nature portfolio

Peer Review File

Realistic tropical cyclone wind and pressure fields can be reconstructed from sparse data using deep learning

Web links to the author's journal account have been redacted from the decision letters as indicated to maintain confidentiality

3rd May 23

Dear Mr Eusebi,

Your manuscript titled "Physics-Informed Neural Networks for Hurricane Reconstruction and Data Assimilation" has now been seen by 3 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they have raised quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version that fully addresses these serious concerns.

We hope you will find the reviewers' comments useful as you decide how to proceed. Should additional work allow you to address these criticisms, we would be happy to look at a substantially revised manuscript. If you choose to take up this option, please either highlight all changes in the manuscript text file, or provide a list of the changes to the manuscript with your responses to the reviewers.

We hope you will find the reviewers' comments useful as you decide how to proceed. For the publication of a revised manuscript in Communications Earth & Environment to be appropriate, we would need you to:

1) Provide new insights into how the PINN model can be applied to predict tropical cyclone intensity, and discuss the novelty and relevance of your work compared to the literature.

2) Demonstrate the robustness of your approach by applying it to other hurricanes, and compare predictions with more established forecasting models.

3) Compare the effects of different data models on reconstruction accuracy, especially around high-gradient regions.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

If the revision process takes significantly longer than three months, we will be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We understand that due to the current global situation, the time required for revision may be longer than usual. We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning. Of course, if you are unable to estimate, we are happy to accommodate necessary extensions nevertheless.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the reviewers' comments with a list of your changes to the manuscript text (which should be in a separate document to any cover letter) and any completed checklist:

[link redacted]

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further. Thank you for the opportunity to review your work.

Best regards,

Rahim Barzegar, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0002-1941-2991

EDITORIAL POLICIES AND FORMAT

If you decide to resubmit your paper, please ensure that your manuscript complies with our editorial policies and complete and upload the checklist below as a Related Manuscript file type with the revised article:

Editorial Policy Policy requirements (Download the link to your computer as a PDF.)

For your information, you can find some guidance regarding format requirements summarized on the following checklist:(https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf) and formatting guide (https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf).

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

Authors use PINN to reconstruct a hurricane based on a small number of observation data. They complete this for 2D and 3D simulated and real cases to estimate wind speeds throughout. They compare it to the modern SHiELD reconstruction technique for hurricanes.
Summary: The method is not novel and is simply the standard PINN method. However, the application is interesting. But the current results are very preliminary.

Major comments:
- Nothing is new in terms of methods. So, more analysis is required, such as how many data points are needed, where are those data points most useful, etc.

- More than 1 hurricane should be tested. A reconstruction of category 1 through 5 would be a much stronger argument that this works.
- The current study only reconstructs the initial condition. But the objective is to predict the hurricane. Authors should use the reconstructed initial condition for forecasting, and see how well the result is.

Minor comments:
- It is not clear from Figure 2 on page 7 how points are sampled from target. This is also not entirely evident in the text, please be more specific.
- All error calculations are not clear, on page 9 0.05% of what? L2 is the standard, this should be included as well in all trials. How about tables for the error to be easily visible.
- It is not clear in the Real Case section on page 10 how SHiELD and real data are combined, again be more specific.
- Figure 4 page 11, is not R^2 more common than R?

Reviewer #2 (Remarks to the Author):

The main idea of the paper is to introduce Physics-Informed Neural Networks (PINNs) as an alternative data assimilation approach for tropical cyclone (TC) by using a set of sparse observations together with physical models, which are described by a simplified version of horizontal Navier-Stokes equations shown in Eq. (3)-(6). The authors demonstrate that PINNs can accurately reconstruct full 2- and 3-dimensional velocity and pressure fields using synthetic training data from hurricanes simulated in a forecast model. The method is also tested in the case of Hurricane Ida using real-time observation.

It is noted that the idea of using PINNs as data assimilation has been applied to fluid dynamics, subsurface transport, and weather modeling, so the methodological novelties of the present study are not significant. Instead, this study focuses on expanding the potential applications of PINNs. But I think the application to TC is very interesting and with a great societal impact.

Overall, the manuscript is clear in describing the methods and presenting the results. I have the following comments that might be considered to improve the manuscript:
1. Whether the training data used for PINNs include both wind speed and pressure? For example, the pattern in Figure 1B is used to collect both wind speed and pressure data. It will help to fairly assess the performance of PINNs by providing the numbers of training data and collocation points used in the numerical examples.

2. The authors could provide more details on the specific normalization techniques used for the input data and equations, especially for those temporal quantities. For example, how to handle the time interval [-6,6]? This is an important step in applying neural networks to complex industrial problems, and readers would benefit from a more thorough explanation of how the authors approached this issue.

3. It will be of great interest to provide a simple comparison of the effects of different data patterns on reconstruction accuracy assuming a similar number of data collected for each pattern. I wonder if increasing collocation points around this high-gradient region will improve the accuracy.

4. The PINN reconstruction results seem to be over-diffusive compared to the target solution and lack accuracy near the storm center (e.g., Figure 2).

5. The Extended Data Figure 6 indicates that the PINN reconstructed speed is sensitive to relative weights γ? I am wondering if the authors try to use any regularization techniques (e.g., L2 regularizer) to stabilize the results. Besides, it would be helpful if the authors provide some guidance on how to select this parameter method in practice when the ground truth is not known.

6. Page 9: Compared to the 2D case, randomly sampled data are used in the 3D case. It would be great to provide more descriptions about whether a randomly distributed observation is realistic to collect in practice. Second, it is not clear whether sampled data at different time instances are used. Please clarify. Lastly, given that dense sampled data points are used but the distribution of reconstructed velocity has a high error (see Extended Data Figure 7), could the authors comment on how well PINNs perform compared to the conventional DA methods?

Other comments on methods:
P27, Line 441: The Coriolis parameter f seems a function of \phi? What are the definition of \phi and the relation to \beta in (3)? Please clarify.

P27, Line 449: Is the notation for vertical velocity a typo?

P32, Line 525: The statement about \Gamma >0.5 seems to be the opposite.

Reviewer #3 (Remarks to the Author):

Review of "Physics-Informed Neural Networks for Hurricane Reconstruction and Data Assimilation " by Eusebim, Vecchi, Lai, and Tong

By Altug Aksoy (University of Miami/CIMAS and NOAA/AOML/HRD)

28 April 2023

Recommendation: Major revision

Synopsis:

I applaud the authors for this exciting work and their demonstration that a physically informed neural network (PINN) can provide realistic hurricane vortex structures using available observations. The manuscript is written well and is very easy to follow. However, there are some critical questions that need to be addressed. For the most part, I thought that some of the claims made did not have the quantitative backing to justify them. I believe that this manuscript can be a useful contribution for the scientific community when these issues are resolved. The manuscript can be in publishable form only after my major and minor comments are addressed below.

