Motivation

Human dynamics are major contributing factor to urban environments. Throughout each day, humans consume energy, whether they are traveling, at home, or in their workplace. Integrated systems improve our understanding of energy, emissions, and other human impacts on their environment to develop sustainable community strategies.

Data

We provide a variety of data for a 2017 scenario in downtown Chicago:

- Vehicle data
    - Simulation snapshot for morning commute from TRansportation ANalysis SIMulation System (TRANSIMS). This snapshot contains vehicle traces (in Universal Transverse Mercator Coordinates) at 30 second intervals for one simulated day. At each time step, we also have the link (road segment) ID, driver ID, and vehicle speed.
    - Schedule for morning commute from National Household Travel Survey (NHTS). This is an extract of the official NHTS data (https://nhts.ornl.gov/) which only contains survey responses from Chicago.
    - Vehicle type distribution. Simplified Federal Highway Association (FHWA) classifications of vehicles in Chicago, which was derived from NHTS data.
- Emissions data
    - Road-level traffic volumes (aggregated from TRANSIMS outputs).
    - Road-level emissions generated using MOVES, an emissions simulator. This simulation is based on traffic volumes and weather patterns throughout a year.
    - Weather data from DarkSky
- Road network
    - The road network has the link IDs for each road segment, as well as road type etc.
    - GeoJSON file of the road network used for the TRANSIMS and MOVES runs.
    - Definition of different link types.
- Building data
    - Building footprints from Microsoft
        - Microsoft has published building footprints for the entire US https://github.com/Microsoft/USBuildingFootprints
        - Participants can download the full Illinois data themselves or use the version that's clipped to Chicago which we provide for this challenge. https://usbuildingdata.blob.core.windows.net/usbuildings-v1-1/Illinois.zip
    - Land Use data from Chicago Metropolitan Agency for Planning (CMAP)
        - GeoJSON file containing polygon data with land use attributes
        - Codebook defining the land use codes
- Socioeconomic data:

- Population from CMAP/Census (2010)
  - Census data summarized to community areas https://datahub.cmap.illinois.gov/dataset/2010-census-data-summarized-to-chicago-community-areas
  - Spreadsheet (CSV) of census data by community area
  - GeoJSON of community area polygons
- Community Area Snapshots (2017)
  - Additional information such as employment, travel mode choice, housing types, job types in community (held by residents, available in community), walkabilty, etc. https://datahub.cmap.illinois.gov/dataset/community-data-snapshots-raw-data
  - Spreadsheet (CSV) of 2017 snapshot data.
  - Data dictionary explaining the different fields.

Challenges

One of the main challenges in coupled or integrated systems is the disparity of data sources. For this data challenge, we would like participants to address the following tasks:

1. Develop an algorithm to efficiently assign vehicle occupants to nearby buildings.
   - We have performed an initial weighted quadtree-based approach to map vehicles to buildings.
     - A. Berres, P. Im, K. Kurte, M. Allen-Dumas, G. Thakur, J. Sanyal: A Mobility-Driven Approach to Modeling Building Energy. 5th IEEE Workshop on Big Data Analytics in Supply Chains and Transportation. Los Angeles (2019).
   - The ideal algorithm should be efficient and accurate. Consider the trade-off.
   - The resulting mapping should be realistic. Consider building size, use type (the vehicle traces are only for commute) etc.
2. Perform an area-wide correlation analysis of vehicle emissions.
   - Determine spatial variation, and variation based on other factors, such as land use of surrounding areas, population, network classification (road type), weather, etc.
   - Correlate the provided emissions data with other provided datasets.
3. Characterize traffic patterns from the simulation:
   - What are the traffic hot spots? Is there any congestion?
   - What are the travel times? (How) do they vary throughout the day?
   - What are busy times? How well do they match the commute pattern from NHTS?
   - How do speeds vary spatially and temporally?
   - What are the most popular roads?
   - Can you draw conclusions about the simulation setup from the output?