R for Travel Demand Modelers: Experience at the Oregon DOT

Brian Gregor, P.E.
Oregon Department of Transportation
2/17/2011

Oregon Department of Transportation

Overview

- Background
- History of the use of R at the Oregon DOT
  - Examples
  - Lessons learned
- Wrap up
  - Overall lessons learned
  - Ideas for future

Background

My Background

- Worked for ODOT for 26+ years
- Worked as a transportation analyst and modeler for 14+ years
- Prior to travel modeling, worked on transportation planning, land use planning, and transportation/land use policy development
- No formal training in computer programming (although I do read about programming and computer languages on the side)

Travel Modeling at ODOT

- Transportation Planning Analysis Unit (17 persons)
  - Modeling
  - Facility Analysis and Simulation
  - Systems Analysis
- Travel modeling for 3 small MPOs and over a dozen smaller urban areas
- Statewide modeling
  - Integrated statewide modeling
  - Transportation sector greenhouse gas modeling

5

Oregon Department of Transportation

Transportation Planning and Modeling in Oregon

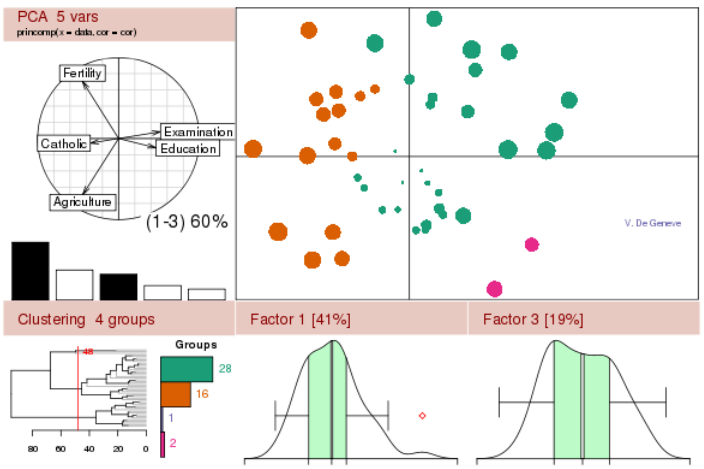
- Strong state planning requirements
  - Land use
  - Transportation and land use coordination
  - Rules and legal challenges
- High expectations for modeling
  - Public and decision makers expect models to be used to inform policy
  - Support for advancing model development
- Good coordination between state and regional agencies

6

Discovering R

- Dissatisfaction with toolset being used to build and implement models
- Interested in more modern numerical computing environment (e.g. MATLAB)
- Interest in data visualization
- Interest in open source software
- Came across R in an article about geostatistical analysis using GRASS and R

The R Project for Statistical Computing

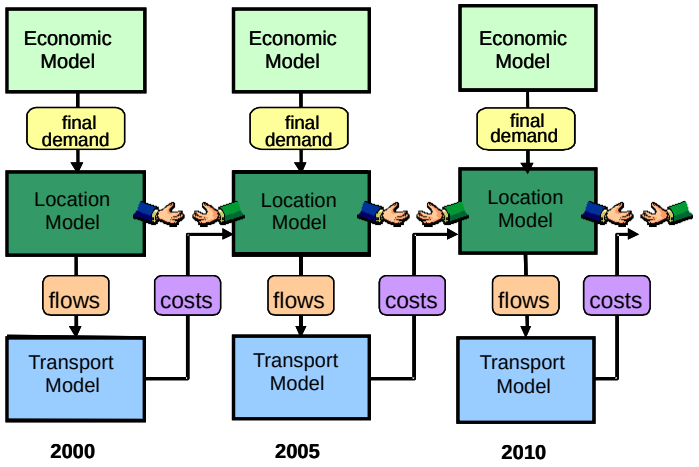


History and Examples of Use

Beginnings

- Started using in 2001
- Learned the basics by using it to analyze survey data
- First big win was using R to automate running our 1st generation statewide model

