EEMS0015

U.S. DEPARTMENT OF ENERGY

SMARTMOBILITY

Systems and Modeling for Accelerated Research in Transportation

Calibration of Activity-Based Transportation System Simulation Tools using High-Performance Computing

Dominik Karbowski, Josh Auld, Randy Weimer, Aymeric Rousseau – *Argonne National Lab.*
Vadim Sokolov, Laura Schultz – *George Mason University*

2017 DOE Hydrogen Program and Vehicle Technologies Annual Merit Review – June 7, 2017

Project Overview

Timeline	Barriers
• Project start date : Oct-2016 • Project start date : Sep-2019 • Percent complete : 15%	• Calibration of transportation models is costly and inaccurate • Transportation models are complex

Budget	Partners
• FY17-FY19 Funding: $495K • FY17 Funding Received : $175k	• Argonne National Laboratory (Lead) • George Mason University (Sub) • Berkeley Lab (data)

Project Relevance

Challenge

- Building a **transportation model** for a given city is **costly** and **lengthy** (years of development and millions of dollars).
- There is no generalized approach to calibration, leading to **inaccurate forecasting**
- Models are rarely updated, and quickly become **obsolete**
- No existing rigorous methodologies to incorporate **new types of transportation data** into models, e.g. cell phone data, GPS trajectories, etc.

Build a framework for transportation system model calibration:
⇒ Automate and speed-up the model building process
⇒ Provide better forecast reliability and certainty

