**DARPA Progress, Status, and Management Report**

**Grant # N66001-17-1-4038**

**8th Quarterly report**

**Feb-Apr 2019**

**Oregon State University**

**Technical Point of Contact:** James R. Watson

**Administrative Point of Contact:** Kelsey L. Herman

**Title**:

Comparing Micro-Macro Dynamics and Control Across Social-like Systems Using Equation Free Modeling

**Technical Area:**

***YFA 2017 Topic Area 6****: Characterizing Micro-Macro Dynamics in Social-like Systems*

**DARPA Mentor**

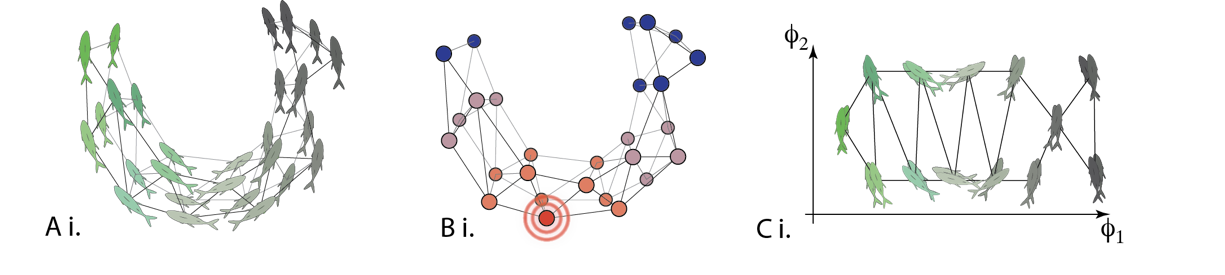
*Adam Russel*

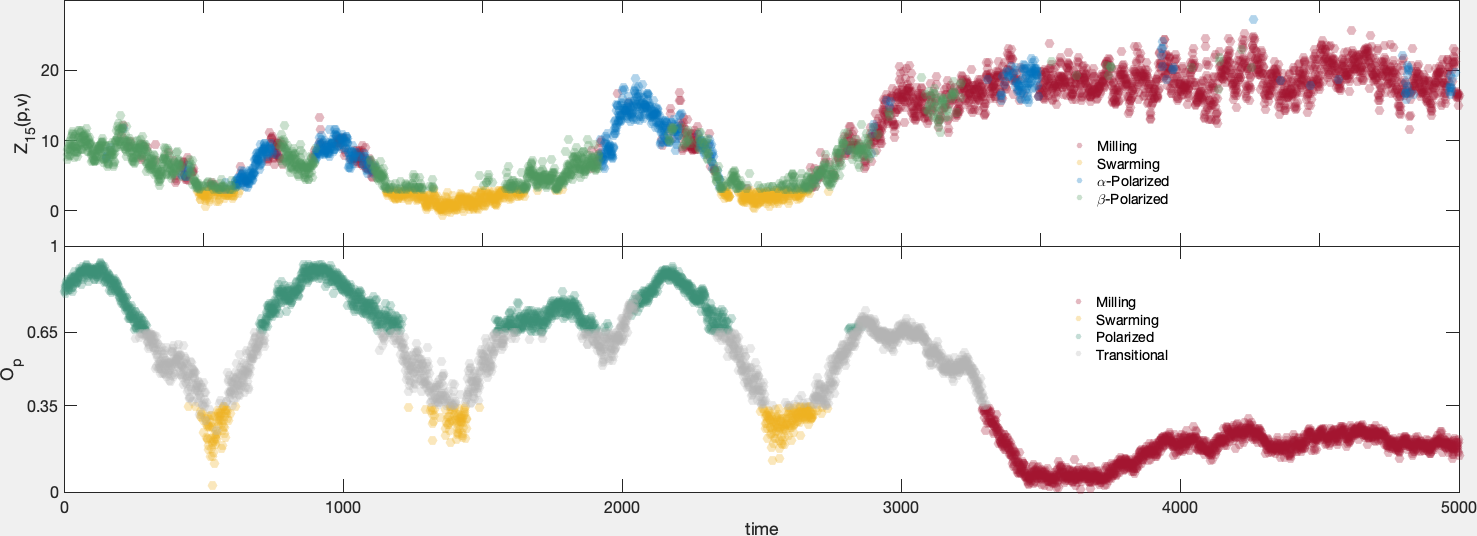
**I Executive Summary**

In the last quarter of the DARPA YFA project titled “Comparing Micro-Macro Dynamics and Control Across Social-like Systems Using Equation Free Modeling”, PI Watson and his team (post-doctoral research associates Drs. Zach Gelbaum and Mathew Titus) have written the following manuscripts that showcase new multiscale methods for analyzing complex social-like systems:

1. The application of Manifold Learning to the study collective behavior in nature. The study system is fish schooling, and a new technique that we have developed -- sequential diffusion maps – has been advanced to identify different forms of collective behavior exhibited by these fish (without any prior knowledge of the system), and to identify the few “influencers” in the school that create transitions in the collective behavior of the fish. This paper is in review at PLoS Computational Biology. This new computational approach is general, with utility to a range of complex systems. To explore its utility, we have applied it to financial market, housing market and oceanographic data, and results identify herd-like behavior preceding the 2008 financial crash, coherent geographic regions affected by the 2006/7 housing market bubble, and important modes of climate variability for the world’s oceans. These additional analyses are leading to three further papers on these topics.
2. The advancement and application of Spectral Graph Wavelets to the characterization and prediction of human migration from cell-phone data. The study system is Senegal in 2013, for which we have cellular text data between cell-towers. This work has revealed new insight about the cross-scale nature of human migration: mass human mobility starts at small spatial scales, and then cascades up to large spatial scales. This work has identified dominant modes and scale of various forms of human mobility (daily commuting, seasonal migration for agriculture, punctuated events such as holidays) and is in review (minor revisions) at the journal PeerJ Computer Science.
3. A new early-warning signal of critical transitions in (complex) dynamical systems has been created and termed Critical Speeding Up (CSU). CSU differs from classical early-warning signals of large changes in complex system based on “critical slowing down” and provides new predictive ability for systems undergoing large change. This new theory has been published in the journal Theoretical Ecology.

In sum, the goal of this DARPA YFA was to find general and recurring multiscale features of social-like systems. To do so new mathematical theory and computational tools were developed and applied to a broad range of complex systems that exhibit multiscale change. Even while at the end of this YFA project, publications and impact will continue to be produced and realized as this future work is conducted. For example, to extend the impact of this DARPA YFA research, PI Watson has secured funding from NASA for a 3-year project where these new computational techniques have successfully been used to predict instances of illegal activity (e.g. narcotics trafficking, illegal fishing) from spatial data on the location of vessels at sea.



******

***Figure 1****) Top - cartoon illustration showing the spatial location of fish schooling. Edges between individuals are defined using different metrics, for example the correlation in heading or even spatial proximity. B) With this metric between individuals, the graph Laplacian quantifies the diffusion of heat on the affinity graph, which can then be used to C) recast the system in a different (reduced) coordinate system, that allows for the identification of different modes of collective behavior. Bottom two panels – results from this analysis identify changes in collective behavior in fish schools over time.*