Major comments:

1. My first major comment is about some of the claims made in the manuscript about how good the analyses obtained with PINN are without necessarily providing contextual quantitative evidence. This becomes especially problematic because in the introduction, the relatively slow progress made in improving intensity forecasts is clearly emphasized, which leads the reader to the expectation that the new methodology (PINN) would provide some improvements in this regard. However, in further reading the results, it becomes clear that this is not necessarily the case with the three examples given. I believe that one potential issue here is that the authors themselves fluctuate in their expectations from PINN. If this is only meant to be a demonstration that a PINN is a viable alternative to a much more complex and computationally expensive data assimilation system, then there is no need to try to claim an absolute accuracy in the analyses obtained. It is sufficient, in my mind, that it can be demonstrated that the analyses are physically realistic and reproduce the observed hurricane structures well qualitatively. Therefore, a demonstration such as this one can still be claimed to be successful even when there are some quantitative deficiencies in its ability to reproduce observed structures. I strongly recommend that the authors streamline their language in this regard, without shying away from insightful quantitative comparisons.

2. In analyzing their results, the authors make use of forecast fields obtained from a numerical model called SHiELD. This is a relatively obscure model, the performance of which is not known well by the hurricane community at large. In the examples discussed, SHiELD's performance also appears to be rather inconsistent, especially in the real-case example. For example, the horizontal wind speed structure generated in the SHiELD 12-h forecst(Fig. 4c) appears far inferior to the observed structure in the Tail Doppler Radar (TDR) horizontal wind speed structure (Fig. 4e). This, in turn, negatively impacts the quality of the comparison of the PINN-generated wind structure in the regression analysis (Fig. 4f). It is thus curious why the authors chose SHiELD as their reference model. Almost all operational centers provide their analysis and forecast fields to the research community, and I think these comparisons would benefit from including comparisons to more established forecast models.

3. Finally, some of the claims about the PINN in the discussion (L194-199) appeared inaccurate to me. While the computational advantages are extraordinary, it would help put this into better context since the number of observations that were considered in this PINN study is minimal compared to typical operational applications. More specifically, in the real-data example, only dropsonde and flight-level observations were considered, which amounts to a number of observations in the order O(3). However, in typical operational applications, this number is greatly increased both by radar and satellite observations and can easily reach order O(6) or higher. Therefore, it's not clear whether the computational expense would still not be greatly impacted by the number of observations in these scenarios. Furthermore, it is also mentioned in this paragraph that a PINN does not need an initial first-guess vortex. However, in the real-data case, the authors clearly mention that "The sparse observational data (Fig. 4d) we use is alone not enough to accurately reconstruct a realistic vortex, so we rely on the 12-hour SHiELD forecast to fill in the gaps" (L158-160). This is clearly inconsistent with the statement that a PINN does not need an initial first-guess structure. While this may be true hypothetically, in practical applications, it is clear that this is not the case in the application the authors describe. Finally, while a PINN may not need an expensive ensemble, this nevertheless doesn't deem ensembles useless. Quite to the contrary, ensembles are extremely useful and critical to predict the probabilistic aspects of tropical cyclones and will likely not be eliminated from our portfolio of prediction tools in the foreseeable future. What would be a more useful discussion here is whether techniques such as PINNs can be used to replace expensive

ensembles to predict the probabilistic aspects of forecasts.

Minor Comments:

L14: “for data assimilation for TC initial conditions” is awkward to read.

L42: I suggest either “TC *intensity* forecasting” or “the full wind and thermodynamic fields are” here because the TC structure obviously also involves very important thermodynamic variations.

L53-54: The constraint of linear adjustment applies to variational schemes, not ensembles.

L63-66: It would be useful for the reader to also mention here the common terminology “Observing System Simulation Experiment (OSSE)”.

L113-116: It’s not clear why the “observing” pattern was varied between “cross” and “plus” patterns at different times. Can the authors explain their reasoning here? Also, at the end of the sentence, it is mentioned that the patterns are motivated by the flight paths used for recon missions. I think this should be explained earlier when data points are first introduced in Fig. 1b. Otherwise, the choice of patterns seems arbitrary.

L117-130: Some of the claims of accuracy here are quite vague without providing better context for the errors that were obtained. One of the main motivations for this manuscript was that intensity (i.e., maximum surface wind speed) forecast errors remain a challenge for available forecasts. However, PINN itself greatly underestimates intensity by 12 m/s (57 m/s analyzed versus 69 m/s observed). This corresponds to an observed category-4 hurricane versus an analyzed category-3 hurricane, a significant difference. It should be also remembered that domain-averaged errors greatly underestimate the overall severity of the errors in the high-wind regions in a narrow band around the radius of maximum wind, so it is not clear at all whether the overall RMSE of 2.2 m/s is large or small in this context. From what I can tell in Fig. 2d between the dotted red (PINN) and blue (observed) lines, there also appears to be a discernable difference for what radius the maximum wind speed (RMW) occurs at. But it’s clear that the PINN analysis is greater than the observed, which would also have important consequences for the minimum sea-level pressure that PINN could attain through the wind-pressure relationship. But the authors’ description of these results generally gives the impression that this is a very successful analysis. I don’t necessarily dispute that, but the results need to be placed in better context to make these claims.

L138: For context, can the authors please provide the 2D equivalent of the percentage of grid points selected as data points?

Figure 3d/e: I strongly suggest changing these plots to radius-height-mean (r-z-mean) equivalents to avoid introducing arbitrary localized radial fluctuations into the comparison. This would also be a more direct comparison to Fig. 2d which is generated in the same manner, but for two dimensions. It would also allow the assessment of RMW, which I mentioned in my previous comment for L117-130 as a potential source of error in the PINN analyses.

L140-141: Can the authors please quantify the intensity error here?

L168-173: It is surprising to see that the forecast model SHIELD does not reproduce the inner-core high-wind region of Ida in this case (Fig. 4c), even though there clearly are dropsonde observations in that region that PINN seems to be responding to (Fig. 4b). Can the authors comment further here how SHIELD generates analyses and why this structure is not reproduced in its 12-h forecast? This is a very atypical structure that I usually don't observe in advanced numerical models at such short forecast times.

L176-177: There are two regions of where the horizontal wind speed is high that would typically contribute to the correlation between the TDR and analysis winds. The first is the inner core, where one typically expects the highest wind speed around RMW. This region is captured well in the PINN analysis thanks to the corresponding dropsonde observations, but surprisingly not captured well in the SHIELD forecast. Meanwhile, the second high-wind region, likely occurring within the primary rainband, is captured well in the SHIELD forecast but is absent in the PINN analysis. The correlation figure (Fig. 4f) is presumably the combination of these two regions and the inner-core region that PINN captures plays a more important role in the overall correlations. It would be useful to point this out for the reader.

