Dear anonymous Referee1:

I am very happy to receive your recommendation and very grateful for your advice. We have followed your comments to revise this manuscript. Then, due to the stupid organization and poor English make readers' understand difficulty, we have made efforts to revise and hope that you could be satisfied. In the resubmitted paper, new text is emphasised as red text. The Referee Comments is abbreviated to “RC”, and the Authors’ Response is abbreviated to “AR”.

The following are the responses to each major comment:

RC 1:

The language used in many cases is really bad and confusing to the reader. Please take careful care of the syntax and rewrite the manuscript where needed. I have also pointed out several cases in the attached pdf file.

AR 1:

I agree with the advice and have revised this problem in my manuscript. We will call for a professional company to polish the manuscript before formal publication.

RC 2:

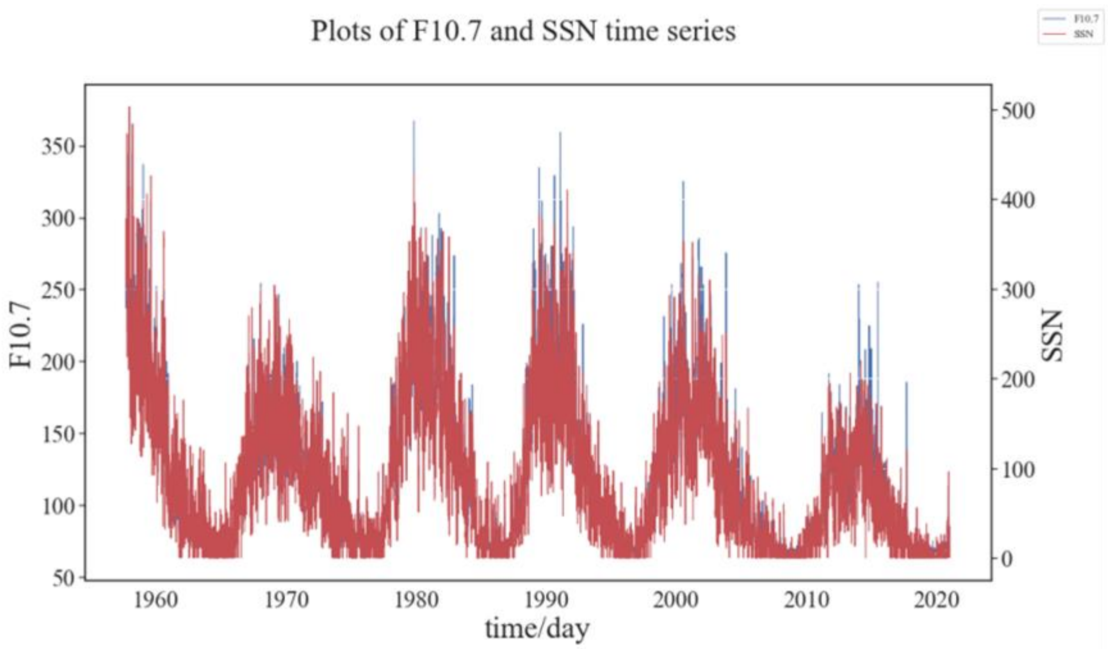
The authors use approximately 4 solar cycles for the training of the ML scheme and 1 solar cycle (solar cycle 24) as a test dataset. Even though this is a pretty usual technique to validate a model, it can potentially lead to significant misconceptions. This is because the solar cycle 24 is quite weak compared to previous cycles (this is also something that is not discussed in the text at all). A more robust technique would be to use an iterative leave-one-out method, which is described in detail in Aminalragia-Giamini et al. 2020 (https://doi.org/10.1051/swsc/2019043). A suggestion could be that the authors leave iteratively one solar cycle out as a test dataset and rerun the model each time (e.g. keep SC23 as test dataset and train the model with the rest SCs, then keep SC22 as test dataset and train the model with the rest SCs, etc.). In the end, they can evaluate the metrics (MAE, RMSE, etc.) using the predictions of all solar cycles

AR 2:

I agree with your advice. Solar cycle 24 is relatively weaker compared to previous cycles, which may lead to smaller errors. Both the control group of our chosen TCN model (forecast results from SWPC and AR models) and the test dataset are within the solar cycle 24. Figures 6 and 7 show that the predictions are not as good as the TCN model. Similarly, we selected the BP model and LSTM model forecasts for the high solar activity year (2003-2004) in solar cycle 23 for comparison. The results show that the TCN model has better forecasts than the BP and LSTM models (as shown in Table 4). However, we didn't consider whether choosing a different dataset would result in a huge difference in model performance. Therefore, we took the advice you gave and used the leave-one-out method to select the test set. The results of the tests are shown in the table 1.

Table.1The prediction errors (MAE, RMSE) and R of the TCN model for the F10.7 data during different solar cycles.

Solar activity cycle	1-Day ahead			2-Days ahead			3-Days ahead		
	MAE (sfu)	RMSE (sfu)	R	MAE (sfu)	RMSE (sfu)	R	MAE (sfu)	RMSE (sfu)	R
19	4.35	9.03	0.9880	4.29	7.84	0.9908	4.42	8.51	0.9897
20	3.35	5.16	0.9924	3.86	5.76	0.9928	3.40	5.37	0.9926
21	4.59	7.51	0.9921	4.48	7.16	0.9927	4.65	7.45	0.9930
22	4.71	7.89	0.9908	5.36	8.57	0.9908	4.75	8.05	0.9903
23	3.76	6.46	0.9917	4.30	7.01	0.9915	3.91	6.73	0.9912
24	3.03	5.60	0.9846	2.78	5.49	0.9833	3.23	5.52	0.9850
Mean	3.97	6.94	0.9899	4.18	6.97	0.9903	4.06	6.94	0.9903



Combined with the variation of sunspot number vs. F10.7 in the above figure. Cycles with stronger solar activity are found to have relatively poor model forecast errors. For cycles with weaker solar activity, the results are relatively better. Solar cycles 20 and 24 have about the same intensity of solar activity and are both weaker. The model forecasts are relatively better, Solar cycles 21 and 22 have about the same intensity of solar activity and are both stronger. The model forecasts are relatively poorer. However, the overall average prediction results do not change much compared to solar cycle 24. Therefore, the TCN model does not affect the final F10.7 forecasts due to the specific properties of the data. I think this is an excellent suggestion. We have revised this problem in my manuscript. You can see more detailed information **in lines 189-199.**

RC3:

Line 72, the"processed data".What do you mean by processed? If you have indeed processed the dataset used you have to explain how.

