In section 2, mainly regarding Algorithmic information, Kolmogorov complexity and the encoding I have a few comments to improve the manuscript. The major drawback of Kolmogorov complexity is not being computable in the sense that it is not possible for a given string and a fixed Turing machine to compute its value. This is why it is inaccurate and depends largely on the compressor that one uses to approximate it.

Furthermore, since we are dealing with TM, in line 139 the expression "not easy to state" should be replaced by "impossible to state" as knowing whether a program stops for a given input is uncomputable.

Answer:
Thank you the reviewer for this suggestion. It is exact that the correct expression to use is «impossible to state». It has been changed in the text.

How exactly is the approximation of \( m \) do you consider? This needs to be properly explained for sake of clarification of the reader. Just saying that \( m \) can me approximated by any probability distribution is not enough. It must be computable and chose properly to feat the data.

Answer:
Thank you the reviewer for this question. First of all, the AP method is used here as a benchmark method to compare with the KC method extended to short time series. As mentioned in the text, we compute the complexity formula \( K(S) \approx -\log_2 m(S) \). We have given the reference about the operational use of this method in the text


Moreover as mentioned in the text, we have not done a huge number of Turing Machine runs to obtain the set of probability but we were using the probability distribution « In practice, we can use freely available probability distribution to compute the complexity of short binary string [13] »

Nevertheless, we agree that it could valuable to add some information about this method. In the new version (NV) of the paper we inserted the following sentence:

«To obtain the probability distribution for a string \( S \), we need to run huge number of randomly selected Turing machines. However, this computation is very challenging theoretically and numerically. In practice for this study, we have used freely available probability distribution to compute the complexity of short binary string [13]. There are available numerical tables with values of the algorithmic complexity dedicated for short strings. For more details on this subject, the reader either can find in [13] or consult an Online Algorithmic Complexity Calculator (OACC) following the web site link http://www.complexitycalculator.com.»
In subsection Kolmogorov complexity I think the order of the operations is reversed, first you apply the threshold and obtain a binary sequence which then is compressed with LZ. The introduction to normalized compression should be explained in the context of the paper, especially if you are using short binary strings.

Answer:

Thank you the reviewer for the comment. It is correct that there is some confusion about this sentence which means that we used Lempel-Ziv before the encoding process to binarized the time series. In fact, there is unfortunately a missing word « before » in the text. We have changed the sentence to:

« To compute KC of the hourly solar radiation time series \( x(t) \), we need firstly to encode the time series before using the Lempel-Ziv algorithm [7] in order to replace it by a new set of binary values such as: »

Please correct equation (1).

Answer:

Thank you the reviewer for the suggestion. According to the changes done in the previous sentence, equation (1) is then correctly written to show how we encode the time series to convert it into a binary time series.

The encoding presented in line 179 looks strange for me especially if you are concatenating the strings.

Answer:

Thank you the reviewer for the comment. Thank you the reviewer for this comment. The sentence is here to explain the step before the concatenation process. As mentioned in the text these are integer values which is conceptually represented by the following expression

\[ \{83|515|758|906|910|857|909|887|733|526|219|25\} \]

Then preserving this conceptual representation, we convert each integer number by its binary representation in a conceptual manner using same symbolic separator

\[ \{1010011|100000001|101110111|110000110|110101101|110111001|111000110|11011101|1101011|110000001|11011011|11001\} \].

Doing that, we can see with this representation the length of each binary string as motioned in the text (« The total length is simply the sum of the twelve individual binary strings respectively of length \{7, 10, 10, 10, 10, 10, 10, 10, 10, 8, 5\} »). This information about each binary string length can be used to recover the integer value which is not possible to do with the final binary string

\[ 1010011000000111101110110110001010111000111101111011001100111001111110111110111001111110111101101101111001. \]

We agree that the explanation given in the text of the previous version of the paper can introduce some confusion in understanding the encoding process between the conceptual and operational view. Therefore, in the NV, instead the aforementioned text we inserted the following sentence:

