Response to the Reviewers for Sensors-2011287 entitled "DTTrans: PV Power Forecasting using Delaunay Triangulation and TransGRU"

Keunju Song, Jaeik Jeong, Jong-Hee Moon, Seong-Chul Kwon and Hongseok Kim

December 8, 2022

1 General response

We appreciate the reviewers' comments and suggestions. Below we present point-by-point discussions about the reviewers' concerns, along with the changes that were made. To enhance readability, the comments from each reviewer are in italicized font while our discussion is not. In addition, the revised texts are quoted in a blue font. We hope that the discussions below and revisions address the substantive concerns of the reviewers. If we have overlooked an aspect that would improve the paper we are willing to do so.

Sincerely,

Keunju Song
Jaeik Jeong
Jong-Hee Moon
Seong-Chul Kwon
Hongseok Kim

2 Response to the first review

The proposed technique is well explained and worth publishing in this journal.
To improve the quality of the paper, a few suggestions are as follows.

⇒ We appreciate the careful review. Our point-by-point answers are provided below.

1) The literature review presented here is highly insufficient and generalized. Please improve it using recent papers.

⇒ Following the reviewer's suggestion, in the introduction we further included recent papers to improve the literature review as follows.

> Moreover, the work in [13] additionally considered historical PV power. They used solar radiation, temperature, relative humidity, wind speed and historical PV power as an input of CNN-bidirectional GRU (BiGRU) model to improve the performance.
>
> The authors in [20] used cloud images from ground-based cameras and historical solar irradiance for input of CNN model to predict solar irradiance by extracting spatio-temporal features.
>
> [13] Zhang, C.; Peng, T.; Nazir, M. S. A novel integrated photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables. Electr. Power Syst. Res., 2022, 213, 108796.
>
> [20] Zhang, X.; Zhen, Z.; Sun, Y.; Wang, F.; Zhang, Y.; Ren, H.; Ma, H; Zhang, W. Prediction Interval Estimation and Deterministic Forecasting Model Using Ground-based Sky Image. IEEE Trans. Ind. App., 2022

> However, PV power output is affected by the stochastic nature of weather condition such as temperature, humidity, wind speed, and cloudiness, so it has uncertainty problem. Considering on this inherent property, various solar PV systems have been designed to provide stable power quality for grid integrated distributed resources system [2-4]. Furthermore, in operating smart grid system, accurate power generation forecasting of distributed energy resources has been required for scheduling and balancing power systems.
>
> [2] Kumar, N.; Singh, B.; Panigrahi, B. K. LLMLF-Based Control Approach and LPO MPPT Technique for Improving Performance of a Multifunctional Three-Phase Two-Stage Grid Integrated PV System. IEEE Trans. Sustain. Energy, 2020, 11, 371-380.
>
> [3] Kumar, N.; Singh, B.; Panigrahi, B. K.; Xu, L. Leaky-least-logarithmic-absolute-difference-based control algorithm and learning-based InC MPPT technique for grid-integrated PV system. IEEE Trans. Ind. Electron., 2019, 66, 9003-9012.
>
> [4] Kumar, N.; Singh, B.; Wang, J.; Panigrahi, B. K. A framework of L-HC and AM-MKF for accurate harmonic supportive control schemes. IEEE Trans. Circuits Syst. I, Reg. Papers, 2020, 67, 5246-5256.

2) Eqn.2 is not clear. Please elaborate.

⇒ Following the reviewer's suggestion, we added the explanation of Eqn.2 as follows.

Eq.2 is about how the Voronoi cell is constructed. Arbitrary point p and seed points p_i are on the same plane, which are the locations of PV power plant (subspace of p) and weather station in our case. Under this status, the distance between p and seed point (e.g., p_i) is calculated and compared with other seed points to find which seed point is closest.

3) Few variables are not defined. Please correct it.

⇒ Following the reviewer's suggestion, we defined the variables that were previously not defined as follows.

where the projections are parameter matrices $W^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W^K \in \mathbb{R}^{d_{model} \times d_k}$, $W^V \in \mathbb{R}^{d_{model} \times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{model}}$. The variables d_k, d_v and d_{model} denote the dimension of keys, values and outputs of Transformer encoder.

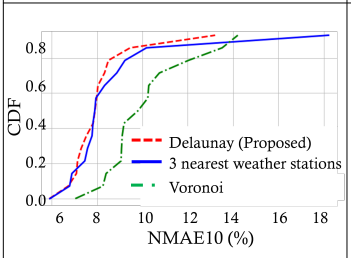
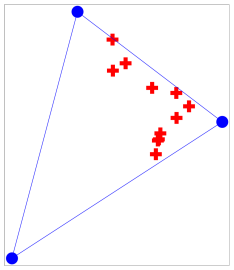
$$\begin{cases} r(t) = \sigma_g(W_r x(t) + U_r h(t-1) + b_r) \\ z(t) = \sigma_g(W_z x(t) + U_z h(t-1) + b_z) \\ h(t) = (1 - z(t)) \odot h(t-1) + z(t) \odot \hat{h}(t) \\ \hat{h}(t) = \sigma_h(W_h x(t) + U_h(r(t) \odot h(t-1)) + b_h) \end{cases} \tag{1}$$

where $z(t)$ is the update gate vector at time step t, $r(t)$ is the reset gate vector, W, U and b are parameter matrices and vector. The operator $\odot$ denotes the Hadamard product. σ_g is a sigmoid function and σ_h is a hyperbolic tangent.

4) The picture quality of waveforms should improve.

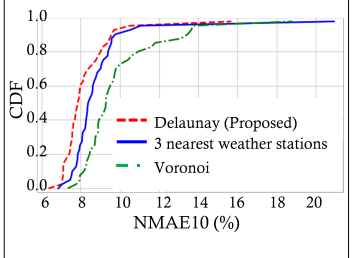
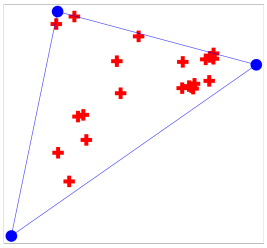
⇒ Following the reviewer's suggestion, we improved our figures by adjusting the size of axis length and axis font as follows.

