

Long-range forecasts using data clustering and information theoryDimitrios Giannakis[†]; Andrew Majda[†] New York University, USALeading author: dimitris@cims.nyu.edu

Even though forecasting the weather beyond about two weeks is not possible, certain climate processes (involving, e.g., the large-scale circulation in the Earth's oceans) are predictable up to a decade in advance. These so-called climate regimes can influence regions as large as the West coast of North America over several years, and therefore developing models to predict them is a problem of wide practical impact. An additional central issue is to quantify objectively the errors and biases that are invariably associated with these models. We present methods based on data clustering and information theory to build and assess probabilistic models for long-range regime forecasts. With reference to a simple ocean simulation mimicking the Gulf Stream in the Atlantic (or the Kuroshio current in the North Pacific), we demonstrate that details of the initial state are not needed in order to make skillful long-range predictions, provided that an appropriate coarse-grained partitioning of the set of possible initial conditions is employed. Here, that partitioning is constructed empirically using running-average coarse graining and K-means clustering of observed data, and optimized by means of relative-entropy measures. We apply the same tools in a related formalism for quantifying errors in imperfect climate models. Together, these techniques provide a framework for measuring predictive skill and model error in a manner that is invariant with respect to general transformations of the prediction observables.