AC3:

Line 72, I am sorry for not explaining the process of processing the data, I would add the following. “The 10.7 cm daily solar flux data were obtained from the website of the National Oceanic and Atmospheric Administration. Three flux determinations are made each day. Each 10.7cm Solar Flux measurement is expressed in three values: the observed, adjusted, and URSI Series D values(absolute values). The observed value is the number measured by the solar radio telescope. This is modulated by two quantities: the level of solar activity and the changing distance between the Earth and Sun. Since it is a measure of the emissions due to solar activity hitting the Earth, this is the quantity to use when terrestrial phenomena are being studied(Tapping, 1987). When studying the Sun, it is undesirable to have the annual modulation of the 10.7cm Solar Flux caused by the changing distance between the Earth and the Sun. However, during the ephemeris calculations required for the solar flux monitors to accurately acquire and track the Sun, one of the byproducts obtained is the distance between the Sun and the Earth. Therefore, we generate an additional value called the adjusted value, which takes into account the variations in the Earth-Sun distance and represents the average distance. Absolute measurements of flux density are quite difficult. Astronomers attempt to match the solar flux density data at various frequencies with a frequency spectrum by applying a scale factor. By combining each wavelength with the calibrated spectrum, a series of D Flux is obtained, where D Flux equals 0.9 multiplied by the adjusted flux(Tanaka et al.,1973).

Between March and October measurements are made at 1700, 2000 (local noon) and 2300UT. However, the combination of location in a mountain valley and a relatively high latitude makes it impossible to maintain these times during the rest of the year. Consequently, from November through February, the flux determination times are changed to 1800, 2000, and 2200, so that the Sun is high enough above the horizon for a good measurement to be made. Therefore, we chose the adjusted flux value of F10.7 measured at 8:00 p.m. UT (local noon).”I agree with your advice and have revised this problem in my manuscript. You can see more detailed information **in lines 69-90**.

Lines 11-14, the sentence“The F10.7 series from 1957 to 2019 are used, which the datasets from 1957 to 2008 are used for training and the datasets from 2009 to 2019 are used for testing. The results show that the TCN model of prediction F10.7 with a root mean square error (RMSE) from 5.03 to 5.44sfu and correlation coefficients (R) as high as 0.98 during solar cycle 24.” is replaced by“The F10.7 series from 1957 to 2019 are used. The data during 1957–1995 are adopted as the training dataset, the data during 1996–2008 (solar cycle 23) are adopted as the validation dataset, and the data during 2009–2019 (solar cycle 24) are adopted as the test dataset. The leave-one-out method is used to group the data set for multiple validations. The prediction results for 1-3 days ahead during solar cycle 24 have a high correlation coefficient (R) of 0.98 and a root mean square error (RMSE) of only 5.03~5.44 sfu.” The sentence has been revised **in lines 12-16.**

Line 28, we have added more recent references to support this idea. The cite “(Swarup et al., 1963; Tapping and DeTracey, 1990; Henney et al., 2012)” is replaced by “(Katsavrias et al.,2021; Simms et al.,2023).” The citation has been revised **in line 31.**

Lines 29-30, the sentence“Time-series data is data where observations of some process are recorded over the same time interval, and the F10.7 index is a typical type of time-series data" is removed.

Line 55, the unit "sfu" is the unit of solar flux (F10.7). The quantities are separated by commas. The 10.7cm Solar Flux is given in solar flux units (a sfu = 10^{-22}W m^{-2} Hz^{-1}). We have revised this problem in my manuscript. You can see more detailed information **in line 69.**

Line 59, the"TCN" has already been mentioned in line 10, and where a proper noun is mentioned before, it will be used as an abbreviation afterwards.

Line 62, the "RNN" means Recurrent Neural Network. I apologise for not giving a definition. I agree with your advice and have revised this problem in my manuscript. You can see more detailed information **in line 62.**

Lines 67-71, the words"radiation"and "download", the sentences"F10.7 is one of the longest-running indices that records the level of solar activity", " from this URL", "from 1957 to 2019 are used, which the datasets" and "and the datasets" are removed. I agree with your advice and have revised this problem in my manuscript. You can see more detailed information **in lines 69-90.**

Line 75, the word "value" is replaced by "values" This error has been corrected **in line 91.**

Line 77, I think this table is important, so I'm going to give an explanatory note about the contents of the table. The addition is as follows: "The parameters related to the hardware and software environment for this experiment are shown in Table 1. We build a model through Python and utilize some efficient frameworks including Pandas, Matplotlib, Tensorflow, and Sklearn. Pandas supports us to complete data processing and Matplotlib supports us to display graphics. Tensorflow and Sklearn are essential frameworks for building various prediction models." You can see more detailed information **in lines 94-97.**

Lines 80-81, the sentence "some scholars believe it will replace RNN as the king of the temporal prediction field" is replaced by "Some scholars has demonstrated that TCN not only achieve better performance but also reduce the computational cost for training, compared to that of RNN (Lea et al.,2016; Bai et al.,2018; Dieleman et al.,2018)" As suggested by the reviewer, we have added more references to support this idea(Lea et al.,2016; Bai et al.,2018; Dieleman et al.,2018) The sentence has been revised **in lines 101-103.**

