Machine Learning of Committor Functions for Predicting High Impact Climate Events

D. Lucente[1], S. Duffner[2], J. Rolland[1], C. Herbert[1] & F. Bouchet[1]

[1] Laboratoire de Physique, Ens de Lyon, Université Claude Bernard, Université de Lyon, CNRS, F-69342 Lyon, France
[2] Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, Villeurbanne, France
dario.lucente@ens-lyon.fr

The low frequency modes of variability of the climate system, for instance ENSO, have a huge impact on nature and human societies, through their local or global signatures. Rare events, such as heat waves, floods, or hurricanes, may also have a huge impact. Predicting the occurrence of such events is thus a major challenge . Because the dynamics of the climate system is chaotic, one usually distinguishes between time scales much shorter than a Lyapunov time for which a deterministic weather forecast is relevant, and time scales much longer than a mixing times beyond which any deterministic forecast is irrelevant and only climate averaged or probabilistic quantities can be predicted. However, for most applications cited above, the largest interest is for intermediate time scales for which some information, more precise than the climate averages, might be predicted, but for which a deterministic forecast is not relevant. We call this range of time scales *the predictability margin*. As a paradigmatic example, we study in this work the probability that El-Niño might occur following year. Another example could be : What is the probability of a heat wave of a given amplitude to happen next summer, given the state of the atmosphere, ocean, and soil moisture, in Spring ? We stress in this talk that the prediction problem at the predictability margin is of a probabilistic nature. Indeed, such time scales might typically be of the order of the Lyapunov time scale or larger, where errors on the initial condition and model errors limit our ability to compute deterministically the evolution. However, we stress that the Lyapunov time scale, a global quantity, is clearly not the relevant dynamical quantity for this predictability problem. By contrast, at the predictability margin, the predictability clearly depends on the current state of the system. What is then the relevant mathematical concept ? Our first aim is to introduce in the field of climate the notion of the committor function. A committor function is the probability that an event will occur or not in the future, as a function of the current state of the system. For the El-Niño case, this committor function will be the probability that an observable $\mathcal{O}$ of the system reaches a given threshold within a time T. The first result of this talk is to demonstrate, using the committor function, that a predictability margin exists for El-Niño. This demonstration is performed within the Jin and Timmermann model, a low dimensional model proposed to explain the decadal amplitude changes of El-Niño. From the computed committor function for the Jin and Timmerman model, we obtain the second main result. This result is the characterisation of regions of the phase space with qualitatively different predictability properties. Firstly, regions of perfect predictability, where the event will occur with probability 0 or 1, respectively. Secondly, regions with good predictability properties where a value of the probability $0 < q < 1$ can clearly be predicted with very mild dependance with respect to initial condition. We call this area the *probabilistically predictable region*. Thirdly, regions which are unpredictable in practice, because the strong dependance with respect to the initial condition prevents any practical prediction, either deterministic or probabilistic. The existence of such features, and especially the new and most interesting *probabilistically predictable region*, should be generic for most prediction problems in climate dynamics. We will explain that committor functions solve Dirichlet problems. However such partial differential equations are extremely difficult to solve especially for high-dimensional systems. Could we compute it directly from data ? There is currently a growing interest to estimate relevant dynamical quantities directly from available data, for instance using machine learning techniques. The second aim of this talk is to propose two different machine learning approaches to compute committor functions. We conclude by discussing the feasibility of the computation of a committor function using machine learning techniques for the Jin and Timmerman model, and for more complex data sets related to other climate applications.