University of Zurich[UZH]

**Department of Informatics**

University of Zurich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zurich

First_name Last_name

Switzerland

**Prof. Dr. Michael Böhlen**
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

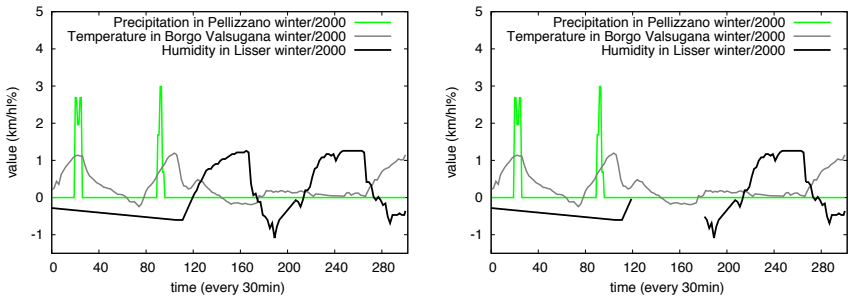
Zürich, 25. September 2012

**Missing Block Recovery using Centroid Decomposition Method**

Master Projektarbeit (18 ECTS):

Work overview:

The aim of this Master project is to investigate and implement the Centroid Decomposition method to recover blocks of missing values in shifted time series. As an example for the block recovery in shifted time series, consider the humidity, precipitation and temperature time series shown in Figure 1. The peaks of the humidity time series are shifted in time with respect to the peaks of the two other time series as shown in Figure 1(a). We propose to use the Centroid Decomposition to restore the missing block in the time range $]120, 175[$, as shown in Figure 1(b).




(a) Complete Shifted Time Series (b) Shifted Time Series with Missing Block

Figure 1: Example of Shifted Time Series with Missing Block

The Centroid Decomposition is a matrix decomposition technique that decomposes an input matrix $\mathbf{V}$ of data into a sum of the product of the loading vectors denoted as $L$ and the centroid vectors denoted as $C^T$. Formally, a matrix $\mathbf{V} = [V_1|V_2|\ldots|V_n] \in \mathcal{R}^{m\times n}$ can be decomposed as follows: $\mathbf{V} = \sum_{i=1}^{d} L_i C_i^T$ where:

- $L$: Loading vectors
- $C^T$: Centroid vectors
- $d$: number of centroid vectors to compute

Work tasks:

1. Understand and implement the Centroid Decomposition algorithm
2. Apply the Centroid Decomposition for the block recovery in hydrological shifted time series. The datasets are already loaded in an Oracle server.
3. Identify classes of shifted time series that can be recovered
4. Empirical comparison of the recovery accuracy and scalability of the implemented algorithm with respect to REBOM: `http://www.ifi.uzh.ch/dbtg/sw/rmvtp.html`
5. Report of 15-20 pages
6. Oral exam (approx. 25 min)

Literature:

1. Chu, M.T., and Funderlic, R.E., : *The Centroid Decomposition: Relationships Between Discrete Variational Decompositions and SVDs*, in SIAM J. Matrix Analysis and Applications, 2002
2. Oh, S.M., *A Note on Singular Value Decomposition*, Technical report, 2005
3. Benjamin, M.M., *Missing Data Problems in Machine Learning*, PhD thesis, 2008

Task assignment and supervisor:

- Mourad Khayati

Starting date of thesis: TBD

Ending date of thesis: TBD

University of Zurich
Department of Informatics

Prof. Dr. Michael Böhlen
Professor