Mobility Transition Model (MoTMo) Data Set Analysis

Meghan Kane, Monica Soto
Data Sets from Past and Present Social Systems
Thematic Einstein Semester (TES) 2022

August 31, 2022

Introduction

The Mobility Transition data set [1] is the output resulting from an agent-based model of the same name (MoTMo from now on) that describes how certain interacting agents affect their mobility choices and emissions produced based on a set of over 500 scenarios. This model aims to describe private mobility demand in Germany.

For each run of the MoTMo, various intervention options can be switched `on` and `off` (there are 10 available such options), relative to the business as usual case (BAU), where no interventions are made. The intervention options span 3 categories: **investment**, **policies**, and **events**. Each of these runs yields a different scenario.

In the MoTMo , there are 5 mobility options that agents can choose: Combustion Cars, Electric vehicles, Public Transport, Car sharing and Non-motorized "active" mobility (e.g. biking). These mobility options comprise the output of the MoTMo alongside emissions caused by each mobility type. These are recorded at each time step (once every other month) over the course of 30 years (2005-2035) yielding 180 total time steps.

The core question that our analysis seeks to give insights into, which was stated in [1], is:

"Given the structure of combinations of options underlying the 500+ scenarios: what can be said about the effects of single options and their combinations?"

To analyze this data set, we will use a method called *variance-sensitivity analysis*, which tells us how sensitive the model is due the inputs individually or combined.

1 Dataset

For each given run of the MoTMo, various options across investment, policy, and event categories can be turned `on/off` with the restriction that at most 2 options can be chosen from a given category. There are 10 options total spanning these 3 categories, which are shown in Table 1.

Category	Option	Label
Investment	Charging infrastructure	CH
	Public transport subsidy	SP
	Electric vehicle subsidy	SE
Policy	Car weight regulation	WE
	Bike friendliness	BP
	Urban combustion restriction	RE
Event	Higher gas price	CO
	Intermodal digitalisation	DI
	EV world market tour	WO
	Increased car sharing availability	CS

Table 1: Available options with their respective category

2 Sensitivity Analysis

One approach to explore the input/output solution space is to use "human-guided exploration" based off of intuition. This has been cited as an approach that can be successful, especially if it involves guidance by domain experts [2] and can be helpful with preliminary analysis, but highly complex output data necessitates analyses that are more systemic and unbiased.

Sensitivity analysis allows us to explore the input/output space in a way that gives us insight into the behavior of the ABM output under various input parameters alone and in interaction. The 2016 research paper by Guus ten Broeke, "Which Sensitivity Analysis Method Should I Use for My Agent-Based Model?" [3], claims that with sensitivity analysis for ABMs we are able to gain insight in how patterns and emergent properties are generated in the ABM, examine the robustness of emergent properties, and quantify the variability in ABM outcomes resulting from model parameters.

2.1 Variance Sensitivity Analysis

The main idea of this method is to decompose the variance output of a given model that can be seen as a function with a certain input X and an output Y, that is, $Y = f(X)$, where $X \in \mathbb{R}^d$ and d is the number of inputs. It aims to find out those input parameters that have the most effect on the variance output and give us a theoretical intuition of how sensitive the model is to certain inputs or factors.

If the inputs are independent and mutually uncorrelated, the output function has an unique decomposition and thus so does the variance output, which leads to express it as (for further details, see [4], section 1.2.6)

$$Var(Y) = V_i + \sum_{i<j}^{d} V_{ij} + \ldots + V_{12\ldots d}$$

where $V_i = Var_{x_i}(E_{X_{\sim i}}(Y|X_i))$, $V_{ij} = Var_{X_{ij}}(E_{X_{\sim ij}}[Y|X_i, X_j]) - V_i - V_j$ and so on. $X_{\sim i}$: is the set of all variables except X_i. More specifically, $E_{X_{\sim i}}(Y|X_i)$ computes the average of all-except X_i, while keeping X_i fixed, and then we proceed to compute the variance of such means, thus getting the value of $Var_{x_i}(E_{X_{\sim i}}(Y|X_i))$.

2.1.1 Sensitivity Indices

These indices measure the contribution of the interaction of the variables to the total variance $V(Y)$, and thus, one can define the *first-order* and *second-order* Sobol indices (resp.) as follows:

$$S_i := \frac{V_i}{Var(Y)} \qquad \text{and} \qquad S_{ij} := \frac{V_{ij}}{Var(Y)} \tag{1}$$

The 1st order indices, S_i, measure the influence of the main effects, so a big S_i suggests that the variable X_i is an *important* factor; by construction, it can take any value between 0 and 1. The 2nd order indices S_{ij} measure the contribution of the interactions of pairs of variables X_i, X_j (for $i \neq j$). Furthermore, the *total-effect index* quantifies the total effect of the factorX_i by measuring all variance caused by its interactions, and it is given by:

$$S_{T_i} = 1 - \frac{Var_{\sim X_i}(\mathbb{E}_{X_i}(Y|X_{\sim i}))}{Var(Y)} = \frac{\mathbb{E}_{\sim X_i}(Var_{X_i}(Y|X_{\sim i}))}{Var(Y)} \tag{2}$$

This equation computes the variance of the mean of all the terms of any order that do not include factor X_i.

3 Implementation

The first challenge we faced when implementing the sensitivity analysis was to decide the input/output space. Specifying the input was not immediately obvious since the core assumption for this analysis to be valid is that the input variables are independent and uniformly distributed. However, under the condition that each category cannot have more than two 1s, this implies that taking the 10 options as inputs are not mutually independent. Instead, we made a variable transformation (reduction) by taking the categories as inputs, which are indeed mutually independent and uniformly distributed. The natural outputs were taken as the total emissions produced and the different mobility choices.