**II Detailed Research Summary**

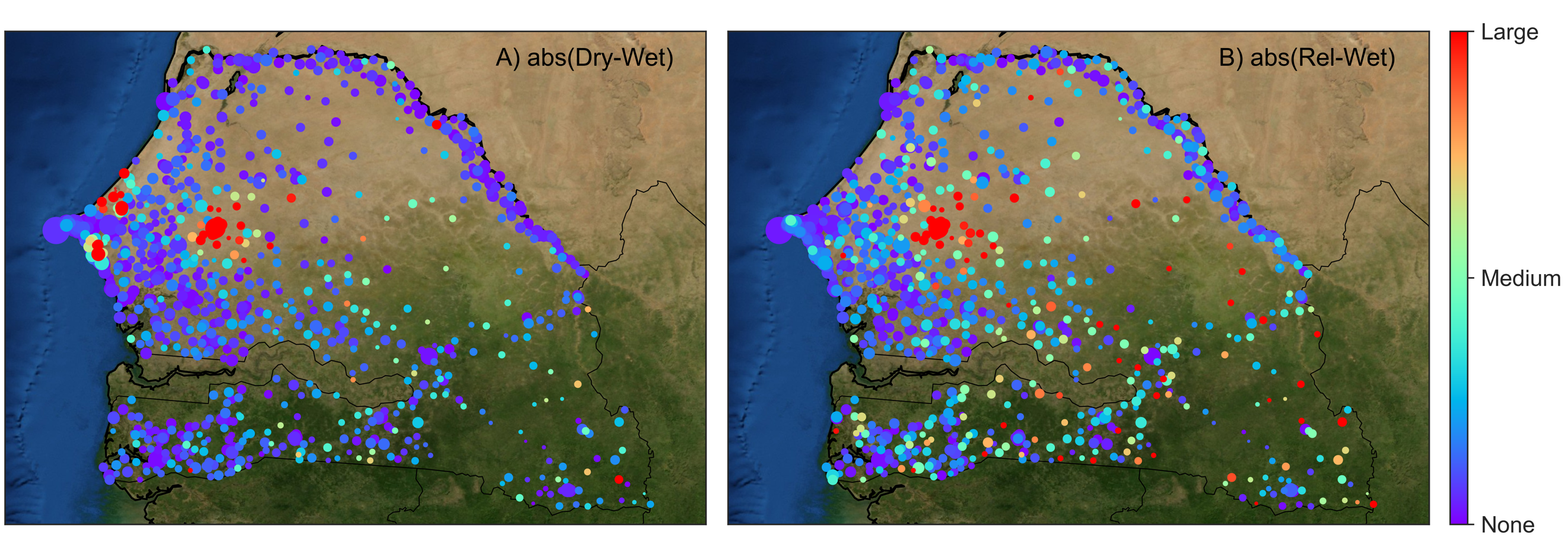
The YFA team at Oregon State is advancing a new methodology for measuring and detecting distinct macroscopic dynamical regimes in social-like systems, from observed microscopic data. Their work is also leading to new abilities to predict abrupt and large-scale shifts in macroscopic variables of interest, based on microscopic measurements. Their methods are based on Manifold Learning techniques; specifically, they model microscopic data as a series of mathematical manifolds evolving through time. This approach is distinct to most others for studying social-like systems in that changes in the geometry of a system is studied, which circumvents the need to develop dynamical equations. Hence this new approach is in the “equation-free” class of techniques and can be applied to systems where agents are not easily codified (i.e. when they are heterogeneous and strategic).

**Manifold Learning of Collective Behavior**

The new Manifold Learning methods from this YFA team is being applied to empirical fish-schooling data. These data describe the location of 300 golden shiner fish (small, about 1-inch long), confined to a shallow pool. As a consequence, these fish are essentially maintained in a 2D system. A video system records each fish’s individual location through time, and these fish exhibit three distinct forms of collective behavior: swarms (where the fish move randomly), polarized (where the fish all move linearly in the same direction) and milling (where the fish rotate as one school). The fish constantly transition between these states). Diffusion maps (a method from Manifold Learning) applied sequentially to these data can accurately identify these three different states (Fig. 1A), and furthermore identify individual fish whose anomalous behavior proceeds the state transition in collective behavior. This is in essence a cross-scale “micro-to-macro” analysis.

Sequential diffusion maps work as follows: given data on the position of fish in a school at a given time. a metric is put on the data, resulting in an “affinity matrix” that describes the similarity between agents at a given time. This could be the correlation in velocity for example. Then, the graph Laplacian is calculated, which describes the diffusion of heat on the graph and quantifies the geometry of the system. The resulting eigenvectors describe new “diffusion coordinates” that can be used to characterize a social-like system and anticipate change (see Figure 1 for an illustrative figure describing this process).

Doing this using different metrics of agent-agent similarity then allows us to build a **map alignment statistic** which quantifies how the different diffusion coordinates, produced from the different metrics, align. This map alignment statistic reveals the different modes of collective behavior exhibited by the fish, and also in other systems such as financial markets. The map alignment statistic also identifies anomalous individuals in a complex system, whose changing behavior can forewarn of system-wide change.

