Manifold Learning of the Dominant Modes of Human Mobility

James R. Watson[a,1], **Zach Gelbaum**[a,b], **Mathew Titus**[a,b], **Grant Zoch**[a], **and David Wrathall**[a]

[a]College of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA. 97330; [b]The Prediction Lab LLC, Corvallis, Oregon, USA, 97330

This manuscript was compiled on February 5, 2019

When, where and how people move is a fundamental part of how human societies organize around every-day needs as well as how people adapt to risks, such as economic scarcity or instability, and natural disasters. Our ability to characterize and predict the diversity of human mobility patterns has been greatly expanded by the availability of Call Detail Records (CDR) from mobile phone cellular networks. The size and richness of these datasets is at the same time a blessing and a curse: while there is great opportunity to extract useful information from these datasets, it remains a challenge to do so in a meaningful way. In particular, human mobility is multiscale, meaning a diversity of patterns of mobility occur simultaneously, which vary according to timing, magnitude and spatial extent. To identify and characterize the dominant modes of human mobility that emerge over time and space we examined CDR data from the Orange mobile network in Senegal using a new form of spectral graph wavelets, an approach from manifold learning. This unsupervised analysis reduces the dimensionality of the data to reveal seasonal changes in human mobility, as well as mobility patterns associated with large-scale but short-term religious events. The novel insight into human mobility patterns afforded by manifold learning methods like spectral graph wavelets have clear applications for urban planning, infrastructure design as well as hazard risk management, especially as climate change alters the biophysical landscape on which people work and live, leading to new patterns of human migration around the world.

Manifold Learning | Human Mobility | Complex System | Dimension Reduction | Prediction | Geography

Introduction

Human mobility is a fundamental part of how individuals, households and communities organize to meet every-day needs, and to respond to infrequent risks and shocks like economic instability and environmental hazards. Human mobility is multiscale in nature (1), that is for any given type of mobility, such as commuting, seasonal migration or holiday travels, individuals move as part of social collectives of varying size and interconnectivity, which span different magnitudes of spatial and temporal scale. Human mobility also has multiple modes of variability: people go to work each day, they go on holiday during specific programmed periods within the year, they may migrate before and after key agricultural seasons, or they may evacuate during floods or other environmental hazards (2). For these reasons and others, it is a continuing challenge to identify, categorize and anticipate the various patterns of human mobility (3). Anticipating and planning for human mobility is a non-trivial task for organizations whose core functions provide critical services to and address the needs of moving people, such as urban planning and transport agencies, disaster first-responders and international aid organizations (4).

To overcome these challenges and generate fundamental insight on human mobility, novel data generated by users of the digital infrastructure (e.g. mobile phone subscribers) is now being used. So-called Big Data, routinely collected from a range of sources, most notably the explosion of mobile phone usage throughout the world, provides rich information on users' locations through time (5). Mobile network operators collect records of their users' calling patterns, a type of data called Call Detail Records (CDR), which include the location of the receiving tower where each voice call or text message is made, as well as the location of the recipient. Over time, each user's calling patterns can be used to reconstruct a detailed record of their location history. The collective mobility history of all users' movements through time provides insight on total population flows between all cellular network locations during any specified period of time. This enables the study of users' behaviors at very high spatiotemporal resolution over both local and system-level spatial scales at time scales of minutes to months to years (e.g. 6). As each phone is embedded within an existing social fabric,

Significance Statement

Mobile phone data can provide insight about patterns of human mobility as people go about their every-day lives and as they respond to economic and environmental risk. The challenge is that these data are large, and as a consequence difficult to use and interpret, and any signal of a importance can be lost in the noise. To overcome this challenge, we have developed a new approach to human mobility data analysis using tools from manifold learning. Specifically, spectral graph wavelets were used to extract the signal of different modes of human mobility, based on changes in mobility-network scale. This novel approach to characterizing and predicting human mobility has practical applications for urban planning, environmental risk management, and humanitarian intervention.

Z.G. and M.T. performed the mathematical analysis. Z.G. and J.W performed the data analysis and wrote the paper. G.Z. and D.W. provided the data and contributed to the design of the analysis and the writing.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: jrwatson@coas.oregonstate.edu

CDR allow the analysis of the changing structure of social organization as people (i.e. individuals, social networks, communities, religious and ethnic groups, etc.) respond to a diversity of stimuli. CDR data has been used for urban planning (e.g. 7), and to describe and evaluate commuting (e.g. 8) and environmental displacement resulting from earthquakes and hurricanes (e.g. 9, 10). They have also been used to evaluate mobility as a vector for the spread of infectious diseases such as ebola (e.g. 11–14).

CDR data offer a vastly clearer and more detailed description of the communication and mobility behaviors of people as they go about their daily lives, compared to traditional data on human mobility, such as surveys and censuses, which are collected at relatively coarse time-scales, i.e. months and years (i.e. 15). CDRs are compiled by network operators principally for the purposes of billing customers for their use of the network, not for scientific analysis; therefore, the main challenge with analyzing CDR is the size, complexity and richness of datasets. CDR are inherently high dimensional and noisy (16), and quantitative analyses must incorporate dimensional reduction and denoising approaches. Once done, the remaining challenge is retrieving a relevant signal from the data at the appropriate spatial and temporal scale for each specific mobility pattern (6).

