Jaynesian Analysis of Environmental Chemistry: Systems Model Component Integration via the Arctic Aquatic Carbon Cycle

Authors: S. Elliott, A. Piliouras, B. Nadiga, F. Hoffman, J. Kumar, J. Tang, U. Mishra at the national labs, plus students and the Soil-Organics-Sea working group (SOS)

Focal Areas: All three are touched upon from the OBER call page 4

We will consider formal logical inference and entropic information analysis for sparse but intensive chemical measurements, as optimal drivers for managing process complexity. As a centerpiece for the discussion we take soil-to-sea organic structural evolution within the Arctic hydro-geochemosphere, assessing detrital negentropy then algorithmic greed to generate ranked, reduced mechanisms for system integration.

Challenge:

Aquatic carbon chemistry connects northern continental ecosystems, soils, rivers, deltas, plumes and regional marine water properties extending to the sea ice edge -and culminating in biophysical effects spanning most of the hyper-sensitive upper Arctic Ocean. This is the most intense regional hydro-geocycling on the planet, with complex bioreactivity controlling coastal influences ranging from vertical energy redistribution through nutrient delivery to chelation. Permafrost degradation and terrestrial ecotone shifts promise to alter aquatic organic structures dramatically, with critical impacts on the ice-free marine environment. Yet in current generation regional-global systems models, links in the soil-to-sea chain remain separated, and the chemistry is usually treated even internally via the generic lump sum quantity DOC (Dissolved Organic Carbon).

The authors of this white paper are breaking down such conceptual barriers by considering detailed transport of molecular bonding along detrital reactivity sequences. We are testing a daisy chain of organo-kinetic fuzzy solvers arrayed along the total Arctic hydro-course. Mechanisms cross all significant systems model boundaries from headwaters on, and they therefore conjoin what would otherwise be independent biogeochemical zones. Our team is already providing manageable reduced schemes for assessing and ranking oceanographic ecosystem components and processes. Results indicate strong influence of carbon bonding structure on the transfer depth of coastal photosynthetic radiation, vertical shifts in heat energy deposition, control on ecodynamics through amino nitrogen injection, and slicking of peripheral seas with suppression of momentum flow. Examples of specific functions under scrutiny vary from enzymatic proteins to diagenetically intermediate hetero-polycondensates to humics.

Rationale

DOI: https://doi.org/10.2172/1769731
DOE BER, AI4ESP 2021 #AI4ESP1039 (https://www.ai4esp.org)

Our ultimate plan is to spring from the recent synthesis of E.T. Jaynes marrying inferential statistics with entropic information, to derive maximum benefit from current soil-sea organic measurements -e.g. those associated with NGEE. We will wrap "infer-matic" logic around a particular class of chemical data, one that is readily available but normally underexploited. Wet analytical, local instrumental or logistically intensive observations often prove difficult to obtain and interpret, and this is especially true for chained carbon in the high Arctic. The Jaynes approach blends historical concepts from Laplace through Boltzman to Shannon and descendants, as documented in several monographs of our reference list. Specifically we will discuss synthesis of data distributions through 1) a formal basis in the calculus of probabilities, 2) combinatoric path analysis as a greedy Gantt pictogram of modeling trends, 3) analogies with statistical mechanics and the convenience of the Stirling approximation, 4) Tesler, Teller and related modern tests of chemical complexity, 5) unit hypercube definitions of entropic opportunity and metered priority, 6) definition of biogeochemical zones in support of process exploration, 7) rank ordering of schemes on tailored chemo-receptor zones.

These concepts have strong parallels in fields ranging from economics to neuroscience, but to our knowledge they have not yet been applied to processes in environmental modeling systems. In a Jaynesian "infer-matic" framework, the Tesler Theorem of Artificial Intelligence defines chemistry as a logical search-simplification game and machine learning as the mode of grid (zone) redefinition. Our "hypothesis" (in a meta-sense) is that interpretation-resistant observations resulting from extreme field conditions, sample delicacy, transport to lab, wet manipulation etc. create an activation barrier for model incorporation, but it can and must be ascended. Unique process information is stored by intensive analysis points in a way that cannot otherwise be matched. Inferential cycling through (neg-) entropic information can both quantitatively and unambiguously identify primary processes and relationships, while enabling viable reduced mechanisms.

The concepts concern efficiencies which may be achieved through informed scanning of an environmental chemical knowledge base. Our lexicon is thus a key to time-resource savings. For example, we will discuss the Cromwell curse while simultaneously drawing on ETJ statistical mechanics. We aim initially for a reasoned, nonbiased and optimal program for layering organo-chemical complexity into boreal climate models, recognizing that this framework is applicable beyond the Earth System as well, e.g. given industrial, defense security and planetary impact applications.

