



Liz Bradley <liz.bradley88@gmail.com>

---

## Fwd: Your\_manuscript ER11308 Garland

---

Joshua Garland <garland.joshua@gmail.com>

Wed, May 28, 2014 at 10:34 AM

To: Liz Bradley <liz.bradley88@gmail.com>, Ryan James <ryan.gregory.james@gmail.com>

What do you guys think?

----- Forwarded message -----

From: <[pre@aps.org](mailto:pre@aps.org)>

Date: May 28, 2014 7:35 AM

Subject: Your\_manuscript ER11308 Garland

To: <[joshua.garland@colorado.edu](mailto:joshua.garland@colorado.edu)>

Cc:

Re: ER11308

Quantifying time-series predictability through structural complexity  
by Joshua Garland, Ryan James, and Elizabeth Bradley

Dear Mr. Garland,

The above manuscript has been reviewed by two of our referees.  
Comments from the reports appear below.

The comments of the second referee suggest that the present version of the manuscript is not suitable for publication in the Physical Review. However, if you feel that you can refute the criticisms, we will give further consideration. Please accompany any resubmittal by a summary of the changes made and a brief response to all recommendations and criticisms.

Yours sincerely,

Margaret Malloy  
Associate Editor  
Physical Review E  
Email: [pre@aps.org](mailto:pre@aps.org)  
<http://journals.aps.org/pre/>

APS Partners with ORCID  
Link your identifier via <https://authors.aps.org/Profile>

---

Report of the First Referee -- ER11308/Garland

---

In this manuscript the authors present a way to explore the relationship between predictability and complexity across a broad (but

not complete) array of forecast strategies such as random walk (last value), naïve (average value), regression based (ARIMA) and a nonlinear method based on state space reconstruction (LMA). They use weighted permutation entropy as a criterion/index of predictive structure in the time series and illustrate their approach on eight real-world datasets derived from three different systems (computer performance traces).

Overall the manuscript is well written and clearly understandable. The analysis seems to be carried out carefully and the conclusions drawn are quite reasonable. While I don't get the impression that this is a breakthrough paper I can still recommend the manuscript for publication in PRE if the issues listed below are addressed thoroughly.

#### Major issues:

I don't get the reason for normalizing the three methods by the random walk method (for example in Table I). As the authors write themselves this makes the indicator vulnerable to a bias caused by the high influence of this (arbitrarily selected) reference method. Why not just use a non-referential estimator, which then would also allow to compare the inherent predictability of different systems (as it is now the values obtained for each systems are strongly influenced by the random walk estimate on that particular system and thus can not be compared). This will also change the results of the fitting (the authors explicitly exclude one value because of the effect caused by the normalization). And it is not that the random walk predictor is the one standard method that all other methods have to compete against.

In many parts the article reads almost like a review. There is a lot of unnecessary redundancy (the entropy rate is quite low). In particular, the main point ('The aim of the paper is not ... but ...') is repeated far too many times. The paper while easy to read could be shortened considerably without any loss of information.

#### Minor issues:

Page 1: Why mention the 'halting problem'? I don't see a straightforward connection to the problem at hand. The only connection is 'undecidability' but there are many other examples for that.

Page 10: "error between the predictions and the true continuations" --> "error of the predictions with respect to the true continuations" Figure 6 on Page 11: It would be more intuitive to turn this into a 'table of figures' with the four methods as row labels on the left and the three systems as column labels at the top. Also: Caption 'for forecast of ... and all four prediction strategies'.

First equation on Page 14: Shouldn't the index  $j$  run from  $i+1$  to  $i+l$

(instead of from  $i$  to  $i+1$ )? You normalize by  $I$  (not  $I+1$ ) values.

Page 19: "There are some exceptions: [...] Conversely, forecast errors for the col major and dgesdd5 signals are higher than the corresponding WPEs suggest—except for the nonlinear LMA predictor." This is not really an exception. As mentioned by the authors before, in general, the result only holds under the assumption that a reasonably good model has been used for the prediction. But for ill-matched datasets and models like in the cases mentioned above this is just not the case so it is not a surprise that the forecast errors are higher.

---

Report of the Second Referee -- ER11308/Garland

---

The problem of time series prediction is considered in the paper of J. Garland et al. Namely, the authors suggest to quantify predictability in advance (before using any concrete prediction scheme) based on the complexity estimate (weighted permutation entropy, WPE). They hypothesize that WPE is related to the best possible prediction error in a simple way. For several benchmark time series (reflecting variations of a computer performance during execution of different tasks), they obtain estimates of WPE and prediction error (the least one over four techniques), and build a linear regression using seven or eight resulting data points. They suppose that the obtained linear function can be used to "predict" the best prediction error from the model-free WPE estimate. Thus, the authors finally assume (as could be seen from the last section) that WPE and optimal prediction error for all systems are related via the same universal linear function with fixed coefficients. This is what I can understand from the entire paper, though the problem is not so definitely described in the Introduction.

I think that the possibility to estimate the best prediction error in advance from a relatively simple WPE characteristic would be interesting and useful, if it could hold true for a wide class of systems. Such a question would deserve a careful study as that started in the presented work. However, I have a number of critical remarks and strongly doubt that the conjecture of the authors can be valid for a reasonably wide class of systems (not even speaking of "all systems"). Other aspects of the paper are not so important, in my view, despite some of them are fresh and interesting (e.g. taking a computer as an object of prediction). My critical remarks are listed below.

1) Introduction is too long, but the main problem is not formulated definitely enough. I see that the main point is the approximately linear (or at least one-to-one) relationship between WPE and the best prediction error. This aim is not claimed explicitly in the Introduction. All other aspects of the manuscript do not seem important enough to be published in a separate article.