Here, we describe a data-driven approach to CDR analysis that explicitly addresses the multiscale nature of the mobility patterns embedded in the data and reflected in the system under study. We analyzed CDR from users of the Orange/Sonatel cellular network, collected in Senegal between 1 January and 31 December 2013 (see Fig. 1 for an illustration of mobility in Senegal for a given day). De-identified data entries included information on the time of the call, the mobile phone tower used and the duration of call. To identify and characterize different modes of human mobility captured in CDR, we developed a novel computational approach based on spectral graph wavelets, an extension of classical wavelet analysis to the setting of networks. There are now numerous examples of where CDR data has been used to understand patterns of human mobility (e.g. 17–19), including those that also explicitly address multiscale patterns (20) using methods from statistical physics (e.g. 3, 21). Here, we have made new mathematical advances to spectral graph wavelets to improve upon these state of the art approaches to the multiscale decomposition of CDR data. In doing so we were able to identify and characterize various modes of human mobility in Senegal for the year 2013.

Our approach is localized in nature, that is multiscale information is produced for each node in a given human-mobility network. No information is lost in doing this transformation, indeed clarity is gained as to the spatio-temporal patterns of mobility, and to demonstrate the utility of this approach we focused on dynamic multiscale mobility patterns to/from a specific city in Senegal - Touba, a market town and religious center in Senegal's agricultural breadbasket. Our goal was to identify and characterize the different modes of human mobility that contrast in overall spatial scale and temporal duration, focusing on inter-city forms of mobility to/from Touba, including seasonal migration relating to changes in agriculture and a mobile labor force, as well as punctuated patterns relating to calendar festivals, local elections, or religious holidays. The identification and interpretation of these dominant modes, and their effective scale, is of value to the overarching goal of extracting key dynamics from complex system in general, and not just limited to those given by CDR data.

Human Mobility Data

To analyze the multiscale nature of human mobility, we developed a new approach to spectral graph wavelets which we then applied to de-identified, pre-processed extracts of CDR from the Orange/Sonatel mobile network in Senegal generated between 1 January and 31 December 2013. These data include a number of variables on individual mobility behaviors/locations for 300,000 randomly sampled users on a rolling 2-week basis. These data were arranged into a pairwise origin-destination network whose vertices, or nodes, are cell towers and whose edges are weighted by the volume of phones moving/flowing between each tower pair for each 24-hour period over the entire year of 2013. To highlight inter-city mobility, we first aggregated the towers of key cities into single nodes. Then, for each 24 hour period, a one was added to a given edge i, j if a phone moved from node i to node j. To further filter out intra-city mobility and patterns of daily commuting, repeated trips between the same towers for the same user within a 24-period were not counted. Thus, mobility was defined and quantified as an asymmetric affinity matrix A varying through time, where $A_{ij}(t)$ is the total number of unique trips made by users between towers i and j on day t of 2013 (see Fig. 1A for a geographic representation for one of these daily affinity matrices). In order to normalize the data so that the signal from high volume mobility in areas such as Dakar, the capital of Senegal, did not wash out all other signals, we calculated relative mobility by dividing the entries of $A(t)$, by their respective row sums. In order to perform the SGW analysis, we then converted this asymmetric affinity matrix to a symmetric one by taking the average of two-way mobility patterns between pairs of cell-towers. Last, so as to not include intra-city traffic, we set the diagonal of A to zero. The result was a symmetric affinity matrix/network with zeros on the diagonal for every day of 2013.

Multiscale Analysis using Spectral Graph Wavelets

Networks of human mobility and communication are naturally modeled by graphs, sets of nodes connected by edges, each with an associated weight representing the strength of the connection. Here, higher edge weights indicate a greater flow of human migration between the edge's two endpoints. Graphs can be viewed as discrete analogues of smooth objects such as geometric manifolds and surfaces, and the geometric content of these graphs is often analyzed using an associated Laplacian operator (22, 23), as it encodes the geometric structure of the graph in a natural way. This is the starting point of our analysis.

Given that human mobility networks are inherently multiscale as well as highly heterogeneous, a method is required that is both multiscale and localized. In the classical settings of time series analysis and image analysis, wavelets are the proper tool for conducting localized, multiscale decompositions and analysis. Here we employed a generalization of wavelets to graphs, spectral graph wavelets. Spectral graph wavelets have been previously used to study human mobility, specifically automobile traffic (24), in analyzing human mobility from photo activity via Flickr (25), and in general clustering and community detection (26).

An important first choice to make when employing spectral graph wavelets is the shape of the wavelet to be used, or in other words the "wavelet kernel".This can have a profound effect on the analysis. Here, we developed a wavelet kernel based specifically on the heat kernel yielding what we call Hermitian Graph Wavelets (for details of the mathematical analysis see: 27). By using the heat kernel, we were able to produce a data analysis method that faithfully and efficiently extracted key geometric information from the time-varying human mobility networks $A(t)$. The outputs of this analysis are a set of wavelet functions associated to each vertex (i.e. for every cell tower in Senegal), as well as a single *dominant scale* for each vertex representing the scale containing the most information for that vertex. These wavelet functions are rich in multiscale geometric content, and we used them to identify and characterize different forms or modes of human mobility that occur throughout Senegal over the period 2013.