«These are integer values \{83|515|758|906|910|857|909|887|733|526|219|25\}. The vertical bar between integer values is used here as a symbolic separator to explain how we conceptually proceed
to encode the time series set. Each value can be coded into its binary representation to build the set of the twelve binary strings. This set of binary strings is conceptually a juxtaposition of twelve separated strings by a symbolic separator (vertical bar) as it is symbolically shown in the following way

\{1010011|100000011|1011110110|111000111|1101011001|1110001110|1101110 11|101101110|1000001110|11011011|11001|\}. Formally, the concatenation of the encoded values of hourly solar radiation into one binary string is significantly longer than the original integer’s values set. In this example, the length of the final binary string is of length 110 and represented by a long string which contains a set of successive of binary value without symbolic separator and bracket 101001110000001110111101101100010101110001110110101100111100011011101110 01110110000011101101111001. The total length is simply the sum of the twelve individual binary strings having the following lengths \{7|10|10|10|10|10|8|5\}, respectively.

Notice that if you use different size for representing the values in binary how would you recover the initial values from the encoding?

**Answer:**

Thank you the reviewer for the question. As explained above, for each integer value, we know the length used to convert into binary representation. This information can be used if we want to recover the initial values which is not the case using only the whole binary string alone to recover the separated initial values.

**Section 3**

In my opinion, section 3 is too extensive and should focus on the main aspects that are relevant for the analysis. For example it can be explained succinctly that the data used is related with solar exposition and there are climate factor that influence that value. The full details of it affect could be omitted for the sake of understanding and readability of the paper.

**Answer:**

Thank you the reviewer for the comments. First of all, this is an significant issue to deal with a subject which the use of algorithmic complexity to assess the reliability of such method to tropical solar radiation which is mainly triggered by cloud patterns. Our goal is to demonstrate how our method and AP method can catch solar variability and complexity over the tropical Indian Ocean. We have chosen the satellite which is the only available solar radiation with such spatial covering (South west Indian Ocean) and hourly temporal resolution. This data is freely available only for the 2004-2006 period(http://www.soda-pro.com). The two previous studies [8] and [9] mentioned in the paper are studies about daily solar radiation and ground-based observation over a small island. This study is about application of Kolmogorov complexity on hourly solar radiation variability over a regional area. We have chosen longitudinal transect which is representative for the marine equatorial and tropical band. The South West India Ocean is an area in which there are numerous meteorological activities at different time and spatial scales. In our opinion, it is relevant to assess the ability of the both methods (KC-ES and AP) to catch the complexity of three of the main tropical well-know meteorological perturbations prevailing over the southwest Indian Ocean and which are well-known to affect the weather and thus the solar radiation received on the ground. AP is a benchmark method which has never been used for hourly solar radiation. The goal of study is to demonstrate that the AP and KC-ES
methods are robust methods to catch the huge complexity range of solar radiation over a regional area.

It’s why we have presented the latitudinal solar radiation variability along the four longitudinal transect (Figure 2) and the variability of daily variability of solar radiation within three latitude band (Figure 3) and finally the infra-day solar radiation for the selected longitude and latitude band (Figure 4). We thought that these figures significantly help the reader to figure out how it is challenging to catch the variability and complexity of solar radiation in space and in time over a tropical regional scale. Moreover, we have showed one another satellite observation of convective activity which gives a physical link between solar radiation, cloud cover associated with tropical meteorological perturbation (MJO, IOD).

