Investigation of Arctic and Antarctic spatial and depth patterns of sea water in CTD profiles using chemometric data analysis

In this paper we examine 2- and 3-way chemometric methods for analysis of Arctic and Antarctic water samples. Standard CTD (conductivity–temperature–depth) sensor devices were used during two oceanographic expeditions (July 2007 in the Arctic; February 2009 in the Antarctic) covering a total of 174 locations. The output from these devices can be arranged in a 3-way data structure (according to sea water depth, measured variables, and geographical location). We used and compared 2- and 3-way statistical tools including PCA, PARAFAC, PLS, and N-PLS for exploratory analysis, spatial patterns discovery and calibration. Particular importance was given to the correlation and possible prediction of fluorescence from other physical variables. MATLAB’s mapping toolbox was used for geo-referencing and visualization of the results. We conclude that: 1) PCA and PARAFAC models were able to describe data in a satisfactory way, but PARAFAC results were easier to interpret; 2) applying a 2-way model to 3-way data raises the risk of flattening the covariance structure of the data and losing information; 3) the distinction between Arctic and Antarctic seas was revealed mostly by PC1, relating to the physico-chemical properties of the water samples; and 4) we confirm the ability to predict fluorescence values from physical measurements when the 3-way data structure is used in N-way PLS regression.
In this paper, we introduce a new method, based on spherical principal component analysis (S-PCA), for the identification of Rayleigh and Raman scatters in fluorescence excitation–emission data. These scatters should be found and eliminated as a prestep before fitting parallel factor analysis models to the data, in order to avoid model degeneracies. The work is inspired and based on a previous research, where scatter removal was automatic (based on a robust version of PCA called ROBPCA) and required no visual data inspection but appeared to be computationally intensive. To overcome this drawback, we implement the fast S-PCA in the scatter identification routine. Moreover, an additional pattern interpolation step that complements the method, based on robust regression, will be applied. In this way, substantial time savings are gained, and the user's engagement is restricted to a minimum, which might be beneficial for certain applications. We conclude that the subsequent parallel factor analysis models fitted to excitation–emission data after scatter identification based on either ROBPCA or S-PCA are comparable; however, the modified method based on S-PCA clearly outperforms the original approach in relation to computational time.
Authors: Kotwa, E. K. (Intern), Jørgensen, B. M. (Intern), Brockhoff, P. B. (Intern), Frosch, S. (Intern)
Pages: 3-11
Publication date: 2013
Main Research Area: Technical/natural sciences

**Publication information**

**Journal:** Journal of Chemometrics

**Volume:** 27

**Issue number:** 1-2

**ISSN (Print):** 0886-9383

**Ratings:**

- BFI (2018): BFI-level 1
- Web of Science (2018): Indexed yes
- BFI (2017): BFI-level 2
- Web of Science (2017): Indexed Yes
- BFI (2016): BFI-level 2
- Scopus rating (2016): SJR 0.5 SNIP 0.873 CiteScore 1.71
- BFI (2015): BFI-level 2
- Scopus rating (2015): SJR 0.553 SNIP 0.984 CiteScore 1.82
- Web of Science (2015): Indexed yes
- BFI (2014): BFI-level 2
- Scopus rating (2014): SJR 0.536 SNIP 1.235 CiteScore 1.72
- BFI (2013): BFI-level 2
- Scopus rating (2013): SJR 0.749 SNIP 1.316 CiteScore 2.2
- ISI indexed (2013): ISI indexed yes
- Web of Science (2013): Indexed yes
- BFI (2012): BFI-level 2
- Scopus rating (2012): SJR 0.66 SNIP 1.342 CiteScore 2.03
- ISI indexed (2012): ISI indexed yes
- BFI (2011): BFI-level 2
- Scopus rating (2011): SJR 0.742 SNIP 1.037 CiteScore 1.66
- ISI indexed (2011): ISI indexed yes
- BFI (2010): BFI-level 2
- Scopus rating (2010): SJR 0.591 SNIP 0.75
- BFI (2009): BFI-level 2
- Scopus rating (2009): SJR 0.931 SNIP 1.008
- Web of Science (2009): Indexed yes
- BFI (2008): BFI-level 1
- Scopus rating (2008): SJR 0.868 SNIP 0.903
- Web of Science (2008): Indexed yes
- Scopus rating (2007): SJR 0.829 SNIP 1.097
- Scopus rating (2006): SJR 1.073 SNIP 1.509
- Web of Science (2006): Indexed yes
- Scopus rating (2005): SJR 1.435 SNIP 1.358
- Web of Science (2005): Indexed yes
- Scopus rating (2004): SJR 1.432 SNIP 1.371
- Web of Science (2004): Indexed yes
- Scopus rating (2003): SJR 1.572 SNIP 1.421
- Web of Science (2003): Indexed yes
- Scopus rating (2002): SJR 0.872 SNIP 1.175
- Web of Science (2002): Indexed yes
- Scopus rating (2001): SJR 1.903 SNIP 1.274
- Web of Science (2001): Indexed yes
- Scopus rating (2000): SJR 1.61 SNIP 1.272
- Web of Science (2000): Indexed yes
Robust procedures in chemometrics

