Measuring and Monitoring Progress toward Sustainability

Chris Jones, UNC-CH and Warwick

Background:

I have felt for some time that a (and perhaps "the") major challenge to applied mathematics is the reconciling of data and models. Technology is leading us to state where we are overwhelmed with both massive sets of observational data and output from computational models. With regard to environmental issues in particular both inform decisions made but usually do so separately.

There is a disciplinary split between the responsibility for the organization, analysis and inference from models, on the one hand, and data on the other. The former task falls on the shoulders of applied mathematicians, while the latter on statisticians. Models and data are, however, becoming less easy to distinguish from each other. Models themselves are resulting in massive databases which require organization and analysis and are also used to fill in data. A neat distinction, and disciplinary separation, between the analysis of data and that of models is thus becoming less viable. My contention is that applied mathematics will need a radical resetting in which problems are formulated to include observational data as part of the fabric, in the way that initial and boundary conditions are at the moment. As I thought about the issue of what mathematics is needed to support measuring and monitoring toward a sustainable world, I kept being brought back to this same fundamental issue.

Sustainability:

The path to sustainability will not be straightforward. There are likely to be unanticipated consequences to our well-intentioned actions. We will need to assess the impact of changes made and their knock-on effects on climate, food sources, vegetation, animal habitats and many other environmental factors. Just as an example (off the top of my head), we could potentially get an enormous amount of energy from covering large parts of the world's deserts with solar cells, but do we know what impact this would have on the climate?

Seen in this light, the issue is not simply about data but also about models which would necessarily be used to project the impacts of any changes measured. A central issue will then be having data and models work together optimally to produce the most, and most important, information for decision support. Some guiding issues are:

- Estimation of impacts based on measured changes in energy sources and usage;
- Assessment of optimal measurements for determination of impacts;
- Articulation of critical information for assessing impacts;
- Assessment of new energy strategies in terms of projected impacts.

Projections from models and data:

For each issue mentioned above, a projection based on observational data plays the key role. Of interest might be assessing the data for the purpose of redirecting measurement strategies, or the impacts due to changes in the energy landscape up to that point. While the purposes might be different, the fundamental ingredient is the same: optimizing the information gained from data, from measurements, and theory, through the models used. The models may range from global climate or Earth system models to models of specific process or impact models.

There are various ways in which data can be used within the model framework to produce projections that reflect the impact of changes, which are in turn reflected in the data themselves. The data can be used to craft forcing in the model or to estimate parameters. But a more interesting approach is to use data assimilation. If only information from independent observations is used during the construction of the model, then the assimilation of data into an otherwise free-running model can offer a clear relation between effects reflected in the data and impacts seen through modified projections of the model. This is potentially a very powerful technique for assessing causal relations between active manipulation of the energy system and environmental impacts.

Challenges:

The challenge in realizing this approach is that the methodology of data assimilation is not tailored to dealing with complex, high-dimensional systems. Methods that assume some quasi-linearity coupled with Gaussian representations of error are well developed, and in operation in, for instance, weather centers all over the world. But systems with highly nonlinear effects and high dimensions, which come for instance from discretizing continuum models, will readily force the breakdown of such a data assimilation scheme.

Part of the issue is that the established approach is to minimize a certain cost function in the process of mediating between data and models. In (Bayesian) statistical terms, this is equivalent to finding the mode of the posterior distribution. If the distribution were Gaussian, then this is the same as finding the mean and the only information missing would be the variance. In a chaotic system, the posterior might, however, be extremely complicated and it is questionable whether the mode or mean alone, even supplemented by the variance, is information that can form the basis of a decision. For instance, if the distribution has many local maxima of approximately the same height then we would need to know this.

We need therefore:

- Robust statistical sampling methods for high-dimensional problems;
- An understanding of how nonlinearity impacts the generation of uncertainty pdfs;
- Techniques for handling high-dimensional, nonlinear systems; and
- Metrics for assessing adequacy of information.