Thus, we agree that this section could be a little bit specialized but we thought at conversely to the both previous studies [8] and [9] on Kolmogorov complexity of daily solar radiation over small island which is an area of 2 512 km² and this study which about the complexity of hourly solar radiation over the South-West Indian Ocean 43°E- 62°E/ 30°S -15°N (8, 550 000 km²)

*The new scheme presented by the authors to analyze the short time series is hard to identify in the paper and therefore it is difficult to understand the main difference from this approach to the previous one [8] and [9].*

**Answer:**

Thank you the reviewer for the comments. As mentioned, the previous studies [8] and [9] have demonstrated the ability of the Kolmogorov complexity method to give a measure about complexity of daily GHI solar radiation over a small island. The use of Kolmogorov complexity on hourly GHI solar radiation has never been done before this study. With the second study, we have also assessed the reliability of this method to estimate the influence of the interannual oscillation on the predictability of the solar radiation inside the island. It is a local-scale study about daily solar radiation. This study is focusing on the regional scale hourly GHI solar radiation and how its complexity is triggered by the major well-known intraseasonal perturbations prevailing in this area (IOD, MJO) and the regularly extreme meteorological events occurring during austral summer season over the South West Indian Ocean (tropical cyclone) which have not been done in both previous studies for such an area and time scale.

Section 4

*In Section 4 the figures are too big and should be rescaled. Furthermore, there is a repetition of the text to describe again the climate influences on light exposure. I have a comment on line 399 and the respective figure. In Figure 6 the red line does not represent a Gaussian distribution as it does not have tail values. This must be revised and presented properly. Also, the importance of this picture as well as Figure 7 for the presentation of the results must be clear for the reader. Caption of figure 6 cut in two pages is terrible. The reader cannot distinguish caption from text.*

**Answer:**

Thank you the reviewer for the comments. We agree that the figures are too big. We have rescaled them in a smaller size to be much more readable for the reader.
To avoid the repetition in the text about the climate influences on solar radiation, we propose inserting the following sentence at the beginning of the paragraph:

*Hourly solar radiation variability during each daytime is mainly triggered by weather and directly by cloud cover.* Hourly clear-sky index is used here to evaluate the atmospheric attenuation due to clouds and their impact on the stochasticity of the fluctuations [34].

We think that the comment about the line 399 is a mistyping number and maybe the reviewer would like to mention line 339 which sounds the most appropriate line to the following comments.

We agree that Figure 6, the red line represents the PDF of Gaussian distribution. Normally, if we represent, for example, the probability density function of time series which we suppose as Gaussian, we must have two tails which represent the rare event of extreme values. Here the purpose is a little bit different, as in turbulence field, we compute the pdf not about the GHI value but the pdf distribution of the normalized GHI increments for different time lag. The definition of the pdf of GHI increments has been referenced in the text.

« The departure from the Gaussian distribution is an indicator of the intermittency [9] »

Generally, in such representation of the Gaussian distribution is not a profile with tail but a reverse parabola as shown in Figure 6. Departure of the GHI increment distribution from the Gaussian gives an indication of the existence of the intermittency in the variability of GHI solar radiation which is the case here for the large GHI increments.

As the title of the chapter is « *Hourly solar variability and algorithmic probability method* »

Pursuing Figure 6 which demonstrates the existence of intermittence in the hourly GHI solar radiation, Figure 7 is here to show how AP method catches the solar radiation variability. The main point is that AP method is built with the concept of algorithmic probability and theorem code for short string to catch the physical feature. It is done with the CV and shown in Figure 2 as mentioned in the text. We agree about the importance of this Figure 7 too, thus we propose inserting the following sentence in the NV to improve the presentation of the results and more appropriate link with Figure 6.

« In addition, beyond the intermittency of hourly clear sky index, the question is how complex is the variability of solar radiation during day time and how it could be catch by the AP method. Over the tropical band 30°S-15°N, the length of the day time is in average of 12h. It means that for hourly time sampling, the data set length per day is short. Thus, computing KC is not possible for time series consisting of only twelve values. In this case, the use of the coding theorem (CT) is useful to compute complexity using the concept of AP method [13, 35].

As mentioned previously, complexity of a binary string is deduced from its probability distribution produced by the huge number of Turing machines run [12, 33]. Before computing the frequency distribution, we have binarized for each day the hourly clear sky index by using the mean as a threshold value [8, 9]. Figure 7 depicts dependence of hourly solar radiation complexity on latitude. South to 5°S, the complexity is roughly constant for the 52°E-62°E longitude band. »

*Many claims are not proven, calculated or are given any insight of their validity. For example:*
1) In line 362 the authors claim that the decrease of complexity is explained by more complex flows. How can this be? If the flows are more complex one would expect that their description of their effect to be increasing the complexity. This also needs more explanation.