We thank the reviewers for their very helpful and insightful comments and suggestions. Here's a summary of some of the main revisions:

1. We improve the model to better handle the eyewall/maximum winds, and described the methods used to do this
   a. One improvement/analysis we performed was finding out where to best have the collocation / data points while using the same number of those points. We find that more data points near the center and less collocation points near the center was better in all cases. We include details in the paper about this fact, and added a table with results that show evidence of this impact.
   b. Another improvement is we modified the cost function so the data points are weighted by their wind speed magnitude - i.e. the PINN will prioritize the strongest winds in its training. We describe this in detail in the manuscript.
2. Per reviewer 3's comments, we streamline the language throughout to focus more on qualitative evaluations of the model instead of quantitative. I still include the quantitative metrics for reference, but I don't use them as evidence for the model's success. I mostly emphasize that the PINN can recover large-scale characteristics, but struggles with small-scale features and high-gradient regions. We also emphasize that the PINN can be used for many applications besides initial conditions for forecast models, including storm surge modeling and risk assessment. The introduction and discussion sections have been substantially rewritten to include these clarifications and more clearly emphasize our goals and objectives with the PINN.
3. We include a performance table showing how the training time scales with changing the number of data points, collocation points, and the network structure to address the reviewers' concerns about how the performance scales with increasing training points.
4. In the Real Case, we included the reconstructed vortex for the initial conditions from the HAFS model, an established forecast model used by NOAA/NCEP and compare this reconstructed vortex to the TDR observations. The PINN actually has a much better correlation with TDR than the HAFS reconstruction, even though the HAFS reconstruction uses more data, including TDR data. This mainly happens because the HAFS reconstruction overestimates the extent/location of the maximum winds in the eyewall. I describe this in the paper, but emphasize this is only one example and it only offers an encouraging result that the PINN has some potential over the established DA methods.
5. I compare the training times/computational resources of the HAFS DA method and the PINN method in the manuscript.

REVIEWER COMMENTS and Author Responses:

Reviewer #1 (Remarks to the Author):

Authors use PINN to reconstruct a hurricane based on a small number of observation data. They complete this for 2D and 3D simulated and real cases to estimate wind speeds throughout. They compare it to the modern SHiELD reconstruction technique for hurricanes. Summary: The method is not novel and is simply the standard PINN method. However, the application is interesting. But the current results are very preliminary.

Major comments:
- Nothing is new in terms of methods. So, more analysis is required, such as how many data points are needed, where are those data points most useful, etc.
Thank you for the feedback - we put in more analysis to better understand the model and how to improve the model training. We improved the model from the previously submitted version of the paper and described the methodologies behind these improvements. We describe in the “Physics Informed Neural Networks” and “Methods” section how we weighted the data points in the cost function by their respective wind speed magnitudes to motivate the model to prioritize the strong winds which only covered a small area, and the eyewall of the storm generally. We also describe in the “Real Case” section how we weighted the real observations twice as much as the data points sampled from SHiELD. Regarding collocation and data points, we describe and show (Fig. 1c, 3b, and 3e) where they were best located. As expected the model performed better with more data points near the center of the storm, but performed better when there were less collocation points near the center of the storm.

- More than 1 hurricane should be tested. A reconstruction of category 1 through 5 would be a much stronger argument that this works.
Thank you for the suggestion - we note that while in all sections we are reconstructing Hurricane Ida, in the 2D and 3D case we are reconstructing category 4 Ida, while in the Real case we are constructing a strong tropical storm / category 1 hurricane Ida. Despite being the same storm, it had very different wind fields and maximum winds at the different times we look at. So this is evidence of the PINN handling different intensities well. Nonetheless, we agree looking at more storms, especially in the real case, would be a more robust argument that the method works. In line with reviewer 3’s comments, we focus more on the qualitative success of the model, and we believe it would be very difficult to focus qualitatively on multiple storms, and still produce a concise writeup. The PINN method is very general, and in principle should work on other storms, and its behavior as described in the “Real Case” section seems predictable. We plan to leave further analysis of the model’s performance for future work.

- The current study only reconstructs the initial condition. But the objective is to predict the hurricane. Authors should use the reconstructed initial condition for forecasting, and see how well the result is.
The objective of the current study is actually just to reconstruct the initial condition (generally, to reconstruct the full flow field of the storm at a given time). We apologize for any confusion in the paper, the language used throughout the paper – especially in the abstract and introduction – has been modified to emphasize that the main goal is to reconstruct the wind and pressure fields of the storm, and that initial conditions for forecasting is an important application of these reconstructed fields (in addition to other important applications). This work is intended to be a proof-of-concept that this method works in this very important and topical setting (hurricane wind fields) and we hope it will generally promote more cross-disciplinary applications of machine learning to hurricane forecasting.

Using the reconstructed fields in a forecast model and is a very large undertaking - discussions are underway about starting this project of using the PINN in a forecast model, but we reserve it for future papers.

Minor comments:
- It is not clear from Figure 2 on page 7 how points are sampled from target. This is also not entirely evident in the text, please be more specific.

More information has been added to the figure caption and the text of the 2D case section to describe how the points are sampled. We sample data points at 3 time points: hours -3, 0, and 3 (where hour 0 is Aug 29 12z in the forecast). At hours -3 and +3, a cross pattern is used, and at hour 0 a plus pattern is used (these patterns are shown in figure 2c and figure 1c).

- All error calculations are not clear, on page 9 0.05% of what? L2 is the standard, this should be included as well in all trials. How about tables for the error to be easily visible.

The sentence you are referring to states "unlike the 2D case, the PINN's data points consist of 0.05% of the grid points randomly sampled from the 3D SHiELD output." The 0.05% is not a measurement of error, we are just stating that we are using a certain percentage of the available data from the target output to train the model.

However, we do agree that some of our error calculations were not clear. When providing wind speed RMSE in the 2D and 3D case, we now avoid domain-integrated averages and focus on radial wind speed RMSE (which better factors how the model does in regions of stronger winds). For equation loss metrics, we emphasize the equation loss should be compared to the equation term magnitudes from the PDES, and these equation term magnitudes are shown in supplementary figures 4 and 5. However, in line with the criticisms from reviewer 3, we generally deemphasize the quantitative arguments in this iteration of the paper and emphasize qualitative evidence that the model is working.

- It is not clear in the Real Case section on page 10 how SHiELD and real data are combined, again be more specific.

We clarify in the Real Case section that the points sampled from the SHiELD forecast and the real observations are treated identically as "data points" in the model (they are all part of 1 big array the PINN sees as its training data points). However, we did add one modification in this version of the paper/model that weighs the real observations twice as high as the shield data points so the model can prefer the observations over the SHiELD data where there are any discrepancies (this was not a feature of the model in our last submitted manuscript).

- Figure 4 page 11, is not R^2 more common than R?

Thanks for the comment. It is true that R^2 is commonly used when evaluating the strength of a model. However, our intention is not to imply that the correlation is indicative of the strength of the model. Our main goal is to show how correlated the patterns in the different models are with the TDR observations, and R is most commonly used for spatial patterns.

