Improving the quality of ecosystem data time series for integrated assessment in the Baltic Sea

Svetlana Kasatkina, _Pavel Gasyukov_

Atlantic Research Institute of Marine Fisheries and Oceanography (AtlantNIRO), Kaliningrad, Russia. Presenter contact details: ks@atlant.baltnet.ru, Phone +7 4012 925469

Summary

Specific particular of existing ecosystem data time series (biotic and abiotic) being is a lot of missed observations. It is therefore important to rebuild the missed observations and estimate uncertainty in data time series. The authors shown availability of using Principal Component Analysis (PCA) based on versions differed   from classic PCA: imputation method (IM) accompanied with cross-validation procedure and Variation Bayesian Principal Component Analysis (VBPCA).   The advantages of proposed PCA versions are the possibility estimating   uncertainty in data time series and statistical characteristics for rebuilding missed observations. Under example of the data for the Central Baltic area it was shown that rebuilding missed observations and estimating uncertainty in data time series may improve ecosystem parameters.

Introduction

Integrated analysis based on ecosystem data time series provides information both for assessing pelagic ecosystem structure and dynamic with focus on long term-trends and regime shifting and for supporting stock assessment and management advisory activity. All datasets currently used are derived from historical and ongoing multidisciplinary pelagic surveys and monitoring programs with variable sampling level. These ecosystem data can has a lot of missed observations. The practice of integrated assessments in different areas of the Baltic Sea indicated that the percent proportion of the missing observations may be considerable (Diekmann, R., 2010). The authors discussed the urgent problems on rebuilding missed observations. under example of time series for the Central Baltic area. applying the modern statistical methods for data processing.

Material and methods

The authors proposed to use Principal Component Analysis (PCA) based on versions differed   from classic PCA: imputation method (IM) and Variation Bayesian Principal Component Analysis (VBPCA). Imputation method (Ilin & Raiko, 2008) was accompanied with cross-validation procedure for determining number of principal components. This number of components should provide the prescribed accuracy (Gasyukov, P. and Kasatkina, 2010.). Variational Bayesian PCA (Ilin & Raiko 2008) belongs to the group of probabilistic PCA methods, including: probabilistic PCA, maximum a posteriory estimation and Variational Bayesian PCA. These methods have several advantages as compared to the imputation method: a) incorporation of the noise component into the model, b) the algorithm has the normalization property, c) the algorithm reduces overfitting, and d) provides statistical characteristics of estimates and confidential intervals. The practical application of proposed data processing was presented under example of 19 ecosystem parameters time series for Central Baltic Sea (1975-2008). . The variances estimates and CV are known only for international bottom trawl surveys indices.

Results and Discussion

Figure 1 shows existing time-series of several ecosystem indices   and correspondent time-series recalculated by Imputation method. The considered method allows recovering missing observations filtration of noises.