The general aim of the thesis was to contribute to the improvement of data analytical techniques within the chemometric field. Regardless the multivariate structure of the data, it is still common in some fields to perform uni-variate data analysis using only simple statistics such as sample mean and variance. Recent instrumental developments in chemometrics often result in high-order data, for which uni-variate tools do not suffice and multivariate data analysis is required. Moreover, many multivariate models assume normality of the residuals (which in many cases is far from reality) and are not resistant towards outliers (which are known to be more the rule than the exception for empirical data). That is the reason for robust methods being a valuable tool for both semi-automated detection of outliers and model building.

The approach adapted in this thesis, can be split in two main parts: 1. applying a multivariate and multi-way data analytical frame-work in fields where less sophisticated data analysis methods are currently used, and 2. developing new, more robust alternatives to already existing multivariate tools.

The first part of the study was realised by applying two- and three-way chemometrical methods, such as PCA and PARAFAC models for analysing spatial and depth profiles of sea water samples, defined by three data modes: depth, variables and geographical location. Emphasis was also put on predicting fluorescence values, as being a natural measure of biological activity, by applying and comparing the Partial Least Squares (PLS) regression technique with its multi-way alternative, N-PLS. Results of the analysis indicated superiority of the three-way frame-work, potentially constituting a novel assessment of the sea water measurements. Particularly in the case of regression models there is a clear preference towards the more complex model, delivering more reliable predictions than a classical 2-way PLS. Therefore, using multi-way data analysis tools is recommended, in order to extract the full information from multi-way data structures.

The second part of the thesis targeted qualitative properties of the analysed data. The broad theoretical background of robust procedures was given as a very useful supplement to the classical methods, and a new tool, based on robust PCA, aiming at identifying Rayleigh and Raman scatters in excitation-mission (EEM) data was developed. The results show clearly that robust methods can significantly contribute to the improvement of existing analytical techniques used commonly in chemometrics, for example by providing excellent outlier detection tools. It is therefore advised to apply robust and classical procedures simultaneously, at least to determine if contamination in the data is present. For this becoming a standard procedure, further work is required, aiming at implementing reliable robust algorithms into standard statistical programs.

General information
State: Published
Organisations: Mathematical Statistics, Department of Informatics and Mathematical Modeling
Authors: Kotwa, E. (Intern), Brockhoff, P. B. (Intern)
Number of pages: 174
Publication date: 2012

Publication information
Place of publication: Kgs. Lyngby
Publisher: Technical University of Denmark (DTU)
Original language: English
Series: IMM-PHD-2012
Number: 284
ISSN: 0909-3192
Main Research Area: Technical/natural sciences
Electronic versions:
phd284_Kotwa_E.pdf
Publication: Research › Ph.D. thesis – Annual report year: 2012

Spatio-temporal analysis and modeling of short-term wind power forecast errors
Forecasts of wind power production are increasingly being used in various management tasks. So far, such forecasts and related uncertainty information have usually been generated individually for a given site of interest (either a wind farm or a group of wind farms), without properly accounting for the spatio-temporal dependencies observed in the wind generation field. However, it is intuitively expected that, owing to the inertia of meteorological forecasting systems, a forecast error made at a given point in space and time will be related to forecast errors at other points in space in the following period.
The existence of such underlying correlation patterns is demonstrated and analyzed in this paper, considering the case-study of western Denmark. The effects of prevailing wind speed and direction on autocorrelation and cross-correlation patterns are thoroughly described. For a flat terrain region of small size like western Denmark, significant correlation between the various zones is observed for time delays up to 5 h. Wind direction is shown to play a crucial role, while the effect of wind speed is more complex. Nonlinear models permitting capture of the interdependence structure of wind power forecast errors are proposed, and their ability to mimic this structure is discussed. The best performing model is shown to explain 54% of the variations of the forecast errors observed for the individual forecasts used today. Even though focus is on 1-h-ahead forecast errors and on western Denmark only, the methodology proposed may be similarly tested on the cases of further look-ahead times, larger areas, or more complex topographies. Such generalization may not be straightforward. While the results presented here comprise a first step only, the revealed error propagation principles may be seen as a basis for future related work.