Reviewer #2 (Remarks to the Author):

The main idea of the paper is to introduce Physics-Informed Neural Networks (PINNs) as an alternative data assimilation approach for tropical cyclone (TC) by using a set of sparse observations together with physical models, which are described by a simplified version of horizontal Navier-Stokes equations shown in Eq. (3)-(6). The authors demonstrate that PINNs can accurately reconstruct full 2- and 3-dimensional velocity and pressure fields using synthetic training data from hurricanes simulated in a forecast model. The method is also tested in the case of Hurricane Ida using real-time observation.

It is noted that the idea of using PINNs as data assimilation has been applied to fluid dynamics, subsurface transport, and weather modeling, so the methodological novelties of the present study are not significant. Instead, this study focuses on expanding the potential applications of PINNs. But I think the application to TC is very interesting and with a great societal impact.

Overall, the manuscript is clear in describing the methods and presenting the results. I have the following comments that might be considered to improve the manuscript:
1. Whether the training data used for PINNs include both wind speed and pressure? For example, the pattern in Figure 1B is used to collect both wind speed and pressure data. It will help to fairly assess the performance of PINNs by providing the numbers of training data and collocation points used in the numerical examples.

Thank you for the suggestion. Figure 1 was modified to stress that the training data consists of the 2 horizontal components of the wind field and the geopotential (which is essentially information about the pressure field). The language in the “Physics-Informed Neural Networks” section and “Methods” section was also updated to clarify this.

The numbers of data and collocation points have been explicitly added in the text for each of the three cases shown in the paper.

2. The authors could provide more details on the specific normalization techniques used for the input data and equations, especially for those temporal quantities. For example, how to handle the time interval [-6,6]? This is an important step in applying neural networks to complex industrial problems, and readers would benefit from a more thorough explanation of how the authors approached this issue.

This information has been added to the “Physics Informed Neural Networks” section (end of paragraph 3) and the methods section. We implement a very straightforward normalization technique to get all values to be between 0 and 1 (similar magnitudes). This is done to ensure that the different variables and different PDEs in the loss function all have similar magnitudes so one isn’t preferred over another.

3. It will be of great interest to provide a simple comparison of the effects of different data patterns on reconstruction accuracy assuming a similar number of data collected for each pattern. I wonder if increasing collocation points around this high-gradient region will improve the accuracy.

For conciseness, we have omitted that figure for now and focused more on showing the results for what worked. If the reviewers still believe that is a priority and there is space for it, we can add another extended data figure showing the results with some of the other data patterns that didn’t work.

We also thought increasing collocation points in the high-gradient region would improve the accuracy. We tried it for the revisions and found that using less collocation points near the center and allowing it to prioritize the data points in that region was better. Essentially, the PDEs were acting as a regularizer to prevent the PINN from attaining the higher wind speeds –we have written about this effect in this new manuscript. Thanks for the suggestion.

4. The PINN reconstruction results seem to be over-diffusive compared to the target solution and lack accuracy near the storm center (e.g., Figure 2).

You are correct - since the last submitted manuscript, we have improved the model for this latest version and have written about the changes we made to help. These changes include having less collocation points near the center of the storm and weighting data points in the cost function by their wind speed magnitude. We write about these changes in the paper. However, the PINN still struggles with high-gradient regions and general small-scale features, so in this revised manuscript we stress this potential downside of the model.

5. The Extended Data Figure 6 indicates that the PINN reconstructed speed is sensitive to relative weights γ? I am wondering if the authors try to use any regularization techniques (e.g., L2 regularizer) to stabilize the results. Besides, it would be helpful if the authors provide some guidance on how to select this parameter method in practice when the ground truth is not known.

Thank you for the comment. The PDEs in our PINN essentially act as a regularizer to constrain our set of possible solutions that fit our data points to a more realistic, physically realizable set of solutions. This unintentionally makes it harder for the PINN to attain the higher wind speed values, since stronger regions in the flow field will have higher term magnitudes in the equations (see extended data fig 4 and 5) and thus higher equation losses. Adding L2 regularization to the model would inevitably make it even harder for the PINN to reach those higher wind speed values.

Generally, if you focus too much on the equation loss, then the “regularizer” in our model will be too strong and the PINN won’t be able to get the higher wind speeds (note the colorbars in the plots are different for the different gamma values). But if our model prioritized the data loss too much on the other hand, then it would overfit our sparse data and we would get an unrealistic solution. Generally, it takes experimentation to find the optimal gamma value, and this could depend on how the variables and equations are non-dimensionalized. This fact has been added to the paper.

6. Page 9: Compared to the 2D case, randomly sampled data are used in the 3D case. It would be great to provide more descriptions about whether a randomly distributed observation is realistic to collect in practice. Second, it is not clear whether sampled data at different time instances are used. Please clarify. Lastly, given that dense sampled data points are used but the distribution of reconstructed velocity has a high error (see Extended Data Figure 7), could the authors comment on how well PINNs perform compared to the conventional DA methods?

Thanks for the suggestion. We have added details that express that the randomly sampled point, and especially having those randomly sampled points cover the full domain, is unrealistic. However, we express that it is realistic to have more observations in the center of the storm as opposed to outside the center of the storm since that is the region most observing systems focus on. We also stress that there are many other forms of data, such as TDR and satellite data, which could provide more information about the flow field throughout the whole domain, in a similar manner that our random observations do. The purpose of this section was more to show that in theory with sufficient observations, the field can be reconstructed. In the Real Case section, our observations are not enough, and so we make use of forecast data to fill in the gaps.

Sorry for the confusion, we stress in the text now that data points are sampled from three time points: hour -3, 0, and +3, like the 2D case. The results shown in the figure are at hour 0.

The extended data figure you are referring to is an example where we train the PINN using just the SHiELD forecast data and no real observations, so the high error is expected in this case. This figure was added to demonstrate how the addition of the real observations affects the PINN output. We have removed this figure in this iteration of the manuscript since it is more clear the role the real observations are playing in the PINN's output. In Figure 4, we show TDR data at the 850hPa level, which could be regarded as the ground truth data that we would like to reconstruct, even though this data is not seen during training (just the flight-level and dropsonde observations). For this version of the manuscript, we have also added the reconstructed vortex at this time from the DA system used by the HAFS model (a popular forecast model used by NOAA/NCEP) to be compared with the PINN reconstruction. In panel f, we print the correlations of the PINN and the HAFS reconstruction with the TDR data – the PINN has a higher correlation, despite the HAFS system using much more data (including the TDR data). We stress this is only one example, but that it highlights the potential of the PINN for reconstructing the large-scale flow of the storm.

Other comments on methods:
P27, Line 441: The Coriolis parameter f seems a function of \phi? What are the definition of \phi and the relation to \beta in (3)? Please clarify.

