Modeling Population Human Mobility with Dynamic Mode Decomposition

Liantao Li and Yang Yang

Abstract

Human mobility research concerns spatiotemporal individual and population movement. Accurate modeling and prediction of human mobility can provide opportunities to monitor, manage and optimize human movement for improved social-economic benefit. In this paper, we adopt the dynamic mode decomposition algorithm to model population human mobility using visitor flow data between different states in the United States from 2019 to 2021. We train multiple DMD models with different low rank structures, and evaluate their modeling accuracy and predictability on novel testing data.

1 Introduction

Human mobility research studies individual and collective human movement in space and time. Examples of human mobility include livelihood mobility, urban migrations, and traffic flows. Understanding human mobility is essential to multiple social-economic aspects such as mitigating urban traffic congestion and improving public health. For instance, human mobility plays a significant role in understanding the dispersal of infectious diseases such as COVID-19. Comprehension of human mobility could help predict the variation of diseases to enable optimal decision making and public health interventions at early stages.

An important aspect of human mobility study is modeling and prediction of human mobility patterns. Accurate modeling and prediction can provide opportunities to monitor, manage and optimize human movement. For example, effective modeling of urban traffic flow could suggest strategies to mitigate traffic jam, promoting road safety and reducing environmental impacts. Human mobility at different scales exhibit different natures. Mobility on the individual scale typically depends on a large number of uncertainty factors such as individual mood or emotions; it thus often shows a stochastic nature. On the other hand, human mobility on the scale of groups demonstrates a deterministic nature, and the nature can be understood as an average/expectation of a large number of individual mobility.

Based on the scale of mobility, current models can be classified into two categories: individual mobility model and general population model. Individual models aims at reproducing individual mobility patterns. Typical examples include Brownian motion model, Levy flight model, continuous time random walk model, and preferential return model. In contrast, general population models seeks to describe the aggregate mobility of many individuals. Representatives include the gravity model proposed by Zipf, which was inspired by Newton’s Law of Gravitation; intervening opportunities models introduced by Stouffer, and the radiation model proposed by Simini et al., which is parameter-free and can be used without mobility measurement. We refer readers to the reference for a comprehensive review of different models as well as their weakness and strength.

As the average of numerous individual behaviors, mobility of populations usually features significantly higher stability and predictability. We therefore focus on modeling population mobility in this paper. All the aforementioned population mobility models are based on the intrinsic physics, thus depending on accurate prior knowledge to achieve accurate modeling. On the other hand, the development of information and communication technologies has made large-scale mobility-related
datasets more accessible than ever before. With the advance of machine learning methodology, powerful data-driven machinery becomes available to extract statistically meaningful patterns from these large datasets. The goal of this paper is to evaluate a data-driven, equation-free modeling approach known as the dynamic mode decomposition.

The dynamic mode decomposition (DMD) is a data-driven method that can provide effective modeling of dynamical systems [7]. The method was initially proposed by Schmid and Sesterhenn [8] and Schmid [9] in fluid dynamics, yet later found wide applications in a variety of areas including video processing [10], epidemiology [11], neuro-science [12], and financial trading [13]. The idea of DMD is to learn a temporal low rank structure of the dynamic systems. It is therefore also a dimension reduction technique, which makes it ideal for processing large-scale data. In this paper, we will apply DMD to model the population human mobility between different states in the US from 2019 to 2021 using the publicly available dataset [1]. We will build multiple DMD models with different low rank structures, and evaluate their accuracy in predicting future human mobility patterns.

The paper is organized as follows. We describe the structure of the human mobility data set in use in Section 2. Section 3 presents the idea of DMD in detail and formulates the precise algorithm. Section 4 includes multiple numerical experiments to validate the efficacy of DMD in modeling population human mobility with different training and testing datasets.

