We will consider the relationship between data and models for a specific application to interdisciplinary studies of aquatic ecosystems.

At the same time, I will focus on the need to understand the past in order to correctly define the models we use and the data we collect for studying the present and preparing for the future. This is an essential topic for any course on interdisciplinary ecological modeling.

- “The addition of a deep historical dimension to analyze and interpret ecological problems requires that we sacrifice some of the apparent precision and analytical elegance prized by ecologists. (…) Clearly, we cannot generate realistic null hypotheses about the composition and dynamics of ecosystems from our understanding of the present alone, since all ecosystems have almost certainly changed due to both human and natural environmental factors.” (Jackson et al., 2001, Science, 293, p.630)

We usually don’t know the history of the ecosystem we are dealing with, and this is one of the reasons to consider natural systems as “open”, which makes all their models “unverifiable” (see Oreskes et al., 1994, Science, 263, p.641).

- Models cannot be verified (= proved to be true) because (a) they require input parameters that are incompletely known, (b) the scale at which they are parameterized is often different from the scale at which they are applied, (c) they include inferences and assumptions on what is important and what is negligible, (d) measurements inevitably include errors, (e) different models can produce similar results.

- Models can only be validated (= shown to be effective and legitimate) with respect to their internal logic and procedures, but not with regard to their results because of all the reasons that make them unverifiable.

- Models can be tested, and eventually confirmed, by comparing their predictions with observational data, although this confirmation always relies on a majority (but not all) of the data.

- Models are useful to (a) summarize existing knowledge, (b) corroborate existing hypotheses and/or formulate new ones, (c) perform sensitivity analysis to determine research priorities and develop different scenarios.

- “The primary value of models is heuristic” (Oreskes et al., 1994, Science, 263, p. 644), which means they can stimulate interest and further investigation. Another definition of heuristic is “a plausible or reasonable approach that has often proved to be useful, a rule of thumb.” (Starfield et al., 1994, as cited by Nicolson et al., 2002, Ecosystems, 5, p.377).

Model confirmation can be obtained using a number of approaches, which can be categorized (in decreasing order of scientific accuracy) as follows:

- Experiments (controlled conditions; random assignment of treatments to subjects).
- Observational studies (conditions are not entirely controlled; assignment of treatments to subjects is not random).
- Iterative studies (no treatments are possible, but models are constantly refined: forecasting, reconstructions, opinion polls).
- Monitoring studies (surveys, inventories, automated data recording).
No matter how accurately we use the terms verification, validation, calibration, and confirmation of models, we still can obtain a more reliable representation of ecosystems if we incorporate knowledge on their past. This is analogous to the common practice in medicine of using the anamnesis (= medical history of a patient) as an essential element for any diagnosis.

—“Historical perspectives increase our understanding of the dynamic nature of landscapes and provide a frame of reference for assessing modern patterns and processes.” (Swetnam et al., 1999, Ecological Applications, 9, p.1189).

Let’s then consider what I’d like to call the “paleo conundrum”:

—Having retrospective information on the history of an ecosystem allows us to better understand its present conditions (see Swetnam et al., 1999, for terrestrial ecosystems, and Jackson et al., 2001, for marine ones).
  - In particular, recent observations cannot represent the full range of variability of natural systems, so that “proxy” records of climate (such as tree rings on land) are used to extend the instrumental record and obtain a more comprehensive picture of situations in which the system can be found.
  - A good example is provided by the Colorado River Compact, which was signed in 1922 to allocate Colorado river water among western states (and Mexico) using streamflow data starting in 1905, a period that has been characterized as unusually wet by comparison with both the instrumental record to date as well as proxy records for the past few centuries.
  - A model is used to link the proxy with the instrumental data, and this model is used to extrapolate the proxy climate information into the past.
  - A major assumption (and guiding principle) in obtaining retrospective information on the past history of ecosystems is provided by the uniformitarian principle: “the present is the key to the past”, originally stated by James Hutton in 1785.
  - If natural systems may change in unanticipated ways (and reducing uncertainty on this issue is the very reason to develop longer histories), then the uniformitarian principle may not hold (but this is one of the arguments used by creationists to dismiss evolution!)
  - As in other iterative studies (forecasting methods, opinion polls), we are constantly refining what we know about the past.
  - However, in those other iterative studies (forecasting, polls) there is normally an outcome that can be used to determine the model skill (i.e., the actual weather or stock market price or election result that was predicted by the models), whereas in retrospective studies we cannot know exactly what happened, and confidence in the results is increased only by comparing multiple proxies that corroborate one another.

A good example of this conundrum is the use of tree-ring records in hydrology (sometime called “dendrohydrology”), such as the work by Meko and Woodhouse, 2005, Journal of Hydrology, 308, pp. 196–213. These studies employ regression techniques to extend instrumental records of streamflow, usually at seasonal to annual time steps over several centuries. Having these long histories allows for a more comprehensive definition of the hydrological properties of watersheds and aquatic ecosystems. An advantage of the tree-ring approach over stochastic simulation is that one can develop dry and wet periods with characteristics outside the bounds of our instrumental observations.

—However, there is usually no direct physical connection between growth of sampled trees
and streamflow. Thus, any statistical relationship is based on connections that wood accumulation and runoff have with climate. Because of this, dendrohydrological reconstructions cannot easily incorporate the influence of watershed factors that can change streamflow measured at a certain point even when upstream precipitation remains the same, such as:

- stream channel profile (affected by incision, alluvial deposition, beaver activity, etc.);
- vegetation cover (affected by plant species dynamics, wildfire, landslides, etc.);
- land use (due to human activities, such as cattle or sheep grazing, clearcutting, crop production, urban development, etc.);
- diversions and their return flow (caused by either natural or human agents).

Dr. Saito and I have therefore proposed to combine dendroclimatic data with a watershed model, such as the Hydrologic Modeling System (HMS) developed by the Hydrologic Engineering Center (HEC) of the U.S. Army Corps of Engineers (USACE) for simulating the land components of the hydrologic cycle in dendritic watersheds. We plan to use tree-ring records for reconstructing precipitation at long time scales, which then becomes input for a deterministic watershed model (such as HEC-HMS) to calculate streamflow. Using model experiments, we can then study the influence of those watershed factors mentioned above. Such sensitivity studies will provide error bounds for the long-term runoff estimates that would otherwise be impossible to compute. We have chosen HEC-HMS because it is in the public domain, well-documented, widely used by both researchers and managers, constantly updated (e.g., a recent development is HEC-GeoHMS, which is an extension for ArcView Geographic Information System), able to simulate long periods of streamflows, and driven only by precipitation as meteorologic data input.

In conclusion, progress in the interdisciplinary combination of data and models for studying aquatic ecosystems will require novel ideas and approaches. A stimulating point of view on this topic is provided by Harte, 2002, Physics Today, pp. 29-34. He points out that natural systems will best be investigated using a synthesis of the “Newtonian” and “Darwinian” approaches to science, and provides examples and guidelines for ways in which such synthesis may be achieved.