**Figure 2**. Maps of Senegal where markers and the location of cell-towers, color coded by the spectral wavelet function centered on Touba (large red dot in the middle of the country), a place of religious worship. Our analysis has identified two forms of human migration to/from Touba: A) seasonal migration to/from the coast, and B) punctuated migration to/from small towns in the interior associated with religious festivals.

**Dominant Modes of Human Mobility from Spectral Graph Wavelets**

In addition to sequential diffusion maps, wavelet signal processing on graphs has been explored as a viable manifold learning tool. The motivation is that diffusion maps produces information that is global in nature, and a computational method for producing information at a range of scales, from local to global (in the network of agents) would truly provide micro-to-macroscale information. Graph wavelets are designed precisely for this. To test their utility, we have applied them to cellular text data from Senegal in 2013 (see Fig. 2A). The goal is to identify and characterize the different modes of variability in human mobility/migration.

Spectral graph wavelets are analogues of classical wavelet analysis applied to complex networks.  They enable multi scale analysis of signals on complex networks, and also enable the study of the structure of complex networks themselves.  There are many possible constructions of wavelets, and our key contribution is a construction of a particular wavelet function based on the heat kernel. Similarly to diffusion maps, we start by creating an affinity matrix amongst agents comprising a system (in this case cellular towers). The graph Laplacian is calculated from this affinity matrix, to which a wavelet kernel is applied (i.e. the time derivative of the heat kernel). The resulting wavelet coefficients identify the dominant scales of variation of a signal across the network. Doing this sequentially through time allows us to track the dominant modes of variability in the signal, as the system changes.

**A close up of a map

Description automatically generated**

Fig. 3. A) The standard well potential diagram, where the ball identifies the state of the system, and the valleys of the landscape identify (multiple) basins of attraction. Here, the red line identifies the tipping point or separatrix delineating the basins of attraction. **b** When undergoing a bifurcation, a slowly changing parameter causes a shallowing of a given basin of attraction. This leads to increasing variance in measured variables and critical slowing down. **c** When a parameter change leads to a narrowing of the occupied basin of attraction, the variance in measured variables diminishes and there is critical speeding up. In some cases, the separatrix is moved left. In any case, well narrowing leads to an increased chance of stochastic regime shift.

With this computational tool have tracked how the structure of a complex network changes through time at multiple scales, and decompose dynamics taking place on the vertices of the network into dominant scales of variation over the network. Applied to the Senegal sms data, our results identify different modes of variability in sms texts between cellular towers (see Fig. 2). This is the first time these kinds of data have been used to do so, and the resulting characterization of human mobility reveals cross-scale and punctuated events, relating to religious holidays, seasonal migration for work and political elections.

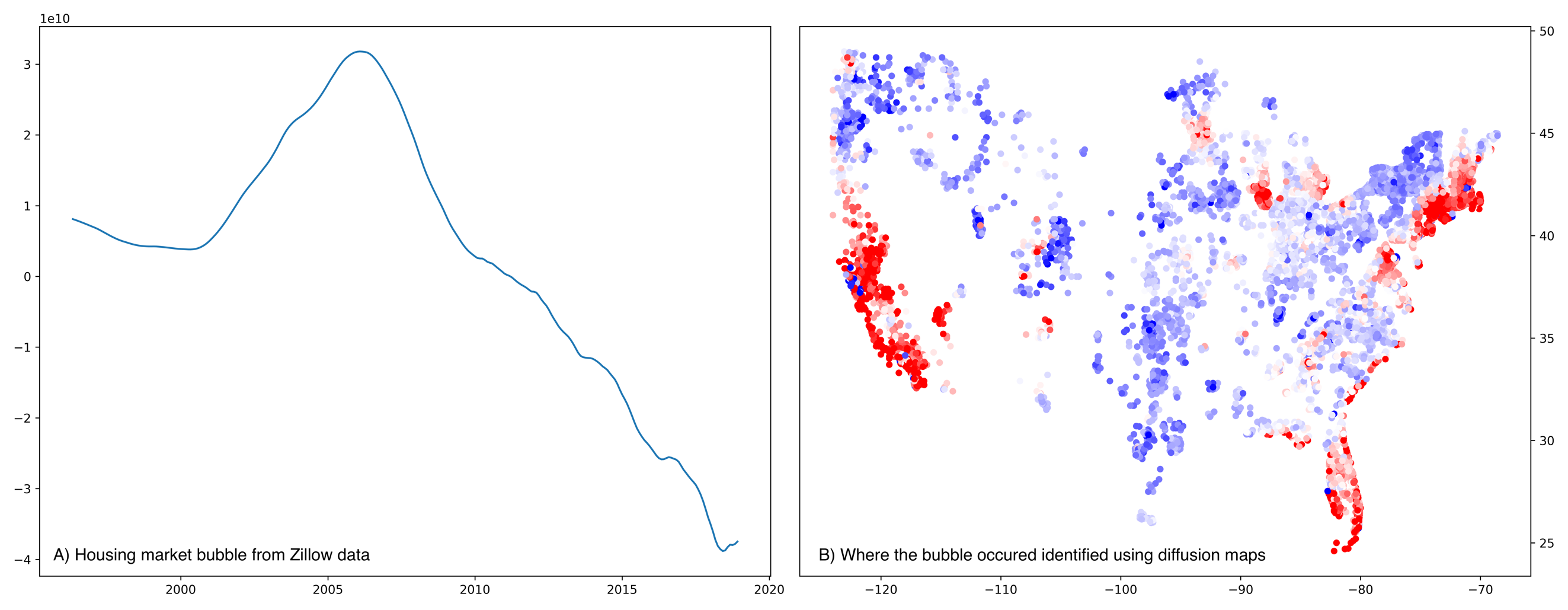
**Novel Early Warning Signals of Critical Transitions**