Line 81, the "RNN" means recurrent neural network and the "CNN"means convolutional neural network. I am sorry for not giving a definition. I agree with your advice and have revised this problem in my manuscript. This error has been corrected **in lines 103-104.**

Line 86 the sentence “Because of its a long sequence can be treated as a whole in TCN” is replaced by “Long input sequences can be processed as a whole in TCN.” The sentence has been revised **in lines 108-109.**

Line 115, the sentence “Figure 3: Expanded causal convolution” is replaced by “Expansion causal convolutional structure diagram.” TCN uses a one-dimensional convolutional network consisting of the expansion causal convolutional and residual modules. Figure 3 represents the expansion convolution of the TCN model, the principle can be found in Bai et al(arXiv:1803.01271v2). The sentence has been revised **in lines 137-138.**

Line 131, the word “Relu” is a commonly used activation function in deep learning and is a nonlinear function. It is defined as $f(x) = \max(0, x)$. I agree with your advice, and have revised this problem in my manuscript. You can see more detailed information **in lines 155-156.**

Line 139, The word“SWPC” is mentioned in line 15 of this article, so the abbreviation is used here. The URL for SWPC is https://www.swpc.noaa.gov/sites/default/files/images/u30/F10.7%20Solar%20Flux.pdf. I agree with your advice and have revised this problem in my manuscript. You can see more detailed information **in line 227.**

Lines 171-172, Fig.5 is intended to represent the difference in the performance of the TCN model in high solar activity years versus low solar activity years. To indicate that even in high solar activity years the TCN model can predict well. Especially during the peak of F10.7, the TCN model's predictions align well with the actual values, and it performs exceptionally well during periods F10.7 of high solar activity. The prediction performance of 1-3 days ahead for each year in solar cycle 24 has been shown in Table 3. Therefore, we do not think it is necessary to show the forecast charts for each year 1-3 days ahead again here. You can see more detailed information **in line 223.**

Line 193, I am very sorry that Figure 6 does not give specific information. Replace the specification of Figure 6 with the following: “Comparison of the prediction performance of SWPC and TCN. Panel (a) is a comparison of the prediction performance of SWPC and TCN 1-day ahead. Panel (b) shows the performance comparison between SWPC and TCN 2-days ahead. Panel (c) shows the performance comparison between SWPC and TCN 3-days ahead.” The sentence has been revised **in lines 246-248.**

Line 203, the sentence “Figure7: Comparison of the prediction performance between AR and TCN” is replaced by “Comparison of the prediction performance of AR and TCN. Panel (a) is a comparison of the prediction performance of AR and TCN 1-day ahead. Panel (b) shows the performance comparison between AR and TCN 2-days ahead. Panel (c) shows the performance comparison between AR and TCN 3-days ahead.” I agree with your advice and have revised this problem in my manuscript. The sentence has been revised **in lines 257-259.**

You can see the detailed changes in the resubmitted manuscript. If you have any problems, please contact me immediately. I am very grateful for your comment. Thank you very much.

Best Regard
LuYao Wang
The 1st author of this manuscript

Dear anonymous Referee2:

I am very happy to receive your recommendation and very grateful for your advice. We have followed your comments to revise this manuscript. Then, due to the stupid organization and poor English making readers understand the difficulty, we have made efforts to revise and hope that you will be satisfied. In the resubmitted paper, new text is emphasised as red text. The Referee Comments is abbreviated to "RC", and the Authors' Response is abbreviated to "AR".

The following are the responses of each major comment:

RC1:

L. 29-30: I disagree with the author's definition of a time series. A time series does not necessarily have a fixed interval between points.

AR1:

L. 29-30:The sentence"Time-series data is data where observations of some process are recorded over the same time interval, and the F10.7 index is a typical type of time-series data" is removed. I agree with your advice, and have revised this problem in my manuscript.

RC2:

L. 33-35: The authors enumerate several institutions that are presumed to be the primary sources for F10.7 forecasts. Instead, the authors could simply list and reference the forecast models, as they do shortly afterward. Alternatively, if they wish to mention institutions offering operational F10.7 forecasts, they should rephrase their sentence to eliminate any implication of ranking. Without a credible source, this ranking lacks scientific validity and holds no relevance in a scientific article.

AR2:

I would like to mention the organizations that provide F10.7 operational forecasts. But to avoid unnecessary ambiguity .The sentence "The forecast models of F10.7 are mainly time series models, and the main research institutions include the SWPC, the US National Oceanic and Atmospheric Administration (NOAA), the National Space Science Center of the Chinese Academy of Sciences and the National Astronomical Observatory of China, etc." is removed.

RC3:

L. 58: The authors should provide a reference to back up their claims about RNNs (which, as far as I know, are correct).

AR3:

I am agree with your advice, the added references are Zachary et al (2015) and Yang et al(2021). I have revised this problem in my manuscript. You can see more detailed information **in line 59.**

Zachary C. Lipton, John Berkowitz, Charles Elkan.: A Critical Review of Recurrent Neural Networks for Sequence Learning. arXiv:1506.00019,2015.

YANG H J，SUN Y Q，ZHU W， et al.: Prediction method of dissolved gas concentration in transformer oil based on CEEMD-TCN model［J］.Electronic Devices，44(4) : 887-8922021，2021.

RC4:

L. 59-60: I fail to see how the TCN's capacity for parallelizing calculations could be a result of reading data faster (in fact, the causal relationship might be the reverse). If the authors intend to highlight that TCN trains faster than an LSTM, they should consider rephrasing their sentence.