In Section 4.3, Figure 9.

CDF Comparison	Delaunay Triangle for Test Region
	

Mean of NMAE10	
Method	**Value (%)**
Delauany (proposed)	**8.06**
3 nearest weather stations	8.64
Voronoi	10.07

(a) Daegu

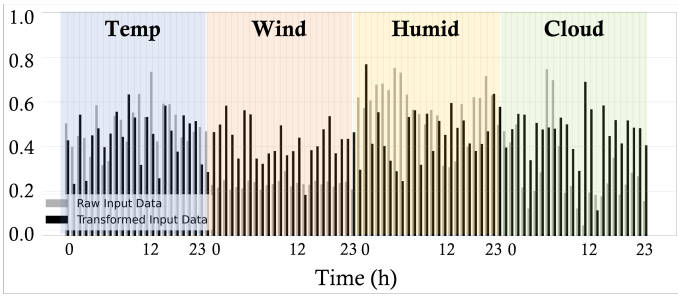
CDF Comparison	Delaunay Triangle for Test Region
	

Mean of NMAE10	
Method	**Value (%)**
Delauany (proposed)	**8.22**
3 nearest weather stations	8.84
Voronoi	9.86

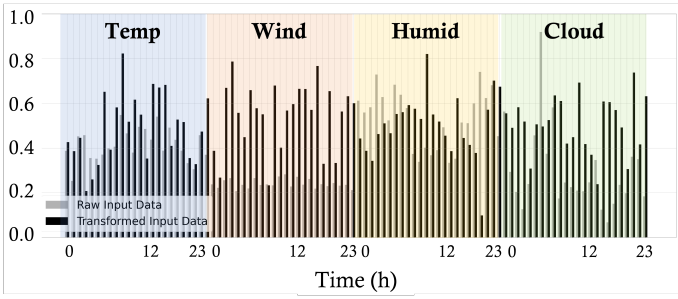
(b) Gwangju

Figure 9. An example of Union-Inner Triangles test results.

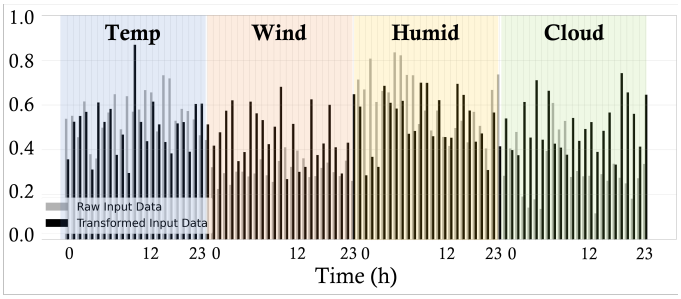
In Section 4.3., Figure 10.



(a) 4 meteorological factors of the weather station A



(b) 4 meteorological factors of the weather station B



(c) 4 meteorological factors of the weather station C

Figure 10. Comparison of raw input data and transformed input feature.

5) Few short forms have been used without giving full forms. Please cross-check throughout the paper properly.

⇒ Following the reviewer's suggestion, we corrected a few short forms by providing full names when they are first introduced as follows.

Furthermore, hybrid models also have been considered such as CNN-LSTM and wavelet packet decomposition-LSTM (WPD-LSTM) [16,17].

For GRU, we select the single GRU unit and set the size of hidden state as 64. Two fully connected (FC) layers are followed behind, and we set 256 dimension of each layer.

We compare the proposed TransGRU model with two well-known models: multi-layer perceptrons (MLP) and GRU [32].

6) To improve the introduction and reference sections, you should follow quality papers. A few suggestions are as follows.

doi: 10.1109/TSTE.2019.2891558
doi: 10.1109/TIE.2018.2890497
doi: 10.1109/TCSI.2020.2996775

⇒ Following the reviewer's suggestion, in the introduction we have added additional explanation including three recommended papers as we answered in the second comment.

7) Please go through those papers, and include and improve your literature review portion of the paper.

⇒ Following the reviewer's suggestion, we comprehensively revised the literature review in the introduction as follows.

Traditionally, physical models for solar energy prediction have been developed based on specific conditions of region such as meteorological, climatological, and geographical parameters [5]. In addition, statistical models such as the auto-regressive integrated moving average (ARIMA) and the auto-regressive moving average with exogenous inputs (ARMAX) models have been applied to predict PV power generation [6,7]. However, physical models can only estimate solar energy for a longer time period, e.g., monthly, so they are hard to predict in detailed time interval such as hours and minutes. Moreover, meteorological factors have nonlinear relationship with PV power generation, so physical models and statistical models may not perform well.

Machine learning based PV power forecasting methods can be summarized as follows by the feature data that are related with PV power generation: 1) meteorological factors and historical PV power generation, 2) spatio-temporal features and 3) satellite or ground-based sky images. First, in [12], the authors used temperature, dew point, humidity and wind speed as input data to predict solar irradiance by LSTM. Moreover, the work in [13] additionally considered historical PV power. They used solar radiation, temperature, relative humidity, wind speed and historical PV power as an input of CNN-bidirectional GRU (BiGRU) model to improve the performance. Secondly, several researchers considered about spatio-temporal features, which are extracted from meteorological factors and geomorphology of PV site. The work in [18] additionally used digital elevation map of the solar PV power plant, solar irradiation, temperature, precipitation and wind speed to extract spatio-temporal features by CNN. Without using meteorological factors, the authors in [15] used location of multiple PV sites to construct a space-time matrix with only historical PV power data, and placed it as an input of CNN to extract spatio-temporal features. Thirdly, satellite and ground-based

sky images are other approaches for extracting spatio-temporal features of PV power generation. In [19], the authors suggested two-stage predictions: the first is about cloud amount prediction by satellite images, and the second is about PV power prediction using the predicted cloud amount as one of the features. The authors in [20] used cloud images from ground-based cameras and historical solar irradiance for input of CNN model to predict solar irradiance by extracting spatio-temporal features.

