Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-565-AC1, 2018


Hydrology and Earth System Sciences
Discussions

# *Interactive comment on* "Automatic design of basin-specific drought indexes for highly regulated water systems" *by* Marta Zaniolo et al.

**Marta Zaniolo et al.**

andrea.castelletti@polimi.it

Received and published: 15 February 2018

The study presents a new framework for determining basin drought indicators (target index) by coupling numerous models conditioned to select and weight hydro-meteorological variable states (predictors) in an automated fashion. The manuscript is topically of interest and relevance to HESS readers, generally well written, and logically presented. Most comments and suggestions provided request clarification in the manuscript, although some additional (minor) analysis is perhaps warranted. Comments below.

1. Introduction: What is the motivation for selecting these 4 objectives, other than 'common' or 'convenient'? Additional justification or rationale is warranted.



FRIDA procedure bases its objectives formulation on information theory, as suggested in Taormina et al., (2016). We consider the use of such objectives combination the most suitable to design an operational index, while providing informative insights about the dynamics driving drought evolution in the basin. In particular, the maximization of accuracy ensures a precise reproduction of the data, while the minimization of cardinality aims at simplifying the final models. These characteristics are key for an operational index, expected to balance precision and ease-of-use.

Relevance and redundancy objectives are instead an asset for an effective subset search process as they foster the diversification of the solutions explored within the MOEA algorithm, while guaranteeing low intra-subset similarity and high information content of the solutions. In particular, a predictor can be strongly relevant, when its removal from the input set causes a significant drop in the model accuracy; irrelevant, when its presence or absence from the input set does not affect the model accuracy; and weakly relevant, when there exists a combination (namely, a Markov blanket) of other predictors carrying analogous information about the target variable (Yu and Liu, 2004). An optimal subset is thus composed of strongly relevant features, and non-redundant weakly relevant features. A weakly relevant predictor is non-redundant when its Markov blanket is not included in the input subset. Depending on the problem at hand, various combinations of weakly-relevant predictors can exist, producing quasi-equally informative models, and requiring the optimization of relevance and redundancy objectives to be entirely identified (Liu et al., 2015).

Following the reviewer suggestion, we will clarify the rationale behind the objectives selection in the Methods and tools chapter, specifically in section 2.2.

2. Methods: Why is only f4 (accuracy) selected to discriminate among subsets (Fig 2, step 2)? Why this one and perhaps not others as well? Subsequently, all 4 objectives /assessment metrics are used for presumably final selection. Does this ultimately indicate that accuracy is the most important objective? Or somehow give it more weight?

The objectives of relevance and redundancy are essential to support the search process towards the finding of a diversified and comprehensive set of solutions, which will not be achieved optimizing cardinality and accuracy only. A two-objective search based on cardinality and accuracy would, in fact, identify optimal solutions, but at the same time disregard a number of quasi-equally informative subsets with an almost identical operational behavior. The identification of such alternative solutions, nevertheless, grants flexibility and multiple options for the expert-based choice of the preferred subset, where certain combinations of predictors can be favored according to case-specific operative purposes, such as a more robust or less costly data gathering process, or enhanced acceptability and immediacy of the index.
Once the search is completed, although, the specific relevance and redundancy score of each variable combination is rarely of interest for the design of an operational index, while its accuracy and cardinality are crucial. When the dataset of candidate variables presents significant redundancy and correlation among features, numerous subsets characterized by a wide range of cardinalities are generally available to achieve a relative small range of accuracies. This is often the case in environmental problems, where spatial and temporal correlation of hydro-meteorological variables and associated indicators is significant. For this reason, the accuracy metric is initially used to discriminate among subsets, in order to limit the number of solutions that undergoes a deeper examination to the highly accurate solutions, provided they feature different cardinalities and predictors combinations.
We will discuss this point in the Methods section of the revised manuscript.

3. Results: The target variable (supply deficit) requires a more clear description earlier in the manuscript. A later statement (p17, L421) indicates that agricultural demand is used to computer the target deficient, however there could be many definitions (deficit in reservoir storage, deficit in long-term groundwater levels, deficit in meeting total demand, etc.) Why is ag demand used?



In a report issued by the Jucar Hydrological Confederation (CHJ, 2007) the Ie is validated against the supply deficit with respect to agricultural and urban demand, the validation result is considered satisfying, and the index is declared operative. The choice of employing the same supply deficit as a target variable for the FRIDA index is thus required to ensure comparability between the two indexes. As the reviewer suggests, we will add a sentence clarifying this point earlier in the manuscript.

4. Results: A linear model is ultimately selected although a non-linear mode is recommended and compared. In terms of R2, there is little difference, however it may also be interesting to compare the weights given to each input/predictor. If there is a significant difference, this may not be intuitive (a statistical modeling artifact?)