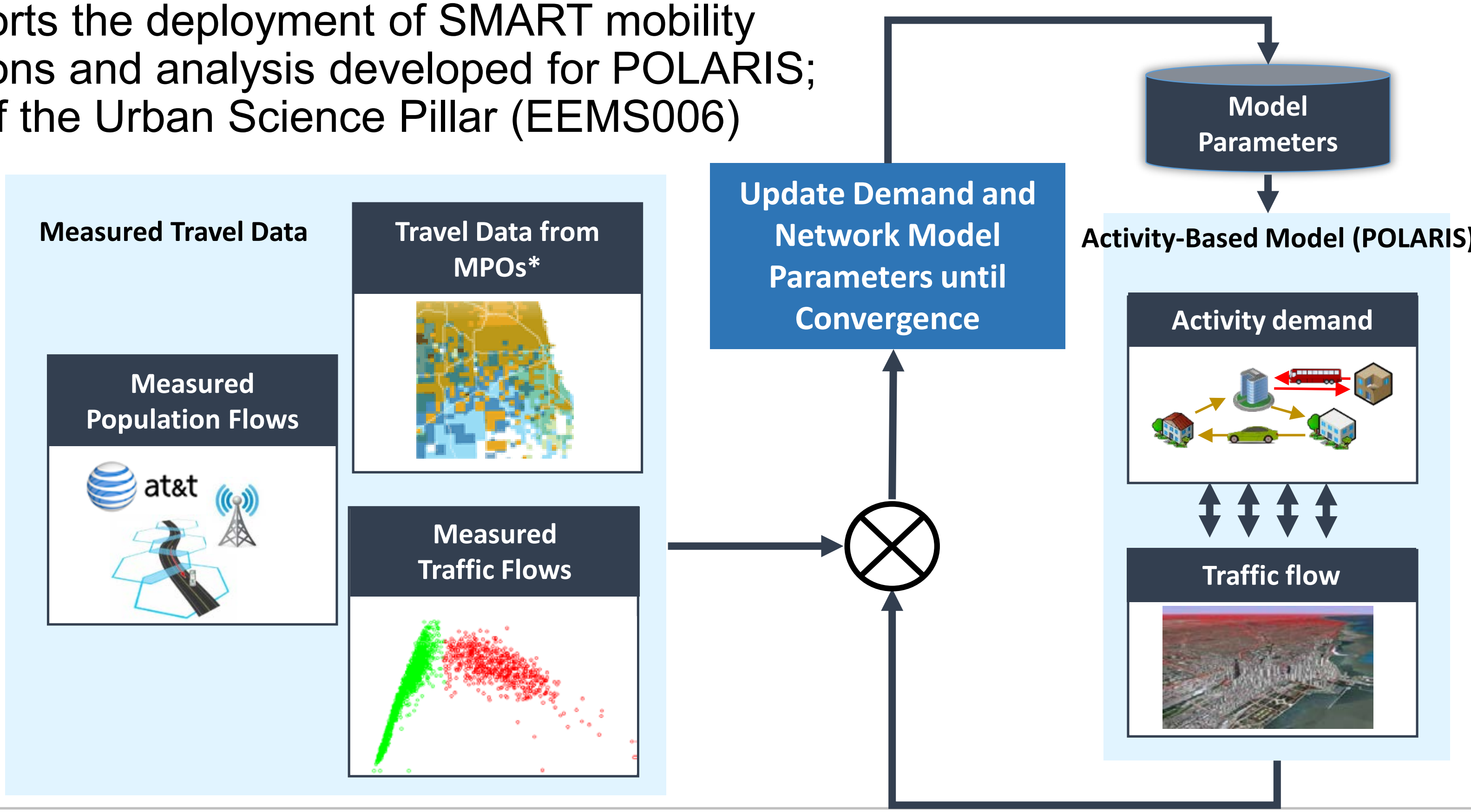
Key Tasks within SMART

- Framework will be demonstrated for POLARIS. POLARIS is used across SMART pillars, and this work will improve forecast accuracy of POLARIS case studies
- This project will greatly facilitate SMART models/analysis deployment beyond the cities currently modeled in POLARIS (e.g. Chicago, Detroit)

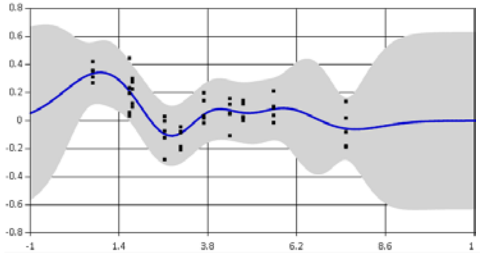
APPROACH

Combine Big Data, Simulation and High-Performance Computing

- Use new types of mobility datasets to calibrate transportation system models
- Develop and implement computational algorithms for automating the calibration process that can be executed on HPCs
- Supports the deployment of SMART mobility solutions and analysis developed for POLARIS; part of the Urban Science Pillar (EEMS006)



Simulation-Based Optimization

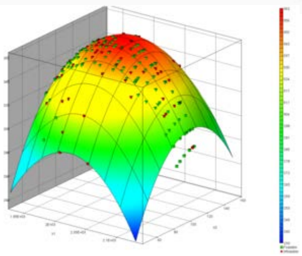


Calibration Techniques: Bayesian Optimization

Build on techniques used to calibrate models of other complex systems (e.g. space rockets, hospital networks…) to transportation systems.

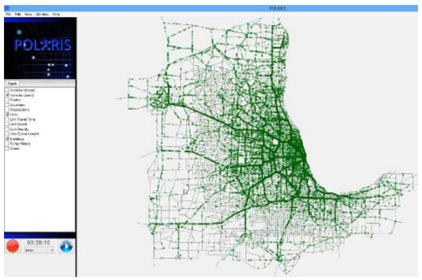
Leverage High-Performance Computing (HPC)

Build on state-of-the art expertise in high performance computing and machine learning.



Dimensionality Reduction

Address problems of computational complexity via dimensionality reduction techniques and adopting existing algorithms to run in parallel on HPCs



Transportation Modeling

Use computationally efficient, multithread integrated agent-based transportation simulation tool

Calibration Framework Will Be Applied to POLARIS (EEMS014)