8) Elaborate discussions of results. Try to point out each waveform using proper justification.

⇒ Following the reviewer's suggestion, we have added additional explanation in the experiment results as follows.

In Figure 12b, as *weather* forecast error goes from 5% to 20%, TransGRU further outperforms MLP and GRU. More specifically, for 5% *weather* forecast error case, differences between TransGRU and MLP, GRU in terms of NMAE10 are 3.81% and 10.31%. For 20% *weather* forecast error case, differences between TransGRU and MLP, GRU get higher, which show 7.28% and 11.57%. In addition, the differences between 5% and 20% *weather* forecast error case for each MLP, GRU, and TransGRU itself are 12.72%, 10.20%, and 8.65%, which show TransGRU has smallest decrease. This confirms that Transformer encoder is robust for weather forecast error.

In addition to CDF comparison, the average of NMAE10 in Gwangju and Daegu are shown in Table 4. In addition, we provide the result using VT to highlight the performance improvement using DT+ over VT in Table 4. As described in the table, DT+ outperforms VT for all the cases of weather forecasting error and the choice of learning model. For example, DT+ shows 19.9%, 16.0%, and 10.3% improvement compared to VT in 20% weather forecast error for Gwangju. In addition, TransGRU outperforms MLP and GRU as well. For example, TransGRU with DT+ shows 5.9% and 15.1% improvement compared to MLP and GRU with DT+ in 20% weather forecast error for Gwangju.

9) Rewrite the conclusion section in the summarized form.

⇒ Following the reviewer's suggestion, we revised the conclusion section in the summarized form as follows.

In this paper, we proposed a novel short-term PV power forecasting technique called DTTrans; based on DT and TransGRU. Unlike using weather data from the closest weather station, DT selects three weather stations that enclose PV sites. Since the effect of DT weather stations depends on the shape of DT and the inside location of the PV site, the proposed Union-Inner Triangles algorithm adaptively handles whether to use DT weather stations or 3-nearest weather stations. The proposed method is then combined with an interpretable AI model called TransGRU, and becomes robust against weather forecast error since it selectively considers the weather stations and

meteorological factors that are highly related with PV generation. Extensive experiments show that the proposed DTTrans is highly effective in forecasting individual PV sites as well as aggregated PV sites. As the weather forecast error increases from 5% to 20%, TransGRU further outperforms the other methods based on MLP and GRU. Furthermore, DTTrans considerably improves the performance when multiple PV sites are aggregated in the VPP form and finally achieves 3−4% forecasting error on average.

3 Response to the second review

In this manuscript, the authors use GRU combined with a Transformer to predict the power output of PV. The topic is relevant to the aim of the journal Sensors. However, the current version cannot be accepted for publication. Detailed information and other comments are listed below:

⇒ We appreciate the careful review. Our point-by-point answers are provided below.

1) is the term Delaunay Triangulation developed by the authors or adopted from elsewhere, please provide the reference if cited from other research work.

⇒ Delaunay Triangulation is an existing algorithm that separate the regions on the plane in a triangle form based on nearest-neighbor rule. To the best of our knowledge, we first adopted this algorithm for PV power forecasting domain. Following the comment, we additionally cited some papers that used Delaunay Triangulation in some other applications.

> Voronoi Tessellation and Delaunay Triangulation have been used in practical and theoretical ways in many fields such as science, technology and visual art [26, 37-41].
>
> [37] Hu, J.; Niu, H.; Carrasco, J.; Lennox, B.; Arvin, F. Voronoi-based multi-robot autonomous exploration in unknown environments via deep reinforcement learning. IEEE Trans. Veh. Technol. 2020, 69, 14413-14423.
>
> [38] Lei, H. Y.; Li, J. R.; Xu, Z. J.; Wang, Q. H. Parametric design of Voronoi-based lattice porous structures. Mater. Des. 2020, 191, 108607.
>
> [39] Luo, Y.; Mi, Z.; Tao, W. DeepDT: Learning geometry from Delaunay triangulation for surface reconstruction. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2–9 February 2021; pp. 2277-2285.
>
> [40] Wang, W.; Wang, H.; Fei, S.; Wang, H.; Dong, H.; Ke, Y. Generation of random fiber distributions in fiber reinforced composites based on Delaunay triangulation. Mater. Des. 2021, 206, 109812.
>
> [41] Liu, Z.; Liu, H.; Lu, Z.; Zeng, Q. A dynamic fusion pathfinding algorithm using Delaunay triangulation and improved a-star for mobile robots. IEEE Access 2021 9, 20602-20621.

2) The authors should state why choosing ML method instead of a physical model to make prediction/regression. Are there any obstacles to adopting physical methods?

⇒ In PV power forecasting, the accuracy of forecasting model can be enhanced by using many relevant inputs which are highly dependent on local weather conditions. However, meteorological factors and PV power generation have nonlinear relationship. Moreover, physical models were used to estimate solar irradiance for a longer period, e.g., monthly, so it is hard to predict detailed time interval such as hours or minutes. Following the comment, we have added additional explanation about obstacles of physical methods in the introduction as follows.