2) In particular, two primary findings mentioned on page 2 (Introduction) are formulated quite vaguely. The first part of the first finding reads "(i) complexity of a noisy real-value time series is quantifiable by permutation entropy". However, it seems obvious from the previous research, including that cited by the authors. The authors do not suggest PE or WPE here, they exploit them as previously known approaches. It was previously known that PE is an estimate of KS entropy as the authors also state. It was well known that KS entropy relates to predictability and, in this sense, it is a measure of complexity (see e.g. the work [G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani, "Predictability: a way to characterize complexity" // Physics Reports 356 (2002) 367–474] which is not cited by the authors). Thus, the first part of the first finding does not provide any new information. The second part of the first finding reads "(ii) complexity of the noisy real-valued time series is correlated with prediction accuracy of an appropriate predictor". However, it seems also obvious and directly follows from the definition of KS-entropy and its discussion in many papers including [Boffetta et al.] cited above.

3) By the way, the authors call "fully complex" the white noise process, i.e. they equate predictability and complexity. However, they do not even mention another approach to the notion of complexity which ascribes low complexity to the white noise [the works e.g. by C.R. Shalizi and J. Crutchfield carried out in the Santa Fe Institute from where one of the authors is]. It would be appropriate to mention that ambiguity of the notion of complexity if the latter plays an important role in many formulations of the authors.

4) The second finding reads "The way information is generated and processed internally by a system plays a crucial role in the success of different forecasting schema - and in the choice of which one is appropriate for a given time series". This statement also looks trivial. Indeed, an optimal predictor for a linear stochastic system and low-dimensional nonlinear deterministic system are quite different (an AR model versus a local or global nonlinear model) due to different properties of the original systems. Thus, there is nothing to be proven here. Probably, the authors imply their particular benchmark signals (reflecting a computer behavior) but they do not claim that. Thus, both primary findings seem to provide nothing new.

5) The benchmark system - a computer as an object of prediction - seems to be an unexpected and, probably, interesting choice but just as an additional illustration of a statement which should be first shown for well-controlled and well-understood paradigmatic systems (like low-dimensional maps or ordinary differential equations). Indeed, it is not known in advance what error one should expect for such a complicated object using a certain predictive technique, what technique is optimal or close to optimal, etc. If the main point of the research is the relationship between WPE and the best prediction error (if any), selection of such a complicated object as computer

performance variations only introduces its own difficulties into the problem.

6) The description of the choice of the benchmark system in the Introduction is too long. On the other hand, if the computer signals are the main interest of the work, then it should be claimed explicitly. Then, the paper should be rewritten and submitted to a kind of engineering journals where it might well be of interest.

7) As for the choice of the four prediction techniques, some of them seem superfluous. E.g. naive approach is just a global AR model of the zero order. Random-walk is similar to the AR model of the first order where the coefficient is not estimated by the least-squares but set equal to 1 (or this is an ARIMA model with the first difference and "zero ARMA part"). Inefficiencies of the ARIMA model as compared to those choices in the paper can represent an inappropriate procedure for the order selection rather than indicate inappropriateness of the entire ARIMA approach. Thus, linear ARIMA and nonlinear LMA might be sufficient (in combination with mathematical benchmark examples) to represent close-to-optimal prediction errors.

8) Conditions for inefficiency of simple prediction schemes (like naive and random-walk) discussed in the paper look rather obvious as well and do not give new information. Is it the purpose of the paper to discuss when and where a concrete simple prediction scheme fails?

9) A better justification of the prediction error metrics could be also in order if this work is continued. At least, comparison of prediction errors quantified with different metrics seems necessary. Especially, taking into account the following major concerns.

General remarks.

I) Strict relationship between WPE and prediction error applicable to all systems is hardly possible. In particular, there is a notion of epsilon-entropy which tends to the KS entropy if epsilon (the cell size) tends to zero (see e.g. [Boffetta et al.] cited above). For stochastic systems, the KS entropy is infinite and epsilon-entropy varies strongly with the scale epsilon. Seemingly, the permutation entropy estimate in this case will strongly depend on the choice of the word length  $L$ . The authors concentrate on  $L = 6$  and obtain coefficients of the regression between WPE and prediction error only for that choice and for their particular signals. However, other systems and other choices of  $L$  may easily lead to other regression coefficients so that deciding about the best possible prediction error from WPE seems impossible in general. It is good if such a decision is possible for a certain narrow class of systems. However, the authors do not even discuss to which class of systems their results apply.

II) PE is invariant under an invertible nonlinear change of variables. The prediction error measured as the authors suggest is not invariant. Thus, such a simple change of variables can lead, at least, to a

change in regression coefficient between WPE and prediction error. Thus, the authors should also discuss to which variable for that class of systems their result apply. I think that the fact that the authors observe a certain regression line at all seems to be due to small number of different signals they consider (only 7 or 8 different data points, such a small number of points may lie approximately near a straight line by chance).

To summarize, the paper shows that the authors are well familiar with the entire topic and highly qualified, they overview some well-known facts (WPE, various prediction techniques, etc) and "invent" an unexpected test system (computer performance data), they report about their attempts to play around with all that stuff. However, their results do not seem to provide any new information to a specialist. Probably, this work can be continued to get some more reliable results about concrete relationship between WPE and prediction error for certain classes of systems (however, I doubt that it is possible to derive any simple and widely applicable relationship like the linear function suggested by the authors). Seemingly, it should then be a new paper which can be hardly considered as a revised version of the present manuscript.