Prof. Themis Palpanas
Paris Descartes University
themis@mi.parisdescartes.fr

Dr. Niklas Boers
École Normale Supérieure
nboers@lmd.ens.fr

2 Funded Research Internships:
Very Large Time Series Analysis for Predictive Maintenance

There is an increasingly pressing need, by several applications in diverse domains, for developing techniques able to index and mine very large collections of sequences, or time series. The purpose of this project is to design techniques for mining truly massive collections of time series, focusing on the task of predictive maintenance for electric power plants. Existing techniques to identify statistical precursors of imminent critical transitions in the system under consideration are limited to comparably small sets of time series (due to their computational complexity).

Internship Position 1: We will develop efficient algorithms for outlier detection in time series, making use of advanced indexing and similarity querying techniques [1]. This will allow us to scale to very large collections of time series, being able to efficiently analyze large historical datasets, and identify a richer and more complete set of abnormal patterns. The resulting abnormal patterns will be verified by the domain experts, and could then be used to assist the analysts in detecting abnormalities in real-time [2]. In our analysis, we will also use classification algorithms that allow us to determine (or label) some phenomenon represented by the series as normal, or abnormal.

Internship Position 2: We will develop efficient and scalable algorithms to identify precursors of abrupt transitions. In cases where the system under consideration is described in terms of a large set of dependent time series, spatial networks can be used for the forecast of extreme events within the system, but also to forecast large-scale regime shifts of the entire system [3,4]. Particular care has to be taken in order to remove spurious artifacts arising from the underlying correlation analysis: Graph edges should only represent conditionally dependent correlations, leading to the framework of Bayesian Networks.

This work will be conducted in close collaboration with the Électricité de France (EDF), which will also provide access to real data from sensors that monitor a multitude of observables in power plants.

Internship:

Accepting any of these two projects will make you part of an enthusiastic team working on real, challenging problems!
Prerequisites: experience with probability theory and statistics, file and data structures, excellent programming skills.
The internships will last between 3-6 months, and are fully funded.

The Team:

Themis Palpanas is a professor of computer science at the Paris Descartes University, where he is co-director of diNo, the data management group. He received the MSc and PhD degrees from the University of Toronto, Canada. His team has developed world-wide expertise on data series management and analysis.

Niklas Boers is a postdoctoral researcher at the Department of Geosciences of the École Normale Supérieure in Paris, funded by the Alexander von Humboldt foundation. He graduated in Physics and Mathematics at LMU and TU Munich, and obtained a PhD in theoretical Physics from HU Berlin. His main research interests include complex network theory, probabilistic prediction of extreme events, and statistical techniques to predict abrupt (critical) transitions in time series with large uncertainties.

References:

[1] Kostas Zoumpatianos, Stratos Idreos, Themis Palpanas. Indexing for Interactive Exploration of Big Data Series. SIGMOD 2014.
[2] Katsiaryna Mirylenka, Alice Marascu, Themis Palpanas, Matthias Fehr, Stefan Jank, Gunter Welde, Daniel Groeber. Envelope-Based Anomaly Detection for High-Speed Manufacturing Processes. APC|M, 2013.
[3] Niklas Boers et al.: Prediction of Extreme Floods in the Eastern Central Andes based on a Complex Network Approach, Nature Communications, 2014.
[4] Niklas Boers et al.: Complex networks identify spatial patterns of extreme rainfall events of the South American Monsoon System, Geophysical Research Letters, 2013.