The black-box nature of the non-linear ELM model does not allow an analysis of the weight given to each predictor. ELM belongs to a family of models called Artificial Neural Networks (ANN), typically employed in regression problems and labeled universal approximators. In ANN models, the input features are manipulated through one or more layers of multiple hidden neurons performing sigmoidal transformations. The optimal number of layers, and of neurons in each layer, is problem dependent and generally needs manual calibration. The neurons' output is then weighted and summed to compose the final output. As a result of the non-linear manipulations, it is not possible to determine the weights assigned to each predictor in a ANN regression, and a comparison with the linear model is not possible in these terms.

5. Results: other comparison metrics (between SI and FRIDA linear model) besides R2 may be warranted. How does the RMSE (or other) compare? Is it better to error above or below the target deficit?

We consider the point of the reviewer well taken. In the revised version of the paper

we will report additional metrics to aid the comparison between indexes. In particular we will include the Pearson coefficient and the RMSE alongside R2. As it is evident from the result reported below, both FRIDA indexes (linear and ELM) outperform the State Index quite significantly, while the different among them is negligible, although the non-linear index is always the top performing.
We consider erroring above or below the target deficit of equal importance as, on the one hand, the underestimation of a deficit value may find the water users unprepared to face a serious drought. On the other hand, the overestimation of drought conditions may ignite repeated false alarms that will compromise the index trustworthiness and its efficacy in triggering an alert state. Therefore, rather than penalizing an error above or below the target trajectory, we find more compelling to focus on errors in the most crucial drought situations i.e., at the maximum level of deficit recorded. In order to do so, we included two additional metrics: R4MS4E, penalizing errors in the deficit peaks, and a confusion matrix reporting the classification performance of critical droughts, arbitrarily defined as months reporting deficit values above the 85th percentile. The rows of the confusion matrix represent the instances in a predicted class while the columns represent the instances in an actual class. Consequently, the main diagonal reports the number correctly classified points. Outside the diagonal the errors are reported: the value in the first column and second row indicates a situation in which the index does not recognize an ongoing drought, while the value in the first row and second column indicates the number of false alarms. These additional metrics substantially confirm the previously obtained results, with the exception of the FRIDA-ELM confusion matrix, that seems to significantly exceed the competitors' performances by erroring only 0.57% of the times, as opposed to the 10,91% of Ie, and the 6,3% of FRIDA-linear.
We will add the newly computed metrics in the paper in the form of tables (Tables: 1, 2, 3, 4 ), as reported below.

6. Results: What are the FRIDA results using the exact set of 12 indicators included in the State Index? And associate weights? This may be useful for comparison (and discussion with water users.)



**Table 1.** Accuracy metrics

| Metric | State Index | Frida Linear | Frida ELM |
|---|---|---|---|
| Pearson | 0.8601 | 0.9506 | 0.9533 |
| R2 | 0.7396 | 0.9036 | 0.9074 |
| RMSE | 0.2066 | 0.1135 | 0.1014 |
| R4MS4E | 0.2549 | 0.1475 | 0.1299 |

**Table 2.** Confusion matrix State Index

| SI-deficit | critical drought | normality |
|---|---|---|
| **critical drought** | 131 | 18 |
| **normality** | 1 | 24 |

**Table 3.** Confusion matrix Frida Linear

| Frida Linear-deficit | critical drought | normality |
|---|---|---|
| **critical drought** | 138 | 11 |
| **normality** | 0 | 25 |

**Table 4.** Confusion matrix FRIDA-ELM

| Frida Linear-deficit | critical drought | normality |
|---|---|---|
| **critical drought** | 147 | 2 |
| **normality** | 1 | 24 |

A linear model calibrated with the whole set of State Index inputs produces a very similar result to the FRIDA model in terms of accuracy, as it is evident from the metrics reported below (Table 5. This result is not surprising if we analyze the weights assigned by the calibration to each input: all negligible except for the storage and piezometer predictors included in the FRIDA selected subset (see Table 6). The use of the 12-predictors thus seems to bring no advantage, as on the one hand it complexifies the model, the data retrieval process, and the index computation by adding unnecessary predictors; and on the other hand it compromises the tool's acceptability. The official SI, in fact, is the outcome of a participatory process where variables and weights were negotiated with stakeholders. In particular, the weights carry a specific physical meaning as they are proportional to the demand class associated to each partial Ie (as detailed in section 3 of the paper). Thus, redefining the input weights will invalidate the outcome of the participatory process, while providing no benefit from an operational and modeling viewpoint.

**Table 5.** Performance of the Linear Model calibrated with the whole set of State Index inputs

| **Metric** | **Linear Model** |
|---|---|
| Pearson | 0.9505 |
| R2 | 0.9013 |
| RMSE | 0.1188 |
| R4MS4E | 0.1499 |

7. Conclusion: The authors make mention of a changing climate. What does this mean for the reliability and accuracy of the framework conditioned on historical (relatively stationary) data? Please discuss.