In order to predict the PV power for the three classes above, various forecasting techniques have been proposed. Traditionally, physical models for solar energy prediction have been developed based on specific conditions of region such as meteorological, climatological, and geographical parameters [5]. In addition, statistical models such as the auto-regressive integrated moving average (ARIMA) and the auto-regressive moving average with exogenous inputs (ARMAX) models have been applied to predict PV power generation [6,7]. However, physical models can only estimate solar energy for a longer time period, e.g., monthly, so they are hard to predict in detailed time interval such as hours and minutes. Moreover, meteorological factors have nonlinear relationship with PV power generation, so physical models and statistical models may not perform well. Recently, machine learning based models have been widely used since they can solve nonlinear problems well and outperform linear techniques. Artificial neural networks (ANN) [9,10], recurrent neural networks (RNN) such as long short-term memory (LSTM), GRU [11-13] and convolutional neural networks (CNN) [14,15] are the typical models. Furthermore, hybrid models also have been considered such as CNN-LSTM and wavelet packet decomposition-LSTM (WPD-LSTM) [16,17]

3) Why only use GRU? There are many other time series-based deep learning algorithms, and the authors should better illustrate why not use this instead of others.

⇒ In our original submission, we gave the rationale of choosing GRU over others including LSTM, but it seems that it was not conveyed well, which is our fault of presentation.

In Section 2.4.2, we had provided the reason about the choice of RNN models; LSTM and GRU. We already did experiments about using LSTM, however, it was found that the performance of LSTM and GRU are similar. Therefore, either can be a flexible choice, but we choose GRU considering model complexity and generalization; with GRU has the smaller number of model parameters than LSTM. The reason of about comparing only LSTM with GRU is because these methods are most well-known, and the basis of many other RNN models.

4) It would be better to show both training and test results in the same figure to better view overfitting and underfitting issues.

⇒ Following the reviewer's suggestion, we have added additional figure to discuss about overfitting and underfitting issues as follows.

By setting hyperparameters in this way, Figure 7 shows that our models are trained well on the dataset wihtout overfitting and underfitting problems.

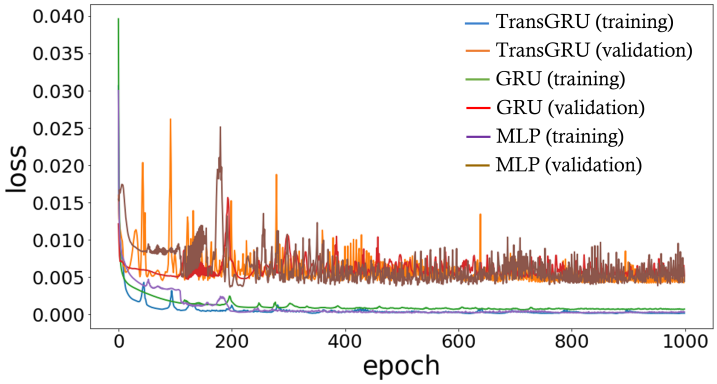


Figure 7. Learning curve comparison during training and validation.

5) Please summarize the literature in the introduction section as appropriate, which would avoid making your literature reviews like stacking papers and also help the audience better view your review work and better transaction to the next paragraph.

⇒ We appreciate the reviewer's comment. Following the suggestion, we completely rewrite the literature review part in the introduction as follows.

Machine learning based PV power forecasting methods can be summarized as follows by the feature data that are related with PV power generation: 1) meteorological factors and historical PV power generation, 2) spatio-temporal features and 3) satellite or ground-based sky images. First, in [12], the authors used temperature, dew point, humidity and wind speed as input data to predict solar irradiance by LSTM. Moreover, the work in [13] additionally considered historical PV power. They used solar radiation, temperature, relative humidity, wind speed and historical PV power as an input of CNN-bidirectional GRU (BiGRU) model to improve the performance. Secondly, several researchers considered about spatio-temporal features, which are extracted from meteorological factors and geomorphology of PV site. The work in [18] additionally used digital elevation map of the solar PV power plant, solar irradiation, temperature, precipitation and wind speed to extract spatio-temporal features by CNN. Without

using meteorological factors, the authors in [15] used location of multiple PV sites to construct a space-time matrix with only historical PV power data, and placed it as an input of CNN to extract spatio-temporal features. Thirdly, satellite and ground-based sky images are other approaches for extracting spatio-temporal features of PV power generation. In [19], the authors suggested two-stage predictions: the first is about cloud amount prediction by satellite images, and the second is about PV power prediction using the predicted cloud amount as one of the features. The authors in [20] used cloud images from ground-based cameras and historical solar irradiance for input of CNN model to predict solar irradiance by extracting spatio-temporal features. Lately, graph neural networks (GNN) were considered to exploit spatio-temporal correlations in PV power generation, and further studies are still on going [21-23]. The work in [21] used the PV generation as signals on a graph to infer the part of cloud dynamics. In [23], the authors focused on interconnections among meteorological factors by factor-based graph modeling to capture complex weather variations.

[13] Zhang, C.; Peng, T.; Nazir, M. S. A novel integrated photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables. Electr. Power Syst. Res., 2022, 213, 108796.

[20] Zhang, X.; Zhen, Z.; Sun, Y.; Wang, F.; Zhang, Y.; Ren, H.; Ma, H; Zhang, W. Prediction Interval Estimation and Deterministic Forecasting Model Using Ground-based Sky Image. IEEE Trans. Ind. App., 2022

6) From the authors' introduction and paper review, it is not clear about the gap, the method is a maturely established model Transformer/GRU, and the target problem is already been taken care of in several papers already. The authors should differentiate theirs from the other authors.

⇒ We appreciate the suggestion. Our proposed model has the following advantages as follow: 1) effective method for small portion of data and scarcity of relevant inputs, 2) Interpretable AI model for PV power forecasting.

For input data, our model uses 4 meteorological factors: temperature, wind, humidity, and cloud, which are relevant factors with PV power generation and collected far from PV site, i.e., 3 vertices of Delaunay triangle. Many other research use high quality meteorological factors such as solar radiation at the land surface and net solar radiation at the top of the atmosphere, collected from local PV site [42]. Furthermore, some experiments assume that weather forecast is correct, so the models under this assumption use historical weather data in test phase [43]. Unlike this naive assumption, we use weather forecast data with various forecast error cases considering real cases.