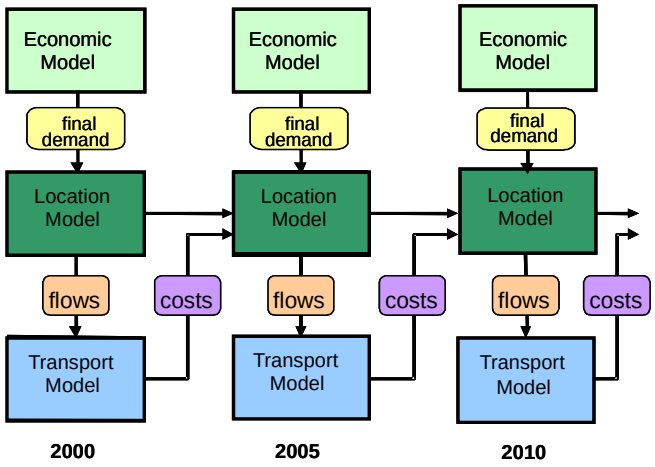
Automating the 1st Generation Statewide Model



11

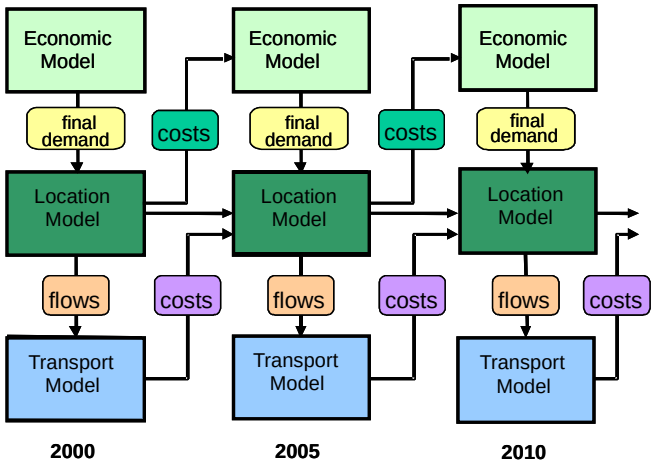
Oregon Department of Transportation

Automating the 1st Generation Statewide Model



12

Extending the 1st Generation Statewide Model



13

Oregon Department of Transportation

Beginning Lessons

- R is a full featured programming language
  - Can be used to control other programs
  - Can parse and manipulate text files
- A modest programming effort can yield big productivity gains
- R can be used to graft extensions onto models written in other platforms

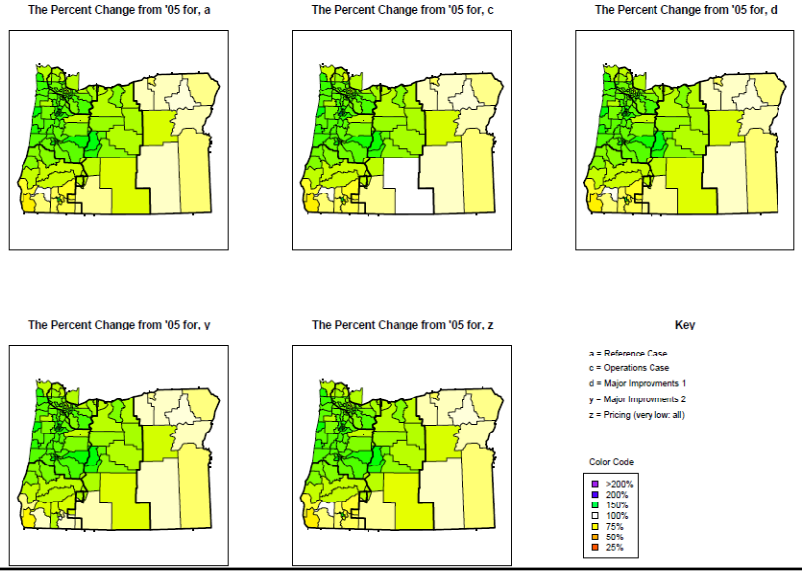
14

Managing Inputs and Outputs

- Since R could produce well formatted text files, we could use it to automate the input file production
  - Bridge study – identifying possible weight restricted links
- Using the excellent graphical capabilities of R, we produced a large number of graphical outputs
  - Done automatically after each model run
  - Developed new ideas and extended ideas over time