2 Data Structure

The dataset we will utilize to model human mobility is the multiscale dynamic human mobility flow dataset collected in [1] and made publicly available on Github [14]. It consists of daily and weekly human mobility flow data collected across the United States from 2019 to 2021. The dataset is collected with the help of the mobile phone application “SafeGraph” and pre-processed in several steps. First, selected mobile phone users from an origin to various destinations are tracked to generate visit information. Next, users in selected regions are grouped to form multiple clusters, and daily/weekly visit information between two clusters are calculated by summing up the number of individual users. Next, data are aggregated into three spatial scales: census tract, county and state by grouping all records based on the origin-destination (O-D) pairs. Finally, the dynamic population flow from the geographic origin “o” to the geographic destination “d” is inferred from the visitor flow using the following relation:

\[ \text{pop}_{\text{flows}}(o, d) = \text{visitor}_{\text{flows}}(o, d) \times \frac{\text{pop}(o)}{\text{num}_{\text{devices}}(o)} \]  \hspace{1cm} (1)

where \( \text{pop}_{\text{flows}}(o, d) \) is the estimated dynamic population flows from the origin \( o \) to the destination \( d \), \( \text{visitor}_{\text{flows}}(o, d) \) is the computed mobile phone-based visitor flow from \( o \) to \( d \), \( \text{pop}(o) \) is the population at the origin \( o \) extracted from the American Community Survey [15], and \( \text{num}_{\text{devices}}(o) \) is the number of unique mobile devices residing in \( o \). Details description of these steps are described in [1].

In this paper, we will make use of weekly visit flow data between different states. The data includes 52 distinct places as origins and destinations: 50 states, Washington D.C., and Puerto Rico. The weekly visitor flow data is recorded in a data matrix [14] with nine columns of information consisting of unique identifiers for origins and destinations, geographic locations of origins and destinations, data range of the records, estimated visitor flow, and inferred population flow based on the relation [1]. The information that is directly used in this paper is identifiers for origins and destinations, data range, and estimated visitor flow. We attempt to utilize the information to model and predict collective human mobility patterns.

3 DMD Algorithm

The dynamic mode decomposition (DMD) is an equation-free and data-driven method in modeling dynamic systems using high dimensional data [2, 16, 17, 12, 18]. It is designed to learn coherent
patterns of dynamic systems. Specifically, for a dynamical system
\[ \frac{dx}{dt} = f(x, t; \mu) \]  
(2)
where \( x(t) \in \mathbb{R}^n \) is the state of the system at time \( t \), \( \mu \) represents parameters of the system, and \( f(x, t; \mu) \) represents the dynamics, its corresponding discrete-time form can be obtained by evolving (2) by a fixed time step \( \Delta t \). Denote \( x_k := x(k \Delta t) \) where \( x \) is the solution of (2), then the Fundamental Theorem of Calculus implies
\[ x_{k+1} = x_k + \int_{k \Delta t}^{(k+1) \Delta t} f(x, \tau; \mu) d\tau \]  
(3)
It is natural to introduce the following discrete-time flow map \( F \):
\[ x_{k+1} := F(x_k) := x_k + \int_{k \Delta t}^{(k+1) \Delta t} f(x, \tau; \mu) d\tau. \]  
(4)
Given snapshots of the solution \( x_1, x_2, \ldots, x_m \), one can use these snapshots to construct two \( n \times (m - 1) \) data matrices:
\[
X = \begin{pmatrix} x_1 & x_2 & \ldots & x_{m-1} \\ \end{pmatrix}, \quad X' = \begin{pmatrix} x_2 & x_3 & \ldots & x_m \\ \end{pmatrix}.
\]  
(5)
The idea of DMD is to compute the leading eigendecomposition of the best-fit linear operator \( A \) relating the data \( X' \approx AX \):
\[ A = X'X'^\dagger \]  
(6)
where \( X'^\dagger \) is the pseudoinverse of \( X \). The eigenvectors of \( A \) are said to be the DMD modes. More specifically, the DMD algorithm \[7\] is summarized in Algorithm 1.

4 Numerical Experiments
We will apply the DMD algorithm to the weekly visit flow data \[13\] to model human mobility patterns. The 52 distinct places (50 states plus Washington D.C. and Puerto Rico) are assigned unique identifiers from 1 to 52. The weeks are arranged in the chronological order, starting from the first week of 2019. For the \( t \)-th week, we form a \( 52 \times 52 \) matrix \( A^t \) whose \((i, j)\)-entry is the visitor flow data from the origin \( i \) to the destination \( j \). Next, we symmetrize each matrix \( A^t \) to obtain a matrix \( S^t \) whose \((i, j)\)-entry is the total visitor flow from \( i \) to \( j \) or vice versa. If we view the matrix \( A^t \) as the adjacent matrix of a directed graph defined by the visitor flow, then \( S^t \) is the adjacent matrix of the underlying undirected graph. Then we re-shape each matrix \( S^t \) to a column vector and use it as a temporal snapshot of the underlying dynamic system that dictates human mobility to apply the DMD algorithm. By the end of these procedures, we obtain a data matrix
\[
M = \begin{pmatrix} S^1 & S^2 & \ldots & S^m \\ \end{pmatrix}
\]  
(13)
where the columns are reshaped symmetric matrices \( S^t, t = 1, 2, \ldots, m \). The procedures are summarized in Algorithm 2.