2) Line 377. "Thus, AP method and statistical variability are well positively correlated which means that computational complexity AP is in accordance with physical variability." Why is that?

Furthermore, how can you deduce that the results observed can be explained by the semi-annual cycle? Why is that?

Answer:

Thank you the reviewer for the comments.

1) We agree that it could be some confusion using words “complexity” and “complex” in the same context although they have not the same meaning in the sentence. Along the equatorial band and the western part of the Indian Ocean the resulting atmospheric flow over this area is a complex interaction between monsoon flow coming from the northwest, the obstacle flow around the big island Madagascar and the ITCZ. Thus, this area is often a very cloudy area with a lot of convection. The variability of the weather during the day is less than other area as shown in Figure 2 with the coefficient of variation. Thus, the algorithmic complexity of the hourly time series of solar radiation is less complex due to the predominance of regular persistent cloudy days over the equatorial band. In the NV, we propose the following revised sentence:

“Northward to the equator, there is an eastward decrease of complexity which could be explained by the persistent cloudy days induced by the interaction flows northward Madagascar as monsoon and trade wind splitting interaction producing ITCZ (Inter Tropical Convergence Zone). Moreover, complexity is higher over land (horn Africa) than over ocean as it is also seen in Figure 2 with coefficient of variation.”

2) We assess with Figure 8 that AP method can catch the physical variability of the hourly solar radiation observed with CV which is a classical statistical measure of variability in analysis of time series. Generally, the CV value is representative of the amplitude of the variability inside a time series in regard to the mean of the time series. Figure 8 shows that AP method gives a measure of complexity which is positively correlated with another measure of variability. The link with the physical aspect is that CV is computed using the hourly GHI solar radiation value which is directly connected to the physical observation whereas AP value is computed using Turing Machine concept applied on solar radiation time series. It means that using two different measures we have the same temporal behavior and physical aspect is related conceptually to the solar radiation parameter. We agree that it needs more explanation in the text. In the NV, we propose the following revised sentence:

“Thus, AP method and statistical variability are well positively correlated which means that computational complexity AP is in accordance with CV which is a statistical measure of the solar radiation variability. The time variation of the hourly solar radiation can be tracked by the computational complexity AP method.”

3) As mentioned in the text, we use as an example the seasonal and semi-annual cycle of solar radiation to show the ability of AP method to catch the solar radiation variability. The annual and semi-annual solar radiation is well observed through latitude band. Due to astronomical consideration, the semi-annual cycle of solar radiation is well pronounced equatorward and less
pronounced poleward. Conversely, the annual cycle of solar radiation is well pronounced poleward and less pronounced equatorward. These two cycles are the most visible variability in the solar radiation temporal variability. It’s why we have choose as an example these two cycles to emphasize the ability of the AP method to track this seasonal and semi-annual variability as done with CV parameter. Nevertheless, we agree that there is a need to give more explanation. In the NV, we propose the following revised sentence:

4) «As it is known, the annual and semi-annual are the most dominant time mode of solar radiation on the earth. There is a latitudinal dependency of these two modes. The semi-annual mode is most pronounced over the equatorial band whereas the annual mode is most pronounced over the tropical and extratropical bands. Both CV and AP method exhibit the annual mode south to 15°S (Figure 8a) and semi-annual mode over the equatorial band (Figure 8b and Figure 8c). For the former, it is in accordance with Bessafi et al. [9] for La Réunion (21°S/55°E) located in the South tropical band.»

Analyzing figure 8 it seems to me that AP and CV are correlated but this is some evidence that is already presented in the previous works. Why this is presented here? The focus of the paper should be comparing the new approach to this one, not giving evidences that AP is good to characterize CV.