Applying Spectral Graph Wavelets to Mobility Networks. Given the set of daily affinity matrices, $A(t)$, the application of Spectral Graph Wavelets is as follows. First, for each of the daily affinity matrices, we constructed their associated Laplacian matrices:

$$\Delta(t) = D(t) - A(t),$$

where $D(t)_{ii} = \sum_j (A(t))_{i,j}$ and $D(t)_{ij} = 0$ for $i \neq j$. Once done, the eigenvectors and eigenvalues of $\Delta(t)$, $\{\phi_{k,t}\}$ and $\{\lambda_{k,t}\}$ respectively, were then computed. Then, Hermitian graph wavelets were formed as:

$$\psi_{s,x}(y) = \sum_k s\lambda_k(t)e^{-s\lambda_k(t)}\phi_{k,t}(x)\phi_{k,t}(y), \qquad [1]$$

for a chosen set of scales $s \in \{s_n\}$ (note that $\psi_{s,x}(y) = \psi_{s,y}(x)$). These functions exhibit several properties that make them ideal for decomposing signals measured on large and complex networks: they are localized, as described above, and the power of each wavelet (i.e. its norm, as a vector) gives us a measure of its importance with respect to the global network structure (27). This is analogous to classical Fourier analysis where large Fourier coefficients in the decomposition of a function indicate the major modes comprising the function. Rather than forming an orthonormal set, the wavelet functions $\{\psi_{s_n,x}\}$ form a *frame* and rather than the classical Parseval equality, an approximate Parseval equality holds: for a function f on the network there are constants $0 < A < B < \infty$ with

$$A\|f\|^2 \leq \sum_{s_n,x} |\langle f, \psi_{s_n,x}\rangle|^2 \leq B\|f\|^2,$$

where

$$\|f\|^2 = \sum_x |f(x)|^2,$$

with the sum taken over all nodes x in the network. The values of A and B depend both on choice of wavelet kernel and the scales chosen. It is important to note that it is up to the investigator to define the set of scales $\{s_n\}$ to be input into the above algorithm. This determination is largely ad hoc and some trial and error is required. However some general points can be made: the chosen scales should always be positive and the largest should not be too much bigger than the largest eigenvalue. The goal is to get a good partitioning of the interval $[0, \max\{\lambda(t)\}]$ relative to the spacing of the eigenvalues, but as the eigenvalues are in general not uniformly spaced some experimentation may be required to find the resolution that is most informative. It is always possible to choose scales and (by renormalizing if necessary) achieve a frame with the values of A and B as close to 1 as desired (i.e. 28).

With an appropriately chosen set of scales, if we let f be a delta function at node z, $f = \delta_z$ (i.e. f(z)=1 and f=0 for all other nodes), the above Parseval bounds imply that

$$1 = \|f\|^2 \approx \sum_{s_n,x} |\langle f, \psi_{s_n,x}\rangle|^2 = \sum_{s_n}\sum_x |\psi_{s_n,z}(x)|^2 = \sum_{s_n} \|\psi_{s_n,z}\|^2.$$

Thus the values of $\|\psi_{s_n,z}\|^2$ serve to indicate the relative importance of each scale with respect to the vertex z. In this way, large values of $\|\psi_{s_n,z}\|^2$ correspond to the major scales of importance at the node z. We utilize this intuition and define the dominant scale at each vertex to be

$$S(z,t) = \arg\max_{s_n} \|\psi_{s_n,z}\|^2. \qquad [2]$$

This measures the scale at which a given vertex is most well-connected to the rest of the network.

The above Hermitian graph wavelet functions yield a complete multiscale analysis at each vertex in the graph (i.e. for every cell tower in Senegal). To highlight the ability of spectral graph wavelets to identify multiscale patterns, we chose to build our wavelets with a fixed base point at the vertex corresponding to the city of Touba (denoted x_T), and track the dominant scale $S(x_T, t)$ as above, through time. We abbreviate this as $S(t)$. We thus obtain a sequence of dominant wavelet functions on the network $\psi_{S(t),x_T}(t, y)$, where the first argument indicates the dependence on the changing network structure encoded in $A(t)$ and the second argument y corresponds to vertices of the graph on which the function takes its values.

Results

To demonstrate the utility of spectral graph wavelets for identifying dominant modes of human mobility, we calculated dominant wavelet functions centered on Touba, the market city in Senegal's central breadbasket and an important site for religious festivals, and compared results with those produced from a basic analysis of the original population flow data. More specifically, for any given day, human mobility to/from a given location can be extracted from the mobility matrixes $A(t)$ and visually inspected (Fig. 2A). Doing so for Touba, one finds that large cities such as Dakar on the west coast account for most of the total daily flow in and out of Touba (i.e. compare the location of large red nodes in Fig. 2A).

In contrast, Fig. 2B depicts the dominant wavelet function for the same day as plotted in Fig. 2A. The function $\psi_{S(t),x_T}(t, y)$ reveals a spatially distinct pattern, with a positive value centered on Touba, that rapidly diminishes to negative values in surrounding nodes, before approaching zero at geographically remote nodes less linked by patterns of mobility. While the wavelet function was calculated from the population flow data, and did not evaluate spatial distance between location pairs explicitly, the spatial dependence of human mobility (i.e. the proximity to Touba to other locations) has an imprint on the resulting wavelet function.

To characterize changes in human mobility through time, focusing on Touba, we first examined changes in the total traffic to/from Touba over time (Fig. 3A); this is calculated by summing the mobility values to/from Touba for all locations for each day. Log_{10} total mobility to/from Touba over time reveals a variety of qualitative features, most striking is a large peak in traffic corresponding to the commemoration of Touba's mosque's 50th anniversary, which occurred May 30th, 2013 (day 150), as well as the end of Ramadan August 7th, 2013 (day 219) and Eid al-Adha on October 15 (day 287).

We then clustered days based upon origin-destination flow values in A(t), that is for every day, the vector of mobility values to/from Touba were extracted from the mobility matrices $A(t)$. We clustered days based on these values, using the Louvain method for community detection (29). This involved calculating the Euclidean distance between days based on their mobility values. We chose the Louvain method primarily because it provides an objective determination of the number of clusters. This is in contrast to other common clustering approaches which require the number of clusters to be chosen (e.g. k-means).