AR4:

L.59-60: I intend to highlight that TCN trains faster than an LSTM .The sentence "The TCN model can read data at a faster rate and therefore has a strong capability of parallel computation" is replaced by "TCN uses convolutional operations for efficient parallel computation." The sentence has been revised **in lines 60-61.**

RC5:

Section 2.1: To complete their presentation of the F10.7 index, we suggest that the authors include a histogram of the distribution of possible F10.7 values, along with an autocorrelation plot. This would likely be very helpful for readers who are not very familiar with the study of this index.

AR5:

I have restated the data description section. I believe that a more detailed description would be helpful to readers unfamiliar with the F10.7 index study. In addition, the article has shown plot of having F10.7 over time. Therefore, I did not choose the F10.7 autocorrelation diagram. You can see more detailed information **in lines 69-90.**

RC6:

Section 2.1: The authors only use a train set and a test set, that is also used for validation. I strongly recommend that the authors use a split into three sub-sets: training, validation and testing. Indeed, when the same set is used for both validation and testing, it can lead authors to select an architecture and model hyperparameters that yield optimal results on this particular set. This approach is occasionally adopted when the dataset is too small to be divided into three adequately sized subsets, but this is not the situation here.

AR:6

I am agree with your advice, and have revised this problem in my manuscript. Divide the data into three parts: training set, validation set, and test set. The data during 1957–1995 are adopted as the training dataset, the data during 1996–2008(solar cycle 23) are adopted as validation set and the data during 2009–2019(solar cycle 24) are adopted as test set. I have revised this problem in my manuscript. You can see more detailed information **in lines 87-90.**

RC7:

Section 2.1: The authors use solar cycle 24 as their test case. This solar cycle is known as a very low-activity solar cycle. It would be more interesting to have test results on another cycle, such as cycle 23. To achieve this, the authors could implement cross-validation or, at the very least, train the model using a separate data split distinct from the initial one. It appears that the authors used the years 2003 and 2004 for testing as well, but it is unclear if they used the results from the training set, or if they trained a different model (see comment 24).

AR7:

I am agree with your advice. Solar cycle 24 is relatively weaker compared to previous cycles. We retrained the model again. Solar cycle 23 is used as the validation set, and solar cycle 24 is used as the test set. The results obtained for solar cycle 23 is shown in table 1.We didn't consider whether choosing a different dataset would result in a huge difference in model performance. Therefore, we used the leave-one-out method to select the test set. We leave iteratively one solar cycle out as a test dataset and rerun the model each time (e.g. keep solar cycle 23 as test dataset and train the model with the rest solar cycles, then keep solar cycle 22 as test dataset and train the model with the rest solar cycles, etc.). The results of the tests are shown in the table 2.

Table 1 .The prediction errors (MAE, RMSE) and R of the TCN model for the F10.7 data during 1996–2008.

Year	**1 Day in Advance**			**2 Day in Advance**			**3 Day in Advance**		
	MAE (sfu)	**RMSE (sfu)**	**R**	**MAE (sfu)**	**RMSE (sfu)**	**R**	**MAE (sfu)**	**RMSE (sfu)**	**R**
1996	1.20	1.69	0.9530	1.01	1.65	0.9525	1.04	1.57	0.9540
1997	1.81	2.76	0.9710	1.83	2.78	0.9693	1.80	2.79	0.9686
1998	4.26	5.78	0.9680	4.23	5.83	0.9670	4.03	5.56	0.9684
1999	7.07	10.30	0.9547	6.73	10.08	0.9545	6.43	9.49	0.9574
2000	7.57	11.27	0.9424	7.50	11.29	0.9433	6.76	10.20	0.9466
2001	7.45	10.36	0.9704	8.11	11.00	0.9687	7.14	9.80	0.9670
2002	6.97	8.97	0.9624	6.99	9.04	0.9606	6.12	8.07	0.9619
2003	4.88	6.43	0.9715	4.91	6.80	0.9705	4.71	6.31	0.9704
2004	3.62	5.12	0.9603	3.41	5.12	0.9598	3.41	4.98	0.9612
2005	2.90	4.19	0.9600	2.88	4.16	0.9588	2.92	4.20	0.9588
2006	1.72	2.63	0.9414	1.67	2.61	0.9385	1.66	2.59	0.9397
2007	1.28	1.67	0.9663	1.02	1.44	0.9712	1.02	1.44	0.9703
2008	1.35	1.69	0.9002	0.71	1.13	0.9228	0.77	1.17	0.9198
Total	4.03	6.63	0.9916	3.93	6.67	0.9914	3.68	6.18	0.9918

Table.2 The prediction errors (MAE, RMSE) and R of the TCN model for the F10.7 data during different solar cycles.

Solar cycle	1-Day ahead			2-Days ahead			3-Days ahead		
	MAE (sfu)	RMSE (sfu)	R	MAE (sfu)	RMSE (sfu)	R	MAE (sfu)	RMSE (sfu)	R
19	4.35	9.03	0.9880	4.29	7.84	0.9908	4.42	8.51	0.9897
20	3.35	5.16	0.9924	3.86	5.76	0.9928	3.40	5.37	0.9926
21	4.59	7.51	0.9921	4.48	7.16	0.9927	4.65	7.45	0.9930
22	4.71	7.89	0.9908	5.36	8.57	0.9908	4.75	8.05	0.9903
23	3.76	6.46	0.9917	4.30	7.01	0.9915	3.91	6.73	0.9912
24	3.03	5.60	0.9846	2.78	5.49	0.9833	3.23	5.52	0.9850
Mean	3.97	6.94	0.9899	4.18	6.97	0.9903	4.06	6.94	0.9903

Cycles with stronger solar activity are found to have relatively poor model forecast errors. For cycles with weaker solar activity, the results are relatively better. Solar cycles 20 and 24 have about the same intensity of solar activity and are both weaker.The model forecasts are relatively better, Solar cycles 21 and 22 have about the same intensity of solar activity and are both stronger.The model forecasts are relatively poorer. However, the overall average prediction results do not change much compared to solar cycle 24.Therefore, the TCN model does not affect the final F10.7 forecasts due to specific properties of the data. The performance of the model may be affected by the intensity of solar activity.We trained the model using data from 1957-1995 and tested it using 2003 and 2004. You can see more detailed information **in lines 189-199.**

RC8:

Section 2.1: The authors mention they use processed data, but do not explain how they processed it. Please elaborate.