4.1 Experiment 1.
In the next few subsections, we conduct numerical experiments by applying Algorithm 2 to model the human mobility patterns.
Firstly, we use the visitor flow data of the year 2019 (52 weeks in total). We implement Algorithm 1 to obtain the DMD model \( x(t) \) as defined in (12). The singular values of the matrix \( M \) are plot in Fig. 1 in log scale. The distribution decreases slower after the 10th singular value. The DMD models \[12\] with target rank \( r = 5, 10, 15 \) are calculated following Algorithm 1.
Algorithm 1 DMD Algorithm [7]

1. Arrange the data \( \{x_1, \ldots, x_m\} \) into matrices

\[
X = \begin{pmatrix}
    x_1 & x_2 & \cdots & x_{m-1} \\
    \upharpoonright & \upharpoonright & & \upharpoonright \\
    x_2 & x_3 & \cdots & x_m
\end{pmatrix},
\quad
X' = \begin{pmatrix}
    x_2 & x_3 & \cdots & x_m \\
    \upharpoonright & \upharpoonright & \upharpoonright & \upharpoonright \\
    x_3 & x_4 & \cdots & x_{m+1}
\end{pmatrix}
\tag{7}
\]

where \( X \) and \( X' \) are defined in equation (5) and equation (6) respectively.

2. Compute the (reduced) SVD of \( X \), writing

\[
X = U\Sigma V^*
\tag{8}
\]

where \( U \) is an \( n \times r \) matrix, \( \Sigma \) is an \( r \times r \) diagonal matrix, \( V \) is an \( m \times r \) matrix, and \( r \) is the rank of \( X \) (known as the target rank).

3. Define the matrix

\[
\tilde{A} := U^*X'V\Sigma^{-1}
\tag{9}
\]

4. Compute the eigendecomposition of \( \tilde{A} \), writing

\[
\tilde{A}W = W\Lambda
\tag{10}
\]

where columns of \( W \) are eigenvectors and \( \lambda \) is a diagonal matrix containing the corresponding eigenvalues \( \lambda_k \).

5. Reconstruct the eigendecomposition of \( A \) from \( W \) and \( \Lambda \). The eigenvalues of \( A \) are given by \( \Lambda \) and the eigenvectors of \( A \) (DMD modes) are given by columns of \( \Phi \):

\[
\Phi = X'V\Sigma^{-1}W
\tag{11}
\]

6. Finally, the approximate solution at all future times is given by

\[
x(t) \approx \sum_{k=1}^{r} \phi_k \exp(\omega_k(t))b_k = \Phi \exp(\Omega t) b,
\tag{12}
\]

where \( \omega_k = ln(\lambda_k)/\Delta t \), \( b_k \) is the initial amplitude of each mode, \( \Phi \) is the matrix whose columns are the DMD eigenvectors \( \phi_k \), and \( \Omega = diag(\omega) \) is a diagonal matrix whose entries are the eigenvalues \( \omega_k \). The eigenvectors \( \Phi \) are the same size as the state \( x \), and \( b \) is a vector of the coefficients \( b_k \).
**Algorithm 2 Generate Data Matrix M**

for $t = 1, \ldots, m$

1. Create Origin-Destination Matrix $A^t_{k \times k}$ for the t-th week

   $$A^t_{ij} = \text{visitor}\_\text{flows}_t(i,j)$$  \hspace{1cm} (14)

   where $\text{visitor}\_\text{flows}_t(i,j)$ is the t-th week visitor flow from state i to state j and there are k states in total.