Interpretable AI model for PV power forecasting has not been actively studied so far. Considering this, we design an interpretable AI model called TransGRU, which is robust against weather forecast error by using Transformer encoder. TransGRU learns input data through Transformer encoder, which builds feature importance for PV power generation in terms of feature representation learning. From the transformed feature data, our model can selectively exploit the weather stations and meteorological factors that are highly related with the target PV site.

[42] Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R. J. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. Int. J. Forecast. 2016 32, 896-913.

[43] Mitrentsis, G.; Lens, H. An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting. Appl. Energy, 2022 309, 118473. 481

7) There are ample references that can be used in the introduction for the expansion of ML algorithms applied for regression, the authors can take the following reference as examples for successfully implementing ML for regression tasks: Comparing deep learning models for multi energy vectors prediction on multiple types of building; A general model for flow boiling heat transfer in microfin tubes based on a new neural network architecture. Please consider referring to them and summarizing these papers as appropriate.

⇒ We appreciate your helpful references that are relevant to our paper. We have thoroughly checked our references and completely rewrite the introduction, which were collectively in response 5.

8) Some attention needs to be given to the limitations associated with the methods used in the current work.

⇒ The limitation of our method is that Union-Inner Triangles algorithm is a heuristic way to select 3 weather stations, which is simply comparing the angles of triangle that depend on the locations of PV site and weather stations.

This algorithm has the limitation that it is a heuristic way to select 3 weather stations, however, it can be the one of examples which capture most relevant weather stations for PV power forecasting.

4 Response to the third review

⇒ We greatly appreciate that the reviewer thoroughly reviewed our manuscript and gave us very constructive and valuable comments, which substantially improved the quality of our manuscript. Our point-by-point answers are provided below.

1) Line 2- . . . scheduling power system's'.

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

In an era of high penetration of renewable energy, accurate photovoltaic (PV) power forecasting is crucial for balancing and scheduling power systems.

2-3) Line 11: Aggregated is misspelt. Line 12: 'on' average not 'in' average.

⇒ Following the reviewer's suggestion, we corrected a spelling error as follows.

Our framework demonstrates 41−60% improvement when PV sites are aggregated and achieves as low as 3−4% of forecasting error on average.

4) Lines 19-20 – '. . . global total installed PV capacity was 167.8 . . . ' Give the correct reference link.

⇒ Following the reviewer's suggestion, we corrected the reference link properly.

1. Global Solar Council. Global market outlook for solar power 2022–2026. Available online: https://www.solarpowereurope.org/insights/market-outlooks/global-market-outlook-for-solar-power-2022 (accessed on 24 August 2022).

5) Line 20: . . . . representing 'a' 36% ...

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

In 2021, global total installed PV capacity was 167.8 GW, representing a 36% growth in one year [1].

6) Line 21: . . . by 'the' stochastic...

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

However, PV power output is affected by the stochastic nature of weather condition such as temperature, humidity, wind speed, and cloudiness, so it has uncertainty problem.

7) Line 23: ... of power system's'.

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

Furthermore, in a operation of smart grid system, accurate power generation forecasting of distributed energy resources has been required for scheduling and balancing power systems.

8) Line 24: ... many 'architectures'...? Approaches perhaps?

⇒ Following the reviewer's suggestion, we corrected the term as follows.

In the literature, many approaches have been suggested to improve forecast accuracy.

9) Line 26: 'an' hourly...

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

Short-term forecasting is performed on an hourly or daily basis and used for load balancing.

10) Line 32: ... techniques have been 'proposed' or 'used'.

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

In order to predict the PV power for the three classes above, various forecasting techniques have been proposed.

11) Line 33: ... since 'they' can 'solve' nonlinear...

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

Recently, machine learning based models have been widely used since they can solve nonlinear problems well and outperform linear techniques.

12) Lines 34 & 35: 'networks' not network

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

Artificial neural networks (ANN) [9,10], recurrent neural networks (RNN) such as long short-term memory (LSTM), GRU [11-13] and convolutional neural networks (CNN) [14,15] are the typical models.

13) Line 37: What is WPD in full?

⇒ WPD is wavelet packet decomposition. We corrected a term with full form as follows.

Furthermore, hybrid models also have been considered such as CNN-LSTM and wavelet packet decomposition-LSTM (WPD-LSTM) [16,17].

14) Line 46: The work In [9] .... What is ST?

⇒ ST is space-time. We improved the sentence with providing full form as follows.

Without using meteorological factors, the authors in [15] used location of multiple PV sites to construct a space-time matrix with only historical PV power data, and placed it as an input of CNN to extract spatio-temporal features.

15-16) Lines 47-48: Lately Graph neural networks (GNN) were ....
Line 49: .. generation, and further ... on going [16-18].

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

Lately, graph neural networks (GNN) were considered to exploit spatio-temporal correlations in PV power generation, and further studies are still on going [21– 23].

17) Line 55: ... be installed further, developing a light model... What does light mean? Computationally light?

⇒ Yes, 'light model' means computationally light.

18) Line 60: Delaunay is misspelt

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

First, we construct Voronoi cells and Delaunay triangles based on the locations of weather stations.

19) Line 61: ... weather stations of

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

From these constructed regions, PV sites can exploit three nearby weather stations of which are the vertices of Delaunay triangle that covers the target PV site.

20) Line 62: Vertices of 'the' Delaunay...

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

From these constructed regions, PV sites can exploit three nearby weather stations of which are the vertices of the Delaunay triangle that covers the target PV site.

21) Line 65-66: ... because it transforms noisy input data to more useful time domain features.... Reference?

⇒ We referred this content in Section 4.4 about feature representation learning, with paper [30].

22) Line 83: 'nationwide' – mention the country here

⇒ Following the reviewer's suggestion, we corrected a term in specific as follows.

Specifically, we provide extensive experiments using nationwide 1034 PV sites and 86 weather stations in Korea.

23) Line 84: '... 7−15% improvement...'. Improvement relative to what?