Thanks for the comment. We clarify in the text that f_o is the value of f at the latitude of our storm center. We define \beta as the meridional gradient in f (df/dy), and we define our approximation of f as f_o + \beta * y as the commonly used beta-plane approximation.

P27, Line 449: Is the notation for vertical velocity a typo?

Sorry for the confusion, it is common practice to denote the vertical velocity in pressure coordinates as ω, as opposed to the usual ‘w’ when working with ‘z’ as the vertical coordinate.

P32, Line 525: The statement about $\Gamma >0.5$ seems to be the opposite.

Thanks for the catch, you’re right. This statement had been written regarding a previous iteration of the model. The language has been updated to reflect the current model/results shown in the paper.

Reviewer #3 (Remarks to the Author):

Review of “Physics-Informed Neural Networks for Hurricane Reconstruction and Data Assimilation " by Eusebim, Vecchi, Lai, and Tong

By Altug Aksoy (University of Miami/CIMAS and NOAA/AOML/HRD)

28 April 2023

Recommendation: Major revision

Synopsis:

I applaud the authors for this exciting work and their demonstration that a physically informed neural network (PINN) can provide realistic hurricane vortex structures using available observations. The manuscript is written well and is very easy to follow. However, there are some critical questions that need to be addressed. For the most part, I thought that some of the claims made did not have the quantitative backing to justify them. I believe that this manuscript can be a useful contribution for the scientific community when these issues are resolved. The manuscript can be in publishable form only after my major and minor comments are addressed below.

Major comments:

1. My first major comment is about some of the claims made in the manuscript about how good the analyses obtained with PINN are without necessarily providing contextual quantitative evidence. This becomes especially problematic because in the introduction, the relatively slow progress made in improving intensity forecasts is clearly emphasized, which leads the reader to the expectation that the new methodology (PINN) would provide some improvements in this regard. However, in further reading the results, it becomes clear that

this is not necessarily the case with the three examples given. I believe that one potential issue here is that the authors themselves fluctuate in their expectations from PINN. If this is only meant to be a demonstration that a PINN is a viable alternative to a much more complex and computationally expensive data assimilation system, then there is no need to try to claim an absolute accuracy in the analyses obtained. It is sufficient, in my mind, that it can be demonstrated that the analyses are physically realistic and reproduce the observed hurricane structures well qualitatively. Therefore, a demonstration such as this one can still be claimed to be successful even when there are some quantitative deficiencies in its ability to reproduce observed structures. I strongly recommend that the authors streamline their language in this regard, without shying away from insightful quantitative comparisons.

Thank you for this suggestion. We agree with your sentiments and have modified/streamlined the language throughout the paper to emphasize the qualitative success of the model in reproducing large-scale features of the storm it is being trained on. We kept quantitative measures of how it fits the target storm for reference, but removed language that implies that these quantitative metrics alone imply a successful reconstruction. We added more language to stress weaknesses in the model, including how it struggles with high-gradient regions and maximum winds, and generally doesn't reproduce the small-scale features seen in the true fields. Let us know if there are any other ways in which we should modify/streamline the language.

Note that since the previous manuscript, we have improved the model to address some of its previous weaknesses. For instance, we modified the cost function by weighting data points by their wind speed - this caused an increase of 10-20% in the PINN's maximum winds in all our cases, closer to the target's maximum wind speeds.

2. In analyzing their results, the authors make use of forecast fields obtained from a numerical model called SHiELD. This is a relatively obscure model, the performance of which is not known well by the hurricane community at large. In the examples discussed, SHiELD's performance also appears to be rather inconsistent, especially in the real-case example. For example, the horizontal wind speed structure generated in the SHiELD 12-h forecst(Fig. 4c) appears far inferior to the observed structure in the Tail Doppler Radar (TDR) horizontal wind speed structure (Fig. 4e). This, in turn, negatively impacts the quality of the comparison of the PINN-generated wind structure in the regression analysis (Fig. 4f). It is thus curious why the authors chose SHiELD as their reference model. Almost all operational centers provide their analysis and forecast fields to the research community, and I think these comparisons would benefit from including comparisons to more established forecast models.

Thank you for the comment. While the SHiELD model is relatively new, we believe it is still of high quality. We clarify that we are specifically using T-SHiELD. We added reference 38 which is a study that directly compares T-SHiELD with the more well-established HAFS model, in addition to the other publications cited which provide other assessments of T-SHiELD's performance.

Nonetheless, we believe the quality of SHiELD doesn't necessarily affect our conclusions in the Real Case section and figure 4. We stress that the SHiELD data we are showing is a 12-hour forecast at 12z Aug 27 that was initialized at 0z Aug 27. The PINN is just a DA scheme that is trying to reconstruct the flow at 12z aug 27. Our main objective in showing the comparison between the PINN and SHiELD is to highlight how the PINN is capable of combining forecast data with real observations (much like most modern DA schemes) to produce a physically consistent TC vortex that matches observations well. In showing the correlations in panel f, we are just highlighting that the PINN can take the SHiELD representation and adjust it to match the observations. We were not trying to imply that since the PINN is "more accurate than SHiELD," then it is "good." It is expected that the PINN representation should be better than the SHiELD representation since it is directly trained with the observational data at that time. The language in the paper has been updated to clear up any confusion about this.

We also added the analysis from the HAFS model DA scheme at this time for direct comparison with the PINN. We stress that these models can be compared with each other, but it doesn't make too much sense to compare SHiELD with them. The HAFS model is much more developed (it even uses the TDR data we compare against) and the PINN is still in its early stages of development, but the PINN is still much more correlated with the TDR data than the HAFS DA scheme. We stress that this is just one example, but it offers the encouraging result that the PINN can produce realistic TCs that correlate well with the real observations. We also hope that it gives the readers an idea of how are PINN differs from modern DA schemes (modern DA schemes reproduce lots of fine-scale features but they aren't always in the right spot. The PINN doesn't reproduce fine-scale features, but it seems to get the large-scale features more correctly, possibly due to the way our PDEs and collocation points are enforcing a physically consistent solution throughout the domain).

3. Finally, some of the claims about the PINN in the discussion (L194-199) appeared inaccurate to me. While the computational advantages are extraordinary, it would help put this into better context since the number of observations that were considered in this PINN study is minimal compared to typical operational applications. More specifically, in the real-data example, only dropsonde and flight-level observations were considered, which amounts to a number of observations in the order O(3). However, in typical operational applications, this number is greatly increased both by radar and satellite observations and can easily reach order O(6) or higher. Therefore, it's not clear whether the computational expense would still not be greatly impacted by the number of observations in these scenarios. Furthermore, it is also mentioned in this paragraph that a PINN does not need an

initial first-guess vortex. However, in the real-data case, the authors clearly mention that "The sparse observational data (Fig. 4d) we use is alone not enough to accurately reconstruct a realistic vortex, so we rely on the 12-hour SHiELD forecast to fill in the gaps" (L158-160). This is clearly inconsistent with the statement that a PINN does not need an initial first-guess structure. While this may be true hypothetically, in practical applications, it is clear that this is not the case in the application the authors describe. Finally, while a PINN may not need an expensive ensemble, this nevertheless doesn't deem ensembles useless. Quite to the contrary, ensembles are extremely useful and critical to predict the probabilistic aspects of tropical cyclones and will likely not be eliminated from our portfolio of prediction tools in the foreseeable future. What would be a more useful discussion here is whether techniques such as PINNs can be used to replace expensive ensembles to predict the probabilistic aspects of forecasts.