Figure 2 shows existing time-series of several ecosystem indices   and correspondent time series recalculated by Variational Bayesian PCA. All calculations used Bayes formula. The VBPCA method rebuild the missed values and gave the confidence intervals. The deviations between observed and recalculated values in some cases were significant.
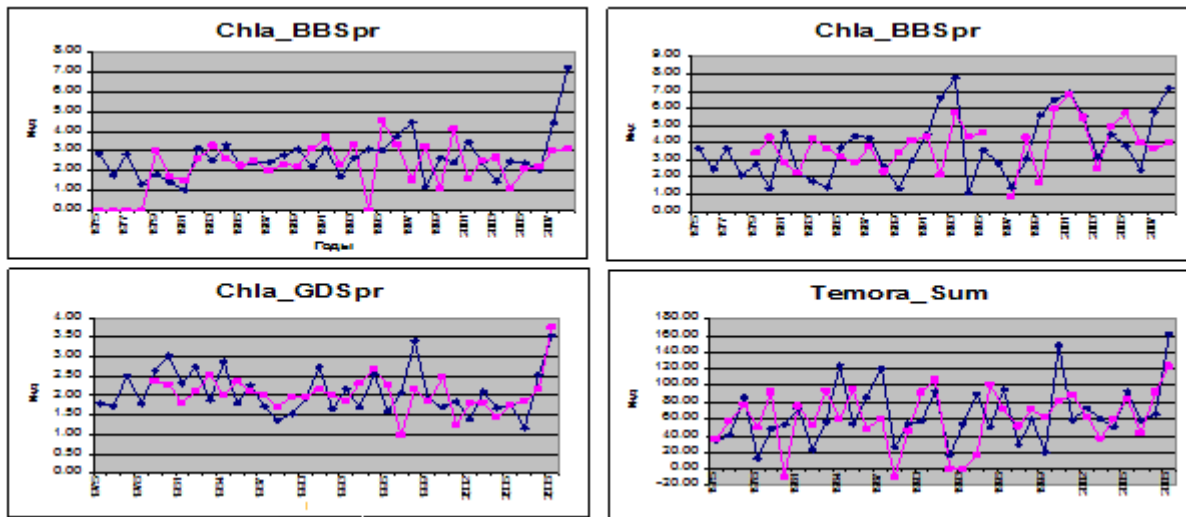
Figure 1. Examples of existing time-series of several ecosystem indices (pink color) and correspondent time-series recalculated by Imputation method (black color).
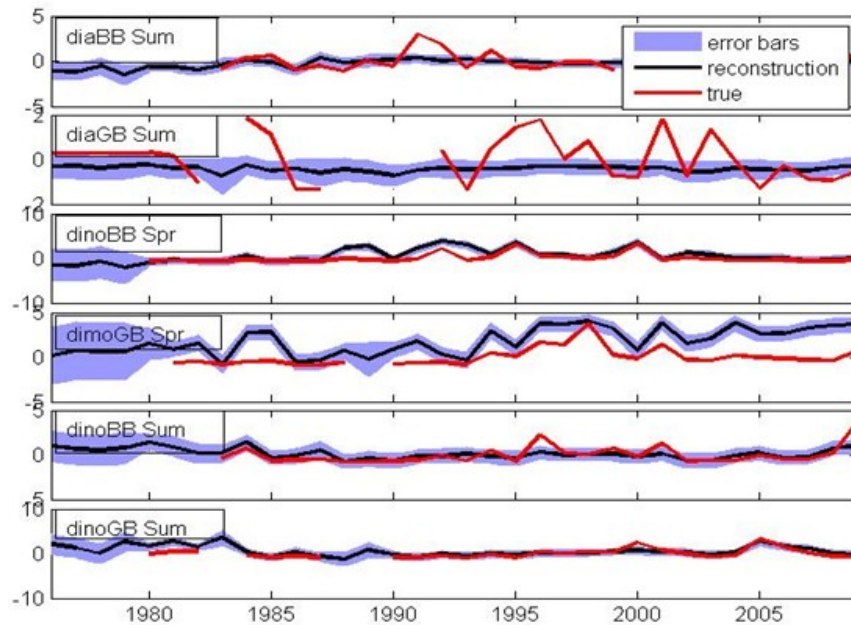

Figure 2. . Examples of existing time-series of several ecosystem indices (red color) and correspondent time-series recalculated by applying Variation Bayesian PCA (black color). Confidence intervals for recalculated data are shown by blue color.

Refereces

Diekmann, R. and Mцllmann, C. 2010. Integrated ecosystem assessments of seven Baltic Sea areas covering the last three decades. ICES Cooperative Research Report No. 302. 90 pp.

Gasyukov, P. And Kasatkina S. 2010. Application of the principle-component analysis in research of spatial-temporal distribution of the east cod in the Baltic sea. Document ICES CM 2010/G: 07, 15 pp.

Ilin A. And Raiko T. 2008. Practical approaches to principal component analysis in the presence of missing values. Helsinki University of Technology, Faculty of Information and Natural Sciences, Department of Information and Computer Science. – 48 pp