2. Get Symmetric Matrix $S^t_{k \times k}$

   $$S^t = (A^t + A^t^T)/2$$  \hspace{1cm} (15)

3. Reshape Matrix $S^t_{k \times k}$ and Fill Data Matrix $M_{k^2 \times m}$

   Finally, we convert square matrix $S^t_{k \times k}$ to a $k^2$-dimensional column vector $M_t$ by column and assign the vector $M_t$ to t-th column of matrix M.

end for

To evaluate the model accuracy, we randomly picked the origin with ID=2 and the destination with ID=32. The visitor flow data as well as DMD models between these two places from Week 1 to Week 105 (that is, from 2019 to 2020) are illustrated in Fig. 2. We observe that the DMD models with $r = 5, 10$ remains to fit well on the testing data from Week 57 to Week 62 (that is, the first 5 to 10 weeks in 2020), yet the DMD model with $r = 15$ flows up more rapidly than others. On the other hand, all the models seem to deviate from the real data after Week 62 (that is the 10th week in 2020).

Various errors between the real data and DMD model predictions with target rank $r = 10$ are shown in Table 1. Here $M$ refers to the data matrix, $M_{\text{dmd}}$ consists of predictions computed by (12), $\|M_i\|_2$ is the $L^2$ norm of the visitor flow of Week i, $\|M_{\text{dmd}}\|_2$ is the the $L^2$ norm of the DMD prediction of Week i, $\|M_i - M_{\text{dmd}}\|_2$ and $\|M_i - M_{\text{dmd}}\|_{\infty}$ are the relative $L^2$ and $L^\infty$ prediction error, respectively. It is clear that, DMD predictions can be far away from real data in the long run, as the model (12) grows exponentially in time $t$.

4.2 Experiment 2.

In this experiment, we use 2019 and 2020 data (104 weeks in total) to compute the DMD mode to predict 2021 data. The singular values of the data matrix $M$ is shown in Fig. 3. We will choose target rank $r = 20, 33, 45$ to compute different DMD models. We choose the places with ID=2 and ID=32 again. The real visitor flow data and DMD predictions from the first week of 2019 to the last week of 2021 (156 weeks in total) are shown in Fig. 4. Various errors with rank $r = 33$ are summarized in Table 2. This time, none of the DMD models seems to give reasonable predictions on testing data.

4.3 Experiment 3.

We repeat the experiment using visitor flow data from March 9, 2020 to May 31, 2021 (65 weeks in total) as the training data to compute DMD models. The singular values of the data matrix are shown in Fig. 5. The real data and DMD predictions with target rank $r = 12, 22, 33$ are displayed in Fig. 6 and the errors are summarized in Table 3.

We observe that the predictions are better than those in Experiment 2. The difference between the training data for Experiment 1 & 3 and that for Experiment 2 is that the latter contains a sharp change in March 2020. This was because of the presidential declaration of the state of national emergency on March 13, 2020, which leads to a drastic decay of the number of domestic visitors. The DMD models, as is shown in (12), consist of mostly low-frequency content, thus finds it difficult to learn this high-frequency content.
Figure 1: Singular Value of 2019 Data Matrix X

Figure 2: Real Visitor Flow and DMD Predictions in 2019 and 2020

Table 1: The Weekly Visitor Flows Data and Errors in 2020

| Week of 2020 | $||M_i||_2$ | $||M_{dmd_i}||_2$ | $||M_i - M_{dmd_i}||_2/||M_i||_2$ | $||M_i - M_{dmd_i}||_\infty/||M_i||_\infty$ |
|-------------|------------|------------------|----------------------------------|----------------------------------|
| 1st         | 2.2066e+07 | 2.0900e+07       | 0.0776                           | 0.0508                           |
| 2nd         | 2.1741e+07 | 2.2276e+07       | 0.0679                           | 0.0694                           |
| 3rd         | 2.1410e+07 | 2.2861e+07       | 0.0669                           | 0.0732                           |
| 4th         | 2.1071e+07 | 2.2923e+07       | 0.0713                           | 0.0742                           |
| 5th         | 2.2923e+07 | 2.2719e+07       | 0.0776                           | 0.0558                           |
Figure 3: Singular Value of 2019 and 2020 Data Matrix X

Figure 4: Real Visitor Flow and DMD Computation Visitor Flow Data from 2019 to 2021

