel plot and lets the user limit them. This could allow a deep exploration of the simulation outputs without involving an optimization method.

Define process to support a Key Path analysis

ODOT has used a "Key Paths" approach to filter the many scenarios produced by their strategic model. The filtering process begins by imposing certain constraints that may represent goals (e.g., high GHG reduction), that are defined as being constrained to within x% of current levels. The resulting filtered set of scenarios (Key Paths) that fit the criteria are evaluated by the local agency/stakeholders, and a selection of Key Paths are chosen to guide future plans. TMIP-EMAT could fit into this process, by adding uncertainty and outcome across various goals/metrics for just those key paths might further aid that prioritization conversation. In Figure 7 the lower right sub-chart shows outcomes for Key Paths that all meet the GHG reduction goal.



Source: ODOT

Figure 7: ODOT Strategic Assessment Example

Management of sub/analysis-scopes to simplify analysis setup and support structured-file control

Setting up a robust search requires defining the optimization function manually. It isn't obvious how to test a subset of levers from the overall scope, e.g. optimize for levers 1, 2, and 4, but ignore 3. A YAML or similar format text file could interact with the tool as opposed to scripting. It could link the analysis to sub-scopes (scopes A, B, C…), that would have to obey the master scope used in the design, but that could have changes for more tailored analysis and digging, without having to alter the original scope. As new functions / visuals are considered, these "analysis scopes" could potentially get a few additional fields, so that the user could interact and design the visuals / analysis through a text file YAML, as opposed to custom scripting. Custom scripting could always be an option too.

Documentation

Where reasonable, consistent use of XLRM should be used throughout the wiki and scripting and scope files. Example, the scope file examples could add a header or comment in each section that reminds the user of X, L, and M elements (sections of the scope). Headers in the online documentation could also be preceded with the letter code.

A wiki page with common functions / examples would be helpful to have as a reference. More tools and resources with more demos and walkthroughs would help reduce having to learn two programs at once (i.e. Jupyter notebook and TMIP-EMAT). It would be useful to have additional guidance on how to choose the optimal set of performance metrics and levers (i.e. more guidance on scoping).

Usability

Jupyter notebooks can be challenging and not easy to use. Thus, it would be beneficial to have a robust software program that includes the TMIP-EMAT functionality currently deployed in Jupyter notebook. Ideally, EMAT could be run without expert support.

Analysis guidance

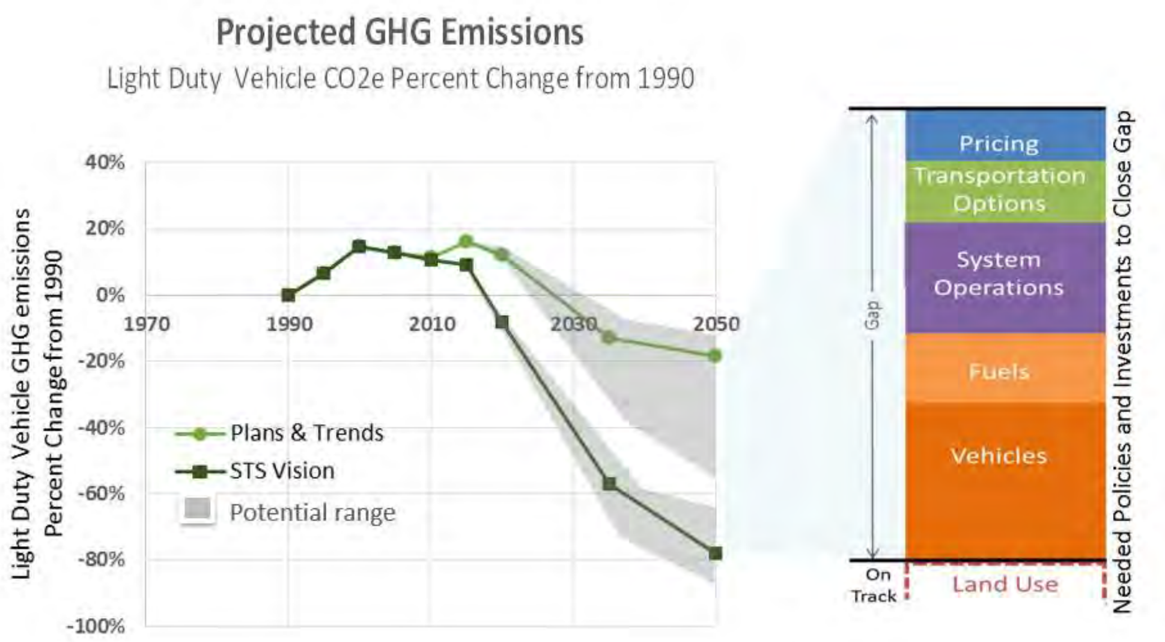
It would be helpful to have better guidance on how to build an analytical story through the analysis tools. There is also a desire to more clearly align TMIP-EMAT with the planning process. This may involve formalizing methodologies for how to tell the story, demonstrating how to analyze the tipping point, and laying out who is the right group of people to engage during scoping and analysis.

Simpler visualizations

Figure 8 and Figure 9 are examples of charts that have been useful in translating results from scenario analysis into easy to understand outcomes. These two charts may be useful for incorporating into TMIP-EMAT, if possible.