Clustering these mobility data for Touba identified two dominant time periods, corresponding to seasonal changes in the Senegalese weather and agriculture, as the rainy season begins around September. There is a third cluster corresponding to a short intermediate period (see Fig. 3A, changes in the marker color: green, grey and blue correspond to the dry, intermediate and wet seasons respectively). Interestingly, the clustering only picked out these long-term changes in mobility, and not the short-punctuated events relating to religious festivals or political events. This is because while these kinds of short-term events are associated with a change in total traffic to/from Touba, any topological changes in the inter-city traffic profile are obscured by the complexity and irregularity of the data.

In contrast to these results created by analyzing the raw mobility values, clustering dominant wavelet functions over time revealed more nuanced information about mobility in Senegal in 2013. Clustering was done using $\psi_{S(t),x_T}(t,y)$ in place of population flow values: we calculated the Euclidean distance between daily pairs of Touba-centered wavelet functions, before using the Louvain community detection algorithm to identify clusters. Like the analysis of the relative population flow values described above, this clustering of daily wavelet functions identifies changes in human mobility relating to the dry and wet season. Importantly however, now the presence of relatively short-term and large-scale events were identified throughout the year. These events are marked by a short-term widening of the wavelet function centered on Touba (in network space), and correspond to the three religious migration events listed above. In addition, other events are identified: the Maouloud/Gamou celebration that occurred on January 23, 2013 (day 23) and Independence Day that occurred April 4th, 2013 (day 94). There are two extra events that this clustering approach identified, around day 50 and at the end of the year. The former is an unknown event, which speaks to the value of the spectral graph wavelet analysis in identifying new and interesting features, and the latter is likely associated with new-years celebrations/holidays. Crucially, the total traffic to/from Touba does not change during many of these events, and they are not identified by clustering the relative population flow data; the dominant-scale wavelet functions fluctuate due to changes in connectivity within the network and this manages to distinguish between typical traffic patterns and migration events.

Averaging the daily wavelet functions associated with each cluster (i.e. producing an average dominant wavelet function centered on Touba for the dry and wet seasons, and for each short-term / large-scale event) reveals stark spatial differences (Fig. 4). For example, the absolute difference between the average wavelet function relating to the dry and wet seasons (Fig. 4A) identifies change in several coastal towns (along with changes in mobility in and around Touba). These changes reflect two processes. First, in Senegal there is a large and mobile agricultural work-force. Seasonally employed farm laborers from the coast use Touba as a stepping stone to rural locations in the interior of the country. Second and also associated with the start of the rainy season are floods, and in 2013 the coast experienced significant flooding. These differences in the average wavelet functions associated with the dry and wet clusters also reflect the impact of the floods on human mobility.

Additional nuances emerge in the punctuated modes of mobility associated with the religious festivals noted earlier. In contrast to the seasonal changes in mobility, the absolute difference between the average wavelet function associated with short-term / large-scale events, in this case Eid al-Adha on October 15 (day 287), and the wet season identifies change in mobility to/from many small rural towns in the far east of Senegal (Fig. 4B). These differences in average dominant wavelet functions clearly identify the long-distance travel that people make as they go to/from Touba for the religious event.

This kind of geographically meaningful information can be used to explore differences in the various modes of human mobility found by clustering daily wavelet functions. In general, this approach more readily reveals heterogenous spatial and network features, relative to results produced from the analysis of the relative population mobility values, which simply described major seasonal changes in flows to/from Touba. Importantly, it bears noting that while we present results from Touba, the spectral graph wavelet analysis also characterizes the dominant modes of human mobility *for every location within the network*. Hence, while we have demonstrated its utility with regards to Touba, there is additional spatially meaningful information that we did not show or analyze.

Discussion

To improve upon current abilities to characterize changes in human mobility over time we developed and applied a new manifold learning method based on spectral graph wavelets (i.e. Hermitian graph wavelets) to a CDR dataset. These data describe the origin-destination mobility patterns for people living in Senegal, daily for the year 2013. Our approach is non-linear, localized and scale explicit, that is it can be used to identify multiscale patterns of human mobility for each node in a mobility network. This is key to disentangling multiscale patterns with big and rich datasets like those produced by CDRs. Spectral graph wavelets applied to these data allowed us to identify seasonal changes in Senegalese human mobility, as well as punctuated large-scale events relating to religious migration and a national holiday. These short-term but large-spatial-scale events were not identified by a standard approach applied to the original mobility values themselves (i.e. $A(t)$), because these data contained too much noise. As a consequence, our spectral graph wavelets approach provides a new method for the multiscale decomposition of human mobility data, and expands the utility of CDR data for anticipating and preparing for changes in modes of human mobility ranging from day-to-day commutes to large-scale migrations and displacements caused by natural disasters.