AR8:

I am sorry for not explaining the process of processing the data, I would add the following. “The 10.7 cm daily solar flux data were obtained from the website of the National Oceanic and Atmospheric Administration. Three flux determinations are made each day. Each 10.7cm Solar Flux measurement is expressed in three values: the observed, adjusted, and URSI Series D values(absolute values).The observed value is the number measured by the solar radio telescope. This is modulated by two quantities: the level of solar activity and the changing distance between the Earth and Sun. Since it is a measure of the emissions due to solar activity hitting the Earth, this is the quantity to use when terrestrial phenomena are being studied(Tapping, 1987). When studying the Sun, it is undesirable to have the annual modulation of the 10.7cm Solar Flux caused by the changing distance between the Earth and Sun. However, during the ephemeris calculations required for the solar flux monitors to accurately acquire and track the Sun, one of the byproducts obtained is the distance between the Sun and the Earth. Therefore, we generate an additional value called the adjusted value, which takes into account the variations in the Earth-Sun distance and represents the average distance. Absolute measurements of flux density are quite difficult. Astronomers attempt to match the solar flux density data at various frequencies with a frequency spectrum by applying a scale factor. By combining each wavelength with the calibrated spectrum, a series of D Flux is obtained, where D Flux equals 0.9 multiplied by the adjusted flux(Tanaka et al.,1973).

Between March and October measurements are made at 1700, 2000 (local noon) and 2300UT. However, the combination of location in a mountain valley and a relatively high latitude makes it impossible to maintain these times during the rest of the year. Consequently, from November through February, the flux determination times are changed to 1800, 2000, and 2200, so that the Sun is high enough above the horizon for a good measurement to be made.Therefore, we chose the adjusted flux value of F10.7 measured at 8:00 p.m. UT (local noon).”I agree with your advice, and have revised this problem in my manuscript. You can see more detailed information **in lines 69-90**.

RC9:

Section 2.2: Surprisingly, this section consists of just one table. We suggest that the authors merge this section with sections 2.3 and/or 2.4, or put this table in the Appendix. On a side note,

it's nice to see that this model can run on a "normal" PC and doesn't require a large compute server, which should make it easier to replicate.
AR9:
I think this table is important, so I'm going to give an explanatory note about the contents of the table. The addition is as follows: "The parameters related to the hardware and software environment for this experiment are shown in Table 1. We build a model through Python and utilize some efficient frameworks including Pandas,Matplotlib, Tensorflow,and Sklearn. Pandas supports us to complete data processing and Matplotlib supports us to display graphics. Tensorflow and Sklearn are essential frameworks for building various prediction models." You can see more detailed information **in lines 94-97**.

RC10:
L. 131: Define ReLU.
AR10:
Line 131, the word "ReLU" is a commonly used activation function in deep learning and is a nonlinear function. It is defined as $f(x) = \max(0, x)$. I am agree with your advice, and have revised this problem in my manuscript. You can see more detailed information **in lines 155-156.**

RC11:
L. 132: Are weights normalized with Batch Normalization? Please clarify.
AR11:
No, feature normalisation is used. The data is normalised using the max-min scaling method before model training. In order to solve the problem of gradient disappearance or gradient explosion during the training process due to different features in the data.

RC12: Table 2: I find this table very confusing.
RC12-(1): How is the batch size "None"?
AC12-(1):We don't set the value of batch size, we directly take an entire training set for training. So, the batch size is none.
RC12-(2):What unit is the step length?
AC12-(2):The unit of F10.7 is day. We chose a step size of 20 days to predict F10.7.
RC12-(3):Clarify what input "dimension" and "shape" are.
AC12-(3): F10.7 for single variable, the data in F10.7 at the time of reading is 1 dimension. Model training requires processing the F10.7 data into the desired shape size. In this paper, we are using 20 days as a step to predict F10.7. The shape size of the input data in the TCN model is 20.
RC12-(4):Why do "tcn_layer.receptive_field" and "Dense(1)" parameters not have a value?
AC12-(4): The "tcn_layer.receptive_field" determines the size of the region of the input layer that corresponds to an element in the output of a particular layer. The receptive field tells you how far the model can see in terms of timesteps. I apologise for forgetting to give the size of the receptive field. The size of the receptive field is 253.I have revised this problem in my manuscript.
Dense(1) is a fully connected neural network that acts as a "classifier" in the whole

convolutional neural network. Dense(1) is one of the constituent structures of the model in this paper and therefore has no value.

RC13:

Section 2.5: Although MAE, RMSE, and R are indeed three commonly used measures, they may not offer a comprehensive assessment of a model's quality.

I recommend that the authors consider following the guidelines provided by [1] and include additional bias and/or discrimination metrics, such as Probability of Detection and False Alarm Rate, to enhance the completeness of their evaluation.

It would be even more beneficial if the authors could assess the timeliness of their forecasts using Dynamic Time Warping (DTW)-based metrics, as demonstrated by [2], [3].

AR13:

I took the advice you gave and chose relative error to test the performance of the model. The relative error is defined as:

$$\sigma = \frac{|f_i - F_i|}{f_i} * 100\%$$

Where f_i denotes forecast and F_i denotes observation.