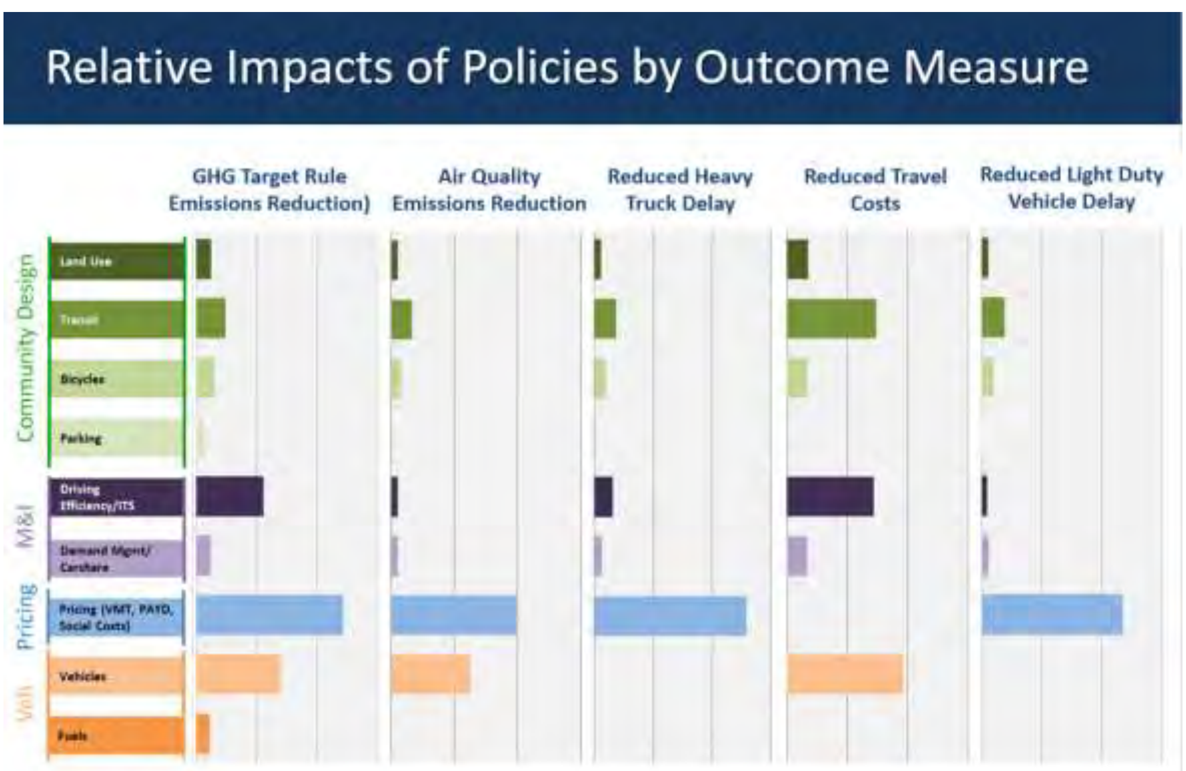
Figure 8 first shows monitoring to a vision. It found that business as usual "Plans & Trends" scenario did not meet the "STS Vision" scenario, even after accounting for key uncertainties (shaded region). The gap is further explained with the bar chart that identifies the impact of the policies in the Vision but not the Trend scenario in meeting the GHG Vision scenario/goal. Running these one-at-a-time on top of the Trend scenario was utilized to develop the bar chart on the right.

Figure 9 is similar (single bar broken into parts), but looks at the relative impact of each policy across multiple goals.



Source: ODOT

Figure 8: ODOT Vision Monitoring Example



Source: ODOT

Figure 9: ODOT Policy Impact Example

A.2.8 Next Steps

Beyond this initial beta-test, ODOT anticipates further utilization of TMIP-EMAT. ODOT did its own scoping exercise on emerging technologies using the materials provided by FHWA with the OMSC (Oregon Modeling Steering Committee). The scoping exercise considered goals from the Regional Transportation District (RTD) and identified levers that would help reach those goals.

ODOT will be looking for the potential to link a future application of VisionEval (VE) with TMIP-EMAP. An application in the Southern Oregon area which could allow data from both the Rogue Valley regional and VisionEval models to be used to feed TMIP-EMAT. This would allow ODOT to explore the ability of TMIP-EMAT to integrate model outcomes, thereby harnessing the strength of various models (often with long run times) by performing selected linked runs and mapping the broader solution space between these selected inputs and outcomes, e.g., land use changes from one model, vehicle choice and GHG impacts from another model, roadway reliability and operations from a third.

ODOT also has longer-term interest in exploring connections between TMIP-EMAT and information from Oregon's Statewide Model (SWIM), which has different strengths than the ABM and VisionEval, such as land use model results, economic measures, and statewide commerce flows. All of these interests are in addition to ODOT's plans to utilize TMIP-EMAT more regularly to help understand ranges of information, uncertainty and resiliency across the different requests and questions that ODOT is called upon to provide information for.

ODOT plans to do a scoping workshop and TMIP-EMAT implementation with VE as an in-house effort. ODOT has also been considering contracting making the ABM and TMIP-EMAT setup more production-ready. This means a tighter (more automated) linkage between the core model and TMIP-EMAT. This work could be done internally, but would likely involve some consultant support.

Another extension would be to connect to the land use model, although changes in land use are hard to represent in a continuous variable. One option would be to use four categorical variables and develop "tag-team models" to explore land use and transportation impacts; another option would be to connect to the VE population synthesizer.