⇒ 7−15% improvement means that our proposed DTTrans (i.e., DT+ based TransGRU) framework outperformed the baselines: VT based MLP, GRU, TransGRU and DT+ based MLP, GRU. We calculated the improvements between DTTrans and each of baselines in terms of NMAE10, and expressed them as a range 7−15%.

24) Figure 1: Give the meaning of FC.

⇒ FC means 'fully connected'. Following the reviewer's suggestion, we corrected a term with full form in Section 3.2 as follows.

Two fully connected (FC) layers are followed behind, and we set 256 dimension of each layer.

25) Line104: ... performance 'for the' test set...

⇒ Following the reviewer's suggestion, we corrected a grammatical error as follows.

In the final step, we estimate the performance for the test set for various cases.

26) Line 158: Do the circles in Equation 5 indicate multiplication? Use solid circles in that case, for clarity.

27) Line 159: '... W, U and ??? are...'. And what?

28) Line 160: Remove the ' . ' after vector

29) Figure 6: Use dots to show interconnection points for the different lines/parameters. Also the relation between Fig. 6 and Equations (5) is not clear. The σ terms in Fig. 6 have no subscripts, and where are the b_r and b_z terms? Redraw Fig. 6 better.

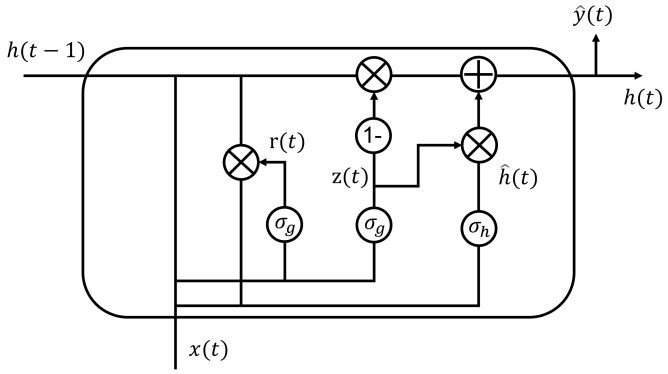
⇒ Following the reviewer's suggestions, we corrected Equation 5 and Figure 6 as follows.

Figure 6 shows the architecture of GRU unit. The relationship between input and output of GRU can be described as follows:

$$\begin{cases} r(t) = \sigma_g(W_r x(t) + U_r h(t-1) + b_r) \\ z(t) = \sigma_g(W_z x(t) + U_z h(t-1) + b_z) \\ h(t) = (1 - z(t)) \odot h(t-1) + z(t) \odot \hat{h}(t) \\ \hat{h}(t) = \sigma_h(W_h x(t) + U_h(r(t) \odot h(t-1)) + b_h) \end{cases} \tag{2}$$

where $z(t)$ is the update gate vector at time step t, $r(t)$ is the reset gate vector, W, U and b are parameter matrices and vector. The operator $\odot$ denotes the Hadamard product. σ_g is a sigmoid function and σ_h is a hyperbolic tangent.

Figure 6 in Section 2.4.2



30) Lines 170-174: Give a summary of the general statistics of the 1034 PV plants – e.g. average, minimum, maximum, standard deviation of the plant sizes, number of plants in various plant size ranges...

⇒ Following the reviewer's suggestion, we have added a summary of the 1034 PV plants as follows.

We use the PV generation data of Korea, provided by Korea Electric Power Corporation (KEPCO). There are 1034 sites located in all parts of the country, and PV generation data from newly built sites are from December 1, 2019 to October 28, 2021 with every hour. Table 1 shows the statistical information (maximum, mean, conditional mean i.e., during effective day time, and standard deviation) of sampled PV data from each location.

Table 1: Statistical information of the PV sites in Korea.

Location	**# of PVs**	**Statistical Parameter of PV data from each location**			
		Maximum	**Mean**	**Conditional Mean**	**Standard Deviation**
Gangwon	72	0.9439	0.1418	0.4334	0.2377
Gyeonggi	77	0.4336	0.0697	0.1932	0.1059
Gyeongnam	112	1.3644	0.2189	0.6407	0.3493
Gyeongbuk	147	0.8453	0.1374	0.3798	0.2102
Gwangju	39	0.9845	0.1499	0.4373	0.2324
Daegu	27	1.7818	0.3044	0.8400	0.4550
Daejeon	5	0.8604	0.0595	0.3635	0.1419
Busan	12	0.0475	0.0086	0.0229	0.0129
Seoul	8	1.1770	0.1986	0.5493	0.3030
Sejong	4	3.2630	0.5486	1.5490	0.8344
Ulsan	4	0.8950	0.1454	0.4104	0.2245
Incheon	10	0.7373	0.1287	0.3522	0.1932
Jeonnam	243	0.5957	0.0985	0.2836	0.1544
Jeonbuk	110	0.4478	0.0700	0.2030	0.1113
Chungnam	113	1.8684	0.3018	0.8458	0.4673
Chungbuk	51	1.3418	0.2063	0.5972	0.3247

31) Line 178: What is the interval of the ASOS weather data?

⇒ The interval of the ASOS weather data is one hour. The revised text is provided in our response of 33.

32) Line 180: Cleaning is misspelt

⇒ We found that data cleansing is a word that has the same meaning to data cleaning. Nevertheless, we revised the sentence.

The linear interpolation method is used to fill missing values after data cleaning.

33) Line 181: '... sampled every hour.' Sampled how, or on what basis? Using the hourly mean?

⇒ As mentioned at response 31, the interval of the ASOS weather data is one hour. Following the reviewer's careful review, we corrected the data description as follows.

> ASOS weather data provides 16 meteorological factors with every hour, and we select 4 meteorological factors: temperature, wind speed, humidity and cloud amount.

34) Line 184-185: Why the choice of 96 and 288?