Thank you for the comment. First, we note that the PINN we train in the "Real Case" uses a combination of real observations from flight-level/dropsonde and synthetically sampled data points from SHiELD. These two sources of data are treated nearly identically in the model as "observations." So while we only use O(10^3) flight-level/dropsonde points, the PINN uses O(10^4) data points in total. Specifically, it uses 54,663 data points (this number has been added to the text). We have added a performance table (Extended Data Table 2) showing how well the PINN performs (training time) using various amounts of data/collocation points and different neural network architectures. Generally, the PINN's training time does not drastically increase until it reaches about (10^6) data points, which we note in this version of the manuscript. It doesn't change much at lower amounts since collocation points are more expensive than data points (requires calculations of numerous derivatives), and therefore increasing data points doesn't have too strong of a computational impact. However, the PINN is limited by memory space in GPUs, which is why the training time sharply increases around (10^6). With more careful care, these performance issues can be alleviated by using more advanced / newer GPUs which could offer order of magnitude speed increases, or by designing the code to train using multiple GPUs to fix the memory error. Some of these notes have been included in the manuscript. We also stress that the PINN does not generally pick up the very small-scale features, so it would generally not be necessary to use 10^6 data points, unless the PINN eventually becomes sophisticated enough to handle all of those features.

Regarding the initial-guess vortex, you are correct that our statement is misleading since we did use data sampled from an initial-guess vortex. The language in the manuscript has been updated to stress that while in this case we do make use of an initial vortex, with sufficient observations, it wouldn't be necessary. And without sufficient observations, we would only need information from the areas without observations. Current schemes, as we understand them, must start from a full initial vortex grid, and perform corrections to it. So even if, for example, the eyewall was fully observed, the DA scheme would still need to start with a potentially erroneous initial vortex including the eyewall, which could limit/hurt the assimilation accuracy. On the other hand, the PINN would be able to use just the observations of the eyewall in this case and only use sampled data from the forecast model in regions away from the eyewall that are not fully observed.

You are correct that ensemble forecasts are very valuable, and our statement was not intended to diminish their importance. We note that we distinguish between the ensembles used in the DA method to determine flow covariances and the ensembles used when running the actual forecast. We believe the ensembles of the forecast model runs will always be important, but the ensembles required solely to generate the initial conditions are expensive and an upside of the PINN is that it does not require expensive ensemble runs to generate its reconstruction. However, we have omitted this argument from the new manuscript now to avoid any confusion (or in case we are wrong and the ensemble element of other DA schemes is not a big computational factor).

Minor Comments:

L14: “for data assimilation for TC initial conditions” is awkward to read.

Thanks for the suggestion, this has been changed to “for TC data assimilation”

L42: I suggest either “TC *intensity* forecasting” or “the full wind and thermodynamic fields are” here because the TC structure obviously also involves very important thermodynamic variations.

Thank you - the language has been changed to your latter suggestion.

L53-54: The constraint of linear adjustment applies to variational schemes, not ensembles.

Thank you, the text has been updated to indicate this is only a weakness for the variational schemes.

L63-66: It would be useful for the reader to also mention here the common terminology “Observing System Simulation Experiment (OSSE)”.

Thank you, the terminology has been added.

L113-116: It’s not clear why the “observing” pattern was varied between “cross” and “plus” patterns at different times. Can the authors explain their reasoning here? Also, at the end of the sentence, it is mentioned that the patterns are motivated by the flight paths used for recon missions. I think this should be explained earlier when data points are first introduced in Fig. 1b. Otherwise, the choice of patterns seems arbitrary.

An explanation of the alternating pattern was added to the 2D case section – essentially alternating the patterns allows the PINN to get more information about the full spatial extent of the storm while still using the same amount of data at each time point. We also clarify that the data points are sampled at hours -3, 0, and 3 relative to the time point we focus on in the plots.

We still kept the full explanation of the data patterns and the flight paths in the 2D case section – Figure 1 in the “Physics-Informed Neural Network” section is supposed to be general, and applicable for the 2D, 3D, and real case. However, we needed to show an example of the data points, so we chose the 2D case example and stressed in the caption that these are the data points just for the 2D case. Generally, we wanted to keep the “Physics-Informed Neural Network” section as general as possible and keep implementation details specific to a certain case in that case’s section. Let us know if you still disagree and think an earlier explanation would be useful.

L117-130: Some of the claims of accuracy here are quite vague without providing better context for the errors that were obtained. One of the main motivations for this manuscript was that intensity (i.e., maximum surface wind speed) forecast errors remain a challenge for available forecasts. However, PINN itself greatly underestimates intensity by 12 m/s (57 m/s analyzed versus 69 m/s observed). This corresponds to an observed category-4 hurricane versus an analyzed category-3 hurricane, a significant difference. It should be also remembered that domain-averaged errors greatly underestimate the overall severity of the errors in the high-wind regions in a narrow band around the radius of maximum wind, so it is not clear at all whether the overall RMSE of 2.2 m/s is large or small in this context. From what I can tell in Fig. 2d between the dotted red (PINN) and blue (observed) lines, there also appears to be a discernable difference for what radius the maximum wind speed (RMW) occurs at. But it’s clear that the PINN analysis is greater than the observed, which would also have important consequences for the minimum sea-level pressure that PINN could attain through the wind-pressure relationship. But the authors’ description of these results generally gives the impression that this is a very successful analysis. I don’t necessarily dispute that, but the results need to be placed in better context to make these claims.

Thank you for the comment. We agree that the full domain-averaged RMSE stat is misleading, and have removed it from the main text (we only show the full domain averaged RMSE in Extended Data Table 1). You are correct that the PINN underestimates the maximum winds, and the language in the text has been updated to stress this weakness. This is largely a result of the PINN not handling sharp gradients well – since these strongest winds only occupy a few grid points, it is very difficult for the PINN to detect them. Nonetheless, the model has been improved since the last paper to better handle these higher winds (descriptions in the paper about how). The maximum winds in the 2D case are up to 62 m/s, and 58 m/s in the 3D case. We note that better results can be obtained by using larger neural network structures at the expense of a longer training time. However, there are many performance upgrades the PINN can receive, including using more advanced GPUs or more GPU cores, so these better results are feasible.

