Data-driven stochastic model discovery of organized clouds dynamics

Mickael D. Chekroun\(^1\), Tom Dror\(^2\), Orit Altaratz\(^2\), and Ilan Koren\(^2\)

\(^1\)UCLA, Department of Atmospheric and Oceanic Sciences, United States of America (mchekroun@atmos.ucla.edu)
\(^2\)Weizmann Institute of Science, Rehovot, Israel (michael-david.chekroun@weizmann.ac.il)

The discovery of dynamical equations governing time-evolving observations issued from a complex dynamical system requires a statistical formulation, since information concerning neglected variables or unobserved degrees of freedom is necessarily incomplete. At the same time, an equation that is closed within a small number of observables is often obtained only by approximations. Thus, the relevance of approximations must be understood before any attempt to derive a closed set of equations. This is where closure formalisms are of usefulness and the corresponding mathematical structures serve as a guide for knowing what to approximate. Many such formalisms are available from turbulence theory, quantum field theories, to statistical physics.

Observables of interest often include response functions, spectra of fluctuations, or low-order moments, etc. These quantities correspond to moments of the full probability density function (PDF), the mother of all system’s statistics but itself beyond the reach of standard closure theories, except in special cases. Yet, to have, for a given choice of observables, a (good) class of closure models able to produce out-of-sample reliable occurrences, is of prime importance. When derived on a firm basis, such closure models may indeed allow for analyzing in greater details certain features of a given phenomenon for which available data are limited, by e.g. drawing a large ensemble of statistical emulations of this phenomenon, from the closure model.

This is the goal that will be pursued here for a special but common class of clouds, namely continental shallow cumulus (Cu) that can be found from low to mid/high latitudes, across a wide range of scales, and that play a growing role in the Earth’s radiative budget. These clouds typically organize through a variety of patterns such as cloud streets, clusters, or mesoscale arcs. Based on observables suitably extracted from high-resolution satellite observations, it will be shown that the efficient learning of hidden, stochastic, variables along with their interaction laws with the observed variables is key for the derivation of relevant stochastic data-driven models. To do so, our approach will rely on the Mori-Zwanzig closure theory to guide the search of the constitutive elements, on one hand, while their learning will exploit recent advances in data-driven stochastic modeling techniques, on the other.

As a byproduct, dynamical equations involving a few variables are learned from high-resolution satellite observations of continental shallow Cu. These equations will be shown to take the form of...
differential equations that include lagged effects, and are driven by a spatially correlated white noise. It will be finally shown that the combined effects of these terms allow to generate easily statistical ensembles of shallow Cu that exhibit a wide range of spatio-temporal variability while displaying consistency with the shallow Cu's organizational and multiscale features, from observations. Based on such large ensembles, new physical insights are attainable and their interpretation will be discussed. This work is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program [Grant Agreement No. 810370].