Earth Syst. Dynam. Discuss.,
doi:10.5194/esd-2016-51-RC2, 2017

Earth System Dynamics Discussions

Interactive comment on "Multivariate Anomaly Detection for Earth Observations: A Comparison of Algorithms and Feature Extraction Techniques" *by* Milan Flach et al.

R. V. Donner (Referee)

reik.donner@pik-potsdam.de

Received and published: 7 February 2017

Flach et al. present a detailed inter-comparison between a selection of recently applied methodological approaches for detecting multivariate anomalies in Earth observation (EO) data, including a performance assessment based on artificially generated time series data that capture some of the essential features (and complications) of real-world observation. The topic is timely and important, since with the fastly growing amount of big data from remote sensing, the automated identification of key features and particularly unexpected behaviors becomes a crucial task. In this regard, I warmly welcome this study and believe that it can be an important milestone in its field, even though it necessarily presents just a case study and can thus not be complete by

definition.

Prior to accepting this very interesting work for final publication in Earth System Dynamics, I would like to ask the authors to address a couple of questions I came up with when working through their material.

1. It would be good if the authors could clarify already in the abstract which kind of anomalies they aim to address. From Figure 2, it is evident that the four considered types of "events" (or better, episodic behaviors) – base shift, trend onset, change in mean seasonal cycle amplitude, change in variance – affect predominantly the basic statistical features of the data, while their dynamical characteristics (respectively, those of the residuals after removing the seasonal variability component) are largely unaffected. Since this is in contrast to some recent works (including papers by the reviewer's group) which have particularly focused on "dynamical anomalies", it might be worth clarifying this from the beginning. In this context, it is interesting that the authors also consider recurrence characteristics, which are commonly used for detecting changes in the dynamical patterns. However, what they consider here is just a variant of the recurrence rate, which is essentially a statistical characteristic again, as opposed to more sophisticated complexity measures that can be defined within this framework as well.

2. The motivation for choosing the specific settings in the artificial data could be further clarified, especially regarding explicit statements on typical features of real-world EO data. In this context, I was wondering why the authors study only short-term correlated noises, whereas much of the stochastic background signals in common geophysical variables exhibits long-term memory, which might strongly complicate the anomaly detection. Do the authors consider their white/red noises mostly reflecting additive measurement uncertainties or "true" dynamical components, e.g., due to variables and/or scales not resolved by the measurement process.

3. To me, the idea behind mwVAR is not fully clear. Subtracting the median mwVAR

just removes a constant factor from the time series as it is described now. Maybe it should be better explained here how this specific “preprocessing” step works.

4. On p.8, l.22, the authors address the model parameters. However, these parameters have not been introduced before, so it is hard to grasp their meaning at this point.

5. Quite a bit of potentially interesting material is “not shown” by the authors. I understand and agree the need to focus on the most important aspects, but maybe the authors could consider preparing some supplementary material containing these additional results.

6. In general, the parameter selection in the different methods is not well motivated (e.g., embedding delay and dimension, number of nearest neighbors, outlier ratio). Some more words on these aspects would be helpful. The authors shortly address the subjectivity of parameter selection on p.16, ll.3-6, but do not mention that there are established ways to make (some of) these parameter selections at least a bit more objective. I do not request a detailed discussion on this aspect, but it would be worth mentioning it at least.

Technical comments:

- p.1, keywords: please capitalize the names Mahalanobis, Foley and Sammon

- p.3, ll.1-3: The papers by Donges, Rammig, Zscheischler et al. use only a bivariate form of event coincidence analysis. Since the authors refer her to the “truly multivariate” case, a better reference would be Siegmund et al., Front. Plant Sci., 7, 733, 2016, who introduced a multivariate version of event coincidence analysis.

- p.3, l.21: The term “data cube” should be explicitly defined here – it is intuitively clear (especially in connection with Fig. 1), but especially the spatial component (2d vs. 3d) could differ from what is considered in this paper.

- p.3, ll.28-30: It is not clear if the authors wish to consider “multivariate events” or “compound events” (i.e., such that are anomalous with respect to the marginal feature

distribution of a single variable or the joint feature distribution of a (sub)set of variables.

- p.4, l.15: Why is Appendix B referenced in the paper before Appendix A. I think that changing the order of both Appendices would be more logical.