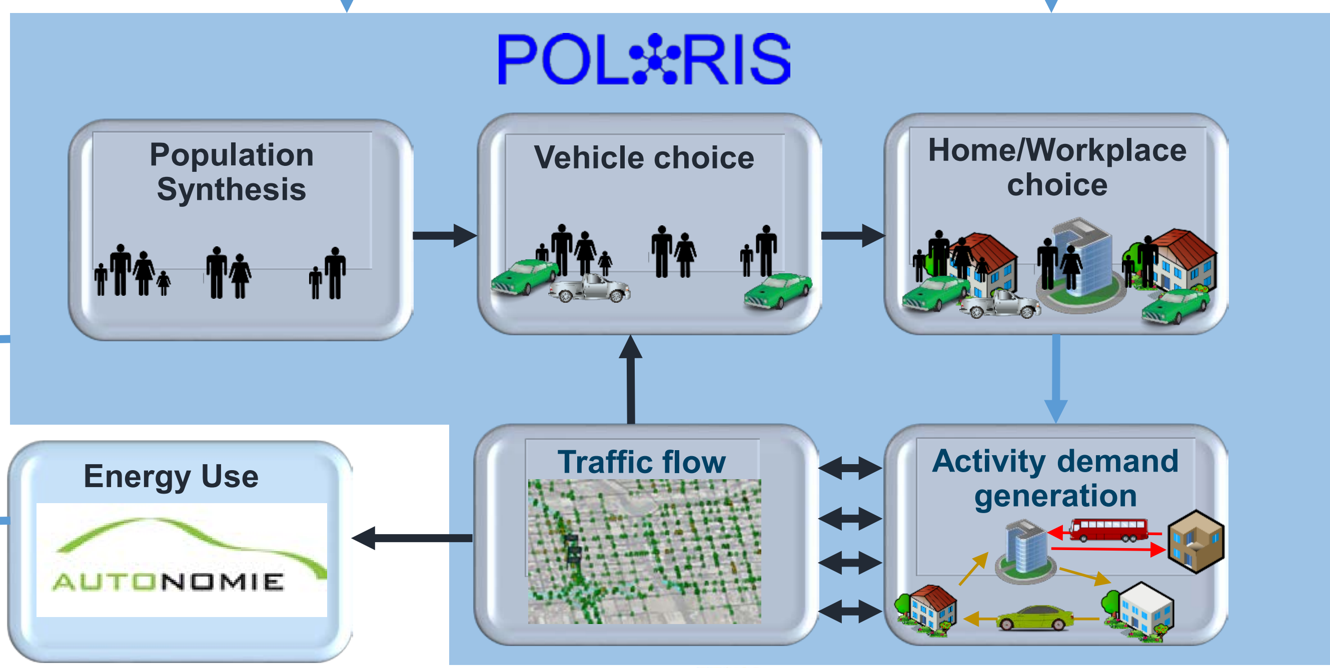
POLARIS is:
- *Activity-based*
- *Agent-based*
- *Integrated (demand + traffic)*

Input Data

Travel surveys
Employment data
Transp. network
Vehicle registrations
Land use, …

Forecast Scenarios

Infrastructure improvements
Population growth
Demographic shifts
New technologies
Policy environment

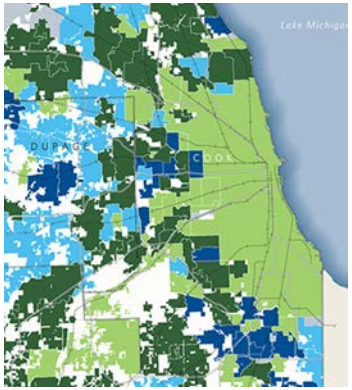


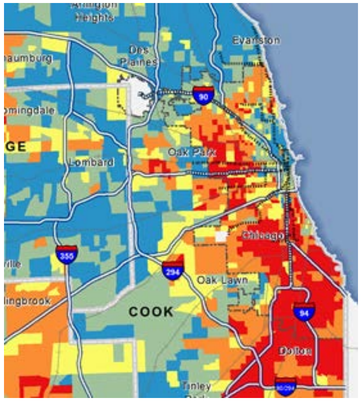
Results

Energy consumption
VMTs
Travel times, …

U.S. DEPARTMENT OF ENERGY SMARTMOBILITY

Transportation Model Requires a Large Number of Parameters to be Calibrated





- **Land Use**
  - e.g. destination attractiveness described by job opportunities, residential density and retail parameters.
  - ~20 parameters per traffic analysis zone (TAZ), 2000 TAZs in the model, i.e. approximately 40k total land use parameters.
  - Known with high level of certainty, but adjustments might be needed.
- **Population parameters**
  - e.g. socio-demographic and vehicle ownership model parameters.
  - Coming from large sample (Census) and usually known accurately enough a priori.
- **Behavioral model parameters**
  - Govern choices travelers make regarding transportation mode, activity types, route preferences.
  - Examples: value of time, required activity frequency.
  - ~ roughly 70 parameters for the entire population.