Suitable Applications of TMIP-EMAT

EMAT would be best suited for guiding plans and strategies, like regional transportation plans (RTPs), or statewide plans, or multi-million dollar projects (i.e., big efforts, not little traffic impact assessments):

- Highway Plan Updates
- More tactical elements: "What ifs" operational policies - ITS, highway management issues, classifications, overlays
- Regional model application

- Other statewide plans that may have different goals for different regions, particularly in geographically diverse areas such as Oregon
- Usually do 3-4 scenarios and get to policy-makers and ask did you test X, Y, Z - this would help us deal w/ that issue (be proactive)
- Strategic models (VisionEval)
    - Leverage the strategic model to inform the important aspects to model more carefully with the regional and disaggregate models
- Transit investment, including HB 2017
    - Expand service on weekends, evenings, headway, transit prioritization, expansion
- Evaluating packages of projects
- ITS (though limited by model), demographics (in-migration uncertainties due to climate change), cost of roads, social costs, vehicle/fuel type
- Connect with MOVES for emissions analysis (TMIP-EMAT can work with any core-model + post processing routine)

A.3 *San Diego Association of Governments Test Summary*

Disclaimer: San Diego Association of Governments (SANDAG) conducted the TMIP-EMAT beta-test using the cross-border model in SANDAG's Activity-Based Model (ABM) to evaluate policy and investment uncertainties associated with border crossings. The cross-border model predicts travel demand generated by Mexican residents who make north-bound crossings into San Diego county. To help achieve that purpose a realistic, but fictitious, set of border-crossing inputs was developed. As part of this beta test, a couple of sensitivity testing oddities with regard to performance metrics were investigated, with one transit fare sensitivity issue unresolved. Information in this data and analysis serves as an example for how to use TMIP-EMAT using realistic data. This dataset and analysis should not be used to draw any specific conclusions about transportation policy's impact on system performance and outcomes.

San Diego Association of Governments (SANDAG) conducted a TMIP-EMAT beta-test to explore the possibility of using TMIP-EMAT to evaluate investment, pricing, and regulatory policies in regional transportation planning. In particular, SANDAG wanted to explore using TMIP-EMAT to analyze the impact of uncertainties associated with emerging technologies, policies, and transportation investments. Ultimately, SANDAG is motivated to partner with FHWA in their development of new modelling techniques, particularly those around uncertainty.

SANDAG decided to use the regional model cross-border component for the TMIP-EMAT beta-test in order to investigate application of the utility without requiring the full model runtime. There are a variety of policy levers and uncertainty variables that can be tested with this component, including border policies, emerging technology uncertainties, and infrastructure projects.

Border policies have a significant impact on the San Diego regional transportation system, given the unique location of the region along the U.S and Mexico border. San Ysidro is the largest border crossing in the U.S. Uncertainties related to border policies include the border wait time; auto operating cost; and Mexican land use changes.

Emerging technologies, including autonomous vehicles and new mobility forms such as hyperloop, may affect decision-making in long range regional planning. Even micro-mobility could have a substantial impact on transit access and egress paths.

San Diego is considering transportation infrastructure decisions specific to point of entry access including both transit and toll road projects.

A.3.1 Broader Modeling Challenges

Beyond the cross-border model perspective, SANDAG described a set of modeling and planning challenges that a utility such as TMIP-EMAT could help alleviate. Some of these will be better addressed by improvements to SANDAG's regional model that are currently in development.

- Land use changes and responses to policy.
- First-Mile, Last-Mile transit challenges and how new technologies will complement and/or compete with transit.
- How to best respond to rapidly changing greenhouse gas emissions targets.
- Balancing pricing impacts on environmental justice with efficacy at reducing VMT.

- Supporting both a visionary, fiscally constrained plan as well as the legal requirements to generate and meet certain point forecasts.

A.3.2 FHWA Workshops and Support

FHWA produced two workshops to support EMAT scoping and analysis as well as bi-weekly calls to answer questions, lay out next steps, share information across the beta-testers, and collect information on the experience using TMIP-EMAT.

Scoping Workshop

The scoping workshop was held at SANDAG's offices in San Diego, CA on March 21, 2019 from 10:00AM to 4:00PM and March 22, 2019 from 8:00AM to 10:00AM. The workshop participants are identified in Table 16.

The participants included representatives from SANDAG's modeling and planning groups.

Table 16: SANDAG Scoping Workshop Participants

Person	Affiliation
Wu Sun	SANDAG Modeling
Yun Ma	SANDAG Modeling
Tom King	SANDAG Modeling
Rick Curry	SANDAG Modeling
Ziying Ouyang	SANDAG Modeling
Ray Major	SANDAG Modeling Director
Elisa Arias	SANDAG Planning
Phil Trom	SANDAG Planning
Laurie Gartrell	SANDAG Planning
Krystal Ayala	SANDAG Planning
Sarah Sun	FHWA
Martin Milkovits	FHWA TMIP-EMAT Consultant
Rachel Copperman	FHWA TMIP-EMAT Consultant

Analysis Workshop

The analysis workshop was held at SANDAG's offices in San Diego, CA on September 16, 2019 from 9:00AM to 4:00PM. The workshop participants are identified in Table 17.

The participants included representatives from SANDAG's modeling and planning groups as well as the SANDAG modeling consultant.

Table 17: SANDAG Analysis Workshop Participants

Person	Affiliation
Wu Sun	SANDAG Modeling
Yun Ma	SANDAG Modeling
Cherry Liu	SANDAG Modeling
Tom King	SANDAG Modeling
Rick Curry	SANDAG Modeling
Susan Xu	SANDAG Modeling
Elisa Arias	SANDAG Planning
Phil Trom	SANDAG Planning
Nagendra Dhakar	SANDAG Modeling Consultant
Sarah Sun	FHWA
Martin Milkovits	FHWA TMIP-EMAT Consultant
Rachel Copperman	FHWA TMIP-EMAT Consultant
Jeffrey Newman	FHWA TMIP-EMAT Consultant

Bi-weekly Calls

FHWA and their consultants hosted calls with SANDAG at least every two weeks between the scoping and analysis workshops. Wu Sun and Yun Ma were the main SANDAG participants on these calls and reported progress against the task list shown in Table 3.