We have updated the language in the paper to stress that the PINN might be able to improve intensity forecasts by obtaining a better representation of the flow field overall (getting the large-scale characteristics of the storm very well) for the initial conditions of a forecast model, not necessarily by reconstructing those strongest winds perfectly for the initial conditions. The results in the Real Case suggest the PINN can produce qualitatively accurate vortices which capture the large-scale structure, and future work is needed to determine if this can improve forecasts (especially intensity forecasts) when this reconstructed TC is used as an initial condition in a forecast model. We also stress in the paper that there are other applications of the PINN reconstruction outside of initial conditions for forecast models (storm surge modeling and risk assessment).

The language has also been updated to acknowledge that the reconstruction does have flaws and limitations, and these flaws could hurt its success as an initial condition in a forecast model

L138: For context, can the authors please provide the 2D equivalent of the percentage of grid points selected as data points?

This percent was approximately 2.4%, and has been added to the text in the 2D case section.

Figure 3d/e: I strongly suggest changing these plots to radius-height-mean (r-z-mean) equivalents to avoid introducing arbitrary localized radial fluctuations into the comparison. This would also be a more direct comparison to Fig. 2d which is generated in the same manner, but for two dimensions. It would also allow the assessment of RMW, which I mentioned in my previous comment for L117-130 as a potential source of error in the PINN analyses.

Thank you for the suggestion, the plot has been modified to show radius-height-means. You are correct that the PINN does struggle a bit with the radius of maximum wind in the 3D case, but less now than in the previous iteration of the model shown in the last submitted manuscript. Some changes were made to the model which are described in the text that allowed the model to handle the eyewall much better. In the 2D case, it seems to get the RMW nearly exactly right, and the 3D case it seems to be struggling a little bit. The PINN definitely still struggles with the high-gradients, however, causing it to overestimate winds closer to the eye and underestimate winds in the exact eyewall – these errors are clear in this new radius-height-RMSE plot.

L140-141: Can the authors please quantify the intensity error here?

The error has been included

L168-173: It is surprising to see that the forecast model SHIELD does not reproduce the inner-core high-wind region of Ida in this case (Fig. 4c), even though there clearly are dropsonde observations in that region that PINN seems to be responding to (Fig. 4b). Can the authors comment further here how SHIELD generates analyses and why this structure is not reproduced in its 12-h forecast? This is a very atypical structure that I usually don't observe in advanced numerical models at such short forecast times.

Thank you for this comment. We have realized that the way in which we were using the SHiELD forecast in this section might have been vague/confusing. We stress that this was a 12hr forecast that was initialized at 0z Aug 27, and so the forecast data we are showing is at 12z Aug 27. The storm looked different at 0z and never would have seen the dropsonde observations at 12z.

Our main objective in this section is to show that a PINN can combine forecast data with real observations to produce a physically consistent TC which recovers much of the key TC characteristics indicated by the observations. We are not directly comparing the PINN against SHiELD because they are doing two very different things – SHiELD is running a forecast that was initialized on 0z Aug 27, and the PINN is just a DA scheme that we are using to reconstruct the flow at 12z Aug 27. It is expected that the PINN would perform better than SHiELD in the TDR analysis since it is essentially taking the best of SHiELD and the real observations, and merging them in a physically realizable way. Our main objective is to show that like many modern DA schemes, it can take forecast data from a forecast model and adjust it to a vortex which better matches the current observations. The quality of SHiELD should not affect our conclusion that the PINN can merge SHiELD data and observations to reconstruct a more accurate vortex, and we include the correlation values with TDR to stress this.

We added a panel to figure 4 showing the reconstructed vortex at this time (12z) by the DA scheme from the HAFS model, a popular model used by NOAA/NCEP, to show what a conventional scheme in a more trusted forecast model would look like, and we show the correlation of this analysis with the TDR data. We stress that the HAFS DA scheme is much more sophisticated and developed than the PINN (and even assimilates much more data including the TDR data), but the PINN still has a better correlation with the wind field in TDR.

L176-177: There are two regions of where the horizontal wind speed is high that would typically contribute to the correlation between the TDR and analysis winds. The first is the inner core, where one typically expects the highest wind speed around RMW. This region is captured well in the PINN analysis thanks to the corresponding dropsonde observations, but surprisingly not captured well in the SHIELD forecast. Meanwhile, the second high-wind region, likely occurring within the primary rainband, is captured well in the SHIELD forecast but is absent in the PINN analysis. The correlation figure (Fig. 4f) is presumably the combination of these two regions and the inner-core region that PINN captures plays a more important role in the overall correlations. It would be useful to point this out for the reader.

Thank you for the comment. We have added more precise descriptions of the ways in which the PINN and HAFS reconstructions and the SHiELD forecast are close or not close to the TDR data.

8th Nov 23

Dear Mr Eusebi,

Your manuscript titled "Physics-Informed Neural Networks for Hurricane Reconstruction and Data Assimilation" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment under the open access CC BY license (Creative Commons Attribution v4.0 International License).

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. In particular, we ask you to:
• Specify and justify the choice of the coordinate system used in the PINN analyses.
• Clarify the rationale behind using correlation as the main performance metric in the analysis and consider additional metrics that could provide a more nuanced evaluation of the PINN model's effectiveness.

At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

Please note that it may still be possible for your paper to be published before the end of 2023, but in order to do this we will need you to address these points as quickly as possible so that we can move forward with your paper.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at https://www.nature.com/documents/commsj-file-checklist.pdf .

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication under a CC BY license (Creative Commons Attribution 4.0 International License). This license allows maximum dissemination and re-use of open access materials and is preferred by many research funding bodies.

For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit https://www.nature.com/commsenv/article-processing-charges

At acceptance, you will be provided with instructions for completing this CC BY license on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:
[link redacted]
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Alienor Lavergne, PhD
Associate Editor
Communications Earth & Environment

Rahim Barzegar, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0002-1941-2991

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

Although the PINN method is not novel, the application of PINN in hurricane is interesting. The revised paper has answered my technical questions, but I am not a domain expert in hurricane and cannot comment too much on the contribution to hurricane's part.

Reviewer #2 (Remarks to the Author):

The authors have addressed most of my comments, and I am satisfied with the revised manuscript and willing to recommend for publication.

I only have minor comments regarding Figure 4, which presents a comparative analysis between the PINN and the conventional HAFS DA method. I'm wondering if using correlation as an indicator is appropriate for assessing modal performance because, in my opinion, it is difficult to tell which model works better from the wind speed reconstruction. Additionally, it would be valuable to include more instances, considering different times and locations, for a comprehensive analysis of their performance.