The use of critical slowing down as an early warning indicator for regime switching in observations from stochastic environments and noisy dynamical models has been widely studied and implemented in recent years (Fig. 3). Some systems, however, have been shown to avoid critical slowing down prior to a transition between equilibria. Possible explanations include non-smooth potential driving the dynamic or large perturbations driving the system out of the initial basin of attraction. To advance predictions of large change in complex systems, we have explored a phenomenon analogous to critical slowing down, where a change in a slow parameter leads to a high likelihood of a regime shift and creates signature warning signs in the statistics of the process’s sample paths. In short, if a basin of attraction is compressed under a parameter change then the potential well steepens, leading to a drop in the time series’ variance and autocorrelation; precisely the opposite warning signs exhibited by critical slowing down. This effect, which we call “critical speeding up,” is demonstrated using a simple ecological model exhibiting an Allee effect. The fact that both dropping and rising variance / autocorrelation can indicate imminent state change should underline the need for reliable modeling of any empirical system where one desires to forecast regime change

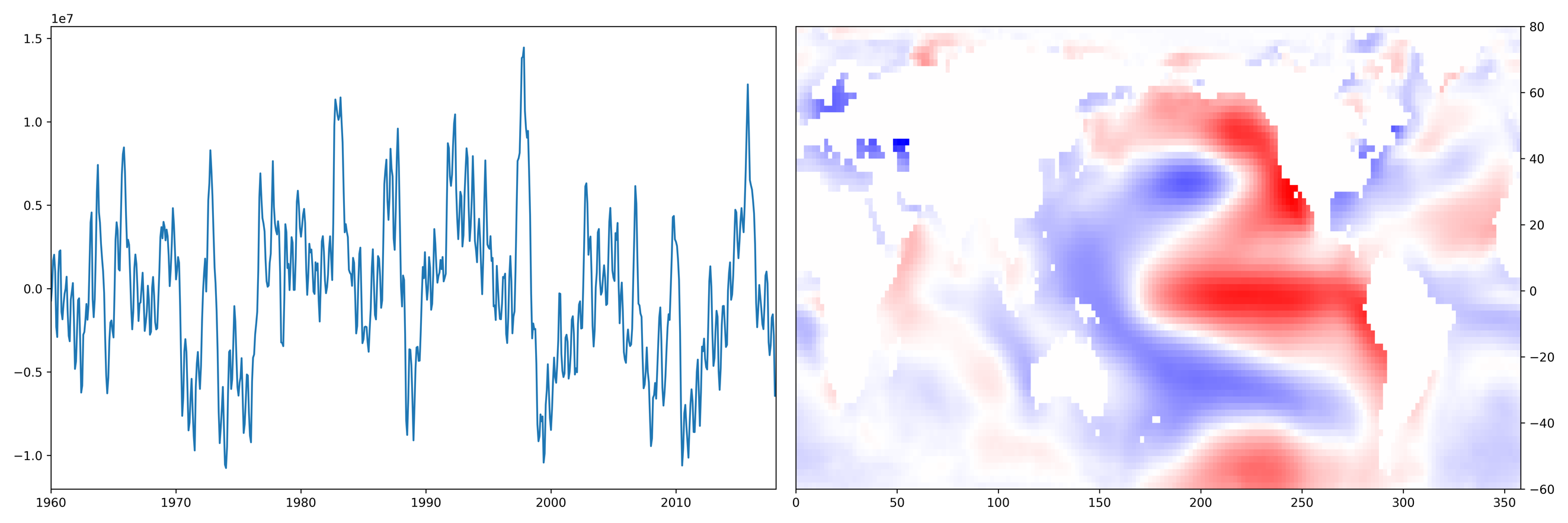
**Extended Analysis Using New Multiscale Tools**

