Wiener Biometrische Sektion
der Internationalen Biometrischen Gesellschaft
Region Österreich – Schweiz
http://www.meduniwien.ac.at/wbs/

Einladung zum

HERBSTSEMINAR DER WIENER BIOMETRISCHEN SEKTION „ANGEWANDTE STATISTIK AN DER UNIVERSITÄT FÜR BODENKULTUR“

am **Donnerstag, 6. November 2014** um **15:00 Uhr**
im Gebäude AQWU der Alten Wirtschaftsuniversität,
Augasse 2-6, 1090 Wien, 2.Obergeschoß, Raumnummer: D2.30
(Plan siehe Anhang)

Vortragende des Instituts für Angewandte Statistik und EDV, Universität für Bodenkultur Wien:

BERNHARD SPANGL

EXTREME VALUE STATISTICS AND ROBUST FILTERING FOR HYDROLOGICAL DATA

GREGOR LAAHA

GEOSTATISTICAL MODELS FOR RIVER NETWORKS

RAINER DANGL UND FRIEDRICH LEISCH

THE BENCHMARK DATA LIBRARY PROJECT: A METADATA REPOSITORY FOR ARTIFICIAL DATA

Wir freuen uns auf zahlreichen Besuch.

Franz König
Präsident

Stephan Lehr
Sekretär

ABSTRACTS

Extreme value statistics and robust filtering for hydrological data

Spangl Bernhard, Dipl.-Ing. Dr.techn.

Institut für Angewandte Statistik und EDV, Universität für Bodenkultur Wien (bernhard.spangl@boku.ac.at)

River discharge data are extreme events with strong seasonal and regional variations. Modern research pools data collected at different times and locations to improve risk prediction at a certain point and time. Besides extreme value statistics, this calls for filtering techniques to remove seasonalities, and regional effects. Among the extreme events of such data, it is not clear whether they are recurrent, so non-robust approaches attributing overly high influence to single extreme events could impair future risk predictions. This motivates the use of robust variants of filtering and extreme value procedures. We analyze data comprising daily average discharge time series of rivers from various sites in Germany collected over the last 35 years at least.
In the presentation, we limit ourselves to considering seasonal effects. For filtering, we fit a linear, time-invariant, time-discrete state space model to the data by means of a robust EM algorithm. By filtering we extract a detrended and desaisonalized series, to which we robustly fit a Generalized Pareto distribution.

Keywords: extreme values, robust filtering, EM algorithm, hydrology, river discharge

Geostatistical models for river networks

Gregor Laaha, Assoc.Prof. Dipl.-Ing. Dr.techn.

Institut für Angewandte Statistik und EDV, Universität für Bodenkultur Wien (gregor.laaha@boku.ac.at).

Abstract:
Geostatistical methods have become popular in various fields of hydrology, and typical applications include the prediction of precipitation events, the simulation of aquifer properties and the estimation of groundwater levels and quality. Surprisingly little effort has been undertaken to apply geostatistics to stream flow variables. This is most likely because of the tree-like structure of river networks, which poses specific challenges for geostatistical regionalisation. Notably, the shape of catchments (irregular block support), the nestedness of catchments along the river network (overlapping support), and the definition of a relevant distance measure between catchments pose specific challenges. This paper attempts an annotated survey of models proposed in the literature, stating contributions and pinpointing merits and shortcomings. Two conceptual viewpoints are distinguished, (1) one-dimensional models which use covariances along a stream network based on river distance [1], and (2) two-dimensional models where stream flow is conceptualised as the integral of the spatially continuous local runoff process over the catchment area [2]. Both geostatistical concepts are evaluated relative to geostatistical standard methods based on Euclidean distances. It is shown how the methods perform in various examples including spatial prediction of low stream flows [3], stream temperatures [4] and nitrate loads.

References:
[1] Ver Hoef JM, Peterson E, Theobald D (2006) Spatial statistical models that use flow and stream distance Environmental and Ecological Statistics 13:449-464.
[2] Skøien J, Merz R, Blöschl G (2006) Top-kriging – geostatistics on stream networks. Hydrology and Earth System Sciences 10: 277–287.
[3] Laaha G, Skøien J, Blöschl G (2010) Spatial prediction on a river network: Comparison of Top-kriging with regional regression. Hydrological Processes, 28(2), 315–324, doi:10.1002/hyp.9578, 2014.
[4] Laaha G, Skøien J, Blöschl G (2010) Spatial prediction of stream temperatures using Top-kriging with an external drift. Environmental Modeling & Assessment, 18(6), 671–683, doi:10.1007/s10666-013-9373-3, 2013.
[5] Garreta, V, Monestiez, P & Ver Hoef, JM (2009) *Environmetrics* 439–456.

The Benchmark Data Library Project: a metadata repository for artificial data

Rainer Dangl, Mag. and Friedrich Leisch, Univ.Prof. Dipl.-Ing. Dr.techn.

Institut für Angewandte Statistik und EDV, Universität für Bodenkultur Wien (rainer.dangl@boku.ac.at)

Introducing new methods of model validation in unsupervised learning requires a lot of testing. New algorithms need to be thoroughly validated before they are put to use on real world problems. Hence, benchmarking plays an important part in the development process. For this purpose, artificial data is generally used. Usually one would create a data set from scratch that should illustrate the capabilities of the new method - yet a more practical approach would be to use data utilized in previous studies (or to share the newly developed data sets with others) in order to facilitate the comparison of methods by using the same data for testing.
The talk will focus on the ongoing development of a web application and an accompanying R package that offer on the one hand the ability to generate data sets that were used in previous benchmarking experiments and on the other hand the possibility to upload a newly created experimental setup. Three data types are avaliable at the moment: metric, functional and ordinal data. The web application will serve as an exchange for the metadata descriptions, the R package is used for generating the actual data sets.

Keywords: Benchmarking, Model Validation, Clustering, Unsupervised Learning, Artificial Data