Reviewer #3 (Remarks to the Author):

Review of "Physics-Informed Neural Networks for Hurricane Reconstruction and Data Assimilation " by Eusebim, Vecchi, Lai, and Tong

By Altug Aksoy (University of Miami/CIMAS and NOAA/AOML/HRD)

26 October 2023

Recommendation: Minor revision

Synopsis:

This is the first revised version of the manuscript and I find it much improved. The authors have addressed my concerns regarding the original version, and I only have one major comment along with some minor editorial ones. It shouldn't take the authors too much effort to address these and therefore I recommend that the manuscript be accepted pending minor revisions.

Major comments:

1. In re-reading the modified manuscript, one thing that I realized wasn't clear is whether the PINN analyses are carried out in Earth-relative or storm-relative coordinates. In other words, do the PINN wind speed errors contain errors due to a misalignment of the observed versus analyzed storm? I think this should be clarified better in the manuscript, which can potentially also account for some of the wind speed errors observed in the analyses. If necessary, some of the discussion should also be modified to reflect this.

Minor Comments:

L39: Omit "maximum" since intensity already refers to maximum 10-m wind speed

L56-60: Please rephrase this sentence

L67-68: This sentence is pure speculation and needs to be modified or softened, along with supporting references

L75: Please provide the year for Hurricane Ida in parentheses

L77: It is not clear what "This" here refers to, it would be beneficial to clarify here that you're referring to the fact that you have a Nature Run available

L83: Please be more specific by what you mean by "the real Hurricane Ida" as I'm sure you're not recovering all aspects of the storm

L93: PINN -> PINNs

L100: and -> with

L100-101: Suggestion: that were within 10% of the wind speeds observed

L102 and L261: It's not clear what you mean by "planetary scale". In atmospheric sciences, planetary scale typically refers to synoptic and larger scales, but here the application is to the tropical cyclone inner core. This should be reworded.

L114: Can you please define/clarify what you mean by the "inner core"?

L125: Please rephrase "this trick". Clearly, it's not trickery that you're implying here. Perhaps "approach" is more appropriate?

L146: Omit one "in"

L156 and elsewhere: Please clarify or reword "large-scale structure" here. What scale do you mean exactly by this? Do you mean the vortex-scale structure or even larger structures that are relevant for the storm's environment?

Figures 3a and 4a: Reverse the orientation of the z-axis label?

Figure 3f: It would be very useful here to indicate RMW. You can either directly point out to it with a vertical line or replot this figure in RMW-normalized horizontal coordinates. This would enable a direct comparison to Fig. 2d where it was clear that the highest errors were inside the RMW and associated with the wind speed gradient. Is this also the case in the 3D analyses?

L226: It's not clear what you mean by "the core of dropsondes" here.

L228: real-time -> real ?

L232: shield -> SHiELD

REVIEWERS' COMMENTS (responses in red font):

Reviewer #1 (Remarks to the Author):

Although the PINN method is not novel, the application of PINN in hurricane is interesting. The revised paper has answered my technical questions, but I am not a domain expert in hurricane and cannot comment too much on the contribution to hurricane's part.

Reviewer #2 (Remarks to the Author):

The authors have addressed most of my comments, and I am satisfied with the revised manuscript and willing to recommend for publication.

I only have minor comments regarding Figure 4, which presents a comparative analysis between the PINN and the conventional HAFS DA method. I'm wondering if using correlation as an indicator is appropriate for assessing modal performance because, in my opinion, it is difficult to tell which model works better from the wind speed reconstruction. Additionally, it would be valuable to include more instances, considering different times and locations, for a comprehensive analysis of their performance.

Thank you for the feedback, we have included MSE in the text of the Real Case section as an alternative metric. We don't have any other storm cases to compare against at the moment – as such, the primary argument in this section isn't that the PINN is better than the other methods, but rather that it shows promise based on this one example. Future work will include much more thorough analysis on the accuracy of the PINN across multiple cases, and how forecast models are impacted by using the PINN output for initial conditions.

Reviewer #3 (Remarks to the Author):

Review of "Physics-Informed Neural Networks for Hurricane Reconstruction and Data Assimilation " by Eusebim, Vecchi, Lai, and Tong

By Altug Aksoy (University of Miami/CIMAS and NOAA/AOML/HRD)

26 October 2023

Recommendation: Minor revision

Synopsis:

This is the first revised version of the manuscript and I find it much improved. The authors have addressed my concerns regarding the original version, and I only have one major comment along with some minor editorial ones. It shouldn't take the authors too much effort to address these and therefore I recommend that the manuscript be accepted pending minor revisions.

Major comments:

1. In re-reading the modified manuscript, one thing that I realized wasn't clear is whether the PINN analyses are carried out in Earth-relative or storm-relative coordinates. In other words, do the PINN wind speed errors contain errors due to a misalignment of the observed versus analyzed storm? I think this should be clarified better in the manuscript, which can potentially also account for some of the wind speed errors observed in the analyses. If necessary, some of the discussion should also be modified to reflect this.

Thank you for the comment. The SHiELD observations were in storm-centered coordinates (relative to SHiELD storm center) and the real observations are also in storm-centered coordinates (relative to the NHC observed Ida storm center according to the Best Track dataset). Using storm-centered coordinates ensured that the PINN would not suffer from SHiELD data if the SHiELD storm track had not been accurate. We have clarified these coordinate systems in the text.

Minor Comments:

All of the below edits/suggestions have been incorporated, thank you!

L39: Omit "maximum" since intensity already refers to maximum 10-m wind speed

L56-60: Please rephrase this sentence

L67-68: This sentence is pure speculation and needs to be modified or softened, along with supporting references

L75: Please provide the year for Hurricane Ida in parentheses

L77: It is not clear what "This" here refers to, it would be beneficial to clarify here that you're referring to the fact that you have a Nature Run available

L83: Please be more specific by what you mean by "the real Hurricane Ida" as I'm sure you're not recovering all aspects of the storm

L93: PINN -> PINNs

L100: and -> with

L100-101: Suggestion: that were within 10% of the wind speeds observed

L102 and L261: It's not clear what you mean by "planetary scale". In atmospheric sciences, planetary scale typically refers to synoptic and larger scales, but here the application is to the tropical cyclone inner core. This should be reworded.

L114: Can you please define/clarify what you mean by the "inner core"?

L125: Please rephrase "this trick". Clearly, it's not trickery that you're implying here. Perhaps "approach" is more appropriate?

L146: Omit one "in"

L156 and elsewhere: Please clarify or reword "large-scale structure" here. What scale do you mean exactly by this? Do you mean the vortex-scale structure or even larger structures that are relevant for the storm's environment?

Figures 3a and 4a: Reverse the orientation of the z-axis label?

Figure 3f: It would be very useful here to indicate RMW. You can either directly point out to it with a vertical line or replot this figure in RMW-normalized horizontal coordinates. This would enable a direct comparison to Fig. 2d where it was clear that the highest errors were inside the RMW and associated with the wind speed gradient. Is this also the case in the 3D analyses?

L226: It's not clear what you mean by "the core of dropsondes" here.

L228: real-time -> real ?

L232: shield -> SHiELD