⇒ The reason of the choice of 96 is that we used 4 meteorological factors in 24 hours a day from the single weather station (i,e, Voronoi cell). Therefore, the dimension of input data becomes $24 \times 4 = 96$. Then, the choice of 288 is that we used 4 meteorological factors in 24 hours a day from three weather stations (i,e, Delaunay triangle). In this case, the dimension of input data becomes $24 \times 4 \times 3 = 288$.

35) Line 198: Give MLP in full

⇒ MLP means multi-layer perceptrons. Following the reviewer's suggestion, we corrected a term with full form in Section 3.3. as follows.

> We compare the proposed TransGRU model with two well-known models: multi-layer perceptrons (MLP) and GRU [32].

36) Line 221-222: Why did you not use 5%, 10%, 15% and 20% of σ rather than 5%, 10%, 15% and 20% of the maximum value of each factor? What is the scientific basis for the latter choice? Is that typical of Korea's weather variations? Could you for example use information spanning longer periods e.g. from ASOS or other weather data sources to come up with the sensitivity factors for σ?

⇒ Please note that we used 5%, 10%, 15% and 20% of the *maximum* value of each factor, which coincides with the reviewer's comment as follows.

> To consider diverse weather forecast errors, we vary σ to the rate of 5%, 10%, 15% and 20% of the *maximum* value of each meteorological factors.

37) Line 236: '... are located in the way...'. Unclear what is meant by in the way

⇒ It means that PV site is located like Figure 7b, as an example, which is the case that DT weather stations can be very different from the 3-nearest weather stations. To clarify this, we added the following sentence.

We can infer that if PV sites are located in the way like green plus points in Figure 8b, which is the case that DT weather stations can be very different from the 3-nearest weather stations, using nearest 3 weather stations can be better than using DT.

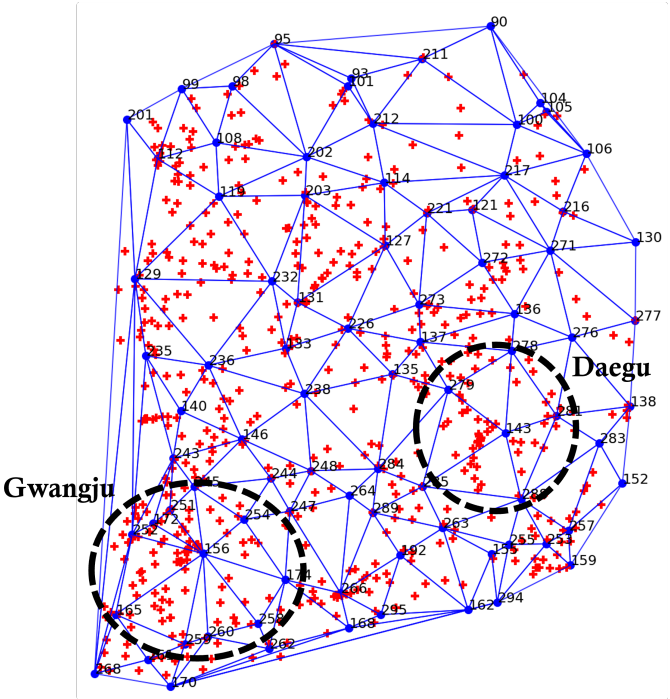
38) Line 240: '... $\epsilon_1 = 1.3$ and $\epsilon_2 = 2$...' What is the basis for these choices of ϵ_1 and ϵ_2?

⇒ The choices of ϵ_1 and ϵ_2 is based on validation results of comparison between DT weather stations and 3 nearest weather stations, which shows the pattern of the specific location of PV sites. To clarify this, we added the following sentence.

In the implementation of Union-Inner Triangles algorithm, we set the degree parameters $\epsilon_1 = 1.3$ and $\epsilon_2 = 2$. The choices of ϵ_1 and ϵ_2 is based on validation results of comparison between DT weather stations and 3 nearest weather stations, which shows the pattern of the specific location of PV sites.

39) Figure 8: Could Figure 8 results be due to location-specific parameters? Show e.g. on the map of Korea and in Figure 2 where these 2 sites (Daegu and Gwangju) are located. Include corresponding data for some other regions aside from just these 2 to show how consistently your method outperforms the other 2 methods.

⇒ We chose one from the east (Daegu) and the other from west (Gwangju) as shown in the figure below. We confirm that other sites also have similar results that our method outperforms the other 2 methods. However, it is hard to show the result of all 1034 PV sites.

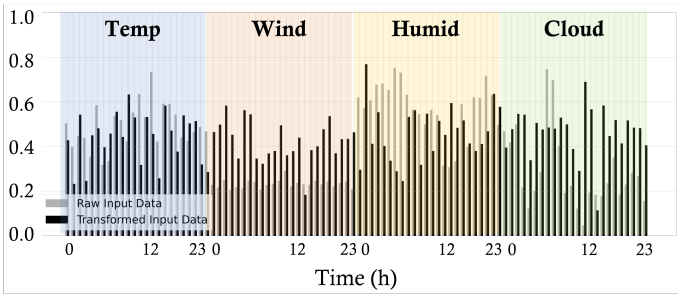


40) Line 251: 'selected weather stations A, B, and C.' Any basis for choosing specifically these 3 stations?

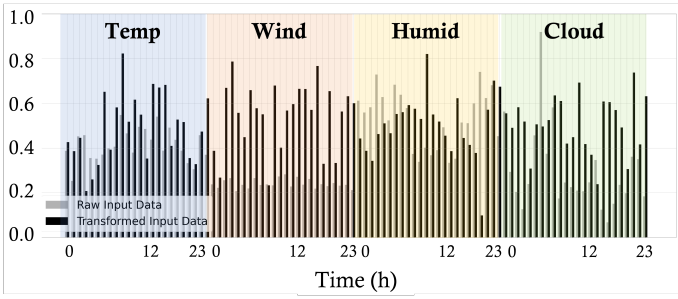
⇒ In this experiments, we chose randomly 3 weather stations and denoted by A, B, and C for illustration purpose.