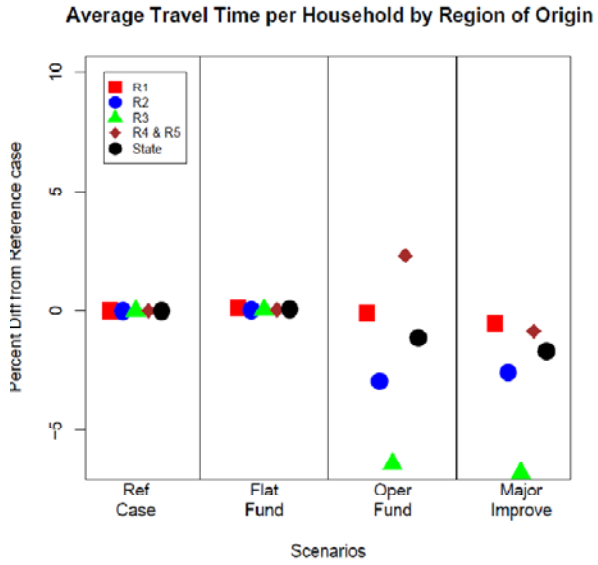
15

Mapping Output Example



16

Custom Chart Example



17

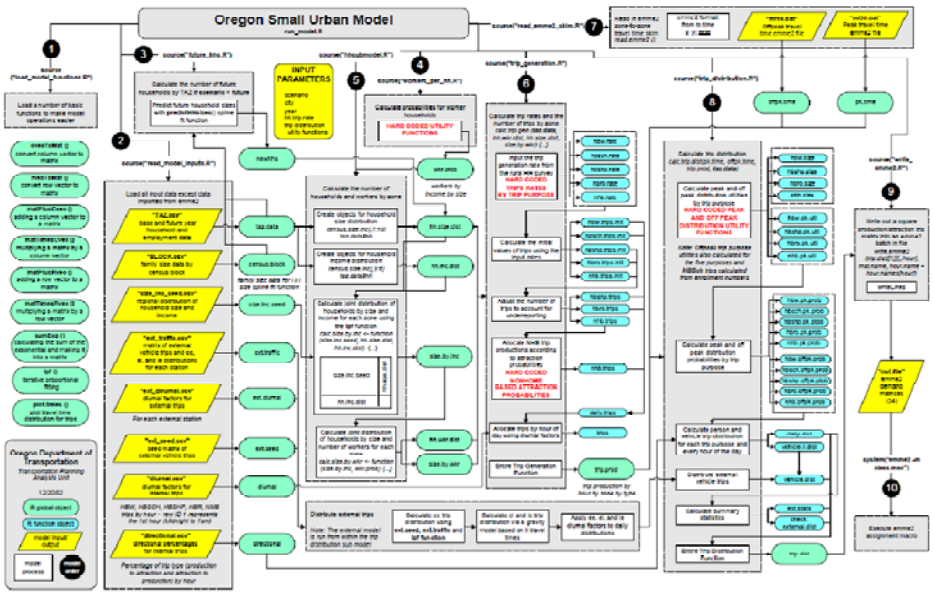
Lessons: Managing Inputs and Outputs

- Using R to generate input files:
  - Reduces tedium for data entry
  - Improves capabilities and reduces error
  - Can improve documentation and reproducibility
- Using R for visualizing outputs:
  - Can generate standard graphics automatically
  - Can visualize whatever you can imagine
- Collaboration and using a common object naming approach

18

Implementing a New Travel Demand Model

- Oregon Small Urban Model
  - New model, based on recent household survey
  - 3-step, destination choice
- Use R's built in matrix capabilities to do all of the model computations (except for traffic assignment)
- Use R to control the process, including calling the traffic assignment software



Lessons: Implementing a New TDM

- R can be used to develop relatively large model applications
- The combination of specifying a common small urban area model structure and implementing it in R:
  - Increased the speed at which we could develop and deploy small urban models
  - Simplified model applications
  - Increased our productivity