Identifying the dominant modes of human mobility, and the spatio-temporal scales at which they occur, is an important part of developing policies for a whole host of operations, ranging from traffic infrastructure to disaster response planning (4). These kinds of spatial policies have previously been made using far coarser data, both spatially and temporally (e.g. 15). Here, making use of relatively high resolution CDR data, we were able to tease out the spatial signal of punctuated large-scale events, in contrast to a standard approach applied to data on population flow values. The rate at which new data is created – always at higher temporal frequency and greater spatial resolution – is ever growing, and the quantitative tools used to extract useful information from monthly or even annual datasets are rapidly diminishing in utility (30). For the best use of these new data (i.e. next-gen CDR datasets) new tools are required, and indeed, new tools that can articulate the complex and adaptive nature of human mobility have extreme utility (e.g. 31). Like human mobility, other complex adaptive systems are multiscale by nature (32), and in general there is a growing need to extract information about the micro-scale agents that comprise these systems, from which meso- and macro-scale patterns emerge (33). Data-analytic tools designed with the multiscale nature of complex adaptive systems in mind will help policy makers develop plans that explicitly account for the emergence of patterns over a continuum of scales, like in this case the various modes of human mobility in Senegal, and their associated network/geographic scale.

The use of spectral graph wavelets allowed us to essentially transform the origin-destination mobility data (i.e. $A(t)$) to a form that better highlights the differences of human mobility patterns. In that spirit, this analysis can be thought of as a process of dimensional reduction or a denoising of the raw data, in a manner that accounts explicitly for network scale. Analyzing the mobility data did not provide the same kinds of scale-dependent information because it is noisy. Admittedly, when analyze the mobility matrices $A(t)$ we used a very basic approach to classification. There are indeed many other more sophisticated approaches that we could have been employed, in order to contrast with the results produced from analysis of the dominant wavelets functions. These approaches vary from traditional dimensional reduction techniques that rely on linear correlations, such as Principal Components Analysis (PCA), to machine learning approaches for feature identification (e.g. 34). Indeed, we see that there is a great opportunity for using the wavelet transformed data in combination with machine learning approaches to classification and feature extraction. No information is lost in the process of our scale-explicit denoising of raw data using spectral graph wavelets, and combining these results with machine learning approaches to classification is likely a fruitful vein of future research.

The multiscale and localized information that spectral graph wavelets provides can be used in many other ways. Here, we have analyzed human mobility information gained from the CDR data, but the CDR dataset can also be used to construct human communication networks through time. Performing the same analysis on both sets of data would produce concurrent wavelet functions through time. A comparison of changes in the dominant modes / scales of human communication and mobility might reveal early-warning signals of migration/displacement. Simply put, as people prepare to move they are likely to call their ultimate destination, and this information can help policy makers prepare for changes in population density at specific nodes/places. There is an opportunity to utilize methods from manifold matching (e.g. 35) to make these comparisons. Manifold matching has been used in image recognition to match photos of the same person, for example. Here, instead of a set of photos from a person's face, the manifolds that would be compared are those associated with a complex system (the Senegal cellular network) described in two ways (i.e. communications and mobility).

Additional early-warning signals of mass human-mobility can also be sought from the dynamics inherent to mobility networks alone. These kinds of early-warning signals come from dynamical systems and bifurcation theory (36) and are measured by changes in the variance and autocorrelation in macroscopic variables (37), for example changes in the density of people in a certain neighborhood. For human mobility CDR data, there is an opportunity to advance new early-warning signals of multiscale change using manifold learning. Specifically, spectral graph wavelets is one way to learn changes in the geometry of the manifold on which dynamics occur, but there are others, for example diffusion maps (23) and Laplacian eigenmaps (38). Systems undergoing a bifurcation driven by some macroscopic variable should *a fortiori* exhibit changes in geometry at small and intermediate scales as well; a multiscale analysis may then allow one to directly address how these kinds of large and abrupt changes in complex systems are related to changes in the behavior of micro-scale agents (i.e. in this case, how individual people move from place to place).

Identifying the dominant modes of human mobility, and potentially early-warning signals of changes between them, is of principal interest of groups tasked with managing human communities (4), as they go about their everyday lives as well as respond to infrequent but impactful events like a natural hazard. In Senegal, flooding is a persistent problem and indeed in 2013 the capital Dakar was severely hit. These kinds of events can lead to the permanent displacement of people from their homes, and similarly to identifying dominant modes of human mobility as done here, there is value to identifying where and when this displacement occurs (39). Indeed, displacement is not necessarily instantaneous with regards to the perturbation, but it may take a relatively long time for people to "realize" their displacement (40). Multiscale methods like spectral graph wavelets applied to CDR data can help distinguish these additional modes of human mobility, and further, methods from manifold matching are likely to be useful too.

In sum, we have made advances to spectral graph wavelets (specifically Hermitian graph wavelets) for analyzing CDR human mobility data. Our approach extracts useful information that is localized and scale-explicit, and we identified seasonal changes in human mobility as well as punctuated large-scale mobility events associated with religious celebrations and a national holiday. Here, we focused our multiscale analysis on one place in Senegal - Touba - a place of religious significance. However, the spectral graph wavelets analysis produces information for all nodes in the network, and there is rich vein of scale-explicit information in the full wavelet transform of the origin-destination mobility data. Last, while the growth in data obtained for complex adaptive systems is daunting (36), there are opportunities to employ new localized and scale-explicit dimensional reduction techniques, like we have done so here, to greatly improve our ability to characterize and predict multiscale change. This ability is vital if we are to maintain welfare from the complex systems in which we are embedded.

ACKNOWLEDGMENTS. The authors would like to acknowledge the DARPA Young Faculty Award YFA N66001-17-1-4038, and the Orange D4D competition for providing the data.