Use Large Travel Datasets as Inputs

- **Traffic flow data**
  - Data from loop detectors: **speed**, **flow** and **occupancy** averages for every 5-minute interval (continuous monitoring) on highways
  - Average **traffic counts** on arterial roads (longitudinal data), available for most major roads
  - Sources: State and city DOTs; county and city traffic management centers
- **Human activity data**
  - **Origins and destination** estimated from:
    - Tollway transactions
    - Travel surveys led by Metropolitan Planning Organizations (MPO)
    - Phone localization services aggregation
  - **Activity patterns** estimated from cell phone data
  - Sources: MPOs, tollway authorities, data vendors (e.g. StreetLight Data) telecommunication companies (e.g. AT&T)
- **Transit Data**
  - **Origins and destinations** estimated from:
    - on-board surveys
    - automated passenger count sensors
    - data from fare card transactions
  - Sources: Transit authorities, payment system operators, e.g. Cubic

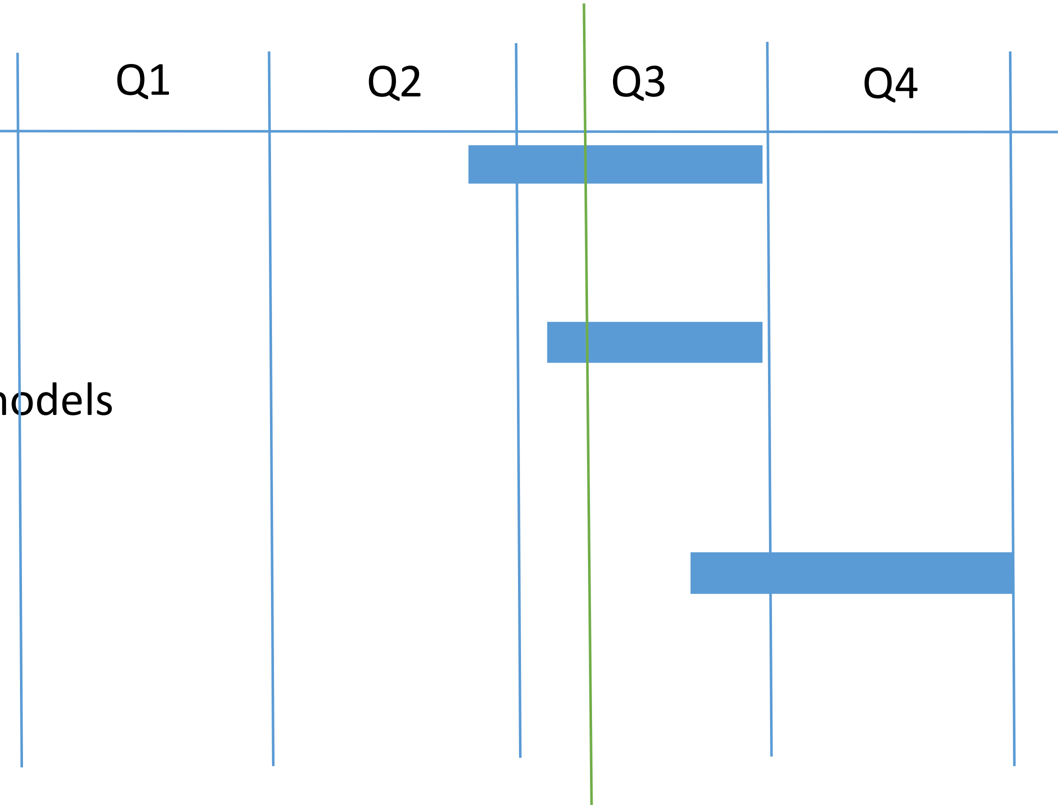
Milestones

Activities

Identify components of the mathematical framework

Develop prototype calibration framework based on lightweight models

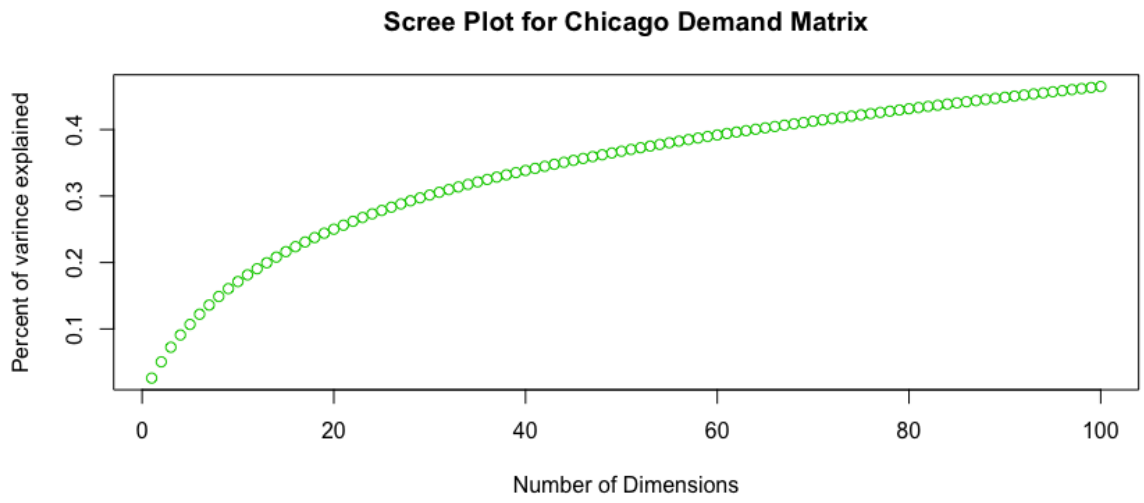