The model was based on a training set from 1957 to 1995.The results are shown in Figure 1 and Table 3. Figure 1 shows the frequency distribution of the difference between the observed values and the model predictions. Table 3 shows the statistical values of the probability of the difference between the observed values and the predicted values of the model. As can be seen from Figure 1 and Table 3,most predictions (88% of the 1-3 days ahead forecast) were located within ±6sfu of error. You can see more detailed information **in lines 200-214.**

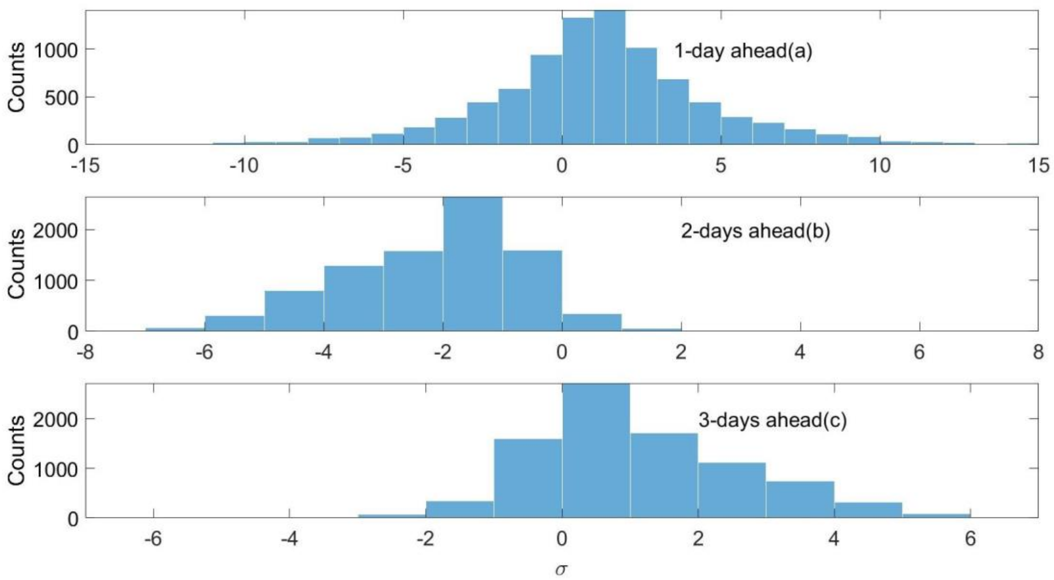


Figure 1 shows the frequency distribution of the difference between the observed values and the model predictions during 1996-2019(solar cycles 23and24).

Table 3 shows the statistical values of the probability of the difference between the observed values and the predicted values of the model.

Relative Error	1-Day ahead	2-Days ahead	3-Days ahead
2sfu	4254(48.53%)	4647(53.01%)	6351(72.45%)
4sfu	6686 (76.27%)	7548(86.11%)	8301(94.70%)
6sfu	7722(88.09%)	8670(98.90%)	8722(99.50%)
8sfu	8272(94.36%)	8753(99.85%)	8751(99.83%)
10sfu	8530(97.31%)	8762(99.95%)	8756(99.89%)

RC14:

Section 3:

RC14-(1):It is very surprising that the model seems to perform sometimes better for 3-days ahead forecasts rather than 2-days or even 1-day ahead forecasts. Can the authors please comment?

AR14-(1): The TCN model forecasts 1-day ahead are also more effective compared to the other models. The other models lead to larger errors with the number of days in advance of the forecast. The TCN model does not show a large change in its effectiveness in the short-term forecasting time. You can see more detailed information **in line 188.**

RC14-(2):Since the authors indicated earlier that one of the advantages of a TCN is its training time, I suggest that the authors indicate the training time of their model.

AR14-(2): Thanks to its parallel architecture, the TCN model reduces training time to approximately 7.5 minutes.

RC15:

Figure 5:The Figure is too small to be useful. Any "persistence-like" (i.e. using the last true observation as the forecast) behaviour would be hidden. Please resize and zoom in, one or two month(s) of data is probably enough.

In addition, sub-figure vertical axes should start at 0.

AR15:

Figure 5 in my manuscript has been modified to the image shown in Figure 2.I have revised the typing errors in my manuscript. You can see more detailed information **in line 223.**

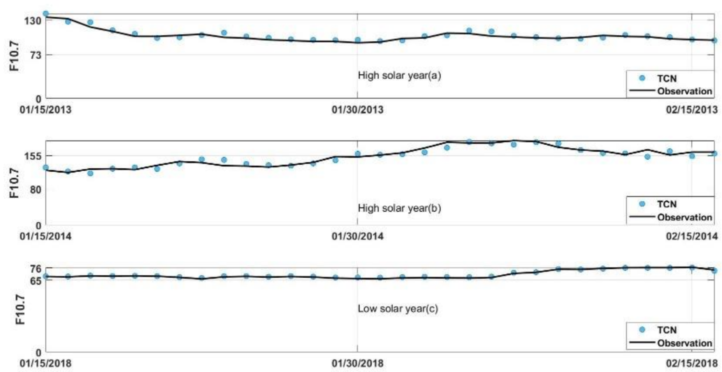


Figure 2:shows the predicted effects for solar activity high years in the Panel(a)-(b) and solar activity low year in the panel(c) for 1-day ahead in solar cycle 24.

RC16:Section 3 and Figure 6:

RC16-(1):The authors should describe and reference the model used by the SWPC to forecast the F10.7 index.
AR16-(1): SWPC is the US Weather Prediction Centre, I'm sorry I don't know what model they used to predict F10.7.