In the short term, one can imagine that, despite a change in the drivers' statistics due to climate change, their interactions and relative role in causing a drought holds un-



**Table 6.** Weights assigned to each predictor of the Linear Model calibrated with the whole set of State Index inputs

| **Predictor** | **Weight** |
|---|---|
| **S1** | **0.826** |
| S2 | 2.04E-10 |
| F1 | 2.04E-10 |
| F2 | 1.96E-10 |
| F3 | 3.55E-10 |
| F4 | 2.06E-10 |
| Pl1 | 3.17E-10 |
| Pl2 | 5.72E-10 |
| Pl3 | 2.82E-10 |
| Pz1 | 2.66E-10 |
| **Pz2** | **0.174** |
| Pz3 | 2.37E-09 |

changed. In this case, the index formulation remains valid in the context a changing climate. In the long term, nevertheless, this hypothesis may cease to hold, thus requiring the reiteration of FRIDA procedure to identify new drivers and dynamics leading to a drought in the basin.

For example, in a groundwater dominated system as the Jucar basin, the piezometer information is likely to remain essential in a future climate, but, at the same time, we can expect evapotranspiration processes to increase their drought-propelling role, as climate change induces a general increase of temperatures. In other contexts, e.g., snow dominated catchments, the role of snow may lose priority due to a diminishing winter snowpack reserve.

FRIDA will thus represent a valuable tool to support a thorough analysis on the role of each driver in drought evolution under a changing climate.

We consider this a good point and we will expand the paper's conclusion to clarify the matter.

8. Discuss: If a subset of the 4 objectives are selected, or different objectives, how might this change the outcomes?

FRIDA procedure bases its objectives formulation on information theory, as suggested in Taormina et al., (2016). Information theoretic criteria (e.g., SU, Mutual information, and Partial Mutual Information) do not assume any functional relationship between the variables and thus result well suited to quantify the dependence between two variables in any modeling context (McKay, 2003). Other objectives formulations could be explored, for instance substituting the use of Symmetric Uncertainty with more traditional correlation coefficients, although with the risk of losing generality by assuming linear dependence between variables. The use of a subset of objectives could, in principle, be a viable option in case of a two-objectives search using accuracy and cardinality only. Such optimization will require less computational time, but on the other hand, will return a poorer set of solutions with respect to the four-objective search (see answer at point 3 for more details). We will include this comment in the Methods and tools section of the revised version of the paper.

9. Discuss: What influence may the selection of the learning machine, MOEA algorithm, etc. have on outcomes? Are they sensitive to choices or not?

The Extreme Learning Machine and Borg MOEA were selected as they were proven to perform well under a suite of different problems, ensuring applicability, scalability, accuracy and, in the case of ELM, very limited computational burden. ELM bypasses the time consuming gradient-based search of optimal neurons parameters required by traditional ANN techniques, by performing a random selection of hidden nodes, followed by the optimization of their output weights. Such optimization is solved through a one-step matrices product and essentially amounts to learning a linear model.
However, we do not expect the choice of the learning machine or MOEA to be crucial for the attainment of the result. A different benchmark MOEA (e.g., NGSAII, MOEAD,



eps-MOEA) is likely to come to a comparable result, although requiring a possibly significant effort in the manual calibration of the evolution parameters, which is automated in Borg MOEA. Similarly, other ANN techniques could in principle be substituted to ELM, although running the risk of incrementing the computational time to unbearable levels, given the multiple calibration and validation processes reiterated in WQEISS.
We will include this comment in the Methods and Tools section of the revised paper.

10. Discuss: Would the selected inputs/predictors change substantially is the target deficit were defined differently? It may not be overly surprising that reservoir volume and groundwater levels are most important for a target deficit focused on agricultural irrigation demand.

The selection of the target variable is indeed a critical step for the FRIDA procedure and requires an expert consultation to select the most appropriate target for the basin, and the operational aim of the index. We agree with the reviewer that the result of the variable selection step is not surprising given the basin climate and the physical meaning of the target variable. However, the aim of the paper was to demonstrate the validity of a completely automated procedure (i.e., that requires no information on system topology or basin characteristics) in recognizing the main drought drivers, and predicting a deficit with accuracy and limited computational effort. The selection of the case study was tailored to this purpose, as the Jucar basin successfully relies on a drought index to activate restraining measures. The analysis of the Jucar State Index provided guidelines for our work, firstly in terms of target choice and candidate variable retrieval, and secondly for validating FRIDA in both the variable selection step and index design outcome.
We thank the reviewer for bringing up this point, and we will specify the matter in the revised version of the paper.

11. What are the prospects for projecting out the State Index, based on the state of some features (e.g. reservoir volume) and predictions of other features (e.g.

precipitation or recharge)? This is hinted at in the very end of the manuscript, but may warrant more discussion.

Following the reviewer suggestion, we will expand the closing sentence of the paper to provide some clarification on the matter. We expect that the projection of a drought index fosters the adoption of proactive (as opposed to the current reactive) approach in facing a drought. Proactivity translates in a shift from costly and often belated mitigation measures, to preventive actions, thus granting flexibility to timely prepare to upcoming droughts, while reducing costs associated to drought impacts and restrictions.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-565, 2017.