Develop calibration software to run on HPC



TECHNICAL ACCOMPLISHMENTS

Preliminary Analysis: Researched Dimensionality Reduction Techniques for Input Data

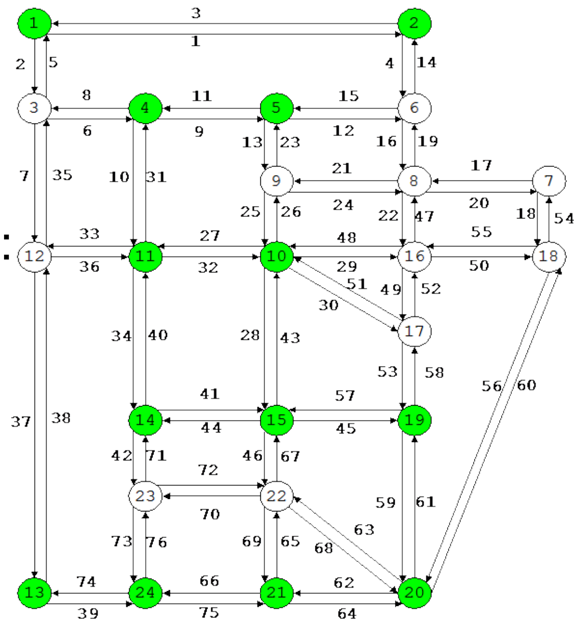
- High dimensional problems (many input parameters) are hard to optimize, due to curse of dimensionality
- Dimensionality reduction using principal component analysis (PCA) applied to origin-destination flow matrix (one of the inputs):
  - Original dimension is 2000x2000 = 4 million parameters
  - If reduced to 100x2000 = 200k parameters, only 40% of variance is explained



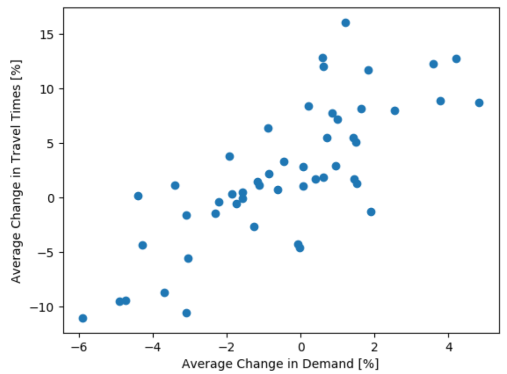
- Further research on dimensionality reduction is needed

Preliminary Analysis: Performed Sensitivity Analysis Using a Simplified Simulator

- Developed a framework for quick prototyping of the calibration techniques, using a lightweight, small-scale model widely used for debugging and prototyping transportation algorithms
- Performed sensitivity analysis of model outputs to inputs:
  - allows to exclude unnecessary input parameters from analysis
  - allows to develop an intuition for appropriate surrogate models to approximate the simulator
  - applied perturbation to input (travel demand)
  - **significant sensitivity observed**