41) Figure 10: Increase the horizontal axis length - the axis values' font size is too small

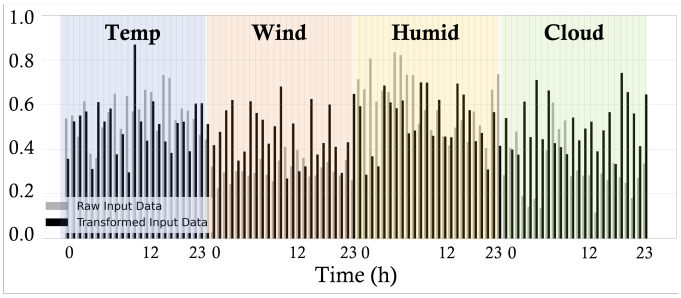
⇒ Following the reviewer's suggestion, we improved Figure 10 as follows.



(a) 4 meteorological factors of the weather station A



(b) 4 meteorological factors of the weather station B



(c) 4 meteorological factors of the weather station C

Figure 10. Comparison of raw input data and transformed input feature.

42) Line 254: Explain more how the inference is made.

⇒ Figure 9 shows the comparison of raw input data and transformed input feature. From this result, we can infer that TransGRU considers the importance of weather stations and meteorological factors, because transformed input feature is made by a part of TransGRU, i.e., Transformer encoder. The role of the Transformer encoder is to transform the input data to the learning-based feature data with same dimension.

From the transformed data, we can specifically infer which meteorological factor is highly considered at a certain time step because the role of the Transformer encoder is to learn and transform the input data to the feature data with the same dimension.

43) Line 255: ... at 'a' certain time step.

⇒ Following the reviewer's suggestion, we corrected a grammatical error in response 42.

44) Line 258-260: Show the location of these 3 weather stations and the target PV site in a figure.

⇒ We show the location of 3 weather stations and the target PV site in Figure 11 as follows.

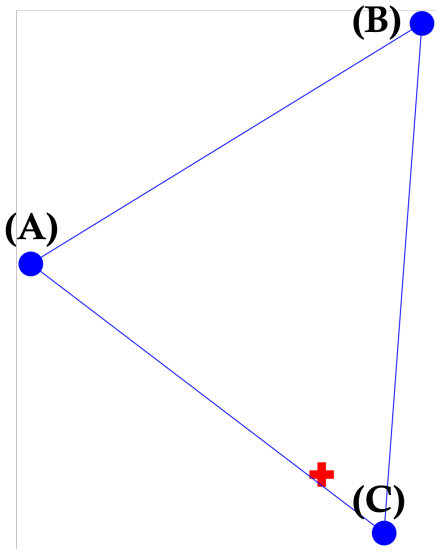


Figure 11. The location of 3 weather stations and the target PV site.

45) Line 272-273: How do you simulate limited data? By having e.g. a smaller training set? What is the effect e.g. of reducing the training set proportion?

⇒ Line 272-273 means that time-series based deep learning models, GRU and LSTM, require sufficient data to perform well. For example, in PV power forecasting, LSTM or GRU requires historical PV data for 3 years at least. However, our PV data have less than 2 years, so it does not satisfy the requirement of GRU or LSTM. As mentioned in Section 3.1, we split our PV data into training set (80%), validation set (10%), and test set (10%).

46) Figure 10: Why is the TransGRU gap more significant in Fig. 10 a for 15% weather forecast error than the other 3 cases of 5%, 10% and 20% error?

⇒ We generated weather forecast data by using additive white gaussian noise to ASOS weather data. Therefore, there may be slight difference such as Figure 10 due to the randomness of weather forecast data.

47) Line 278-279: 'For example, DT+ shows 19.9%, 16.0%, and 10.3% improvement compared to VT in 20% weather forecast error for Gwangju.' Unclear what you are referring to here, and which figures you are comparing

⇒ We revised the discussion of results as follows.

> In addition to CDF comparison, the average of NMAE10 in Gwangju and Daegu are shown in Table 4. In addition, we provide the result using VT to highlight the performance improvement using DT+ over VT in Table 4. As described in the table, DT+ outperforms VT for all the cases of weather forecasting error and the choice of learning model. For example, DT+ shows 19.9%, 16.0%, and 10.3% improvement compared to VT in 20% weather forecast error for Gwangju.

48) Line 292: Is this 30-45% improvement for all NMAE10% values in Table 4 relative to those in Table 3?

⇒ 30-45% improvement is for all NMAE10% values in Talbe 4 relative to those in Table 3.

49) Line 284: Provide some additional detail on the concept of VPPs

⇒ Virtual Power Plant (VPP) is a cloud-based distributed power plant that aggregates the capacities of heterogeneous distributed energy resources (DER) for the purposes of enhancing power generation, as well as trading or selling power on the electricity market. In this case, PV power generation of VPP form becomes smoother as more PV sites are aggregated, because individual PV sites have severe fluctuation due to stochastic weather conditions. Therefore, PV power forecasting in VPP form can be different from the individual PV sites case because forecasting errors from one site can be compensated by other sites.

> VPP is a cloud-based distributed power plant that aggregates the capacities of heterogeneous distributed energy resources (DER) for the purposes of enhancing power

generation, as well as trading or selling power on the electricity market. Since aggregating PV sites can improve the forecasting accuracy substantially, it is of interest to see the VPP performance of the proposed DTTrans model. Power generation from individual PV sites may severely fluctuate due to stochastic weather conditions, but power generation becomes smoother as more PV sites are aggregated. Therefore, PV aggregation can achieve economical profit by stable PV power generation. In this case, PV power forecasting can be different from the individual PV sites case because forecasting errors from one site can be compensated by other sites.

50) Line 320: '. . . as weather forecast error increases'. Unclear, rephrase

⇒ Following the reviewer's suggestion, we specifically revised the sentence as follows.

As the weather forecast error increases from 5% to 20%, TransGRU further outperforms the other methods based on MLP and GRU.