A Hybrid Origin-Destination Trip Matrices Estimation Model Using Machine Learning Techniques

Yohan Chang, *Ph.D. student* and **Praveen Edara**, *Ph.D., P.E., University of Missouri - Columbia*

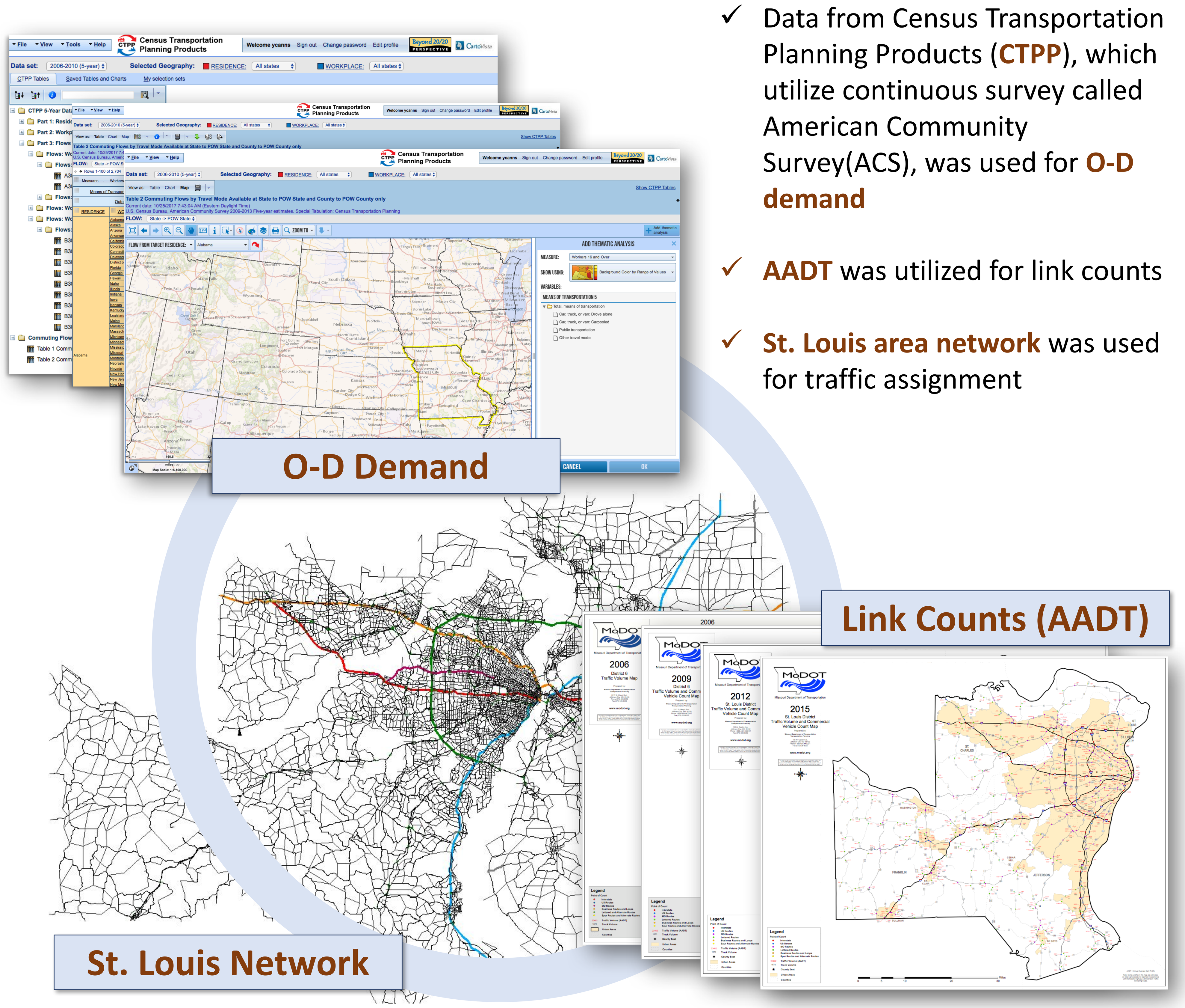
TRB TRANSPORTATION RESEARCH BOARD OF THE NATIONAL ACADEMIES

Applying Census Data for Transportation: 50 Years of Transportation Planning Data Progress