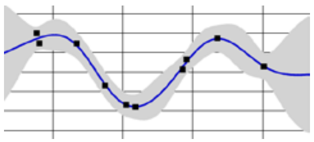
Lightweight model:
24 nodes, 76 links, >360k trips



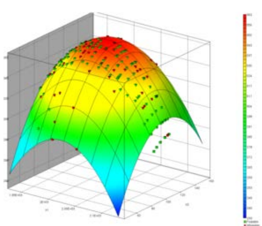
Performed Literature Review on Calibration of Complex Models

Statistics	Calibration	Design of Experiments	Optimization Techniques
Acquisition Functions Gaussian Processes Kernel Variations Bayesian Inference Information Theory Conjugate Functions	Computer Modeling Bayesian Deep Learning	Experimental Design Adaptive Exploration Particle Learning High-Dimensional Use Cases Supercomputer Experiment Analysis	Bayesian Dynamic Programming Particle Learning

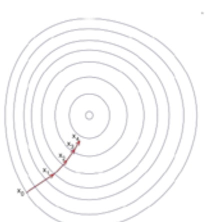
3 distinct approaches identified:



- **Bayesian optimization** (most promising)
  - Successfully used to calibrate systems with high input parameter uncertainty and large approximation errors.
  - Represents inputs and simulator as random variables to find correct distributions over parameter inputs.
  - As a side product, outputs uncertainty quantification.



- **Approximation**
  - Methods based on curve fitting
  - Good and efficient for slowly changing functions
  - Unlikely to be appropriate for transportation



- **Optimization**
  - Based on deterministic optimization machinery.
  - Restrictive about types of uncertainties we can assume about parameters, thus less rigorous and less efficient estimation, i.e. estimates of input parameters are biased.

Defined Bayesian Optimization Formulation for Calibrating Model Using Traffic Flow Data

- Objective: minimize discrepancy between simulated and observed data:

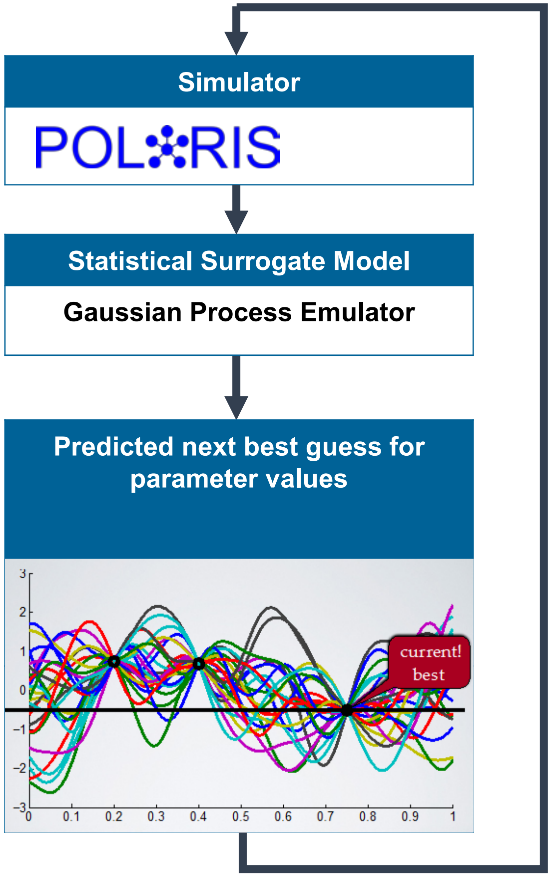
$$\min_{\theta}(\|\varphi(\theta) - D\|)$$