- p.5, ll.6-9: replace a, b, c, d by (a), (b), (c), (d)

- p.5, l.14: I think that it is not the Earth observations (EOs) that are driven by extrinsic forcings, but the EO variables.

- p.6, l.24: In fact, what you study is the maximization of the rate of correct detections at simultaneous minimization of false detections (this is essentially what the ROC analysis does).

- p.6, l.26: “the anomaly time series becomes the feature then” – maybe the authors should explicitly state here what they “define” (consider) to be meant by a feature.

- Figure 3: Since the authors allow for combining different feature extraction techniques, they should emphasize here that their application might be non-commutative in some cases. For example, TDE must be performed after sMSC, otherwise, the signal would be dominated by seasonality and potentially reflect different features than those one is actually interested in.

- p.7, l.9: “This theoretical consideration does not hold true for high dimensional multivariate data.” Do the authors have a reference for this? I am not convinced that this statement is correct in general. In particular, one may refer to multi-channel SSA (mSSA), which essentially combines TDE for multivariate data with PCA. What is the difference between mSSA and “dynamic PCA” mentioned in p.7, ll.18-19?

- p.8, l.2: To my knowledge, there are various variants of ICA, and the one maximizing the negentropy is just one version among several others.

- p.8, l.12: “in the literature”

- p.8, l.21: “we fix the model parameters”

- p.8, ll.22-23: “model parameters... and the models themselves... are estimated” – better use the terms model selection and parameter estimation separately- p.8, ll.24-25: Do the authors mean “more resampling is NOT affordable...”?

- p.8, l.26: “a resampling of 3” – 3 what?

- p.8, l.30: “if one or several of the univariate variables are below or above a certain quantile threshold” – again: do the authors mean marginal quantiles or multivariate quantiles (i.e., multivariate or compound extremes)? Page 9, ll.2-3 suggests that they refer to extremes in the marginals.

- p.9, ll.1-2: The event coincidence analysis the authors refer to here is a bivariate (or, in its extension, multivariate) statistical method. Its relevance in the context of the present work is not clear, since I do not find information that statistical interrelationships between anomalies in different variables are considered here.

- p.9, l.3: Details on the definition of the threshold exceedance score should be given.

- p.9, l.12: I suppose that the authors are using standardized variables; otherwise, defining distances across different variables might not make much sense in the real-world data case. I recommend elaborating a bit more on this aspect.

- p.9, ll.15-16: This formulation should be checked again; for me, the difference between the two measures does not become obvious from the given description.

- p.9, ll.18-19: In how far do the authors really “take advantage” here? Isn't it rather that you wish to exclude trivial information due to autocorrelation in your variables?

- p.9, l.29: “An \epsilon-hyperball”

- p.9, l.31: $\zeta\cdotT{-1}$

- p.9, l.32: “degree of centrality” is not the proper network theoretic term (it would be “degree centrality” or just “degree”); however, what the authors consider here is not the degree, but the “degree density” (cf. Donges et al., Phys. Rev. E, 85, 046105, 2012).

- p.9, l.33: !quantiles of the distribution of elements of the distance matrix” (also on p.10, ll.3-4)

- p.10, l.20: “of the one-class support vector machine”

- p.10, l.28: “that is fixed”

- p.11, ll.7-9: Temperature extremes represent strong deviations from the mean rather than “changes in the mean”.

- p.14, l.17: Do the authors mean “mean length of the vectors”?

- p.16, l.1: “that these findings” or “that their findings”

- p.17, l.22: “parameterise”

- p.18: It is a bit unusual to write the Conclusions completely in present tense. Maybe you wish to consider using present perfect here?

- p.18, ll.12-13: Maybe it is worth clarifying here again that the results apply for the considered types of anomalies?

- Figure A2: It would be interesting to see these charts detailed for the different detection algorithms (e.g. using different colors for the respective bars). Maybe the authors could add some corresponding figure as supplementary material?

- p.21, l.4: I suggest putting the two equations in brackets.

- The authors should check/revise/complete the following citations: Bintanja and van der Linden (2013), Faranda and Vaienti (2013) [remove publisher], Pfeifer et al. (2011) [capitalization of “Earth”], Pinheiro et al. (2016) [capitalization of “R”], Poincaré (1890) [incomplete reference], Smetek and Bauer (2007), van der Maaten (2009), Webber and Marwan (2015) [page numbers].

Interactive comment on Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-51, 2016.