1. Song C, Koren T, Wang P, Barabasi AL (2010) Modelling the scaling properties of human mobility. Nature Physics 6(10):818.
2. Widhalm P, Yang Y, Ulm M, Athavale S, Gonzalez MC (2015) Discovering urban activity patterns in cell phone data. Transportation 42(4):597–623.
3. Simini F, Gonzalez MC, Maritan A, Barabasi AL (2012) A universal model for mobility and migration patterns. Nature 484(7392):96.
4. Jiang S, Ferreira J, Gonzalez MC (2017) Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. IEEE Transactions on Big Data 3(2):208–219.
5. Giannotti F, et al. (2011) Unveiling the complexity of human mobility by querying and mining massive trajectory data. The International Journal on Very Large Data Bases 20(5):695–719.
6. Song C, Qu Z, Blumm N, Barabasi AL (2010) Limits of predictability in human mobility. Science 327(5968):1018–1021.
7. Becker RA, et al. (2011) A tale of one city: Using cellular network data for urban planning. IEEE Pervasive Computing 10(4):18–26.
8. Iqbal MS, Choudhury CF, Wang P, Gonzalez MC (2014) Development of original destination matrices using mobile phone call data. Transportation Research Part C: Emerging Technologies 40:63 – 74.
9. Lu X, et al. (2016) Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in bangladesh. Global Environmental Change 38:1 – 7.
10. Lu X, Bengtsson L, Holme P (2012) Predictability of population displacement after the 2010 haiti earthquake. Proceedings of the National Academy of Sciences 109(29):11576–11581.
11. Wesolowski A, et al. (2014) Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. Scientific Reports 4:1–7.
12. Balcan D, et al. (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences 106(51):21484–21489.
13. Belik V, Geisel T, Brockmann D (2011) Natural human mobility patterns and spatial spread of infectious diseases. Physical Review X 1(1):011001.
14. Wesolowski A, et al. (2014) Commentary: containing the ebola outbreak - the potential and challenge of mobile network data. PLoS Currents 6:1–20.
15. Bell M, Ward G (2000) Comparing temporary mobility with permanent migration. Tourism Geographies 2(1):87–107.
16. Chen CP, Zhang CY (2014) Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences 275:314–347.
17. Jarv O, Ahas R, Witlox F (2014) Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. Transportation Research Part C: Emerging Technologies 38:122 – 135.
18. Calabrese F, Diao M, Lorenzo GD, Ferreira J, Ratti C (2013) Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. Transportation Research Part C: Emerging Technologies 26:301 – 313.

19. Dobra A, Williams NE, Eagle N (2015) Spatiotemporal detection of unusual human population behavior using mobile phone data. PLOS ONE 10(3):1–20.
20. Phithakkitnukoon S, Smoreda Z, Olivier P (2012) Socio-geography of human mobility: A study using longitudinal mobile phone data. PLOS ONE 7(6):1–9.
21. Lambiotte R, et al. (2008) Geographical dispersal of mobile communication networks. Physica A: Statistical Mechanics and its Applications 387(21):5317 – 5325.
22. Chung FRK (1997) Spectral graph theory, CBMS Regional Conference Series in Mathematics. (Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI) Vol. 92, pp. xii+207.
23. Coifman RR, Lafon S (2006) Diffusion maps. Appl. Comput. Harmon. Anal. 21(1):5–30.
24. Mohan DM, Asif MT, Mitrovic N, Dauwels J, Jaillet P (2014) Wavelets on graphs with application to transportation networks in 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). pp. 1707–1712.
25. Dong X, Ortega A, Frossard P, Vandergheynst P (2013) Inference of mobility patterns via spectral graph wavelets in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 3118–3122.
26. Tremblay N, Borgnat P (2013) Multiscale community mining in networks using spectral graph wavelets in 21st European Signal Processing Conference (EUSIPCO 2013). pp. 1–5.
27. Gelbaum Z, Titus M, Watson J (2019) Multi-scale analysis on complex networks using hermitian graph wavelets. arXiv arXiv:1901.07051.
28. Hammond DK, Vandergheynst P, Gribonval R (2011) Wavelets on graphs via spectral graph theory. Appl. Comput. Harmon. Anal. 30(2):129–150.
29. Clauset A, Newman ME, Moore C (2004) Finding community structure in very large networks. Physical review E 70(6):066111.
30. Lee JG, Kang M (2015) Geospatial big data: challenges and opportunities. Big Data Research 2(2):74–81.
31. Scheffer M, et al. (2018) Quantifying resilience of humans and other animals. Proceedings of the National Academy of Sciences 115(47):11883–11890.
32. Levin SA (1998) Ecosystems and the biosphere as complex adaptive systems. Ecosystems 1(5):431–436.
33. Folke C (2006) Resilience: The emergence of a perspective for social–ecological systems analyses. Global environmental change 16(3):253–267.
34. Bi J, Bennett K, Embrechts M, Breneman C, Song M (2003) Dimensionality reduction via sparse support vector machines. Journal of Machine Learning Research 3(Mar):1229–1243.
35. Shen C, Vogelstein JT, Priebe CE (2017) Manifold matching using shortest-path distance and joint neighborhood selection. Pattern Recognition Letters 92:41–48.
36. Scheffer M, et al. (2009) Early-warning signals for critical transitions. Nature 461(7260):53.
37. Boettiger C, Ross N, Hastings A (2013) Early warning signals: the charted and uncharted territories. Theoretical ecology 6(3):255–264.
38. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation 15(6):1373–1396.
39. Xie M, Neal J, Burke M, Lobell D, Ermon S (2016) Transfer learning from deep features for remote sensing and poverty mapping. arXiv arXiv:1510.00098.
40. Black R, Arnell NW, Adger WN, Thomas D, Geddes A (2013) Migration, immobility and displacement outcomes following extreme events. Environmental Science & Policy 27:S32–S43.