In summary, our plan is to harness the fundamentals of formal logic, inference, and computer science, to identify for the boreal aquatic regime the controlling organic chemistry mechanisms from temporal and spatial data sets. Our motivation is to understand upcoming Arctic coastal response as the northern water-carbon interface comes undone over the next century. Additionally, our methods carry complete generality, and will benefit systems simulation across the board.

Relevance

The chemistry scanning recipe outlined here may prove a strong match for DOE interests at several levels. For example, our plan includes next-generation biogeochemistry, permafrost destabilization, coastal marine science, Arctic climate sensitivity and model component integration. Perhaps most timely is our emphasis on in situ Goldilocks data - those which are rare while high in value and tend to be expensive operationally then go underutilized. We will consider the (negative or Brillouin) entropy inherent to wet and field laboratory endeavor, to prioritize and characterize opportune processes and drive model growth, supplementing remotely sensed "big data" such as vegetation indices or ocean color maps. We expect that Jaynesian inference can unlock the potential.

The method promises to be highly complementary of cutting edge Earth System modeling in that new schemes are prescribed then rank-ordered precisely for underserved regions, which emerge from identification of biogeochemical zones. Likewise, the method will be applicable for generalized model packages that must also deal with chemical complexity. Flexibility required to navigate arbitrary reactive landscapes is part of the design. Our concepts further supplement existing DOE benchmarking packages, such as ILAMB (International Land Model Benchmarking). Though target diagrams and intercomparisons may also be arrayed on the unit interval, they tend to point toward our highlighted negentropics only indirectly, as for example via grey area "shame" factors.

Definitions:

Although Jaynes' logical and mathematical forms predate modern computational workflows, they are universal and blend seamlessly into contemporary worlds of artificial intelligence and machine learning. For example, to play a Turing game, rules must first be acknowledged, but to date the community approach to complex chemistry has been disjoint. We can establish a requisite "gaming board" through close attention to probability theory while also calling upon the power of machine learning. Specifically, our group will survey the potential to formalize then automate chemical intuition in an Arctic aquatic context. The goal will be to explore that which is most critical to changing climate, enabling selection of next generation processes and in particular those which knit components together (here involving soil through river course to sea). We target the most sophisticated experimental analyses available for organics. Refinement through Laplacian cycling takes place on a learned chemo-cartography distilled from standard geophysical dynamics. Pattern searches may be unsupervised and draw liberally on traditional big-data sets. We aim to improve rigor in the stepwise application of environmental chemical inference. Our planetary-scale process game dwarfs chess, Go or Jeopardy in its intricacy, but it cannot be avoided.

Addenda:

Current membership of our informal SOS working group: A. Aguilar-Islas (UA Fairbanks), R. Amon (TAMU Galveston), V. Alexeev (IARC), M. Debolskiy (Moscow), K. Frey (Clark U.), G. Gibson (IARC), Q. Kang (LANL), J. Clement-Kinney (NPS Monterey), J. Kumar (ORNL), Y. Liu (TAMU), P. Matrai (Bigelow Lab Maine), U. Mishra (Sandia and recently ANL), J. McClelland (UT Port Aransas), B. Nadiga (LANL), A. Piliouras (LANL), M. Rawlins (UMass), R. Spencer (Florida State), J. Tang (LBL), H. Viswanathan (LANL)

Hoping soon to recruit: S. Tank (U. Alberta), J. Vonk (Utrecht), K. Wickland (USGS)

Sample References:

AGRO, Arctic Great Rivers Observatory Metadata, NSF 2018

Batiuk, Chesapeake Bay Submerged Aquatic Vegetation and Water Quality, EPA 2000

Bevensee, Maximum Entropy Solutions to Scientific Problems, Prentice Hall 1993

Bianchi et al., Biogeochemical Dynamics at Coastal Interfaces, Academic 2014

Brillouin, Science and Information Theory, Academic 1956

Denbigh and Denbigh, Entropy in Relation to Incomplete Knowledge, Cambridge 1985

Droppo, Flocculation in Natural and Engineered Aquatic Systems, CRC 2005

Eckschlager and Danzer, Information Theory in Analytical Chemistry, Wiley 1994

Hald, History of Mathematical Statistics, Wiley 1998

Hansell and Carlson, Marine Dissolved Organic Matter, Academic 2002

Jaynes, Probability Theory: The Logic of Science 2003

Kosko, Fuzzy Engineering, Prentice Hall, 1997

Leifer, Environmental Aquatic Photochemistry, ACS 1988

Lindsey, Bayesian Statistics, SIAM 1972

Stein and MacDonald, Organic Carbon Cycle in the Arctic Ocean, Springer 2004