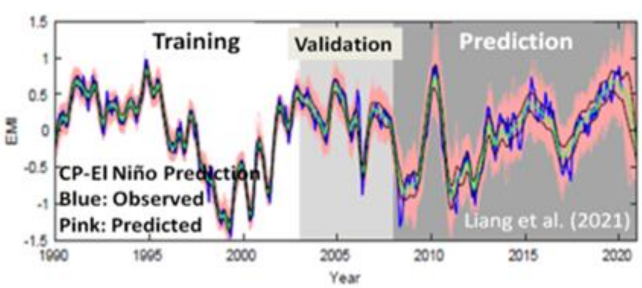
Causality as a real physical notion ab initio, and causality analysis in climate and environmental sciences

X. San Liang

Fudan University, Shanghai, China, [2]The AI Group, Southern Marine Laboratory, Zhuhai, China

Causality analysis is an important and old problem lying at the heart of scientific research. Causality analysis based on data, however, is a relatively recent development. Traditionally causal inference has been classified as a field in statistics. Motivated by the predictability problem in physical science, it is found that causality in terms of information flow/transfer is actually a real notion in physics that can be derived ab initio, rather than axiomatically proposed as an ansatz, and, moreover, can be quantified. A comprehensive study with generic systems (both deterministic and stochastic) has just been fulfilled, with explicit formulas attained in closed form (Liang, 2016). These formulas are invariant upon nonlinear coordinate transformation, indicating that the so-obtained information flow should be an intrinsic physical property. The principle of nil causality that reads, an event is not causal to another if the evolution of the latter is independent of the former, which all formalisms seek to verify in their respective applications, turns out to be a proven theorem here. In the linear limit, its maximum likelihood estimator is concise in form, involving only the commonly used statistics, i.e., sample covariances. An immediate corollary is that causation implies correlation, but the converse does not hold, expressing the long-standing philosophical debate ever since Berkeley (1710) in a transparent mathematical expression. The above rigorous formalism has been validated with benchmark systems like baker transformation, Hénon map, stochastic gradient system, and with causal networks in extreme situations such as those buried in heavy noises and those with nodes almost synchronized (e.g., Liang, 2021), to name a few. They have also been applied to real world problems in the diverse disciplines such as climate science, dynamic meteorology, turbulence, neuroscience, financial economics, quantum mechanics, etc., with interesting new findings. For example, Stips et al. (216) found that, while CO2 emission does drive the recent global warming, on a paleoclimate scale, it is global warming that drives the CO2 emission; PNA, a teleconnection pattern related to the inclement weather in North America, may trace a part of its origin to a rather limited local marginal sea far away in Asia. Besides, with the above causality analysis, pollution sourcing (particularly PM2.5) can be conducted in an effective way via causal graph reconstruction. If time permits, I will also present an ongoing application to the development of causal AI algorithms to overcome the interpretability crisis, and a recent remarkable exercise with such an algorithm in the forecasting of El Niño Modoki, a climate mode linked to hazards in far-flung regions of the globe (cf. the figure).



El Niño prediction has become a benchmark problem for the testing of machine learning algorithms. The present wisdom is that El Niño may be predicted at a lead time of **1-2 years.** Shown here are 1000 predictions (pink) of the El Niño Modoki index (EMI) using a causal AI algorithm as mentioned in the abstract. Overlaid are the observed EMI (blue), the mean of the realizations (cyan). The light shading marks the period for validation, while the darker shading marks the prediction period---It is more than **10 years** long (from Liang et al., 2021).

References

[1] X.S. Liang, Unraveling the cause-effect relation between time series. Phy. Rev. E, 90, 052150 (2014).
[2] X.S. Liang, Information flow and causality as rigorous notions ab initio. Phys. Rev. E, 94, 052201 (2016).
[3] X.S. Liang, Normalized multivariate time series causality analysis and causal graph reconstruction. Entropy, 23, 679 (2021).
[4] X.S. Liang et al., El Niño Modoki can be mostly predicted more than 10 years ahead of time. Nature Sci. Rep. 11:17860 (2021).