INTRODUCTION

- ✓ The **Origin-Destination trip matrix (O-D matrix)** is an essential ingredient in a variety of transportation planning and analysis studies
- ✓ The **traditional O-D Matrix estimation models** used license plate surveys, home interviews, roadside surveys, etc., but these methods have a disadvantage in the view of **cost-effectiveness**
- ✓ In this research, we want to propose a **hybrid O-D estimation** approach, **combining both a mathematical based model** (Cho 2008; O-D estimation model with partial information) **and a group of machine learning models**, such as Random forests (**RF**), neural networks (**NNs**), and Deep neural networks (**DNNs**).
- ✓ **St. Louis area** was chosen for testing and **Census data** was used for seed O-D matrix and annual average daily traffic (**AADT**) value was used for seed volume and calibration

DATA PREPARING and PROCESSING



- ✓ Data from Census Transportation Planning Products (**CTPP**), which utilize continuous survey called American Community Survey(ACS), was used for **O-D demand**
- ✓ **AADT** was utilized for link counts
- ✓ **St. Louis area network** was used for traffic assignment

OBJECTIVES

- ✓ Provide additional O-D matrix estimation method with using Census data and other public data sources
- ✓ Provide informational data resources to local jurisdictions for planning purposes
- ✓ **O-D matrix estimation using several machine learning methods**

STRATEGIES for CREATING TRAINING SET

O-D Demand

- ✓ All ACS published MOE are based on a 90 percent confident level (*"American Community Survey Multiyear Accuracy of the Data 2013"*, American Community Survey Office 2014)
  - Standard error = MOE / 1.645
  - Lower Confidence Bound = Counts - MOE
  - Upper Confidence Bound = Counts + MOE
- ✓ Create a set of perturbed O-D matrices based on actual flows and margin of error (MOE) in CTPP data
  - O-D matrix for standard error
  - O-D matrix for Lower Confidence Bound
  - O-D matrix for Upper Confidence Bound

Link Counts (AADT)

- ✓ Create a set of perturbed AADT traffic counts based on actual AADT and other research findings for Urban area (Gadda et al., 2007, "*Estimates of AADT: Quantifying the Uncertainty*")
  - Min error : 4.89% ~ 5.62%
  - Max error : 37.77% ~ 81.14%
  - Mean error : 11.47% ~ 14.28%
  - Standard deviation : 6.06% ~ 17.08 %

Data combinations using both seed information

- ✓ Current O-D and its traffic assignment (TA) results
- ✓ perturbed current O-D and its TA results
- ✓ Future AADT
- ✓ Estimated O-D with current AADT
- ✓ Estimated O-D with perturbed future AADT
- ✓ Network attributes (e.g., capacity, length)

Acknowledgement

We thank **Gloria Bradley**, *Traffic Management Operations Team in Missouri Department of Transportation* for assistance with providing multiple AADT map and **Jennifer Reiman**, *GIS Manager in East-West Gateway Council of Governments* for providing St. Louis area network