21

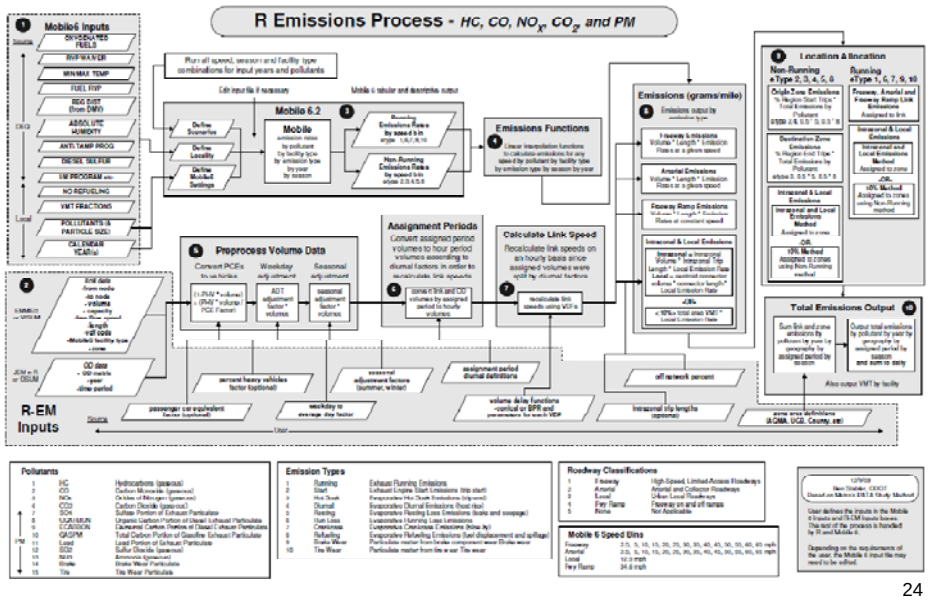
Lessons: Implementing a New TDM

- Features of the language simplify programming and reduce need to program loops
  - Built in vectorization of the language
  - Variety of data types and corresponding iterators
- Data structures are important to the language
  - The choice of data structure can simplify understanding and programming
  - Important to communicate the data structure in the program code

22

Automating Post-Processing for Air Quality Analysis

- OSUM interested MPO modelers in the potential of R for travel modeling
- MPOs were interested in having the air quality analysis process automated
- Requires managing data from multiple sources and running multiple applications
- Uses R to control the process



Lessons: Air Quality Modeling

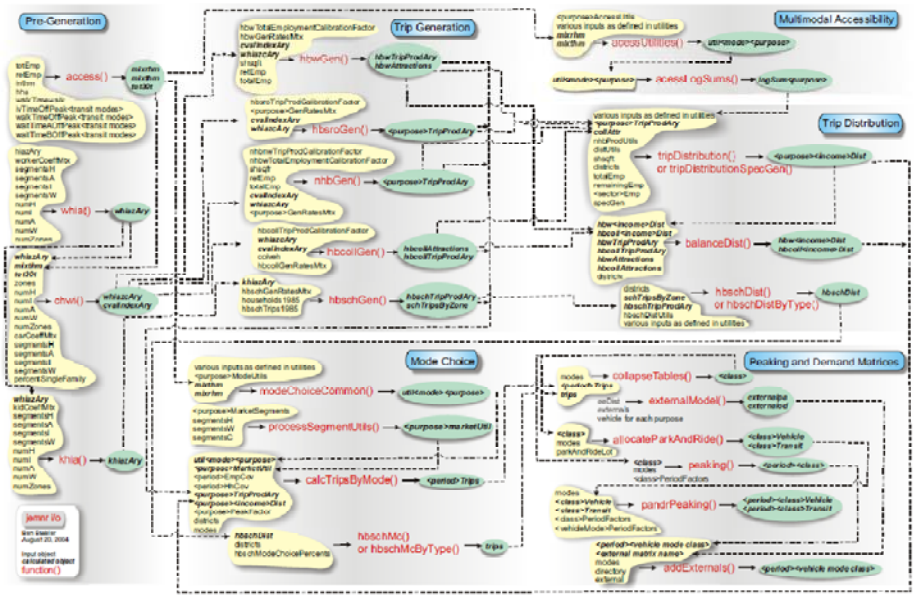
- R very capable for tying together multiple data sources and model software and managing the process
- R facilitated rapid development
- Reduced the amount of time required for air quality analysis
- The flexible structure of lists makes it easy to tie together related data that is represented in several different forms
  - Caution: don't make lists too complicated or it can be difficult for others to understand the code