A.3.3 Model

The core model used in SANDAG's Regional Plans is a CT-RAMP activity-based model developed in Emme. The current version of the model is ABM2, which has a San Diego resident model and a few special market models including a cross-border model. The model runtime is between 32 to 40 hours, depending on the modeled year. The cross-border model as a stand-alone component takes about 2 hours to generate new cross-border trips. These are then combined with the other model results and the highway and transit assignments, taking about 5 hours. The beta-test will be conducted using the cross-border model.

A.3.4 Scope

This section begins with the preliminary scopes that SANDAG's planners developed in advance of the workshop, then presents the materials developed during the scoping workshop, and concludes with the selected EMAT scope.

Preliminary High-Level Scoping

SANDAG produced high-level scopes inspired by four different goals:

- Deploy and understand on-demand, shared AV fleets that provide point to point service
- Dynamic complete streets utilization—Active Transportation and Demand Management (ATDM) for all roads
- High-speed/flexible/modular transit
- Smart and dynamic parking management

The levers, metrics and uncertainties identified for each of these goals are listed in Table 18.

Table 18: SANDAG Preliminary High-Level Scopes

Goal	**Deploy and understand on-demand, shared AV fleets that provide point to point service**	**Dynamic complete streets utilization (ATDM for all roads)**	**High-speed/flexible/modular transit**	**Smart and dynamic parking management**
Levers	• Subsidies (pricing) • Toll relief • Lane dedication/prioritization	• Use thresholds/street criteria • Roadway design & infrastructure; Lane dedication/prioritization • Pricing	• Capital and operations needs/investments • Fares & subsidies (pricing) • Partnerships for advanced delivery • Multimodal integration/partnerships with new mobility services	• Pricing • Availability/ scarcity • Costs • Curb utilization/policy (flexible curb space)
Metrics	• Travel time & distance (equivalency) • Wait time • Modeshare • Cost per mile/trip • Value of time • Occupancy • Cost-benefit	• Cost-benefit • Modeshare • Safety • Street utilization	• Fleet characteristics • Trip characteristics (travel time, travel length, wait time, occupancy, etc.) • Use propensity • Cost—benefit	• Cost—benefit • Parking occupancy/turnover • Curb utilization for pick up and drop off
Uncertainties	• Penetration rate • Fleet size • Costs • Infrastructure needs (hard and soft including Information & Communications Technology, roadway re-design, wayfinding, etc)	• Data & technology availability • Infrastructure needs (adaptability/conversion) • Land use integration (built environment)	• Mode integration Multimodal sensitivity Infrastructure needs Environmental implications	• Data availability/local parking inventory Availability of technology

High-Level Scoping

The preliminary scopes gave the group a broad motivation for the scoping workshop, which then narrowed the focus area to the border points of entry (POE), as the cross-border model will be used in the beta test.

In the workshop, the group first started with the goal to promote non-SOV border crossings to reduce congestion and identified a variety of levers to support that goal. The potential levers discussed were:

- Improve transportation options [PEVs, improved transit/trolley services]
  - Transit LOS
  - Facilities & improved micro-mobility availability
- Incentivize shared auto crossing
  - HOT/HOV lane facility on roadways to border
- SR125 Toll Policy / Rate [occupancy specific, dynamic]
- New crossing with toll road access

The next step was to enumerate the metrics that could be collected to evaluate the efficacy of these levers. The potential metrics discussed were:

- Border crossing mode share (auto, non-auto)
- Number of non-auto crossings
- Vehicle occupancy
- VMT / VHT
- Transit boardings at border
- Emissions (regional)
- Toll road revenue
- Some type of freight measure

Finally, the exogenous uncertainties that could affect the lever efficacy were identified as:

- Travel behavior (perceptions of new options, weather impacts)
  - Perception of micro-mobility
- Auto competitiveness
- Border crossing wait times by mode (auto vs. non-auto)
  - Arrival behavior not suited to transit service (metered by customs)
  - Crossing wait times at particular stations are independent
- Ability to invest in new options
- Land use changes shift destinations away from transit services

    - Mexico land use and transportation
- Economic factors changing type and level of travel
    - VOT of Mexican residents
- PnR availability and cost (Mexico & U.S.)

The group also explored a second high-level scope supporting the goal to maximize mobility through the opening of a new point of entry at the border with exclusive toll road access. The primary metric defined for this lever would be toll revenue and the group discussed the potential for TMIP-EMAT to be useful for toll and revenue forecasting. The uncertainties related to this lever were identified as:

- Travel behavior (perceptions of new options, weather impacts)
    - Perception of micro-mobility
- Auto competitiveness
- Border crossing wait times by mode (auto, non-auto)
    - Arrival behavior not suited to transit service (metered by customs)
    - Crossing wait times at particular stations are independent
- Ability to invest in new options
- Land use changes shift destinations away from transit services
    - Mexico land use and transportation
- Economic factors changing type and level of travel
    - VOT of Mexican residents
- PnR availability and cost (Mexico & U.S.)

Identify Model Functionality

The next step in the workshop was to identify the existing or needed model functionality to represent each lever and uncertainty variable in the model. As part of the discussion, the group reviewed the estimated level of effort involved to develop any new model functionality.