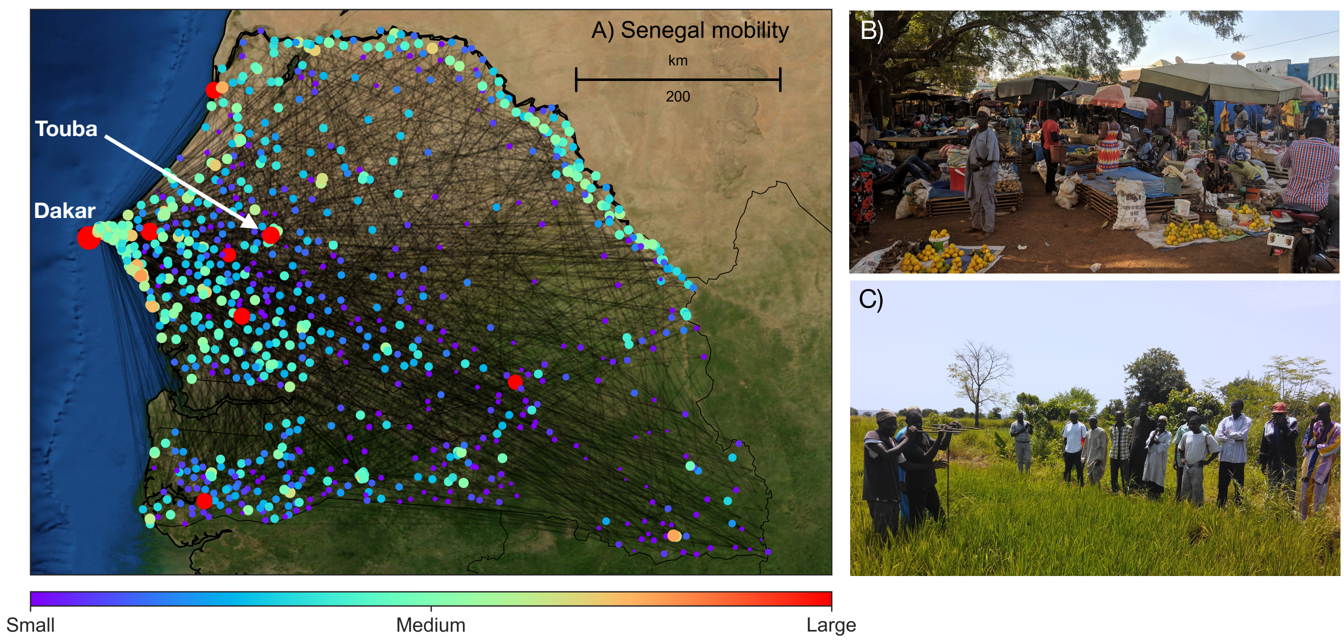

Fig. 1. A) Map of Senegal with major cellular communication towers (i.e. human mobility nodes) color coded and sized by population density (i.e. town/city size) and edges connecting them highlighting the density and complexity of human communication and mobility networks. These networks are changing over time, and in Senegal there are large-scale modes of mobility relating to religious events (often held at Touba) and relating to seasonal changes and agriculture (B and C).

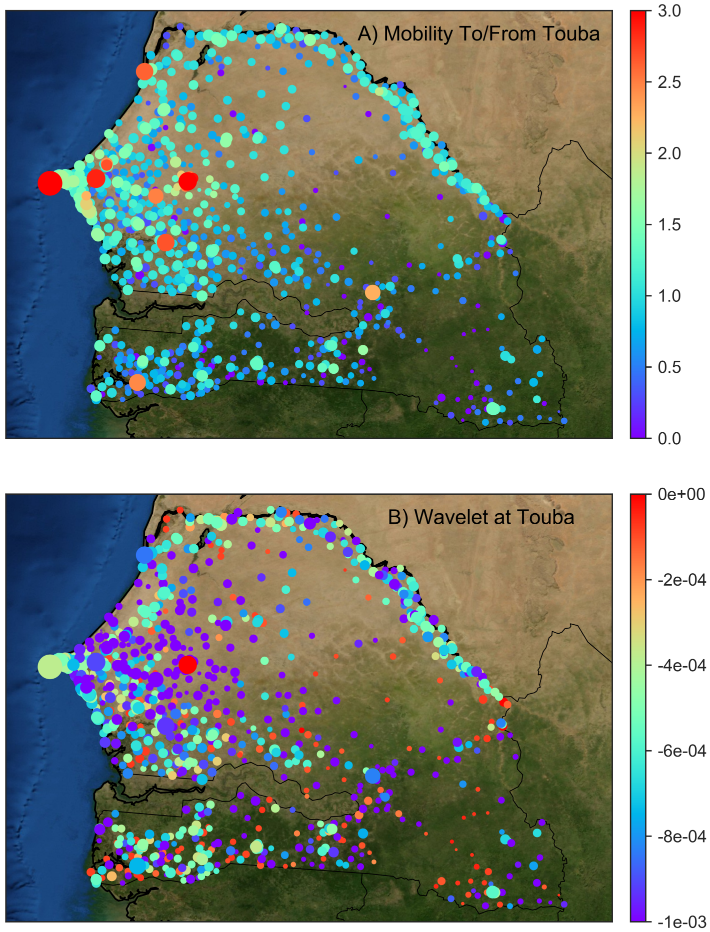

Fig. 2. A) Cell-towers (i.e. human mobility nodes) in Senegal color coded by the log_{10} number of people moving to/from Touba on a random day in 2013, and sized by local population density. B) In contrast, colors now denote the dominant wavelet function centered on Touba for the same day (node size remains proportional to local population density). The two maps reveal very different information. In (A) mobility to/from the major urban hubs, such as Dakar the capital, are identified and in (B) the shape of the wavelet function is highlighted (positive at Touba, the large red node near the center of Senegal, decreasing to large negative values in nearby towns, before becoming less negative at far-off towns).