25

Implementing a Metropolitan 4-step Model

- Next big project was to implement a metropolitan model in R
  - Joint Estimated Model in R (JEMnR)
- Planned for it to be a joint development effort
- Mode choice increased the complexity substantially

26



27

Lessons: JEMnR

- A structured approach is very important for developing model code that is efficient and maintainable: data, functions, modules
- Make model code a calculation engine. Changeable parts of the model should be input data.
  - Example: All utilities in JEMnR are input data. Utilities can be changed without making any code changes
- JEMnR enabled TPAU staff to handle the workload of developing 3 new MPO models
- R does have a learning curve

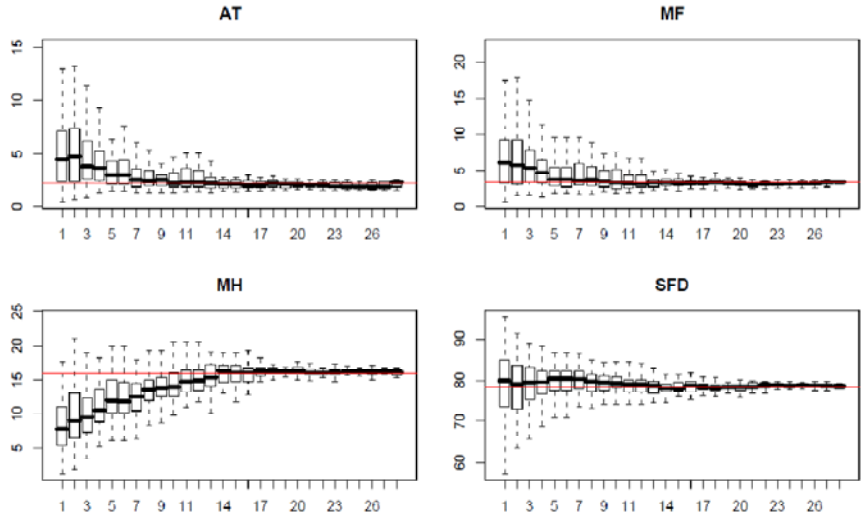
28

Aggregate Land Development Model (ALD)

- ALD is a module in our 2nd generation StateWide Integrated Model (SWIM2)
- Due to the complexity of SWIM2, it was decided to simplify the land development component, using the SWIM1 approach (implemented in R)
- R was used to estimate the parameters of the model as well as implement the model

29

Using Evolutionary Algorithms to Estimate Model Parameters



30

Lessons: ALD

- R facilitated tying the development of the model implementation code and model estimation processes together and enabled more advanced estimation procedures to be used
- R is a good platform for learning about, experimenting with, and implementing new methods
- An R-based model can be managed as a component of a model that is based in another computer language