RC16-(2):They should also detail how they got these SWPC results (did they reproduce it? Did they download it?).
AR16-(2): I am sorry for not providing documentation of the SWPC forecast results, which can be obtained at https://www.swpc.noaa.gov/sites/default/files/images/u30/F10.7%20Solar%20Flux.pdf 。 I have revised this problem in my manuscript. You can see more detailed information **in line 227.**

RC16-(3):It is also unclear why the comparison is limited to the years 2009 to 2013 and does not extend until 2019. Please provide an explanation for this choice and, if possible, complete the evaluation with the remaining years.
AR16-(3): Documentation of forecast F10.7 provided by SWPC. Only forecasts from 1989 to 2013 are available.

RC17:Section 3 and Figure 7:
RC17-(1):it is unclear to me which model is the so-called "AR" model. Please clarify.
AC17-(1):The "AR" means autoregressive model .AR is mentioned in line 14,and where a proper noun is mentioned before, it will be used as an abbreviation afterwards.
RC17-(2):It is again unclear why the model is evaluated only for those specific years. Please explain and if possible complete the evaluation with the remaining years.
AC17-(2): The models used for comparison only gave forecasts of F10.7 in these years.

RC18:Section 3 and Table 4:
RC18-(1):The authors are now using the years 2003 and 2004 to evaluate and compare their model. Did they train another version of the model with a different train/test split? Please comment.
AC18-(1): YES. We retrained the model again. 1957-1995 for the training set and 1996-2019 for the test set. Then the results of the F10.7 forecasting model already studied by previous authors((Wang et al., 2009;Zhang et al., 2020) were selected for comparison. The reason why 2003-2004 was chosen for comparison is that they both happen to have results for those two years.

RC18-(2):Please comment why the RMSE is the only measure used for comparison purposes. In my opinion, the authors could additionally use at least one measure of correlation and one of bias (or even a measure of training time) in their comparison.
AC18-(2): Most of the previous evaluators have adopted RMSE as the evaluation metric for F10.7 forecast, and other space weather centers such as SWPC have also included RMSE as the evaluation metric for F10.7 forecast performance. Therefore, I choose RMSE as the evaluation metric. But your advice is also very necessary. So I added more years and metrics for comparison. The model was based on a training set from 1957 to 1995.The comparison

results are shown in Table 3. You can see more detailed information **in line 280.**

Table3. Results of the TCN model's forecast performance 3-days ahead compared to other models

Year	**BP/TCN**			**LSTM/TCN**		
	RMSE (sfu)	**MAPE (%)**	**R**	**RMSE (sfu)**	**MAPE (%)**	**R**
2003	14.82/4.71	8.15/3.63	0.9937/0.9704	7.04/4.71	3.70/3.63	/
2004	9.74/3.14	7.05/3.12	0.9960/0.9612	5.14/3.14	3.22/3.12	0.9603/0.9612
2008	2.15/1.22	2.11/1.11	0.9996/0.9198	1.22/1.22	1.20/1.11	0.9200/0.9198
2009	1.84/1.03	1.91/1.01	0.9996/0.9295	1.05/1.03	1.07/1.01	/

RC18-(3):Please clarify if all the BP and LSTM results were reproduced or obtained from their original authors. If the results were reproduced, it would be beneficial to have a description of the models' architectures and training procedures (at least as an Appendix).

AC18-(3): All the BP and LSTM results were obtained from the original authors.

RC19:

I find it rushed to assert that the TCN architecture is intrinsically better than an LSTM for predicting the daily value of F10.7 when only one evaluation metric is used, over only two test years (even if these are years of high activity), without knowing anything about how the LSTM was trained (same sequence length?). In my opinion, the authors should comment on these points, and probably be more careful when asserting that TCN is a significant improvement over an LSTM. Additional metrics, or even figures to back up their claim, would be more convincing.

AR19:

Both LSTM and TCN models are based on the same dataset for model training. F10.7 data before 1995 were chosen as the training set and 1996-2019 as the test set. Finally, 2003-2004 data was chosen for comparison. I agree with your advice. I added the LSTM model for more forecasting results. The model was based on a training set from 1957 to 1995.The results are shown in Table 4.

Table 4 Results of the TCN model's forecast performance 3-days ahead compared to LSTM

Year	RMSE			R	
	LSTM	TCN	Error reduction	LSTM	TCN
1996	1.72	1.57	8.72%	0.958	0.9540
1998	5.72	5.56	2.79%	0.9662	0.9684
2001	10.02	7.14	28.7%	0.9694	0.9670
2003	7.04	4.71	33.1%	/	0.9704
2004	5.14	3.41	33.7%	0.9603	0.9612
2008	1.22	1.17	4.1%	0.9200	0.9198
2011	5.36	5.14	4.1%	0.9757	0.9744

2014	8.65	4.96	42.7%	0.9497	0.9463
2016	3.05	2.95	3.27%	0.9644	0.9659
2019	1.33	1.19	10.6%	0.8969	0.9022
Total	6.20	5.12	17.4%	0.9889	0.9855

RC20:

Conclusion: The conclusion should probably be reworked and tempered after the above points have been addressed.

AR20:

I have revised conclusion in my manuscript. You can see the detailed information in the conclusion section. You can see more detailed information **in lines 286-289.**

Technical Comments

RC1):

Many typing errors, e.g. on lines: 22; 25; 37; 42; 48; 50; 53; 166; 175; 176; 183; 188; 191; 195; 204; 268; etc.

AR1):

I have revised the typing errors in my manuscript. You can see the detailed information in my manuscript.

RC2):

The authors should indicate in their abstract what is the forecast horizon associated with the provided performance metrics.