Table 19: SANDAG Model Variable Identification by Lever

Lever	Potential Model Variables to Represent Lever
Transit System Enhancements	• Transit mode represented in network (alignment and stop coding) • Transit travel times • Transit headway • Transit fares • Availability and preference for KnR transit alternatives (representing TNC connections) • Access and egress speed (representing a micro-mobility complement to transit)
Managed lane facilities	• SR 125 toll policy (tolling by auto-occupancy and time of day • Definition of dedicated lanes in highway network
Personal Electric Vehicles (micro-mobility)	• Increase bike and walk speeds and distance thresholds
New Point of Entry (POE)	• Toll rates by time-of-day with correlation to POE waiting times by mode

Table 20: SANDAG Model Variable Identification by Exogenous Uncertainty

Exogenous Uncertainty	Potential Model Variables to Represent Uncertainty
Travel behavior	• Bias constant of auto, transit, and/or walk mode alternatives • Value of time coefficient for Mexico residents • Walk speed reflecting increased use of micro-mobility
Auto competitiveness	• Auto operating costs (i.e., lower for electric vehicles)
Border crossing wait time by mode	• Border crossing wait time by TOD, mode, POE and visa documentation • Market share of visa holders
Land use	• Retail employment (single site development or proportional shift) • Mexico land use (vary population distribution and density)

Exogenous Uncertainty	Potential Model Variables to Represent Uncertainty
Border PnR availability and cost	• No suitable variable available

Select and Define Scope

After considering the work required to implement each lever, uncertainty variable and metric, the group selected the model variables that would be leveraged through TMIP-EMAT keeping in mind that the total number of policy-levers and independent uncertainty variables have a linear relationship with the required number of Core Model runs, for example: 10 policy-levers/uncertainty variables require 100 Core Model runs, while the number of metrics has no impact on the number of Core Model runs required.

Table 21 summarizes the selected levers and exogenous uncertainties. Originally, nine levers and uncertainties were scoped. Through the development and testing process, one lever and one uncertainty variable were dropped.

Table 21: SANDAG Selected Levers and Uncertainty Variables

Policy-Lever/Uncertainty Variable	Minimum	Default	Maximum	Distribution (applies to Exogenous Uncertainties only)	Unit/Correlations/Other Notes
LEVER: SR125 toll policy	0.5	1.0	1.5	NA	Applied as a factor to existing tolls on SR125
LEVER: transit headway	0.5	1.0	1.5	NA	Applied as a factor to all transit service connecting to the POEs.
LEVER: transit mode	Local Bus	Base (Trolley)	High-Speed	NA	Lever changes the existing Blue Line trolley connection between downtown and San Ysidro.
LEVER: transit fare	Half	Base	Double	NA	Applied as a factor to transit fares, to simplify the implementation this was implemented as a categorical variable with 3 levels: half, base, double rather than a continuous factor. *This lever was dropped after sensitivity testing*
EX. UNCERTAINTY: Boarder Crossing Wait Time by Poe/Hour/Mode/Visa	0.3	1.0	2.8	PERT—relative peak 0.33, gamma 4	Applied as a factor to all wait times across POE, mode, and hour. Magnitude of factor was based on historical data.
EX. UNCERTAINTY: Walk speed (micro-mobility)	0.5	1.0	2.0	Triangle, relative peak 0.5	Implemented as a change in walk speed and distance for transit access / egress and non-motorized modes
EX. UNCERTAINTY: Auto operating costs	0.5	1.0	1.5	Triangle, relative peak 0.5	Factor applied to auto operating costs.
EX. UNCERTAINTY: Mexico land use	Lower density in Mexico	Base	Higher density in Mexico	Uniform	High density scenario shifted population primarily to the San Ysidro POE.
EX. UNCERTAINTY:VOT factor for Mexico residents	0.5	1.0	1.5	PERT—relative peak 0.33, gamma 4	Applied as a factor to VOT of Mexico residents. *This variable was dropped after sensitivity testing*

The scoped metrics and their optimization hint are shown in Table 22. Note that the regionwide emissions metric was later dropped for two reasons: it did not show much response to the variables and it was not an accurate measure given that the entire model stream was not run. For example, the change in auto operating costs would have an impact on regionwide mode shift and thus the regional emissions, but this test would only show the mode shift on cross-border trips.

Table 22: SANDAG Scoped Metrics

Metric	Optimization Hint
Border Crossing Drive Alone Mode Share	minimize
Border Crossing Number of Walk Tours	maximize
Border Crossing Shop Purpose Tours	informational
Border Crossing VMT	minimize
Border Crossing VHT	minimize
SR125 Toll Road Daily Flow	maximize
Transit Ridership (Boardings at Border Station)	maximize
Transit Travel Time from Downtown to Border (Southbound)	minimize
Transit Travel Time from Downtown to Border (Northbound)	minimize
Regionwide Emissions *removed from later analysis*	minimize