EXPERIMENTAL DESGIN

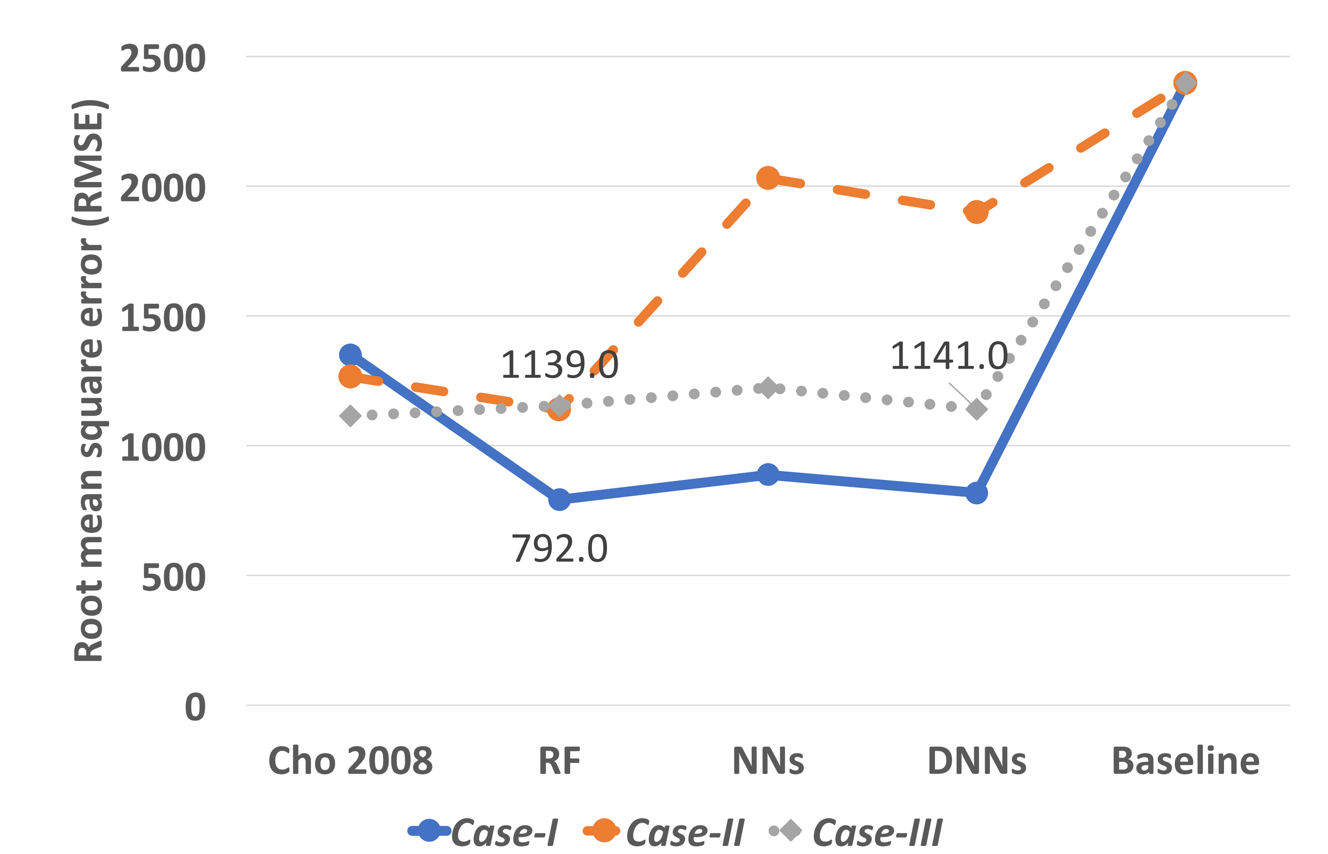
Scenario Design

Scenario 1 – Estimate accurate O-D matrix with **seed AADT** counts and **seed O-D matrix**
Scenario 2 – Estimate accurate O-D matrix with **20% error in seed AADT** counts and **seed O-D matrix**
Scenario 3 – Estimate accurate O-D matrix with **seed AADT** counts and **20% error in seed O-D matrix**

Optimization for Machine learning models

RF model - three critical parameters were fine-tuned and calibrated using Tabu search
DNNs model - A Genetic algorithm was used for optimizing parameters

EXPERIMENTAL RESULTS



CONCLUSION and FUTURE STUDIES

- ✓ This study firstly tried **a hybrid O-D matrix estimation models** using publically accessible data source, including **Census data**
- ✓ **RF model** outperformed other models in most of the cases
- ✓ Predicted machine learning model offers **another view of demand acquisition** in addition to the ACS data

- ✓ Increased accuracies considering other variables
- ✓ Specific zones (under county level) and network