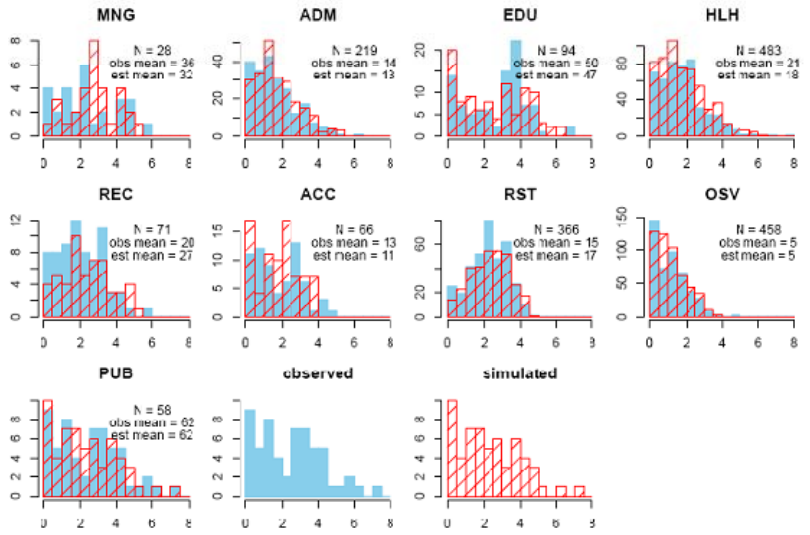
31

Land Use Scenario DevelopeR (LUSDR)

- LUSDR was developed to support a regional planning process to designate urban reserves
- LUSDR uses a different approach to land use modeling
  - Develops many scenarios to assess risks
- R was used for entire process, from data development to implementation
- R used exclusively for data visualization