Answer:

Thank you the reviewer for the comments. We agree that the AP and CV are correlated. In the previous study [8] we also have computed the coefficient of variation of the Kolmogorov time (inverse of Kolmogorov complexity) for daily solar radiation for different window size. Here, the coefficient of variation CV is computed with the original hourly solar radiation time series. The AP method has not been used in the previous study and the correlation computed in this study is between the coefficient variation CV of solar radiation whereas in the previous there was no correlation analysis similar to one which is done in this study.

We are agree that the focus of this paper is to compare the new approach KC-ES with AP method but the use of AP method to hourly solar radiation has never be done before. Thus, it is relevant to check how the AP method, which is used as a benchmark, can catch the hourly solar radiation complexity before comparing the new approach KC-ES with AP.

In section 4.2.1 the authors compare AP with KC-ES and claim that they have similar shape (graphically visible) but the variability is twice for KC-ES than for AP, but this is not clear from the text nor the graphic. Furthermore the scale of the two graphics in Fig 11 are different (~30 for AP and ~1 for KC-ES). The same issue arises in Figure 14. I do not see why the phenomena is important in the comparison between the two methods.

Answer:

Thank you the reviewer for the comments. We agree that the AP and KC-ES have not the same scale value. The scale between the two methods is different because we use for KC-ES a normalization process following the compression step. It is explained in the previous study [are correlated and in a previous [8], [9] and other works [10]. Concerning AP method, we have mentioned that the complexity values are available through numerical tables which give another kind of complexity measure. We agree that it must be much clearer in the text. In the NV, we propose the following revised sentence:
«In addition, same complexity pattern is highlighted by KC-ES complexity and AP method. The scale value of AP and KC-ES are two different measure of complexity. The former is obtained through numerical tables of probability distribution and the latter is a normalized value computed after the Lempel-Ziv process [8, 9, 10]. We can notice that over Mozambique Chanel and Africa Horns (Longitude 43°E), the variability of complexity is highest for both AP and KC-ES methods than the remaining longitude transect (52°, 57°E and 62°E).»

Regarding the Figure 14, we agree that the scale is different between the two methods and as we mentioned these two methods give different measure of complexity. We show that the variability of the complexity of hourly solar radiation calculated by AP or KC-ES during tropical events is embedded within a higher broadband range of complexity induced by a huge variety of solar radiation variability triggered by numerous cloud patterns and convective activities. We agree to emphasize this aspect of scale in the text. In the NV, we propose the following revised sentence:

«On the right side, we have plotted the complexity only for days where a tropical cyclone track crosses the 30°S-15°S/43°-62°E area during austral summer season 2004-2005 and 2005-2006. Despite the different scale, between AP and KC-ES methods, the variability of the complexity during tropical cyclone events is embedded inside the natural complexity of clear sky index prevailing during days without tropical cyclones events over the western part of the Indian Ocean.»

In the conclusions the authors suggest why the analysis was performed in the strong meteorological events but I do not understand why this useful. I would expect to use this approach to identify possible occurrences of this phenomena or to predict day light exposure.

Answer:

Thank you the reviewer for the comments. We agree that this study is performed using strong meteorological events. As mentioned, IOD, MJO and tropical cyclones are the major meteorological perturbations prevailing over the Indian Ocean and especially the southwest area. We think that it is relevant to assess AP and KC-ES methods through such event especially if we intend to have a measure of the predictability of hourly solar radiation over area where there are such often phenomena. We agree to mention this aspect as a perspective in the text. We propose the following sentence in the concluding part:

«This measure of complexity could also be used as an index in day-by-day hourly solar radiation classification and predictability.»

Summing up, the paper is to extensive to support the idea of how to use KC meteorological data. I do not fully understand the goal of the paper and why in fact this new approach is better than the ones presented in [8] and [9]. Short-time series is indeed a good motivation but in my opinion the paper fails to properly present that motivation and why this is really useful in this context.

Answer:

Thank you the reviewer for the valuable and helpful comments and questions. We expect that our answers, our suggested additional and removed sentences or figures in the text are in accordance with the requested changes by the reviewer.