**Leaders and Followers in Teams:** The application of spectral graph wavelets to the human mobility data has led to new mathematical theory that proves a relationship between dominant wavelet functions and “leadership centrality”, a relatively new and important graph theoretic metric for identifying influential nodes in a network. By couching leadership centrality in terms of wavelets, our new definition extends its utility to identify not only influencers, but also nodes that are susceptible to influence – the followers. We are applying this new theory to an additional fish schooling dataset that we have obtained. Here, certain fish have been trained to know where food is, and others are uninformed. We are testing this approach with these data to identify informed individuals/fish, with the goal of then applying this new method to sporting team data which is in hand, specifically soccer player location data. An extension of this YFA research is to develop new ways to quantify the “gel” amongst team mates in competitive sports, with application to military settings.

**The 2006/7 Housing Bubble:** The manifold learning approach based on diffusion maps, which we applied to characterize fish schooling behaviors, has been extended and applied to housing value data from Zillow. The mathematical extension was to allow diffusion maps to identify Empirical Orthogonal Functions (EOFs) from a multivariate time-series. EOFs have traditionally been identified using linear approaches like Principal Components Analysis, but the diffusion map approach is non-linear and is better suited for complex social-like systems. Results from this analysis have identified geographically coherent structures that showed bubble like features in their housing values over time (see Fig. 3). The goal of this work is to identify multi-scale contagion in the 2006/7 housing bubble.

**Figure 3**. Example result from the application of diffusion maps to Zillow housing data: A) diffusion map EOFs have an associated time-series, and in this case we clearly identify the 2006/7 housing bubble as a peak in the timeseries; B) this bubble is associated with geographically coherent locations in the US. These places identify at-risk neighborhoods where sub-prime mortgages and defaults were common.

**Dominant Modes of Oceanographic Variability:** The diffusion map approach to EOF analysis was also applied to global sea surface data. While not a social-like system, global sea surface data is rich with dynamical features and we wanted to push our methods to see how much information they could provide about a fluid system. Diffusion map EOFs successfully managed to identify various modes of oceanographic variability such as El Nino, the North Atlantic Oscillation and other important modes. This is a new approach to geophysical data science, based on Manifold Learning.



**Figure 4**. Example result from the application of diffusion maps to sea surface temperature data: Left) diffusion map EOFs have an associated time-series, and in this case we identify El Nino / La Nina oscillations through time; Right) these oceanographic changes are associated with geographically coherent features in the Pacific (i.e. a warm pool in the eastern pacific).

**III Management Summary**

The YFA team was comprised of PI Watson and post-docs Zach Gelbaum and Mathew Titus. PI Watson continues as faculty at Oregon State University and is now PI on a new DARPA Seedling project. Gelbaum and Titus have transitioned to a new job and in particular, Titus and Watson continue to work together under a data-science company that they formed (The Prediction Lab LLC). Collaborations with the Princeton University mentors and collaborators continues to be a source of help as the final   
YFA products are created, specifically through Dr. George Hagstrom and Prof. Simon Levin.

**IV Financial Status Report**

