

# Forecasting long-term ecological dynamics using open paleodata

Jason S. McLachlan<sup>1</sup> and the PalEON Project<sup>2</sup>

As humans alter the environment in unprecedented ways, forecasts of the future state of ecosystems become increasingly important. Good forecasts require skilled models, and paleoecological data have played an important role validating retrospective model hindcasts. Modern analytical approaches, like data assimilation, now allow paleodata to be explicitly incorporated into forward-looking model forecasts. In particular, paleodata can provide unique empirical constraints on forecasts of slow or infrequent events that are difficult to constrain with more recent instrumental measurements.

“Forecasting” has a specific meaning here, known from meteorology (Dietze 2017). An ecological forecast is a set of quantitative predictions about the most-likely future state (or reconstructed hindcast) of an ecosystem. A forecast is comprised of both models and data, each of which is incomplete and flawed: Models are simplified and imperfect representations of reality, and paleodata are noisy, geographically sparse, usually indirect, measurements of past ecosystems. Forecasting estimates the most-likely set of predictions of ecosystem state by weighting an ensemble of model predictions by the likelihood that they match statistical estimates of empirical data (Fig. 1A).

Ecological models can be informed by paleodata via initial conditions, drivers, state variables, and parameters, each of which helps improve scientific inference (Fig 1B).

**Initial conditions** can have persistent impacts on ecosystem state in both models and in nature (Turney et al. 2016). Paleoecological data can thus help ensure that model runs do not entrain the consequences of flawed initialization, e.g. by initializing from well-calibrated empirical estimates of historical vegetation (Paciorek et al 2016).

**Drivers** of ecosystem models include reconstructions of climate and other external forces driving ecological processes. For retrospective studies, empirically estimated drivers (Tipton et al. 2016) can be assimilated into climate models using data-assimilation approaches similar to those advocated in Hakim et al. (this issue).

**State variables** describe the state of the ecosystem being modeled over time. Plant biomass, for instance, is a state variable whose long-term dynamics can be modeled using paleoecological observations (Fig. 1A).

**Model parameters**, like the growth rate ( $r$ ) in a population growth model, establish links among variables. Paleodata can validate predictions of long-term ecosystem dynamics based on a particular model parameterization,

or they can identify the best among a set of competing parameterizations (Fig. 1B)

In data-model assimilation, discrepancies between model predictions and paleo-observations are resolved by considering their respective uncertainties; highly certain observations will exert a correspondingly stronger constraint on state variables or parameters. Hence, accurate representation of uncertainty is paramount. In Figure 1A, data from a fossil-pollen network, calibrated against vegetation survey data, produce a statistical reconstruction of changing plant biomass, accounting for uncertainty in pollen counts, taphonomic processes, etc. (Dawson et al. 2016). The mechanistic linkages between biomass and soil carbon in an ecosystem model then allow the empirically constrained reconstruction of biomass to improve estimates of soil carbon, an unobserved state variable. By narrowing uncertainty about long-term ecosystem dynamics in the past, this approach improves the model generally and thereby reduces uncertainty in forecasts of future ecosystem dynamics.

The suite of approaches to paleodata-model fusion outlined here pose opportunities and challenges for the producers and synthesizers of open data. Win-win opportunities emerge from the iterative coupling between models and data (Dietze 2017), for example,

by motivating new data campaigns to meet model demands. To capitalize on such opportunities, data stewards should work with statisticians and modelers to ensure that data are useable: For instance, when derived quantities, say temperature reconstructions, are archived, the raw data underlying them should also be archived, along with the code underlying all analyses. The rewards for this inconvenience will be new collaborations and increased predictive power!

## ACKNOWLEDGEMENTS

This work is funded by the NSF Macrosystem Biology Program (EF 1241874). The ideas presented here derive from seven years of conversation among the 100+ members of the PalEON Project.

## AFFILIATIONS

<sup>1</sup>Department of Biological Sciences, University of Notre Dame, USA

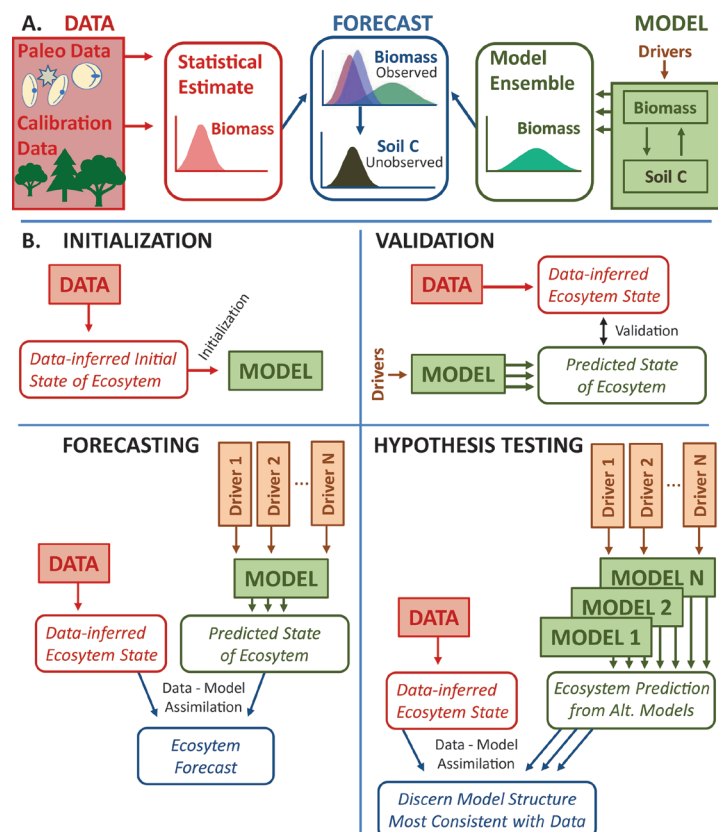
<sup>2</sup>sites.nd.edu/paleonproject

## CONTACT

Jason S. McLachlan: jmclachl@nd.edu

## REFERENCES

- Dawson A et al. (2016) *Quat Sci Rev* 137: 156-175  
 Dietze MC (2017) *Ecological Forecasting*, Princeton U. Press, 228 pp  
 Turney CSM et al. (2016) *PAGES Mag* 24: 3  
 Paciorek CJ et al. (2016) *PLoS ONE* 11: e0150087  
 Tipton J et al. (2016) *Environmetrics* 27: 42-54



**Figure 1:** (A) Assimilating paleoecological estimates of aboveground plant biomass to forecast unobserved soil carbon. (B) Experimental designs for integrating paleoecological datasets with ecosystem models.