Table 2: The Weekly Visitor Flows Data and Errors in 2021

| Week of 2021 | $||X_i||_2$ | $||M_{dmd,i}||_2$ | $\frac{||M_i-M_{dmd,i}||_2}{||M_i||_2}$ | $\frac{||M_i-M_{dmd,i}||_\infty}{||M_i||_\infty}$ |
|--------------|-------------|-----------------|---------------------------------|---------------------------------|
| 1st          | 2.1651e+07  | 1.9726e+07      | 0.1040                          | 0.1088                          |
| 2nd          | 2.1370e+07  | 1.9628e+07      | 0.0958                          | 0.1073                          |
| 3rd          | 2.0825e+07  | 1.9530e+07      | 0.0788                          | 0.0782                          |
| 4th          | 2.0556e+07  | 1.9432e+07      | 0.0779                          | 0.0901                          |
| 5th          | 2.1252e+07  | 1.9335e+07      | 0.1074                          | 0.1192                          |
Figure 5: Singular Value of Partly 2019 and 2020 Data Matrix X

Figure 6: Real Visitor Flow and DMD Computation Visitor Flow Data From March 9, 2020 to December 31, 2021

Table 3: The Weekly Visitor Flows Data and Errors in 2021

| Week of 2021 | $||X_t||_2$ | $||M_{dmdt}||_2$ | $||M_t-M_{dmdt}||_2/||M_t||_2$ | $||M_t-M_{dmdt}||_\infty/||M_t||_\infty$ |
|--------------|-------------|----------------|-------------------------------|----------------------------------|
| 23rd         | 2.6570e+07  | 2.8383e+07     | 0.0733                        | 0.0576                           |
| 24th         | 2.6968e+07  | 2.8500e+07     | 0.0640                        | 0.0633                           |
| 25th         | 2.6561e+07  | 2.8617e+07     | 0.0841                        | 0.0754                           |
| 26th         | 2.6720e+07  | 2.8735e+07     | 0.0824                        | 0.0703                           |
Acknowledgement

The research of L. Li and Y. Yang is partially supported by the NSF grant DMS-1715178, DMS-2006881, and the start-up fund from Michigan State University.

References

[1] Yuhao Kang, Song Gao, Yunlei Liang, Mingxiao Li, and Jake Kruse. Multiscale Dynamic Human Mobility Flow Dataset in the U.S. during the COVID-19 Epidemic. *Scientific Data*, pages 1–13, 2020.

[2] Hugo Barbosa, Marc Barthelme, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.

[3] Zhicheng Zheng, Zhixiang Xie, Yaochen Qin, Kun Wang, Yan Yu, and Pinde Fu. Exploring the influence of human mobility factors and spread prediction on early covid-19 in the usa. *BMC Public Health*, 21(1):1–13, 2021.

[4] George Kingsley Zipf. The *P*1/P2/D hypothesis: on the intercity movement of persons. *American sociological review*, 11(6):677–686, 1946.

[5] Samuel A Stouffer. Intervening opportunities: a theory relating mobility and distance. *American sociological review*, 5(6):845–867, 1940.

[6] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.

[7] J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic Mode Decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[8] P. J. Schmid and J. Sesterhenn. Dynamic mode decomposition of numerical and experimental data. In *61st Annual Meeting of the APS Division of Fluid Dynamics*, November 2008.

[9] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010.

[10] Jacob Grosek and J Nathan Kutz. Dynamic mode decomposition for real-time background/foreground separation in video. *arXiv preprint arXiv:1404.7592*, 2014.

[11] Joshua L Proctor and Philip A Eckhoff. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International health*, 7(2):139–145, 2015.

[12] Bingni W Brunton, Lise A Johnson, Jeffrey G Ojemann, and J Nathan Kutz. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of neuroscience methods*, 258:1–15, 2016.

[13] Jordan Mann and J Nathan Kutz. Dynamic mode decomposition for financial trading strategies. *Quantitative Finance*, 16(11):1643–1655, 2016.

[14] [https://github.com/GeoDS/COVID19USFlows](https://github.com/GeoDS/COVID19USFlows)

[15] Bruce H Webster Jr and Alemayehu Bishaw. US Census Bureau, American Community Survey Reports, ACS-02, Income, Earnings, and Poverty Data From the 2005 American Community Survey. *Washington, DC: US Government Printing Office*, 2006.

[16] Jonathan H. Tu, Clarence W. Rowley, Dirk Martin Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and Applications. *ACM Journal of Computer Documentation*, 1:391–421, 2014.

[17] Diana A Bistrian and I Michael Navon. An improved algorithm for the shallow water equations model reduction: Dynamic Mode Decomposition vs POD. *International Journal for Numerical Methods in Fluids*, 78(9):552–580, 2015.

[18] Neethu Mohan, KP Soman, and S Sachin Kumar. A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. *Applied energy*, 232:229–244, 2018.