**General information**

State: Published  
Organisations: Mathematical Statistics, Department of Informatics and Mathematical Modeling  
Authors: Tastu, J. (Intern), Pinson, P. (Intern), Kotwa, E. (Intern), Madsen, H. (Intern), Nielsen, H. A. (Intern)  
Pages: 43-60  
Publication date: 2011  
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Wind Energy  
Volume: 14  
Issue number: 1  
ISSN (Print): 1095-4244  
Ratings:  
BFI (2018): BFI-level 2  
Web of Science (2018): Indexed yes  
BFI (2017): BFI-level 2  
Web of Science (2017): Indexed yes  
BFI (2016): BFI-level 2  
Scopus rating (2016): CiteScore 3.37 SJR 1.104 SNIP 2.306  
Web of Science (2016): Indexed yes  
BFI (2015): BFI-level 2  
Scopus rating (2015): SJR 1.196 SNIP 2.086 CiteScore 3.06  
Web of Science (2015): Indexed yes  
BFI (2014): BFI-level 2  
Scopus rating (2014): SJR 1.272 SNIP 3.75 CiteScore 3.42  
Web of Science (2014): Indexed yes  
BFI (2013): BFI-level 2  
Scopus rating (2013): SJR 1.275 SNIP 2.464 CiteScore 2.75  
ISI indexed (2013): ISI indexed yes  
Web of Science (2013): Indexed yes  
BFI (2012): BFI-level 2  
Scopus rating (2012): SJR 1.126 SNIP 2.39 CiteScore 2.36  
ISI indexed (2012): ISI indexed yes  
Web of Science (2012): Indexed yes  
BFI (2011): BFI-level 2  
Scopus rating (2011): SJR 1.024 SNIP 2.718 CiteScore 2.49  
ISI indexed (2011): ISI indexed yes  
Web of Science (2011): Indexed yes  
BFI (2010): BFI-level 2  
Scopus rating (2010): SJR 1.487 SNIP 2.013  
Web of Science (2010): Indexed yes  
BFI (2009): BFI-level 2  
Scopus rating (2009): SJR 1.124 SNIP 1.448  
Web of Science (2009): Indexed yes  
BFI (2008): BFI-level 2
Spatio-temporal modeling of wind power prediction errors

General information
State: Published
Organisations: Mathematical Statistics, Department of Informatics and Mathematical Modeling
Authors: Vlasova, J. (Intern), Kotwa, E. (Intern), Nielsen, H. A. O. T. 3. (Intern), Madsen, H. (Intern), Pinson, P. (Intern)
Publication date: 2008

Host publication information
Title of host publication: Proceedings of the European Wind Energy Conference 2008
Main Research Area: Technical/natural sciences
Conference: 2008 European Wind Energy Conference and Exhibition, Brussels, Belgium, 31/03/2008 - 31/03/2008
Source: orbit
Source-ID: 223278
Publication: Research › Article in proceedings – Annual report year: 2008

Spatio-temporal modelling of short-term wind power prediction errors: 02004/FU5766 - Improved wind power prediction

General information
State: Published
Organisations: Mathematical Statistics, Department of Informatics and Mathematical Modeling
Authors: Vlasova, J. (Intern), Kotwa, E. (Intern), Nielsen, H. A. (Intern), Madsen, H. (Intern)
Publication date: 2007

Publication information
Publisher: Informatics and Mathematical Modelling, Technical University of Denmark, DTU
Original language: English
Number: 18
ISSN: 1601-2321
Main Research Area: Technical/natural sciences
Electronic versions: tr07_18.pdf
Links:
Projects:

Analysis and Modelling of Chain Data

Department of Informatics and Mathematical Modeling
Period: 01/10/2007 → 05/09/2014
Number of participants: 5
Phd Student:
Kotwa, Ewelina Katarzyna (Intern)
Main Supervisor:
Brockhoff, Per B. (Intern)
Examiner:
Kulahci, Murat (Intern)
Rinnan, Åsmund (Ekstern)
Westad, Frank Ove (Ekstern)

Financing sources
Source: Internal funding (public)
Name of research programme: Globaliseringsmidler
Project: PhD