AR2):

I agree with your advice. I have revised abstract in my manuscript.Lines11-14: The sentence "The F10.7 series from 1957 to 2019 are used, which the datasets from 1957 to 2008 are used for training and the datasets from 2009 to 2019 are used for testing. The results show that the TCN model of prediction F10.7 with a root mean square error (RMSE) from 5.03 to 5.44sfu and correlation coefficients (R) as high as 0.98 during solar cycle 24."is replaced by"The F10.7 series from 1957 to 2019 are used.The data during 1957–1995 are adopted as the training dataset, the data during 1996–2008 (solar cycle 23) are adopted as the validation dataset,and the data during 2009–2019 (solar cycle 24) are adopted as the test dataset. The leave-one-out method is used to group the data set for multiple validations. The prediction results for 1-3 days ahead during solar cycle 24 have a high correlation coefficient (R) of 0.98 and a root mean square error (RMSE) of only 5.03~5.44 sfu." The sentence has been revised **in lines 13-17.**

RC3):

The authors should specify in the abstract that their model is an autoregressive model (it only takes past values of the F10.7 index as inputs).

AR3):

I agree with your advice. I have revised abstract in my manuscript.Lines11-14: The sentence "The overall accuracy of the TCN forecasts is better than those of the widely used autoregressive (AR) models and the results of the US Space Weather Prediction Center (SWPC) forecasts especially for 2 and 3 days ahead."is replaced by"The overall accuracy of the TCN forecasts is better than the autoregressive (AR) model (it only takes past values of the F10.7 index as inputs)and the results of the US Space Weather Prediction Center (SWPC) forecasts, especially for 2 and 3 days ahead." The sentence has been revised **in lines 17-19.**

RC4):The abstract is written in a rather poor and confusing language (in particular Lines 11 – 15), and should be rewritten.

AR4):

I agree with your advice. I have revised abstract in my manuscript. You can see the detailed information in the in my manuscript. You can see more detailed information **in lines 13-19.**

RC5):

L. 30: Did the authors meant "correlation" instead of "link"? This sentence should be rephrased.

AR5):

The sentence "The link between F10.7 at the current moment and F10.7 at the previous moment would be decreasing as the time interval increases."is replaced by"The correlation between F10.7 at the current and previous moments decreases as the time interval increases." The sentence has been revised **in line 32.**

RC6):

L. 45-46: This sentence is very poorly written and should be completely reworded.

AR6):

The sentence "With the rapid development of machine learning and neural networks, the powerful learning capabilities of machine learning are increasingly being intrigued by researchers and used to in the study variaiton of solar activity"is replaced by"With the rapid development of machine learning and neural networks. Researchers are increasingly intrigued by the powerful learning capabilities of machine learning and neural networks, using them to study variations in solar activity." The sentence has been revised **in lines 45-46.**

RC7):

In the manuscript, the authors often refer to a so-called Back-Propagation (BP) network. I understand that this denomination comes from the article by Xiao et al., 2017 cited by the authors. I find it a misnomer because "back-propagation" is the method used to make the neural network learn, and does not depend on the architecture of the network. For example, the vast majority of convolutional networks also use "back-propagation" during their training phase (and the author's TCN model also does). Here, the network referred to by the authors is simply a "feedforward artificial neural network" (also sometimes called "multi-layer perceptron"). This is why I suggest that the authors change the BP and BPNN denomination to a more standard and understandable one.

AR7):

I agree with your suggestion. I have replaced "Back-propagation (BP) network" in my manuscript with "back propagation neural network".

RC8):

L. 55-56: This sentence is very poorly written and should be reworded. "sfu" should be introduced.

AR8):

The sentence “The model of the root mean square error (RMSE) of the forecast was only 6.20-6.35sfu, and the high correlation coefficient (R) was 0.98”is replaced by“The forecast had a high correlation coefficient (R) of 0.98 and a low root mean square error (RMSE) range of 6.20-6.35 sfu.” The sentence has been revised **in lines 55-56.**

Line 55 ,the unit “sfu” is the unit of solar flux (F10.7). The quantities are separated by commas. The 10.7cm Solar Flux is given in solar flux units (a sfu = 10^{-22}W m^{-2} Hz^{-1}). I have revised this problem in my manuscript. You can see more detailed information **in line 69.**

RC9):

L. 56: The authors mention “RNN-based” architectures without introducing the meaning of the acronym and explaining that “RNN-based” methods include “LSTMs”.

AR9):

The “RNN” means Recurrent Neural Network. I am sorry for not giving a definition. I am agree with your advice, and have revised this problem in my manuscript. You can see more detailed information **in line 56.**

RNN(recurrent neural network) and LSTM(long short-term memory network)are not the focus of this paper, which instead presents methods based on CNN. Therefore, there was no presentation.

RC10):

L. 58: The right reference is probably Bai et al., 2018 and not 2017.

AR10):

I have revised this problem in my manuscript. This error has been corrected **in line 59.**

RC11):Section 2.3: Notation for vector is inconsistent. Sometimes the vector "x" is referred as "x" sometimes as $\vec{x}$ (with an upper arrow).

AR11):

I have revised this problem in my manuscript. This error has been corrected **in line 128.**

RC12):

L. 124: Please add a reference for L1 loss.

AR12):

The L1 norm loss function is extensively utilized in deep learning tasks(Zhao et al.,2017).You can see more detailed information **in line 149.**

H. Zhao, O. Gallo, I. Frosio and J. Kautz.:Loss Functions for Image Restoration With Neural Networks,in IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 47-57,doi: 10.1109/TCI.2016.2644865,2017.

RC13):L. 138-139: “Business sector” is probably a confusing way of referring to operational space weather centers or space agencies. Please clarify.

AR13):

" Business sector " means the Space Weather Prediction Center. I'm deleting it.I have revised this problem **in lines 161-163.**

RC14):
Figure 5: “Practical” is unclear. Please consider changing it to “Observations” or something similar.
AR14): I have revised this problem in my manuscript. Modified as shown in Figure 3. You can see more detailed information **in line 223.**

You can see the detailed changes in the resubmitted manuscript. If you have any problems, please contact me immediately. I am very grateful for your comment. Thank you very much.

Best Regard
LuYao Wang
The 1st author of this manuscript