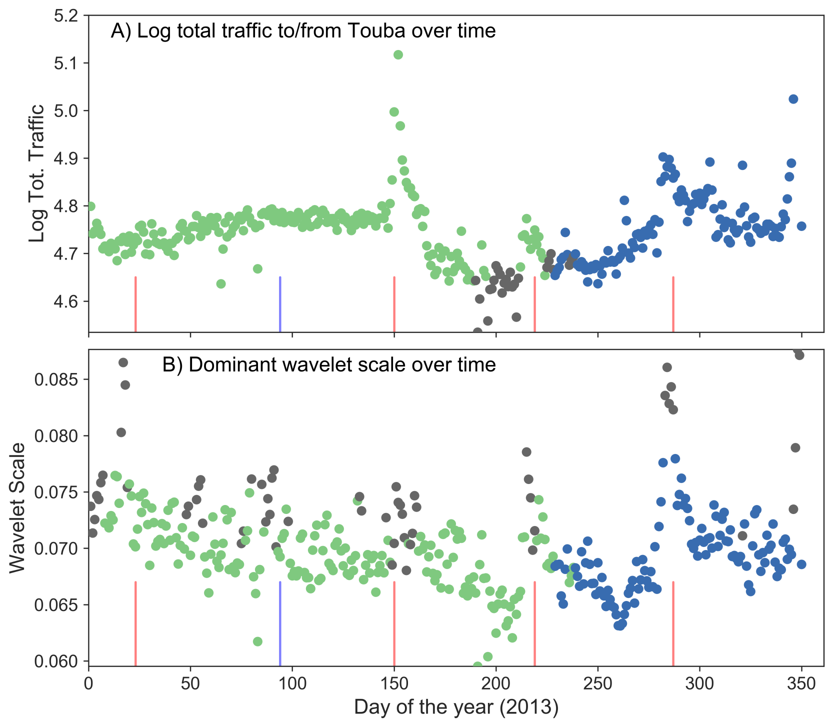

Fig. 3. A) Time-series of the log_{10} total mobility to/from Touba for 2013, colored by group ID produced from clustering days based on their mobility values to/from Touba. This cluster analysis identifies mobility associated with the dry (green) and wet (blue) seasons (with an transitionary phase in grey), and there are peaks in total traffic to/from Touba that correspond with major religious events at day 150, 219 and 287; however, the cluster analysis does not recognize these events as distinct. B) In contrast, changes over time in the dominant wavelet scale (centered on Touba) better reveals the punctuated and large-scale mobility events related to religious celebrations (vertical red lines) and Senegal's independence day (blue vertical line). In addition, using these dominant wavelet functions to cluster days extracts meaningful information about seasonal migration, as well as the short-term / large-spatial scale events (in grey).

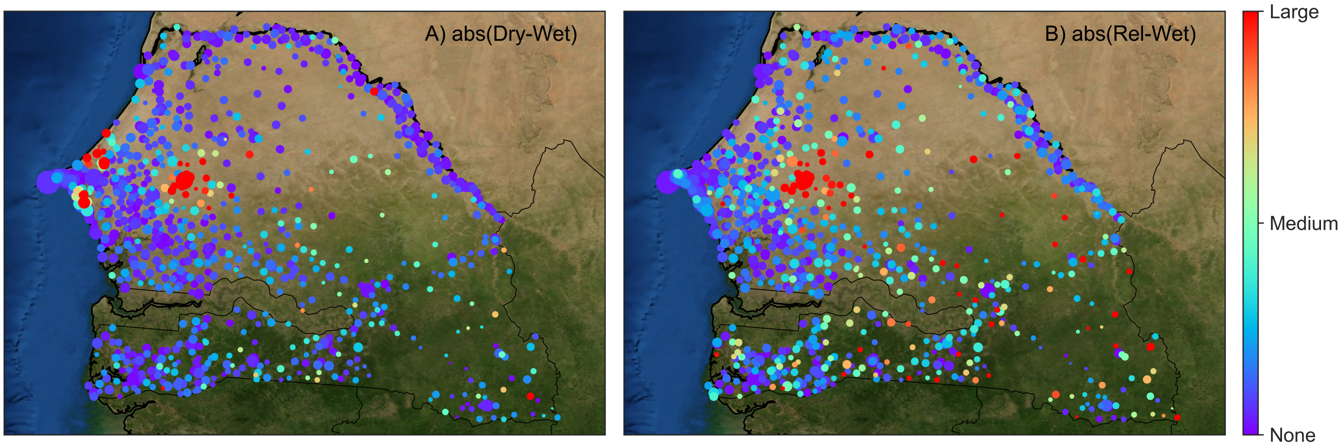

Fig. 4. Dominant modes of human mobility can be identified by averaging the dominant wavelet functions associated with the cluster groups shown in Fig. 3B. Then, these modes of human mobility can be compared by calculating their absolute difference and plotting as a map. A) The absolute difference in the average dominant wavelet function (centered on Touba) associated with the dry and wet seasons highlights changes in mobility to/from the coast and Touba. B) The absolute difference in the average wavelet function (centered on Touba) associated with religious events and the wet season reveals changes in mobility in far-off towns in the east of Senegal, and Touba.