Input space: investment: X_I, policies: X_P and events: X_E. Outputs: total emissions and mobility choices between cars: public transport (**stock_P**), electric cars (**stock_E**), combustion cars (**stock_C**), non-motorized (**stock_N**) and shared vehicles (**stock_S**)

3.1 Results

We computed the first, second and total-effect indices with input variables being the categories (X_I, X_P and X_E) and for each of the mobility choices and total emissions. Our code can be found in our Github repository [5]. The tables below show the results.

	stock_C	stock_E	stock_N	stock_P	stock_S	total_emissions
S_I	0.144434	0.487024	0.514822	0.534556	0.058319	0.005789
S_P	0.426486	0.150791	0.380571	0.160698	0.229341	0.939900
S_E	0.411576	0.280636	0.089221	0.288213	0.538039	0.042712

Figure 1: First-order indices S_i

	stock_C	stock_E	stock_N	stock_P	stock_S	total_emissions
(I, P)	0.000557	0.035216	0.000989	0.003023	0.023244	0.000508
(I, E)	0.006729	0.037781	0.002485	0.004842	0.012958	0.001033
(P, E)	0.007902	0.004284	0.009361	0.006122	0.084962	0.007984

Figure 2: Second-order indices S_{ij}

	stock_C	stock_E	stock_N	stock_P	stock_S	total_emissions
ST_I	0.154036	0.564289	0.520846	0.544966	0.147658	0.009404
ST_P	0.437261	0.194559	0.393472	0.172389	0.390685	0.950466
ST_E	0.428523	0.326969	0.103618	0.301723	0.689096	0.053803

Figure 3: Total-effect indices S_{T_i}

Recall that the higher the index, the more "important" the variable is. The results shown in the first table above (1st order indices) indicate that the choice of the agents towards **combustion cars** (stock_C) is mainly affected by the implementation of policies, X_P and events, X_E. However, the analysis alone does not tell us whether the change increases or decreases this choice. Some experimental results showed that indeed when at most one policy option was ON, the mobility choices of combustion cars increased substantially, so this might suggest that creating policies like car weight regulation, bike friendliness policies, or urban combustion restrictions makes people use these vehicles less.

Furthermore, the second-order index of stock_C shows that the policies combined with events leads the *best* results in terms of people choosing to drive combustion cars less. The events might be related to higher gas prices, improvement of the convenience for using electric vehicles, public transport, car sharing, bikes, etc., a higher electric car market, or more availability of car sharing.

The indices of the mobility choices of **electric cars**, stock_E, suggest that changes in investments make the output variance higher for this type of mobility. Thus, with actions such as adding more charging stations, and subsidies for public transport and/or electric vehicles make the choice of this cars more appealing. The pair of categories that optimizes this mobility choice are the investment and the events category.

Analogously, the indices of the **non-motorized** choices, stock_N, imly that X_I is a factor of importance (investment), followed by X_P (policies); but the interestingly, 2nd order index indicates that the combination of changes in investment and events (not policies) is the most influential pair of factors. For the **public transport** choice, stock_P, the category that is the most influential alone is investment, X_I, but the pair of policies and events combined are also influential for this output. For the **shared car** option, stock_S, the most influential input variable seems to be X_E (event), and when combined with policies, it seems to describe the output variance the most.

For the **total emissions**, it is the policy category the most descriptive input variable, some runs of certain models showed that changing (or turning off) options in investment and events does not affect as much as changes in policies, so the targeting policy changes may yield the highest benefits for influencing the total emissions production.

4 Further Research

While we have preliminary results that yield some insights into the MoTMo dataset, there are many ideas that we would like to pursue to improve on our analysis. First, we would like to explore ways to incorporate the full 10 options. Saltelli[4] suggests that the implementation of this analysis for correlated inputs is possible by incorporating noise input variables, so we intend to explore this in detail.

This sensitivity analysis alone did not yield and insight on how do the "most important" input variables affect for *good* or for *bad* the output, other than doing it by brute force and exploraton. We would like to investigate this in a more practical and theoretical way. Furthermore, Lo Piano et al. [6], suggest that these indices are also very expensive and proposes to use other indices and redesign the entire method.

References

[1] Sarah Wolf, Steffen Fürst, Andreas Geiges, Manfred Laublichler, Jahel Mielke, Gesine A. Steudle, Konstantin Winter, and Carlo C Jaeger. The decision theatre triangle for societal challenges insights from decision theatres on sustainable mobility and resulting research needs. *Global Climate Forum*, Feb 2021.

[2] Ju-Sung Lee, Tatiana Filatova, Arika Ligmann-Zielinska, Behrooz Hassani-Mahmooei, Forrest Stonedahl, Iris Lorscheid, Alexey Voinov, Gary Polhill, Zhanli Sun, and Dawn C. Parker. The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation*, 18(4), 2015.

[3] Guus ten Broeke, George van Voorn, and Arend Ligtenberg. Which sensitivity analysis method should i use for my agent-based model? *Journal of Artificial Societies and Social Simulation*, 19(1), 2016.

[4] Andrea Saltelli, M. Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global Sensitivity Analysis. The Primer*, volume 304. 01 2008.

[5] Monica Soto and Meghan Kane. Motmo data set analysis, github repository. `https://github.com/MoniSoto/MoTMo`.

[6] Samuele Lo Piano, Federico Ferretti, Arnald Puy, Daniel Albrecht, and Andrea Saltelli. Variance-based sensitivity analysis: The quest for better estimators and designs between explorativity and economy. *Reliability Engineering & System Safety*, 206:107300, 2021.