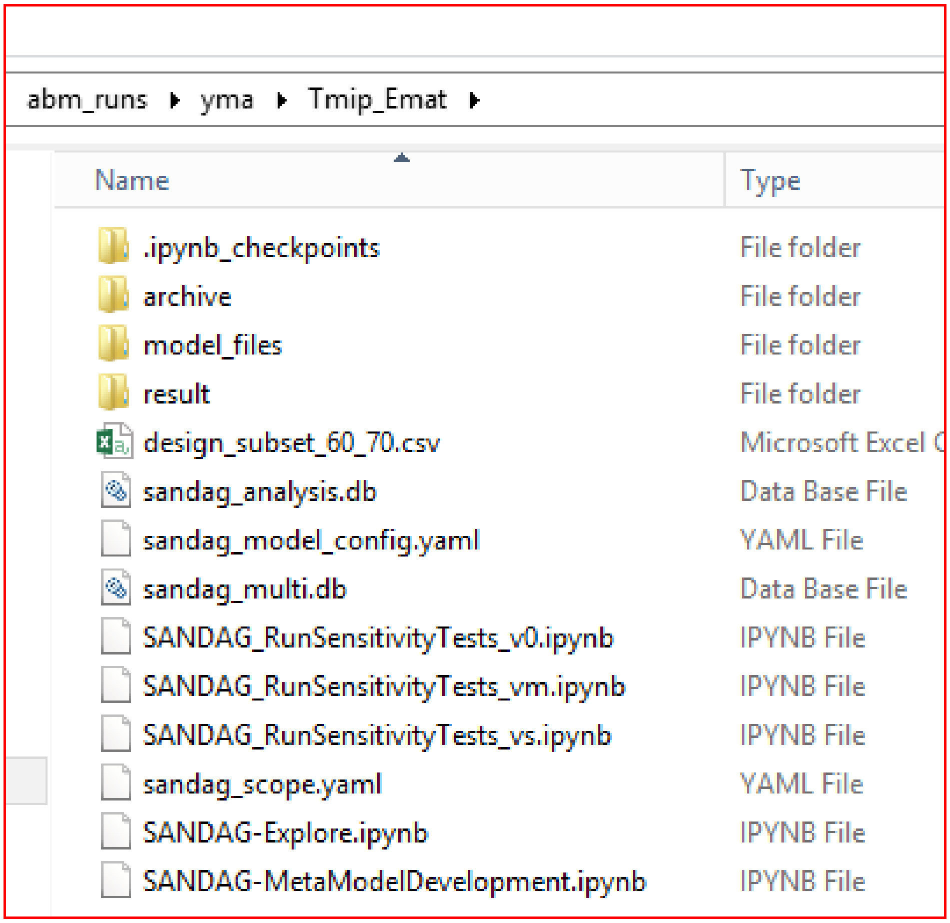
A.3.5 Implementation and Model Run Experience Report and Feedback

This section summarizes the TMIP-EMAT API development, ABM model extensions, and experiences conducting the core-model runs.

API Development and Model Extensions

The SANDAG regional model was recently updated to a new travel demand model software platform, Emme. SANDAG initially engaged INRO for support to programmatically modify networks and set parameters within the model. The variables that required this included the toll road setting and the transit modes. Later in the beta-test, SANDAG leveraged their modeling support contract for assistance in developing the functionality to set levers and uncertainty variables. SANDAG staff implemented the metric processing.

TMIP-EMAT was installed within the ProgramData folder on each model workstation to be consistent with the Anaconda installation location. SANDAG structured the model such that a working folder of the TMIP-EMAT configuration and API support files was organized and the model was configured for each experiment run (Figure 10).



Source: SANDAG

Figure 10: SANDAG TMIP-EMAT Configuration and Support File Organization

Core Model Runs

The univariate sensitivity tests were conducted on a single machine first manually and again when the API was fully functional. There were three rounds of testing and troubleshooting against the univariate sensitivity tests before the full LHS experiment design was run successfully.

The seventy LHS experiments were distributed across six computers with the results written to a common SQLite database. The distributed workload allowed the LHS experiments to be complete more than twenty days of model runs in less than four days (Table 23).

Table 23: SANDAG Experiment Run Statistics

Server	Experiment ID Start	Experiment ID End	Total Run Time	Number Completed	Average Time
tiger	1	12	84:29:00	12	7:02:25
monty	13	24	88:21:00	12	7:21:45
nittany	25	36	84:00:00	12	7:00:00
wildcat	37	48	82:14:00	12	6:51:10
aztec	49	60	82:10:00	12	6:50:50
triton	61	70	69:43:00	10	6:58:18
Total			**490:57:00**	**70**	**7:00:49**

The full databank from the model is 100GB and was deemed to be too large to be a feasible archive. A 250MB archive set of outputs from each experiment were saved locally on each machine within the archive folder to be consolidated on a network shared drive archive as a manual process following the model runs.

Run and Setup Issues and Challenges

While SANDAG modelers had experience with Python, they were relatively new to Jupyter Notebooks. There were challenges to set up Anaconda and get the correct packages installed. Once everything was installed, there were several hurdles to operate the model platform through TMIP-EMAT. These were related to the Emme Python environment and parameter settings, the report generation procedure, and conducting a hybrid model run along with miscellaneous issues.

Although Emme is programmed in Python, it requires the legacy 2.7 version of Python. TMIP-EMAT runs on the latest version of Python and thus Emme and TMIP-EMAT must operate separately. This required the development of scripts that would configure and launch the regional model, as opposed to calling Emme directly from TMIP-EMAT. Most of the API functionality was implemented in stand-alone python scripts outside of the TMIP-EMAT codebase with the TMIP-EMAT API simply calling the python script. This method is similar to the mechanism that ODOT developed to manage their VISUM model.

The implementation of stand-alone scripts to setup the model led to a challenge to restore the default parameter values, especially when a variable is applied as a factor to the default (e.g., auto operating cost). The Emme databank maintains a schema log that could not be easily overridden so SANDAG extended their model definition to maintain the default values of parameters within the databank. Instead of duplicating the databank, which is very large, there are a set of “original” fields that maintain the default value. The “originals” are not part of the normal model design, but this extension will be useful to programmatically run other scenarios.

SANDAG has a SQL Server-based summary report generation process that runs at the end of the model to produce metrics. This process was elected to not be used for TMIP-EMAT because of the added runtime and the complexity to access metrics through SQL Server. Instead, SANDAG built custom procedures to summarize metrics into csv files that could be easily imported using TMIP-EMAT’s standard parsing utilities.