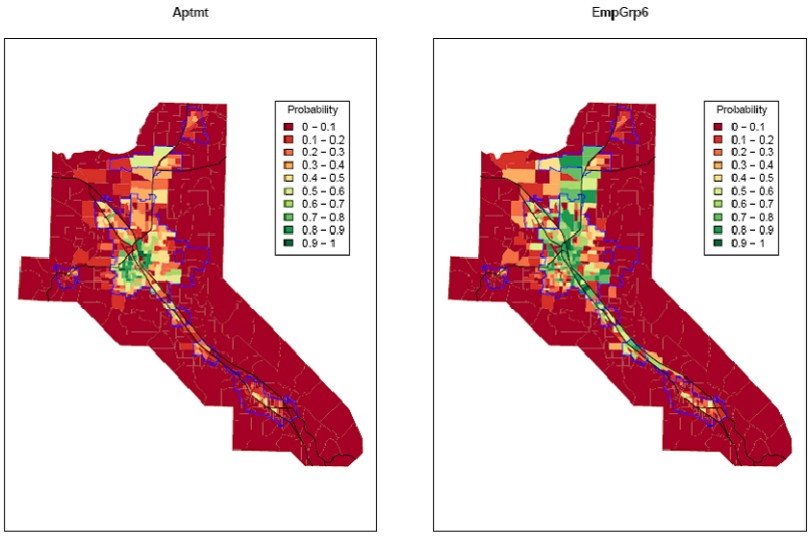
32

Using Graphics in Model Development



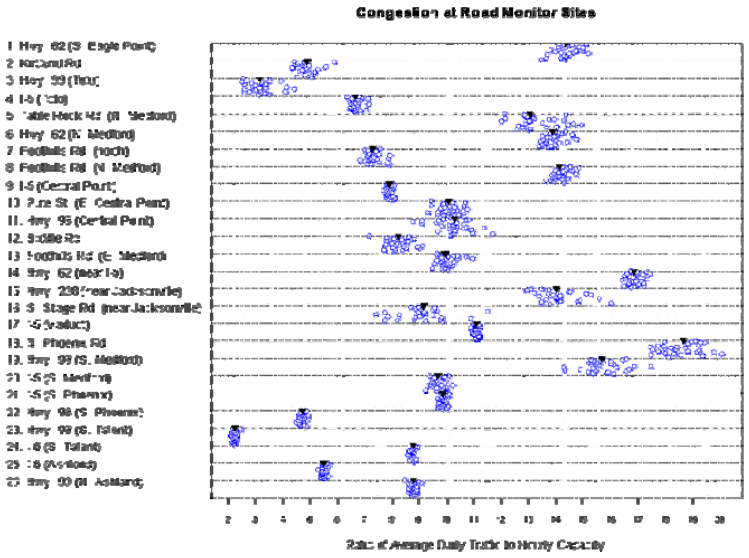
33

Using Graphics in Model Development

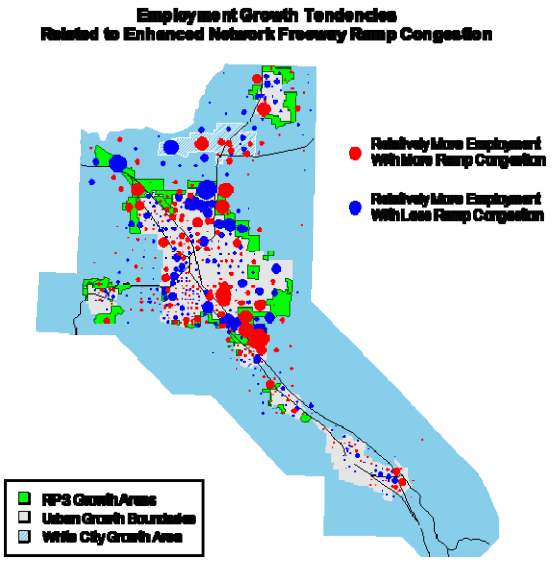


34

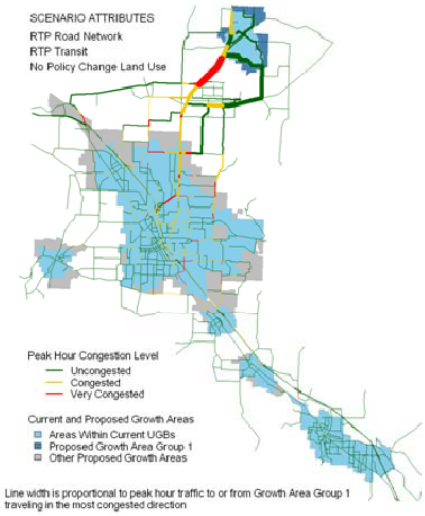
Showing Results of Multiple Scenarios

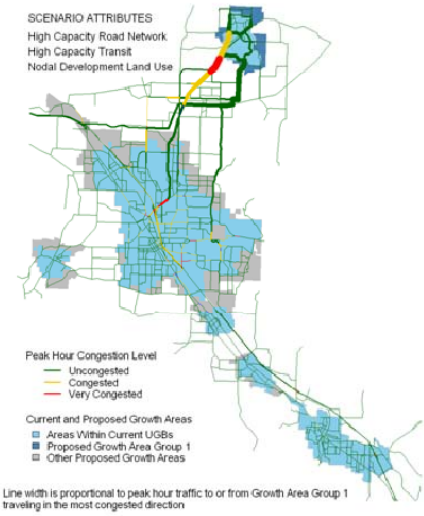


Analyzing Results Graphically



Comparing Scenarios





37

Oregon Department of Transportation

Lessons: LUSDR

- R facilitated development of a new model quickly
- Wealth of packages enables testing and deployment of different methods
- Agile development process using R enabling successive refinement
- Vectorization helpful for list-based processes as well matrix-based processes

