From:	Anderson, Chauncey
To:	zwick@pdx.edu
Cc:	James D Crammond
Subject:	Fwd: environmental data
Date:	Wednesday, April 01, 2015 10:31:59 PM

Hi Marty,
Your email to Alan made it's way to me. Our databases are replete with datasets that could be mined. I can hopefully work with you on it in a couple weeks, but I'm about to launch on a vacation for a week and know I'll have some high priority tasks waiting for me when I return. In the meantime, you might consider looking at the our offices continuous water quality and discharge data, available online through several avenues including these two: USGS Data Grapher for Oregon, and this mapping interface for our continuous data. At a number of our continuous sites, there are various instantaneous water quality sample data, available (with a bit of a learning curve) at NWISWeb.

Here's an example set of issues or types of analyses I know are waiting to be looked at:
One thing we've done a fair amount of work with is the use of turbidity as a surrogate for suspended sediment concentrations (SSC). Quite a few sites in Oregon have point data for SSC, and accompanying continuous data for turbidity, and for many of these we've established regressions in the past. We use these regressions to determine concentrations and loads, together with collocated stream discharge. A couple questions that have come up in the past are (1) how well might we mix data from multiple sites, if they have compatible turbidity instrumentation (not a trivial issue), and get a regression that could apply more broadly than our conservative approaches have allowed in the past? What if we stratified the sites by hydrologic influences (e.g. regulated or unregulated, valley floor vs higher gradient mountain streams, etc.)? What would the changes in uncertainty be in order to combine multiple sites like this, and would those changes be acceptable.
And, (2) What sorts of hysteresis effects are we seeing, at multiple temporal scales (e.g. event, seasonal, annual...). Are those hysteresis effects consistent, and what do they imply about sediment sources or hydrologic forcing factors?

Below are a couple of reports where we've used these methods on individual stream systems, and we have two projects wrapping up now that will add to these data sets.
http://pubs.usgs.gov/sir/2010/5008/
http://pubs.usgs.gov/sir/2010/5038/
http://pubs.usgs.gov/sir/2007/5187/
http://pubs.usgs.gov/sir/2007/5164/
http://pubs.usgs.gov/sir/2012/5200/

Regards,
Chauncey

Chauncey Anderson
Water Quality Specialist
US Geological Survey, Oregon Water Science Center
2130 SW 5th Avenue
Portland, OR 97201
tel: 503-251-3206 (w), 503-251-3470 (f)
chauncey@usgs.gov

---------- Forwarded message ----------
From: **Crammond, James** <crammond@usgs.gov>
Date: Wed, Apr 1, 2015 at 3:13 PM
Subject: Fwd: environmental data
To: GS-W-OR PROFESSIONALS <gs-w-or_professionals@usgs.gov>

Please read this carefully. Looks like a unique opportunity to analyze data sets with some real-world big data tools.

---------- Forwarded message ----------
From: **Alan Yeakley** <yeakley@pdx.edu>
Date: Wed, Apr 1, 2015 at 2:59 PM
Subject: Fwd: environmental data
To: SOE-Affiliated-Group <SOE-Affiliated-Group@pdx.edu>, SOE-Faculty-Group <soe-faculty-group@pdx.edu>, Martin Zwick <zwick@pdx.edu>, Wayne Wakeland <wakeland@pdx.edu>

Hi all,

Please see the below from Marty Zwick, and contact him directly if this is of interest. Sounds like a very good offer to me!

Alan

---------- Forwarded message ----------
From: **Martin Zwick** <zwick@pdx.edu>
Date: Mon, Mar 30, 2015 at 6:30 PM
Subject: environmental data
To: Alan Yeakley <yeakleya@pdx.edu>

Alan,

As you probably know, I teach my Data Mining with Information Theory (DMIT) class this quarter, where students bring data that we analyze with my information theory methodology (and the Occam software that we've developed here in SySc). Occasionally, students take the course without any data of their own, and ask me if I have data that they could analyze. At the moment, if anyone asks me this, I mention that I have data from Andrew Fountain on glaciers and that I also have in principle access to the hydrological data that Michael Fienen of the Wisconsin USGS is working on (he's indicated an interest/willingness to collaborate). I've also referred one student who is interested in forest fires to Rob Scheller in case Rob might have some data.

I just wanted to ask you if there are other people in SoE/ESM who have or might have data

that DMIT students could possibly use for their projects if they don't have data of their own. The kind of data that I can analyze is 'rectangular' spreadsheet data, where columns are variables, either continuous valued or discrete (nominal), and where rows are 'cases', typically members of some population or time (or even space) points. The number of rows, i.e., the sample size, should be in the 100s ideally, or certainly greater than 50. Values of N in the 1000s or more are, of course, even better. My methods can handle up to a few 100s of variables. They make sense if one has at least 4 or 5 variables. The approach is exploratory, i.e., aimed at discovering relations that are unknown, i.e., not hypothesized in advance. (People can use Occam to do confirmatory tests of theory-based hypotheses, but that isn't the point – or the advantage -- of the software.)

Please let me know if you know of SoE people who have data and aren't totally satisfied with the methods they're using to analyze this data. (If a person is totally satisfied in the analytics techniques they currently use, obviously they wouldn't be interested in trying a new method.) It would be nice to be able to add more names of possible data sources, if anyone comes to me looking for data for a project.

And, of course, I'd love to have more SoE students taking DMIT (and the theory course, DMM), so please mention this course to any students you think might be interested.

-Marty

--
Alan Yeakley, PhD
Director, School of the Environment
Professor, Environmental Science & Management
Portland State University
Portland, OR 97201 USA
E: yeakley@pdx.edu
W: http://web.pdx.edu/~yeakley/
T: 1-503-725-8040

To unsubscribe from this group and stop receiving emails from it, send an email to SOE-Affiliated-Group+unsubscribe@pdx.edu.

--
Dar Crammond, Director
USGS Oregon Water Science Center
2130 SW 5th Avenue
Portland, OR 97201

(503) 251- 3204
(503) 251-3470 (fax)

Learn more science for a changing world:
*USGS Podcasts: http://or.usgs.gov/podcasts
*USGS on Twitter: http://twitter.com/USGS_OR
*USGS on Facebook: http://www.facebook.com/USGeologicalSurvey