There were unanticipated challenges to conducting a partial run of the regional model. A full model run was necessary to provide the required inputs to the cross-border model and a new model sequence was defined to run only that model (see Figure 11). Developing the API was found to test the limits of understanding around the regional model scripts and structures, while conducting a hybrid model run even more so.

Set / reset all	☐	☐	☐
Skip highway assignments and skims	☐	☐	☑
Skip transit skims	☐	☐	☑
Skip core ABM	☐	☐	☑
Skip other simulation model	☐	☐	☐
Skip commercial vehicle sub-model	☐	☐	☑
Skip truck sub-model	☐	☐	☑
Skip external-internal sub-model	☐	☐	☑
Skip external-external sub-model	☐	☐	☑
Skip trip table creation	☐	☐	☐
Skip final highway assignments			☐
Skip final transit assignments			☐
Skip data export			☐
Skip data load request			☑
Skip delete intermediate files			☑

Source: SANDAG

Figure 11: SANDAG Model Configuration

Issues encountered during the initial rounds included:

- Work structure/path was not consistent
- Emme bugs—#name issue
- TMIP-EMAT bug in univariate experiment design[7]
- Master_run procedure has to be changed
- Variable setting issue—link attribute built up
- Inconsistency issue—scope variables, PM not consistent with the code
- Some input variable extreme values caused model crashes and more constrained values needed to be used
- Improper operations

[7] Resolved in EMAT version 0.2.4

However, once the hurdle of scripting the API and testing the model response was cleared, model operation through TMIP-EMAT worked very smoothly. Following the example Jupyter notebooks provided, experimental design, running experiments across multiple systems, and visualizing the results all worked well.[8]

A.3.6 Analysis Feedback

This section summarizes SANDAG's feedback and the workshop discussion by analysis type.

Univariate Sensitivity Test Review

The sensitivity tests illuminated variables that were not very sensitive in the model as well as variables that had a high sensitivity, such as the border crossing wait time. Manual testing in particular was very important in making sure the analyst knew the model inside and out.

Through the sensitivity tests, SANDAG gained better insight into the model operation and is more confident on what the model effectively represents in application. For example, the model response to transit fare changes motivated a closer review of the cross-border operation. Once the transit fare was confirmed to be set correctly, a review of the model parameters and border crossing mode to trip mode was conducted. The border crossing mode was found to be influenced primarily by border wait time and thus a change in the trip mode logsum due to changes in transit fare did not substantially shift the utility of non-auto border crossing modes. As a result, the transit fare lever was removed from the EMAT scope.

The sensitivity tests also demonstrated where more specific metrics would be useful to confirm correct operation and better explain the inner workings of the model. For example, the increase in POE waiting time led to both a decrease in drive-alone mode share due to the larger actual wait time difference between walk and auto modes but also an increase in VMT (see Table 24). Further investigation revealed that the extreme wait times create a shift in POE choice to ones with a lower wait time, but a longer travel distance. A POE specific measure of mode share and number of trips would help illuminate these effects.

[8] Example notebooks are available here: https://tmip-emat.github.io/source/emat.examples/index.html

Table 24: SANDAG Border Crossing Wait Time Sensitivity Test Results

Metric	Base	Border Crossing Wait Time Factor = 0.3	Border Crossing Wait Time Factor = 2.8
Border Crossing Drive Alone Mode Share	25.0%	26.9%	17.8%
Border Crossing Number of Walk Tours	26,611	20,947	46,143
Border Crossing Shop Purpose Tours	59,817	60,150	59,000
Border Crossing VMT	1,754,027	1,725,957	1,811,215
Border Crossing VHT	3,001,253	2,975,777	3,039,822
SR125 Toll Road Daily Flow	29,264	29,166	29,630
Transit Ridership (Boardings at Border Station)	15,213	13,824	19,952
Transit Travel Time from Downtown to Border,PM	61	61	61

Metric	Base	Border Crossing Wait Time Factor = 0.3	Border Crossing Wait Time Factor = 2.8
Border Crossing Drive Alone Mode Share	0	7.7%	-28.9%
Border Crossing Number of Walk Tours	0	-21.3%	73.4%
Border Crossing Shopping Tours	0	0.6%	-1.4%
Border Crossing VMT	0	-1.6%	3.3%
Border Crossing VHT	0	-0.8%	1.3%
SR125 Toll Road Daily Flow	0	-0.3%	1.3%
Transit Ridership at Border Station	0	-9.1%	31.2%
Transit Travel Time from Downtown to Border,PM	0	-0.5%	0.3%

For SANDAG, the univariate sensitivity tests were sufficient to flush out any issues with the lever or uncertainty variable definitions and to gauge the impact of changes in the model inputs on metrics.

Visualization and Feature Scoring using Meta-Model Results

The group found the scatter plot visualization and feature scoring to be the most useful presentations of the model results. The group discussed how the Mexico Land Use uncertainty variable is a major contributor to change in border crossing VMT and VHT, but is not strongly correlated to the other variables. This was a surprising result given that the univariate sensitivity test implied that the Land Use shift also impacted the SR125 Daily Flow and Transit Ridership.

Feature scoring also demonstrated the relatively low importance of auto operating costs and high importance of the border wait time. SANDAG concluded that feature scoring showed the most intuitive results that could be taken to policy makers.

The high statistics on the meta-model results prompted a discussion of how TMIP-EMAT could develop the meta-models with fewer experiments. The group discussed the purpose of the LHS experiment and the value in when it reveals unexpected model results. However, where the relationships between variables and metrics are well understood, the sampling does not necessarily help. But, the modeler should not limit model inputs unless the relationship within the core model is so simplistic as to guarantee the change in metrics.