  $\varphi(\theta)$: the simulator, e.g. POLARIS

  θ: parameters to be calibrated

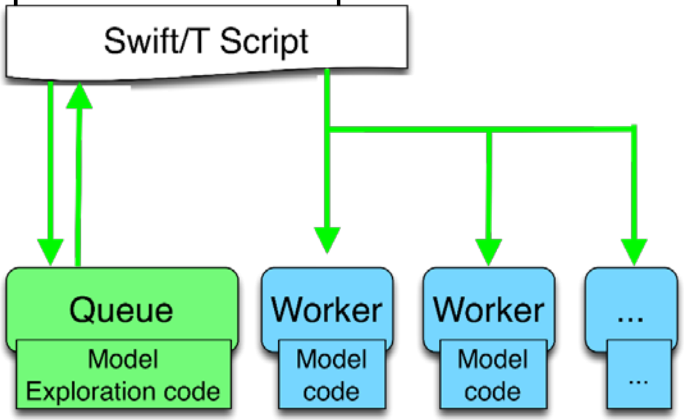
  D: Field data

- Current approaches are ad-hoc and not transferable
- We use Bayesian optimization to solve the problem
  - Relies on Gaussian Processes (GP) to approximate the uncertainties about simulator's outputs
  - $\varphi(\theta)(t)$ is represented as a GP, i.e. every point in time has a normal distribution
- GPs are common models which can be leveraged as a Bayesian prior for simulator input parameters
- GP allows to incorporate prior information about parameter uncertainties, e.g. expert opinion or uncertainties from statistical fitting procedures
- GP efficiently uses computing resources, by selecting next sample that leads to largest uncertainty reduction



Prototyped a Computational Framework for Calibration Using HPC

- Model calibration on HPC requires:
  - **Exploration code**: performs parameter sampling, Bayesian optimization, data analysis, and eventually calibration (Python or R code).
  - **Model**: POLARIS instance with parameters defined by exploration code
  - **HPC workflow management** (EMEWS): manages the flow of information between exploration code and model
- Leverage **Extreme-scale Model Exploration with Swift** (EMEWS) framework developed at Argonne:
  - Exploration code sends “jobs” to the queue for execution
  - EMEWS assigns and launches job using model on available computer worker
  - EMEWS transfers output from worker to Exploration code
  - Exploration code computes new set of parameters



Response to Previous Year Reviewers' Comments

Project was not reviewed in the past

Partnerships and Collaborations

Argonne National Laboratory	Prime recipient. POLARIS development, including transportation modeling and adaption of core code to HPC machines
George Mason University	Subcontractor. Develops the calibration framework for POLARIS
Berkeley Lab	Collaboration on large-scale traveler activity data
CDOT Chicago Department of Transportation	Exchange of transportation data

Remaining Challenges and Barriers

- Build a **reliable software** infrastructure to run thousands of simulations simultaneously on large scale HPCs from prototype code
- Developing **asynchronous sampling** schemes so that calibration algorithms can be run on HPCs.
  - Multiple best guesses need be generated in each iteration, one per computing node.
  - Simulation run times are different from one set of parameters to another, we need to start new simulations without waiting for previous batch to be completed.
- Address issue of **dimensionality reduction:**
  - The system is highly nonlinear and "off-the-shelf" linear techniques did not lead to satisfactory results.
  - Non-linear dimensionality reduction techniques need to be used
- Access to large-scale transportation **datasets** is difficult due to cost, privacy and contracting issues

Proposed Future Research

- Move **from prototype to the usable implementation** of the computational framework for running large number of simulations on HPC (by the end of FY17)
- Run **sensitivity and perturbation** simulations, similar to those performed for quick prototyping framework so far (by the end of FY17)
- FY18: finish implementation of **mathematical models** for calibration
- FY19: **application** of calibration techniques using cell phone data and origin-destination flows estimated from toll transactions and license plate matching.

"Any proposed future work is subject to change based on funding levels."

Summary

- An **automated calibration** process of transportation models using large dataset is essential:
  - Development of models for new cities is **expensive**
  - Transportation models (e.g. POLARIS) can be used to assess policies and technologies for a more a energy efficient transportation environment
- Development of an automated process for transportation model calibration is a **challenging task** from methodological and computational points of view: large models (POLARIS), large number of parameters to calibrate.
- This research relies on statistical, computing and transportation modeling methodologies
- Main achievements (end of FY17 Q2):
  - Preliminary analysis highlights **high dimensionality** of the problem and high **sensitivity to perturbation** of inputs
  - Selected and formulated **Bayesian optimization** as main calibration algorithm
  - Developed a **prototype framework** to run calibration on **HPC**
- Preliminary results confirm our initial guess, that “off-the shelf” tools do not lead to satisfactory solutions and further research and development is required.