A Case Study:

High Resolution Soil Water from Regional Databases and Satellite Images

Robin D. Morris, Vadim N. Smelyanskii
in collaboration with
Joseph Doughan, Jennifer Dunbar

Combating uncertainty:

Before you can combat uncertainty, you have to honestly quantify the uncertainty present.

Representing uncertainty in terms of probabilities is the natural framework, and allows a huge body of statistical tools to be readily applied.

It also corresponds to what we do — "quantiﬁed common sense".

The theory of probabilities is at bottom only common sense made precise.

Tools now exist that allow for computation in complex statistical models.

Case study:

- Soil water is often a limiting factor in plant growth.
- Ecologists are therefore interested in soil water over large areas.
- Directly measuring soil water is difficult, involving extensive ﬁeldwork and laboratory work.
- Satellite data can be used to infer plant growth.
- Under certain circumstances, plant growth can be linked to soil water.

Inputs:

- Satellite observations
  - AVIRIS data and clear infrared bands at 1.35um wavelength
- STATSGO databases
  - State soil geographic database
    - provides index information on soil water from an unknown (and most probably unique) sample and regular sampling of the area
- Task is to relate AVIRIS data to soil water and combine with the STATSGO information.

Probabilistic formulation:

- Start with the distribution of what we do know (AVIRIS data), conditional on everything we don't know (NDVI, LAI, AWC)
  \[ p(\text{avirs} | \text{ndvi, lai, awc}) \]
- Invert using Bayes' Theorem
  \[ p(\text{ndvi, lai, awc} | \text{avirs}) = p(\text{avirs} | \text{ndvi, lai, awc}) p(\text{ndvi, lai, awc}) / p(\text{avirs}) \]
- Look at the conditional independence structure of the problem
  \[ \text{avirs} \rightarrow \text{NDVI} \rightarrow \text{leaf area index} \rightarrow \text{soil water} \]
Probabilistic formulation (cont):

- \( p(\text{aviris} \mid \text{ndvi}, \text{lai}, \text{awe}) \) and \( p(\text{ndvi} \mid \text{lai}, \text{awe}) \)
- \( p(\text{aviris} \mid \text{ndvi}) \) and \( p(\text{ndvi} \mid \text{awe}) \)
- \( p(\text{aviris} \mid \text{ldvi}) \) and \( p(\text{ndvi} \mid \text{lai}) \)

Sources of uncertainty:

- \( p(\text{aviris} \mid \text{ndvi}) \) and \( p(\text{ndvi} \mid \text{lai}) \)
- \( p(\text{aviris} \mid \text{ndvi}) \) and \( p(\text{ndvi} \mid \text{awe}) \)
- \( p(\text{aviris} \mid \text{ldvi}) \) and \( p(\text{ndvi} \mid \text{lai}) \)
- \( p(\text{ndvi} \mid \text{lai}) \) and \( p(\text{lai} \mid \text{awe}) \)
- \( p(\text{aviris} \mid \text{ndvi}) \) and \( p(\text{ndvi} \mid \text{awe}) \)
- \( p(\text{aviris} \mid \text{ldvi}) \) and \( p(\text{ndvi} \mid \text{lai}) \)
- \( p(\text{ndvi} \mid \text{lai}) \) and \( p(\text{lai} \mid \text{awe}) \)
- \( p(\text{aviris} \mid \text{ndvi}) \) and \( p(\text{ndvi} \mid \text{lai}) \)
- \( p(\text{aviris} \mid \text{ldvi}) \) and \( p(\text{ndvi} \mid \text{lai}) \)
- \( p(\text{ndvi} \mid \text{lai}) \) and \( p(\text{lai} \mid \text{awe}) \)

- \( \text{ndvi} \) is derived directly from AVIRS read and near infrared bands.
- \( \text{awc} \)

- Compared with the other terms, the uncertainty here is negligible.

- More fieldwork - measuring LAI for a number of plots, and also making laboratory measurements of soil water.

- Nemani & Running (1988) provide the data points for this graph:

- Again, sample from \( p(A, B, C \mid \text{data}) \) where \( \text{data} = \text{A x ave} + B + e \)

- This is the second source of information - the STATSGO database provides prior information on the distribution of AWC values.
Putting it all together:

- \( p(\text{awc}, \text{lai}|\text{ndvi}) = p(\text{ndvi}|\text{lai}) p(\text{lai}|\text{awc}) p(\text{awc}) \)

- What we're really interested in is
  \( p(\text{awc}|\text{ndvi}) = \int p(\text{awc}|\text{lai}) p(\text{lai}|\text{ndvi}) \)

- This can be approximated by sampling from \( p(\text{awc}, \text{lai}|\text{ndvi}) \) and then ignoring the lai values.

Results:

Oregon NDVI

Polygon OR149

Results (cont):

Distribution of AWC from sampling, excluding the STATSGO prior

Discussion:

- The awc distributions shown are for sampling \( p(\text{ndvi}|\text{lai}) p(\text{lai}|\text{awc}) \), i.e. not including the prior \( p(\text{awc}) \).
- Compare the distribution from sampling with the STATSGO prior.

Further refinements:

- Improving \( p(\text{ndvi}|\text{lai}) \) and \( p(\text{lai}|\text{awc}) \) can be done by incorporating other types of information, e.g.
  - tree species,
  - age,
  - time since last clearcut, etc.
  - but all of these require more fieldwork.

- Including a spatial prior will reduce variation across the polygons.
Representing distributions using samples

- We wish to represent $p(x)$, and then to compute expectations

$$E(f(x)) = \int f(x)p(x)\,dx$$

- If we sample the domain of $x$ uniformly, and compute $p(x)$ for each $x$ value, then the integral can be approximated by

$$E(f(x)) = \sum f(x(i))p(x(i))$$

- However, this becomes inefficient if $x$ is in more than a few dimensions (curse of dimensionality)

- Instead, if we sample from $p(x)$, is concentrate the samples in the high probability regions, then we can approximate the integral by

$$E(f(x)) = \sum f(x(i))$$

Markov chain Monte Carlo:

- MCMC is a method of generating samples from $p(x)$.

A Markov chain is a sequence of states, where the probability of the next state depends only on the current state.

The method constructs a Markov chain that converges to the distribution of interest, $p(x)$.

To do this, it is necessary to determine the Transition Probabilities.

The Metropolis Algorithm is the simplest scheme for doing this:

- Initialize $X$ randomly
- Propose a new value $y$, where one of the elements of $X$ is changed by drawing it from a symmetric distribution
- Accept $y$ as the new value with probability $P_y = \min\left(1, \frac{p(y)}{p(x)}\right)$
- Otherwise return the current value, $x$
- Store the realizations

The realizations are an approximate sample from the posterior, from which we can compute quantities of interest (means, variances etc).