Directed / Robust Search

The directed search analysis tool did a good job demonstrating how maximizing SR125 Toll Road Daily Flow is contradictory to the transit and non-motorized related metrics. The parallel coordinate lines showed how giving up some benefit on the toll road volume would improve the other metrics.

The group did some investigation into the extreme points of the robust search output and concluded that some outliers were leading to counter-intuitive results. For example, a low-SR 125 Toll Charge factor was included in the solution to minimize worst-case VHT, which does not seem reasonable. Lower tolls should create more auto travel and higher VHT, unless the toll facility is generally uncongested, which is not the case for SR125. Use of a 99th percentile value could be useful to avoid outliers and confirm results.

A.3.7 Level of Effort

The level of effort exceeded SANDAG's expectations. It was challenging to develop and integrate input variable controls as well as to develop a new metric summary procedure to support the TMIP-EMAT API. In-depth knowledge of the core model script and structure was required to properly implement and debug the API operation.

Overall, the implementation took about 4-months' work of a SANDAG employee working 50% on this project with the addition of a modeling consultant for about 1.5 months at 30% time. Repeating the process with a different set of variables and metrics given the existing API is anticipated to require about 1-2 weeks, excluding testing.

If SANDAG were to utilize TMIP-EMAT again, they would most likely engage consultant support due to staff availability and existing workload.

A.3.8 Lessons Learned and Overall Feedback

SANDAG was able to enlist six modeling workstations to complete the 70 model runs defined in the beta-test experiment design within four days. Yet, the biggest challenge SANDAG cited throughout this process was model run time. SANDAG worked around the runtime of the full model by only applying TMIP-EMAT with a sub-component of their overall model, however this component runtime including assignment was still 7 hours. SANDAG does not anticipate that TMIP-EMAT could be effectively utilized with their full model, which has a runtime between 32 and 40 hours. The group discussed methods to reduce the overall model runtime such as reducing the number of experiments required and developing an option to configure TMIP-EMAT as a background process. However, SANDAG found the runtime with the 7 hour subcomponent to be prohibitive and does not see the full regional model as being a feasible option to utilize TMIP-EMAT. However, the ability to programmatically operate the full model is attractive, particularly for sensitivity testing, and may be extended to the full model.

Overall, SANDAG found that the methodology needs to be further vetted and modeling assumptions investigated by a greater community before it could be integrated into their planning process. Due to the beta-test timeframe and difficulty implementing the programs to set levers and uncertainty variables, the test results were coarser than would have been implemented under a planning study. The value in TMIP-EMAT running the model against multiple dimensions many times was to demonstrate the implications of the modeling assumptions. Model assumptions were highlighted through this test and the group discussed how they restricted the ability of planners to pursue questions, such as:

- Fixed trip table for border crossing demand.
- Fixed relationship between auto, walk, and visa wait times.
- Sensitivity to transit cost.
- Metric limitations due to entire model not being run.

SANDAG concluded that there is still work to be done to get decision-makers to adopt Robust Decision-making. The group reported that it is hard enough to utilize scenario analysis over point forecasts.

Scoping

SANDAG found the scoping workshop to be helpful to understand the TMIP-EMAT process, develop the levers, metrics and uncertainties, and set the direction for the workplan. However, SANDAG recommends that the learning curve for TMIP-EMAT is too steep to have conducted this workshop without guidance and would not conduct it independently.

To support the visualization, FHWA consultants implemented a method to define shorter names for the metrics than were originally specified. This allowed for cleaner graphics in the visualization and future scopes should utilize shorter names if possible.

Suggestions for improvement

Better integration with Travel Demand Model software

The custom coding required to specify the model input variables and produce metrics was challenging for SANDAG. An approach that reduces the need for custom coding would require tighter integration of TMIP-EMAT with the travel demand model software packages, or development of the meta-model procedures within the TDM software.

Simpler visualizations

Translating the information to more intuitive presentations that would be accessible to less technical staff would be valuable.

Adaptive experiment design to limit core model runs

Run time being the largest concern, methods that would reduce the number of model runs for variables that have little variance would be advantageous.

Integrate with a strategic model

The simpler structure of a strategic model combined with the faster run time would obviate the key challenges SANDAG experienced in testing TMIP-EMAT.

Suitable Applications of TMIP-EMAT

SANDAG is currently updating their regional model to include capabilities to represent AVs and TNCs. The application of these capabilities, i.e., the specific parameters to use in each scenario, will be guided by a set of research white papers and applied in a scenario planning type approach. SANDAG anticipates 40-50 scenarios to be tested, but does not plan to utilize TMIP-EMAT with the enhanced regional model.

Based on the runtime concerns and difficulty to extend the regional model with programmatic controls necessary to be run through an API, SANDAG sees the highest potential for future application of TMIP-EMAT to be with a sketch planning model, such as Delloitte's FutureScape[9] or VisionEval[10]. A strategic model could be used to develop the preferred scenario for running through the full travel demand model. TMIP-EMAT coupled with a strategic model would be closer to a real world application in SANDAG's view.

A.3.9 References

SANDAG model documentation

https://www.sandag.org/index.asp?classid=32&fuseaction=home.classhome

[9] https://www.futurescape.ai/

[10] https://visioneval.org/

NOTICE

U.S. Department of Transportation
Federal Highway Administration
Office of Planning, Environment, and Realty
1200 New Jersey Avenue, SE
Washington, DC 20590

December 2019

FHWA-HEP-20-015

U.S. Department of Transportation
Federal Highway Administration