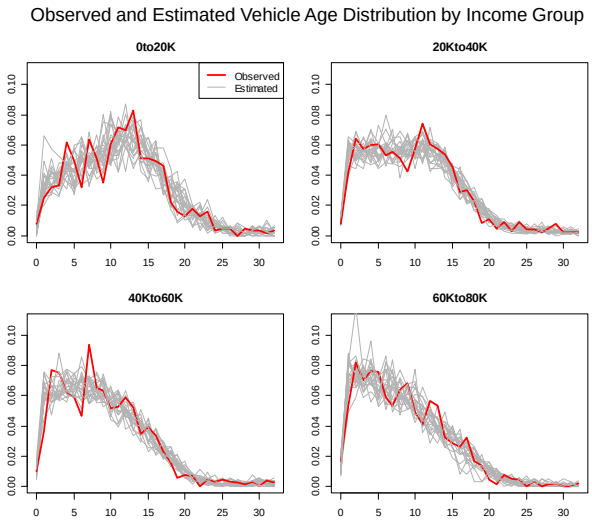
38

GreenSTEP

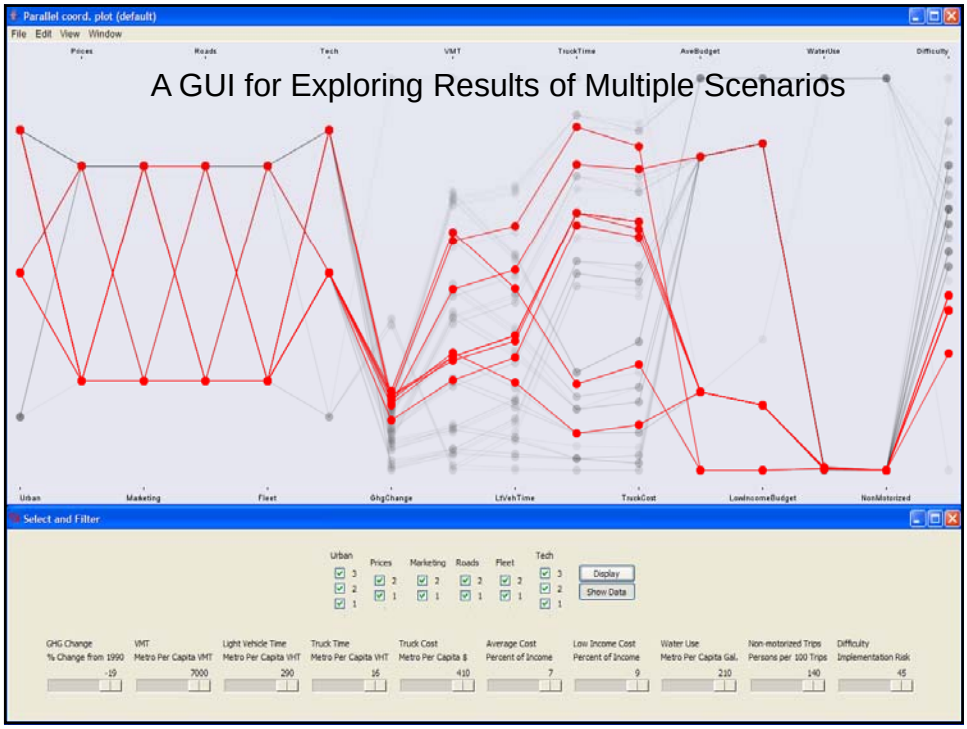
- A model for calculating greenhouse gas emissions from the transportation sector considering land use, transportation, pricing, vehicles, fuels
- Entire development chain in R: data prep, estimation, implementation, calibration & validation, outputs
- Because model estimation with R scripts, it is all replicable and easily modified
- Build script to assemble model from components

39

Comparing Observed and Estimated Values Considering Stochastic Effects



40



Oregon Department of Transportation

Lessons: GreenSTEP

- Development in R enabled development of a new model quickly
- Development in an integrated way across all steps enable rapid reestimation and revision
- Agile development process using R enabling successive refinement
- Reproducibility

42

Overall Lessons Learned and Ideas for Future

Lessons Learned

While R is not the best language for systems programming, it has other advantages that outweigh limitations

- Good interpreted programming environment facilitates rapid learning and development
- Built for data analysis. Easy to do things that should be easy to do.
- It can be used every day for routine data tasks so it is easy to stay current
- Very large number of add on packages
- Open source

Lessons Learned

- R is easier to learn to the point of being productive than other programming languages, but it is a programming language nevertheless
  - Can be challenging to get people out of Excel and into R
  - Requires time to learn and experiment
  - Need to build capabilities over time
- Even so, all of our major projects paid off in time savings and improved capabilities in a short amount of time

45

Lessons Learned

- R is a very useful tool to help modelers at state DOTs, MPOs, etc. to develop advanced models without being advanced computer programmers
- The environment facilitates learning and testing of modeling concepts and bringing those concepts into models
- Models developed in this way are not black boxes to the modelers and can more readily maintained and extended by the modelers

46

Ideas for the Future

- Develop a community of transportation modelers using R
  - Website to share ideas and R scripts
  - Collaboration on projects
- Develop a structure supported by packages of functions to facilitate the development of transportation models in R

Brian Gregor
Oregon Dept. of Transportation
Brian.J.Gregor@odot.state.or.us
http://www.oregon.gov/ODOT/TD/TPAU